



Investigation on uncertainty and sensitivity analysis of complex systems

Yueying Zhu

► To cite this version:

Yueying Zhu. Investigation on uncertainty and sensitivity analysis of complex systems. General Physics [physics.gen-ph]. Le Mans Université, 2017. English. NNT : 2017LEMA1021 . tel-01801799

HAL Id: tel-01801799

<https://theses.hal.science/tel-01801799>

Submitted on 28 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de Doctorat

Yueming ZHU

*Mémoire présenté en vue de l'obtention du
grade de Docteur de Le Mans Université
sous le sceau de l'Université Bretagne Loire*

École doctorale : 3M

Discipline : Physique
Spécialité : Atomes, Molécules et Plasmas
Unité de recherche : IMMM

Soutenue le 23 Octobre 2017
Thèse N° :

Investigation on uncertainty and sensitivity analysis of complex systems

JURY

Rapporteurs :	Jérôme MORIO , Professeur, ONERA Olivier LE MAITRE , Professeur, CNRS, LIMS
Examineurs :	Benoît ROBYN , Professeur, HEI-Yncréa Hauts de France Marina KLEPTSYNA , Professeur, Le Mans Université Michel BASSET , Professeur, Université de Haute Alsace (UHA) Wei LI , Professeur, Central China Normal University
Directeur de Thèse :	Q. Alexandre WANG , Professeur, Ecole Hautes Etudes Ingénieurs et Le Mans Université
Co-directeur de Thèse :	Alain BULO , Professeur CNRS, IMMM- Le Mans Université

Abstract

By means of Taylor series expansion, a general analytic formula is derived to characterise the uncertainty propagation from input variables to the model response, in assuming input independence. By using power-law and exponential functions, it is shown that the widely used approximation considering only the first order contribution of input uncertainty is sufficiently good only when the input uncertainty is negligible or the underlying model is almost linear. The method is then applied to a power grid system and the EOQ model.

The method is also extended to correlated case. With the extended method, it is straightforward to identify the importance of input correlations in the model response. This allows one to determine whether or not the input correlations should be considered in practical applications. Numerical examples suggest the effectiveness and validation of our method for general models, as well as specific ones such as the deterministic HIV model.

Our method is then compared to Sobol's one which is implemented with sampling based strategy. Results show that, compared to our method, it may overvalue the roles of individual input factors but underestimate those of their interaction effects when there are nonlinear coupling terms of input factors. A modification is then introduced, helping understand the difference between our method and Sobol's one.

Finally, a numerical model is designed based on a virtual gambling mechanism, regarding the formation of opinion dynamics. Theoretical analysis is proposed by the use of one-at-a-time method. Sampling-based method provides a global analysis of output uncertainty and sensitivity.

Keywords: Uncertainty analysis, Sensitivity analysis, Variance decomposition, Sampling, Correlation, Sensitivity measure, Opinion dynamics

Résumé

Par un développement en série de Taylor, une relation analytique générale est établie pour calculer la propagation des incertitudes des variables d'entrée sur la réponse du modèle, en assumant l'indépendance des entrées. En utilisant des relations puissances et des relations exponentielles, il est démontré que l'approximation souvent utilisée consistant à ne considérer que la contribution du premier ordre sur l'incertitude d'entrée permet d'évaluer de manière satisfaisante l'incertitude sur la réponse du modèle pourvu que l'incertitude d'entrée soit négligeable ou que le modèle soit presque linéaire. La méthode est appliquée à l'étude d'un réseau de distribution électrique et à un modèle d'ordre économique.

La méthode est étendue aux cas où les variables d'entrée sont corrélées. Avec la méthode généralisée, on peut déterminer si les corrélations d'entrée doivent ou non être considérées pour des applications pratiques. Des exemples numériques montrent l'efficacité et la validation de notre méthode dans l'analyse des modèles tant généraux que spécifiques tels que le modèle déterministe du VIH.

La méthode est ensuite comparée à celle de sobol. Les résultats montrent que la méthode de sobol peut surévaluer l'incidence des divers facteurs, mais sous-estimer ceux de leurs interactions dans le cas d'interactions non linéaires entre les paramètres d'entrée. Une modification est alors introduite, aidant à comprendre la différence entre notre méthode et celle de sobol.

Enfin, un modèle numérique est établi dans le cas d'un jeu virtuel prenant en compte la formation de la dynamique de l'opinion publique. L'analyse théorique à l'aide de la méthode de modification d'un paramètre à la fois. La méthode basée sur l'échantillonnage fournit une analyse globale de l'incertitude et de la sensibilité des observations.

Mots-clés: Analyse d'incertitude, Analyse de sensibilité, Décomposition de variance, Echantillonnage, Corrélation, Mesure de sensibilité, Dynamique d'opinion

Acknowledgements

My Ph.D. study of three years is near its ending. I was continuously supported and encouraged by my supervisors, colleagues, friends, and family members. At this special and important moment, I would like to express my sincere gratitude to all of them.

First and foremost, I would like to show the greatest appreciation to my supervisor, Prof. Q. Alexandre Wang, who provided me an exciting opportunity to study in France, particularly, in ISMANS at the beginning and then at Le Mans University. He guided me to a more international research environment and also offered me a special platform to communicate and discuss with scholars from different fields and countries. It has been a great honor to be one of his Ph.D. students. I deeply appreciate his continuous support, innovative ideas, valuable suggestions and timely guidance, which all are useful and beneficial to my Ph.D. study and related research, and also make my study experience productive and stimulating. The tremendous interests and enthusiasm he has for his research, as well as his immense knowledge, are contagious and motivational for me, even during tough times in the Ph.D. pursuit. I am also grateful for the great help that he and his wife offered for my daily life in France.

I am especially thankful to my supervisor, Prof. Xu Cai, working at Central China Normal University. He introduced me to the research field of Complexity Science and provided me many chances to attend international academic conferences. He encouraged me to persistently move forward and to work independently. As a mentor, except for the knowledge related to research, he also taught me how to positively face setbacks in study and life, how to communicate with others, how to build good personality, how to "conduct oneself", etc. His encouragement and unlimited support incited me to do better both in study and life.

My sincere thanks also go to Prof. Wei Li, my another supervisor working at Central China Normal University. He guided me to do numerical simulations and theoretical analysis. He taught me how to write a research paper, how to reply the referee's report, etc. I am particularly indebted to him for his constant encouragement and endless support throughout the whole stage of my postgraduate study. His insightful comments and valuable ideas have particularly contributed to the writing of this thesis.

I would like to thank my associate supervisor, Prof. Alain Bulou, for his constant encouragement and support when I studied at IMMM. He provided me financial support to attend a particular international academic conference, and permission to use facilities of the laboratory. He is very gentle and always offers help to me when I am in troubles.

I would also like to thank the rest members of my thesis committee: Prof. Olivier Le Maître, Prof. Jérôme Morio, Prof. Benoît Robyns, Prof. Marina Kleptsyna, and Prof. Michel Basset, for their insightful comments and hard questions which stimulate me to extend my research from various perspectives.

I am hugely appreciative to Prof. Liping Chi from Central China Normal University. She introduced me to the interesting research field of public opinion dynamics at the beginning of my master study. She is more like a sister who always encourages and cares for me. Special mention should also go to her generous assistance in polishing the thesis.

Profound gratitude goes to Prof. Chunbin Yang at Central China Normal University for his helpful discussions about theoretical analysis. Gratitude also goes to Profs. Wenping Bi and Paul Bourguine, the CST external scientific members, for their insightful reports and comments about my study progress. Special thanks should be expressed to Prof. Jean-Marc Greneche, for his support, encouragement, and warm assistance both in academic research and daily life, to Prof. Florent Calvayrac, for his consideration to my study and valuable suggestions about providing an oral presentation, and to Francis Chavanon, for his kind help in using the server machine of the laboratory.

I would especially like to thank Saïda Ménard. It is quite hard to study alone in a foreign country. As my best friend in France, she consorted with me for numerous moments of helplessness and loneliness. I am deeply indebted to her for the persistent support, encouragement, contribution of time, and anything else she did for me in daily life. I would also like to thank Michèle Houdusse, my reception family in France, for her kind help and consideration during my stay here.

Sincere thanks also go to my colleagues and other friends in Le Mans: Anna Kharlamova, Fan Wang, Lei Xiong, Liyang Zheng, Lizhong Zhao, Ru Wang, Van Tang Nguyen and Wenting Guo, for all the fun and pleasure we have had within the last three years; and to all the previous and current members in Complexity Science Center of Central China Normal University, for their constant support, warm assistance, and productive discussions. In particular, I am grateful to Dr. Jian Jiang for his valuable comments in the writing of the thesis.

I gratefully acknowledge the financial support that enabled me to finish smoothly my Ph.D. study in France. I was funded by China Scholarship Council. My work was also partially supported by the Programme of Introducing Talents of Discipline to Universities of China, the National Natural Science Foundation of China, the Ecole Doctorale 3M at Le Mans University, and the Institut des Molécules et Matériaux du Mans (IMMM).

Last but not the least, I would like to take this opportunity to express my deepest appreciation to my parents, parents-in-law, brother, sister-in-law, nephew and niece, for

their selfless love in my life, continuous support and encouragement in all my pursuits. And most of all, I wish to thank my husband, for his faithful love, constant support, unconditional pay, trust, and understanding during the stage of my Ph.D. study in France.

Le Mans, France

October, 2017

Abbreviations

AIDS	A cquired I mmune D eficiency S yndrome
CC	Pearson C orrelation C oefficient
CCDF	C omplementary C umulative D istribution F unction
CDF	C umulative D istribution F unction
CL_2	C entered L_2 -discrepancy
EOQ	E conomic O rders Q uantity
HIV	H uman I mmunodeficiency V irus
IFFD	I terated F ractional F actorial D esign
KRCC	K endall R ank C orrelation C oefficient
LHS	L atin H ypercube S ampling
MC	M onte C arlo
PCC	P artial C orrelation C oefficient
PDF	P robability D ensity F unction
PRCC	P artial R ank C orrelation C oefficient
QMC	Q uasi M onte C arlo
RCC	R ank C orrelation C oefficient
SIR	S usceptible- I nfectious- R ecovered
SIS	S usceptible- I nfectious- S usceptible

Contents

Abstract	i
Résumé	ii
Acknowledgements	iii
Abbreviations	vi
General introduction	1
1 Introduction	4
1.1 Characterisation of uncertainty in model input	6
1.2 Presentation of uncertainty in model output	8
1.3 Methods of sensitivity analysis	11
1.3.1 One-at-a-time method	12
1.3.2 Regression analysis method	14
1.3.3 Response surface method	16
1.3.4 Differential-based method	20
1.3.5 Variance-based methods	22
1.3.6 Moment independent method	25
1.3.7 Sampling-based method	26
1.4 Determination of analysis results	35
1.4.1 Scatter plots	35
1.4.2 Correlation measures	39
1.4.3 Sensitivity indices	43
2 The analytic analysis for models with independent input variables	48
2.1 Taylor series	48
2.2 Variance propagation for univariate case	49
2.2.1 Uniform distribution	51
2.2.2 Normal distribution	53
2.3 Generalisation of the analytic formula	55
2.4 Applications to the analysis of complex systems	56
2.4.1 Power grid system	57

2.4.2	Economic system	61
3	The analytic analysis for models with correlated input variables	64
3.1	Variance propagation	64
3.2	Estimation of sensitivity indices	66
3.2.1	Generation of correlated variables	67
3.2.2	Sensitivity indices	69
3.3	Numerical examples and a practical application	72
3.3.1	Additive linear model	72
3.3.2	Nonlinear models	73
4	The establishment of sampling-based strategy	83
4.1	A comparison of our method with Sobol's one	83
4.2	Analysis of SIR and SIS models	86
5	Opinion formation based on a gambling mechanism and its sensitivity analysis	91
5.1	Introduction	91
5.2	Modeling	93
5.3	Methods	95
5.4	Results and discussion	96
5.4.1	Theoretical analysis for special cases	97
5.4.2	The distribution of Ω_s in general situations	103
5.4.3	Opinion Clustering	106
5.4.4	The fraction of winners	107
5.4.5	Uncertainty and Sensitivity analysis	109
6	Conclusions and Prospectives	111
6.1	Conclusions	111
6.2	Future work plan	113
A	Central moments	115
A.1	Uniform distribution	115
A.2	Normal distribution	116
B	Generation of correlated variables	117
B.1	Two correlated variables	117
B.2	Three correlated variables	118
B.3	Four correlated variables	119
C	Derivation of Eqs. (3.51-3.52)	121
C.1	First-order contributions	121
C.2	Second-order contributions	123
C.3	Third-order contributions	124
D	Derivation of the fourth-order central moment with four different variables	126

E	Partial variance contributions for Ishigami function	128
E.1	First-order contributions	128
E.2	Second-order contributions	130
F	Special case for opinion formation: $p = \gamma = 0$	134
	Bibliography	139
	Publications	160

General introduction

The concept of global uncertainty and sensitivity analysis has proposed for a long time. It attracts the considerations of a large number of researchers from various of areas. The global uncertainty and sensitivity analysis aims at analysing the uncertainties of output variables (also called observations or model responses) according to the uncertainties in input variables (or named factors, parameters, covariates), and the sensitivity of each output variable with respect to individual input parameters, as well as to their interactions. Undoubtedly, global uncertainty and sensitivity analysis is advantageous for gaining insight into how input variables should be ranked according to their importance in establishing the uncertainties in different output variables. While the strategies for uncertainty and sensitivity analysis are quite extensive, a general analytic method is still limited, especially for models of present input correlations. In this dissertation, we mainly focus on the establishment of a general theoretical framework for global uncertainty and sensitivity analysis in the modelling of complex systems.

Mathematical models have a wide range of applications in diverse disciplines. They can help explain a system by visualized data and/or figures and analyse the possible effects of different parameters, and also, if necessary, make predictions about the underlying behaviour. With a deterministic mathematical model of general form $\mathbf{y} = f(\mathbf{x})$ with \mathbf{y} denoting the output vector and \mathbf{x} indicating the input vector. When \mathbf{y} is calculated from \mathbf{x} through a specified function or some natural or artificial rules connecting \mathbf{y} and \mathbf{x} , uncertainties in the elements of input vector, if exist, will propagate through the calculation to the members of output vector \mathbf{y} dependent on \mathbf{x} [1, 2]. This process is called variance propagation (or uncertainty propagation). Variance propagation, often regarded as the fundamental ingredient of sensitivity analysis for complex models, mainly considers the determination of output's variance via uncertainties in input factors [3, 4].

At present, many strategies have been built for the determination of variance propagation, such as simulation-based methods[5, 6], most probable point-based methods[7, 8], functional expansion-based methods[9], numerical integration-based methods[10–13]. Simulation-based methods, also called sampling-based methods, are regarded as both

effective and widely used, especially for those models with the functional relationship connecting \mathbf{y} and \mathbf{x} absent [14–16]. These briefly mentioned strategies, however, are computationally expensive, especially in the presence of a high number of input variables. For a general model with given functional form, the procedure will be much easier and numerically cheaper for determining the output’s variance if an analytic formula associated with variance propagation can be provided. More information associated with other methods for variance propagation can be found in some reviewed papers [17–19].

A simple analytic formula has been appeared since 1953. It approximately computes the variance of the product of two independent random variables [20]. In 1966, this approximation was extended by engineers and experimentalists to more general multivariate cases [21]. This formula, also called Taylor series approximation, restricted to the first-order terms [22], has gained a wide applications thanks to its simplicity and convenience [23]. However, it can satisfactorily estimate the output’s variance only when the functional relationship between output and input variables is almost linear or the uncertainty of each input variable is negligible [17]. For most models, y highly nonlinearly depends on \mathbf{x} having large uncertainties. This suggests the necessity of an exact analytic formula in calculating the output’s variance and evaluating its sensitivities with respect to individual input factors, as well as to their interactions.

Furthermore, many methods have also been designed for performing sensitivity analysis, including the traditional approach of changing one factor at a time [24, 25], local method [26], regression analysis [27], variance-based method [28], etc. Among the various available strategies, variance-based sensitivity analysis has been assessed as versatile and effective for uncertainty and sensitivity analysis of model response. The consideration of variance-based importance measures can be traced back to over twenty years ago when Sobol characterised the first-order sensitivity measures on the basis of deposing the variance in model response into different partial contributions attributable to individual input variables and to their combinations (called variance decomposition) [29]. Then extensive relevant investigations are carried out around this Sobol’s work, boiling down to the improvements in analysis strategies and to their applications to the sensitivity and reliability analysis of complex systems [30, 31]. However, these frameworks, as well as above mentioned strategies for the determination of variance propagation, are often proposed when the input variables are assumed to be statistically independent.

Recently, the interest in extending sensitivity analysis strategies from uncorrelated case to the correlated one is increasing as correlated factors are often happened in practical applications. Previous investigations about sensitivity analysis of models in the presence of input correlations only provided overall sensitivity indices with respect to individual input factors. However, the correlated and independent variance contributions were

absent [32]. In practical applications, the distinction between independent and correlated contributions is quite important. It allows one to decide whether or not the correlations among input factors should be considered.

Both correlated and independent variance contributions were firstly considered by C. Xu *et al* [33]. They proposed a regression-based strategy to decompose partial variance contributions into independent and correlated parts by assuming approximate linear dependence between model response and input variables. To overcome the limitation of their method, many frameworks on sensitivity analysis are recently developed in the presence of input correlations, contraposing the investigation of more effective and universal technics for sensitivity analysis in general correlated situations [34–36]. Still, a theoretical framework of the determination of partial variance contributions and of relative correlated and independent effects is limited, especially when a single input is correlated with many others simultaneously.

Consequently, in this dissertation, we mainly focus on the establishment of a theoretical framework for uncertainty and sensitivity analysis. The applications of sampling-based method are also proposed to the uncertainty and sensitivity analysis of epidemic spreading and opinion formation systems.

The manuscript is organised as follows. The first chapter introduces in detail the background of uncertainty and sensitivity analysis of complex systems. It also provides the implementation of uncertainty and sensitivity analysis by using different strategies. In the second chapter, a systematic theoretical framework is established for the uncertainty and sensitivity analysis of general models with given functional forms, by assuming that input factors are statistically independent of each other. In the third chapter, the theoretical method for uncertainty and sensitivity analysis is generalised to more universal models of input correlations. The fourth chapter concerns the difference of our method from Sobol’s one. A rough sampling-based approach that is coincident with our analytic method is then established by introducing a modification to the Sobol’s method, in assuming input independence. A systematic framework on uncertainty and sensitivity analysis of a numerical model is described in chapter 5. This model considers the formation of public opinion dynamics based on a virtual gambling mechanism. Finally, a general conclusion and future work plan are given.

Chapter 1

Introduction

Mathematical models are of great importance in the natural sciences. They have been diffusely utilized in many disciplines as diverse as mathematics [37, 38], physics [24, 25, 39–42], chemistry [43], etc. With mathematical models, one can explicate a system in mathematical language, analyse the roles of linked factors by physical methods, and then make reasonable predictions of underlying behaviors. In general a model contains three major elements: the input vector, the output vector, and associations between them. In practical applications, the elements of input vectors are rarely deterministic but contain uncertainty following some distribution laws [44, 45]. Consequently, the determination of the variations in input variables, the investigation of their propagating through the model, as well as the quantification of the sensitivities of model outputs with respect to input variables are of crucial importance for establishing reliable and robust models [3, 33, 46, 47]. The implementation of these procedures is known as uncertainty and sensitivity analysis.

A view of modeling that may help illustrate the role of uncertainty and sensitivity analysis in the scientific process is offered in Fig. 1.1, taken from the work of Robert Rosen, an American theoretical biologist [48]. The figure shows two systems, a natural system \mathbf{N} which forms the subject under investigation, and a formal system \mathbf{F} which indicates the modeling of this subject. Each system has its own internal entailment structures and the two systems are connected by the encoding and decoding processes.

The uncertainty under discussion here is often referred to as epistemic uncertainty (also known as systematic uncertainty). Epistemic uncertainty derives from a lack of information or non-accuracy in measurement about the appropriate value used for specifying a quantity that is assumed to be constant in the context of the analysis for a particular problem. In the conceptual and computational designation of an analysis, epistemic uncertainty is regarded in general to be distinct from aleatory uncertainty, which, also

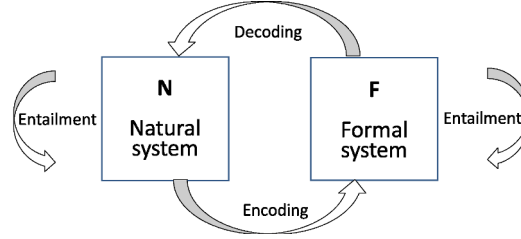


FIGURE 1.1: Modeling after Rosen (1991)

known as statistical uncertainty, arises from an inherent randomness in the behavior of the system under study.

Uncertainty and sensitivity analysis are essential parts of analyses for complex systems. Specifically, sensitivity analysis considers the determination of variance contributions of individual input variables to the elements of output vectors [49]. Uncertainty analysis, preceding sensitivity analysis, mainly focuses on the determination of the uncertainties in output variables that derive from uncertainties in input factors. Conceptually, uncertainty and sensitivity analyses should be run in tandem. They work together to help determine:

- (1) Which input factors contribute most to the variation of model output.
- (2) Which parameters are significant and which ones can be eliminated from the model;
- (3) How to efficiently reduce the uncertainty in model output by strengthening the knowledge base concerning input parameters.

Quantifying the impact of a variable under sensitivity analysis could be useful for a series of purposes, such as deep understanding of the relationships between input and output variables in a system or model, fixing model inputs that have no effect on the output, identifying and removing redundant parts of the model structure, and avoiding useless time consumption on non-sensitive variables in models of a large number of parameters. In models consisting of a large number of input variables, sensitivity analysis constitutes essential ingredient of model building and quality assurance. Sensitivity analysis has also extended its application to national and international agencies involved in impact assessment studies, including the European Commission [50, 51], Australian pathology laboratories [52], the Intergovernmental Panel on Climate Change [53], and US Environmental Protection Agency's modelling guidelines [54].

The framework of uncertainty and sensitivity analysis is easily performed when only a single input factor is involved in the model under discussion, which is known as univariate situation. It requires a straightforward one-dimensional analysis by presenting results in figures in a two-dimensional space. When two or more input factors are under

assessment, however, the problem is much more complicated, especially if input factors do not have a separable monotonic effect on the output variable of interest.

The early framework of sensitivity analysis for multivariate models was established by using local analysis. Local sensitivity analysis aims at assessing the local effects of uncertainties in individual input factors on output variables, by concentrating on the sensitivity in vicinity of a set of special factor values [55]. Such sensitivity is usually evaluated by the use of gradients or partial derivatives of functions connecting output and input vectors at these special factor values. This means the values of the rest factors are fixed while studying the local sensitivity of model response with respect to a single factor. Local sensitivity analysis is most frequently employed for the analysis of complex models, especially when a large number of input factors are involved. This is common because of the simplicity and low computational cost in its implementation. However, it abortively quantifies the global impact of individual input factors and of their interactions on output variables. Of importance to a part of model analysis practitioners (mostly working in the fields of statistics, risk and safety assessment, and reliability detection) is understanding the sensitivity of an output variable with respect to simultaneous variations of several input factors [47, 56]. The global uncertainty and sensitivity analysis provides such sensitivity information. It evaluates the influence of individual input factors by looking at the entire input space rather than at a specified point.

Generally, the process of global uncertainty and sensitivity analysis can be decomposed into: (a) specifying the model under study and defining its input and output variables; (b) characterising the uncertainty in input variables; (c) determining the uncertainty in model output; and (d) quantifying the importance of individual input variables in the estimation of output variables. In establishing the framework of sensitivity analysis for a given model with defined input and output variables, the main goal is to handle the remaining procedures.

1.1 Characterisation of uncertainty in model input

Quite often, some or all of the model inputs are subject to sources of uncertainty, including errors of measurement in experiments, absence of information and poor or partial understanding of the driving forces and mechanisms. The most essential practice in uncertainty and sensitivity analysis is to characterise the uncertainty in input variables. The definition of model input, however, depends upon the particular model under investigation.

A model can be stated as diagnostic or prognostic. Diagnostic models are used for understanding a law. They are often built by wild speculations applied to play what-if games, such as models designed to study the emergence of an agreement in a population [57–59], models used to investigate organizational change, etc. In the investigation of organizational change, as an example, three diagnostic models are high potential candidates to highlight the problem areas and provide structure for solution development. The first one is an analytic model, also known as the difference-integration model. It focuses on thorough analytical diagnosis as the foundation for organisational change [60]. The second one is the force-field analysis model, originally developed by Kurt Lewin in the early 1950s. It regards the organisation as the result of internal forces that drive change or maintain the current status [61]. The third one is developed on the bases of cause maps and social network analysis. It provides a mathematical approach to organisation diagnosis [62]. Regarding diagnostic models, input variables are pre-defined by model designers and often assumed to follow particular distribution laws in fixed real ranges (e.g. uniform and Gaussian distributions). Prognostic models can be viewed as accurate and trusted predictors of a system. They mainly focus on the estimation (prediction) of the probability that a particular event or outcome will happen. Prognostic models are often developed for the clinical practice, where the risk of disease development or disease outcome (e.g. recovery from a specific disease) can be calculated for individuals by combining information across patients. In the clinical practice, prognostic models can be presented in the form of a clinical prediction rule [63–66]. Prognostic models also find their applications in other fields, such as risk and safety evaluation in engineering [67, 68], problems solving of water dynamics in estuaries [69]. It is often preferable that input variables in prognostic models, in contrast to those in diagnostic models, are easily determined by social experience or practical examples for ensuring the applicability of a prognostic model in practical applications.

Furthermore, models can also be classified as data-driven or law-driven. A data-driven (or inverse) model tries to derive properties statistically by empirical study. Advocates of data-driven models like to describe social behaviours with a minimum of adjustable parameters, for instance, models helping understand the spreading of really happened epidemics [70–73], models designed for explaining the generation of traffic jams [74, 75], and models proposed to describe financial time series [76, 77]. Law-driven (or forward) models, on the other hand, aim at employing appropriate laws which have been attributed to the system to predict its behaviour. For example, people can use Darcy’s and Fick’s laws to understand the motion of a solute in water flowing through a porous medium [78, 79]. In building design, as another example, building energy simulation models are generally classified as prognostic law-driven models by which the behaviour of a complex system can be predicted in terms of a set of well-defined laws

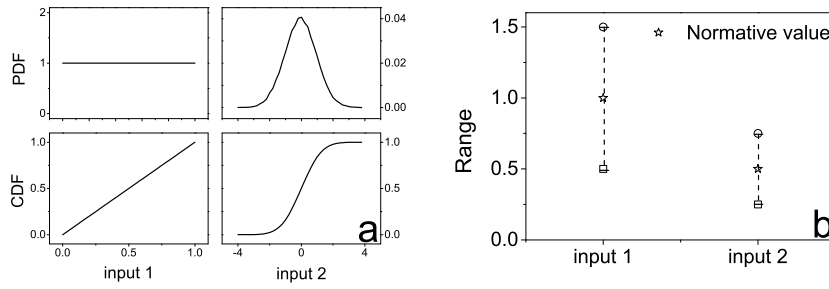


FIGURE 1.2: Uncertainty characterisation of independent input variables. (a): an example of given distribution laws of input variables: input 1 is uniformly distributed in the real range $[0, 1]$; input 2 follows the standard normal distribution ($\mu = 0, \sigma = 1$). (b): an example of deterministic input variables: input 1 is specified at 1; input 2 is fixed at 0.5. Their uncertainties are represented by changing 50% around their normative values.

(e.g., mass balance, energy balance, conductivity, and heat transfer, etc.) [80]. In data-driven models, input parameters are introduced based on special situations, which could be deterministic and attached with an artificially defined uncertainty. For law-driven models, input factors are often imported by the laws we employed. Characterising the uncertainty in input factors is also dependent upon the situation under analysis.

With given distribution laws, the uncertainty in input variables can be specified by their mean values (e.g., arithmetic mean or mathematical expectation, geometric mean, median), standard deviations, PDFs, CDFs and CCDFs. Particularly, when input variables are deterministic (often happening in agent-based systems where input parameters can be determined based on practical experience), their uncertainties are frequently represented by artificially introducing fixed variation around their normative values or few typical scenarios (e.g., scenarios corresponding to any possible combinations of specified low, medium, high values of input factors) for performing the uncertainty and sensitivity analysis of the system under discussion [16, 81–84]. The analysis framework of deterministic situations is often designed according to the variation in model output driven from the independent variation in each input factor. This method is known as one-at-a-time method and will be discussed below in detail. Examples, as shown in Fig. 1.2, present the characterisation of uncertainties in input factors for both kinds of situations, in the absence of input correlations.

1.2 Presentation of uncertainty in model output

As already mentioned at the beginning of this chapter, uncertainty and sensitivity analysis should be run in series, with uncertainty analysis preceding in current practice. Uncertainty analysis is, through a certain way, to determine the uncertainty in model output based on the uncertainty in model input.

One popular way of establishing uncertainty analysis is dependent upon the computer, which is also known as Monte Carlo (MC) method. Consider a general model of the form $y = f(\mathbf{x})$ with $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ indicating an input vector of n -dimensional variables. All elements of input vector are assumed to be independent of each other. By given PDFs of individual input factors, a sample of size M , indicated by an $M \times n$ matrix, can be generated as

$$\begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^{n-1} & x_1^n \\ x_2^1 & x_2^2 & \cdots & x_2^{n-1} & x_2^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{M-1}^1 & x_{M-1}^2 & \cdots & x_{M-1}^{n-1} & x_{M-1}^n \\ x_M^1 & x_M^2 & \cdots & x_M^{n-1} & x_M^n \end{bmatrix}. \quad (1.1)$$

Run independently the model for all points that are sampled in n -dimensional input space. A set of values of the model output y are then generated accordingly:

$$y = (y_1, y_2, \dots, y_M)^T. \quad (1.2)$$

It is straightforward to state the uncertainty in output y according to its values presented in Eq. (1.2).

Presentation formats of the uncertainty in model output include mathematical expectation, standard deviation, the percentiles of its distribution, confidence bounds, PDF, CDF, CCDF and box plot [85–89]. In general, the last four presentation patterns are usually preferable to the first several indices which will make large amount of uncertainty information neglected in implementation. Furthermore, box plot is definitely beneficial for displaying the uncertainty in model output with normative input factors and comparing the uncertainties in a number of related variables. The box plot is a standardised way of displaying the distribution of data. It is often generated by a box and whisker plots. The bottom and top of the box are, in general, the first and third quartiles of all of the data. The band inside the box is always the second quartile (the median). The ends of the whiskers can represent several possible alternative values including: the minimum and maximum of all of the data, the 9th percentile and the 91st percentile, the 2nd percentile and the 98th percentile [90, 91]. Figure 1.3(a) exhibits an example of uncertainty analysis of a simple model with functional form given by

$$y = x_1^2 + x_2^2. \quad (1.3)$$

The uncertainties in input factors are defined by Fig. 1.2(a), that is, x_1 follows a uniform distribution in the real range $[0, 1]$ and x_2 the standard normal distribution. Box plot, as

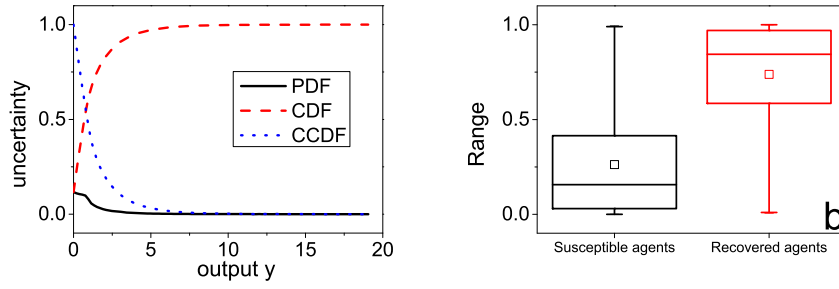


FIGURE 1.3: Representation of uncertainty in model output. (a): uncertainty analysis for the model presented in Eq. (1.3) whose input uncertainty is defined by Fig. 1.2(a). (b): box plot for the normalized susceptible and recovered agents at equilibrium state of SIR model with input factors assumed to be uniformly distributed between 0 and 1. Bars show the full range of the ensemble distribution of values; boxes show the range encompassed by the 25th and 75th percentiles; the horizontal line and square within each box show the median and mean, respectively.



FIGURE 1.4: Progression of population for SIR model.

another example, for the equilibrium state of SIR model is presented in Fig. 1.3(b) where the normalized susceptible and recovered agents are analysed. Three input factors: $s(0)$ (initial proportion of susceptible agents), γ (recovered rate) and β (infectious probability) are assumed to be uniformly distributed between 0 and 1 [92, 93]. SIR model is one of the compartmental models in epidemiology, serving as a base mathematical framework for understanding the complex dynamics of the disease spread. The model consists of three compartments: susceptible agents, infectious agents, and recovered (or immune) agents. Each member of the population typically progresses from susceptible to infectious to recovered, as shown in Fig. 1.4.

In practical applications, the model output is not always a scalar but could also be a function. For example, in the investigation of epidemic spreading, the system of interest is time-dependent. Uncertainty in input factors will be propagated to the uncertainty in the dependence of model output upon the time parameter. For such situations, an effective presentation format of the uncertainty in model output is to use two graphical frames, with first one displaying any possible dependence of model output upon a relevant parameter and second one presenting statistical results for the outcomes in the first one [30]. Figure 1.5 displays the uncertainty analysis for the normalized infected agents in SIR model. Three input factors of SIR model: $s(0)$ (initial proportion of susceptible agents), γ (recovered rate) and β (infectious probability) are assumed to be uniformly distributed between 0 and 1. Having performed uncertainty analysis we can then move on to the sensitivity analysis. Sensitivity analysis allows one to understand how uncertainty in the model output can be attributed to different sources of uncertainties in

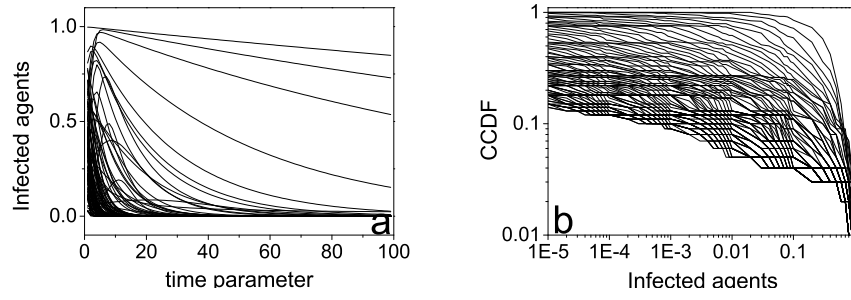


FIGURE 1.5: Representation of uncertainty in model output which acts a function of time parameter. Analysis of the time-dependent normalized number of infected agents generated in SIR model is presented as a particular example. Three involved input factors are assumed to be uniformly distributed between 0 and 1.

input factors.

1.3 Methods of sensitivity analysis

Sensitivity analysis is a primary part of model development. It involves importance evaluation of input parameters in the estimation of the model output. At present, a large number of approaches have been built for performing the sensitivity analysis. Regarding the complexity of models, many methods are developed to address one or more constrains. For example, most common sensitivity analysis methods assume input factors are independent of each other [30, 94]; approaches based on linear regression are valid only for linear models; virtually all sensitivity analysis methods consider a single univariate model output, by which sensitivity measures are hard to be interpreted for models with correlated outputs.

Methods of sensitivity analysis can be classified based on the methodology as mathematical, statistical or graphical [95]. Mathematical methods evaluate the sensitivity of model output with respect to the range of variation of each input factor. Typically, they involve the calculation of output variable according to a few values of each factor that represent the possible variation range of the factor. Mathematical methods can identify the influences of individual factors in their variation ranges on an output variable. However, they do not indicate the variance of output variable propagated from the uncertainties in input factors but represent, for example, the sensitivity of model output as the magnitude of percentage change compared to its nominal value. In some cases, especially when input factors are almost deterministic, mathematical methods are helpful in recognising the most important factors [81, 96]. One-at-a-time method (discussed below) is one of most widely used mathematical methods. Statistical methods assess the variance contribution of input factors to the output variable with given

probability distributions related to input factors. By employing statistical methods, the variation of one or more input factors can be considered simultaneously. This allows one to identify the interaction effects among multiple input factors on the uncertainty in output variable. Some statistical methods often considered are introduced here, including differential-based method, response surface method, regression analysis, variance-based analysis, and sampling-based method. Graphical methods, in general, mainly focus on the representation of sensitivity analysis results in the form of graphs, charts, or surfaces. They provide a visual understanding of how an output variable is affected by the variation in input factors. Graphical methods can be used to complement the analysis results of mathematical and statistical methods in establishing systematical framework of sensitivity analysis. Some other classifications of sensitivity analysis methods may focus on the capability of a specific technique, which aids in understanding the applicability of a specific method to a particular model and analysis objective [97].

1.3.1 One-at-a-time method

One-at-a-time method for sensitivity analysis, also known as nominal range sensitivity analysis, local sensitivity analysis or threshold sensitivity analysis, is one of the simplest and most widely used approaches. This method is individually varying only one of the model inputs across its entire range of plausible values at a time while fixing the others at their base-case or mean values, to see what effect this exerts on the model output. The sensitivity of the model output with respect to a particular input variable can be identified by the difference in the model output contributed by the variation of the variable. Regarding deterministic models, sensitivity measures can be typically represented as a positive or negative percentage change of the output variable compared to its normative value.

For linear models, one-at-a-time analysis is advantageous for recognising the most important factors because of its simplicity and low computational cost in implementation. This approach, however, does not consider the simultaneous variation of input variables. This makes it fail in identifying the impacts of interaction effects among multiple input factors on the uncertainty in an output variable. Accordingly, the analysis results of one-at-a-time strategy are potentially misleading for models other than linear ones.

One-at-a-time method is quite often used for sensitivity analysis of models with a large number of input variables (e.g., climate models, contagious disease spreading models) since it is easily operated, by only repeating sensitivity analysis process for any number of individual model inputs [81, 98–100]. Murphy *et al.*, for example, applied the method to address the range of climate changes resulted from the variation of input factors [101].

With one-at-a-time strategy, they obtained statistical predictions of climate prediction and sensitivity indices in assuming that the effects of perturbations of 32 input parameters combine linearly and independently. In a deterministic HIV model with 20 uncertain input parameters involved and supposed to be independent of each other, Blower and Dowlatabadi employed PRCC to assess the statistical relationship between each input parameter and the outcome variable while fixing the remaining input factors at their nominal values [3].

Regarding one-at-a-time method for sensitivity analysis, one often considers the Morris method [102]. Morris method is stated as effective to screen a few important input factors from a large number controlling a model. In this method, the input space (for simplicity, defined as an n -dimensional unit hypercube) is discretized to n -dimensional p -level grid. Each input x_i may take values from a sequence $\{0, 1/(p-1), 2/(p-1), \dots, 1\}$. For a given value of input vector \mathbf{x} , the elementary effect of the i th input factor is defined based on Morris method as

$$d_i(\mathbf{x}) = \frac{f(\mathbf{x} + \mathbf{e}_i \Delta) - f(\mathbf{x})}{\Delta} = \frac{\partial f}{\partial x_i}, \quad (1.4)$$

where Δ is a predetermined multiple of $1/(p-1)$, and \mathbf{e}_i a column vector where the i th entry is set to 1 and the rest ones are set to zero. The finite distribution of elementary effects is estimated by randomly sampling different \mathbf{x} from input space. For input x_i , the distribution is denoted by F_i . The mean value u_i and standard deviation v_i of the distribution F_i are then estimated. u_i characterises the effect of input x_i on the model output and v_i the nonlinear effect of x_i as well as the interaction effects associated with x_i :

$$u_i = \int \frac{\partial f}{\partial x_i} d\mathbf{x}, \quad v_i = \left[\int \left(\frac{\partial f}{\partial x_i} - u_i \right)^2 d\mathbf{x} \right]^{1/2}. \quad (1.5)$$

When the model under discussion is non-monotonic, distribution F_i contains positive and negative elements. Averaging rule may cancel some effects so as to make u very small or even zero. For this reason, an improved sensitivity measure is considered by Campolongo *et al*, called u^* , which is defined as the mean of the distribution of the absolute values of the elementary effects [103, 104]:

$$u^* = \int \left| \frac{\partial f}{\partial x_i} \right| d\mathbf{x}. \quad (1.6)$$

Measure u^* can help identify out the input factors of important overall influence on the model output.

1.3.2 Regression analysis method

Regression analysis is a statistical process for providing an algebraic representation of the relationship between output variable and one or more of input parameters [105]. It allows one to understand how the output variable changes when any one of the input factors is varied while the remaining factors are fixed.

The earliest form of the regression was the method of least squares, which was considered by Legendre in 1805 [106] and also by Gauss in 1809 [107]. However, the term "regression" was proposed by Francis Galton in the late of nineteenth century, with which, a biological system was described [108]. It was later extended by Udny Yule and Karl Pearson to a more general statistical context [109].

In the context of sensitivity analysis, regression analysis usually involves the construction of linear relationship connecting output variable and input parameters. The standardized regression coefficients are then directly used for assessing the sensitivity of model output with respect to individual input factors. Regression analysis contains three groups of variables: the unknown regression coefficients, denoted as b_i with $i = 0, 1, \dots, n$, the input factors x_i with $i = 1, 2, \dots, n$, and the model output y . y could be a vector. But for simplicity, we consider y as a scalar.

Regression analysis is most properly performed by independent random samples which constitute the mapping from input factors to the output variable. By linear regression, the model under study is approximated as

$$\hat{y} = b_0 + \sum_{i=1}^n b_i x_i, \quad (1.7)$$

where b_i is the regression coefficient for input x_i , which can be interpreted as the change in output y when the input factor x_i increases by one unit in keeping the remaining factors constant [110]; \hat{y} denotes the predicted value of output variable for a given point in the n -dimensional input space when regression coefficients are determined. The coefficients b_i are determined by least squares: minimizing the sum of squares of deviation from the true values:

$$\sum_{j=1}^M (y_j - \hat{y}_j)^2 = \sum_{j=1}^M \left[y_j - \left(b_0 + \sum_{i=1}^n b_i x_j^i \right) \right]^2, \quad (1.8)$$

where M is the number of samples (experimental points), y_j the j th output data point given by the j th n -dimensional input data point, x_j^i the j th sampled value of input x_i [111]. The deviation of the prediction of the regression model (Eq. (1.7)) from the exact

values given by the original model can be evaluated as

$$R^2 = \sum_{j=1}^M (\hat{y}_j - \bar{y})^2 \bigg/ \sum_{j=1}^M (y_j - \bar{y})^2, \quad (1.9)$$

which provides a measure of the amount of uncertainty in output variable explained by linear regression model [112, 113]. Particularly, $R^2 \rightarrow 1$ indicates that the developed regression model accounts for most of the uncertainty in output variable. Conversely, $R^2 \rightarrow 0$ means that the regression model is not satisfied in explaining the output uncertainty [30].

To some degree, the regression coefficients can reflect the sensitivity of model output to input factors. If a coefficient b_i is close to 0, then there is not a statistically significant linear relationship between input x_i and the output y . Conversely, if b_i is significantly different from 0, then the output y can be regarded as being sensitive to x_i . However, b_i is influenced by the units of x_i . To reduce the dimensional effects of input factors, the regression model represented by Eq. (1.7) is commonly standardised to make the variance of output and input variables equal to 1:

$$(\hat{y} - \bar{y})/\hat{\sigma} = \sum_{i=1}^n (b_i \hat{\sigma}_i / \hat{\sigma})(x_i - \mu_i) \hat{\sigma}_i, \quad (1.10)$$

where

$$\hat{\sigma} = \left[\frac{1}{M} \sum_{j=1}^M (y_j - \bar{y})^2 \right]^{1/2}, \quad (1.11)$$

$$\hat{\sigma}_i = \left[\frac{1}{M} \sum_{j=1}^M (x_j^i - \mu_i)^2 \right]^{1/2}, \quad (1.12)$$

and

$$\bar{y} = \frac{1}{M} \sum_{j=1}^M y_j, \quad (1.13)$$

$$\mu_i = \frac{1}{M} \sum_{j=1}^M x_j^i. \quad (1.14)$$

The coefficients $b_i \hat{\sigma}_i / \hat{\sigma}$ are referred to as standardised coefficients, taking values between -1 and 1. The standardised coefficients are helpful in identifying which of input parameters have greater effects on the output variable when the input variables are measured in different units of measurement [114]. Linear regression analysis is most suitable when

the model of interest is in fact linear as it is difficult to interpret the standardised coefficients when nonlinear regression analysis is involved. In some analyses, nonlinear regression provides an alternative to linear regression for more accurate estimation of the relationship between output and input variables [115, 116].

Because of simplicity and low computational cost, regression analysis as a strategy of sensitivity analysis has been adopted by many researchers from various of fields, such as medical science [117, 118], bioscience [119], human science [120], and food science [121, 122].

1.3.3 Response surface method

Response surface method consists of a group of mathematical and statistical techniques used in the development of an adequate model function connecting an output variable and a number of input parameters. With the established functional relationship, response surface method can identify curvature in the response surface by accounting for high-order effects produced by input parameters. The method was introduced by Box and Wilson in 1951 [123]. The main idea of this approach is to use a sequence of designed experiments to obtain an optimal response. Considering the complexity in implementation, response surface method, therefore, is commonly used for the analysis of models with limited number of input factors.

In general, the functional relationship between model output and input parameters is unknown but can be approximated by a low-degree polynomial model of the form

$$y = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + \epsilon, \quad (1.15)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, the input vector of n -dimensional variables; $\mathbf{f}^T(\mathbf{x})$ is a vector function of a group of elements, consisting of powers and cross-products of powers of individual input parameters up to a certain degree d (≥ 1); $\boldsymbol{\beta}$ is a vector of unknown constant coefficients; ϵ is a random experimental error and assumed to have a zero mean [124].

Currently two important models are used in response surface method, with one being linear and the other, nonlinear. The linear one is also classified as the first-degree model ($d=1$) constructed in terms of the first-order terms of input parameters:

$$y = \beta_0 + \sum_{i=1}^n \beta_i x_i + \epsilon. \quad (1.16)$$

The nonlinear one is classified as the second-degree model ($d=2$), which still involves second-order effects of input parameters, except for first-order ones:

$$y = \beta_0 + \sum_{i=1}^n \beta_i x_i + \sum_i^n \sum_{j>i}^n \beta_{ij} x_i x_j + \sum_{i=1}^n \beta_{ii} x_i^2 + \epsilon. \quad (1.17)$$

The application of response surface method to sensitivity analysis of models can be concluded as three procedures:

- (1) To approximately establish the functional relationship connecting output variable and input factors.
- (2) To quantify, through hypothesis testing, importance of individual factors.
- (3) To determine the optimum settings of input factors that result in the maximum or minimum output value over a certain range of interest.

In general, the first-degree model is sufficient to determine which of input factors affect the model output of interest most. For a deep understanding of the effects produced by input factors on the model output, however, a more complicated design should be implemented to estimate a second-degree polynomial model.

A series of experiments should be first designed to perform response surface analysis, helping generate the mapping from input factors to the output variable. The design, denoted by \mathbf{D} , can be represented by an $M \times n$ matrix, as displayed in Eq. (1.1), where M is the number of experiments (the size of a design) and n the number of input variables. Each row of \mathbf{D} represents a point in the n -dimensional input space. Designs used for estimating the first-degree model are usually referred to as first-order designs and those used for estimating the second-degree model, second-order designs.

In the estimation of the first-degree model, an easy but most common design is 2^n factorial design [125]. In a 2^n factorial design, each input variable is measured at two levels which are commonly coded as -1 for the low level and +1 for the high level. A factorial design consists of all possible combinations of previously defined levels of n input factors. In practical applications, the points in a two-level factorial design are frequently represented by plus and minus signs, conventionally, $-$ for the first (or low) level, and $+$ for the second (or high) level. Take the case of three input factors as an example. The corresponding 2^3 design is a 8×3 matrix of the form

$$\mathbf{D} = \begin{bmatrix} - & - & - & + & - & + & + & + \\ - & - & + & - & + & - & + & + \\ - & + & - & - & + & + & - & + \end{bmatrix}^T. \quad (1.18)$$

If n is large, a large number of points will be introduced by the 2^n factorial design, thereby high cost is required in the computer simulation. For this case, fractions (e.g., one-half fraction, one-fourth fraction) of a 2^n design are often considered to reduce the cost of computer simulations in the estimation of the first-degree model. In general, a 2^{n-m} th fraction of a 2^n design contains 2^{n-m} points of a 2^n design. Here m is an integer number such that $2^{n-m} \geq n+1$ for guaranteeing all $n+1$ parameters (elements of vector β) included in the first-degree model (Eq. (1.16)) can be estimated. Particular manners for the construction of fractions of a 2^n design can be found in Refs. [126, 127]. Some other commonly discussed designs for fitting the first-degree model are Plackett-Burman design [128, 129] and simplex design [130]. The Plackett-Burman design allows two levels of each input factor, analogous to the 2^n factorial design, but requires a much smaller number of design points, especially for large n . The number of design points required by the Plackett-Burman design is equal to the number of parameters ($=n+1$) in the first-degree model. Specifically, this design can be employed only when the number of input variables, n , is a multiple of 4 [128]. The simplex design also contains $n+1$ experimental points. These points are located at the nodes of an n -dimensional regular-sided figure [131, 132].

In the fitting of the second-degree model, one of the most frequently discussed designs is the 3^n factorial design [133]. A 3^n factorial design is formed from all possible combinations of the levels of all n input variables. Each input variable has three levels that are commonly coded as -1 for the low level, 0 for the intermediate level, and +1 for the high level. In practice, the matrix of a 3^n factorial design simply consists of plus and minus signs, and also 0. Analogously, the number of experimental points for this design ($=3^n$) will be very large when a large number of input factors are involved into the original model. Following the phenomenon, fractions of a 3^n factorial design are often employed to save the cost of computer simulations. In the construction of fractions, the number of experimental points must at least equal the number of parameters ($=2n+1+n(n-1)/2$, the number of elements of vector β) included in the second-degree model (Eq. (1.17)) [127, 134]. Another most widely used second-order design is the central composite design. It is stated as the most popular design for building a second-degree model. The central composite design was first introduced by A. I. Khuri in 1988 [135], consisting of three distinct sets of experimental runs:

- (1) A full (or a fraction of) 2^n factorial design. This is called the factorial portion. Two levels of each input factor are coded as -1 and +1. They are often simplified as plus and minus signs in design matrix.
- (2) n_0 central points. Central point, commonly coded as 0, is the median of the values of each factor used in the factorial portion. n_0 replications of central point is used to

improve the precision of the experiment.

(3) $2n$ axial points. Two points are taken on the axis of each input variable at a distance of α from the center of the variable.

The number of experimental runs (design points) in a central composite design is $(2^n + 2n + n_0)$. The design matrix of a simple case with $n = 3$ and $n_0 = 4$ is formed as follows:

$$\mathbf{D} = \begin{bmatrix} - & - & - & + & - & + & + & + & -\alpha & \alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & - & + & - & + & - & + & + & 0 & 0 & -\alpha & \alpha & 0 & 0 & 0 & 0 & 0 & 0 \\ - & + & - & - & + & + & - & + & 0 & 0 & 0 & 0 & -\alpha & \alpha & 0 & 0 & 0 & 0 \end{bmatrix}^T. \quad (1.19)$$

Many strategies have been developed to select a useful value of the axial parameter α . Let F denote the number of points in the factorial portion and $T = 2n + n_0$. Two common values are

$$\alpha = (Q \times F/4)^{1/4} \quad (1.20)$$

with $Q = (\sqrt{F+T} - \sqrt{F})^2$, which makes the central composite design orthogonal, and

$$\alpha = F^{1/4} \quad (1.21)$$

which makes the design rotatable [136]. The value of n_0 is often assigned in terms of some certain desirable properties of the central composite design. For example, n_0 can be set to a value that makes a rotatable central composite design hold orthogonality property or the uniform precision property [130, 137]. The other most frequently used second-order design is the Box-Behnken design devised by G. E. P. Box and D. Behnken in 1960 [133]. In this design, three levels (equally spaced) of each input factor are considered. A Box-Behnken design is formed from a particular subset of the full 3^n factorial design [138, 139]. Some other second-order designs are available in Refs. [140–143].

A representative strategy to develop the expression of Eq. (1.15) is using a least squares regression method to fit a standardized first- or second-order equation to the data obtained from the original model. MC methods are typically borrowed to produce multiple values of each input factor, thereby to calculate corresponding values of model output. Other techniques such as rank-based or nonparametric approaches are also occasionally considered in employing the response surface method to establish the framework of sensitivity analysis [130, 144]. The precision and accuracy of analysis results provided by the response surface method can be evaluated by comparing the predictions provided by the method to the output values of the original model generated by the same values of input parameters. If the precision and accuracy are not satisfactory, an improved fit might be obtained by iterating on values of parameters [145].

Time-consuming and effort requirement in applying response surface approach typically rely on the number of input parameters included and the type of response structure required. Hence, mainly focusing on the effects of those input factors that have been identified as quite important through a screening sensitivity analysis method (e.g., one-at-a-time method) may be advantageous to reduce the complexity and difficulty in the implementation of response surface strategy.

A key advantage of the response surface method is that one can save computational time in computationally intensive model run by simplifying the form of the model under discussion. Furthermore, the functional form of the model that is established by the response surface method and the values of the coefficients included in the form can provide a fruitful information for quantifying the sensitivity of model output with respect to individual parameters. However, most frameworks established by the response surface method only consider the effects of some but not all of the input factors contained in the original model. This may result in absent or non-accurate global sensitivity measures in sensitivity analysis.

1.3.4 Differential-based method

Differential techniques for sensitivity analysis, also referred to as the direct or local methods, involve partial derivatives of output variable with respect to input parameters. In sensitivity analysis, one of the most used differential-based strategies is the first-order Taylor series approximation. It was discussed since 1966 by engineers and experimentalists [21].

Recall a generic model of the form $y = f(\mathbf{x})$ with $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ labeling the input vector of n -dimensional variables. By employing the first-order Taylor series approximation, the variance of output y , denoted as $V(y)$, is calculated as

$$V(y) = \sum_{i=1}^n \left(\frac{\partial y}{\partial x_i} \right)_{X_0}^2 V(x_i), \quad (1.22)$$

where the subscript X_0 indicates that the derivative is taken at a fixed point (often indicated by the central point) in the space of input variables, and $V(x_i)$ the variance of input x_i . The sensitivity coefficient, denoted by s_i , interprets the importance of input x_i in establishing the uncertainty of output y . s_i is determined by (see [146])

$$s_i = \left(\frac{\partial y}{\partial x_i} \right)_{X_0}^2 V(x_i) / V(y). \quad (1.23)$$

This calculation is performed under the assumption that high-order (≥ 2) partial differentials are negligible and input parameters are independent of each other. Consequently, the first-order Taylor series approximation can provide accurate and reliable analysis results only when the model under study is almost linear or the uncertainty in input parameters is negligible [17, 147].

Sensitivity analysis proposed by the differential-based technique is computationally efficient but bound with intensive effort requirement in solving differential equations. When an explicit algebraic equation describes the relationship connecting model output and input parameters, it is straightforward to evaluate sensitivity measures by the use of differential-based strategy. If a large set of equations are involved with the model under discussion, the first-order partial derivative can be approximated as a finite variation in output values driven from a small change in the input parameter [148, 149]. By neglecting non-linearities of models, the sensitivity of model output with respect to an arbitrary input parameter x_i can be approximated as

$$s_i = \frac{\% \Delta y}{\% \Delta x_i} = \frac{[f(\mathbf{x} + \Delta_i) - f(\mathbf{x})] / f(\mathbf{x})}{[x_i - \Delta_i] / x_i}, \quad (1.24)$$

where $(\mathbf{x} + \Delta_i) = (x_1, \dots, x_i + \Delta_i, \dots, x_n)$, and Δ_i is a small change introduced to input x_i .

A derivative-based global sensitivity method has also been proposed by Sobol and Kucherenko, by averaging the square of local derivatives [150, 151]. In this method, the global sensitivity measures are defined as

$$v_i = \int_{C^n} \left(\frac{\partial f}{\partial x_i} \right)^2 d\mathbf{x}, \quad (1.25)$$

where $C^n = (\mathbf{x} | 0 \leq x_i \leq 1; i = 1, 2, \dots, n)$, the n -dimensional unite hypercube. v_i can be regarded as an improvement of the importance criterion u^* (see Eq. (1.6)). The above definition is motivated by the fact that a high value of the derivative of model output with respect to an input variable indicates a robust influence of the input variable on the model output [152]. It is proved that

$$s_{Ti} \leq \frac{v_i}{\pi^2 V(y)}, \quad (1.26)$$

where s_{Ti} are the one-dimensional total sensitivity indices (see Eq. (1.38)) and $V(y)$ the total variance of output y . This states that small v_i imply small s_{Ti} . Unessential input parameters then can be identified out based on computed values of v_i ($i = 1, 2, \dots, n$). For highly nonlinear functions, however, the ranking of influential parameters in terms of the importance criterion v_i may suggest false conclusions [150].

The differential-based strategy is usually more demanding than other methods in the sensitivity analysis of complex models. It requires of model designers to explicitly calculate the first-order partial derivatives of output variable with respect to individual parameters, and yet provides only comparable but not accurate results, especially for nonlinear models.

1.3.5 Variance-based methods

Variance-based techniques have a long history in the aspect of sensitivity analysis. They are often used for determining whether an output variable is statistically associated with one or more input factors, and whether the values of model output vary in a statistically significant manner with the variation in values of one or more input variables. In the seventies, Cukier firstly established variance-based sensitivity analysis of multi-variate systems by Fourier implementation [153]. While the complete variance decomposition strategy was firstly developed by Sobol in 1993 [29]. In 1994, Jansen *et al* introduced an efficient method relying on random sampling to evaluate the partial contributions from input variables of uncertainty to the predicted variance in output variable [154]. A similar strategy was developed by Homma and Saltelli [155] in 1996 to determine global sensitivity measures that quantify the global importance of individual input variables in the estimation of model response. In spite of time consuming in computation, the instrument of complete variance decomposition is known to be useful and informative for uncertainty and sensitivity analysis of complex nonlinear systems [156].

Variance-based sensitivity analysis, often referred to as the Sobol method or Sobol indices, specifies the uncertainty in input and output variables through probability distributions. Working within a probabilistic framework, it decomposes the variance of model output into different partial contributions attributable to individual input variables and to their combinations. By computing the percentage of each partial variance contribution in the global variance of output variable, sensitivity measures are directly interpreted for individual factors and also for their interaction effects. Variance-based sensitivity analysis methods are attractive and widely used because they allow full exploration of input space, analysis of nonlinear models, and consideration of interactions between different input variables.

Without any assumption regarding the type of the model under discussion, variance-based approaches find broad applications across various fields, including scientific models evaluation [49], risk assessment [157], importance assessment [158], economic system analysis [28], behaviour prediction in forest systems [159], etc.

Recall the generic model of the form $y = f(\mathbf{x})$ with $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ labeling the input vector of n -dimensional variables. The model is defined over C^n , the n -dimensional unit hypercube, as defined before. Recalling the classical Hoeffding decomposition [29, 160–163], the output variable can be expanded as

$$y = f_0 + \sum_{i=1}^n f_i(x_i) + \sum_{i=1}^n \sum_{j>i}^n f_{ij}(x_i, x_j) + \dots + f_{12\dots n}(x_1, x_2, \dots, x_n), \quad (1.27)$$

where f_0 is a constant, f_i a function of x_i , f_{ij} a function of x_i and x_j , and so on up to the last term a function involving all input variables. Each term is square integrable over C^n . Summands presented in Eq. (1.27) must satisfy

$$\int_0^1 f_{i_1 i_2 \dots i_s}(x_{i_1}, x_{i_2}, \dots, x_{i_s}) dx_{i_k} = 0, \quad (1.28)$$

where $1 \leq i_1 < i_2 < \dots < i_s \leq n$ and $i_k \in \{i_1, i_2, \dots, i_s\}$. This condition drives

$$f_0 = \int_{C^n} f(\mathbf{x}) d\mathbf{x}, \quad (1.29)$$

$$f_i(x_i) = \int_0^1 \dots \int_0^1 f(\mathbf{x}) d\mathbf{x} / dx_i - f_0, \quad (1.30)$$

$$f_{ij}(x_i, x_j) = \int_0^1 \dots \int_0^1 f(\mathbf{x}) d\mathbf{x} / dx_i dx_j - f_0 - f_i(x_i) - f_j(x_j), \quad (1.31)$$

\vdots

where $d\mathbf{x}/dx_i$ is the product of all the dx_j except dx_i , similar to $d\mathbf{x}/dx_i dx_j$ which indicates the integration with respect to all input variables, except x_i and x_j . By assuming $f(\mathbf{x})$ is square-integrable, the variance of model output can be expressed as

$$V(y) = \int_0^1 \dots \int_0^1 \left[\sum_{s=1}^n \sum_{i_1 < \dots < i_s}^n f_{i_1, \dots, i_s}^2(x_{i_1}, x_{i_2}, \dots, x_{i_s}) dx_{i_1} \dots dx_{i_s} \right]. \quad (1.32)$$

Expanding the above equation yields

$$V(y) = \sum_{i=1}^n V_i + \sum_{i=1}^n \sum_{j>i}^n V_{ij} + \dots + V_{12\dots n}, \quad (1.33)$$

where

$$V_i = \int_0^1 f_i^2(x_i) dx_i, \quad (1.34)$$

$$V_{ij} = \int_0^1 \int_0^1 f_{ij}^2(x_i, x_j) dx_i dx_j, \quad (1.35)$$

\vdots

are in sequence explaining the variances contributed by x_i alone, by the interaction between x_i and x_j , etc. Equation (1.33) is known as the complete variance decomposition, deduced by Sobol in 1993. It shows how the variance of a model output can be decomposed into terms that are attributable to individual input factors and to their interaction effects [29]. Following the concept of Sobol's variance decomposition, sensitivity measures are defined by

$$s_i = V_i / V(y), \quad (1.36)$$

$$s_{ij} = V_{ij} / V(y), \quad (1.37)$$

$$\vdots$$

$$s_{Ti} = \left(V_i + \sum_{j=1; j \neq i}^n V_{ij} + \cdots + V_{12 \dots n} \right) / V(y), \quad (1.38)$$

where s_i is often called the first-order sensitivity index or the main effect index, labeling the fraction of $V(y)$ contributed by x_i alone; s_{ij} the second-order sensitivity index, indicating the fraction of $V(y)$ contributed by the interaction between x_i and x_j ; and so on up to the last label s_{Ti} the total sensitivity index, quantifying the fraction of $V(y)$ contributed by x_i alone and also by interactions of x_i with the remaining factors. They satisfy

$$\sum_{i=1}^n s_i + \sum_{i=1}^n \sum_{j>i}^n s_{ij} + \cdots + s_{12 \dots n} = 1, \quad (1.39)$$

$$\sum_{i=1}^n s_{Ti} \geq 1. \quad (1.40)$$

The equal sign in Eq. (1.40) holds iff the model under analysis is purely additive.

Theoretically, the determination of sensitivity measures depends upon multidimensional integrals. In practice, sampling-based strategies are often employed to carry out this mission. It can become computationally expensive when there are a large number of input variables.

Sobol's definitions of the second and higher order partial variance contributions in Eq. (1.33) hold if input factors are independent of each other. In the presence of input correlations, however, partial variance contributions with dimensionality larger than 1 are contributed not only by the coupling items presented in the functional form of the model under discussion (for independent case), but also by the input correlations. Recently, the interest in extending sensitivity analysis strategies from uncorrelated case to the correlated one is increasing as correlated input variables are of frequent occurrence in practical applications [164, 165]. In general, variance-based sensitivity analysis in the

presence of correlated inputs is carried out by using linear correlation model which regards the correlation part of an arbitrary variable as a linear combination of the rest variables [34–36, 166].

1.3.6 Moment independent method

The moment independent method looks at the influence of input uncertainty on the entire output distribution without reference to a specific moment of the output. The moment independent importance indicator was first introduced by Chun *et al* [167]. They quantified the entire change of CDFs in terms of the normalised Euclidean metric distance between two CDFs. The metric distance (Minkowski distance) of order a between two points $X_1 = (x_1^1, x_1^2, \dots, x_1^n)$ and $X_2 = (x_2^1, x_2^2, \dots, x_2^n)$ is in general defined by

$$D = \left(\sum_{i=1}^n |x_1^i - x_2^i|^a \right)^{1/a}, \quad (1.41)$$

where a is a number no less than 1. The metric distance is typically used with a being 1 or 2, which correspond to the Manhattan distance and the Euclidean distance, respectively. The measure of uncertainty importance is then defined by

$$\text{MD}(i : o) = \left(\int_0^1 [Y_j^i - Y_j^o]^2 dj \right)^{1/2} / E(y^o), \quad (1.42)$$

where Y_j^o is the j th percentile of the output CDF for the base case, Y_j^i the j th percentile of the output CDF after introducing a change to input x_i , and $E(y^o)$ the mean value of output y for the base case. The base case refers to the case where an output distribution is obtained by setting all input distributions to their nominal ones. It is stated that a larger value of $\text{MD}(i : o)$ implies a more important parameter x_i , compared to other input parameters. However, the value of $\text{MD}(i : o)$ depends on the hypothesised change introduced to parameter x_i .

Another moment independent importance measure was presented by Borgonovo, which does not require one pre-suppose any changes associated with input parameters [168]. It follows from his concept that the global sensitivity index of parameter x_i , denoted by δ_i , with respect to the output y is represented by

$$\delta_i = \frac{1}{2} \int P(x_i) \left[\int |P_y(y) - P_{y|x_i}(y)| dy \right] dx_i, \quad (1.43)$$

where $P(x_i) = \int \dots \int P(\mathbf{x}) \prod_{j \neq i} dx_j$ the marginal density (PDF) of input x_i , $P(y)$ is the PDF of output y , and $P_{Y|x_i}(y)$ indicates the conditional PDF of y by assuming that the input parameter x_i is fixed at a constant value. It is proved that δ_i holds some properties,

such as a) $0 \leq \delta_i \leq 1$; b) $\delta_i = 0$ if the output y is independent of x_i ; c) $\delta_{1,2,\dots,n} = 1$ (the importance of all input parameters equals one); etc [169]. The indicator δ helps identify out input parameters that affect output uncertainty the most. In identifying the less relevant parameters with respect to the model output, δ agrees with Sobol's variance-based method. However, discrepancy between two evaluations exists in ranking relevant parameters.

Regarding the difficulty in deriving the PDFS of output variable, a new moment independent measure was recently proposed, called the PAWN index, with the conditional and unconditional distributions characterised by their CDFs [170]. The Kolmogorov-Smirnov statistic ([171]) is employed to measure the distance between conditional and unconditional CDFs:

$$KS(x_i) = \max_y |F_y(y) - F_{y|x_i}(y)|, \quad (1.44)$$

where $F_y(y)$ is the unconditional CDF of output y , $F_{y|x_i}(y)$ the conditional CDF when the input x_i is fixed. As $KS(x_i)$ depends on the value at which x_i is fixed, the PAWN index, denoted by T_i , is then defined by considering a statistic (e.g. maximum or median) over all possible values of x_i :

$$T_i = \text{stat}_{x_i}\{KS(x_i)\}. \quad (1.45)$$

The PAWN index T_i is a global, quantitative and model independent sensitivity index, varying between 0 and 1 (the higher the value, the more influential x_i). T_i ($i = 1, 2, \dots, n$) can be used for ranking input parameters according to their contributions to the output uncertainty, so as to identify which of input parameters are influential, and which are non-influential. Compared to other moment independent sensitivity indices, the PAWN index has the advantage of being very easy to implement and interpret [170]. Regarding the convergence rate and screening of non-influential parameters, the PAWN method is in good agreement with Sobol's variance-based method. However, the PAWN method is better than Sobol's one in quantifying the difference between relative importance of influential parameters [172].

1.3.7 Sampling-based method

In the implementation of uncertainty and sensitivity analysis, the sampling-based (also known as MC) approach has been classified as both effective and widely used. In carrying out the sampling-based approach, what is the most fundamental and critical is to generate and explore the mapping from uncertain input variables to the uncertain model response. In exploration and generation of the underlying mapping, two basic components should be previously considered. The first one is the definition of probability distributions P_1, P_2, \dots, P_n that characterise uncertainties in input variables in

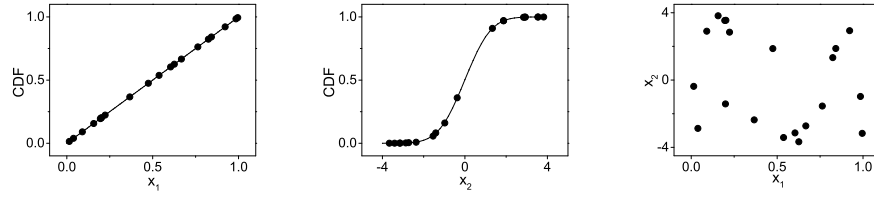


FIGURE 1.6: Example of random sampling to generate a sample of size $M = 20$ from $\mathbf{x} = (x_1, x_2)$ with x_1 uniformly distributed and x_2 normally distributed, defined by Fig.1.2(a). First two panels illustrate sampling of values for x_1 and x_2 , respectively, and the last panel presents a random pairing of the sampled values of x_1 and x_2 in the construction of a random sampling.

the absence of input correlations. The second one is the generation of sample points in n -dimensional input space according to the previously defined distributions of individual input variables. Diversified sampling procedures have been designed. In this section, we discuss some frequently considered sampling strategies: random sampling, sampling for iterated fractional factorial design (IFFD), importance sampling, Latin Hypercube sampling (LHS), and Quasi Monte Carlo (QMC) sampling.

Random sampling

Sample points provided by random sampling are generated by pseudo-random numbers. Values of single uncertain input variables are chosen randomly and entirely by chance from their real ranges, independent of their PDFs [173]. It is stated that random sampling is prone to clustering that yields wanted sample points and empty areas in input space because each new sample is selected randomly without minding the gaps between already generated sample values [174, 175]. Figure 1.6 presents an example of random samples of size $M = 20$ generated from $\mathbf{x} = (x_1, x_2)$. x_1 is uniformly distributed in the real range $[0, 1]$ and x_2 follows the standard normal distribution (Fig.1.2(a)). Clustering effect and empty areas are evident in the distribution of sample points. The uniformity of sample points in input space directly determines the accuracy of analysis results when the number of sample points is finite.

IFFD

The factorial sampling and fractional factorial sampling have been briefly mentioned before. They are widely considered in the response surface method. Sampling for IFFD was also proposed as a direct tool for sensitivity analysis [176, 177]. In this sampling, input variables are sampled at discrete levels. This is the main difference from other sampling strategies where each input variable is sampled at a continuous interval. IFFD was built to identify a few robust parameters within batches of a large number of fragile

ones [177]. Investigation indicates that the sensitivity measure based on IFFD is extremely reproducible, and more robust than rank regression coefficient in that it can also identify quadratic effects of individual parameters, except for linear ones [178].

Specifically, IFFD is generated by multiple iterations of a basic orthogonal fractional factorial design. Given a basic fractional factorial design, three steps of randomization control the generation of a sampling for IFFD:

- (1) Parameters are randomly assigned to columns of the basic fractional factorial design.
- (2) Each parameter is randomly oriented in considering two orientations: positive one and negative one. Parameters with a positive orientation will copy their values from the associated column of the basic design, and those with a negative one will take the opposite values from the associated column of the basic design.
- (3) The orientation value of each parameter is randomly set to zero in a previously defined proportion of the total iterations.

The first two randomization steps are carried out independently for each iteration, and the third one is performed within the whole set of iterations. Each iteration of an IFFD is analysed separately. The obtained results are then combined together for the sensitivity analysis of the system under discussion. Denote the value of the output variable in the i th simulation of the m th iteration by $y^m[i]$, and the input parameter that copies values from the j th column of the basic design by z_j . The main effect in the m th iteration of parameter z_j on the output variable is given by

$$ME_m(z_j, y^m) = \frac{1}{n} \sum_{i=1}^{2n} J_n[i, j] \cdot y^m[i], \quad (1.46)$$

where n is the number of input parameters, J_n the basic fractional factorial design of size $n \times n$, and $J_n[i, j]$ the (i, j) th position in matrix J_n . $ME_m(z_j, y^m)$ is a linear effect as it calculates the difference in average response between two levels (low and high) of z_j . The main effect of a parameter x throughout the entire design of M iterations is given by

$$ME(x, y) = \frac{\sum_{m=1}^M S_x^m \cdot ME(z_{c_x^m}, y^m)}{\sum_{m=1}^M |S_x^m|}, \quad (1.47)$$

where S_x^m is the orientation value of parameter x in the m th iteration, c_x^m (taking a value from 1 to n) the randomly chosen column associated with parameter x . The quadratic effect of parameter x can also be defined by

$$QE(x, y) = \frac{\sum_{m=1}^M (1 - |S_x^m|) \sum_{i=1}^{2n} y^m[i]}{2n \sum_{m=1}^M (1 - |S_x^m|)} - \frac{\sum_{m=1}^M |S_x^m| \sum_{i=1}^{2n} y^m[i]}{2n \sum_{m=1}^M |S_x^m|}. \quad (1.48)$$

Importance sampling

In statistics, importance sampling is a well-known technique for estimating properties of a particular distribution, by using samples generated from a different distribution than the distribution of interest. It is often used for the estimate of rare event probability [179, 180]. Consider a general univariate function $f(x)$. x is a random variable from a previously defined distribution of density $P(x)$. One wishes to estimate the expected value of $f(x)$ under $P(x)$, denoted as $E(f; P)$. A MC estimate of $E(f; P)$ is

$$\begin{aligned}\hat{E}(f; P) &= \int f(x)P(x)dx \\ &= \frac{1}{M} \sum_{j=1}^M f(x_j)\end{aligned}\tag{1.49}$$

where M denotes the sample size. The precision of this estimate depends upon the variance of $f(x)$:

$$V(\hat{E}(f); P) = V(f, P)/M.\tag{1.50}$$

The basic idea of importance sampling is to reduce $V(\hat{E}(f); P)$ for a given sample size by using samples from a different distribution. The principle of importance sampling is

$$\hat{E}(f; P) = \int f(x)P(x)dx = \int f(x)\frac{P(x)}{G(x)}G(x)dx,\tag{1.51}$$

where the newly introduced distribution $G(x)$ should have the same support as $P(x)$. Apparently, sampling x from a distribution $P(x)$ is equivalent to sampling $x \cdot w(x)$ from the new distribution $G(x)$, with importance sampling weight:

$$w(x) \equiv \frac{P(x)}{G(x)}.\tag{1.52}$$

The object in importance sampling is to concentrate the distribution of the sample points in the parts of interval of most importance (e.g. the interval $[0, 1]$ that we discussed in the example) instead of spreading them over the whole region [181]. A good importance sampling function $G(x)$ should hold the following properties [182]

- (1) $G(x) > 0$ whenever $f(x) \neq 0$;
- (2) $G(x)$ should be most closely the shape of $f(x)$ and makes $f(x) \cdot P(x)/G(x)$ bounded;
- (3) It should be easy to simulate values from $G(x)$;
- (4) It should be easy to compute the density $G(x)$ for any values of x that one might realize.

TABLE 1.1: Estimated mean value and its standard deviation based on MC and importance sampling (denoted by IS). Results are obtained by N independent simulations of sample size 100.

N	Simulation	$\hat{E}(f; P)$	$\sigma(\hat{E}(f); P)$
100	MC	0.459	0.051
	IS	0.460	0.020
1000	MC	0.459	0.048
	IS	0.461	0.021
100000	MC	0.460	0.049
	IS	0.460	0.022

Consider a simple example of $f(x) = x$ with x following a standard normal distribution: $x \sim N(0, 1)$. One wants to evaluate the mean of f in the region $[0, 1]$ of x . By MC sampling, a sequence values x_j ($j = 1, \dots, M$) of x can be generated from $N(0, 1)$. The mean value of $f(x)$ is then calculated by

$$\hat{E}(f; P) = \frac{\sum_{j=1}^M c_j \cdot x_j}{\sum_{j=1}^M c_j}, \quad (1.53)$$

with

$$c_j = \begin{cases} 1 & 0 \leq x_j \leq 1, \\ 0 & \text{otherwise.} \end{cases} \quad (1.54)$$

In importance sampling, we set $G(x) = 1$ for $x \in [0, 1]$. M sampled values of x then can be obtained from a uniform distribution in the region $[0, 1]$. For each sampled value x_j , the importance weight is

$$w(x_j) = \frac{P(x_j)}{\int_0^1 P(z) dz}. \quad (1.55)$$

where $P(\cdot)$ is the standard normal distribution. The mean value of $f(x)$ based on importance sampling is then calculated by

$$\hat{E}(f; P) = \frac{1}{M} \sum_{j=1}^M x_j \cdot w(x_j). \quad (1.56)$$

A comparison between MC and importance sampling is presented in table 1.1. Results display that the variance of the mean value of $f(x)$ is reduced by introducing importance sampling.

LHS

LHS is a kind of stratified sampling strategy. It is widely regarded as one of the most popular variance reduction techniques that can be employed to increase the analysis efficiency so as to more possibly achieve the desired accuracy. The LHS was firstly

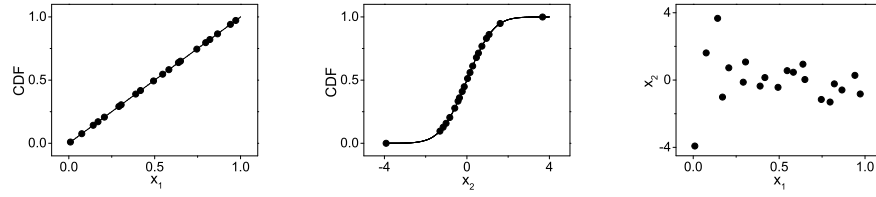


FIGURE 1.7: Example of LHS to generate a sample of size $M = 20$ from $\mathbf{x} = (x_1, x_2)$ with x_1 uniformly distributed and x_2 normally distributed, defined by Fig.1.2(a). First two panels illustrate sampling of values for x_1 and x_2 , respectively, and the last one a random pairing of sampled values of x_1 and x_2 in the construction of a LHS.

designed by McKay *et al* in 1979 for numerically evaluating a multiple integral [174]. It was further elaborated by Iman *et al* in 1981 [183].

Sample points provided by LHS are generated in terms of PDFs of individual input variables, namely P_1, P_2, \dots, P_n . Divide the range of x_1 into M (the size of sample) adjacent intervals with equal probability provided by P_1 . One point is selected at random from each interval to construct a sequence of M points $\{x_j^1\}$, $j = 1, 2, \dots, M$. Another sequence, $\{x_j^2\}$, $j = 1, 2, \dots, M$, can be formed similarly but independently for x_2 . The two generated sequences are then randomly paired without replacement to constitute an $M \times 2$ matrix. These M pairs are continuously randomly combined without replacement with M points in the sequence $\{x_j^3\}$, $j = 1, 2, \dots, M$, to form an $M \times 3$ matrix, and so on until an $M \times n$ matrix is constructed:

$$x_j = [x_j^1, x_j^2, \dots, x_j^n], \quad j = 1, 2, \dots, M. \quad (1.57)$$

Figure 1.7 presents an example of LHS of size 20 for a special two-variate situation $\mathbf{x} = (x_1, x_2)$. x_1 is uniformly distributed between 0 and 1 and x_2 follows the standard normal distribution, as defined in Fig.1.2(a).

The above generated LHS is known as standard LHS (or randomly generated LHS). It can be quite structured: the variables may be highly correlated or the sampled values may not have good uniformity of distribution in the space of input variables. Consequently, some optimal criteria are proposed to avoid these problems and to obtain an optimal sampling which achieves the space-filling property of input factors.

An often considered optimization is designed based on the maximin distance criterion [184]. For an established sampling design \mathbf{D} (an $M \times n$ matrix), the inner-site distance between any two n -dimensional sample points (two rows in the $M \times n$ matrix, as displayed in Eq. (1.57)) is defined by

$$d(\mathbf{x}_t, \mathbf{x}_s) = \left[\sum_{i=1}^n (x_t^i - x_s^i)^\alpha \right]^{1/\alpha}, \quad (1.58)$$

where α takes two values: 1 (corresponding to the rectangular distance) or 2 (related to the Euclidean distance). The first built maximin distance criterion aims at maximizing the minimum inter-site distance:

$$\min_{1 \leq t, s \leq M; t \neq s} d(\mathbf{x}_t, \mathbf{x}_s). \quad (1.59)$$

Later on, an intuitively appealing extension of the maximin distance criterion was introduced by Morris and Mitchell [185]. For a designed sampling \mathbf{D} , they defined a distance list: (d_1, d_2, \dots, d_m) where the elements are the distinct values of inter-site distances and sorted with $d_1 < d_2 < \dots < d_m$. Obviously, m , labeling the number of distinct distance values, can be as large as $\binom{M}{2}$ (the inter-site distances between any two sampling items are unique) and as small as 1 (the inter-site distances between any two sampling items are consistent). Continuously, an index list was accordingly defined: (J_1, J_2, \dots, J_m) in which an element J_k counts the number of pairs of sites in the designed sampling separated by d_k . The maximin distance design criterion tends to make the minimum distance maximal. This implies a sampling driven by this criterion must follow the property: (1) d_1 is maximized; (2) J_1 is minimized. To construct an optimal sampling that satisfies this property, an evaluation function was then introduced:

$$\phi_p(\mathbf{D}) = \left[\sum_{k=1}^m J_k d_k^{-\alpha} \right]^{1/\alpha}, \quad (1.60)$$

where α is a positive integer and \mathbf{D} a designed sampling. An optimal sampling is generated by minimising Eq. (1.60) [185].

The Shannon information entropy defined by Shannon in 1948 was also applied to the optimisation of LHS, known as the entropy criterion [186]. Shannon demonstrated that the lower the information entropy, the more precise the knowledge is. It has been further illustrated that the entropy criterion is equivalent to minimizing the quantity:

$$-\log |\mathbf{R}|, \quad (1.61)$$

where \mathbf{R} is the correlation matrix of the designed sampling with elements given by

$$r_{ts} = \exp \left[\sum_{i=1}^n c_i |x_t^i - x_s^i|^\alpha \right], \quad 1 \leq t, s \leq M; 1 \leq \alpha \leq 2, \quad (1.62)$$

in which c_i ($i = 1, 2, \dots, n$) are correlation coefficients [187]. The entropy criterion can help design an optimal sampling that minimizes correlations among input factors.

In experimental design, L_2 discrepancy is most frequently used in solving the non-uniformity problem. The centered L_2 -discrepancy (CL_2) criterion was developed to optimise both random sampling and LHS [188]. According to this criterion, a constructed sampling \mathbf{D} (an $M \times n$ matrix) is optimal (having space-filling property) if it minimizes the centered L_2 discrepancy given by

$$\begin{aligned}
 CL_2(\mathbf{X}) = & \left(\frac{13}{12}\right)^2 - \frac{2}{M} \sum_{j=1}^M \prod_{i=1}^n \left(1 + \frac{1}{2}|x_j^i - 0.5| - \frac{1}{2}|x_j^i - 0.5|^2\right) \\
 & + \frac{1}{M^2} \sum_{j=1}^M \sum_{k=1}^M \prod_{i=1}^n \left(1 + \frac{1}{2}|x_j^i - 0.5| + \frac{1}{2}|x_k^i - 0.5| - \frac{1}{2}|x_j^i - x_k^i|\right). \quad (1.63)
 \end{aligned}$$

Some other early developed algorithms also exist in constructing optimal LHSs. They have been shown to have a good space-filling property, including, for example, the row-wise element exchange algorithm [189], the columnwise-pairwise algorithm [190], and the threshold accepting algorithm [191]. Considering the requirement of highly computational cost in the implementation of these algorithms, Jin *et al* developed an efficient and flexible method for constructing optimal samplings, by the use of enhanced stochastic evolutionary algorithm and the employment of efficient evaluating strategies of the optimality criteria [192]. Some recently introduced methods for optimising the strategy of LHS can be found in Refs. [193–195].

QMC sampling

QMC method is specifically designed by using low-discrepancy sequences (also called quasi-random sequences or sub-random sequences) to place sampled values as uniformly as possible. This is in contrast to the regular MC method (random sampling) consisting of sequences of pseudo-random numbers. MC and QMC methods are stated in a similar way. The problem is to approximate the integral of a function as the average of the function evaluated at a set of sampled values. The difference between QMC and MC methods is the way sampled values are generated. There are a few well-known and commonly used low-discrepancy sequences, including Halton sequences [196], Sobol sequences [197], and Faure sequences [198]. Sobol low-discrepancy sequences are identified in many aspects as the superior one, compared to other sequences.

Sobol sequences (also called LP_τ sequences or (t, s) sequences in base 2) were first introduced by Sobol in 1967 [199]. Define a real integrable function f over C^n (n -dimensional unit hypercube). The original motivation of Sobol was to generate a sequence \mathbf{x}_j in C^n

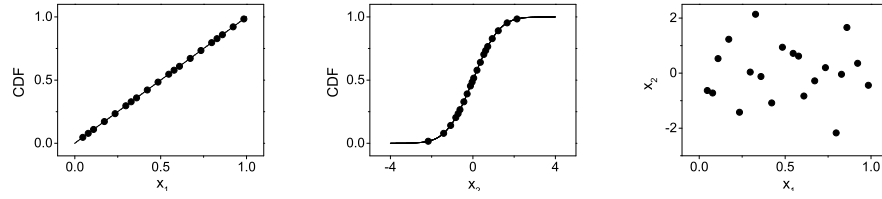


FIGURE 1.8: An example of QMC sampling based on Sobol sequences. The sample size: $M = 20$. x_1 is uniformly distributed and x_2 normally distributed, defined by Fig.1.2(a). The first two panels illustrate sampled values of x_1 and x_2 , respectively, and last one the distribution of sampled values in the construction of QMC sampling.

so that

$$\lim_{M \rightarrow \infty} \frac{1}{M} \sum_{j=1}^M f(\mathbf{x}_j) = \int_{C^n} f(\mathbf{x}) d\mathbf{x}, \quad (1.64)$$

and the convergence is as fast as possible.

Sample values of QMC sampling by using Sobol sequences are uniformly placed in input space, even for rather small sample size. Additional conditions, known as property A and A', were also introduced by Sobol in 1975 to make sample values perfectly uniformly distributed in input space [200]:

Definition: A low-discrepancy sequence is said to satisfy **Property A** if for any binary segment (not an arbitrary subset) of the n -dimensional sequence of length 2^n there is exactly one point in each 2^n hyper-octant that results from subdividing the unit hypercube along each of its length extensions into half.

Definition: A low-discrepancy sequence is said to satisfy **Property A'** if for any binary segment (not an arbitrary subset) of the n -dimensional sequence of length 4^n there is exactly one point in each 4^n hyper-octant that results from subdividing the unit hypercube along each of its length extensions into four equal parts.

QMC method recently attracts much interest from researchers working in the field of mathematical finance or computational finance. In these fields, high-dimensional numerical integrals in hundreds or thousands of variables are frequently evaluated within a previously defined threshold. Hence, QMC method is widely used in the financial sector, helping value financial derivatives [201, 202].

The detailed generation process of Sobol sequences can be found in Ref. [203]. An example of Sobol sequences-based QMC sampling is presented in Fig. 1.8. A sample of size $M = 20$ is generated from $\mathbf{x} = (x_1, x_2)$. x_1 is assumed to be uniformly distributed between 0 and 1, and x_2 normally distributed (Fig.1.2(a)).

A qualitative comparison is also presented in Fig. 1.9 for the uniformity of sampled values provided by different sampling strategies, including random sampling, standard

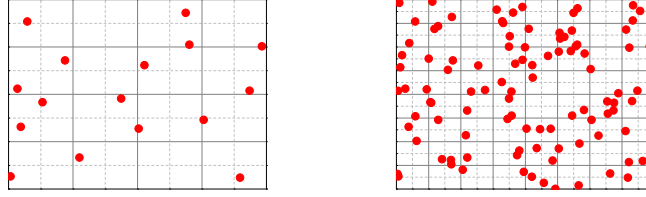
LHS, LHS with the use of maximin distance criterion, LHS with the use of CL_2 criterion, and QMC sampling based on Sobol sequences. A situation of 2-dimensional input vector is exemplified. Two kinds of sample size: $M = 16$ and $M = 100$ are considered, respectively. Results suggest a higher uniformity of sampled values provided by Sobol sequences-based QMC sampling, compared to those given by random sampling and LHSs. Furthermore, the optimised LHS is superior to the standard LHS which precedes random sampling in constructing uniformly distributed sample values, especially when a large number of input factors are included. Some other newly introduced sampling strategies could be found in Refs. [204, 205].

1.4 Determination of analysis results

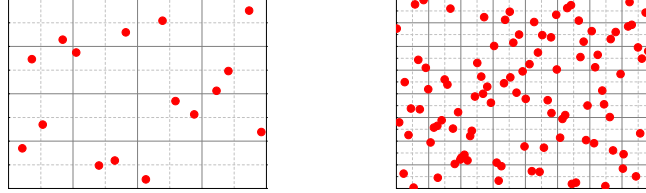
In the determination of uncertainty and sensitivity analysis results, the most fundamental task is to actually explore the mapping from input vector to output variables: $[\mathbf{x}_j, \mathbf{y}(\mathbf{x}_j)]$, $j = 1, 2, \dots, M$. \mathbf{y} labels the output vector. It may contain several independent or correlated observations. For simplicity in statement, we just consider an arbitrary element in \mathbf{y} , denoted by y . The mapping from input vector to the model output can be analytically obtained if the functional form of the model under study is previously provided. With this kind of situation, optionally, the mapping of interest can also be constructed by computational strategies (sampling-based methods) which are most frequently used since their implementations do not require the form of model function to be given at first. In the following, several representations of analysis results are introduced by the use of sampling-based methods.

1.4.1 Scatter plots

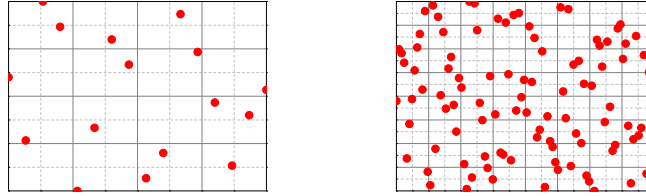
Scatter plots-based representation is a natural starting point in the uncertainty and sensitivity analysis of complex systems. It provides an intuitive visual indication of the dependence of output variable upon individual input factors. Furthermore, it is definitely advantageous for directly understanding the impact of uncertainties in input factors on the uncertainty in model output. Scatter plots of two two-variate simple models are presented as examples: one is constructed in the absence of coupling items between input factors (Fig. 1.10) and the other is designed in the presence of coupling items. Input factors are assumed to be independent of each other and to be uniformly distributed in the real range $[0, 1]$. The scatter plots generated by the standard LHS and by the one-at-a-time method are considered, respectively. For sampling-based methods, the dependence of output y versus a single input factor is plotted in keeping values of the rest factors arbitrary. Regarding one-at-a-time method, however, the relationship of y



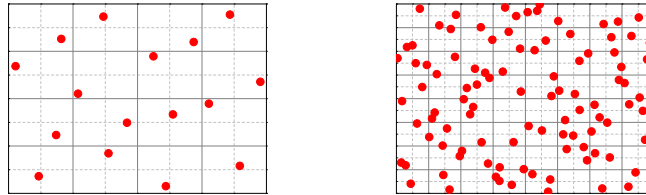
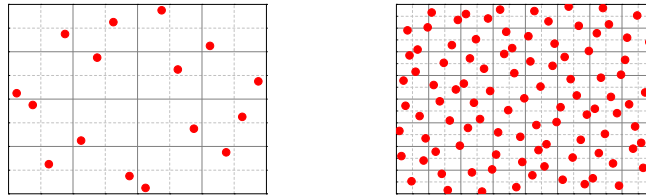
(a) Sampled values based on random sampling.



(b) Sampled values based on standard LHS.



(c) Sampled values based on LHS with the use of maximin distance criterion.

(d) Sampled values based on LHS with the use of CL_2 criterion.

(e) Sampled values of QMC sampling based on Sobol sequences.

FIGURE 1.9: Distribution of sampled values in a 2-dimensional input space. Input factors are assumed to be uniformly distributed in $[0, 1]$. (a): sampled values are generated by employing random sampling based on pseudo-random numbers. (b): sampled values are generated by applying the standard LHS. (c): sampled values are generated based on LHS which is optimised by the maximin distance criterion. (d): sample values are generated based on LHS which is optimised by CL_2 criterion (e): sampled values are generated by Sobol sequences-based QMC sampling. From left to right, the size of sample $M = 16$ and $M = 100$, respectively. The unit square is divided into 64 (on the left) and 256 (on the right) squares.

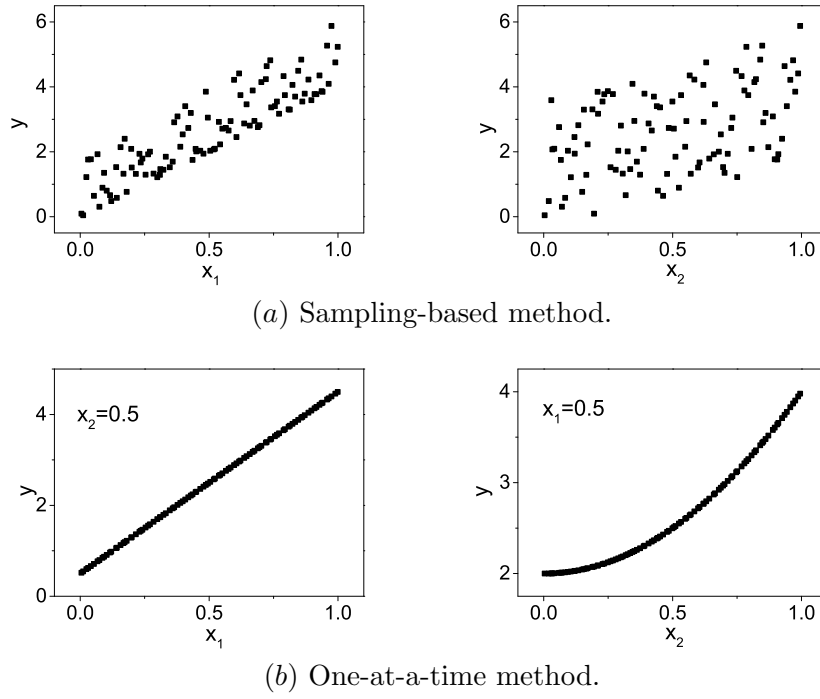


FIGURE 1.10: (Model 1) Examples of scatter plots-based representation for the uncertainty and sensitivity analysis of model: $y = 4x_1 + 2x_2^2$. The input factors are assumed to be independent of each other and to be uniformly distributed in the real range $[0, 1]$. Panel (a) presents the plots generated by standard LHS. Panel (b) illustrates the plots obtained in terms of one-at-a-time method.

with a single input factor is presented while fixing the rest factors (often constant at their mathematical expectation values). It is straightforward to qualitatively state that the factor x_1 is more important than x_2 in establishing the uncertainty in y for both models, according to the sampling-based scatter plots. For the model with absent coupling items between input factors, one-at-a-time method is superior to sampling-based ones in displaying the qualitative relationship between output variable and individual input factors (Fig. 1.10). However, when coupling items between input factors are involved, one-at-a-time method can just present local dependence of output variable upon a single factor by neglecting the interaction effects between different input factors (introduced by coupling items) (Fig. 1.11(b)). Moreover, analysis results associated with each single factor rely on the previously assumed values of the remaining factors. This implies sampling-based scatter plots are definitely beneficial for displaying the global dependence of output variable upon input factors, regarding models involving coupling terms (Fig. 1.11(a)).

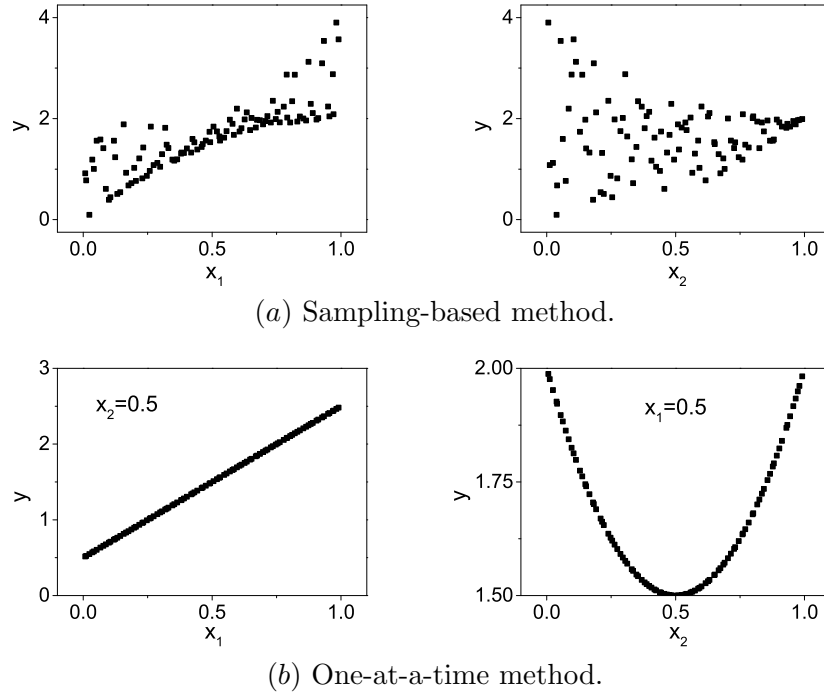


FIGURE 1.11: (Model 2) Examples of scatter plots-based representation for the uncertainty and sensitivity analysis of model: $y = 4x_1 + 2x_2^2 - 4x_1x_2$. The input factors are assumed to be independent of each other and to be uniformly distributed in the real range $[0, 1]$. Panel (a) presents the plots generated by standard LHS. Panel (b) illustrates the plots obtained in terms of one-at-a-time method.

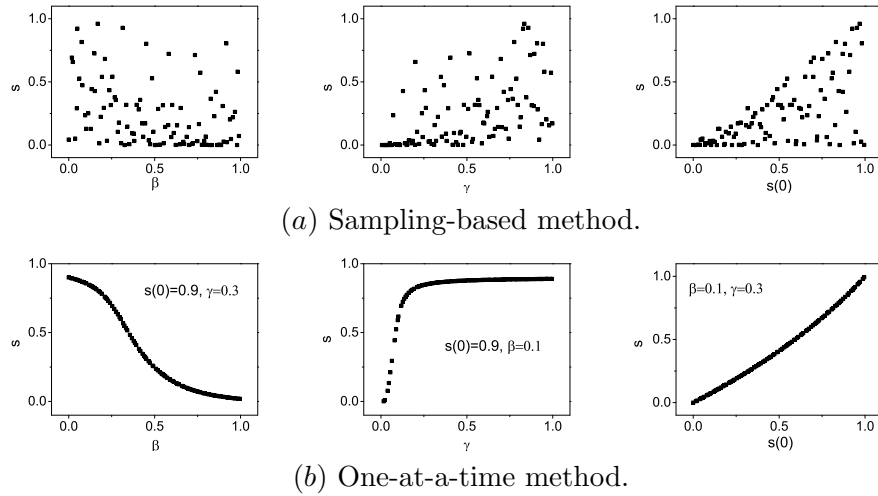


FIGURE 1.12: (SIR model) Examples of scatter plots-based representation for the uncertainty and sensitivity analysis of SIR model. The input factors β (infectious probability), γ (recovered rate), and $s(0)$ (initial proportion of susceptible agents) are assumed to be independent of each other and uniformly distributed in the real range $[0, 1]$. Panel (a) presents the plots generated by standard LHS. Panel (b) illustrates the plots obtained in terms of one-at-a-time method.

1.4.2 Correlation measures

Correlation indices quantify the correlation and independence between two or more values in fundamental statistics. One of the most often considered correlation indices is Pearson correlation coefficient (CC). The CC was introduced by Pearson in the 1880s, following a related idea of Galton [206–208]. It provides a measure of the strength of linear correlation between individual input factor and the output variable. The CC between two variables has values between -1 and 1. Three particular values -1, 0, and 1 separately correspond to the total negative linear correlation, no linear correlation, and total linear correlation. Specifically, the CC between an input factor x_i and output variable y , denoted by $\rho(x_i, y)$, is defined by

$$\rho(x_i, y) = \frac{\frac{1}{M} \sum_{j=1}^M (x_j^i - \mu_i)(y_j - E(y))}{\sigma_i \sigma_y}, \quad (1.65)$$

where μ_i denotes the mathematical expectation of input x_i , $E(y)$ the mathematical expectation of output y , σ_i the standard deviation of x_i , and σ_y the standard deviation of y . They are determined by

$$\mu_i = \frac{1}{M} \sum_{j=1}^M x_j^i, \quad E(y) = \frac{1}{M} \sum_{j=1}^M y_j, \quad (1.66)$$

$$\sigma_i = \left[\frac{1}{M} \sum_{j=1}^M (x_j^i - \mu_i)^2 \right]^{1/2}, \quad \sigma_y = \left[\frac{1}{M} \sum_{j=1}^M (y_j - E(y))^2 \right]^{1/2}. \quad (1.67)$$

The absolute value of $\rho(x_i, y)$, between 0 and 1, corresponds to a trend from no linear relationship to an exact linear relationship between x_i and y . Typically, for linear (purely additive) models, the sum of squared pairwise CCs between each of the input factors and output variable is 1 in the absence of correlations among input factors:

$$\sum_{i=1}^n \rho^2(x_i, y) = 1. \quad (1.68)$$

As an example, the CCs are shown in table 1.2 for two two-variate nonlinear models and SIR model. Their scatter plots are presented in Figs. 1.10, 1.11 and 1.12, by the use of LHS-based method and one-at-a-time method.

The partial correlation coefficient (PCC) provides a measure to characterise the degree of association between a single factor x_i and an output variable y with the effects of the remaining elements of \mathbf{x} removed. By the use of linear regression model, a new variable

TABLE 1.2: The CCs for two-variate nonlinear models and SIR model with scatter plots presented in Figs. 1.10, 1.11 and 1.12. For SIR model, x_1 denotes the parameter β , x_2 the parameter γ , x_3 the parameter $s(0)$, and y the output variable s .

CC	sampling-based method			one-at-a-time method		
	model 1 (Fig. 1.10(a))	model 2 (Fig. 1.11(a))	SIR model (Fig. 1.12(a))	model 1 (Fig. 1.10(b))	model 2 (Fig. 1.11(b))	SIR model (Fig. 1.12(b))
$\rho(x_1, y)$	0.89	0.84	-0.35	1	1	-0.96
$\rho(x_2, y)$	0.44	-0.03	0.53	0.97	0.00	0.61
$\rho(x_3, y)$	—	—	0.62	—	—	1.00

is introduced based on the elements of \mathbf{x} except x_i :

$$\hat{x}_i = c_0 + \sum_{j=1, j \neq i}^n c_j x_j \quad (1.69)$$

And the effects of the remaining elements of \mathbf{x} on y are represented as

$$\hat{y} = b_0 + \sum_{j=1, j \neq i}^n b_j x_j. \quad (1.70)$$

The PCC between x_i and y is defined as the CC between $(x_i - \hat{x}_i)$ and $(y - \hat{y})$. An example of the PCCs for nonlinear models will be presented later, together with a similar measure: the partial rank correlation coefficient.

Spearman's rank correlation coefficient (RCC), named after Charles Spearman, characterises the statistical association between the ranking of two variables of interest. It quantifies how well the dependence between two variables can be explained by a monotonic function. The RCC between x_i and y is defined as the CC between their corresponding ranks. Consider a size of sample M . The values of x_i and y are replaced by their rank numbers, positive integers between 1 and M , in the sorting of x_i and y from the smallest to the largest, namely rx_i and ry , respectively. The RCC between x_i and y is then given by

$$\rho(x_i, y) = \frac{\sum_{j=1}^M (rx_j^i - \bar{rx}_i)(ry_j - \bar{ry})}{\left[\sum_{j=1}^M (rx_j^i - \bar{rx}_i)^2 \right]^{1/2} \left[\sum_{j=1}^M (ry_j - \bar{ry})^2 \right]^{1/2}}. \quad (1.71)$$

When M rank numbers (integers from 1 to M) are distinct from each other for both x_i and y , the mathematical expectation and statistical variance of rank transformed data is calculated as

$$\bar{rx}_i = \bar{ry} = \frac{M+1}{2}, \quad V(rx_i) = V(ry) = \frac{M^2 - 1}{12}, \quad (1.72)$$

TABLE 1.3: RCCs for two-variate nonlinear models and SIR model with scatter plots presented in Figs. 1.10, 1.11 and 1.12. In SIR model, x_1 denotes the parameter β , x_2 the parameter γ , x_3 the parameter $s(0)$, and y the output variable s .

RCC	sampling-based method			one-at-a-time method		
	model 1 (Fig. 1.10(a))	model 2 (Fig. 1.11(a))	SIR model (Fig. 1.12(a))	model 1 (Fig. 1.10(b))	model 2 (Fig. 1.11(b))	SIR model (Fig. 1.12(b))
$\rho(x_1, y)$	0.89	0.89	-0.38	1	1	-1
$\rho(x_2, y)$	0.41	0.06	0.63	1	0.00	1
$\rho(x_3, y)$	—	—	0.58	—	—	1

where the formula $\sum_{j=1}^M j^2 = M(2M+1)(M+1)/6$ was used. The RCC between x_i and y then can be simplified as

$$\rho(x_i, y) = 1 - \frac{6 \sum_{j=1}^M (rx_j^i - ry_j)^2}{M(M^2 - 1)}. \quad (1.73)$$

The RCC between x_i and y has values between -1 and 1, analogous to the CC, with positive values corresponding to the increase or decrease of x_i and y along the same direction and negative values corresponding to the increase or decrease of x_i and y along opposite directions. A RCC of zero value indicates an obvious non-monotonic relationship between x_i and y . The larger the RCC in magnitude, the closer the relationship between x_i and y gets to be monotonic. The RCCs are displayed in table 1.3 for three models with scatter plots presented in Figs. 1.10, 1.11 and 1.12, by the use of LHS-based method and one-at-a-time method.

Another rank correlation coefficient is the Kendall rank correlation coefficient(KRCC), commonly referred to as Kendall's τ coefficient. The KRCC was developed by Kendall in 1938 [209]. It measures the ordinal association between two variables under study. The KRCC between x_i and y is also defined based on their rank numbers in the sorting from the smallest to the largest. It has values between -1 and 1, with high values indicating a similar (or identical when the coefficient is 1) rank between x_i and y and low values labeling a dissimilar (or fully different when the coefficient is -1) rank between them. A KRCC of zero value between x_i and y corresponds to the independent relationship between them. For a size of sample M , an arbitrary pair of mappings from x_i to y : $[x_j^i, y_j], [x_k^i, y_k], j \neq k$ and $j, k = 1, 2, \dots, M$, is said to be concordant if the ranks for both elements follow the same direction, that is, if both $x_j^i > x_k^i$ and $y_j > y_k$ or if both $x_j^i < x_k^i$ and $y_j < y_k$. It is said to be discordant if the ranks for both elements follow opposite directions, that is, if $x_j^i > x_k^i$ and $y_j < y_k$ or if $x_j^i < x_k^i$ and $y_j > y_k$. When $x_j^i = x_k^i$ or $y_j = y_k$, the correspond pair of mappings is neither concordant nor discordant. The definition of KRCC is given by

$$\tau(x_i, y) = \frac{2[(\text{number of concordant pairs}) - (\text{number of discordant pairs})]}{M(M-1)}. \quad (1.74)$$

TABLE 1.4: KRCCs for two-variate nonlinear models and SIR model with scatter plots presented in Figs. 1.10, 1.11 and 1.12. For SIR model, x_1 denotes the parameter β , x_2 the parameter γ , x_3 the parameter $s(0)$, and y the output variable s .

KRCC	sampling-based method			one-at-a-time method		
	model 1 (Fig. 1.10(a))	model 2 (Fig. 1.11(a))	SIR model (Fig. 1.12(a))	model 1 (Fig. 1.10(b))	model 2 (Fig. 1.11(b))	SIR model (Fig. 1.12(b))
$\tau(x_1, y)$	0.71	0.73	-0.27	1	1	-1
$\tau(x_2, y)$	0.29	0.05	0.46	1	0.00	1
$\tau(x_3, y)$	—	—	0.45	—	—	1

TABLE 1.5: The PCCs and PRCCs for two-variate nonlinear models and SIR model with scatter plots presented in Figs. 1.10(a) (model 1), 1.11(a) (model 2) and 1.12(a) (SIR model). For SIR model, x_1 denotes the parameter β , x_2 the parameter γ , x_3 the parameter $s(0)$, and y the output variable s .

measures	PCC			PRCC		
	model 1	model 2	SIR model	model 1	model 2	SIR model
$\rho(x_1, y)$	0.99	0.83	-0.57	0.98	0.90	-0.68
$\rho(x_2, y)$	0.97	-0.02	0.64	0.92	0.17	0.78
$\rho(x_3, y)$	—	—	0.73	—	—	0.76

As an example, the KRCCs are exhibited in table 1.4 for the models with scatter plots presented in Figs. 1.10, 1.11 and 1.12, by the use of LHS-based method and one-at-a-time method.

Similar to PCC, the partial rank correlation coefficient (PRCC) performs a partial correlation on rank-transformed data: rx_i and ry by the use of linear regression models described by Eqs. (1.69) and (1.70). The PRCC provides a robust sensitivity measure for nonlinear but monotonic relationships between x_i and y in the absence of input correlations. The PCCs and PRCCs are presented in table 1.5 for nonlinear models with scatter plots illustrated in Figs. 1.10(a) (model 1), 1.11(a) (model 2) and 1.12 (SIR model), by the use of LHS-based method.

In the absence of coupling terms involving different input factors, rank correlation coefficients, including RCC and KRCC, based on one-at-a-time method is equivalent to the PCC and PRCC based on sampling strategies. They all provide reliable measures in quantifying the monotonic relationship between each individual input factor and the output variable. However, in the presence of coupling items, the PRCC is superior to any other correlation coefficient in assessing how well the relationship between an individual input factor and the output variable can be described by a monotonic function.

Occasionally, multiple correlation is also of interest to be considered. The coefficient of multiple correlation, denoted as R in general, assesses how well a given output variable can be predicted using a linear function of a set of input variables [210]. It provides a global measure of the strength of the association between input variables and the output variable. The coefficient of multiple correlation takes values between 0 and 1. A higher value indicates a stronger linear association connecting input factors and output

variable. A value of zero suggests the absence of linear relationship between input factors and output variable. Unlike the previously introduced correlation coefficients which offer both the strength and direction of the underlying association connecting input factors and the output variable, the coefficient of multiple correlation tells only the strength of the association. Regarding a general model of two input factors, the coefficient of multiple correlation is defined by

$$R = \left[\frac{\rho^2(x_1, y) + \rho^2(x_2, y) - 2\rho(x_1, y)\rho(x_2, y)\rho(x_1, x_2)}{1 - \rho^2(x_1, x_2)} \right]^{1/2}, \quad (1.75)$$

where ρ indicates the pairwise CC (Eq. (1.65)) between involved variables. By assuming input variables to be independent of each other, the above expression can be simplified as

$$R = [\rho^2(x_1, y) + \rho^2(x_2, y)]^{1/2}, \quad (1.76)$$

which is naturally extended to a general situation of n input factors of absent input correlations:

$$R = \left[\sum_{i=1}^n \rho^2(x_i, y) \right]^{1/2}. \quad (1.77)$$

The multiple correlation coefficient is 0.99 for model 1 (Fig. 1.10(a)), 0.84 for model 2 (Fig. 1.11(a)) and 0.89 for SIR model (Fig. 1.12(a))

1.4.3 Sensitivity indices

The sensitivity indices (or measures), also called importance indices (measures), of output variables with respect to input factors are often defined by variance-based methods. One commonly considered variance-based method is the Sobol's variance decomposition [29]. Following his concept, the determination of sensitivity indices depends upon multidimensional integrals, as introduced in Section 1.3. Considering the complexity in evaluating multidimensional integrals, the sampling-based method is frequently applied to the calculation of sensitivity indices in practice. With sampling-based method, the sensitivity indices of different orders are evaluated by using two independently constructed samplings, namely \mathbf{A} and \mathbf{B} , which are $M \times n$ matrices. The mathematical expectation and statistical variance of output y are approximated by any of the constructed samplings, for instance, \mathbf{A} , as

$$E(y) = \frac{1}{M} \sum_{j=1}^M f(\mathbf{A})_j, \quad (1.78)$$

$$V(y) = \frac{1}{M} \sum_{j=1}^M f^2(\mathbf{A})_j - E^2(y). \quad (1.79)$$

TABLE 1.6: An example of construction of both $\mathbf{A}_B^{(i)}$ and $\mathbf{B}_A^{(i)}$ matrices with sample size $M = 5$ and number of input variables $n = 3$. Matrices \mathbf{A} and \mathbf{B} are generated by the use of standard LHS. Three input variables are independently and uniformly distributed in the range $[0, 1]$.

$\mathbf{A} = \begin{bmatrix} 0.747 & 0.359 & 0.098 \\ 0.053 & 0.783 & 0.955 \\ 0.325 & 0.098 & 0.453 \\ 0.899 & 0.936 & 0.220 \\ 0.519 & 0.549 & 0.618 \end{bmatrix},$	$\mathbf{B} = \begin{bmatrix} 0.760 & 0.806 & 0.381 \\ 0.524 & 0.755 & 0.422 \\ 0.342 & 0.194 & 0.180 \\ 0.949 & 0.400 & 0.746 \\ 0.062 & 0.480 & 0.958 \end{bmatrix}.$
$\mathbf{A}_B^{(1)} = \begin{bmatrix} 0.760 & 0.359 & 0.098 \\ 0.524 & 0.783 & 0.955 \\ 0.342 & 0.098 & 0.453 \\ 0.949 & 0.936 & 0.220 \\ 0.062 & 0.549 & 0.618 \end{bmatrix},$	$\mathbf{B}_A^{(1)} = \begin{bmatrix} 0.747 & 0.806 & 0.381 \\ 0.053 & 0.755 & 0.422 \\ 0.325 & 0.194 & 0.180 \\ 0.899 & 0.400 & 0.746 \\ 0.519 & 0.480 & 0.958 \end{bmatrix},$
$\mathbf{A}_B^{(2)} = \begin{bmatrix} 0.747 & 0.806 & 0.098 \\ 0.053 & 0.755 & 0.955 \\ 0.325 & 0.194 & 0.453 \\ 0.899 & 0.400 & 0.220 \\ 0.519 & 0.480 & 0.618 \end{bmatrix},$	$\mathbf{B}_A^{(2)} = \begin{bmatrix} 0.760 & 0.359 & 0.381 \\ 0.524 & 0.783 & 0.422 \\ 0.342 & 0.098 & 0.180 \\ 0.949 & 0.936 & 0.746 \\ 0.062 & 0.549 & 0.958 \end{bmatrix},$
$\mathbf{A}_B^{(3)} = \begin{bmatrix} 0.747 & 0.359 & 0.381 \\ 0.053 & 0.783 & 0.422 \\ 0.325 & 0.098 & 0.180 \\ 0.899 & 0.936 & 0.746 \\ 0.519 & 0.549 & 0.958 \end{bmatrix},$	$\mathbf{B}_A^{(3)} = \begin{bmatrix} 0.760 & 0.806 & 0.098 \\ 0.524 & 0.755 & 0.955 \\ 0.342 & 0.194 & 0.453 \\ 0.949 & 0.400 & 0.220 \\ 0.062 & 0.480 & 0.618 \end{bmatrix}.$

Following the concept of Sobol's variance decomposition, the first-order and total sensitivity indices are estimated, respectively, by

$$s_i = \left[\frac{1}{M} \sum_{j=1}^M f(\mathbf{A})_j f(\mathbf{B}_A^{(i)})_j - E^2(y) \right] / V(y), \quad (1.80)$$

$$s_{Ti} = \left[\frac{1}{M} \sum_{j=1}^M f(\mathbf{A})_j \left(f(\mathbf{A})_j - f(\mathbf{B}_A^{(i)})_j \right) \right] / V(y), \quad (1.81)$$

where $\mathbf{A}_B^{(i)}$ is a matrix with column i in \mathbf{A} substituted by the i th column in \mathbf{B} , and vice versa for matrix $\mathbf{B}_A^{(i)}$ whose column i comes from matrix \mathbf{A} and the remaining $(n - 1)$ columns come from matrix \mathbf{B} [156]. An example of construction of both $\mathbf{A}_B^{(i)}$ and $\mathbf{B}_A^{(i)}$ matrices is presented in table 1.6 with sample size $M = 5$ and three input variables considered. They are assumed to be independently and uniformly distributed in the real range $[0, 1]$.

Analogously, the second- and higher-order sensitivity indices are determined by

$$s_{i_1 i_2} = \left[\frac{1}{M} \sum_{j=1}^M f(\mathbf{A})_j f(\mathbf{B}_A^{(i_1 i_2)})_j - E^2(y) \right] / V(y) - s_{i_1} - s_{i_2}, \quad (1.82)$$

$$s_{i_1 i_2 i_3} = \left[\frac{1}{M} \sum_{j=1}^M f(\mathbf{A})_j f(\mathbf{B}_A^{(i_1 i_2 i_3)})_j - E^2(y) \right] / V(y) - s_{i_1} - s_{i_2} \quad (1.83)$$

$$- s_{i_3} - s_{i_1 i_2} - s_{i_1 i_3} - s_{i_2 i_3}, \quad (1.84)$$

\vdots

where the first two indices separately quantify the fraction of total variance of output y that is contributed by the interaction effect between x_{i_1} and x_{i_2} , and by the interaction effect involving x_{i_1} , x_{i_2} and x_{i_3} .

Alternative formulas for the determination of the first-order and total sensitivity indices are offered by Jansen in 1999 [154, 211], which are expressed as

$$s_i = 1 - \left[\frac{1}{2M} \sum_{j=1}^M \left(f(\mathbf{B})_j - f(\mathbf{A}_B^{(i)})_j \right)^2 \right] / V(y), \quad (1.85)$$

$$s_{Ti} = \left[\frac{1}{2M} \sum_{j=1}^M \left(f(\mathbf{A})_j - f(\mathbf{A}_B^{(i)})_j \right)^2 \right] / V(y). \quad (1.86)$$

By employing a large set of test functions involving different degrees of linearity, additivity and effective dimension, it is stated that Jansen's method is more efficient than Sobol's one in evaluating sensitivity indices [156]. The sensitivity indices of three exemplified models are presented, as examples, in tables 1.7, 1.8, and 1.9, accompanying with a comparison of Sobol's and Jansen's methods. When the size of sample is large enough, Sobol's and Jansen's methods both offer perfect sensitivity indices. However, Jansen's method is more efficient than Sobol's one as it requires less computational cost in implementation. Both the standard LHS and QMC sampling are performed to carry out the analysis process in considering the effect of sample size. QMC sampling can be seen as a deterministic algorithm as the low discrepancy sequence used for the construction of sampling is not random, but deterministic. Specifically, in the implementation of QMC sampling, matrices \mathbf{A} and \mathbf{B} of size $M \times n$, used for the determination of sensitivity analysis results, should be constructed from a quasi-random sequence of size $M \times 2n$. \mathbf{A} is generated by the left half part of the sequence, and \mathbf{B} by the rest half part, as shown in table 1.10. Regarding sampling strategies, average over independent constructions of LHS is of necessity in obtaining accurate sensitivity measures. This states again that QMC sampling is much more efficient for sensitivity analysis of models. It provides accurate analysis results by using a small number of sampled points and also avoids highly computational cost in performing configuration average.

TABLE 1.7: Variance-based sensitivity analysis results for the model of form $y = 4x_1 + 2x_2^2$. The standard LHS and QMC sampling are used, with the effect of sample size M considered. Both Sobol's (Eqs. (1.80)-(1.81)) and Jansen's (Eqs. (1.85)-(1.86)) methods are employed to the evaluation of sensitivity indices. For LHS, analysis results are averaged over 100 independent configurations, with standard deviations displayed in parentheses. QMC sampling, generated based on Sobol sequences, is deterministic and numbers in parentheses indicate the thousandth and ten-thousandth digits.

sampling	M	method	$E(y)$	$V(y)$	s_1	s_2	s_{T1}	s_{T2}
LHS	10^2	Sobol'	2.67(.00)	1.69(.15)	0.79(.05)	0.21(.08)	0.79(.08)	0.21(.05)
		Jansen	2.67(.00)	1.70(.15)	0.79(.03)	0.22(.11)	0.78(.11)	0.21(.03)
	10^4	Sobol'	2.67(.00)	1.69(.01)	0.79(.00)	0.21(.01)	0.79(.01)	0.21(.00)
		Jansen	2.67(.00)	1.69(.01)	0.79(.00)	0.21(.01)	0.79(.01)	0.21(.00)

QMC	10^2	Sobol'	2.66(72)	1.67(50)	0.79(63)	0.21(41)	0.78(60)	0.20(37)
		Jansen	2.66(72)	1.67(50)	0.77(61)	0.22(85)	0.77(15)	0.22(40)
	10^4	Sobol'	2.66(67)	1.68(89)	0.78(90)	0.21(05)	0.78(95)	0.21(10)
		Jansen	2.66(67)	1.68(89)	0.78(95)	0.21(06)	0.78(95)	0.21(05)
	$3 \cdot 10^4$	Sobol'	2.66(66)	1.68(88)	0.78(97)	0.21(03)	0.78(97)	0.21(04)
		Jansen	2.66(66)	1.68(88)	0.78(95)	0.21(05)	0.78(95)	0.21(05)

TABLE 1.8: Variance-based sensitivity analysis results for the model of form $y = 4x_1 + 2x_2^2 - 4x_1x_2$. The standard LHS and QMC sampling are used, with the effect of sample size M considered. Both Sobol's (Eqs. (1.80)-(1.81)) and Jansen's (Eqs. (1.85)-(1.86)) methods are employed to the evaluation of sensitivity indices. For LHS, analysis results are averaged over 100 independent configurations, with standard deviations displayed in parentheses. QMC sampling, generated based on Sobol sequences, is deterministic and numbers in parentheses indicate the thousandth and ten-thousandth digits.

sampling	M	method	$E(y)$	$V(y)$	s_1	s_2	s_{T1}	s_{T2}
LHS	10^2	Sobol'	1.67(.04)	0.46(.04)	0.71(.19)	0.03(.21)	0.97(.21)	0.29(.19)
		Jansen	1.67(.03)	0.47(.04)	0.72(.05)	0.04(.14)	0.96(.12)	0.29(.04)
	10^4	Sobol'	1.67(.00)	0.47(.00)	0.72(.02)	0.05(.02)	0.95(.02)	0.28(.02)
		Jansen	1.67(.00)	0.47(.00)	0.72(.00)	0.05(.02)	0.95(.01)	0.28(.00)

QMC	10^2	Sobol'	1.67(46)	0.48(74)	0.67(71)	0.08(13)	0.91(87)	0.32(29)
		Jansen	1.67(46)	0.48(74)	0.72(51)	0.09(33)	0.85(59)	0.26(22)
	10^4	Sobol'	1.66(67)	0.46(65)	0.71(46)	0.04(78)	0.95(22)	0.28(54)
		Jansen	1.66(67)	0.46(65)	0.71(41)	0.04(73)	0.95(24)	0.28(60)
	$3 \cdot 10^4$	Sobol'	1.66(67)	0.46(68)	0.71(41)	0.04(72)	0.95(28)	0.28(59)
		Jansen	1.66(67)	0.46(68)	0.71(43)	0.04(77)	0.95(25)	0.28(57)

TABLE 1.9: Variance-based sensitivity analysis results for SIR model. The standard LHS and QMC sampling are used, with the effect of sample size M considered. Both Sobol's (Eqs. (1.80)-(1.81)) and Jansen's (Eqs. (1.85)-(1.86)) methods are employed to the evaluation of sensitivity indices. For LHS, analysis results are averaged over 100 independent configurations, with standard deviations displayed in parentheses. QMC sampling, generated based on Sobol sequences, is deterministic and numbers in parentheses indicate the thousandth and ten-thousandth digits.

sampling	M	method	$E(y)$	$V(y)$	s_β	s_γ	$s_{s(0)}$	$s_{T\beta}$	$s_{T\gamma}$	$s_{Ts(0)}$
LHS	10^2	Sobol'	0.24(.01)	0.07(.01)	0.19(.14)	0.20(.12)	0.43(.15)	0.30(.09)	0.33(.12)	0.59(.13)
		Jansen	0.25(.01)	0.07(.01)	0.19(.12)	0.19(.13)	0.42(.10)	0.31(.07)	0.33(.08)	0.60(.09)
	10^4	Sobol'	0.24(.00)	0.07(.00)	0.19(.01)	0.18(.01)	0.42(.01)	0.31(.01)	0.34(.01)	0.59(.01)
		Jansen	0.24(.00)	0.07(.00)	0.19(.01)	0.19(.01)	0.42(.01)	0.31(.01)	0.34(.01)	0.59(.01)

QMC	10^2	Sobol'	0.23(83)	0.07(30)	0.23(21)	0.23(95)	0.31(50)	0.34(66)	0.50(42)	0.54(42)
		Jansen	0.23(83)	0.07(30)	0.31(77)	0.36(61)	0.41(09)	0.32(83)	0.41(59)	0.56(16)
	10^4	Sobol'	0.24(19)	0.07(14)	0.18(56)	0.18(11)	0.41(54)	0.30(97)	0.33(99)	0.59(45)
		Jansen	0.24(19)	0.07(14)	0.18(30)	0.18(08)	0.41(36)	0.30(99)	0.33(97)	0.59(50)
	$3 \cdot 10^4$	Sobol'	0.24(19)	0.07(14)	0.18(61)	0.18(15)	0.41(51)	0.30(82)	0.34(13)	0.59(53)
		Jansen	0.24(19)	0.07(14)	0.18(75)	0.18(30)	0.41(59)	0.30(81)	0.34(08)	0.59(55)

TABLE 1.10: An example of the construction of two independent QMC samplings with sample size $M = 10$ and number of input variables $n = 3$. Three input variables are independently and uniformly distributed in the real range $[0, 1]$.

$\begin{bmatrix} 0.09375 & 0.46875 & 0.84375 & 0.40625 & 0.28125 & 0.34375 \\ 0.59375 & 0.96875 & 0.34375 & 0.90625 & 0.78125 & 0.84375 \\ 0.84375 & 0.21875 & 0.09375 & 0.15625 & 0.53125 & 0.09375 \\ 0.34375 & 0.71875 & 0.59375 & 0.65625 & 0.03125 & 0.59375 \\ 0.46875 & 0.09375 & 0.46875 & 0.28125 & 0.65625 & 0.71875 \\ 0.96875 & 0.59375 & 0.96875 & 0.78125 & 0.15625 & 0.21875 \\ 0.71875 & 0.34375 & 0.71875 & 0.03125 & 0.40625 & 0.96875 \\ 0.21875 & 0.84375 & 0.21875 & 0.53125 & 0.90625 & 0.46875 \\ 0.15625 & 0.15625 & 0.53125 & 0.84375 & 0.84375 & 0.40625 \\ 0.65625 & 0.65625 & 0.03125 & 0.34375 & 0.34375 & 0.90625 \end{bmatrix},$					
$\mathbf{A} = \begin{bmatrix} 0.09375 & 0.46875 & 0.84375 \\ 0.59375 & 0.96875 & 0.34375 \\ 0.84375 & 0.21875 & 0.09375 \\ 0.34375 & 0.71875 & 0.59375 \\ 0.46875 & 0.09375 & 0.46875 \\ 0.96875 & 0.59375 & 0.96875 \\ 0.71875 & 0.34375 & 0.71875 \\ 0.21875 & 0.84375 & 0.21875 \\ 0.15625 & 0.15625 & 0.53125 \\ 0.65625 & 0.65625 & 0.03125 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0.40625 & 0.28125 & 0.34375 \\ 0.90625 & 0.78125 & 0.84375 \\ 0.15625 & 0.53125 & 0.09375 \\ 0.65625 & 0.03125 & 0.59375 \\ 0.28125 & 0.65625 & 0.71875 \\ 0.78125 & 0.15625 & 0.21875 \\ 0.03125 & 0.40625 & 0.96875 \\ 0.53125 & 0.90625 & 0.46875 \\ 0.84375 & 0.84375 & 0.40625 \\ 0.34375 & 0.34375 & 0.90625 \end{bmatrix}.$					

Chapter 2

The analytic analysis for models with independent input variables

Any operations that we perform on a model response dependent upon a number of input variables of uncertainty require us to identify the response uncertainty based on the uncertainty in input variables. The propagation of variance, characterising the effect of input uncertainty on the uncertainty of model response, constitutes the essential ingredient of uncertainty and sensitivity analysis of complex systems. In the present chapter, an analytic formula is derived by using Taylor series to specify the variance propagation from input variables to the model response in the absence of input correlations. With the formula, we can exactly calculate the uncertainty of model response. Furthermore, it also allows one to quantify the partial variance contributions of different orders from input variables to the output one, whereby input variables can be ranked according to their importance in explaining the uncertainty of the output variable.

2.1 Taylor series

The concept of a Taylor series was formulated by a Scottish mathematician James Gregory and formally introduced by an English mathematician Brook Taylor in 1715. In mathematics, a Taylor series is a representation of a function as an infinite sum of terms that are evaluated from the values of the function's derivatives at a single point [212]. In general, a function can be approximated as the sum of a finite number of terms constituting of function's derivatives of different orders by using Taylor series of neglecting higher-order terms.

Taylor series is widely used for the expansion of functions only dependent upon one single variable. Regarding a general univariate function of the form $y = f(x)$, its Taylor series at a point $x = a$ is given by

$$\sum_{i=0}^{\infty} \frac{f^{(i)}(a)}{i!} (x - a)^i, \quad (2.1)$$

where $i!$ denotes the factorial of i , and $f^{(i)}(a)$ the i th derivative of f evaluated at point a . The zeroth-order derivative of f is defined as f itself and $0!=1$. If $a = 0$, the expansion is known as a Maclaurin series, named after the Scottish mathematician Colin Maclaurin who introduced extensive use of this special case of Taylor series in the 18th century [213].

The Taylor series can also be generalised to functions of more than one input variable [214]. For a real function of the form $y = f(\mathbf{x})$ with $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ labeling the input vector of n -dimensional variables, its Taylor series at a specified point $\mathbf{x} = (a_1, a_2, \dots, a_n)^T$ is represented as

$$\sum_{i_1=0}^{\infty} \dots \sum_{i_n=0}^{\infty} \frac{(x_1 - a_1)^{i_1} \dots (x_n - a_n)^{i_n}}{i_1! \dots i_n!} \left(\frac{\partial^{i_1+\dots+i_n} f}{\partial x_1^{i_1} \dots \partial x_n^{i_n}} \right) (a_1, \dots, a_n). \quad (2.2)$$

2.2 Variance propagation for univariate case

Beginning with the univariate function $y = f(x)$, it can be expanded by the use of Taylor series at the central point of x as

$$y = f(\mu) + \sum_{i=1}^{\infty} \frac{1}{i!} \left(\frac{d^i f}{dx^i} \Big|_{x=\mu} \right) (x - \mu)^i, \quad (2.3)$$

in which μ indicates the mathematical expectation of x . Taking the average of both sides of Eq.(2.3) yields

$$E(y) = f(\mu) + \sum_{i=1}^{\infty} \frac{1}{i!} \left(\frac{d^i f}{dx^i} \Big|_{x=\mu} \right) M_i(x), \quad (2.4)$$

where $M_i(x)$ is the i^{th} central moment of variable x , given by

$$M_i(x) = \int (x - \mu)^i P(x) dx. \quad (2.5)$$

$P(x)$, as stated in Chapter 1, denotes the PDF of x and characterises the epistemic uncertainty in x . The variance of y , namely $V(y)$, is defined in mathematics by

$$V(y) = \int y^2 P(x) dx - E^2(y). \quad (2.6)$$

Substituting Eqs. (2.3) and (2.4) into Eq. (2.6) provides

$$\begin{aligned} V(y) &= \int \left[f(\mu) + \sum_{i=1}^{\infty} \frac{1}{i!} \left(\frac{d^i f}{dx^i} \Big|_{x=\mu} \right) (x - \mu)^i \right]^2 P(x) dx \\ &\quad - \left[f(\mu) + \sum_{i=1}^{\infty} \frac{1}{i!} \left(\frac{d^i f}{dx^i} \Big|_{x=\mu} \right) M_i(x) \right]^2 \\ &= f^2(\mu) + \sum_{i,j=1}^{\infty} \frac{1}{i! \cdot j!} \left(\frac{d^i f}{dx^i} \Big|_{x=\mu} \cdot \frac{d^j f}{dx^j} \Big|_{x=\mu} \right) \int (x - \mu)^{i+j} P(x) dx \\ &\quad + 2f(\mu) \sum_{i=1}^{\infty} \frac{1}{i!} \left(\frac{d^i f}{dx^i} \Big|_{x=\mu} \right) \int (x - \mu)^i P(x) dx - 2f(\mu) \sum_{i=1}^{\infty} \frac{1}{i!} \left(\frac{d^i f}{dx^i} \Big|_{x=\mu} \right) M_i(x) \\ &\quad - \sum_{i,j=1}^{\infty} \frac{1}{i! \cdot j!} \left(\frac{d^i f}{dx^i} \Big|_{x=\mu} \cdot \frac{d^j f}{dx^j} \Big|_{x=\mu} \right) M_i(x) M_j(x) - f^2(\mu) \\ &= \sum_{i,j=1}^{\infty} \frac{1}{i! \cdot j!} \left(\frac{d^i f}{dx^i} \Big|_{x=\mu} \cdot \frac{d^j f}{dx^j} \Big|_{x=\mu} \right) [M_{i+j}(x) - M_i(x) M_j(x)], \end{aligned} \quad (2.7)$$

which is equivalent to

$$V(y) = \sum_{i,j=0}^{\infty} \frac{1}{i! \cdot j!} \left(\frac{d^i f}{dx^i} \Big|_{x=\mu} \cdot \frac{d^j f}{dx^j} \Big|_{x=\mu} \right) [M_{i+j}(x) - M_i(x) M_j(x)], \quad (2.8)$$

since $M_0(x) = 1$. This analytic expression mathematically explains the output variance propagated from the uncertainty in input variable through the functional relationship connecting y and x . Furthermore, it can also help identify the partial variance contributions of different orders of x , embodied by the derivatives of different orders of f . If one just considers the first-order variance contribution of x , Eq. (2.8) is then simplified as

$$V(y) \approx \left(\frac{df}{dx} \Big|_{x=\mu} \right)^2 V(x), \quad (2.9)$$

where $M_1 = 0$ and $M_2 = V(x)$ have been used, and $V(x)$ labels the variance of x . Equation (2.9), called the general Taylor series expansion truncated to the first order, is most widely used for the estimation of the uncertainty in y in terms of the uncertainty in x . This approximation, however, is satisfactory for the uncertainty analysis of highly nonlinear models only when the uncertainty in input variable is negligible [17], as stated

in the first chapter.

To identify the partial variance contributions of different orders from input x , a new quantity of interest is proposed, labeled as R_v^γ . It is defined as the ratio of $V(y)$ contributed by first γ orders of uncertainty in x to its exact value:

$$R_v^\gamma = \frac{V_\gamma(y)}{V_T(y)}, \quad (2.10)$$

where $V_\gamma(y)$ is calculated with Eq. (2.8) under the condition $i + j \leq 2\gamma$ and $V_T(y)$ the theoretical value of $V(y)$ obtained from the below integral:

$$\begin{aligned} V_T(y) &= \int y^2 P(x) dx - E^2(y), \\ &= \int f^2(x) P(x) dx - \left(\int f(x) P(x) dx \right)^2. \end{aligned} \quad (2.11)$$

The analytic formula previously deduced (Eq. (2.8)) specifies the variance propagation from input variable to the model output. With this formula, it is natural to quantify the importance of partial variance contributions of different orders from input x in the estimation of uncertainty in output y . In the following, higher-order contributions of input uncertainty to the uncertainty of output y are evaluated for highly nonlinear models, by using the analytic formula. Two widely discussed distribution laws: uniform distribution and normal distribution, as examples, specify the uncertainty in input x .

2.2.1 Uniform distribution

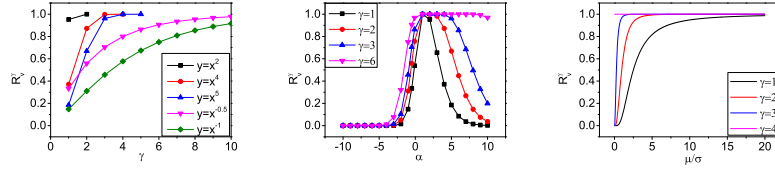
We first suppose x is uniformly distributed. Its PDF is given by

$$P(x) = \begin{cases} 0, & x < x_0, \\ \frac{1}{x_m - x_0}, & x_0 \leq x \leq x_m, \\ 0, & x > x_m. \end{cases} \quad (2.12)$$

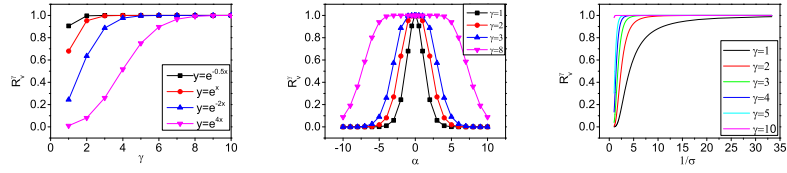
Uniform distribution provides the mathematical expectation of x : $\mu = \frac{x_m + x_0}{2}$ and its central moments:

$$M_k(x) = \begin{cases} 0, & k \text{ is odd,} \\ \frac{3^{k/2}}{k+1} \sigma^k, & k \text{ is even,} \end{cases} \quad (2.13)$$

where k is a positive integer; σ labels the standard deviation of x ; x_0 and x_m denote the lower and upper bounds of x , respectively. Detailed derivation of $M_k(x)$ is presented in Appendix A. Inserting the expression of central moments (Eq. (2.13)) into Eq. (2.8)



(a) Distributions for power-law function



(b) Distributions for exponential function

FIGURE 2.1: Distributions of the observation R_γ^γ with different parameters for power-law function: $y = x^\alpha$ (panel (a)) and exponential function: $y = e^{\alpha x}$ (panel (b)) by assuming input variable is uniformly distributed; To the left in both panels is the distribution of R_γ^γ versus the order index γ for two kinds of nonlinear relationship connecting y and x with constant values of parameter α ; to the middle is the dependence of R_γ^γ on the parameter α in considering different order contributions of uncertainty in x ; to the right is the association of R_γ^γ with the distribution parameter of variable x by setting $\alpha = 4$ for both panels. For a function with specified parameter α , the observation R_γ^γ only depends upon the ratio of μ to σ (power-law function) or σ (exponential function) for each specified γ . In the left and middle plots, $\mu/\sigma = 2$ in panel (a) and $\sigma^2 = 0.5$ in panel (b).

yields

$$V(y) = \sum_{i=1}^{\infty} \sum_{\substack{j=i \\ j=j+2}}^{\infty} C_{ij} \left(\left. \frac{d^i f}{dx^i} \right|_{x=\mu} \cdot \left. \frac{d^j f}{dx^j} \right|_{x=\mu} \right) \sigma^{i+j}, \quad (2.14)$$

where j is summed with an increment of 2 and C_{ij} is defined as follows.

If $i = j$

$$C_{ij} = \begin{cases} \frac{3^{(i+j)/2}}{(i+j+1) \cdot i! \cdot j!}, & i \text{ and } j \text{ are odd,} \\ \frac{3^{(i+j)/2} \cdot i \cdot j}{(i+j+1) \cdot (i+1)! \cdot (j+1)!}, & i \text{ and } j \text{ are even.} \end{cases} \quad (2.15)$$

Else

$$C_{ij} = \begin{cases} \frac{2 \cdot 3^{(i+j)/2}}{(i+j+1) \cdot i! \cdot j!}, & i \text{ and } j \text{ are odd,} \\ \frac{2 \cdot 3^{(i+j)/2} \cdot i \cdot j}{(i+j+1) \cdot (i+1)! \cdot (j+1)!}, & i \text{ and } j \text{ are even.} \end{cases} \quad (2.16)$$

The underlying results of R_γ^γ are displayed in Fig. 2.1, for two widely discussed nonlinear functions: power-law function and exponential function. It is noticed that R_γ^γ will be close to one while γ is large enough for both kinds of functions with different parameters, see left plots in both panels.

By inserting the functions under discussion: $y = x^\alpha$ and $y = e^{\alpha x}$, equation (2.14) can be separately updated by

$$\sum_{i=1}^{\infty} \sum_{\substack{j=i \\ j=j+2}}^{\infty} \frac{C_{ij}(\alpha!)^2}{(\alpha-i)! \cdot (\alpha-j)!} \mu^{2\alpha} \left(\frac{\sigma}{\mu}\right)^{i+j}, \quad (2.17)$$

and

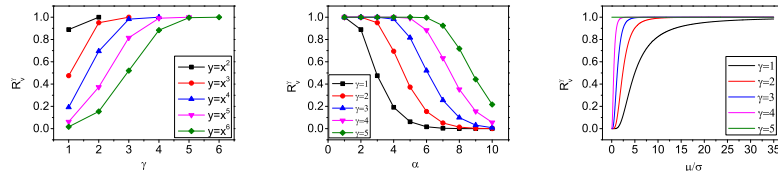
$$\sum_{i=1}^{\infty} \sum_{\substack{j=i \\ j=j+2}}^{\infty} e^{2\alpha\mu} C_{ij}(\alpha\sigma)^{i+j}, \quad (2.18)$$

which demonstrate that, for specified γ (thereby both i and j are specified) and α , the contribution of uncertainty in input x to the observation R_V^γ is embodied in the ratio of σ to μ for power-law relationships connecting y and x and in σ for exponential relationships. In Fig. 2.1, middle plots of both panels display the distribution of R_V^γ versus exponent α under specified μ/σ (for power-law function) and fixed σ (for exponential function), by considering different order contributions of uncertainty in x . Right ones illustrate the dependence of R_V^γ on μ/σ ($1/\sigma$) for power-law function (exponential function) by, particularly, setting $\alpha = 4$. Regarding power-law function, γ should be large for large $|\alpha - 1|$ if μ/σ is constant, or for small μ/σ if α is constant, in order to make R_V^γ more closer to 1, i.e., to make the estimated variance in model output more closer to its exact value. For example, the contribution of uncertainty in x should be considered until up to the 6th order (to ensure $R_V^\gamma > 0.98$) in regions of $\alpha < 0.1$ and $\alpha > 8$. For the exponential relationship connecting y and x , R_V^γ is symmetric with respect to $\alpha = 0$. The contributions of higher orders of input uncertainty are non-ignorable when the functional form deviates from linear law and also σ is non-negligible. The statement is visually verified that the original approximation presented in Eq. (2.9), by only considering the contribution of the first order of input uncertainty, can successfully estimate the uncertainty in output variable only when the input uncertainty is negligible or the model under discussion is almost linear. The contributions of higher order ($\gamma \geq 2$) of input uncertainty, however, are non-ignorable for models behaving non-linearly in the neighbourhood of input variable when σ (for the exponential association connecting input and output variables) or σ/μ (for the power-law association connecting input and output variables) is non-negligible.

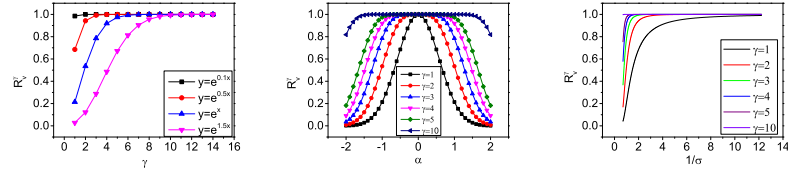
2.2.2 Normal distribution

Optionally, specify the uncertainty in x by normal (Gaussian) distribution. Its PDF is given by

$$P(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \quad x \in (-\infty, +\infty), \quad (2.19)$$



(a) Distributions for power-law function



(b) Distributions for exponential function

FIGURE 2.2: Distributions of the observation R_γ^γ versus different parameters for power-law function: $y = x^\alpha$ (panel (a)) and exponential function: $y = e^{\alpha x}$ (panel (b)) by assuming input variable is normally distributed; To the left in both panels is the distribution of R_γ^γ with order index γ for two kinds of functional form connecting y and x , by considering different values of parameter α ; to the middle is the association of R_γ^γ with parameter α under considering different order contributions of uncertainty in x ; to the right is the dependence of R_γ^γ upon the distribution parameter of variable x by setting $\alpha = 5$ in panel (a) and $\alpha = 1$ in panel (b), respectively. For a function with specified α , the observation R_γ^γ correlates with the ratio of μ to σ (power-law function) or with σ (exponential function) for each fixed γ . In the left and middle plots, $\mu/\sigma = 2$ in panel (a) and $\sigma = 1$ in panel (b).

which provides

$$M_k(x) = \begin{cases} 0, & k \text{ is odd,} \\ \sigma^k (k-1)!!, & k \text{ is even,} \end{cases} \quad (2.20)$$

where $(\star)!!$ is the double factorial with definition given by

$$(\star)!! = \begin{cases} \star \cdot (\star - 2) \cdot (\star - 4) \cdots 5 \cdot 3 \cdot 1, & \star > 0 \text{ odd,} \\ \star \cdot (\star - 2) \cdot (\star - 4) \cdots 6 \cdot 4 \cdot 2, & \star > 0 \text{ even,} \\ 1, & \star = -1, 0. \end{cases} \quad (2.21)$$

Associations of the observation R_γ^γ with order index γ , exponent parameter α , as well as the input uncertainty that is specified by μ and σ , are presented in Fig. 2.2. Panel (a) illustrates the data for the power-law relationship connecting y and x . Panel (b) displays the results for the exponential relationship. The observation R_γ^γ will be close to one when γ is large enough for each constant α , see left plots in both panels. We can notice from the middle plots that the contributions of higher orders of input uncertainty should be considered for highly nonlinear functions. Right plots in both panels suggest that R_γ^γ of $\gamma = 1$ can reach 1 only when the ratio of μ to σ is very large for power-law function ($y = x^5$ as an example) or σ is very small for exponential function ($y = e^x$ as an example). Typically, when the input variable is normally distributed, parameter α

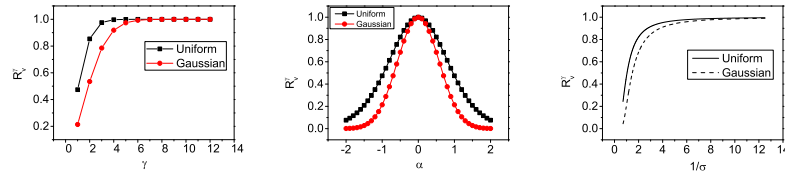


FIGURE 2.3: Results associated with the observation R_γ^γ for exponential relationship connecting y and x : $y = e^{\alpha x}$ by assuming input x is uniformly and normally distributed, respectively. To the left is the distribution of R_γ^γ versus order index γ by keeping $\alpha = \sigma = 1$; To the middle denotes the dependence of R_γ^γ upon parameter α in fixing $\gamma = \sigma = 1$; To the right is the correlation between R_γ^γ and σ by setting $\gamma = \alpha = 1$.

of power-law function should be a positive integer to enable the mean and variance of output y calculable.

To understand the influence of different distribution laws specifying the uncertainty in input variable on the estimation of output variance, a comparison of uniform and normal distributions is provided for the observation R_γ^γ by keeping both μ and σ the same for two kinds of law, see Fig. 2.3. The exponential relationship connecting y and x : $y = e^{\alpha x}$, as an example, is discussed here. Results suggest that, when both α and σ are specified, uniformly distributed input variable will drive the evaluated output uncertainty more close to its exact value, compared to normally distributed one, in considering variance contributions of the same order.

2.3 Generalisation of the analytic formula

The analytic formula Eq. (2.8) for variance propagation is valid only for the model of a single input variable. However, in mathematical modeling, the output quantity often depends upon two or more input variables of uncertainty. In the following, the analytic formula is generalised to the general case of more than one input variable. Recall the real model of n independent input variables of the form

$$y = f(\mathbf{x}) = f(x_1, x_2, \dots, x_n). \quad (2.22)$$

By the use of multi-variate Taylor series, the model response y can be expanded at the central point $(\{\mu\}) = (\mu_1, \mu_2, \dots, \mu_n)$ of input vector as

$$\begin{aligned}
 y = f(\{\mu\}) &+ \sum_{t=1}^n \sum_{i=1}^{\infty} \frac{(x_t - \mu_t)^i}{i!} \left(\frac{\partial^i f}{\partial x_t^i} \right) (\{\mu\}) + \sum_{\substack{t,s=1 \\ t < s}}^n \sum_{i_t, i_s=1}^{\infty} \frac{(x_t - \mu_t)^{i_t} (x_s - \mu_s)^{i_s}}{i_t! \cdot i_s!} \\
 &\times \left(\frac{\partial^{i_t+i_s} f}{\partial x_t^{i_t} \partial x_s^{i_s}} \right) (\{\mu\}) + \dots + \sum_{i_1, \dots, i_n=1}^{\infty} \frac{(x_1 - \mu_1)^{i_1} \dots (x_n - \mu_n)^{i_n}}{i_1! \dots i_n!} \left(\frac{\partial^{i_1+\dots+i_n} f}{\partial x_1^{i_1} \dots \partial x_n^{i_n}} \right) (\{\mu\}),
 \end{aligned} \tag{2.23}$$

where μ_i denotes the mathematical expectation of input x_i , as before. Reviewing the mathematical definition of variance, $V(y)$ is analogously calculated as

$$V(y) = \sum_{\substack{i_1 \dots i_n=0 \\ j_1 \dots j_n=0}}^{\infty} \frac{F_{(x_1)^{i_1} j_1 \dots (x_n)^{i_n} j_n}}{A(i_1, \dots, i_n, j_1, \dots, j_n)} \left(\frac{\partial^{i_1+\dots+i_n} f}{\partial x_1^{i_1} \dots \partial x_n^{i_n}} \cdot \frac{\partial^{j_1+\dots+j_n} f}{\partial x_1^{j_1} \dots \partial x_n^{j_n}} \right) (\{\mu\}), \tag{2.24}$$

with

$$A(\dots) = i_1! \dots i_n! \cdot j_1! \dots j_n!, \tag{2.25}$$

and

$$F_{(x_1)^{i_1} j_1 \dots (x_n)^{i_n} j_n} = M_{i_1+j_1}(x_1) \dots M_{i_n+j_n}(x_n) - M_{i_1}(x_1) M_{j_1}(x_1) \dots M_{i_n}(x_n) M_{j_n}(x_n). \tag{2.26}$$

Equation (2.24) specifies the variance propagation from input variables to the output one for a general model, by assuming input variables are statistically independent of each other.

2.4 Applications to the analysis of complex systems

The power grid system and economic system are considered in this section as examples of the application of our generalised analytic formula, Eq. (2.24), in the uncertainty and sensitivity analysis of complex systems. These two systems play extremely important roles in modern society. Their reliability analyses have attracted many researchers' interest.

Concerning the topic of sensitivity analysis, someone is of most interest to the sensitivity indices. They allow one to quantitatively interpret the importance of individual factors and of their interaction effects in the estimation of model response. Recalling the analytic formula of variance propagation in general models (see Eq. (2.24)), partial variance

contributions of different dimensionality in Eq. (1.33) can be computed through

$$V_t = \sum_{i,j=0}^{\infty} \frac{F_{(x_t)^{ij}}}{A(i,j)} \left(\frac{\partial^i f}{\partial x_t^i} \times \frac{\partial^j f}{\partial x_t^j} \right) (\{\mu\}), \quad (2.27)$$

$$V_{ts} = \sum_{i,j,k,l=0}^{\infty} \frac{F_{(x_t)^{ij}(x_s)^{kl}}}{A(i,j,k,l)} \left(\frac{\partial^{i+k} f}{\partial x_t^i \partial x_s^k} \cdot \frac{\partial^{j+l} f}{\partial x_t^j \partial x_s^l} \right) (\{\mu\}) - V_t - V_s, \quad (2.28)$$

$$\begin{aligned} V_{tsu} &= \sum_{i,j,k,l,p,q=0}^{\infty} \frac{F_{(x_t)^{ij}(x_s)^{kl}(x_u)^{pq}}}{A(i,j,k,l,p,q)} \left(\frac{\partial^{i+k+p} f}{\partial x_t^i \partial x_s^k \partial x_u^p} \cdot \frac{\partial^{j+l+q} f}{\partial x_t^j \partial x_s^l \partial x_u^q} \right) (\{\mu\}) \\ &\quad - V_t - V_s - V_u - V_{ts} - V_{tu} - V_{su}, \quad (2.29) \\ &\vdots \end{aligned}$$

The sensitivity indices of different orders are then quantified accordingly by

$$\begin{aligned} s_t &= \frac{V_t}{V(y)}, \quad s_{ts} = \frac{V_{ts}}{V(y)}, \quad s_{tsu} = \frac{V_{tsu}}{V(y)}, \quad \dots, \\ s_{Tt} &= s_t + \sum_{s=1; s \neq t}^n s_{ts} + \sum_{s,k=1; s > k; s, k \neq t}^n s_{tsk} + \dots \end{aligned} \quad (2.30)$$

2.4.1 Power grid system

In our daily life, the main form of energy that we think of is the power from electricity. Typically, a complex distribution system used for the transmission of electric power is called the power grid. The electricity power grid is a physical system that delivers electricity from the place where it is generated to the site where it is used. Consequently, a power grid system can be segmented into three sections: electric power generation, supply, and transmission, which work together to meet the basic electricity demand of ordinary people. Uncertainty and reliability analysis of the power grid system has been carried out since 1994 by using MC methods [215]. In general, the assessment of power grid system reliability is divided into two aspects: system adequacy and system security which are, respectively, related to the steady-state operation of system and to the ability of system to withstand sudden natural disturbances or to avoid attack.

Facing the increased global energy consumption, electric utility companies are striving to generate wind power to meet the growing electricity demand. Wind power, as a typical example of renewable energy sources, is plentiful, widely distributed, clean, produces no greenhouse gas emissions during operation, consumes no water, and uses little land [216]. In this part, a framework on the uncertainty and sensitivity analysis of the actual wind power output, denoted as P_d , is established by using the above established analytic method.

In theory, the wind output power of the wind turbine generator is always the same when the wind speed is specified. In reality, however, the output power for a fleet wind turbine generator of the same type exhibits considerable variations even when they are operating at the same wind speed [217, 218]. Consequently, P_d is often considered as a random variable that is characterised by two parameters: x and ε through a functional relationship:

$$P_d(x) = f(x) + \varepsilon, \quad (2.31)$$

where x indicates the wind speed, $f(x)$ the deterministic power output from a wind turbine generator, which acts as a function of the wind speed:

$$f(x) = \begin{cases} 0, & 0 \leq x \leq v_{ci}, \\ (A + Bx + cx^2) * P_r, & v_{ci} \leq x \leq v_r, \\ P_r, & v_r \leq x \leq v_{co}, \\ 0, & x \geq v_{co}, \end{cases} \quad (2.32)$$

and ε the variation of the power output, obeying normal distribution with $\mu = 0$ and $\sigma = 0.1P_r$ [219–222]. v_{ci} is called the cut-in wind speed at which the turbine first starts to rotate and generate power, v_r the rated wind speed at which the rated power P_r (the power output limit that the electrical generator is capable of) is reached, and v_{co} the cut-out wind speed at which the wind turbine generator is shut down for safety reasons, as shown in Fig. 2.4. Following the reference [222], $v_{ci} = 3\text{ms}^{-1}$ and $v_r = 12\text{ms}^{-1}$. The constants A , B , and C are determined by v_{ci} and v_r through equations [219]

$$A = \frac{1}{(v_{ci} - v_r)^2} \left[v_{ci}(v_{ci} + v_r) - 4v_{ci}v_r \left(\frac{v_{ci} + v_r}{2v_r} \right)^3 \right], \quad (2.33)$$

$$B = \frac{1}{(v_{ci} - v_r)^2} \left[4(v_{ci} + v_r) \left(\frac{v_{ci} + v_r}{2v_r} \right)^3 - (3v_{ci} + v_r) \right], \quad (2.34)$$

$$C = \frac{1}{(v_{ci} - v_r)^2} \left[2 - 4 \left(\frac{v_{ci} + v_r}{2v_r} \right)^3 \right]. \quad (2.35)$$

Substituting the values of v_{ci} and v_r , we get $A = 0.1215$, $B = -0.0784$ and $C = 0.0126$.

A systematic framework on reliability evaluation of P_d is performed by limiting x to the closed interval $[v_{ci}, v_r]$. With Eq. (2.32), the actual power output can be updated as

$$P_d(x) = (A + Bx + cx^2)P_r + \varepsilon. \quad (2.36)$$

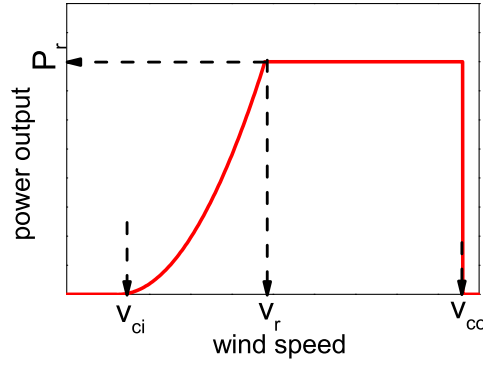
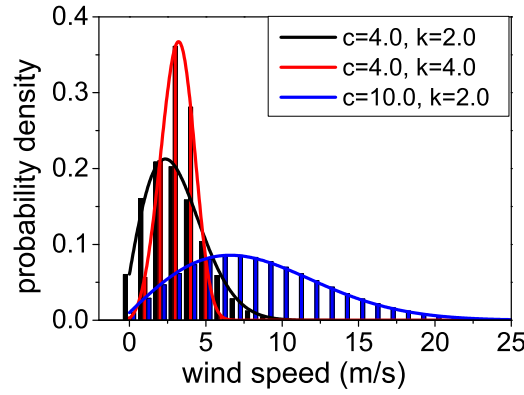


FIGURE 2.4: Typical wind turbine power output with steady wind speed.

FIGURE 2.5: The probability density distribution of wind speed for different combinations of parameters c and k .TABLE 2.1: Parameters associated with the uncertainty in wind speed x .

	μ_x	$V(x)$	$M_3(x)$	$M_4(x)$
$c = 4, k = 2$	4.79	1.97	3.03	16.09
$c = 4, k = 4$	4.1	0.5	0.21	0.73
$c = 10, k = 2$	7.47	6.03	0.37	69.63

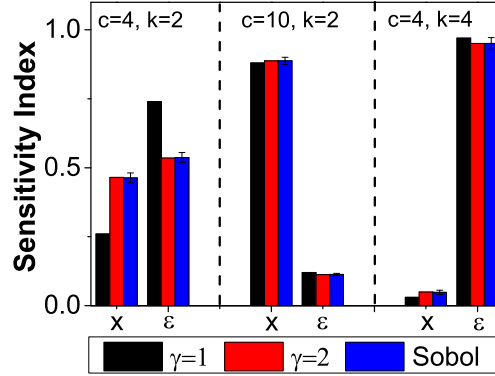
In the power grid system, the uncertainty of the wind speed x is often represented by a Weibull distribution of two parameters as

$$P(x) = \frac{k}{c} \left(\frac{x}{c}\right)^{k-1} e^{-\left(\frac{x}{c}\right)^k}; \quad (k > 0, x > 0, c > 1), \quad (2.37)$$

where c and k are the scale parameter and the shape parameter, respectively, [223]. Three different combinations of c and k are considered here. The underlying probability density function of the wind speed is displayed in Fig. 2.5. In our analysis, the wind speed x is limited to the real range [3m/s, 12m/s]. Parameters characterising the uncertainty in x are displayed in table 2.1.

TABLE 2.2: Analysis results of the power grid system in considering different values of γ (the highest order of Taylor expansion considered in our method).

	γ	$V(P_d)/P_r^2$	s_x	s_ε	$s_{x\varepsilon}$
$c = 4, k = 2$	1	0.01(35)	0.26(07)	0.73(93)	0
	2	0.01(87)	0.46(51)	0.53(49)	0
$c = 4, k = 4$	1	0.01(03)	0.03(01)	0.96(99)	0
	2	0.01(05)	0.04(93)	0.95(07)	0
$c = 10, k = 2$	1	0.08(28)	0.87(92)	0.12(08)	0
	2	0.08(91)	0.88(77)	0.11(23)	0

FIGURE 2.6: A comparison of our method with Sobol's one. Sobol's values are obtained by random sampling-based strategy (sample size: $M = 10000$), with error bars indicating the standard deviation of 100 runs.

The underlying results of uncertainty and sensitivity analysis are presented in table 2.2, by considering different values of the order index γ . Analysis results of $\gamma = 2$ also represent the exact values of our method as the highest order is two in the functional form. The obtained values suggest that, the uncertainty of model output P_d and its sensitivity with respect to individual input variables robustly depend upon the values of c and k . c and k work together to control the uncertainty in input x , thus directly influence the analysis results. For example, in the first case of $c = 4$ and $k = 2$, the second-order variance contribution of x plays a robust role in establishing the uncertainty in output P_d . While in the second case of $c = 4$ and $k = 4$, and also the third one of $c = 10$ and $k = 2$, the second-order variance contribution of x is ignorable. $s_{x\varepsilon} = 0$ indicates a vanishing interaction effect associated with x and ε . This is because the combination term between input variables is absent in the form of model function, see Eq. (2.36). A comparison of our method with Sobol's one is also presented in Fig. 2.6. The exact values obtained by our method are in good agreement with those provided by Sobol's method. This implies our method is equivalent to Sobol's one in the analysis of the model under consideration.

2.4.2 Economic system

Regarding the economic system, one of the oldest classic production scheduling models is the EOQ model. This model was developed by Harris in 1913 [224]. It aims at determining the economic order quantity that minimizes the total system cost. Some analyses about the uncertainty and sensitivity of this model have been proposed in Refs.[19, 168, 225]. However, the discussion of its reliability to each input parameter, especially to the interaction effects between different parameters, is still limited.

This subsection builds an intuitive insight into the uncertainty and reliability of EOQ model in terms of the analytic formula deduced before. In EOQ model, three types of cost are accounted: purchase cost, holding cost and ordering cost. The purchase cost is the variable cost of goods, equaling the product of the price of the units themselves, denoted as P , and the annual demand quantity, denoted as D . The price of the units themselves is assumed to be fixed, regardless of the number of demand quantity. The holding cost indicates the cost of holding units in inventory (management's cost of capital invested in the units, the cost of the space consumed by the units, taxes of insurance, etc). It is expressed as the product of the cost of holding unit, denoted as h , and the average quantity in inventory (between fully replenished and empty), denoted as $Q/2$. Q is the order quantity. The ordering cost represents all the costs associated with placing orders excluding the purchase cost. Suppose each order has a fixed cost K . The number of orders we need to make per year is D/Q . Hence the ordering cost is $K * D/Q$. The total system cost in EOQ model can then be represented as

$$TC = PD + \frac{DK}{Q} + \frac{hQ}{2}. \quad (2.38)$$

EOQ is the order quantity that minimizes the total system cost. It is easy to obtain the value of Q which determines the minimum point of TC , that is

$$Q^* = \sqrt{\frac{2DK}{h}}. \quad (2.39)$$

The uncertainty of Q^* , as well as its sensitivity with respect to independent input variables D , K , h and to their interactions, is quantified in this part by using our analytic method.

TABLE 2.3: Analysis results for the EOQ model while considering different values of γ .

$\gamma =$	$V(Q^*)$	s_D	s_K	s_h	s_{DK}	s_{Dh}	s_{Kh}	s_{DKh}
1	2119	0.385	0.307	0.307	0	0	0	0
2	2192	0.377	0.300	0.314	-0.002	0.006	0.005	0

Input variables are assumed to be independently and uniformly distributed in real ranges as follows [225]:

$$\begin{aligned}
900 \leq D \leq 1600 & \quad \text{unit per year,} \\
\$75 \leq K \leq \$125 & \quad \text{per order,} \\
\$1.5 \leq h \leq \$2.5 & \quad \text{per order and per year,}
\end{aligned} \tag{2.40}$$

which yields

$$\begin{aligned}
\mu_D &= 1250, & \mu_K &= 100, & \mu_h &= 2, \\
V(D) &= 40833.333, & V(K) &= 208.333, & V(h) &= 0.083.
\end{aligned} \tag{2.41}$$

Substituting the functional form and distribution laws of model inputs into Eq. (2.11) gives the exact value of output uncertainty: $V(Q^*) = 2195$. The exact values of sensitivity analysis are determined as

$$\begin{aligned}
s_D &= 0.377, & s_K &= 0.300, & s_h &= 0.314, \\
s_{DK} &= -0.002, & s_{Dh} &= 0.006, & s_{Kh} &= 0.005, & s_{DKh} &= 0.000.
\end{aligned} \tag{2.42}$$

Sensitivity analysis results are listed in table 2.3 for different values of γ . While $\gamma = 2$, the analysis results are almost equal to the exact values. This suggests that the contributions of input uncertainty of third- or higher-order can be neglected and the analysis results of $\gamma = 2$ can truly represent the reliability of EOQ model. Results state that all three parameters are important in establishing the uncertainty in model response. Furthermore, the interaction effect between each two parameters also contributes a small part to the uncertainty in response Q^* . $s_{DK} = -0.002$ says that the interaction effect between D and K will result in a small decrease in the variance of model response. A comparison of our analysis results with Sobol's values is shown in Fig. 2.7. The convergent results of our method are almost the same as Sobol's numbers.

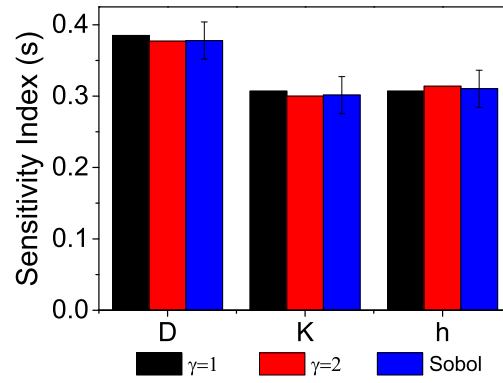


FIGURE 2.7: A comparison of our method with Sobol's one. Sobol's values are obtained by random sampling-based strategy (sample size: $M = 10000$), with error bars indicating the standard deviation of 100 runs.

Chapter 3

The analytic analysis for models with correlated input variables

Over the past few decades, the uncertainty and sensitivity analysis of models is mainly performed by assuming that input variables are independent of each other. A problem is often arisen in the development of the methodology for sensitivity analysis: how to interpret the sensitivity measures when input variables are non-independent? Recently, the interest in uncertainty and sensitivity analysis is increasing in the presence of correlated input variables, facing the existence of correlated factors in practical applications. In this chapter, we focus on the extension of the analytic framework established in the previous chapter. The extended framework allows one to understand, from a theoretical point of view, the variance propagation from correlated input variables to the model response and also the sensitivity of model response with respect to input independence and correlations.

3.1 Variance propagation

Analogously, consider a general model of the form $y = f(\mathbf{x})$ with $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ labeling the input vector of n -dimensional variables. Recalling the multi-variate Taylor series at the central point of input vector, presented in Eq. (2.23), the mathematical expectation of the model response can be represented in the presence of input correlations

as

$$\begin{aligned}
E(y) &= f(\{\mu\}) + \sum_{t=1}^n \sum_{i_t=1}^{\infty} \frac{M_{i_t}(x_t)}{A(i_t)} \left(\frac{\partial^{i_t} f}{\partial x_t^{i_t}} \right) (\{\mu\}) + \sum_{\substack{t,s=1 \\ t < s}}^n \sum_{i_t, i_s=1}^{\infty} \frac{M_{i_t, i_s}(x_t^{i_t}, x_s^{i_s})}{A(i_t, i_s)} \left(\frac{\partial^{i_t+i_s} f}{\partial x_t^{i_t} \partial x_s^{i_s}} \right) (\{\mu\}) \\
&\quad + \cdots + \sum_{i_1, \dots, i_n=1}^{\infty} \frac{M_{i_1, \dots, i_n}(x_1^{i_1}, \dots, x_n^{i_n})}{A(i_1, \dots, i_n)} \left(\frac{\partial^{i_1+\dots+i_n} f}{\partial x_1^{i_1} \dots \partial x_n^{i_n}} \right) (\{\mu\}) \\
&= \sum_{i_1, \dots, i_n=0}^{\infty} \frac{M_{i_1, \dots, i_n}(x_1^{i_1}, \dots, x_n^{i_n})}{A(i_1, \dots, i_n)} \left(\frac{\partial^{i_1+\dots+i_n} f}{\partial x_1^{i_1} \dots \partial x_n^{i_n}} \right) (\{\mu\}),
\end{aligned} \tag{3.1}$$

where $A(\dots)$ is defined by Eq. (2.25), $\{\mu\}$ the central point in input space, and M_{i_1, \dots, i_n} the multivariate central moment with n different variables. Mathematically, the multivariate central moment is defined as

$$\begin{aligned}
M_{i_1, \dots, i_n}(x_1^{i_1}, \dots, x_n^{i_n}) &= E[(x_1 - \mu_1)^{i_1} \dots (x_n - \mu_n)^{i_n}] \\
&= \int (x_1 - \mu_1)^{i_1} \dots (x_n - \mu_n)^{i_n} P(\mathbf{x}) d\mathbf{x},
\end{aligned} \tag{3.2}$$

with $P(\mathbf{x})$ indicating the joint PDF of \mathbf{x} . In the absence of input correlations, $P(\mathbf{x})$ is simplified as

$$P(\mathbf{x}) = \prod_{i=1}^n P_i, \tag{3.3}$$

where P_i denotes the PDF of input x_i , as before. In the existence of correlations among input variables, the variance of model response is derived by the use of Eqs. (2.23) and (3.1) as

$$\begin{aligned}
V(y) &= \sum_{\substack{i_1, \dots, i_n=0 \\ j_1, \dots, j_n=0}}^{\infty} \frac{1}{A(i_1, \dots, i_n, j_1, \dots, j_n)} \left(\frac{\partial^{i_1+\dots+i_n} f}{\partial x_1^{i_1} \dots \partial x_n^{i_n}} \cdot \frac{\partial^{j_1+\dots+j_n} f}{\partial x_1^{j_1} \dots \partial x_n^{j_n}} \right) (\{\mu\}) \\
&\times \left[M_{i_1+j_1, \dots, i_n+j_n}(x_1^{i_1+j_1}, \dots, x_n^{i_n+j_n}) - M_{i_1, \dots, i_n}(x_1^{i_1}, \dots, x_n^{i_n}) M_{j_1, \dots, j_n}(x_1^{j_1}, \dots, x_n^{j_n}) \right].
\end{aligned} \tag{3.4}$$

Apparently, in the absence of input correlations, the above equation is coincident with Eq. (2.24) that explains the variance propagation from independent input variables to the model response.

The concept of complete variance decomposition, presented in Eq. (1.33), is proposed by assuming input independence. Its form is also valid for the correlated case. In the presence of correlated input variables, however, partial variance contributions with dimensionality larger than 1 are contributed not only by the coupling items presented in the functional form of the model under discussion (for independent case), but also by the

input correlations. Regarding the situation of correlated inputs, the impact of a single variable can be represented as the sum of contributions provided by its correlations with the remaining variables and by its independence, respectively. Based on the description, each fractional variance contribution included in the original variance decomposition can be divided into three sections: independent variance contribution (labeled by superscript U), correlated variance contribution (labeled by superscript C), and coupling variance contribution (labeled by superscript UC). Mathematically, the output variance can be decomposed in the presence of input correlations as

$$V(y) = \sum_{i=1}^n (V_i^U + V_i^C + V_i^{UC}) + \sum_{\substack{i,j=1 \\ i < j}}^n (V_{ij}^{U_p} + V_{ij}^{C_p} + V_{ij}^{UC_p}) + \cdots + (V_{12\cdots n}^{U_q} + V_{12\cdots n}^{C_q} + V_{12\cdots n}^{UC_q}), \quad (3.5)$$

where $p \in \{i, j\}$, $q \in \{1, 2, \dots, n\}$, and

$$V_i = V_i^U + V_i^C + V_i^{UC}, \quad (3.6)$$

$$V_{ij} = V_{ij}^{U_p} + V_{ij}^{C_p} + V_{ij}^{UC_p}, \quad (3.7)$$

$$\vdots$$

$$V_{12\cdots n} = V_{12\cdots n}^{U_q} + V_{12\cdots n}^{C_q} + V_{12\cdots n}^{UC_q}. \quad (3.8)$$

V_i^U (V_i^C) is the variance contribution produced by the independent (correlated) section of x_i alone, $V_{ij}^{U_i}$ ($V_{ij}^{C_i}$) the contribution of the interaction between x_j and the independent (correlated) section of x_i , and so on up to $V_{12\cdots n}^{U_q}$ ($V_{12\cdots n}^{C_q}$) the contribution of the interaction associated with the independent (correlated) section of x_q and the rest variables. Coupling variance contributions are produced by the coupling effects between independent and correlated sections for individual input variables.

3.2 Estimation of sensitivity indices

Working within a probabilistic framework, variance-based sensitivity measures are defined on the bases of partial contributions presented in the variance decomposition of model response. In the determination of each partial variance contribution, higher-order covariance embodied in the analytic formula Eq. (3.4) should be concerned for nonlinear models. Consequently, it is necessary to specify the correlated and independent sections of single input variables for the confirmation of fractions contained in Eq. (3.5), whereby the importance of the independent section, correlated section and their coupling effect can be quantified for each individual input variable, in establishing the uncertainty of model response.

3.2.1 Generation of correlated variables

In probabilistic models the dependency between input variables is often represented by the CC that is defined as

$$\rho(x_i, x_j) = \frac{E[(x_i - \mu_i)(x_j - \mu_j)]}{\sigma_i \sigma_j}, \quad (3.9)$$

where $E[*]$ is the expectation operation by returning the average value of $*$. Eq. (3.9) is equivalent to Eq. (1.65) which defines the CC between an arbitrary input variable and the model response. For the sake of simplicity in writing, $\rho(x_i, x_j)$ is simplified as ρ_{ij} in the following discussion.

In the presence of correlations, an arbitrary variable can be represented as the sum of a correlated section and an independent section. Based on the description, we can rewrite the input vector as

$$\mathbf{x} = \mathbf{x}^C + \mathbf{x}^U \quad (3.10)$$

$$= (x_1^C, x_2^C, \dots, x_n^C)^T + (x_1^U, x_2^U, \dots, x_n^U)^T. \quad (3.11)$$

The correlated sections x_i^C with $i = 1, 2, \dots, n$ indicate the correlations of x_i with the remaining variables. By using the linear correlation model (Eq. (3.9)), \mathbf{x}^C can be generated as

$$\mathbf{x}^C = \mathbf{A}\mathbf{x}, \quad (3.12)$$

where \mathbf{A} is an $n \times n$ nonsymmetric hollow matrix (diagonal elements are all equal to zero). The independent sections x_i^U with $i = 1, 2, \dots, n$ denote the independence of x_i . They are often represented by newly introduced random variables. \mathbf{x}^U can be expressed as

$$\mathbf{x}^U = \mathbf{C}\mathbf{r}, \quad (3.13)$$

where \mathbf{C} is an $n \times n$ diagonal matrix and $\mathbf{r} = (r_1, r_2, \dots, r_n)^T$ an n -dimensional vector of random variables. \mathbf{r} must follow the same type distributions as \mathbf{x} and also satisfy

$$\boldsymbol{\mu}(\mathbf{r}) = \mathbf{C}^{-1}(\boldsymbol{\mu} - \mathbf{A}\boldsymbol{\mu}), \quad (3.14)$$

and

$$\boldsymbol{\sigma}(\mathbf{r}) = \boldsymbol{\sigma}, \quad (3.15)$$

in which, \mathbf{C}^{-1} is the inverse of matrix \mathbf{C} , $\boldsymbol{\mu}(\mathbf{r}) = (\mu(r_1), \mu(r_2), \dots, \mu(r_n))^T$ the mean vector of newly introduced random variables, $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$ the mean vector of input variables, and similarly for $\boldsymbol{\sigma}(\mathbf{r})$ and $\boldsymbol{\sigma}$ the vectors of standard deviations. Entries in matrices \mathbf{A} and \mathbf{C} : $\{a_{ij}, c_{ii}; i, j = 1, 2, \dots, n\}$ are called coefficients that specify the

correlated and independent sections of input variables. They are determined by given pairwise correlations through equations

$$a_{ij} = \frac{\sigma_i}{\sigma_j} \left[1 - \sum_{\substack{k < l, \\ k, l \neq i}} \rho_{kl}^2 (1 - \sum_{\substack{h < q, \\ h > k; h, q \neq i, l}} \rho_{hq}^2) + 2 \sum_{\substack{k < l < h, \\ k, l, h \neq i}} \rho_{kl} \rho_{kh} (\rho_{lh} - \sum_{q \neq i, l, h; q > k} \rho_{lq} \rho_{hq}) \right]^{-1} \\ \times \left[\rho_{ij} (1 - \sum_{\substack{k < l, \\ k, l \neq i, j}} \rho_{kl}^2 + 2 \sum_{\substack{k < l < h, \\ k, l, h \neq i, j}} \rho_{kl} \rho_{kh} \rho_{lh}) - \sum_{k \neq i, j} \rho_{ik} \rho_{jk} (1 - \sum_{\substack{h < q, \\ h, q \neq i, j, k}} \rho_{hq}^2) \right. \\ \left. + \sum_{\substack{k < l, \\ k, l \neq i, j}} (\rho_{ik} \rho_{jl} + \rho_{il} \rho_{jk}) (\rho_{kl} - \sum_{h \neq i, j, k, l} \rho_{kh} \rho_{lh}) \right], \quad (3.16)$$

and

$$c_{ii} = \left[1 - \sum_{\substack{k < l, \\ k, l \neq i}} \rho_{kl}^2 (1 - \sum_{\substack{h < q, \\ h > k; h, q \neq i, l}} \rho_{hq}^2) + 2 \sum_{\substack{k < l < h, \\ k, l, h \neq i}} \rho_{kl} \rho_{kh} (\rho_{lh} - \sum_{q \neq i, l, h; q > k} \rho_{lq} \rho_{hq}) \right]^{-1/2} \\ \times \left[1 - \sum_{k < l} \rho_{kl}^2 (1 - \sum_{\substack{h < q, \\ h > k; h, q \neq l}} \rho_{hq}^2) + 2 \sum_{k < l < h} \rho_{kl} \rho_{kh} (\rho_{lh} - \sum_{q \neq l, h; q > k} \rho_{lq} \rho_{hq}) \right]^{\frac{1}{2}}, \quad (3.17)$$

where the fifth- and higher-order terms are neglected. The above expressions, derived according to the analysis of simple cases as shown in Appendix B, constitute essential ingredients of quantifying sensitivity measures associated with correlated section, independent section, and their coupling effect, for each single input variable.

If two variables, say x_i and x_j , simultaneously correlate with a third variable, say x_k , the correlation between x_i and x_j is then not arbitrary assigned. With given correlation coefficients ρ_{ik} and ρ_{jk} , x_i and x_j then can be formulated based on x_k as

$$x_i = \rho_{ik} \frac{\sigma_i}{\sigma_k} x_k + \sqrt{1 - \rho_{ik}^2} u_i, \quad (3.18)$$

and

$$x_j = \rho_{jk} \frac{\sigma_j}{\sigma_k} x_k + \sqrt{1 - \rho_{jk}^2} u_j, \quad (3.19)$$

where u_i and u_j are newly introduced variables that are independent of x_k . Inserting the above formulations into Eq. (3.9) yields

$$\rho_{ij} = \rho_{ik} \rho_{jk} + \sqrt{(1 - \rho_{ik}^2)(1 - \rho_{jk}^2)} \rho(u_i, u_j). \quad (3.20)$$

In statistics, $\rho(u_i, u_j)$, labeling the linear correlation between variables u_i and u_j , can be as large as 1 (total positive correlation) and as small as -1 (total negative correlation). This implies the correlation ρ_{ij} between x_i and x_j will be limited to the range

$$\left[\rho_{ik}\rho_{jk} - \sqrt{(1 - \rho_{ik}^2)(1 - \rho_{jk}^2)}, \rho_{ik}\rho_{jk} + \sqrt{(1 - \rho_{ik}^2)(1 - \rho_{jk}^2)} \right] \quad (3.21)$$

when the correlations of them with a third variable x_k are specified. The relationship $\rho_{ij} = \rho_{ik}\rho_{jk}$ holds iff the newly introduced variables u_i and u_j are set to be independent of each other.

3.2.2 Sensitivity indices

With help of the analytic formula (3.4) that explains the variance propagation in the presence of input correlations, the partial variance contributions of different dimensionality can be calculated by

$$V_i = \sum_{k,l=0}^{\infty} \frac{1}{k! \cdot l!} \left(\frac{\partial^k f}{x_i^k} \cdot \frac{\partial^l f}{\partial x_i^l} \right) (\{\mu\}) \cdot [M_{k+l}(x_i) - M_k(x_i)M_l(x_i)], \quad (3.22)$$

$$\begin{aligned} V_{ij} &= \sum_{k,l,p,q=0}^{\infty} \frac{1}{k! \cdot l! \cdot p! \cdot q!} \left(\frac{\partial^{k+p} f}{x_i^k x_j^p} \cdot \frac{\partial^{l+q} f}{\partial x_i^l \partial x_j^q} \right) (\{\mu\}) \left[M_{k+l,p+q}(x_i^{k+l}, x_j^{p+q}) - \right. \\ &\quad \left. M_{k,p}(x_i^k, x_j^p) M_{l,q}(x_i^l, x_j^q) \right] - V_i - V_j, \\ &\vdots \end{aligned} \quad (3.23)$$

where multivariate central moments with different variables are involved. The (k, l) central moment with any two different variables, say x_i and x_j , is defined as (simplified from Eq. (3.2))

$$M_{k,l}(x_i^k, x_j^l) = E \left[(x_i - \mu_i)^k (x_j - \mu_j)^l \right]. \quad (3.24)$$

It can be derived analytically by formulating one variable on the basis of another:

$$x_i = \rho_{ij} \frac{\sigma_i}{\sigma_j} x_j + \sqrt{1 - \rho_{ij}^2} u_i, \quad (3.25)$$

or

$$x_j = \rho_{ij} \frac{\sigma_j}{\sigma_i} x_i + \sqrt{1 - \rho_{ij}^2} u_j, \quad (3.26)$$

with $u_i(u_j)$ independent of $x_j(x_i)$ and taking the same standard deviation as $x_i(x_j)$. If $k \neq l$, the above two formulating strategies are equivalent in determining $M_{k,l}(x_i^k, x_j^l)$ only when x_i and x_j are normally distributed and hold the same standard deviation.

The k th-order central moment of \mathbf{x} is given by

$$M_{k_1, \dots, k_m}(\mathbf{x}) = E \left[\prod_{i=1}^m (x_i - \mu_i)^{k_i} \right], \quad (3.27)$$

where $k_1 + k_2 + \dots + k_m = k$. For normal distribution, we have

$$M_{k_1, \dots, k_m}(\mathbf{x} - \boldsymbol{\mu}) = \begin{cases} 0, & k \text{ is odd,} \\ \sum (\sigma_{ij} \sigma_{lh} \dots), & k \text{ is even,} \end{cases} \quad (3.28)$$

where the sum is taken over all allocations of the set $\{1, 2, \dots, k\}$ into $k/2$ pairs, and σ_{ij} the covariance of x_i and x_j . For example, for the fourth-order central moment ($k = 4$) with four different variables (calculated in detail in Appendix D), one sums the products of any two covariances:

$$E \left[\prod_{i=1}^4 (x_i - \mu_i) \right] = \sigma_{12} \sigma_{34} + \sigma_{13} \sigma_{24} + \sigma_{14} \sigma_{23}. \quad (3.29)$$

This yields $(k - 1)!!$ (double factorial) terms in the sum. When nonlinear terms of a variable are contained in the central moments, like $(x_i - \mu_i)^k$ in Eq. (3.24), one can expand it as the product of k variables and uses $\sigma_{ii} = \sigma_i^2$. Consequently, for normally distributed x_i and x_j with $\sigma_i = \sigma_j$, $M_{k,l}(x_i^k, x_j^l)$ is invariant in exchanging x_i and x_j , that is

$$M_{k,l}(x_i^k, x_j^l) = M_{l,k}(x_i^l, x_j^k). \quad (3.30)$$

This property (invariant in exchanging any two variables) could go for any order central moments with any number of different variables that follow a normal distribution and have the same standard deviation.

For uniformly distributed variables, however, the fourth-order central moment with four different variables is calculated as

$$E \left[\prod_{i=1}^4 (x_i - \mu_i) \right] = \sigma_{12} \sigma_{34} + \sigma_{13} \sigma_{24} + \sigma_{14} \sigma_{23} - \frac{6}{5} \Delta, \quad (3.31)$$

where Δ is a product of three covariances, dependent upon how to choose a variable to be unformulated in the calculating process. For example, when x_2 is selected to be unformulated, we will get

$$\Delta = \sigma_{12} \sigma_{23} \sigma_{24}. \quad (3.32)$$

Furthermore, if one focuses on the correlated, independent and coupling effects contained in a high-order covariance, see Eq. (3.24), x_i (x_j) should be formulated on the bases of all the remaining input parameters, as discussed before.

The total contributions to the variance of model response, provided by the independence, correlation, and their coupling effect are represented, for an arbitrary variable x_i , as

$$V_i^{\text{TU}} = V_i^{\text{U}} + \sum_{j \neq i}^n V_{ij}^{\text{U}_i} + \cdots + V_{12 \dots n}^{\text{U}_i}, \quad (3.33)$$

$$V_i^{\text{TC}} = V_i^{\text{C}} + \sum_{j \neq i}^n V_{ij}^{\text{C}_i} + \cdots + V_{12 \dots n}^{\text{C}_i}, \quad (3.34)$$

$$V_i^{\text{TUC}} = V_i^{\text{UC}} + \sum_{j \neq i}^n V_{ij}^{\text{UC}_i} + \cdots + V_{12 \dots n}^{\text{UC}_i}. \quad (3.35)$$

The sensitivity (or importance) measures can then be determined by

$$\begin{aligned} s_i^{\text{U}} &= \frac{V_i^{\text{U}}}{V(y)}, & s_i^{\text{C}} &= \frac{V_i^{\text{C}}}{V(y)}, & s_i^{\text{UC}} &= \frac{V_i^{\text{UC}}}{V(y)}, \\ s_i^{\text{TU}} &= \frac{V_i^{\text{TU}}}{V(y)}, & s_i^{\text{TC}} &= \frac{V_i^{\text{TC}}}{V(y)}, & s_i^{\text{TUC}} &= \frac{V_i^{\text{TUC}}}{V(y)}. \end{aligned} \quad (3.36)$$

The first three measures are called the main sensitivity indices which, separately, denote the importance of the independent section, correlated section, and their coupling effect, by neglecting the interaction effects among input parameters. The last three measures are called the total sensitivity indices which denote the importance of each corresponding part in considering the interaction effects of x_i with the remaining inputs. Similar to Sobol's indices, the main and total sensitivity indices are defined as

$$s_i = \frac{V_i}{V(y)} = s_i^{\text{U}} + s_i^{\text{C}} + s_i^{\text{UC}}, \quad (3.37)$$

$$S_{Ti} = \left(V_i + \sum_{j \neq i} V_{ij} + \cdots \right) / V(y) = s_i^{\text{TU}} + s_i^{\text{TC}} + s_i^{\text{TUC}}, \quad (3.38)$$

which evaluate the importance of individual input parameters x_i before and after considering the interaction effects among different input parameters, respectively.

With given correlation coefficients between any two input parameters, our method allows one to evaluate the importance of individual input parameters in the estimation of model output in both the absence and presence of input correlations. The newly introduced sensitivity indices also quantify the importance of input independence and correlations, allowing one to determine whether or not the input correlations should be considered in practice. Quite recently the original Sobol's sensitivity indices are generalised to deal with correlated inputs, by using hierarchically orthogonal functional decomposition [226, 227]. Compared to the generalised Sobol's indices, our indices are easier to understand and interpret.

3.3 Numerical examples and a practical application

In this section, analytic polynomial models, including one purely additive and three nonlinear ones, are taken as examples to illustrate the effectiveness and validation of our established analytic framework. A practical application of the method is also proposed to a deterministic HIV model. Ten involved parameters are then ranked according to their importance in establishing the uncertainty of the basic reproduction number R_0 .

3.3.1 Additive linear model

In the first example a purely additive model is investigated, with functional form given by

$$y = 2x_1 + x_2 + x_3, \quad (3.39)$$

where $(x_1, x_2, x_3) \sim N(\boldsymbol{\mu}, \Sigma)$ with mean vector $\boldsymbol{\mu} = (0, 0, 0)^T$ and covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & 2\rho_{13} \\ \rho_{12} & 1 & 2\rho_{23} \\ 2\rho_{13} & 2\rho_{23} & 4 \end{pmatrix}. \quad (3.40)$$

By the use of Eq. (3.4), we get the exact expression of the variance of model response:

$$V(y) = 9 + 4\rho_{12} + 8\rho_{13} + 4\rho_{23}, \quad (3.41)$$

which is constituted of the following fractional contributions of different dimensionality:

$$\begin{aligned} V_1 &= 4, & V_2 &= 1, & V_3 &= 4, & V_{12} &= 4\rho_{12}, \\ V_{13} &= 8\rho_{13}, & V_{23} &= 4\rho_{23}, & V_{123} &= 0. \end{aligned} \quad (3.42)$$

The vanishing nonlinear problem in the model function suggests the absence of coupling variance contributions but the presence of independent and correlated ones:

$$\begin{aligned} V_i^U &= c_{ii}^2 V_i, & V_{12}^{U_j} &= 0, & V_{13}^{U_k} &= 0, & V_{23}^{U_l} &= 0, \\ V_i^C &= (1 - c_{ii}^2) V_i, & V_{12}^{C_j} &= V_{12}, & V_{13}^{C_k} &= V_{13}, & V_{23}^{C_l} &= V_{23}, \\ V_{123}^{U_i} &= V_{123}^{C_i} = 0, \end{aligned} \quad (3.43)$$

TABLE 3.1: Uncertainty and sensitivity analysis results for linear additive model by assuming different correlations between input variables.

ρ	$V(y)$	x	s_i	s_i^U	s_i^C	s_i^{UC}	s_{Ti}	s_i^{TU}	s_i^{TC}	s_i^{TUC}
$\rho = 0$	9	x_1	0.444	0.444	0.0	0.0	0.444	0.444	0.0	0.0
		x_2	0.111	0.111	0.0	0.0	0.111	0.111	0.0	0.0
		x_3	0.444	0.444	0.0	0.0	0.444	0.444	0.0	0.0
$\rho_{12} = 0.8$	12.2	x_1	0.328	0.118	0.210	0.0	0.590	0.118	0.472	0.0
		x_2	0.082	0.030	0.052	0.0	0.344	0.030	0.314	0.0
		x_3	0.328	0.328	0.0	0.0	0.328	0.328	0.0	0.0
$\rho_{12} = -0.8$	5.8	x_1	0.690	0.248	0.442	0.0	0.138	0.248	-0.110	0.0
		x_2	0.172	0.062	0.110	0.0	-0.379	0.062	-0.441	0.0
		x_3	0.690	0.690	0.0	0.0	0.690	0.690	0.0	0.0
$\rho_{12} = 0.8$	17.8	x_1	0.225	0.072	0.152	0.0	0.629	0.072	0.557	0.0
$\rho_{13} = 0.5$		x_2	0.056	0.020	0.036	0.0	0.326	0.020	0.306	0.0
$\rho_{23} = 0.4$		x_3	0.225	0.169	0.056	0.0	0.539	0.169	0.371	0.0

where $i \in \{1, 2, 3\}$, $j \in \{1, 2\}$, $k \in \{1, 3\}$, $l \in \{2, 3\}$, and c_{ii} , specifying the independence of variable x_i , are determined via Eq. (3.17) as

$$c_{11} = (1 - \rho_{23}^2)^{-1/2}(1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23})^{1/2}, \quad (3.44)$$

$$c_{22} = (1 - \rho_{13}^2)^{-1/2}(1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23})^{1/2}, \quad (3.45)$$

$$c_{33} = (1 - \rho_{12}^2)^{-1/2}(1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23})^{1/2}, \quad (3.46)$$

in which $\{\rho_{12}^2, \rho_{13}^2, \rho_{23}^2\} \neq 1$.

The underlying sensitivity measures are provided in table 3.1 under considering different correlations between input variables. Results indicate the vanishing sensitivity indices associated with the coupling effect between input independence and correlations. In the absence of correlated input variables ($\rho = 0$), the main sensitivity indices sum up to one. By introducing input correlations, however, this summation could be smaller than one (with positive correlations) or larger than one (with negative correlations), contrary to the sum of the total sensitivity indices. A negative sensitivity index explains a negative partial variance contribution produced by the negative input correlation.

3.3.2 Nonlinear models

Trivariate model

In the second example a trivariate nonlinear model is considered, containing the linear, quadratic, and interaction terms:

$$y = 2x_1 + x_2^2 + 4x_1^2x_2 + x_1x_3, \quad (3.47)$$

where $(x_1, x_2, x_3) \sim N(\boldsymbol{\mu}, \Sigma)$ with mean vector $\boldsymbol{\mu} = (0, 0, 0)^T$ and covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}. \quad (3.48)$$

The variance $V(y)$ of model response can be similarly computed by the use of Eq. (3.4) as

$$V(y) = 55 + 48\rho_{12} + 2\rho_{12}\rho_{23} + 192\rho_{12}^2 + \rho_{13}^2, \quad (3.49)$$

which is generated by the partial variance contributions involving

$$\begin{aligned} V_1 &= 4, & V_2 &= 2, & V_3 &= 0, \\ V_{12} &= 48(1 + 4\rho_{12}^2 + \rho_{12}), & V_{13} &= 1 + \rho_{13}^2, & V_{23} &= 0, \\ V_{123} &= 2\rho_{12}\rho_{23}. \end{aligned} \quad (3.50)$$

The independent, correlated and coupling variance contributions divided from the main variance contributions V_1 and V_2 are stated as

$$\begin{aligned} V_1^U &= 4c_{11}^2, & V_1^C &= 4(1 - c_{11}^2), & V_1^{UC} &= 0, \\ V_2^U &= 2c_{22}^4, & V_2^C &= 2(1 - c_{22}^2)^2, & V_2^{UC} &= 4c_{22}^2(1 - c_{22}^2). \end{aligned} \quad (3.51)$$

Regarding the existent higher-order partial variance contributions, we have

$$\begin{aligned} V_{12}^{U_1} &= 48c_{11}^4, & V_{12}^{U_2} &= 48c_{22}^2, \\ V_{12}^{C_1} &= 48(1 - c_{11}^2)(1 - c_{11}^2 + \rho_{12} + 4\rho_{12}^2), & V_{12}^{C_2} &= 48(1 + 4\rho_{12}^2 + \rho_{12} - c_{22}^2), \\ V_{12}^{UC_1} &= 48(2 + \rho_{12} + 4\rho_{12}^2 - 2c_{11}^2)c_{11}^2, & V_{12}^{UC_2} &= 0, \\ V_{13}^{U_i} &= c_{ii}^2, & V_{13}^{C_i} &= 1 + \rho_{13}^2 - c_{ii}^2, & V_{13}^{UC_i} &= 0, \\ V_{123}^{U_j} &= 0, & V_{123}^{C_j} &= 2\rho_{12}\rho_{23}, & V_{123}^{UC_j} &= 0, \end{aligned} \quad (3.52)$$

where $i \in \{1, 3\}$, $j \in \{1, 2, 3\}$ and c_i are determined with Eq. (3.17). A detailed calculation process for the above items are presented in Appendix C. In table 3.2, the exact sensitivity indices are listed, showing a dominated influence of the interaction effect between x_1 and x_2 in the absence of input correlations. In the presence of input correlations, the independent, correlated and coupling effects provided by x_1 and x_2 are all significant in establishing the uncertainty of model response. The convergence of our analysis results with the order of Taylor expansion is also presented in Fig. 3.1. γ is the highest-order of Taylor expansion that is considered in our method. Analysis results are convergent at $\gamma = 3$ (the highest-order of Taylor expansion of the model output), implying the effects of any possible orders of Taylor expansion of the model output are

TABLE 3.2: Exact analytic results for uncertainty and sensitivity analysis of the first nonlinear model with different input correlations.

ρ	$V(y)$	x	s_i	s_i^U	s_i^C	s_i^{UC}	s_{Ti}	s_i^{TU}	s_i^{TC}	s_i^{TUC}
$\rho = 0$	55	x_1	0.073	0.073	0.0	0.0	0.964	0.964	0.0	0.0
		x_2	0.036	0.036	0.0	0.0	0.909	0.909	0.0	0.0
		x_3	0.0	0.0	0.0	0.0	0.018	0.018	0.0	0.0
$\rho_{12} = 0.5$	127	x_1	0.031	0.024	0.007	0.0	0.984	0.242	0.175	0.567
		x_2	0.016	0.009	0.001	0.006	0.960	0.292	0.662	0.006
		x_3	0.0	0.0	0.0	0.0	0.008	0.008	0.0	0.0
$\rho_{12} = -0.5$ $\rho_{13} = 0.6$	79.36	x_1	0.050	0.020	0.03	0.0	0.975	0.117	0.452	0.406
		x_2	0.025	0.009	0.004	0.012	0.932	0.382	0.538	0.012
		x_3	0.0	0.0	0.0	0.0	0.017	0.007	0.010	0.0
$\rho_{12} = 0.4$ $\rho_{13} = 0.5$ $\rho_{23} = 0.8$	105.81	x_1	0.038	0.028	0.01	0.0	0.981	0.291	0.166	0.524
		x_2	0.019	0.002	0.008	0.009	0.950	0.167	0.774	0.009
		x_3	0.0	0.0	0.0	0.0	0.018	0.003	0.015	0.0

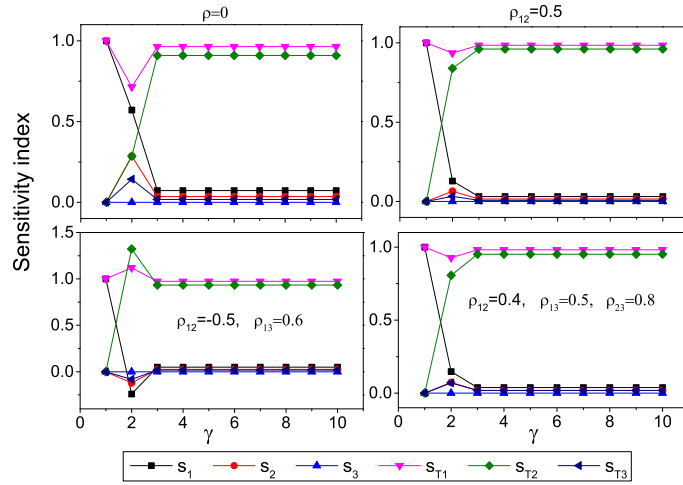


FIGURE 3.1: The convergence of analysis results with the order of Taylor expansion.

influential.

Fourvariate model

Another nonlinear model is designed based on four input variables as

$$y = x_1x_3 + x_2x_4, \quad (3.53)$$

where $(x_1, x_2, x_3, x_4) \sim N(\boldsymbol{\mu}, \Sigma)$ with mean vector $\boldsymbol{\mu} = (1, 2, 2, 1)^T$ and covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{12} & 1 & \rho_{23} & \rho_{24} \\ \rho_{13} & \rho_{23} & 1 & \rho_{34} \\ \rho_{14} & \rho_{24} & \rho_{34} & 1 \end{pmatrix}. \quad (3.54)$$

The total variance of model response is obtained by employing Eq. (3.4) as

$$V(y) = 12 + 4\rho_{12} + 4\rho_{13} + 8\rho_{14} + 2\rho_{23} + 4\rho_{24} + 4\rho_{34} + \rho_{13}^2 + \rho_{24}^2 + 2(\rho_{12}\rho_{34} + \rho_{14}\rho_{23}), \quad (3.55)$$

which is constituted of

$$\begin{aligned} V_1 &= 4, & V_2 &= 1, & V_3 &= 1, & V_4 &= 4, \\ V_{12} &= 4\rho_{12}, & V_{13} &= 1 + 4\rho_{13} + \rho_{13}^2, & V_{14} &= 8\rho_{14}, & V_{23} &= 2\rho_{23}, \\ V_{24} &= 1 + 4\rho_{24} + \rho_{24}^2, & V_{34} &= 4\rho_{34}, & V_{1234} &= 2(\rho_{12}\rho_{34} + \rho_{14}\rho_{23}). \end{aligned} \quad (3.56)$$

In the calculation of V_{1234} , the covariance $\text{cov}(x_1, x_2, x_3, x_4)$ of four variables is involved, whose derivation is presented in Appendix D. The form of model function (only involves the linear problem of each input) suggests the vanishing coupling effect in all partial variance contributions but the existent correlated and independent ones:

$$\begin{aligned} V_i^U &= c_{ii}^2 V_i, & V_i^C &= (1 - c_{ii}^2) V_i, \\ V_{13}^{U_j} &= c_{jj}^2, & V_{13}^{C_j} &= V_{13} - V_{13}^{U_j}, \\ V_{24}^{U_k} &= c_{kk}^2, & V_{24}^{C_k} &= V_{24} - V_{24}^{U_k}, \end{aligned} \quad (3.57)$$

where $i \in \{1, 2, 3, 4\}$, $j \in \{1, 3\}$ and $k \in \{2, 4\}$. The partial variance contributions V_{12} , V_{14} , V_{23} , V_{34} and V_{1234} are all contributed by input correlations. The coefficients c_i are determined with Eq. (3.17). Table 3.3 lists the exact analytic values of the underlying sensitivity indices. Data show a vanishing coupling effect between input independence and correlations. This because the nonlinear problems of single variables are absent in the form of model function. The convergence of our analysis results is displayed in Fig. 3.2, along the direction of the highest-order of Taylor expansion that is considered in our method. Values are convergent at $\gamma = 2$ (the highest-order of Taylor expansion of the model output).

TABLE 3.3: Exact analytic values of uncertainty and sensitivity analysis for the second nonlinear model by assuming uncorrelated and correlated inputs.

ρ	$V(y)$	x	s_i	s_i^U	s_i^C	s_i^{UC}	s_{Ti}	s_i^{TU}	s_i^{TC}	s_i^{TUC}
$\rho = 0$	12	x_1	0.333	0.333	0.0	0.0	0.417	0.417	0.0	0.0
		x_2	0.083	0.083	0.0	0.0	0.167	0.167	0.0	0.0
		x_3	0.083	0.083	0.0	0.0	0.167	0.167	0.0	0.0
		x_4	0.333	0.333	0.0	0.0	0.417	0.417	0.0	0.0
$\rho_{13} = 0.5$ $\rho_{24} = 0.8$	18.09	x_1	0.221	0.166	0.055	0.0	0.401	0.207	0.194	0.0
		x_2	0.055	0.020	0.035	0.0	0.323	0.040	0.283	0.0
		x_3	0.055	0.041	0.014	0.0	0.235	0.083	0.152	0.0
		x_4	0.221	0.080	0.141	0.0	0.489	0.100	0.389	0.0
$\rho_{12} = -0.5$ $\rho_{13} = 0.6$ $\rho_{14} = 0.4$	15.96	x_1	0.251	0.058	0.193	0.0	0.561	0.072	0.489	0.0
		x_2	0.063	0.030	0.033	0.0	0.0	0.060	-0.060	0.0
		x_3	0.063	0.024	0.039	0.0	0.298	0.049	0.249	0.0
		x_4	0.251	0.148	0.103	0.0	0.514	0.185	0.329	0.0
$\rho_{12} = -0.5, \rho_{13} = -0.4$ $\rho_{14} = 0.2, \rho_{23} = 0.3$ $\rho_{24} = 0.4, \rho_{34} = 0.4$	13.84	x_1	0.289	0.110	0.179	0.0	0.208	0.137	0.071	0.0
		x_2	0.072	0.032	0.040	0.0	0.150	0.065	0.085	0.0
		x_3	0.072	0.040	0.032	0.0	0.179	0.080	0.099	0.0
		x_4	0.289	0.121	0.168	0.0	0.699	0.152	0.547	0.0

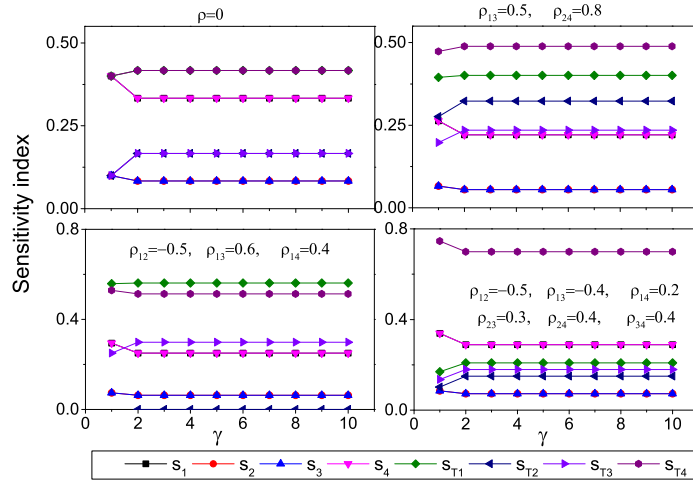


FIGURE 3.2: The convergence of analysis results with the order of Taylor expansion.

Ishigami function

The Ishigami function [228] has been extensively used as a benchmark for sensitivity analysis [229, 230]. Its functional form was defined as

$$y = \sin(x_1) + 7 \sin^2(x_2) + 0.1x_3^4 \sin(x_1), \quad (3.58)$$

where all input variables are uniformly distributed in the interval $[-\pi, \pi]$. The presence of correlation between x_2 and each of the rest does not influence the total variance of

TABLE 3.4: Exact analytic values of uncertainty and sensitivity analysis for Ishigami function by assuming uncorrelated and correlated inputs. In case 1 x_1 is generated based on x_3 , contrary to the second case where x_3 is generated based on x_1 .

ρ	$V(y)$	x	s_i	s_i^U	s_i^C	s_i^{UC}	s_{Ti}	s_i^{TU}	s_i^{TC}	s_i^{TUC}
$\rho_{13} = 0$	13.845	x_1	0.036	0.036	0.0	0.0	0.557	0.557	0.0	0.0
		x_2	0.442	0.442	0.0	0.0	0.442	0.442	0.0	0.0
		x_3	0.0	0.0	0.0	0.0	0.521	0.521	0.0	0.0
$\rho_{13} = 0.5$ (case 1)	12.971	x_1	0.039	0.044	0.039	-0.044	0.528	0.633	1.048	-1.153
		x_2	0.472	0.472	0.0	0.0	0.472	0.472	0.0	0.0
		x_3	0.0	—	—	—	0.489	—	—	—
$\rho_{13} = 0.5$ (case 2)	19.110	x_1	0.026	—	—	—	0.679	—	—	—
		x_2	0.321	0.321	0.0	0.0	0.321	0.321	0.0	0.0
		x_3	0.0	0.0	0.0	0.0	0.653	0.145	0.004	0.505

model response owing to zero partial variance contributions associated with the interaction between x_2 and the rest. Consequently, we just consider here the correlation between x_1 and x_3 . The results of analytic analysis are listed in table 3.4 by assuming independent and correlated input variables. Two formulating strategies are considered in the presence of correlation: x_1 is formulated on the basis of x_3 and vice versa. They are non-equivalent for the uncertainty and sensitivity analysis of the model under discussion as x_1 and x_3 are uniformly distributed.

In the first case, x_1 is formulated on the basis of x_3 as

$$x_1 = \rho_{13} \frac{\sigma_1}{\sigma_3} x_3 + \sqrt{1 - \rho_{13}^2} r_1, \quad (3.59)$$

where the newly introduced random variable r_1 is the element of \mathbf{r} that satisfies Eqs. (3.14) and (3.15). Sensitivity measures show a strong positive variance contribution produced by the interaction effect between x_3 and the correlated part of x_1 , as well as a very strong negative variance contribution caused by the interaction term involving x_3 and both correlated and independent sections of x_1 .

For the second case, we generate x_3 on the basis of x_1 as

$$x_3 = \rho_{13} \frac{\sigma_3}{\sigma_1} x_1 + \sqrt{1 - \rho_{13}^2} r_3, \quad (3.60)$$

where the random variable r_3 is the element of \mathbf{r} that satisfies Eqs. (3.14) and (3.15). Zero mean of x_1 leads to the nonexistence of sensitivity measures evaluating the main effect of x_3 . A dominated contribution to the variance of model response is produced by the interaction effect between x_1 and the coupling of independent section with the correlated one of x_3 .

Analysis results imply that the correlation between x_1 and x_3 , if exists, will play a crucial role in determining the model response. Sensitivity indices of correlated and

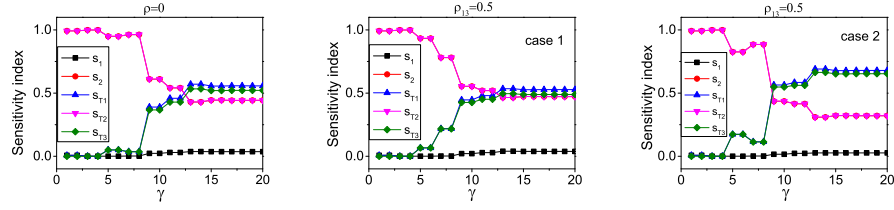


FIGURE 3.3: The convergence of analysis results with the order of Taylor expansion.

independent sections of x_3 are not indicated in the first case because x_3 is considered as a whole variable, analogous to the second case. The derivation process of partial variance contributions of different orders is presented in detail in Appendix E for both cases. The convergence of our analysis results is also discussed, see Fig. 3.3. For three situations, values are all convergent at $\gamma = 13$ (the highest-order of Taylor expansion that is considered in our method).

HIV model

The basic reproduction number, denoted as R_0 , is arguably the most important quantity in infectious disease epidemiology because it helps determine whether or not an infectious disease can spread through a population [231, 232]. R_0 is defined as the average number of new cases of an infection caused by one typical infected individual, in a population consisting of susceptibles only [233–235]. The first application of this metric in epidemiology was introduced by George MacDonald in 1952, who designed agent-based models of the spread of malaria [236]. Generally, the larger the value of R_0 , the harder it is to control the spreading of an epidemic. Typically, when $R_0 < 1$, the disease free equilibrium is locally asymptotically stable and the epidemic will die out in the long run, whereas if $R_0 > 1$, it is unstable and the epidemic will invade the population [237].

Consider a deterministic model of HIV-1 with vertical transmission (from an HIV-infected mother to her child) which was discussed in Ref. [84]. The basic reproduction number R_0 is represented by

$$R_0 = \frac{\beta_0(1-\gamma)\theta_d^2 + \beta_1 n_1 Q_0(\theta_d - \kappa) + \beta_2 n_2 \alpha Q_0 + (1-\gamma)(\kappa + \alpha)\beta_0 \theta_d}{\theta_d(\theta_d + \kappa)(\theta_d + \alpha)}. \quad (3.61)$$

Description and baseline values of parameters included in the above expression are presented in table 3.5.

To identify the importance of individual parameters in establishing the uncertainty of R_0 , each parameter is artificially increased and decreased by 10% of its baseline value.

TABLE 3.5: Description and baseline values of parameters for HIV/AIDS model, see Refs. [84, 238, 239].

Parameter	Symbol	Baseline value
Recruitment rate	Q_0	0.029
Birth rate of infective	β_0	0.03
Fraction of susceptible newborn from infective class	γ	0.4
Contact rate of susceptible with asymptomatic infective	β_1	0.2
Contact rate of susceptible with symptomatic infective	β_2	0.08
Number of sexual partners of susceptible with asymptomatic infective	n_1	2.0
Number of sexual partners of susceptible with symptomatic infective	n_2	2.0
Natural death rate	θ_d	0.02
Removal rate to symptomatic class	α	0.6
Rate of development to AIDS	κ	0.1

Furthermore, for simplicity, uncertainties of parameters are indicated by uniform distribution in their ranges of variation. The mathematical expectations and uncertainties of input parameters and output R_0 are presented in table 3.6. Regarding the uncertainty in R_0 , both independent and correlated situations are discussed. The underlying sensitivity analysis results are displayed in table 3.7. In our analysis, the first-order Taylor expansions are considered only, which explain 99.6% and 98.8% of the exact uncertainty (indicated by the standard deviation) of R_0 for independent and correlated situations, respectively. A ranking of input parameters is displayed in Fig. 3.4, according to the total sensitivity indices. Two values of γ (the highest-order of Taylor expansion that is considered in our method) are considered. Results suggest that the analysis results of $\gamma = 1$ are almost the same as those of $\gamma = 2$ for both the absence and presence of input correlations. This implies our analysis results are convergent at the first-order Taylor expansion for the model under discussion. In both the absence and presence of input correlations, parameters κ (rate of development to AIDS), β_2 (contact rate of susceptible with symptomatic infective), and n_2 (number of sexual partners of susceptible with symptomatic infective) are identified as the most influential, while parameters β_0 (the birth rate of infective) and γ (the fraction of susceptible newborn from infective class) are identified as negligible, in determining the basic reproduction number R_0 . This provides one an opportunity to effectively limit the spread of a disease by controlling the three most influential parameters, and to simplify an HIV model by neglecting the effects contributed by parameters β_0 and γ .

The distribution of our analysis results is also presented in Fig. 3.5, along the direction of the correlation between β_1 and n_1 , and the direction of the correlation between β_2 and n_2 . The underlying results suggest that the effect contributed by the correlation

TABLE 3.6: Uncertainty determination for both input and output parameters. The uncertainty in R_0 is obtained by considering the first-order Taylor expansion, which explains 99.6% and 98.8% of the exact uncertainty of R_0 for independent and correlated situations, respectively.

	parameter	μ (baseline value)	Range of variation	σ (uncertainty)
input	Q_0	0.029	[0.0261, 0.0319]	0.002
	β_0	0.03	[0.027, 0.033]	0.002
	γ	0.4	[0.36, 0.44]	0.023
	β_1	0.2	[0.18, 0.22]	0.012
	β_2	0.08	[0.072, 0.088]	0.005
	n_1	2	[1.8, 2.2]	0.115
	n_2	2	[1.8, 2.2]	0.115
	θ_d	0.02	[0.018, 0.022]	0.001
	α	0.6	[0.54, 0.66]	0.035
	κ	0.1	[0.09, 0.11]	0.006
output	R_0	1.429	–	0.227($\rho = 0$)
		1.432	–	0.252
				($\rho_{\beta_1 n_1} = 0.3, \rho_{\beta_2 n_2} = 0.5$)

TABLE 3.7: The convergent analysis results of the basic reproduction number R_0 for a deterministic HIV-1 model with vertical transmission by assuming independent and correlated input parameters.

x	$\rho = 0$	$\rho_{\beta_1 n_1} = 0.3, \rho_{\beta_2 n_2} = 0.5$		
	$s_{Ti}(= s_{Ti}^U)$	s_{Ti}	s_{Ti}^U	s_{Ti}^C
Q_0	0.101	0.082	0.082	0
β_0	0.002	0.002	0.002	0
γ	0.001	0.001	0.001	0
β_1	0.025	0.032	0.018	0.014
β_2	0.227	0.363	0.138	0.225
n_1	0.025	0.032	0.018	0.014
n_2	0.227	0.363	0.138	0.225
θ_d	0.122	0.098	0.098	0
α	0.027	0.022	0.022	0
κ	0.244	0.197	0.197	0

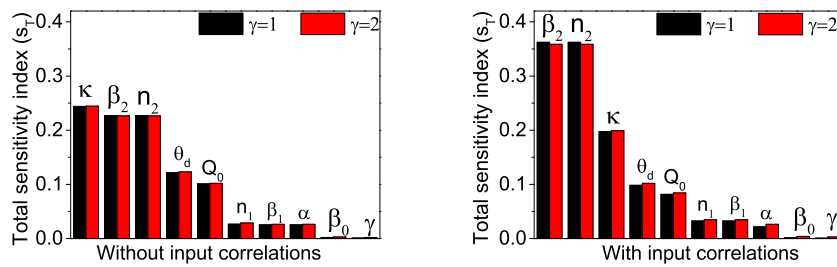


FIGURE 3.4: (Color online) Parameters are ranked according to their global importance (the total sensitivity indices s_{Ti}) in establishing the uncertainty of the basic reproduction number R_0 for HIV-1 model with vertical transmission. The left panel is for independent situation, and the right one for the correlated one with $\rho_{\beta_1 n_1} = 0.3$ and $\rho_{\beta_2 n_2} = 0.5$.

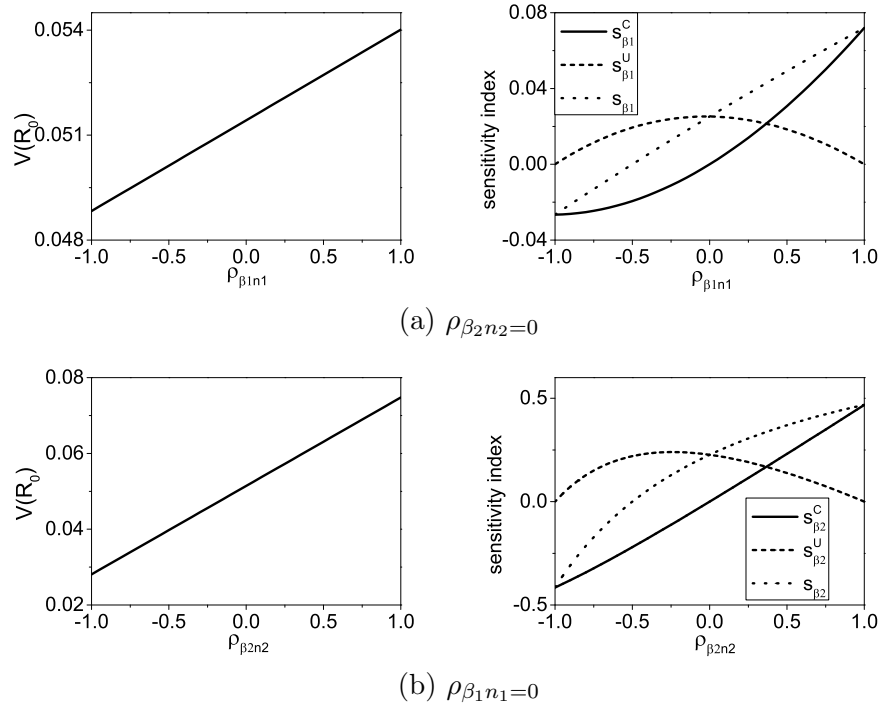


FIGURE 3.5: The distribution of our analysis results along the direction of the correlation between β_1 and n_1 , and the direction of the correlation between β_2 and n_2 . The sensitivity indices for n_1 and n_2 are exactly the same as those for β_1 and β_2 , respectively.

between β_1 and n_1 can be neglected, but the one contributed by the correlation between β_2 and n_2 is quite influential, in the estimation of the basic reproduction number R_0 .

Chapter 4

The establishment of sampling-based strategy

In diverse strategies for uncertainty and sensitivity analysis, the sampling-based method are considered as both effective and widely used in the absence of input correlations. Analysis relying on this method involves the generation and exploration of a mapping from input variables to the model response. The most crucial ingredient is the generation of sample points regarding sampling-based uncertainty and sensitivity analysis. In general, LHS and QMC sampling are often discussed and widely used for the analysis of computationally demanding models. The efficient stratification property of LHS allows for the extraction of a large amount of uncertainty and sensitivity information with a relatively small sample size [30, 240]. Low-discrepancy sequences guarantee a perfect uniformity and uniqueness of QMC sampling in the input space, which avoids large computational cost in configuration average [31, 241].

4.1 A comparison of our method with Sobol's one

A comparison of our method with Sobol's one is discussed in this chapter. Sobol's variance-based method is implemented here with the sampling-based strategy which also allows one to evaluate the sensitivity indices even if the functional form connecting input and output variables is absent. By using sample points, the representations of Sobol's sensitivity indices of different orders are presented in Section 1.3.7. In this part, the difference of our method from Sobol's one is presented via trivariate linear and nonlinear systems of specified functional relationships connecting y and \mathbf{x} , in the

absence of input correlations. Four numerical models of forms below are exemplified:

$$y_1 = 2x_1 - 4x_2 + 6x_3, \quad (4.1)$$

$$y_2 = 2x_1 + 3x_2^2 + 4x_1x_3, \quad (4.2)$$

$$y_3 = x_1x_2x_3^2, \quad (4.3)$$

$$y_4 = x_1x_2/x_3, \quad (4.4)$$

where input variables are assumed to be independent of each other and uniformly distributed in the closed interval $[0, 1]$ ($[0.1, 1.1]$ for the last case).

The underlying results obtained by our analytic method and by the sampling-based Sobol's one are exhibited in tables 4.1 and 4.2, respectively. Two sampling strategies are employed here, including LHS and QMC sampling. Typically, QMC sampling is performed by using high-dimensional Sobol low-discrepancy sequences that hold advanced uniformity properties in the input space. In practice, LHS-based results are averaged over 1000 configurations used for eliminating the randomness in sample generation. And QMC sampling-based values are obtained by unique configuration as low-discrepancy sequences are deterministic for given sample size and input variables. Results indicate a good agreement between our values and Sobol's ones, except the last two examples involving interaction effect of nonlinear problems of x_3 . For these two cases, Sobol's sensitivity indices overestimate the impacts contributed by x_1 , x_2 , as well as by their interaction, but underestimate the roles played by the association of each of them with x_3 , compared to the values provided by our method. This difference is introduced by the following components separately appearing in the analytic expansions of s_{13} , s_{23} , and s_{123} (see the definition of sensitivity indices in the second chapter):

$$\sum_{i=1}^{\infty} \frac{1}{i!} \left(\frac{\partial f}{\partial x_1} \cdot \frac{\partial^{i+1} f}{\partial x_1 \partial x_3^i} \right) (\{\mu\}) \cdot M_2(x_1) M_i(x_3) \Big/ V(y), \quad (4.5)$$

$$\sum_{i=1}^{\infty} \frac{1}{i!} \left(\frac{\partial f}{\partial x_2} \cdot \frac{\partial^{i+1} f}{\partial x_2 \partial x_3^i} \right) (\{\mu\}) \cdot M_2(x_2) M_i(x_3) \Big/ V(y), \quad (4.6)$$

$$\sum_{i=1}^{\infty} \frac{1}{i!} \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \cdot \frac{\partial^{i+2} f}{\partial x_1 \partial x_2 \partial x_3^i} \right) (\{\mu\}) \cdot M_2(x_1) M_2(x_2) M_i(x_3) \Big/ V(y). \quad (4.7)$$

The above components are in sequence regarded as the effects of x_1 , x_2 and of the interaction effect between them in Sobol's definition since artificially moving the above items in our analytic formula for the third case yields $\{s_1=0.14, s_2=0.14, s_3=0.36, s_{12}=0.05, s_{13}=0.13, s_{23}=0.13, s_{123}=0.04\}$ which are in remarkable agreement with Sobol's sampling-based values.

The information provided by item $[\partial y / \partial x_t]$ in Eqs. (4.5) and (4.6) can be approximated

TABLE 4.1: Analytic values for uncertainty and sensitivity analysis to numerical test examples. Numbers in parentheses denote the thousandth and ten-thousandth digits of corresponding items.

y	s_1	s_2	s_3	s_{T1}	s_{T2}	s_{T3}	s_{12}	s_{13}	s_{23}	s_{123}
y_1	0.07(14)	0.28(57)	0.64(29)	0.07(14)	0.28(57)	0.64(29)	0	0	0	0
y_2	0.51(73)	0.31(03)	0.12(93)	0.56(04)	0.31(03)	0.17(24)	0	0.04(31)	0	0
y_3	0.08(52)	0.08(52)	0.36(37)	0.36(36)	0.36(36)	0.80(11)	0.02(84)	0.18(75)	0.18(75)	0.06(25)
y_4	0.08(00)	0.08(00)	0.41(57)	0.32(25)	0.32(25)	0.82(15)	0.01(85)	0.18(18)	0.18(18)	0.04(21)

$E(y_1)=2.00(00)$, $E(y_2)=3.00(00)$, $E(y_3)=0.08(33)$, $E(y_4)=0.86(32)$;
 $V(y_1)=4.66(67)$, $V(y_2)=2.57(66)$, $V(y_3)=0.01(53)$, $V(y_4)=1.04(16)$.

as the square root of the whole information provided by x_t alone which decides the sensitivity index s_t , that is $\sqrt{V_t}$. Similarly, the contribution made by item $[\partial^2 y / (\partial x_1 \partial x_2)]$ in Eq. (4.7) can be approximated as $\sqrt{V_{12}}$. Nevertheless, the impact resulted from the right part in brackets of each equation among Eqs. (4.5), (4.6) and (4.7) should be separately approximated as half of the square root of V_{13} , V_{23} and V_{123} . This because of the only existent even-order central moments of x_3 . x_3 is uniformly distributed, resulting non-existence of contributions from the relevant odd-order derivatives of f with respect to x_3 . Following these approximations, a modification is introduced to the Sobol's method:

$$\begin{aligned}
 s_1^r &= s_1 - \frac{\sqrt{s_1 \cdot s_{13}}}{2}, & s_2^r &= s_2 - \frac{\sqrt{s_2 \cdot s_{23}}}{2}, & s_{12}^r &= s_{12} - \frac{\sqrt{s_{12} \cdot s_{123}}}{2}, \\
 s_{13}^r &= s_{13} + \frac{\sqrt{s_1 \cdot s_{13}}}{2}, & s_{23}^r &= s_{23} + \frac{\sqrt{s_2 \cdot s_{23}}}{2}, & s_{123}^r &= s_{123} + \frac{\sqrt{s_{12} \cdot s_{123}}}{2}.
 \end{aligned} \quad (4.8)$$

The revised results for the last two cases are shown in y_3^r - and y_4^r -th rows in table 4.2. They are in well agreement with analytic values accompanied by ignorable errors. The above modification depends upon the functional form of the model under discussion that provides information about which of measures are overvalued and which are underestimated. Following the modification, it is easy to understand the difference between our method and Sobol's one. For models of absent nonlinear interaction effects of input parameters, our method (in the absence of input correlations) is equivalent to Sobol's one in evaluating the importance of individual input parameters. When the model under discussion involves nonlinear interaction terms, however, the difference between our method and Sobol's one is existed in quantifying some sensitivity indices. By introducing the modification to Sobol's method, a rough sampling-based method is also established, coincident with our analytic concept. This builds a foundation of extending the sampling-based method from independent case to the correlated one. Sampling-based method of input correlations is of great importance in practice, which allows one to decide where or not the input correlations should be considered in models of absent functional forms. Results also suggest that QMC sampling, compared to LHS, can provide more precise measures at lower computational cost, as stated in the first chapter.

TABLE 4.2: Sampling-based results for uncertainty and sensitivity analysis of numerical test examples with $M = 10000$. Rows of y_3^r and y_4^r indicate the values after introducing our proposed modification. Values provided by LHS-based method are averaged over 1000 simulations with the corresponding standard deviation displayed in parentheses. For the results given by QMC sampling-based method, data in parentheses denote the thousandth and tenthousandth digits.

LHS-based method										
y	s_1	s_2	s_3	s_{T1}	s_{T2}	s_{T3}	s_{12}	s_{13}	s_{23}	s_{123}
y_1	0.07(.01)	0.29(.01)	0.64(.01)	0.07(.00)	0.29(.01)	0.64(.01)	0.00(0.01)	0.00(.01)	0.00(.01)	0.00(.01)
y_2	0.52(.01)	0.31(.01)	0.13(.01)	0.56(.01)	0.31(.01)	0.17(.01)	0.00(.01)	0.04(0.01)	0.00(0.01)	0.00(.01)
y_3	0.15(.01)	0.15(.01)	0.37(.01)	0.36(.02)	0.36(.02)	0.64(0.02)	0.05(.01)	0.12(.02)	0.12(.02)	0.04(.01)
y_4	0.17(.01)	0.17(.01)	0.42(.02)	0.32(.02)	0.32(.02)	0.63(.02)	0.04(.01)	0.09(.02)	0.09(.02)	0.02(.01)
y_3^r	0.09(.02)	0.09(.02)	0.37(.02)	0.36(.02)	0.36(.02)	0.64(.02)	0.03(.01)	0.18(.02)	0.18(.02)	0.07(.01)
y_4^r	0.10(.02)	0.10(.02)	0.42(.02)	0.32(.02)	0.32(.02)	0.63(0.02)	0.02(.01)	0.16(.03)	0.16(.03)	0.04(.02)
$E(y_1)=2(.00)$, $E(y_2)=3(.00)$, $E(y_3)=0.08(.00)$, $E(y_4)=0.86(.00)$; $V(y_1)=4.66(.05)$, $V(y_2)=2.59(.03)$, $V(y_3)=0.015(.00)$, $V(y_4)=1.04(.03)$.										
QMC sampling-based method										
y	s_1	s_2	s_3	s_{T1}	s_{T2}	s_{T3}	s_{12}	s_{13}	s_{23}	s_{123}
y_1	0.07(17)	0.28(57)	0.64(29)	0.07(15)	0.28(55)	0.64(28)	0.00(02)	0.00(02)	0.00(02)	0.00(02)
y_2	0.51(75)	0.31(02)	0.12(82)	0.56(12)	0.31(06)	0.17(19)	0.00(04)	0.04(37)	0.00(04)	0.00(04)
y_3	0.15(82)	0.15(33)	0.36(55)	0.36(37)	0.36(15)	0.64(53)	0.04(32)	0.11(49)	0.11(75)	0.04(74)
y_4	0.16(80)	0.16(76)	0.41(48)	0.32(24)	0.32(44)	0.62(96)	0.03(49)	0.09(29)	0.09(53)	0.02(66)
y_3^r	0.09(08)	0.08(62)	0.36(55)	0.36(37)	0.36(15)	0.80(24)	0.02(06)	0.18(22)	0.18(46)	0.07(01)
y_4^r	0.10(55)	0.10(44)	0.41(48)	0.32(23)	0.32(44)	0.77(05)	0.01(96)	0.15(53)	0.15(85)	0.04(19)
$E(y_1)=2.00(00)$, $E(y_2)=3.00(00)$, $E(y_3)=0.08(33)$, $E(y_4)=0.86(34)$; $V(y_1)=4.66(70)$, $V(y_2)=2.57(75)$, $V(y_3)=0.01(53)$, $V(y_4)=1.04(34)$.										

4.2 Analysis of SIR and SIS models

The modeling of infectious diseases provides a tool to study the mechanisms by which diseases spread, to predict the future course of an outbreak and to evaluate strategies to control an epidemic [242]. Mathematical modeling of disease spreading was first carried out in 1766 by Daniel Bernoulli who designed a numerical model to defend the practice of inoculating against smallpox [243]. However, modern theoretical epidemiology did not begin until the research of Ronald Ross into the spread of malaria [244]. This was soon followed by a deterministic model, i.e., SIR model, proposed by W.O. Kermack and A.G. McKendrick in 1927 with the consideration of a fixed population divided into three compartments: susceptible group S who are not infected with disease but can get infected, infectious category I who are infected with disease and can spread the disease to group S , and recovered community R who have been recovered from infection and are immune to reinfection [37]. Another widely considered model for epidemic spreading is SIS model relying on a coarse grained description of individuals in the population within which individuals are immediately susceptible once they have recovered.

The reliability investigation regarding deterministic compartmental models are performed for many years under the consideration of output's dependence upon one input factor at a time in fixing the others [245, 246]. This widely used one-at-a-time method, however, fails to globally quantify the roles played by individual input factors and by their interaction effects in the estimation of model response. Following the deficiency,

a systematical framework is carried out for the reliability analysis of both SIR and SIS models, by using sampling-based strategy and our proposed modification.

Considering a fixed population: $N = S(t) + I(t) + R(t)$, McKendrick and Kermack designed the following nonlinear system of ordinary differential equations:

$$\begin{cases} \frac{ds(t)}{dt} = -\beta s(t)i(t), \\ \frac{di(t)}{dt} = \beta s(t)i(t) - \gamma i(t), \\ \frac{dr(t)}{dt} = \gamma i(t). \end{cases} \quad (4.9)$$

Apparently, we have $s(t) + i(t) + r(t) = 1$. β labels the probability at which a susceptible individual can be infected after communicating with infected individuals, and γ the probability at which an infected individual will be spontaneously recovered from the infection. In Eq. (4.9), dividing the first equation by the third one yields:

$$\ln s(t) = -\frac{\beta}{\gamma}r(t) + \ln s(0) + r(0), \quad (4.10)$$

in which $s(0)$ and $r(0)$ label the initial condition that specifies the infectious invading. In general, $r(0)$ is set to 0, yielding $i(0) = 1 - s(0)$. At equilibrium state of disease spreading, the normalized number of both susceptible and infectious agents satisfy

$$\begin{cases} \beta si & = 0, \\ \beta si - \gamma i & = 0, \\ \gamma i & = 0. \end{cases} \quad (4.11)$$

which, together with Eq. (4.10), provides

$$\begin{cases} i & = 0, \\ \ln s & = -\frac{\beta}{\gamma}(1 - s) + \ln s(0). \end{cases} \quad (4.12)$$

Apparently, SIR model does not lead to any analytic solution for the equilibrium state but just states that s associates with input parameters β , γ and $s(0)$ which are assumed to be independent of each other and uniformly distributed in the real range $[0, 1]$:

$$s = f(s(0), \beta, \gamma); \quad \beta, \gamma, s(0) \in [0, 1]. \quad (4.13)$$

Differential equations for SIS model can be deduced from SIR model by removing the recovered populations into the susceptible compartment:

$$\begin{cases} \frac{ds(t)}{dt} = -\beta s(t)i(t) + \gamma i(t), \\ \frac{di(t)}{dt} = \beta s(t)i(t) - \gamma i(t). \end{cases} \quad (4.14)$$

Its exact solution at equilibrium state is

$$s = \begin{cases} 1, & \gamma \geq \beta, \\ \frac{\gamma}{\beta}, & \gamma < \beta. \end{cases} \quad (4.15)$$

Numerical simulations of Eqs (4.9) and (4.14) are presented in Fig. 4.1 by considering different values of input variables. Results demonstrate that the underlying system will fast reach the equilibrium state for both models. Uncertainty and sensitivity analysis of both models is performed by sampling-based strategies including LHS and QMC sampling. QMC sampling is implemented by the use of high-dimensional Sobol low-discrepancy sequence generator. The underlying analysis results are exhibited in table 4.3 where rows of SIR^r and SIS^r correspond to the values after introducing our proposed modification. LHS-based values are averaged over 1000 configurations for eliminating the randomness in sample generation, and QMC sampling-based ones are provided by unique configuration.

Regarding SIS model yielding the exact solution of equilibrium state, the inaccuracy of the original sampling-based strategy in the implementation of our analytic method is introduced by the following summation set:

$$\sum_{i=1}^{\infty} \frac{1}{i!} \left(\frac{\partial s}{\partial \gamma} \cdot \frac{\partial^{i+1} s}{\partial \gamma \partial^i \beta} \right) (\{\mu\}) \cdot M_2(\gamma) M_i(\beta) / V(y), \quad (4.16)$$

which leads to the overestimation of index s_γ but underestimation of index $s_{\beta\gamma}$. Based on this description, the relevant items are then corrected as

$$s_\gamma^r = s_\gamma - \frac{\sqrt{s_\gamma \cdot s_{\beta\gamma}}}{2}, \quad s_{\beta\gamma}^r = s_{\beta\gamma} + \frac{\sqrt{s_\gamma \cdot s_{\beta\gamma}}}{2}. \quad (4.17)$$

For SIR model with no exact solution of equilibrium state, the association of model response s with each input variable is shown in Fig. 4.2. It is noticed that s nonlinearly depends upon the ratio of γ to β but almost linearly correlates with $s(0)$. Furthermore, the interaction effect of $s(0)$ with each of β and γ can be neglected since the lines in double log plane of s - $s(0)$ are almost parallel to each other for different ratios of γ to β except as $s(0) \rightarrow 1$. In the region of $s(0) \rightarrow 1$, interaction effects of $s(0)$ with other

TABLE 4.3: Sampling-based values for uncertainty and sensitivity analysis of both SIR and SIS models. The sample size is $M = 10000$. Rows of SIR^r and SIS^r indicate the results after introducing our proposed modifications. For LHS-based results, numbers in parentheses denote the standard deviation of each corresponding item, resulting from 1000 independent simulations. For QMC sampling-based values, data in parentheses are the thousandth and ten-thousandth digits.

LHS-based method										
$y = s$	s_β	s_γ	$s_{s(0)}$	$s_{T\beta}$	$s_{T\gamma}$	$s_{Ts(0)}$	$s_{\beta\gamma}$	$s_{\beta s(0)}$	$s_{\gamma s(0)}$	$s_{\beta\gamma s(0)}$
SIR	0.19(.01)	0.18(.01)	0.42(.01)	0.31(.01)	0.34(.01)	0.60(.01)	0.04(.01)	0.06(.02)	0.09(.02)	0.03(.01)
SIS	0.20(.02)	0.64(.01)	—	0.36(.01)	0.80(.02)	—	0.16(.02)	—	—	—
SIR^r	0.15(.02)	0.14(.02)	0.42(.01)	0.35(.01)	0.38(.01)	0.60(.01)	0.11(.03)	0.06(.02)	0.09(.02)	0.03(.01)
SIS^r	0.20(.02)	0.49(.02)	—	0.51(.01)	0.80(.02)	—	0.31(.04)	—	—	—
$E(y)^{SIR}=0.24(.00)$, $E(y)^{SIS}=0.75(.00)$; $V(y)^{SIR}=0.07(.00)$, $V(y)^{SIS}=0.10(.00)$.										
QMC sampling-based method										
$y = s$	s_β	s_γ	$s_{s(0)}$	$s_{T\beta}$	$s_{T\gamma}$	$s_{Ts(0)}$	$s_{\beta\gamma}$	$s_{\beta s(0)}$	$s_{\gamma s(0)}$	$s_{\beta\gamma s(0)}$
SIR	0.18(56)	0.18(11)	0.41(54)	0.30(97)	0.34(00)	0.59(45)	0.03(88)	0.05(90)	0.09(38)	0.02(62)
SIS	0.20(12)	0.64(41)	—	0.35(59)	0.79(88)	—	0.15(47)	—	—	—
SIR^r	0.14(32)	0.13(92)	0.41(54)	0.35(16)	0.38(23)	0.59(45)	0.12(32)	0.05(90)	0.09(38)	0.02(62)
SIS^r	0.20(12)	0.48(62)	—	0.51(38)	0.79(88)	—	0.31(26)	—	—	—
$E(y)^{SIR}=0.24(19)$, $E(y)^{SIS}=0.75(01)$; $V(y)^{SIR}=0.07(14)$, $V(y)^{SIS}=0.10(42)$.										

factors are existed, explaining the non-vanishing high-order sensitivity indices involving $s(0)$. Consequently, for SIR model, the items that give rise to the inaccuracy of the sampling-based strategy in the implementation of our analytic method include

$$\sum_{i,j,k=1}^{\infty} \frac{1}{i! \cdot j! \cdot k!} \left(\frac{\partial^i s}{\partial \gamma^i} \cdot \frac{\partial^{j+k} s}{\partial \gamma^j \partial \beta^k} \right) (\{\mu\}) \cdot M_{i+j}(\gamma) M_k(\beta) \Big/ V(y), \quad (4.18)$$

$$\sum_{i,j,k=1}^{\infty} \frac{1}{i! \cdot j! \cdot k!} \left(\frac{\partial^i s}{\partial \beta^i} \cdot \frac{\partial^{j+k} s}{\partial \gamma^j \partial \beta^k} \right) (\{\mu\}) \cdot M_j(\gamma) M_{i+k}(\beta) \Big/ V(y), \quad (4.19)$$

which yield the modification of sensitivity indices as

$$s_\beta^r = s_\beta - \frac{\sqrt{s_\beta \cdot s_{\beta\gamma}}}{2}, \quad s_\gamma^r = s_\gamma - \frac{\sqrt{s_\gamma \cdot s_{\beta\gamma}}}{2}, \quad s_{\beta\gamma}^r = s_{\beta\gamma} + \frac{\sqrt{s_\beta \cdot s_{\beta\gamma}}}{2} + \frac{\sqrt{s_\gamma \cdot s_{\beta\gamma}}}{2}. \quad (4.20)$$

By introducing modifications presented above, analysis results of SIR model suggest that the initial proportion $s(0)$ of susceptible individuals plays the most important role in the estimation of equilibrium state of epidemic spreading; infectious probability β and recovered rate γ almost explain the same proportion of variance in response s ; the interaction effects of input factors play fragile roles in the decision of model response. Regarding SIS model, parameter γ plays more robust role than β , and their interaction effect also provides a crucial contribution to the uncertainty of equilibrium state, by holding 31% of the decision of s . Sampling-based plots for the association of s with each single factor are exhibited in Fig. 4.3, together with phase diagrams of model response s in any possible two-dimensional input spaces.

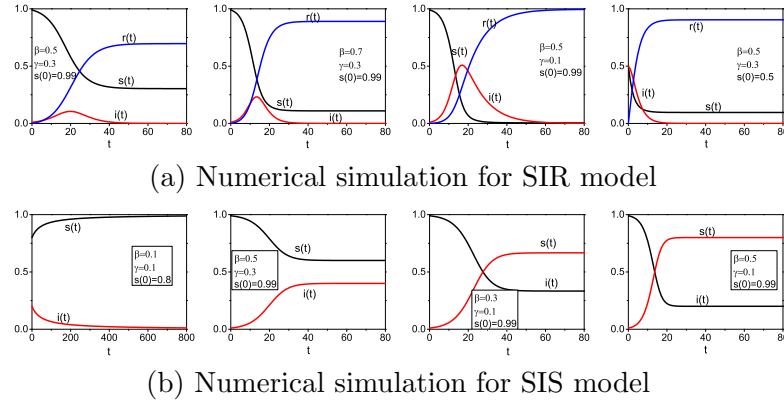


FIGURE 4.1: Numerical simulation results for both SIR and SIS models by considering different values of input parameters.

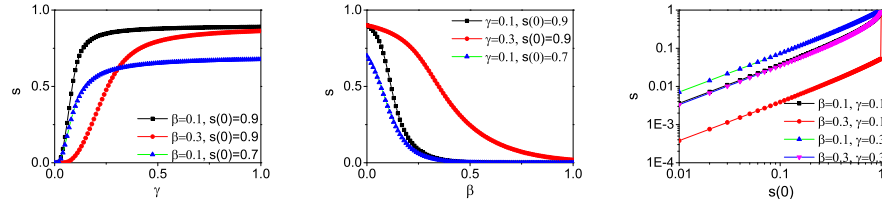


FIGURE 4.2: The normalized number s of susceptible individuals acts as functions of individual input factors at equilibrium state of SIR model.

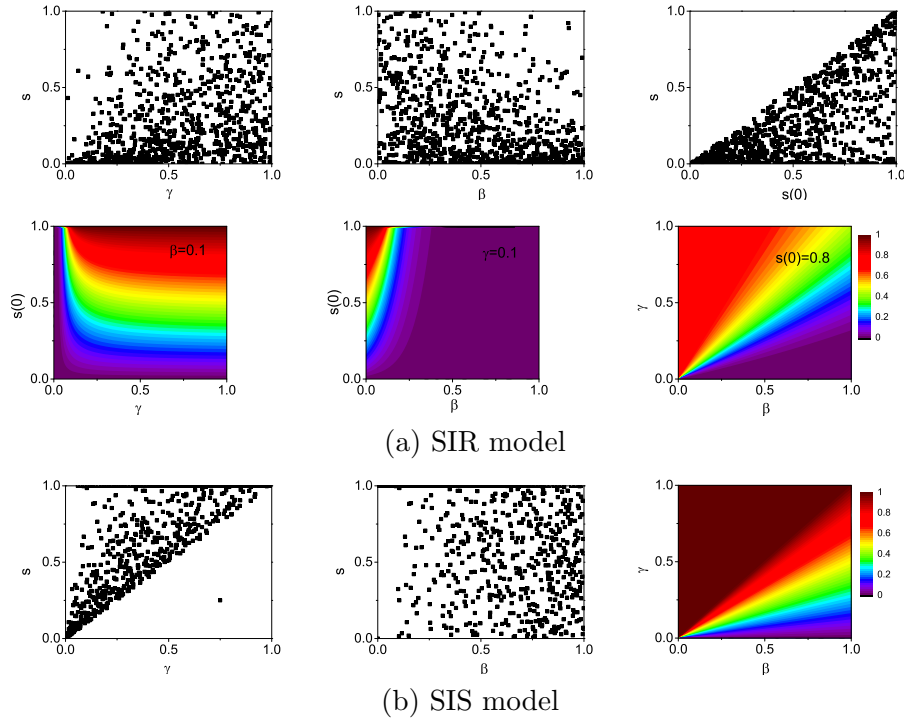


FIGURE 4.3: Scatter plots based on LHS, as an example, with $M = 1000$ for the normalized number of susceptible individuals at equilibrium state, together with its phase diagrams in any possible two-dimensional input space.

Chapter 5

Opinion formation based on a gambling mechanism and its sensitivity analysis

5.1 Introduction

The dynamics formation of public opinion has raised a fair amount of attention from researchers working on employing diverse tools furnished by statistical physics to understand the appearance of some social phenomena [247–249]. As Castellano *et al* have stated, a common theme of social dynamics is the understanding of the transition from an initially disordered state to a configuration that displays order (at least partially) [39]. In opinion dynamics, ordered configuration corresponds to the emergence of an agreement which is necessary to be reached for a community in many situations appeared in daily life. In establishing a framework on opinion dynamics, the most important part includes specifying any possible opinion states of a community and defining the elementary processes or rules that determine opinion transitions of every individual between such states.

The first model of opinion dynamics known so far was proposed by Weidlich to study individual's decision behaviour by reformulating statistical models which are originally used for describing the dynamics of interacting spins [250]. Later on, Ising model made its first appearance in opinion dynamics [251]. In these early models, as well as many other subsequently built ones, such as the widely considered majority-voter model [25, 252], and the Sznajd model [253], the public opinion is simplified as a discrete variable. Discrete opinion often takes two possible values, representing a reasonable description in several instances: yes or no, buying or selling, cooperation or defection, Samsung

or iPhone, etc. The discrete opinion dynamics is favoured over continuous one where agent's opinion can vary smoothly from one extreme of the range of possible choices to another, because of its simplicity in theoretical analysis [24, 254, 255]. However, for the investigation of the evolution of agent's attitudes about a given topic, for example, the political orientation of individual agents, continuous opinion dynamics seems more suitable. Furthermore, the latter is also definitely advantageous for gaining insight into the emergence of opinion clusters.

Opinion clusters could be one (consensus), two (polarization), or more (fragmentation). How social clusters emerge in opinion space is one of the fundamental problems in social science. In small communities, such as a football team, or a research group, each agent knows and can communicate with every other one. For a large community or population, however, many agents do not know or even never heard of each other because of the geographic limitation. Hence, local interactions happen only within groups of friends. In computer simulations, the friendship in a community is defined in general by a connected network with the same size of the community. Moreover, the bounded confidence is also widely considered. It was proposed according to the realistic aspect of human communications happening among agents whose opinions are sufficiently close to each other [256–258]. Mathematical models and statistical physics have offered powerful tools for deeply understanding how and what conditions collective behaviours happen in local individual interaction rules [259, 260]. The roles of some possible social factors, such as leaders and environmental noises [261], mass media [262], agent's conviction and influence strength [263], and informed agents [264], are also identified recently in the formation of opinion dynamics.

In current social opinion research, unidirectional influence from considered social factors to the underlying dynamics is the main focus. However, agent's opinion may also, in turn, react on the associated social environment. Based on the consideration, a virtual gambling mechanism is introduced to explore the characteristics of opinion dynamics behaviour in a social community. At each round of gambling, agents fight for a limited but conserved resource. The social pattern considered here is the resource distribution in the underlying system. Opinions of agents determine directly the resource that each agent can share, thereby deciding winning or losing of agents at our gambling. Agents then can accumulate scores from a reward and punishment system which is proposed based on the winning or losing state of agents. Agent's scores will, in turn, affect his opinion in subsequent interactions with his friends whose opinions must differ from his own not more than a pre-defined confidence threshold. In general, human beings are recognised as "higher animal" because of self-awareness and the freedom from nature's determinism that allows one to choose, whether for good or ill. This implies that agents participating in our gambling will instinctively learn from their friends who have earned

more scores from the activity owing to the primary intend of winning money and/or material goods. A systematic framework is also performed for the uncertainty and sensitivity analysis of the model under discussion. The underlying analysis results help gain insight into how involved parameters can be ranked according to their importance in establishing the uncertainty of model response of our interest.

5.2 Modeling

In our model, we have N social agents. Their friendship is realized on a Barabási-Albert scale-free network (for detailed generation process, please check Ref. [265]) as many social networks have been reported to be scale-free [265–267]. Each agent i has an opinion s_i in the real range $[-1, 1]$. Agents are segregated into two cliques according to the sign of their opinions: G_+ (with positive opinions) and G_- (with negative opinions). The absolute value of s_i is regarded as the degree of attention or the cost that agent i pays for the issue of gambling.

A parameter γ is introduced by the gambling mechanism to control the resource discrepancy between two cliques: $\gamma = R_+/R_-$ in keeping total resource R ($=R_+ + R_-$) conserved. R_+ and R_- label the resource allocated to G_+ and G_- , respectively. In computer simulations, they are specified by γ and R defined previously:

$$R_+ = \frac{\gamma R}{1 + \gamma}, \quad R_- = \frac{R}{1 + \gamma}, \quad (5.1)$$

where the total resource R is set to 1 in our simulations, but the behavior of our model is almost independent of the value of R . For simplicity but without loss of generality, γ is restricted to the closed interval $[0, 1]$ (the system is symmetric around $\gamma = 1$). The resource held by each clique is then shared by their members. Generally speaking, an agent who pays more attention or cost will share more resource according to the general rule of survival: the more you give, the more you receive. On this basis, the resource that an agent shares from his community is proportional to the absolute value of his opinion.

Rewards and punishments system is also introduced. An agent is regarded as a winner and to be rewarded with C scores if the resource he shares is no less than the global average value $A_r = R/N$; otherwise, the agent is considered as a loser and to be punished with losing C scores. C is a constant for each agent and set to 5 in our modelling, but its value does not influence the results of our interest. Actually, similar ideas have been considered in earlier modeling of social behaviors. In the naming game, as an example, the concept of score is introduced to indicate the agent's reputation which is variable

in time and partially controls the behavior of the system [268]. For another example, in the modeling of discrete opinion dynamics based on a majority rule, a parameter indicating the convincing power of each agent, which may increase or decrease with time, analogous to the above mentioned score, is proposed to decide whether the majority opinion succeeds or not in a randomly selected group [269]. More considerations of the concept or similar concept of score are given by [270–272].

In general, gamblers are economic with making decisions by their own and taking the strategy in their favour. Some people, however, may act as mindless while facing some events or subjects possibly because the corresponding events or subjects do not interest them or they do not care about personal gain. Following the consideration, a parameter p is introduced to control the fraction of mindless agents included in the system under consideration.

The initial state of the system is assumed to be fully disordered, that is, at the beginning of the dynamics, each individual has an opinion drawn from the uniform distribution in the range $[-1, 1]$. Mindless agents are selected at random from the system with probability p . At a given time step t , the following microscopic rules control the formation of opinion dynamics:

(1) Agent i is mindless. In this case, the agent takes the average opinion of those friends whose opinions differ from his own no larger than a certain confidence threshold ε to form his opinion for the next time, that is

$$s_i(t+1) = \delta_{|F(i,t)|} s_i(t) + (1 - \delta_{|F(i,t)|}) \frac{1}{|F(i,t)|} \sum_{j \in F(i,t)} s_j(t), \quad (5.2)$$

where $|\cdot|$ for a finite set denotes the number of elements, $\delta_x = 1(0)$ for $x = 0(> 0)$ the Kronecker delta function. The set $F(i, t)$, which denotes the effective friends of agent i , is defined by

$$F(i, t) = \{1 \leq j \leq N, j \neq i \mid a_{ij} = 1 \text{ and } |s_i(t) - s_j(t)| \leq \varepsilon\}, \quad (5.3)$$

in which $|\cdot|$ denotes the absolute value of a real number, a_{ij} the element in adjacency matrix with $a_{ij} = 1(0)$ if there is an edge connecting i and j (otherwise). In the present paper, the confidence threshold ε is taken as constant in time and across the whole population, as discussed in Ref. [273] where the heterogeneity of threshold and the adaptive threshold were also introduced, respectively, yielding the appearance of new features of opinion dynamics.

(2) Agent i is economic. i will update his opinion by taking into account the scores of himself and of his effective friends. s_i keeps unchanged if i gains the highest scores

compared with all $j \in F(i, t)$; otherwise, s_i is updated by the opinion (or average opinion) of the agent(s) who own(s) the highest scores in the set $F(i, t)$. The rule for the evolution of s_i is given by

$$s_i(t+1) = \delta_{c_i(t)m_i(t)} s_i(t) + (1 - \delta_{c_i(t)m_i(t)}) \frac{1}{|M(i, t)|} \sum_{j \in M(i, t)} s_j(t), \quad (5.4)$$

with

$$m_i(t) = \max(c_i(t), \{c_j(t), j \in F(i, t)\}) \quad (5.5)$$

and

$$M(i, t) = \{j \in F(i, t) \mid c_j(t) = m_i(t)\}, \quad (5.6)$$

where $\delta_{xy} = 1(0)$ if $x = y$ (otherwise), and $\max()$ is a function to return the maximum value of a set of elements. $c_i(t)$ labels the scores that i has gained until time t , which will be refreshed accordingly by

$$c_i(t+1) = \begin{cases} c_i(t) + 5, & r_i(t) \geq A_r, \\ c_i(t) - 5, & r_i(t) < A_r. \end{cases} \quad (5.7)$$

$r_i(t)$ is the resource that i shares at time t , given by

$$r_i(t) = \begin{cases} \frac{s_i(t)}{T_{\Omega^+}(t)} \cdot R_+, & s_i(t) > 0, \\ \frac{s_i(t)}{T_{\Omega^-}(t)} \cdot R_-, & s_i(t) \leq 0, \end{cases} \quad (5.8)$$

in which $T_{\Omega^+}(t)$ and $T_{\Omega^-}(t)$ are the sum of positive and negative opinions, respectively,

$$T_{\Omega^+}(t) = \sum_{i=1, s_i(t)>0}^N s_i(t), \quad T_{\Omega^-}(t) = \sum_{i=1, s_i(t)\leq 0}^N s_i(t). \quad (5.9)$$

In our modelling, the unit time ($t = 1$) is defined as N , which corresponds to a single update of each agent. Network parameters are set as $N = 1000$ and $k_a = 4$. k_a indicates the average degree of each agent, that is the average number of friends.

5.3 Methods

Time evolution of the underlying opinion dynamics is firstly presented in Fig. 5.1. The absolute global average opinion and local average opinions in both cliques are represented as

$$\Omega(t) = \frac{1}{N} \left| \sum_{i=1}^N s_i(t) \right|, \quad \Omega^+(t) = \frac{T_{\Omega^+}(t)}{N_+(t)}, \quad \Omega^-(t) = \frac{T_{\Omega^-}(t)}{N_-(t)}, \quad (5.10)$$

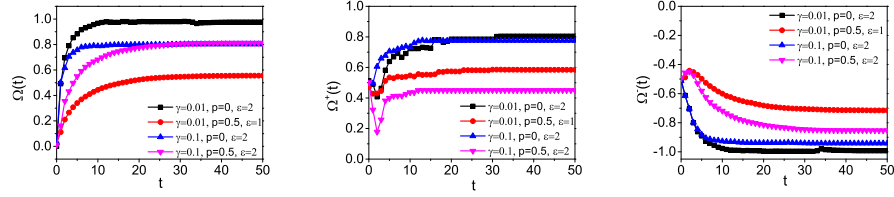


FIGURE 5.1: Time evolution of opinion dynamics for different specified values of parameters with $N = 1000$ and $k_a = 4$. One unit time corresponds to sampling one round of gambling game.

where $N_+(N_-)$ indicates the number of agents in clique $G_+(G_-)$. Numerical results demonstrate the presence of steady state on the opinion evolution process.

To characterise the coherence of the collective state of the population, we employ

$$\Omega_s = \left\langle \frac{1}{N} \left| \sum_{i=1}^N s_i \right| \right\rangle, \quad (5.11)$$

where $\langle \dots \rangle$ denotes average over configurations, computed at steady states. Ω_s is known to be an order parameter which plays the role of the magnetisation in magnetic systems. It is sensitive to the unbalance between positive and negative opinions. A collective state with a significantly nonvanishing value of Ω_s indicates a symmetry-broken distribution of opinions. A very small value of $\Omega_s (\simeq 0)$ means a symmetric distribution of opinions. And an exact value $\Omega_s = 0$ corresponds to an absorbing state where all agents share the neutral opinion $s = 0$ [274].

In practical simulations, first 100 time steps are used to make the system reach its steady state. Time average is then taken over the following 100 steps. 1000 independent simulations are proposed to take the average over configurations.

5.4 Results and discussion

The global steady state is driven from the local steady states of both cliques. This implies that the order parameter can be represented as

$$\Omega_s = |n_+ * \Omega_s^+ + (1 - n_+) * \Omega_s^-|, \quad (5.12)$$

where n_+ indicates the normalized number of agents in clique G_+ ; Ω_s^+ and Ω_s^- denote the average opinions in clique G_+ and G_- at steady state, respectively.

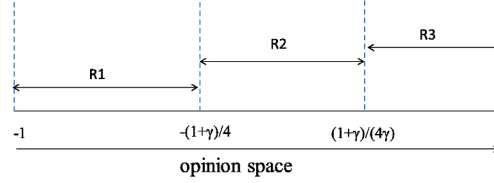


FIGURE 5.2: The generation of three adjacent regions in the opinion space.

5.4.1 Theoretical analysis for special cases

$$p = 0, \varepsilon = 2$$

Firstly, the role of resource allocation parameter γ in the formation of opinion dynamics is focused by fixing both confidence threshold and fraction of mindless agents. For simplicity, we set $\varepsilon = 2$ and $p = 0$. The primary intent of winning money or material goods at gambling will drive agents to learn opinions (or strategies) of their friends who have gained more scores. At the beginning of the dynamics ($t = 0$), rewards and punishments system requires the opinions of winners (have the ability to affect others) to follow

$$|s_i(0)| \geq \begin{cases} \frac{1+\gamma}{4\gamma}, & s_i(0) > 0, \\ \frac{1+\gamma}{4}, & s_i(0) < 0. \end{cases} \quad (5.13)$$

Consider the situation of $\gamma \geq 1/3$ obtained from the inequation

$$\frac{1+\gamma}{4\gamma} \leq 1, \quad (5.14)$$

which guarantees the presence of winners in G_+ at the initial time. We divide the opinion space into three adjacent regions, as shown in Fig. 5.2. An agent i with initial opinion $s_i(0)$ in the region $R2$ will adapt his opinion to $R1$ or $R3$, or continuously to $R2$ by learning from his friends. However, to keep $s_i(t > 0)$ staying in $R2$, i must have friends with initial opinions in neither or both of the regions $R1$ and $R3$. This event is neglected in the following analysis as it occurs with a quite small probability decided by the power function of the fraction of agents in different regions.

Fully disordered initial state also drives uniformly distributed opinions in both regions $R1$ and $R3$ at very beginning of gambling. With agents migrating in opinions from $R2$ to $R1$ or $R3$, the resource shared by each agent will decrease accordingly. This gives rise to a higher requirement for the opinions of agents who can continuously win at our gambling. The final range of opinions in both cliques can be approximated by

$$\frac{s_i}{N_+ \cdot \frac{1}{2}(1 + \frac{1+\gamma}{4\gamma})} \cdot R_+ \geq \frac{R}{N}, \quad (5.15)$$

and

$$\frac{|s_i|}{N_- \cdot \frac{1}{2}(1 + \frac{1+\gamma}{4})} \cdot R_- \geq \frac{R}{N}, \quad (5.16)$$

respectively. Simulation results suggest $N_+/N_- = R_+/R_-$ at steady state (Fig. 5.3(a)). Inserting $R_+ = \gamma R/(1 + \gamma)$ and $R_- = R/(1 + \gamma)$ into Eqs. (5.15) and (5.16) yields

$$|s_i| \geq \begin{cases} \frac{1}{2} + \frac{1+\gamma}{8\gamma}, & s_i > 0, \\ \frac{1}{2} + \frac{1+\gamma}{8}, & s_i < 0. \end{cases} \quad (5.17)$$

which determines the order parameter as

$$\begin{aligned} \Omega_s &= \left| \frac{\gamma}{1+\gamma} \cdot \frac{1}{2} \left(1 + \frac{1}{2} + \frac{1+\gamma}{8\gamma} \right) - \frac{1}{1+\gamma} \cdot \frac{1}{2} \left(1 + \frac{1}{2} + \frac{1+\gamma}{8} \right) \right| \\ &= \frac{3}{2(1+\gamma)} - \frac{3}{4}. \end{aligned} \quad (5.18)$$

Sufficiently biased resource allocation, that is $\gamma < 1/3$ obtained from the inequation

$$\frac{1+\gamma}{4\gamma} \geq 1, \quad (5.19)$$

will result in the absence of winners in G_+ at the initial time. However, some agents will be attracted from G_+ to G_- by communicating with their friends. The decrease in the number of agents with positive opinions will then generate winners in G_+ . The distribution of the fraction of winners presented in Fig. 5.3(b) demonstrates that almost all agents are winner at final state when $10/N < \gamma < 1/3$ ($\gamma = 0$ is also included). Assume agents in G_+ gather finally in opinion s . The condition for winning requires

$$\frac{s}{N_+ s} \cdot R_+ \geq \frac{R}{N}. \quad (5.20)$$

By substituting $R_+ = \gamma R/(1 + \gamma)$, we have

$$\gamma > \frac{1}{N}. \quad (5.21)$$

This explains the absence of winners at quite low γ : $0 < \gamma < 1/N$ (Fig. 5.3(b)).

Furthermore, the fluctuation of the fraction of winners in the region of $1/N \leq \gamma < 10/N$ is apparent. This results from the cyclical change of the opinion of one or few exceptional agents between positive and negative values at steady state (Fig. 5.4(a)). The cyclical behaviour makes the fraction of winners, denoted by n_w , skip continually between one value (close to 0) and another (close to 1) (Fig. 5.4(b)). Nevertheless, it does not influence the main clustering of agents in opinion close to -1, compared to which, opinion

contribution of other agents is negligible. This provides

$$\Omega_s = \frac{1}{1 + \gamma}. \quad (5.22)$$

When we limit γ to $[10/N, 1/3)$, the fraction of winners is close to one. This implies two main clusterings of agents in opinions -1 and 1 because the condition of winning requires the final minimal absolute opinions in both cliques to satisfy

$$\frac{s}{N_+ \cdot \frac{1}{2}(1 + s)} \cdot R_+ \geq \frac{R}{N}, \quad \text{for } s > 0, \quad (5.23)$$

and

$$\frac{|s|}{N_- \cdot \frac{1}{2}(1 + |s|)} \cdot R_- \geq \frac{R}{N}, \quad \text{for } s < 0, \quad (5.24)$$

respectively. Clustering of few agents is ignored in opinion close to 0 here, which is generated by agents who have friends in both above mentioned clusters. The order parameter Ω_s is then specified as

$$\begin{aligned} \Omega_s &= \left| \frac{\gamma}{1 + \gamma} - \frac{1}{1 + \gamma} \right| \\ &= \frac{2}{1 + \gamma} - 1. \end{aligned} \quad (5.25)$$

Now we analytically obtain the distribution of order parameter Ω_s in the whole region of resource allocation parameter γ by keeping $p = 0$ and $\varepsilon = 2$:

$$\Omega_s = \begin{cases} 1, & 0 \leq \gamma < \frac{1}{N}, \\ \frac{1}{1 + \gamma}, & \frac{1}{N} \leq \gamma < \frac{10}{N}, \\ \frac{2}{1 + \gamma} - 1, & \frac{10}{N} \leq \gamma < \frac{1}{3}, \\ \frac{3}{2(1 + \gamma)} - \frac{3}{4}, & \gamma \geq \frac{1}{3}. \end{cases} \quad (5.26)$$

In the present paper, the size of system considered is $N = 1000$. This implies the above equation can be simplified as

$$\Omega_s = \begin{cases} \frac{2}{1 + \gamma} - 1, & 0 \leq \gamma < \frac{1}{3}, \\ \frac{3}{2(1 + \gamma)} - \frac{3}{4}, & \frac{1}{3} < \gamma \leq 1. \end{cases} \quad (5.27)$$

A remarkable agreement between analytic values and numerical results is presented in Fig. 5.3(c).

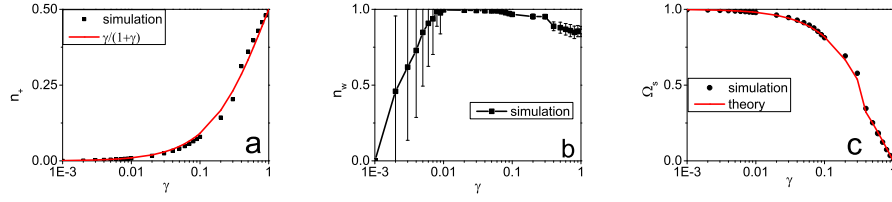


FIGURE 5.3: Corresponding results for the special case with $p = 0$ and $\varepsilon = 2$ after system reaches the steady state. Simulation data are averaged over 1000 simulations with $N = 1000$ and $k_a = 4$. (a): values for the normalised number of agents in G_+ ; (b): distribution of the fraction of winners with error bars denoting its standard deviation derived from independent simulations; (c): a comparison between numerical results and theoretic analysis, represented by Eq. (5.27), for the order parameter Ω_s .

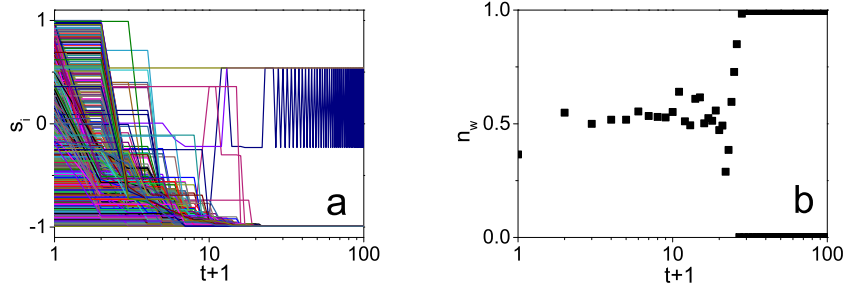


FIGURE 5.4: Show only single run with $N = 1000$, $k_a = 4$. Model parameters: $p = 0$, $\varepsilon = 2$, $\gamma = 0.003$. One unit time corresponds to sampling one round of gambling game.

$$p = \gamma = 0$$

The effect of confidence threshold ε is also assessed on the formation of opinion dynamics by assuming $p = \gamma = 0$. $\gamma = 0$ stipulates the total resource is concentrated in one clique (G_-). Following Eq. (5.13), initial winners hold opinions between -1 and -0.25. By introducing confidence threshold, an agent i with initial opinion larger than -0.25 can be affected at a probability P_i by his friends with opinions between $-(\varepsilon - s_i)$ and -0.25. P_i is straightforwardly specified by the mean field assumption as

$$P_i = \frac{k_a}{2}(\varepsilon - s_i(0) - \frac{1}{4}), \quad (5.28)$$

where $k_a(=4)$ denotes the average degree. Setting $P_i = 1$ yields

$$s_i(0) = \varepsilon - \frac{1}{4} - \frac{2}{k_a} = \varepsilon - \frac{3}{4}. \quad (5.29)$$

We divide the opinion space into four adjacent regions, as shown in Fig. 5.5. Agents with initial opinions in region $R2$ will definitely adapt their opinions to $R1$ by local communications. Agents who take initial opinions in $R3$ will be affected by their effective friends at a certain probability. For the fourth region, opinion distribution will maintain

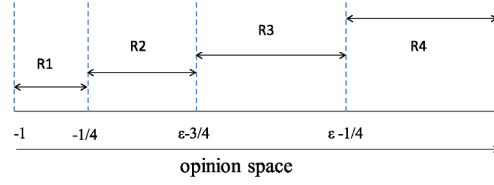


FIGURE 5.5: The generation of four adjacent regions in the opinion space.

its initial configuration throughout the time evolution of opinion dynamics because of confidence limitation.

When $\varepsilon \in (3/4, 5/4]$, the fraction of agents who will migrate in opinion from $R3$ to $R1$ is approximated as

$$\begin{aligned} n_{3 \rightarrow 1} &= \int_{\varepsilon - \frac{3}{4}}^{\varepsilon - \frac{1}{4}} \frac{1}{2} \cdot \frac{k_a}{2} (\varepsilon - x - \frac{1}{4}) dx \\ &= \frac{1}{8}. \end{aligned} \quad (5.30)$$

The fraction of winners (all agents in G_-) in the steady state is then determined by

$$n_w = \frac{1}{8} + \frac{1}{2}(\varepsilon - \frac{3}{4} + 1) = \frac{\varepsilon}{2} + \frac{1}{4}, \quad (5.31)$$

which says the final opinions of agents in G_- must satisfy the inequality

$$\frac{|s|}{N(\frac{\varepsilon}{2} + \frac{1}{4}) \cdot \frac{1}{2}(|s| + 1)} \cdot R_- \geq \frac{R}{N}, \quad (5.32)$$

where opinions in G_- are assumed to be uniformly distributed between -1 and s . Substituting $R_- = R$ drives

$$s \leq \frac{2\varepsilon + 1}{2\varepsilon - 7}, \quad (5.33)$$

which also gives the final average opinion in G_- :

$$\Omega_s^- = \frac{2\varepsilon + 1}{2\varepsilon - 7}. \quad (5.34)$$

The final average opinion in $R4$ is readily calculated:

$$\begin{aligned} \Omega_s^4 &= \frac{1}{2} [1 - (\varepsilon - \frac{1}{4})] \cdot \frac{1}{2} [1 + (\varepsilon + \frac{1}{4})] \\ &= -\frac{\varepsilon^2}{4} + \frac{\varepsilon}{8} + \frac{15}{64}. \end{aligned} \quad (5.35)$$

Regarding region $R3$, some agents will be affected by his friends with initial opinions in $R1$. Furthermore, agents with opinions more closer to $R2$ are more easier to be affected.

This implies the final opinions in $R3$ is non-uniformly distributed and

$$\begin{aligned}\Omega_s^3 &= \int_{\varepsilon - \frac{3}{4}}^{\varepsilon - \frac{1}{4}} \frac{1}{2} \left[1 - \frac{k_a}{2} \left(\varepsilon - \frac{1}{4} - x \right) \right] \cdot x dx \\ &= \frac{\varepsilon}{8} - \frac{5}{96}.\end{aligned}\tag{5.36}$$

The sum of Ω_s^3 and Ω_s^4 yields the final average opinion in G_+ ,

$$\Omega_s^+ = -\frac{\varepsilon^2}{4} + \frac{\varepsilon}{4} + \frac{35}{192}.\tag{5.37}$$

Then we have

$$\begin{aligned}\Omega_s &= |\Omega_s^+ + \Omega_s^-| \\ &= \frac{8}{7-2\varepsilon} + \frac{1}{4} \left(\varepsilon - \frac{1}{2} \right)^2 - \frac{239}{192}.\end{aligned}\tag{5.38}$$

For any other possible values of ε , Ω_s can be analogously determined and is analysed in Appendix F in detail. Finally, we obtain

$$\Omega_s = \begin{cases} \frac{1}{13-4\varepsilon^2} \left(\frac{8}{3}\varepsilon^3 + 3\varepsilon^2 \right), & 0 \leq \varepsilon \leq \frac{1}{4}, \\ \frac{1}{8\varepsilon^2-26} \left(\varepsilon - \frac{391}{12} \right) + \frac{1}{6} \left(\varepsilon - \frac{1}{4} \right)^3 - \frac{5}{4}, & \frac{1}{4} < \varepsilon \leq \frac{1}{2}, \\ \frac{1}{7-2\varepsilon} \left[-\frac{4}{3} \left(\varepsilon - \frac{3}{4} \right)^3 + 2 \left(\varepsilon + \frac{1}{2} \right) \right] + \frac{1}{6} \left(\varepsilon - \frac{1}{4} \right)^3 - \frac{1}{4}, & \frac{1}{2} < \varepsilon \leq \frac{3}{4}, \\ \frac{8}{7-2\varepsilon} + \frac{1}{4} \left(\varepsilon - \frac{1}{2} \right)^2 - \frac{239}{192}, & \frac{3}{4} < \varepsilon \leq \frac{5}{4}, \\ \frac{4}{(\varepsilon - \frac{7}{4})^2 + 2} - \frac{\varepsilon}{6} \left(\varepsilon - \frac{9}{8} \right)^2 + \frac{55}{128} (\varepsilon - 1) - \frac{29}{24}, & \frac{5}{4} < \varepsilon \leq \frac{7}{4}, \\ 1, & \frac{7}{4} < \varepsilon \leq 2. \end{cases}\tag{5.39}$$

Similar analysis is also performed for scale-free networks with $k_a = 32$ (as an example of dense networks) and fully connected networks, that is, $k_a = N - 1$. When $k_a = 32$, the order parameter is expressed as

$$\Omega_s = \begin{cases} 0, & 0 \leq \varepsilon \leq \frac{1}{16}, \\ \frac{128}{105-32\varepsilon} \left[1 + \frac{1}{4} \left(\varepsilon - \frac{1}{4} \right)^2 - \frac{1}{64} \left(\varepsilon - \frac{13}{48} \right) \right] - \frac{5}{4}, & \frac{1}{16} < \varepsilon \leq \frac{1}{4}, \\ \frac{128}{105-32\varepsilon} \left[1 + \frac{4}{3} \left(\frac{5}{16} - \varepsilon \right)^3 \right] + \frac{4}{3} \left(\varepsilon - \frac{1}{4} \right)^3 - \frac{5}{4}, & \frac{1}{4} < \varepsilon \leq \frac{5}{16}, \\ \frac{128}{105-32\varepsilon} + \frac{1}{4} \left(\varepsilon - \frac{1}{4} \right)^2 - \frac{1}{64} \left(\varepsilon - \frac{13}{48} \right) - \frac{5}{4}, & \frac{5}{16} < \varepsilon \leq \frac{5}{4}, \\ \frac{2}{4(\varepsilon - \frac{21}{16})^2 + 1} - \frac{4}{3} \left(\varepsilon - \frac{5}{16} \right)^3 + 4 \left(\varepsilon - \frac{5}{16} \right) - \frac{11}{3}, & \frac{5}{4} < \varepsilon \leq \frac{21}{16}, \\ 1, & \frac{21}{16} < \varepsilon \leq 2. \end{cases}\tag{5.40}$$

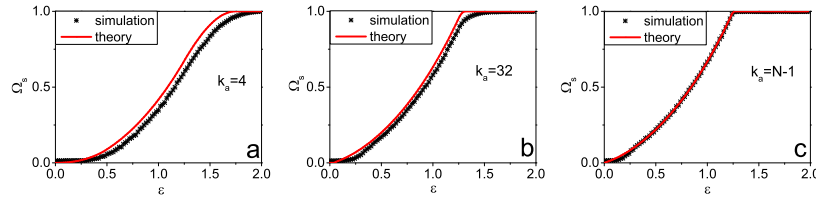


FIGURE 5.6: A comparison between numerical results (star symbols) and analytic values (solid line) for the association of Ω_s with parameter ε in keeping $p = \gamma = 0$. Numerical results are averaged over 1000 simulations with $N = 1000$ and k_a indicated on the figure. Analytic results are represented, from left to right, by Eqs. (5.39), (5.40) and (5.41).

For fully connected networks, we have

$$\Omega_s = \begin{cases} \frac{16}{13-4\varepsilon} \left[1 + \frac{1}{4}(\varepsilon - \frac{1}{4})^2 \right] - \frac{5}{4}, & 0 \leq \varepsilon \leq \frac{1}{4}; \\ \frac{16}{13-4\varepsilon} + \frac{1}{4}(\varepsilon - \frac{1}{4})^2 - \frac{5}{4}, & \frac{1}{4} < \varepsilon \leq \frac{5}{4}; \\ 1, & \frac{5}{4} < \varepsilon \leq 2. \end{cases} \quad (5.41)$$

A remarkable agreement between analytic values and numerical results is presented in Fig. 5.6, attached with a reasonable discrepancy when the friendship network is sparse (Fig. 5.6(a)). The small discrepancy is caused by mean field assumption, which gradually decreases with network parameter k_a (check plots from left to right) and will vanish when the underlying network is fully connected, i.e., when the mean field limit reaches.

5.4.2 The distribution of Ω_s in general situations

Consider different values of both p and γ . We would like to investigate the effect of parameter ε on the formation of opinion dynamics in detail. The underlying numerical results are exhibited in Fig. 5.7: the first panel relates to the distribution lines for different p by keeping $\gamma = 0$, and the second one involves the data for different γ by setting $p = 0$. For any fixed p and γ , Ω_s is positively correlated with ε . This is because higher confidence threshold guarantees the happening of more local communications between an arbitrary agent and his friends in the system under consideration, which also benefits the formation of more ordered opinion dynamics.

Equation (5.13) suggests a vanishing cross-opinion influence between two cliques when $\varepsilon \leq (1 + \gamma)/4$. Consequently, the fraction of agents whose opinions will be affected by

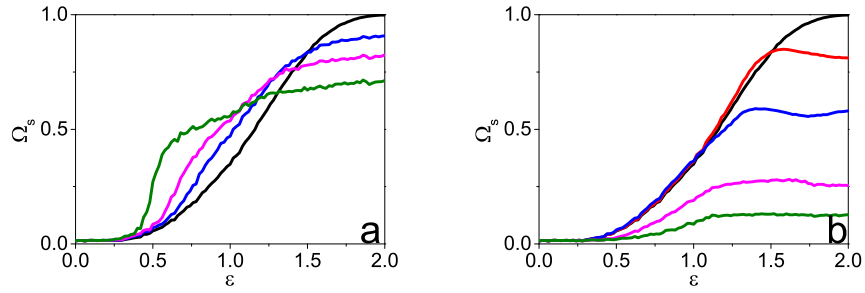


FIGURE 5.7: Steady values of order parameter Ω_s versus confidence threshold ε with the impacts of mindless agents in keeping resource undivided ($\gamma = 0$) (a) and the roles of resource allocation parameter in keeping all agents economic ($p = 0$) (b) considered, respectively. In panel (a), the fraction p of mindless agents is 0, 0.4, 0.6 and 0.8 in sequence from up to down. In panel (b), the resource allocation parameter $\gamma = 0, 0.1, 0.3, 0.5$ and 0.7 from up to down. Results are averaged over 1000 simulations with $N = 1000$ and $k_a = 4$.

others can be approximated as

$$\begin{aligned} n^f &= \int_{-\frac{1+\gamma}{4}}^{\varepsilon - \frac{1+\gamma}{4}} \frac{1}{2} \cdot \frac{k_a}{2} \left(\varepsilon - x - \frac{1+\gamma}{4} \right) dx + \int_{\frac{1+\gamma}{4} - \varepsilon}^{\frac{1+\gamma}{4}} \frac{1}{2} \cdot \frac{k_a}{2} \left(x + \varepsilon - \frac{1+\gamma}{4} \right) dx \\ &= \frac{k_a}{4} \varepsilon^2. \end{aligned} \quad (5.42)$$

Apparently, for sparse networks (k_a is small), the evolution of opinion dynamics is negligible at low confidence threshold: $\varepsilon \leq (1 + \gamma)/4$. When $\varepsilon > (1 + \gamma)/4$, cross-opinion influence between two cliques will result in non-ignorable change in the dynamics. This indicates that the variation of opinion dynamics starts from $(1 + \gamma)/4$ along ε -axis, see Fig. 5.7.

$\gamma = 0$

$\gamma = 0$ guarantees the concentration of winners only in G_- throughout the evolution of dynamics. It is difficult for economic agents in G_+ to be affected at low ε because of their primary intention of winning scores. The introducing of mindless agents then will make positive opinions move along the direction of -1 because of opinion average rule among their effective friends, so as to increase the order parameter: $\Omega_s = |\Omega^+ - \Omega^-| = |\Omega^-| - \Omega^+$. At high ε , economic agents will gradually adapt their opinions to -1 by learning from their friends. The clustering of agents is generated accordingly in opinion -1. This clustering, however, will be broken by the introducing of mindless agents. The corresponding numerical results are displayed in Fig. 5.7(a). Positive and negative correlations associated with Ω_s and p are presented respectively in the region of low ε and of high ε , which coincides with above analysis. More detailed information about opinion evolution can be found in Fig. 5.8 which provides an intuitive comparison

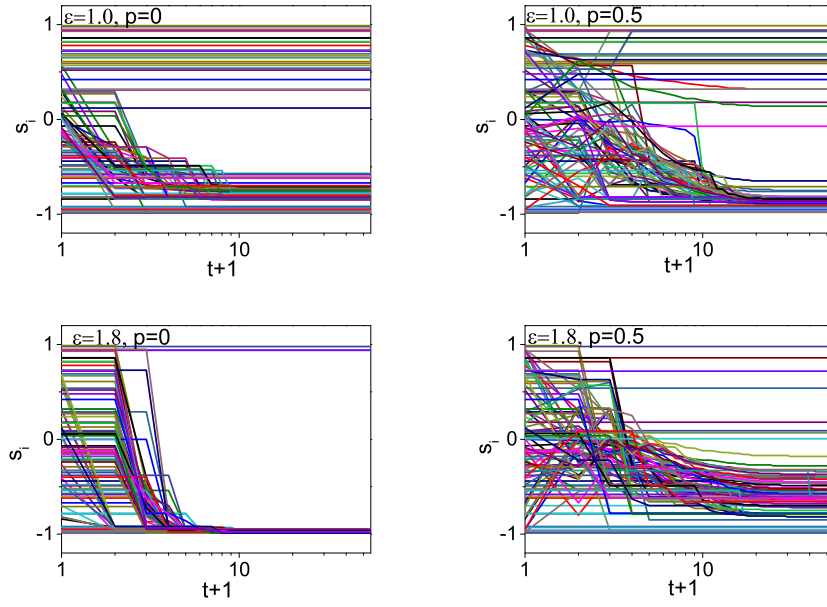


FIGURE 5.8: Show only single runs. Opinion evolution with time step t for $N = 100$, $k_a = 4$. The initial state of opinion distribution keeps the same for four considered cases. One unit time corresponds to one sampling round of gambling game.

between before and after introducing mindless agents by considering both low and high confidence thresholds.

$p = 0$

$p = 0$ says all agents involved in our gambling are economic. Numerical results suggest the presence of platform in the distribution of Ω_s for fixed γ (Fig. 5.7(b)). This implies, when ε is sufficiently high, its further increase will not influence the established dynamics. Denote the critical confidence threshold as ε_c from which a platform is generated.

Reviewing the special case with $\gamma = 0$, agents will try to migrate in opinion from G_+ to G_- for gaining scores. Eq.(5.29) states that the initial opinion of an agent i from G_+ who will be definitely affected by others must follow

$$s_i(0) \leq \varepsilon - \frac{1}{4} - \frac{2}{k_a}. \quad (5.43)$$

To ensure the migration of all agents in opinion from G_+ to G_- , we have

$$1 \leq \varepsilon - \frac{1}{4} - \frac{2}{k_a}, \quad (5.44)$$

which determines

$$\varepsilon_c = \begin{cases} \frac{5}{4} + \frac{2}{k_a}, & k_a \geq \frac{8}{3}, \\ 2, & k_a < \frac{8}{3}. \end{cases} \quad (5.45)$$

This critical value of confidence threshold indicates the emergence of consensus state in keeping $p = \gamma = 0$.

When parameter γ is limited to the range $(0, 1/3]$, initial winners must hold opinions no larger than $-(1 + \gamma)/4$. By changing $1/4$ to $(1 + \gamma)/4$ in Eq. (5.28), Eq. (5.29) is then updated by

$$s_i(0) = \varepsilon - \frac{3 + \gamma}{4}. \quad (5.46)$$

where $k_a = 4$ is inserted. To ensure agent i can be definitely affected by others that hold opinions in $[-1, -(1 + \gamma)/4]$, its initial opinion must satisfy

$$s_i(0) \leq \varepsilon - \frac{3 + \gamma}{4}. \quad (5.47)$$

By ignoring the migration of agents with initial opinions larger than $\varepsilon - (3 + \gamma)/4$ (rare event), the critical confidence threshold then can be approximated as

$$\frac{1}{2}(\varepsilon_c - \frac{3 + \gamma}{4}) = \frac{1}{2} - \frac{\gamma}{1 + \gamma}, \quad (5.48)$$

where $n_+ = \gamma/(1 + \gamma)$ (final fraction of agents in G_+) was used. Now we get

$$\varepsilon_c = \frac{3 + \gamma}{4} + \frac{1}{1 + \gamma}. \quad (5.49)$$

For comparable resource allocation between two cliques: $\gamma \in (1/3, 1]$, winners are produced in both cliques at the beginning of dynamics: $s_i(0) \leq -(1 + \gamma)/4$ or $s_i(0) \geq (1 + \gamma)/(4\gamma)$. The average probability at which an agent i with initial opinion between $-(1 + \gamma)/4$ and $(1 + \gamma)/(4\gamma)$ will be affected is represented by

$$P_i = \frac{1}{2} \left[\frac{k_a}{2}(\varepsilon - s_i(0) - \frac{1 + \gamma}{4}) + \frac{k_a}{2}(\varepsilon + s_i(0) - \frac{1 + \gamma}{4\gamma}) \right]. \quad (5.50)$$

Inserting $k_a = 4$ and setting $P_i = 1$ provide

$$\varepsilon_c = \frac{1}{2} + \frac{(1 + \gamma)^2}{8\gamma}. \quad (5.51)$$

5.4.3 Opinion Clustering

A dispersion index proposed by Derrida and Flyvberg [275] is introduced to identify agents clustering in opinions:

$$Y = \sum_{j=1}^M m_j^2, \quad (5.52)$$

where M labels the number of clusters, m_j the normalized size of j th cluster. That $Y \rightarrow 0$ as $N \rightarrow \infty$ indicates random opinion distribution in the range $[-1, 1]$ since $Y = 1/M$ labels M clusters of equal size. Finite Y corresponds to the emergence of large clusters. The generation of large cluster can be identified by the relative difference between quantity Y and the largest cluster, given by

$$R(Y, m^L) = \frac{Y - (m^L)^2}{Y}. \quad (5.53)$$

$R(Y, m^L) \approx 0$ indicates Y just considers the contribution of m^L compared to which other clusters, if exist, can be neglected, and $R(Y, m^L) \approx 1$, which is equivalent to $Y \approx 0$, indicates a vanishing large cluster but existent small ones with comparable size.

The underlying results regarding Y and $R(Y, m^L)$ are exhibited along ε -axis (Fig. 5.9). It is noticed that, in the absence of economic agents, Y is a fast increasing function of parameter ε . The emergence of large cluster compared to which other clusters, if exist, are negligible ($R(Y, m^L) \approx 0$) happens from $\varepsilon = 0.4$. Furthermore, the consensus state is definitely generated since $\varepsilon = 1$. When economic agents are present, however, the emergence of consensus state requires higher confidence threshold ε and $p = \gamma = 0$. Introducing of mindless agents or increase of the difference in resource between two cliques will destroy the clustering of agents in one opinion so as to generate small clusters with comparable size at high ε . Nevertheless, when ε is not so large (< 1.8), dispersing resource to two cliques is better for opinion clustering, compared to concentrating resource in one clique where comparable clusters in size are produced in adjacent opinions.

Colour fill diagrams of dispersion index Y (Fig. 5.10) present an intuitive interpretation for the roles of both p and γ in the clustering of agents in opinions. For fixed ε , opinion clustering happens only in the field with small p and γ ($p < 0.5$, $\gamma < 0.4$) and when p tends to 1. However, this clustering will fade away with the decrease of ε . Along with p -axis, the smallest value of Y is obtained at $p \approx 0.7$ which corresponds to the most chaotic steady state of opinion dynamics.

5.4.4 The fraction of winners

Regarding gambling game, one of most attractive observations is the fraction of winners: n_w . Numerical results of n_w are presented in γ - ε space by fixing parameter p , as shown in Fig. 5.11(a). In the upper half region of ε , increase of mindless agents will decrease the generation of winners. However, when ε is fixed between 0.5 and 1, a fragile positive correlation between n_w and p exists at low γ . Furthermore, for small p , n_w acts as a positive function of ε for each specified γ except the pattern formed at the position with $\varepsilon \approx 1.5$ and $\gamma \approx 0.1$. The pattern, corresponding to the presence of a small

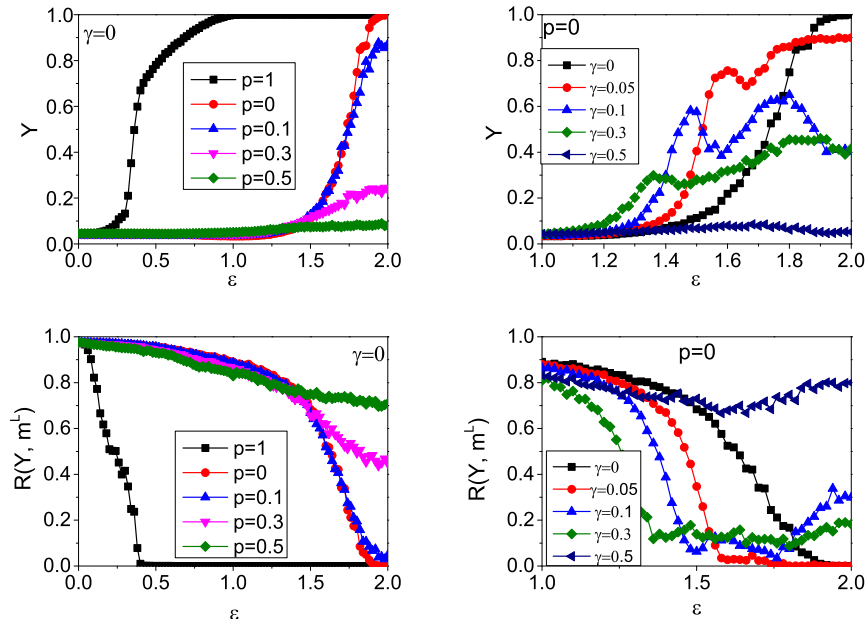


FIGURE 5.9: Steady values of dispersion index Y and relative difference $R(Y, m^L)$ versus parameter ε , for different values of p and γ indicated on the figure. Network parameters used for the results generation averaged over 1000 simulations are $N = 1000$ and $k_a = 4$.

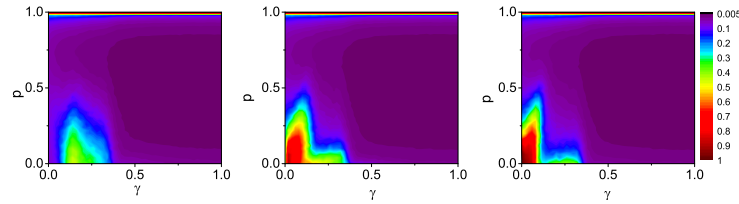
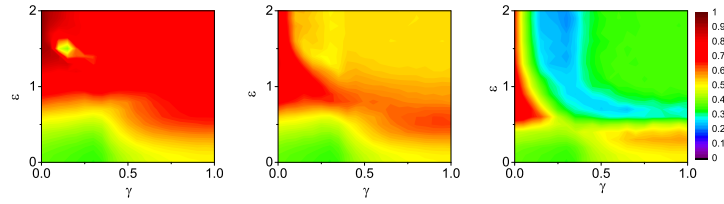


FIGURE 5.10: (Color online) Color fill diagrams of dispersion index Y , generated in γ - p space for different values of ε : 1.4, 1.7, and 2.0 from left to right. Data are averaged over 100 simulations with $N = 1000$ and $k_a = 4$.

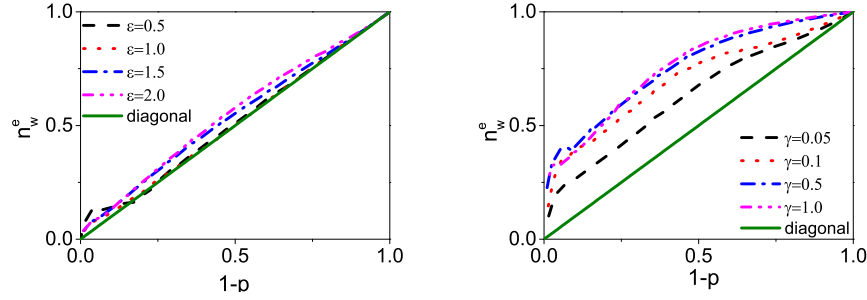
number of winners, will gradually disappear with the increase of p , accompanied by the appearance of a non-monotonic relationship between n_w and ε . Robust association of n_w with parameter γ happens only when p is large and $\varepsilon > 0.5$. A weak positive correlation between n_w and γ appears in the region of low confidence threshold ($\varepsilon < 0.5$) (opinions of almost all agents keep their initial states), which is introduced by fully disordered initial dynamics that gives

$$n_w(0) = \begin{cases} \frac{3-\gamma}{8}, & \gamma \leq \frac{1}{3}, \\ 1 - \frac{(1+\gamma)^2}{8\gamma}, & \gamma > \frac{1}{3}. \end{cases} \quad (5.54)$$

The association of the fraction of economic agents in the group of winners, denoted by N_w^e , with that in the whole system ($= 1 - p$) is also investigated for identifying who, an economic one or a mindless one, is more likely to win at our gambling game (Fig. 5.11(b)). Results display that ε plays a fragile role in deciding the probability of winning



(a) Phase diagrams of n_w generated in γ - ε space for $p=0.2, 0.5$, and 0.8 from left to right.



(b) n_w^e as a function of $(1-p)$ with $\gamma = 0$ to the left and $\varepsilon = 2$ to the right.

FIGURE 5.11: (Color online) Distribution of the fraction n_w of winners (a), and the dependence of the percentage of economic agents in winners upon the percentage of economic agents included in the system under consideration (b). Data are averaged over 100 simulations for panel (a) and 1000 simulations for panel (b) with $N = 1000$ and $k_a = 4$.

for economic or mindless agents. The increase of γ , however, enables a general economic agent to win with higher probability. Furthermore, values of n_w^e are located only in the upside of diagonal ($n_w^e = 1 - p$, corresponding to the equal opportunity of winning for economic and mindless agents). This suggests, in general case, it is more likely for economic agents than mindless ones to win at gambling.

5.4.5 Uncertainty and Sensitivity analysis

Finally, a systematic framework on sensitivity analysis of the model under discussion is performed by the use of sampling-based strategies including LHS and QMC sampling. Particularly, QMC sampling is implemented by using the high-dimensional Sobol low-discrepancy sequence generator. Sobol low-discrepancy sequence holds advanced uniformity properties in the input space, and makes QMC sampling deterministic for previously defined input space, as stated in the first chapter. This avoids highly computational cost in performing configuration average.

Regarding nonlinear models, sensitivity analysis allows one to globally understand the importance of individual input factors and of their interaction effects in establishing the uncertainty of model response. The model response of our interest here includes the order parameter Ω_s of opinion dynamics, the dispersion index Y of opinion clustering, and the fraction n_w of winners at gambling. Analysis results are listed in table 5.1. LHS-based

TABLE 5.1: Sampling-based results for the uncertainty and sensitivity analysis of the model under discussion. The number of samples used here is 10000. Each value offered by LHS-based method is averaged over 100 samplings with number in parentheses denoting the standard deviation resulted from independent samplings. For each sample, the opinion dynamics is established by averaging over 20 random configurations. For QMC method, results are obtained by a deterministic sampling. And the opinion dynamics is established by averaging over 100 random configurations for each sample.

LHS-based method								
y	$E(y)$	$V(y)$	s_p	s_γ	s_ε	s_{Tp}	$s_{T\gamma}$	$s_{T\varepsilon}$
Ω_s	0.20(.00)	0.05(.00)	0.01(.01)	0.41(.01)	0.29(.01)	0.07(.01)	0.68(.01)	0.57(.01)
Y	0.04(.00)	0.01(.00)	0.34(.02)	0.08(.02)	0.05(.01)	0.73(.02)	0.47(.03)	0.51(.03)
n_w	0.50(.00)	0.03(.00)	0.62(.02)	0.03(.02)	0.03(.03)	0.89(.02)	0.16(.01)	0.33(.02)
QMC sampling-based method								
y	$E(y)$	$V(y)$	s_p	s_γ	s_ε	s_{Tp}	$s_{T\gamma}$	$s_{T\varepsilon}$
Ω_s	0.20(87)	0.05(23)	0.00(50)	0.41(06)	0.28(38)	0.06(60)	0.67(95)	0.56(61)
Y	0.03(85)	0.00(80)	0.36(22)	0.07(41)	0.05(08)	0.74(03)	0.45(54)	0.48(88)
n_w	0.49(94)	0.03(35)	0.62(36)	0.03(48)	0.02(61)	0.89(41)	0.15(80)	0.32(03)

results are averaged over 100 samplings used for eliminating the randomness in sample generation. For each sample, the opinion dynamics is established by averaging over 20 random configurations. QMC sampling-based values are provided by a deterministic sampling. And for each sample, the opinion dynamics is established over 100 random configurations. Regarding observation Ω_s , its uncertainty is mainly contributed by the resource allocation parameter γ , confidence threshold ε , and the interaction effect between them. In establishing the uncertainty of dispersion index Y , parameter p , the fraction of mindless agents, acts as the most important factor. Furthermore, the interaction effects among three involved parameters are non-ignorable in determining the uncertainty in Y . Parameter p also plays the most important role in the generation of winners at our gambling, compared to which, the influence contributed by each of the remaining parameters alone is ignorable.

Chapter 6

Conclusions and Prospectives

6.1 Conclusions

The present research focuses on the establishment of a systematic theoretical framework on the uncertainty and sensitivity analysis of complex systems, as well as its applications to some numerical models. Specifically, the remarkable results and potential applications are concluded as below.

Firstly, an analytic formula is deduced by using Taylor series to exactly evaluate the uncertainty in output variable with given uncertainties in input factors in the absence of input correlations. With its applications to univariate power-law and exponential functions, it is analytically demonstrated that the widely used approximation with only the first order contribution of input uncertainty considered can satisfactorily explain the output variance only when the input uncertainty is negligible or the input-output relationship is close to linear. This implies the importance of high-order contributions of input uncertainty in the analysis of highly nonlinear models. Applications of the analytic method are proposed to the power grid system and EOQ model where input factors are assumed to be statistically independent of each other. Input factors of both models are analytically ranked according to their importance in establishing the uncertainty of model response.

The analytic method is then extended to the situation of input correlations. An arbitrary variable can be divided into independent and correlated parts. Coefficients that specify the correlated and independent parts are then derived by using linear correlation model. With given correlation coefficients between any two input parameters, it is straightforward to quantify the sensitivity of model response with respect to the input independence and correlations, as well as to their coupling effects. This allows one to

decide whether or not the input correlations should be considered in practice. Four numerical examples have confirmed the effectiveness and applicability of the extended analytic method. Specifically, the method is applied to a deterministic HIV model. Analysis results provide the importance of ten involved factors in determining the basic reproduction number R_0 . This may help effectively limit the spread of HIV by controlling the first three most important parameters: the rate of development to AIDS (κ), the contact rate of susceptible with symptomatic infective (β_2), and the number of sexual partners of susceptible with symptomatic infective (n_2), and simplify an HIV model by neglecting the effects of two negligible parameters: the birth rate of infective (β_0) and the fraction of susceptible newborn from infective class (γ). Moreover, the correlation between β_2 and n_2 , if exists, also provides a non-negligible effect on the basic reproduction number.

The difference of our method from Sobol's one is also discussed in the absence of input correlations. Sobol's method is implemented with the sampling-based strategy. Results suggest that our method is equivalent to Sobol's one when the nonlinear interaction effects between input factors are absent in the model under discussion. However, when the model involves nonlinear interaction effects between input factors, Sobol's method may overvalue the roles acted by individual factors but underestimate those played by their interaction effects, compared to our analytic results. Following the phenomenon, a modification is introduced to Sobol's method for establishing a sampling-based strategy that well coincides with our analytic method. The modification depends on the functional form connecting input and output variables that tells which of sensitivity indices are overvalued and which are underestimated. Furthermore, both SIR and SIS models regarding the epidemic spreading dynamics are analysed by the modified sampling-based strategy. Results suggest that the equilibrium state of SIR model more robustly depends upon the initial proportion of susceptible individuals than both infectious probability β and recovered rate γ ; their interaction effects totally contribute 30% of the information for the decision of equilibrium state. As for SIS model, parameter γ plays more crucial role than β , by explaining 49% variance in observation s . Their interaction effect also provides non-ignorable (31%) contribution to the determination of epidemic spreading.

Finally, we design a model based on a virtual gambling mechanism to investigate the dynamics formation of public opinion in the presence of limited but conserved resource and of confidence threshold. Theoretical analysis based on the traditional one-at-a-time method provides a deep understanding of the roles of both resource allocation parameter and confidence threshold in the formation of public opinion dynamics. For a sparse network, the evolution of opinion dynamics is negligible in the period of low confidence threshold when mindless agents are absent. Furthermore, the consensus state which also means a win-win state is generated only when the following three conditions are

satisfied simultaneously: mindless agents are absent, the resource is concentrated in one clique, and confidence threshold tends to a critical value which is negatively dependent upon the average number of friends of individual agents. Numerical results also imply that, for specified fraction of mindless agents, the dependence of opinion dynamics upon confidence threshold is weakened with the decrease in resource difference between two cliques. The clustering of agents in opinions is also considered. The introducing of gambling mechanism leads to a higher requirement for confidence threshold (> 1.2) to drive apparent clustering of agents, which also demands the percentage of mindless agents to be less than half and resource allocation between two cliques to be very biased. For the low fraction of mindless agents (< 0.5), resource partitioning is more advantageous to opinion clustering when confidence threshold departs from 2, compared to resource concentration. In addition, for fixed γ and ε , the most chaotic steady state of opinion dynamics is generated at $p \approx 0.7$. It is also stated that, in general cases, it is more likely for economic agents to win at gambling, instead of mindless ones. By the use of sampling-based methods, a systematic framework for the uncertainty and sensitivity analysis of the model under discussion is provided. Results state that the resource allocation parameter, confidence threshold, and their interaction effect all play robust roles in establishing the uncertainty of the order parameter of opinion dynamics. Regarding the dispersion index of opinion clustering and the fraction of winners, their uncertainties are mainly contributed by the fraction of mindless agents, and also by the interaction effects among three involved parameters. This research may contribute to the deep understanding of the generation of some social dynamics, and of the roles played by any parameters that drive the underlying dynamics.

6.2 Future work plan

Following the above research work for uncertainty and sensitivity analysis of complex systems, several aspects may deserve further considerations, whereby the roles of input parameters can be more precisely quantified in both the modeling of complex systems and the analysis of practical problems.

Sampling-based strategy design in the presence of input correlations

A theoretical framework has been established for the uncertainty and sensitivity analysis of general complex systems with given functional relationships connecting input and output variables. However, in the modeling of most complex systems (e.g., the agent-based system), it is often hard to specify the exact functional form of models. In the absence of functional forms connecting input and output variables, the uncertainty and sensitivity analysis should be performed only through sampling-based strategies. At

present, nevertheless, sampling-based strategies regarding uncertainty and sensitivity analysis have been designed by assuming that input variables are independent of each other. Following the presence of input correlations in practical applications, we shall aim at extending the sampling-based strategy to the case of input correlations. This may contribute to the application of uncertainty and sensitivity analysis in the modeling specific to more general practical problems, and help decide whether or not the input correlations should be considered in practice when the functional form connecting input and output variables is absent.

Sensitivity analysis for the economic time series

For now, the uncertainty and sensitivity analysis is mainly implemented for models with scalar output. In many situations, however, the output variable is not always a scalar but could also be a function. This commonly happens in time-dependent systems. A most typical example of time-dependent systems is the economic time series. In official statistics and applied macroeconomics, it has been customary to analyse an economic time series by extracting from it a long term movement, or trend, for separate study and then scrutinizing the residual portion for short term oscillatory movements and random fluctuations. With given time series, we shall try to quantify the roles of some possible parameters in the generation of a typical shape and fluctuations, and in trend extraction, using Sobol's measures.

Sensitivity analysis for the spread of an epidemic disease

The spread of an epidemic disease is referred to as another typical time-dependent problem. In the context of uncertainty and sensitivity analysis of models for the reproduction of disease spreading, what researchers are usually interested in is the equilibrium state of the underlying system, as that we discussed in Chapter 4. Focusing on the equilibrium state of epidemic spreading neglects the influence of temporal dimension. In practical applications, however, the evolution process of epidemic disease with time parameter is much more important than the equilibrium state, which provides fruitful information for identifying the outbreak of a disease and then makes its efficient controlling possible. We shall attempt to quantitatively identify the global roles of different parameters in the propagation process of an epidemic disease along the axis of time, whereby the outbreak of a disease could be understood more deeply.

Appendix A

Central moments

The k th central moment of a continuously distributed variable x is defined as

$$M_k(x) = \int_{-\infty}^{+\infty} (x - \mu)^k P(x) dx. \quad (\text{A.1})$$

A.1 Uniform distribution

Suppose x is uniformly distributed. Its PDF is then given by

$$P(x) = \begin{cases} 0, & x < x_0, \\ \frac{1}{x_m - x_0}, & x_0 \leq x \leq x_m, \\ 0, & x > x_m, \end{cases} \quad (\text{A.2})$$

which provides $\mu = (x_m + x_0)/2$, and standard deviation $\sigma = (x_m - x_0)/(2\sqrt{3})$. Substituting Eq. (A.2) into Eq. (A.1) yields

$$\begin{aligned} M_k(x) &= \frac{1}{x_m - x_0} \int_{x_0}^{x_m} (x - \mu)^k dx \\ &= \frac{1}{(x_m - x_0)(k+1)} \left[(x_m - \mu)^{k+1} - (x_0 - \mu)^{k+1} \right], \end{aligned} \quad (\text{A.3})$$

which can be simplified as follows by inserting $\mu = (x_m + x_0)/2$:

$$M_k(x) = \frac{1}{2(k+1)} \left(\frac{x_m - x_0}{2} \right)^k \left[1 - (-1)^{k+1} \right]. \quad (\text{A.4})$$

The above expression suggests $M_k(x)$ keeps at zero for odd k but acts as a function of σ for even k , that is

$$M_k(x) = \begin{cases} 0, & k \text{ is odd,} \\ \frac{3^{k/2}}{k+1} \sigma^k, & k \text{ is even.} \end{cases} \quad (\text{A.5})$$

A.2 Normal distribution

Regarding normally distributed x , the PDF is defined as

$$P(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \quad x \in (-\infty, +\infty). \quad (\text{A.6})$$

By inserting the above expression into Eq. (A.1), we have

$$\begin{aligned} M_k(x) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x - \mu)^k e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &\stackrel{t=x-\mu}{=} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^k e^{-\frac{t^2}{2\sigma^2}} dt \\ &\stackrel{t=u\sigma}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma^k u^k e^{-\frac{u^2}{2}} du, \end{aligned} \quad (\text{A.7})$$

which fixes at zero for odd k since it involves the integral of an odd function over a real line. While k is even, $M_k(x)$ is obtained as

$$\begin{aligned} M_k(x) &= \frac{2\sigma^k}{\sqrt{2\pi}} \int_0^{+\infty} u^k e^{-\frac{u^2}{2}} du \\ &\stackrel{u=\sqrt{2}v}{=} \frac{2\sigma^k}{\sqrt{2\pi}} \cdot 2^{\frac{1}{2}(k-1)} \int_0^{+\infty} v^{\frac{1}{2}(k-1)} e^{-v} dv \\ &= \frac{2\sigma^k}{\sqrt{2\pi}} \cdot 2^{\frac{1}{2}(k-1)} \int_0^{+\infty} v^{\frac{1}{2}(k+1)-1} e^{-v} dv \\ &= \sigma^k (k-1)!!, \end{aligned} \quad (\text{A.8})$$

where $k!!$ denotes the double factorial of k with $0!! = 1$. The k th central moment of a normally distributed variable x is obtained as

$$M_k(x) = \begin{cases} 0, & k \text{ is odd,} \\ \sigma^k (k-1)!!, & k \text{ is even.} \end{cases} \quad (\text{A.9})$$

Appendix B

Generation of correlated variables

B.1 Two correlated variables

Suppose x_1 and x_2 are only correlated with each other in the input space with given correlation coefficient ρ_{12} . By using the linear correlation model, x_1 can be formulated on the basis of x_2 as

$$x_1 = a_{12}x_2 + c_{11}r_1, \quad (\text{B.1})$$

where r_1 is the element of \mathbf{r} that satisfies Eqs. (3.14) and (3.15). Employing the definition of correlation coefficient, see Eq. (3.9), we have

$$\begin{aligned} \rho_{12} &= \frac{1}{\sigma_1\sigma_2} E[(x_1 - \mu_1)(x_2 - \mu_2)] \\ &= \frac{1}{\sigma_1\sigma_2} E[(a_{12}(x_2 - \mu_2) + c_{11}(r_1 - \mu(r_1)))(x_2 - \mu_2)] \\ &= \frac{\sigma_2}{\sigma_1} a_{12}, \end{aligned} \quad (\text{B.2})$$

which yields

$$a_{12} = \frac{\sigma_1}{\sigma_2} \rho_{12}. \quad (\text{B.3})$$

Furthermore, we also have

$$\begin{aligned} \sigma_1^2 &= E[(x_1 - \mu_1)^2] \\ &= E[(a_{12}(x_2 - \mu_2) + c_{11}(r_1 - \mu(r_1)))^2] \\ &= a_{12}^2\sigma_2^2 + c_{11}^2V(r_1). \end{aligned} \quad (\text{B.4})$$

By substituting $\sigma_1^2 = V(r_1)$, coefficient c_{11} is obtained as

$$c_{11} = \sqrt{1 - \rho_{12}^2}. \quad (\text{B.5})$$

As a consequence, x_1 , if only correlated with x_2 , can be formulated as

$$x_1 = \frac{\sigma_1}{\sigma_2} \rho_{12} x_2 + \sqrt{1 - \rho_{12}^2} r_1. \quad (\text{B.6})$$

B.2 Three correlated variables

If **correlations just exist among three variables, say x_1 , x_2 and x_3** , x_1 , for example, then can be similarly formulated by the use of linear correlation model as

$$x_1 = a_{12}x_2 + a_{13}x_3 + c_{11}r_1. \quad (\text{B.7})$$

Recalling the correlation coefficient between x_1 and x_2 , we get

$$\begin{aligned} \rho_{12} &= \frac{1}{\sigma_1 \sigma_2} E[(x_1 - \mu_1)(x_2 - \mu_2)] \\ &= \frac{1}{\sigma_1 \sigma_2} E[(a_{12}(x_2 - \mu_2) + a_{13}(x_3 - \mu_3) + c_{11}(r_1 - \mu(r_1)))(x_2 - \mu_2)] \\ &= \frac{\sigma_2}{\sigma_1} a_{12} + \frac{\sigma_3}{\sigma_1} a_{13} \rho_{23}. \end{aligned} \quad (\text{B.8})$$

Analogously, the correlation coefficient between x_1 and x_3 can be expressed as

$$\begin{aligned} \rho_{13} &= \frac{1}{\sigma_1 \sigma_3} E[(x_1 - \mu_1)(x_3 - \mu_3)] \\ &= \frac{1}{\sigma_1 \sigma_3} E[(a_{12}(x_2 - \mu_2) + a_{13}(x_3 - \mu_3) + c_{11}(r_1 - \mu(r_1)))(x_3 - \mu_3)] \\ &= \frac{\sigma_2}{\sigma_1} a_{12} \rho_{23} + \frac{\sigma_3}{\sigma_1} a_{13}, \end{aligned} \quad (\text{B.9})$$

which, together with Eq. (B.8), states

$$a_{12} = \frac{\rho_{12} - \rho_{13} \rho_{23}}{1 - \rho_{23}^2} \frac{\sigma_1}{\sigma_2}, \quad a_{13} = \frac{\rho_{13} - \rho_{12} \rho_{23}}{1 - \rho_{23}^2} \frac{\sigma_1}{\sigma_3}. \quad (\text{B.10})$$

c_{11} can be determined by the definition of the variance of x_1 , that is

$$\begin{aligned} \sigma_1^2 &= E[(x_1 - \mu_1)^2] \\ &= E[(a_{12}(x_2 - \mu_2) + a_{13}(x_3 - \mu_3) + c_{11}(r_1 - \mu(r_1)))^2] \\ &= a_{12}^2 \sigma_2^2 + a_{13}^2 \sigma_3^2 + 2a_{12}a_{13}\sigma_2\sigma_3\rho_{23} + c_{11}^2 V(r_1). \end{aligned} \quad (\text{B.11})$$

Inserting the expressions of a_{12} and a_{13} , and $\sigma_1^2 = V(r_1)$, into the above equation, we obtain

$$c_{11} = \sqrt{\frac{1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}}{1 - \rho_{23}^2}}. \quad (\text{B.12})$$

Accordingly, an arbitrary variable x_1 that is correlated with x_2 and x_3 at the same time can be generated with given pairwise correlation coefficients through equation

$$x_1 = \frac{\rho_{12} - \rho_{13}\rho_{23}}{1 - \rho_{23}^2} \frac{\sigma_1}{\sigma_2} x_2 + \frac{\rho_{13} - \rho_{12}\rho_{23}}{1 - \rho_{23}^2} \frac{\sigma_1}{\sigma_3} x_3 + \sqrt{\frac{1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}}{1 - \rho_{23}^2}} r_1. \quad (\text{B.13})$$

B.3 Four correlated variables

Consider a more complex situation where correlations exist among four variables, say x_1, x_2, x_3 and x_4 . As before, employing the linear correlation model, we can represent x_1 as

$$x_1 = a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + c_{11}r_1, \quad (\text{B.14})$$

Reviewing the definition of correlation coefficient, we have

$$\begin{aligned} \rho_{12} &= \frac{1}{\sigma_1\sigma_2} E[(x_1 - \mu_1)(x_2 - \mu_2)], \\ \rho_{13} &= \frac{1}{\sigma_1\sigma_3} E[(x_1 - \mu_1)(x_3 - \mu_3)], \\ \rho_{14} &= \frac{1}{\sigma_1\sigma_4} E[(x_1 - \mu_1)(x_4 - \mu_4)]. \end{aligned} \quad (\text{B.15})$$

Substituting the formulation of x_1 (Eq. (B.14)) into the previous equation, we get

$$\begin{aligned} \rho_{12} &= \frac{1}{\sigma_1\sigma_2} (a_{12}\sigma_2^2 + a_{13}\sigma_2\sigma_3\rho_{23} + a_{14}\sigma_2\sigma_4\rho_{24}), \\ \rho_{13} &= \frac{1}{\sigma_1\sigma_3} (a_{12}\sigma_2\sigma_3\rho_{23} + a_{13}\sigma_3^2 + a_{14}\sigma_3\sigma_4\rho_{34}), \\ \rho_{14} &= \frac{1}{\sigma_1\sigma_4} (a_{12}\sigma_2\sigma_4\rho_{24} + a_{13}\sigma_3\sigma_4\rho_{34} + a_{14}\sigma_4^2), \end{aligned} \quad (\text{B.16})$$

which drive

$$\begin{aligned} a_{12} &= \frac{\rho_{12}(1 - \rho_{34}^2) - (\rho_{13}\rho_{23} + \rho_{14}\rho_{24}) + (\rho_{13}\rho_{24} + \rho_{14}\rho_{23})\rho_{34}}{1 - \rho_{23}^2 - \rho_{24}^2 - \rho_{34}^2 + 2\rho_{23}\rho_{24}\rho_{34}} \frac{\sigma_1}{\sigma_2}, \\ a_{13} &= \frac{\rho_{13}(1 - \rho_{24}^2) - (\rho_{12}\rho_{23} + \rho_{14}\rho_{34}) + (\rho_{12}\rho_{34} + \rho_{14}\rho_{23})\rho_{24}}{1 - \rho_{23}^2 - \rho_{24}^2 - \rho_{34}^2 + 2\rho_{23}\rho_{24}\rho_{34}} \frac{\sigma_1}{\sigma_3}, \\ a_{14} &= \frac{\rho_{14}(1 - \rho_{23}^2) - (\rho_{12}\rho_{24} + \rho_{13}\rho_{34}) + (\rho_{12}\rho_{34} + \rho_{13}\rho_{24})\rho_{23}}{1 - \rho_{23}^2 - \rho_{24}^2 - \rho_{34}^2 + 2\rho_{23}\rho_{24}\rho_{34}} \frac{\sigma_1}{\sigma_4}. \end{aligned} \quad (\text{B.17})$$

Similarly, c_{11} can be obtained by the definition of the variance of x_1 :

$$\sigma_1^2 = E[(x_1 - \mu_1)^2], \quad (\text{B.18})$$

which implies

$$c_{11}^2 = 1 - \frac{1}{\sigma_1^2} (a_{12}^2 \sigma_2^2 + a_{13}^2 \sigma_3^2 + a_{14}^2 \sigma_4^2 + 2a_{12}a_{13}\sigma_2\sigma_3\rho_{23} + 2a_{12}a_{14}\sigma_2\sigma_4\rho_{24} + 2a_{13}a_{14}\sigma_3\sigma_4\rho_{34}), \quad (\text{B.19})$$

where $\sigma^2(r_1) = \sigma_1^2$ was used. Inserting Eq. (B.17) into Eq. (B.19) provides

$$\begin{aligned} c_{11} = & \left[1 - \rho_{23}^2 - \rho_{24}^2 - \rho_{34}^2 + 2\rho_{23}\rho_{24}\rho_{34} \right]^{-1/2} \left[1 - \rho_{23}^2 - \rho_{24}^2 - \rho_{34}^2 + 2\rho_{23}\rho_{24}\rho_{34} \right. \\ & - \rho_{12}^2(1 - \rho_{34}^2) - \rho_{13}^2(1 - \rho_{24}^2) - \rho_{14}^2(1 - \rho_{23}^2) + 2\rho_{12}\rho_{13}(\rho_{23} - \rho_{24}\rho_{34}) \\ & \left. + 2\rho_{13}\rho_{14}(\rho_{34} - \rho_{23}\rho_{24}) + 2\rho_{12}\rho_{14}(\rho_{24} - \rho_{23}\rho_{34}) \right]^{1/2}. \quad (\text{B.20}) \end{aligned}$$

Through the above analysis of simple cases, general expressions of the coefficients (entries of matrices **A** and **C**) that specify the independence and correlation sections separated from arbitrary variables are then derived, see Eqs. (3.16) and (3.17).

Appendix C

Derivation of Eqs. (3.51-3.52)

The form of related model function is

$$y = 2x_1 + x_2^2 + 4x_1^2x_2 + x_1x_3, \quad (\text{C.1})$$

where $(x_1, x_2, x_3) \sim N(\boldsymbol{\mu}, \Sigma)$ with mean vector $\boldsymbol{\mu} = (0, 0, 0)^T$ and covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}. \quad (\text{C.2})$$

C.1 First-order contributions

With Eq. (3.4), fractional variance contribution produced by each input factor alone are represented as

$$V_1 = \left(\frac{\partial y}{\partial x_1}\right)^2(\{\mu\}) \cdot M_2(x_1) = 4M_2(x_1), \quad (\text{C.3})$$

$$V_2 = \frac{1}{2! \cdot 2!} \left(\frac{\partial^2 y}{\partial x_2^2}\right)^2(\{\mu\}) \cdot [M_4(x_2) - M_2^2(x_2)] = M_4(x_2) - M_2^2(x_2), \quad (\text{C.4})$$

$$V_3 = \left(\frac{\partial y}{\partial x_3}\right)^2(\{\mu\}) \cdot M_2(x_3) = 0, \quad (\text{C.5})$$

To determine the independent, correlated and coupling variance contributions that are contained in V_1 , x_1 should be reformed on the bases of x_2 and x_3 as

$$x_1 = a_{12}x_2 + a_{13}x_3 + c_{11}r_1, \quad (\text{C.6})$$

in which r_1 is the element of \mathbf{r} that satisfies Eqs. (3.14) and (3.15). Coefficients $\{a_{12}, a_{13}, c_{11}\}$ are presented in Eqs. (B.10) and (B.12). Substituting the formulation

of x_1 into Eq. (C.3) yields

$$\begin{aligned} V_1 &= 4E [(a_{12}x_2 + a_{13}x_3 + c_{11}r_1)^2] \\ &= 4(a_{12}^2 + a_{13}^2 + 2a_{12}a_{13}\rho_{23} + c_{11}^2) \\ &= 4(1 - c_{11}^2) + 4c_{11}^2, \end{aligned} \quad (\text{C.7})$$

where $\mu_i = 0$ for $i = 1, 2, 3$ were used. The above expression suggests a vanishing coupling variance contribution but the existent independent and correlated ones:

$$V_1^U = 4c_{11}^2, \quad V_1^C = 4(1 - c_{11}^2), \quad V_1^{UC} = 0. \quad (\text{C.8})$$

Analogously, x_2 can be reformed as below based on x_1 and x_3 for the quantification of independent, correlated and coupling variance contributions contained in V_2 :

$$x_2 = a_{21}x_1 + a_{23}x_3 + c_{22}r_2, \quad (\text{C.9})$$

with r_2 being the element of \mathbf{r} that satisfies Eqs. (3.14) and (3.15). Coefficients $\{a_{21}, a_{23}, c_{22}\}$ determined by Eqs. (3.16) and (3.17). Inserting Eq. (C.9) into Eq. (C.4) drives

$$\begin{aligned} V_2 &= E [(a_{21}x_1 + a_{23}x_3 + c_{22}r_2)^4] - E^2 [(a_{21}x_1 + a_{23}x_3 + c_{22}r_2)^2] \\ &= E [(a_{21}^2x_1^2 + a_{23}^2x_3^2 + 2a_{21}a_{23}x_1x_3 + 2a_{21}c_{22}x_1r_2 + 2a_{23}c_{22}x_3r_2 + c_{22}^2r_2^2)^2] \\ &\quad - E^2 [(a_{21}x_1 + a_{23}x_3 + c_{22}r_2)^2] \\ &= [3(a_{21}^4 + a_{23}^4) + 6a_{21}^2a_{23}^2(2\rho_{13}^2 + 1) + 12a_{21}a_{23}\rho_{13}(a_{21}^2 + a_{23}^2)] \\ &\quad + [6c_{22}^2(a_{21}^2 + a_{23}^2 + 2a_{21}a_{23}\rho_{13})] + 3c_{22}^4 - [(a_{21}^2 + a_{23}^2 + 2a_{21}a_{23}\rho_{13}) + c_{22}^2]^2 \\ &= 2(1 - c_{22}^2)^2 + 2c_{22}^4 + 4c_{22}^2(1 - c_{22}^2), \end{aligned} \quad (\text{C.10})$$

where $\mu_i = 0$ for $i = 1, 2, 3$ were used. In the above equation, the first term is produced by the correlations of x_2 with the remaining inputs, the second one by its independence, and the third one by the coupling effect associated with correlations and independence, indicating

$$V_2^U = 2c_{22}^4, \quad V_2^C = 2(1 - c_{22}^2)^2, \quad V_2^{UC} = 4c_{22}^2(1 - c_{22}^2). \quad (\text{C.11})$$

C.2 Second-order contributions

The second-order partial variance contributions produced by the combinations between each pair of inputs are expressed as

$$\begin{aligned} V_{12} &= \frac{1}{2! \cdot 2!} \left(\frac{\partial^3 y}{\partial x_1^2 \partial x_2} \right)^2(\{\mu\}) \cdot M_{4,2}(x_1^4, x_2^2) + \frac{2}{2!} \left(\frac{\partial y}{\partial x_1} \cdot \frac{\partial^3 y}{\partial x_1^2 \partial x_2} \right)(\{\mu\}) \cdot M_{3,1}(x_1^3, x_2) \\ &= 16 [M_{4,2}(x_1^4, x_2^2) + M_{3,1}(x_1^3, x_2)], \end{aligned} \quad (\text{C.12})$$

$$V_{13} = \left(\frac{\partial^2 y}{\partial x_1 \partial x_3} \right)^2(\{\mu\}) \cdot [M_{2,2}(x_1^2, x_3^2) - M_{1,1}^2(x_1, x_3)] = M_{2,2}(x_1^2, x_3^2) - \rho_{13}^2, \quad (\text{C.13})$$

$$V_{23} = 0. \quad (\text{C.14})$$

Considering the reforming of x_1 on the bases of x_2 and x_3 , the partial variance contribution V_{12} is determined as

$$\begin{aligned} V_{12} &= 16E [x_2^2(a_{12}x_2 + a_{13}x_3 + c_{11}r_1)^4] + 16E [x_2(a_{12}x_2 + a_{13}x_3 + c_{11}r_1)^3] \\ &= 16E [x_2^2(a_{12}^2x_2^2 + a_{13}^2x_3^2 + 2a_{12}a_{13}x_2x_3 + 2a_{12}c_{11}x_2r_1 + 2a_{13}c_{11}x_3r_1 + c_{11}^2r_1^2)^2] \\ &\quad + 16E [x_2(a_{12}^3x_2^3 + a_{13}^3x_3^3 + 3a_{12}^2a_{13}x_2^2x_3 + 3a_{12}a_{13}^2x_2x_3^2)] \\ &\quad + 16E [x_2(3c_{11}(a_{12}^2x_2^2 + a_{13}^2x_3^2 + 2a_{12}a_{13}x_2x_3)r_1 + 3c_{11}^2(a_{12}x_2 + a_{13}x_3)r_1^2 + c_{11}^3r_1^3)] \\ &= 48(1 - c_{11}^2)(1 - c_{11}^2 + \rho_{12} + 4\rho_{12}^2) + 48c_{11}^4 + 48c_{11}^2(2 - 2c_{11}^2 + \rho_{12} + 4\rho_{12}^2), \end{aligned} \quad (\text{C.15})$$

which, from the first fraction to the third one, are separately produced by the correlations of x_1 with the rest, the independence of x_1 , and the coupling effect between correlations and independence, specifying

$$\begin{aligned} V_{12}^{C_1} &= 48(1 - c_{11}^2)(1 - c_{11}^2 + \rho_{12} + 4\rho_{12}^2), \\ V_{12}^{U_1} &= 48c_{11}^4, \quad V_{12}^{UC_1} = 48c_{11}^2(2 - 2c_{11}^2 + \rho_{12} + 4\rho_{12}^2). \end{aligned} \quad (\text{C.16})$$

Regarding the formulation of input x_2 , V_{12} can also be calculated as

$$\begin{aligned} V_{12} &= 16 [x_1^4(a_{21}x_1 + a_{23}x_3 + c_{22}r_2)^2] + 16E [x_1^3(a_{21}x_1 + a_{23}x_3 + c_{22}r_2)] \\ &= 16 [15a_{21}^2 + a_{23}^2(12\rho_{13}^2 + 3) + 30a_{21}a_{23}\rho_{13} + 3a_{21} + 3a_{23}\rho_{13}] + 48c_{22}^2 \\ &= 48(1 - c_{22}^2 + \rho_{12} + 4\rho_{12}^2) + 48c_{22}^2, \end{aligned} \quad (\text{C.17})$$

which shows a vanishing coupling variance contribution but the existent correlated (first term with brackets) and independent (last term) ones:

$$V_{12}^{U_2} = 48c_{22}^2, \quad V_{12}^{C_2} = 48(1 - c_{22}^2 + \rho_{12} + 4\rho_{12}^2), \quad V_{12}^{UC_2} = 0. \quad (\text{C.18})$$

Recalling the formulation of x_1 on the bases of x_2 and x_3 , the partial variance contribution V_{13} can be analogously obtained as

$$\begin{aligned}
V_{13} &= E \left[x_3^2 \cdot (a_{12}x_2 + a_{13}x_3 + c_{11}r_1)^2 \right] - \rho_{13}^2 \\
&= a_{12}^2 (2\rho_{23}^2 + 1) + 3a_{13}^2 + 6a_{12}a_{13}\rho_{23} - \rho_{13}^2 + c_{11}^2 \\
&= 1 - c_{11}^2 + 2a_{12}\rho_{23}(a_{13} + a_{12}\rho_{23}) + 2a_{13}(a_{13} + a_{12}\rho_{23}) - \rho_{13}^2 + c_{11}^2 \\
&= 1 - c_{11}^2 + \rho_{13}^2 + c_{11}^2,
\end{aligned} \tag{C.19}$$

which only contains the correlated variance contribution (first three components) produced by the correlations of x_1 with the rest variables, and independent one (last component) produced by the independence of x_1 :

$$V_{13}^{U_1} = c_{11}^2, \quad V_{13}^{C_1} = 1 - c_{11}^2 + \rho_{13}^2, \quad V_{13}^{UC_1} = 0. \tag{C.20}$$

V_{13} can be equivalently obtained by formulating x_3 on the bases of x_1 and x_2 , constituted of

$$V_{13}^{U_3} = c_{33}^2, \quad V_{13}^{C_3} = 1 - c_{33}^2 + \rho_{13}^2, \quad V_{13}^{UC_3} = 0. \tag{C.21}$$

C.3 Third-order contributions

The third-order partial variance contribution associated with the second test case is expressed as

$$\begin{aligned}
V_{123} &= \frac{2}{2!} \left(\frac{\partial^2 y}{\partial x_2^2} \cdot \frac{\partial^2 y}{\partial x_1 \partial x_3} \right) (\{\mu\}) [M_{1,2,1}(x_1, x_2^2, x_3) - M_2(x_2)M_{1,1}(x_1, x_3)] \\
&= M_{1,2,1}(x_1, x_2^2, x_3) - M_2(x_2)\rho_{13}.
\end{aligned} \tag{C.22}$$

By introducing the reforming of x_2 (Eq. (C.9)), we get

$$\begin{aligned}
V_{123} &= E [x_1 x_3 (a_{21}x_1 + a_{23}x_3 + c_{22}r_2)^2] - E [(a_{21}x_1 + a_{23}x_3 + c_{22}r_2)^2] \rho_{13} \\
&= E [x_1 x_3 (a_{21}x_1 + a_{23}x_3)^2] - E [(a_{21}x_1 + a_{23}x_3)^2] \rho_{13} + E [c_{22}^2 x_1 x_3 r_2^2] \\
&\quad - E [c_{22}^2 r_2^2] \rho_{13},
\end{aligned} \tag{C.23}$$

where $\mu_i = 0$ for $i = 1, 2, 3$ were used. In the above equation, the first two items are contributed by the correlations of x_2 with both x_1 and x_3 , and the last two items, summing to zero, are provided by the coupling effect between the independence of x_2

and the correlation of x_1 with x_3 . V_{123} is then computed as

$$\begin{aligned}
 V_{123} &= 2a_{12}^2\rho_{13} + 2a_{23}^2\rho_{13} + 2a_{12}a_{23}\rho_{13}^2 + 2a_{12}a_{23} \\
 &= 2a_{12}\rho_{13}(a_{12} + a_{23}\rho_{13}) + 2a_{23}(a_{12} + a_{23}\rho_{13}) \\
 &= 2a_{12}\rho_{12}\rho_{13} + 2a_{23}\rho_{12} \\
 &= 2\rho_{12}\rho_{23},
 \end{aligned} \tag{C.24}$$

which is totally contributed by input correlations:

$$V_{123}^{\text{U}_2} = V_{123}^{\text{UC}_2} = 0, \quad V_{123}^{\text{C}_2} = 2\rho_{12}\rho_{23}. \tag{C.25}$$

We can analogously state that

$$V_{123}^{\text{U}_1} = V_{123}^{\text{UC}_1} = V_{123}^{\text{U}_3} = V_{123}^{\text{UC}_3} = 0, \quad V_{123}^{\text{C}_1} = V_{123}^{\text{C}_3} = 2\rho_{12}\rho_{23}. \tag{C.26}$$

Appendix D

Derivation of the fourth-order central moment with four different variables

In the second nonlinear test case presented in Chapter 3, the fourth-order central moment with four different variables is involved:

$$M_{1,1,1,1}(x_1, x_2, x_3, x_4) = E[(x_1 - \mu_1)(x_2 - \mu_2)(x_3 - \mu_3)(x_4 - \mu_4)], \quad (\text{D.1})$$

where $(x_1, x_2, x_3, x_4) \sim N(\boldsymbol{\mu}, \Sigma)$ with mean vector $\boldsymbol{\mu} = (1, 2, 2, 1)^T$ and covariance matrix displayed in (3.54). Select x_1 (equivalent to select each of the others) to be reformed on the bases of the remaining factors:

$$x_1 = a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + c_{11}r_1, \quad (\text{D.2})$$

where r_1 is the element of \mathbf{r} that satisfies Eqs. (3.14) and (3.15). Coefficients $\{a_{12}, a_{13}, a_{14}, c_{11}\}$ are presented in Eqs. (B.17) and (B.20). Inserting (D.2) into (D.1) provides

$$\begin{aligned} M_{1,1,1,1}(x_1, x_2, x_3, x_4) &= E[(a_{12}x'_2 + a_{13}x'_3 + a_{14}x'_4 + c_{11}r'_1)x'_2x'_3x'_4] \\ &= E[a_{12}(x'_2)^2x'_3x'_4] + E[a_{13}x'_2(x'_3)^2x'_4] + E[a_{14}x'_2x'_3(x'_4)^2], \end{aligned} \quad (\text{D.3})$$

where $x'_i = x_i - \mu_i$ for $i = 2, 3, 4$ and $r'_1 = r_1 - \mu(r_1)$. In the determination of $E[a_{12}(x'_2)^2x'_3x'_4]$, another variable should also be reformed based on the others enclosed in the brackets. Here we formulate x_3 (equivalent to formulate x_2 or x_4 as input

variables are normally distributed and hold the same standard deviation) as

$$x_3 = a_{32}x_2 + a_{34}x_4 + c_{33}r_3, \quad (\text{D.4})$$

where r_3 is the element of \mathbf{r} that satisfies Eqs. (3.14) and (3.15). Coefficients $\{a_{32}, a_{34}, c_{33}\}$ are determined via Eqs. (3.16) and (3.17) as

$$a_{32} = (\rho_{23} - \rho_{24}\rho_{34})/(1 - \rho_{24}^2), \quad (\text{D.5})$$

$$a_{34} = (\rho_{34} - \rho_{23}\rho_{24})/(1 - \rho_{24}^2), \quad (\text{D.6})$$

$$c_{33} = (1 - \rho_{23}^2 - \rho_{24}^2 - \rho_{34}^2 + 2\rho_{23}\rho_{24}\rho_{34})^{1/2}/(1 - \rho_{24}^2)^{1/2}. \quad (\text{D.7})$$

Then we can obtain

$$\begin{aligned} E [a_{12}(x_2')^2 x_3' x_4'] &= [a_{12}a_{32}(x_2')^3 x_4' + a_{12}a_{34}(x_2')^2 (x_4')^2] \\ &= 2a_{12}\rho_{24}(a_{32} + a_{34}\rho_{24}) + a_{12}(a_{34} + a_{32}\rho_{24}) \\ &= 2a_{12}\rho_{23}\rho_{24} + a_{12}\rho_{34}. \end{aligned} \quad (\text{D.8})$$

The other two average items in Eq. (D.3) can be similarly determined as

$$E [a_{13}x_2'(x_3')^2 x_4'] = 2a_{13}\rho_{23}\rho_{34} + a_{13}\rho_{24}, \quad (\text{D.9})$$

$$E [a_{14}x_2'x_3'(x_4')^2] = 2a_{14}\rho_{24}\rho_{34} + a_{14}\rho_{23}. \quad (\text{D.10})$$

We now get

$$\begin{aligned} M_{1,1,1,1}(x_1, x_2, x_3, x_4) &= 2(a_{12}\rho_{23}\rho_{24} + a_{13}\rho_{23}\rho_{34} + a_{14}\rho_{24}\rho_{34}) + (a_{12}\rho_{34} + a_{13}\rho_{24} + a_{14}\rho_{23}) \\ &= \rho_{34}(a_{12} + a_{13}\rho_{23} + a_{14}\rho_{24}) + \rho_{24}(a_{13} + a_{12}\rho_{23} + a_{14}\rho_{34}) \\ &\quad + \rho_{23}(a_{14} + a_{12}\rho_{24} + a_{13}\rho_{34}) \\ &= \rho_{12}\rho_{34} + \rho_{13}\rho_{24} + \rho_{14}\rho_{23}. \end{aligned} \quad (\text{D.11})$$

Appendix E

Partial variance contributions for Ishigami function

The Ishigami function is expressed as

$$y = \sin(x_1) + 7 \sin^2(x_2) + 0.1 x_3^4 \sin(x_1), \quad (\text{E.1})$$

with input variables uniformly distributed in the interval $[-\pi, \pi]$ which provides $\mu_i=0$ and $\sigma_i = \pi/\sqrt{3}$ for $i = 1, 2, 3$.

E.1 First-order contributions

With help of Eq. (3.4), the main partial variance contribution produced by input x_1 alone is represented as

$$V_1 = \sum_{i,j=0}^{\infty} \frac{1}{i! \cdot j!} \left(\frac{\partial^i y}{\partial x_1^i} \cdot \frac{\partial^j y}{\partial x_1^j} \right) (\{\mu\}) \cdot [M_{i+j}(x_1) - M_i(x_1)M_j(x_1)]. \quad (\text{E.2})$$

The existent partial derivatives of y with respect to x_1 are totally provided by the first term of Eq. (E.1): $\sin(x_1)$ as $\mu_3 = 0$. Zero mean value of x_1 determines

$$\left(\frac{\partial^i \sin(x_1)}{\partial x_1^i} \Big|_{\mu_1=0} \right) = \begin{cases} (-1)^{(i-1)/2}, & i \text{ is odd,} \\ 0, & i \text{ is even.} \end{cases} \quad (\text{E.3})$$

Substituting Eqs. (A.5) and (E.3) into Eq. (E.2) yields

$$V_1 = \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j}}{(2i+1)!(2j+1)!} \cdot \frac{\pi^{2(i+j+1)}}{2i+2j+3} = 0.5. \quad (\text{E.4})$$

Analogously, the main variance contribution produced by x_2 is represented as

$$V_2 = \sum_{i,j=0}^{\infty} \frac{1}{i! \cdot j!} \left(\frac{\partial^i y}{\partial x_2^i} \cdot \frac{\partial^j y}{\partial x_2^j} \right) (\{\mu\}) \cdot [M_{i+j}(x_2) - M_i(x_2)M_j(x_2)], \quad (\text{E.5})$$

which is provided by the second term of Eq. (E.1): $7 \sin^2(x_2)$. Zero mean value of x_2 suggests

$$\left(\frac{\partial^i \sin^2(x_2)}{\partial x_2^i} \right) \Big|_{\mu_2=0} = \begin{cases} 2^{i-1}(-1)^{(i+2)/2}, & i \text{ is even,} \\ 0, & i \text{ is odd.} \end{cases} \quad (\text{E.6})$$

By inserting Eqs. (A.5) and (E.6) into Eq. (E.5), we get

$$V_2 = 7^2 \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j} 4^{i+j+1} \pi^{2(i+j+2)}}{(2i+2)!(2j+2)!} \left[\frac{1}{2i+2j+5} - \frac{1}{(2i+3)(2j+3)} \right] = 6.125. \quad (\text{E.7})$$

Zero mean value of x_1 also suggests a vanishing main partial variance contribution produced by x_3 alone since x_3 just appears in the combination with sin function of x_1 in the form of Ishigami function. x_2 is assumed to be independent. But x_1 and x_3 may correlate with each other. To interpret the independent, correlated and also coupling variance contributions included in V_1 , x_1 should be formulated on the basis of x_3 :

$$\begin{aligned} x_1 &= \rho_{13} \frac{\sigma_1}{\sigma_3} x_3 + \sqrt{1 - \rho_{13}^2} r_1 \\ &= \rho_{13} x_3 + \sqrt{1 - \rho_{13}^2} r_1, \end{aligned} \quad (\text{E.8})$$

with r_1 holding the same distribution as x_1 (since the mean values of x_1 and x_3 are zero). Equation (E.4) is then rewritten as

$$V_1 = \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j}}{(2i+1)!(2j+1)!} \cdot E \left[(\rho_{13} x_3 + \sqrt{1 - \rho_{13}^2} r_1)^{2(i+j+1)} \right], \quad (\text{E.9})$$

where $\mu_3 = \mu(r_1) = 0$ were used. By employing the binomial theory, the above expression is updated as

$$V_1 = \sum_{i,j=0}^{\infty} \sum_{k=0}^{i+j+1} \binom{2(i+j+1)}{2k} \frac{(-1)^{i+j} \cdot \pi^{2(i+j+1)}}{(2i+1)!(2j+1)!} \cdot \frac{\rho_{13}^{2k} (1 - \rho_{13}^2)^{i+j+1-k}}{(2i+2j+3-2k)(2k+1)}. \quad (\text{E.10})$$

The independent and correlated variance contributions are then evaluated by setting $k = 0$ and $k = i + j + 1$, respectively:

$$V_1^U = \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j} \cdot \pi^{2(i+j+1)}}{(2i+1)!(2j+1)!} \cdot \frac{(1 - \rho_{13}^2)^{i+j+1}}{2i+2j+3}, \quad (\text{E.11})$$

$$V_1^C = \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j} \cdot \pi^{2(i+j+1)}}{(2i+1)!(2j+1)!} \cdot \frac{\rho_{13}^{2(i+j+1)}}{2i+2j+3}. \quad (\text{E.12})$$

The coupling variance contribution is then evaluated accordingly by

$$V_1^{UC} = V_1 - V_1^U - V_1^C. \quad (\text{E.13})$$

E.2 Second-order contributions

For partial variance contributions of second-order, the form of Ishigami function provides $V_{12} = V_{23} = 0$ and

$$\begin{aligned} V_{13} = & 0.1^2 \sum_{i,j=0}^{\infty} \frac{1}{i! \cdot j! \cdot 4! \cdot 4!} \left(\frac{\partial^{4+i} y}{\partial x_1^i \partial x_3^4} \cdot \frac{\partial^{4+j} y}{\partial x_1^j \partial x_3^4} \right) (\{\mu\}) \cdot \left[M_{i+j,8}(x_1^{i+j}, x_3^8) - M_{i,4}(x_1^i, x_3^4) M_{j,4}(x_1^j, x_3^4) \right] \\ & + 2 \cdot 0.1 \sum_{i,j=0}^{\infty} \frac{1}{i! \cdot j! \cdot 4!} \left(\frac{\partial^i y}{\partial x_1^i} \cdot \frac{\partial^{4+j} y}{\partial x_1^j \partial x_3^4} \right) (\{\mu\}) \cdot \left[M_{i+j,4}(x_1^{i+j}, x_3^4) - M_i(x_1) M_{j,4}(x_1^j, x_3^4) \right] \\ = & 0.1^2 \sum_{i,j=0}^{\infty} \frac{1}{i! \cdot j!} \left(\frac{\partial^i \sin(x_1)}{\partial x_1^i} \cdot \frac{\partial^j \sin(x_1)}{\partial x_1^j} \right) (\mu_1 = 0) \cdot \left[M_{i+j,8}(x_1^{i+j}, x_3^8) - M_{i,4}(x_1^i, x_3^4) M_{j,4}(x_1^j, x_3^4) \right] \\ & + 2 \cdot 0.1 \sum_{i,j=0}^{\infty} \frac{1}{i! \cdot j!} \left(\frac{\partial^i \sin(x_1)}{\partial x_1^i} \cdot \frac{\partial^j \sin(x_1)}{\partial x_1^j} \right) (\mu_1 = 0) \cdot \left[M_{i+j,4}(x_1^{i+j}, x_3^4) - M_i(x_1) M_{j,4}(x_1^j, x_3^4) \right] \end{aligned} \quad (\text{E.14})$$

By substituting Eq. (E.3), V_{13} is updated as

$$V_{13} = 0.1 \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j}}{(2i+1)!(2j+1)!} \left[0.1 M_{2i+2j+2,8}(x_1^{2i+2j+2}, x_3^8) + 2 M_{2i+2j+2,4}(x_1^{2i+2j+2}, x_3^4) \right]. \quad (\text{E.15})$$

To determine V_{13} , it is necessary to derive the covariance items enclosed in the above square parentheses.

Firstly, consider the generation of x_1 on the basis of x_3 . Inserting Eq. (E.8) into covariance items in Eq. (E.14) yields

$$\begin{aligned}
M_{2i+2j+2,8}(x_1^{2(i+j+1)}, x_3^8) \\
&= E \left[x_3^8 \sum_{k=0}^{i+j+1} \binom{2(i+j+1)}{2k} \rho_{13}^{2k} (1 - \rho_{13}^2)^{i+j+1-k} x_3^{2k} r_1^{2i+2j+2-2k} \right] \\
&= \sum_{k=0}^{i+j+1} \binom{2(i+j+1)}{2k} \rho_{13}^{2k} (1 - \rho_{13}^2)^{i+j+1-k} M_{2i+2j+2-2k}(r_1) M_{8+2k}(x_3) \\
&= \sum_{k=0}^{i+j+1} \binom{2(i+j+1)}{2k} \cdot \frac{\pi^{2(i+j+1)+8} \rho_{13}^{2k} (1 - \rho_{13}^2)^{i+j+1-k}}{(2(i+j+1) - 2k + 1)(8 + 2k + 1)},
\end{aligned} \tag{E.16}$$

and

$$\begin{aligned}
M_{2i+2j+2,4}(x_1^{2(i+j+1)}, x_3^4) \\
&= E \left[x_3^4 \sum_{k=0}^{i+j+1} \binom{2(i+j+1)}{2k} \rho_{13}^{2k} (1 - \rho_{13}^2)^{i+j+1-k} x_3^{2k} r_1^{2i+2j+2-2k} \right] \\
&= \sum_{k=0}^{i+j+1} \binom{2(i+j+1)}{2k} \rho_{13}^{2k} (1 - \rho_{13}^2)^{i+j+1-k} M_{2(i+j+1)-2k}(x_1) M_{4+2k}(x_3) \\
&= \sum_{k=0}^{i+j+1} \binom{2(i+j+1)}{2k} \cdot \frac{\pi^{2(i+j+1)+4} \rho_{13}^{2k} (1 - \rho_{13}^2)^{i+j+1-k}}{(2(i+j+1) - 2k + 1)(4 + 2k + 1)}.
\end{aligned} \tag{E.17}$$

V_{13} is now obtained as

$$\begin{aligned}
V_{13} = 0.1 \sum_{i,j=0}^{\infty} \sum_{k=0}^{i+j+1} \frac{(-1)^{i+j} \pi^{2(i+j+1)+4}}{(2i+1)!(2j+1)!} \cdot \binom{2(i+j+1)}{2k} \\
\times \frac{\rho_{13}^{2k} (1 - \rho_{13}^2)^{i+j+1-k}}{2(i+j-k) + 3} \left[\frac{0.1 \cdot \pi^4}{9 + 2k} + \frac{2}{5 + 2k} \right].
\end{aligned} \tag{E.18}$$

The Independent and correlated partial variance contributions included in V_{13} are specified by setting $k = 0$ and $k = i + j + 1$, respectively:

$$V_{13}^{U_1} = 0.1 \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j} \pi^{2(i+j+1)+4}}{(2i+1)!(2j+1)!} \frac{(1 - \rho_{13}^2)^{i+j+1}}{2(i+j) + 3} \left[\frac{0.1 \cdot \pi^4}{9} + \frac{2}{5} \right], \tag{E.19}$$

$$V_{13}^{C_1} = 0.1 \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j} \pi^{2(i+j+1)+4}}{(2i+1)!(2j+1)!} \rho_{13}^{2(i+j+1)} \left[\frac{0.1 \cdot \pi^4}{2(i+j) + 11} + \frac{2}{2(i+j) + 7} \right]. \tag{E.20}$$

The coupling variance contribution is then spontaneously determined for x_1 by

$$V_{13}^{\text{UC}_1} = V_{13} - V_{13}^{\text{U}_1} - V_{13}^{\text{C}_1}. \quad (\text{E.21})$$

Optionally, x_3 can be also formulated in terms of x_1 as:

$$\begin{aligned} x_3 &= \rho_{13} \frac{\sigma_3}{\sigma_1} x_1 + \sqrt{1 - \rho_{13}^2} r_3 \\ &= \rho_{13} x_1 + \sqrt{1 - \rho_{13}^2} r_3, \end{aligned} \quad (\text{E.22})$$

with r_3 holding the same distribution as x_3 . Then we get

$$\begin{aligned} M_{2i+2j+2,8}(x_1^{2(i+j+1)}, x_3^8) \\ &= E \left[x_1^{2(i+j+1)} \sum_{k=0}^4 \binom{8}{2k} \rho_{13}^{2k} (1 - \rho_{13}^2)^{4-k} x_1^{2k} r_3^{8-2k} \right] \\ &= \sum_{k=0}^4 \binom{8}{2k} \rho_{13}^{2k} (1 - \rho_{13}^2)^{4-k} M_{2(i+j+1)+2k}(x_1) M_{8-2k}(x_3) \\ &= \sum_{k=0}^4 \binom{8}{2k} \cdot \frac{\pi^{2(i+j+1)+8} \rho_{13}^{2k} (1 - \rho_{13}^2)^{4-k}}{(2(i+j+1) + 2k + 1)(8 - 2k + 1)}, \end{aligned} \quad (\text{E.23})$$

and

$$\begin{aligned} M_{2i+2j+2,4}(x_1^{2(i+j+1)}, x_3^4) \\ &= E \left[x_1^{2(i+j+1)} \sum_{l=0}^2 \binom{4}{2l} \rho_{13}^{2l} (1 - \rho_{13}^2)^{2-l} x_1^{2l} r_3^{4-2l} \right] \\ &= \sum_{l=0}^2 \binom{4}{2l} \rho_{13}^{2l} (1 - \rho_{13}^2)^{2-l} M_{2(i+j+1)+2l}(x_1) M_{4-2l}(x_3) \\ &= \sum_{l=0}^2 \binom{4}{2l} \cdot \frac{\pi^{2(i+j+1)+4} \rho_{13}^{2l} (1 - \rho_{13}^2)^{2-l}}{(2(i+j+1) + 2l + 1)(4 - 2l + 1)}, \end{aligned} \quad (\text{E.24})$$

which yield

$$\begin{aligned} V_{13} &= 0.1 \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j} \pi^{2(i+j+1)+4}}{(2i+1)!(2j+1)!} \left[0.1 \sum_{k=0}^4 \binom{8}{2k} \cdot \frac{\pi^4 \rho_{13}^{2k} (1 - \rho_{13}^2)^{4-k}}{(2i+2j+2k+3)(9-2k)} \right. \\ &\quad \left. + 2 \sum_{l=0}^2 \binom{4}{2l} \cdot \frac{\rho_{13}^{2l} (1 - \rho_{13}^2)^{2-l}}{(2i+2j+2l+3)(5-2l)} \right]. \end{aligned} \quad (\text{E.25})$$

The independent and correlated effects contained in the above partial variance contribution can be calculated by setting $k = l = 0$ and $k = 4, l = 2$, respectively:

$$V_{13}^{U_3} = 0.1 \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j} \pi^{2(i+j+1)+4}}{(2i+1)!(2j+1)!} \left[\frac{0.1 \cdot \pi^4 (1 - \rho_{13}^2)^4}{9(2i+2j+3)} + \frac{2(1 - \rho_{13}^2)^2}{5(2i+2j+3)} \right], \quad (\text{E.26})$$

$$V_{13}^{C_3} = 0.1 \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j} \pi^{2(i+j+1)+4}}{(2i+1)!(2j+1)!} \left[\frac{0.1 \cdot \pi^4 \rho_{13}^8}{2i+2j+11} + \frac{2\rho_{13}^4}{2i+2j+7} \right]. \quad (\text{E.27})$$

The coupling variance contribution is then determined for x_3 by

$$V_{13}^{UC_3} = V_{13} - V_{13}^{U_3} - V_{13}^{C_3}. \quad (\text{E.28})$$

Appendix F

Special case for opinion formation: $p = \gamma = 0$

Regarding sparse scale-free networks with $k_a = 4$, the theoretical analysis of the association of opinion dynamics with confidence threshold ε is completed in detail in this part by keeping $p = \gamma = 0$.

When the confidence threshold is quite low: $\varepsilon \leq 1/4$, we can divide the opinion space into four adjacent regions, as shown in Fig. F.1. With time evolution of opinion dynamics, the fraction of agents who will migrate in opinion from $R2$ to $R1$ is

$$\begin{aligned} n_{2 \rightarrow 1} &= \int_{-\frac{1}{4}}^{\varepsilon - \frac{1}{4}} \frac{1}{2} \cdot \frac{k_a}{2} (\varepsilon - x - \frac{1}{4}) dx \\ &= \frac{1}{2} \varepsilon^2, \end{aligned} \tag{F.1}$$

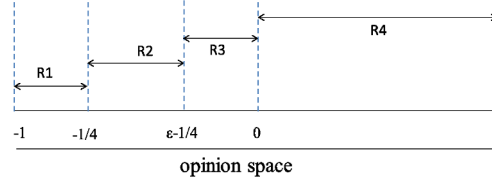
which determines the final fraction of agents in $R1$ as

$$n_1 = \frac{3}{8} + \frac{1}{2} \varepsilon^2. \tag{F.2}$$

Analogous to Eq. (5.36), at steady state, the average opinion in the region $R2$ can be specified as

$$\begin{aligned} \Omega_s^2 &= \int_{-\frac{1}{4}}^{\varepsilon - \frac{1}{4}} \frac{1}{2} [1 - \frac{k_a}{2} (\varepsilon - x - \frac{1}{4})] \cdot x dx \\ &= -\frac{1}{6} \varepsilon^3 + \frac{3}{8} \varepsilon^2 - \frac{1}{8} \varepsilon. \end{aligned} \tag{F.3}$$

With respect to the limitation of confidence threshold, opinion distributions in both $R3$ and $R4$ will keep at their initial configurations (uniformly distributed) throughout the

FIGURE F.1: The generation of four adjacent regions in the opinion space when $\varepsilon \in [0, 1/4]$.

time evolution. This provides

$$\Omega_s^3 = -\frac{1}{4}\left(\varepsilon - \frac{1}{4}\right)^2, \quad (\text{F.4})$$

and

$$\Omega_s^+ = \Omega_s^4 = \frac{1}{4}. \quad (\text{F.5})$$

The primary intend of winning scores at gambling will impel agents with opinions in $R1$ to win. Denote the maximal steady opinion held by winners as s . The steady value of average opinion in G_- then can be determined accordingly by

$$\begin{aligned} \Omega_s^- &= \Omega_s^1 + \Omega_s^2 + \Omega_s^3 \\ &= n_1 \cdot \frac{1}{2}(s-1) - \frac{1}{6}\varepsilon^3 + \frac{3}{8}\varepsilon^2 - \frac{1}{8}\varepsilon - \frac{1}{4}\left(\varepsilon - \frac{1}{4}\right)^2. \end{aligned} \quad (\text{F.6})$$

The definition of winning at our gambling requires

$$\frac{|s|}{N_-|\Omega_s^-|} \cdot R_- \geq \frac{R}{N}, \quad (\text{F.7})$$

which yields

$$\Omega_s^- = \frac{1}{4\varepsilon^2 - 13} \left(\frac{8}{3}\varepsilon^3 + 2\varepsilon^2 + \frac{13}{4} \right). \quad (\text{F.8})$$

The order parameter is now directly stated as

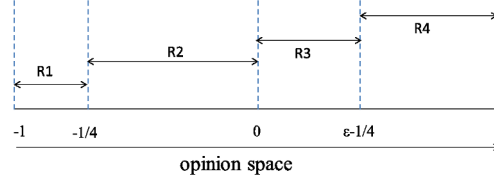
$$\Omega_s = \frac{1}{13 - 4\varepsilon^2} \left(\frac{8}{3}\varepsilon^3 + 3\varepsilon^2 \right). \quad (\text{F.9})$$

When we limit the confidence threshold to the range $(1/4, 1/2]$, the opinion space then can be partitioned as Fig. F.2. The fraction of agents who will migrate in opinion from $R2$ and $R3$ to $R1$ is the same as that who will migrate from $R2$ to $R1$ in the previous situation, that is

$$n_{2,3 \rightarrow 1} = \frac{1}{2}\varepsilon^2. \quad (\text{F.10})$$

This implies the fraction of winners at steady state is given by

$$n_w = \frac{3}{8} + \frac{1}{2}\varepsilon^2. \quad (\text{F.11})$$

FIGURE F.2: The generation of four adjacent regions in the opinion space when $\varepsilon \in (1/4, 1/2]$.

The stable average opinions in $R2$ and $R3$ are given by

$$\begin{aligned}\Omega_s^2 &= \int_{-1/4}^0 \frac{1}{2} \left[1 - \frac{k_a}{2} \left(\varepsilon - x - \frac{1}{4} \right) \right] \cdot x dx \\ &= \frac{1}{32} \left(\varepsilon - \frac{7}{12} \right),\end{aligned}\tag{F.12}$$

and

$$\begin{aligned}\Omega_s^3 &= \int_0^{\varepsilon-1/4} \frac{1}{2} \left[1 - \frac{k_a}{2} \left(\varepsilon - x - \frac{1}{4} \right) \right] \cdot x dx \\ &= -\frac{1}{6} \left(\varepsilon - \frac{1}{4} \right)^2 \left(\varepsilon - \frac{7}{4} \right).\end{aligned}\tag{F.13}$$

respectively. In the region $R4$, opinions are fixed at fully disordered initial positions throughout the evolution of dynamics. This offers

$$\Omega_s^4 = \frac{1}{4} \left[1 - \left(\varepsilon - \frac{1}{4} \right)^2 \right].\tag{F.14}$$

Now we get the final average opinion in clique G_+ :

$$\Omega_s^+ = \Omega_s^3 + \Omega_s^4 = -\frac{1}{6} \left(\varepsilon - \frac{1}{4} \right)^3 + \frac{1}{4}.\tag{F.15}$$

Similarly, denote the maximal opinion among the group of winners as s . The final average opinion in clique G_- then can be represented by

$$\Omega_s^- = n_w \cdot \frac{1}{2} (s - 1) + \Omega_s^2.\tag{F.16}$$

By recalling Eq. (F.7), we obtain

$$\Omega_s^- = 1 - \frac{1}{8\varepsilon^2 - 26} \left(\varepsilon - \frac{391}{12} \right),\tag{F.17}$$

which, together with previously calculated Ω_s^+ , specifies

$$\Omega_s = \frac{1}{8\varepsilon^2 - 26} \left(\varepsilon - \frac{391}{12} \right) + \frac{1}{6} \left(\varepsilon - \frac{1}{4} \right)^3 - \frac{5}{4}.\tag{F.18}$$

If the confidence threshold is limited to the interval $(1/2, 3/4]$, we can divide the opinion

space into five adjacent regions (Fig. F.3). As stated before, all agents holding initial opinions in $R2$ will migrate to $R1$ in opinion by communicating with their friends. And the fraction of agents who will be affected in opinion by their friends from $R3$ and $R4$ to $R1$ is approximated as

$$\begin{aligned} n_{3,4 \rightarrow 1} &= \int_{\varepsilon - \frac{3}{4}}^{\varepsilon - \frac{1}{4}} \frac{k_a}{2} \cdot \frac{1}{2} \left(\varepsilon - x - \frac{1}{4} \right) dx \\ &= \frac{1}{8}. \end{aligned} \quad (\text{F.19})$$

The fraction of agents holding steady opinions in $R1$ is then determined as

$$n_1 = \frac{1}{2}\varepsilon + \frac{1}{4}. \quad (\text{F.20})$$

The steady average opinion in G_+ is the same as that in the previous case, i.e., Eq. (F.15), since the partition of opinion space in the positive interval keeps unchanged. For region $R3$, its average opinion at steady state is analogously computed as

$$\begin{aligned} \Omega_s^3 &= \int_{\varepsilon - \frac{3}{4}}^0 \frac{1}{2} \left[1 - \frac{k_a}{2} \left(\varepsilon - x - \frac{1}{4} \right) \right] \cdot x dx \\ &= \frac{1}{6} \left(\varepsilon - \frac{3}{4} \right)^3. \end{aligned} \quad (\text{F.21})$$

Agents taking opinions in $R1$ will finally win at gambling by further local communications. This suggests the average opinion in G_- will stabilize at

$$\Omega_s^- = n_1 \cdot \frac{1}{2}(s - 1) + \Omega_s^3, \quad (\text{F.22})$$

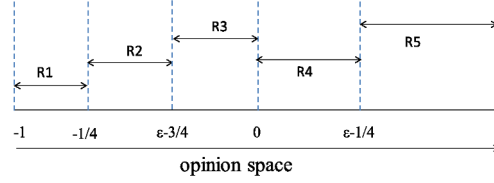
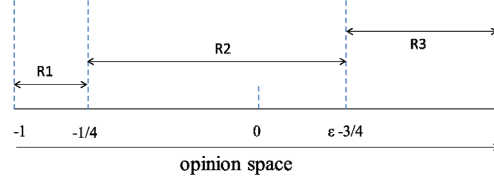
where s denotes the maximal opinion held by winners. Substituting the above expression into Eq. (F.7) yields

$$\Omega_s^- = \frac{1}{2\varepsilon - 7} \left[-\frac{4}{3} \left(\varepsilon - \frac{3}{4} \right)^3 + 2 \left(\varepsilon + \frac{1}{2} \right) \right], \quad (\text{F.23})$$

which, together with Ω_s^+ , provides

$$\Omega_s = \frac{1}{7 - 2\varepsilon} \left[-\frac{4}{3} \left(\varepsilon - \frac{3}{4} \right)^3 + 2 \left(\varepsilon + \frac{1}{2} \right) \right] + \frac{1}{6} \left(\varepsilon - \frac{3}{4} \right)^3 - \frac{1}{4}. \quad (\text{F.24})$$

For high confidence threshold: $\varepsilon \in (\frac{5}{4}, \frac{7}{4}]$, the opinion space can be divided into three regions (Fig. F.4). The fraction of agents who will be affected in opinion by their friends

FIGURE F.3: The generation of five adjacent regions in the opinion space when $\varepsilon \in (1/2, 3/4]$.FIGURE F.4: The generation of three adjacent regions in the opinion space when $\varepsilon \in (5/4, 7/4]$.

from $R3$ to $R1$ is approximated as

$$\begin{aligned} n_{3 \rightarrow 1} &= \int_{\varepsilon - \frac{3}{4}}^1 \frac{1}{2} \cdot \frac{k_a}{2} (\varepsilon - x - \frac{1}{4}) dx \\ &= -\frac{1}{2} (\varepsilon - \frac{5}{4})^2 + \frac{1}{8}. \end{aligned} \quad (\text{F.25})$$

Consequently, the fraction of agents in G_- will fix at

$$\begin{aligned} n_- &= \frac{1}{2} (\varepsilon - \frac{3}{4} + 1) + n_{3 \rightarrow 1} \\ &= 1 - \frac{1}{2} (\varepsilon - \frac{7}{4})^2. \end{aligned} \quad (\text{F.26})$$

The primary intend of winning scores at gambling will guide all agents in G_- to win. Analogously, the final average opinion in G_- can be obtained as below by recalling Eq. (F.7):

$$\Omega_s^- = 1 - \frac{4}{(\varepsilon - \frac{7}{4})^2 + 2}. \quad (\text{F.27})$$

For the clique G_+ , we have

$$\begin{aligned} \Omega_s^+ &= \int_{\varepsilon - \frac{3}{4}}^1 \frac{1}{2} [1 - \frac{k_a}{2} (\varepsilon - x - \frac{1}{4})] \cdot x dx \\ &= \frac{\varepsilon}{6} (\varepsilon - \frac{9}{8})^2 - \frac{55}{128} (\varepsilon - 1) + \frac{5}{24}. \end{aligned} \quad (\text{F.28})$$

The corresponding order parameter is now obtained as

$$\Omega_s = \frac{4}{(\varepsilon - \frac{7}{4})^2 + 2} - \frac{\varepsilon}{6} (\varepsilon - \frac{9}{8})^2 + \frac{55}{128} (\varepsilon - 1) - \frac{29}{24}. \quad (\text{F.29})$$

Consider sufficiently high confidence threshold: $\varepsilon > 7/4$. The partition of opinion space can be simplified as two regions (Fig. F.5). With local communication, almost all agents

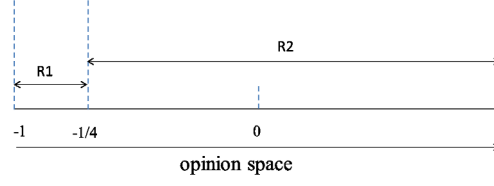


FIGURE F.5: The generation of two adjacent regions in the opinion space when $\varepsilon \in (7/4, 2]$.

having initial opinions in $R2$ will migrate to $R1$ in opinion. Analogously, the agents with opinions in $R1$ then will learn from their friends to win in the opinion evolution process. This requires

$$\frac{|s|}{N \cdot \frac{1}{2}(|s| + 1)} \cdot R_- \geq \frac{R}{N}. \quad (\text{F.30})$$

Solution of the above inequality ($|s| \geq 1$) implies the clustering of all agents in opinion -1 at steady state. So we get

$$\Omega_s = 1. \quad (\text{F.31})$$

Bibliography

- [1] R. M. May. Simple mathematical models with very complicated dynamics. *Nature*, 261(5560):459, 1976.
- [2] J. P. Huang. Experimental econophysics: Complexity, self-organization, and emergent properties. *Phys. Rep.*, 564:1, 2015.
- [3] S. M. Blower and H. Dowlatabadi. Sensitivity and uncertainty analysis of complex models of disease transmission: an hiv model, as an example. *Int. Stat. Rev.*, pages 229–243, 1994.
- [4] K. Gulyuz et al. High precision determination of the β decay q ec value of c 11 and implications on the tests of the standard model. *Phys. Rev. Lett.*, 116(1):012501, 2016.
- [5] R. E. Melchers. Importance sampling in structural systems. *Struct. Safety*, 6:3, 1989.
- [6] A. D. Kiureghian. Structural reliability methods for seismic safety assessment: a review. *Eng. Struct.*, 18:412, 1996.
- [7] A. M. Hasofer and N. C. Lind. Exact and invariant second order code format. *J. Eng. Mech.*, 100:111, 1974.
- [8] B. Fiessler, R. Rackwitz, and H. Neumann. Quadratic limit states in structural reliability. *J. Eng. Mech.*, 105:661, 1979.
- [9] D. Xiu and G. E. Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comput. Phys.*, 187:137, 2003.
- [10] H. S. Seo and B. M. Kwak. Efficient statistical tolerance analysis for general distributions using three-point information. *Int. J. Prod. Res.*, 40:931, 2002.
- [11] H. Xu and S. Rahman. A generalized dimension-reduction method for multi-dimensional integration in stochastic mechanics. *Int. J. Numer. Methods Eng.*, 61:1992, 2004.

- [12] S. Rahman and H. Xu. A univariate dimension-reduction method for multi-dimensional integration in stochastic mechanics. *Probab. Eng. Mech.*, 19:393, 2004.
- [13] S. H. Lee and B. M. Kwak. Response surface augmented moment method for efficient reliability analysis. *Struct. Safety*, 28:261, 2006.
- [14] S. M. Blower, H. B. Gershengorn, and R. M. Grant. A tale of two futures: Hiv and antiretroviral therapy in san francisco. *Science*, 287(5453):650, 2000.
- [15] J. C. Helton et al. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliab. Eng. Syst. Safe.*, 91(10):1175, 2006.
- [16] B. G. Marcot, P. H. Singleton, and N. H. Schumaker. Analysis of sensitivity and uncertainty in an individual-based model of a threatened wildlife species. *Nat. Resour. Model.*, 28(1):37, 2015.
- [17] M. Padulo, M. S. Campobasso, and M. D. Guenov. Comparative analysis of uncertainty propagation methods for robust engineering design. *Guidelines for a Decision Support Method Adapted to NPD Processes*, 2007.
- [18] S. H. Lee and W. Chen. A comparative study of uncertainty propagation methods for black-box-type problems. *Struct. Multidisc. Optim.*, 37(3):239, 2009.
- [19] E. Borgonovo and E. Plischke. Sensitivity analysis: A review of recent advances. *Eur. J. Oper. Res.*, 248:869, 2016.
- [20] F. Yates. *Sampling Methods for Censuses and Surveys*. Charles Griffin & Co., Ltd., London., 2 edition, 1953.
- [21] H. H. Ku. Notes on the use of propagation of error formulas. *J. Res. Natl. Bur. Stand. C*, 70C(4):262, 1966.
- [22] J. C. Helton. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliab. Eng. Syst. Safe.*, 42(2):327, 1993.
- [23] G. H. White and I. Farrance. Uncertainty of measurement in quantitative medical testing a laboratory implementation guide. *Clin. Biochem. Rev.*, 25(4):S1, 2004.
- [24] L. Guo, Z. Luo, and Y. Zhu. Heterogeneity and subjectivity in binary-state opinion formation systems. *J. Stat. Mech.*, 2013(11):P11013, 2013.
- [25] Y. Zhu, W. Li, and X. Cai. Opinion evolution on a ba scaling network. *Physica A*, 392(24):6596–6602, 2013.
- [26] D. G. Cacuci, M. Ionescu-Bujor, and I. M. Navon. *Sensitivity and uncertainty analysis: Applications to Large-Scale Systems*, volume 2. Taylor & Francis, 2005.

- [27] B. Iooss and P. Lemaître. A review on global sensitivity analysis methods. In Dellino G. and Meloni C., editors, *Uncertainty Management in Simulation-Optimization of Complex Systems*, volume 59 of *Operations Research/Computer Science Interfaces Series*, pages 101–122. Springer, Boston, 2015.
- [28] Y. Zhu et al. An analytic method for sensitivity analysis of complex systems. *Physica A*, 469:52–59, 2017.
- [29] I. M. Sobol. Sensitivity estimates for nonlinear mathematical models. *MMCE*, 1(4):407–414, 1993.
- [30] J. C. Helton et al. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliab. Eng. Syst. Safe.*, 91(10):1175–1209, 2006.
- [31] S. Kucherenko, D. Albrecht, and A. Saltelli. Comparison of latin hypercube and quasi monte carlo sampling techniques. arXiv:1505.02350, 2015.
- [32] A. Saltelli, M. Ratto, and S. Tarantola. Model-free importance indicators for dependent input. In *Int. Symp. Sensitivity Analysis of Model Output*, pages 21–25, 2001.
- [33] C. Xu and G. Z. Gertner. Uncertainty and sensitivity analysis for models with correlated parameters. *Reliab. Eng. Syst. Safe.*, 93(10):1563–1573, 2008.
- [34] T. Most. Variance-based sensitivity analysis in the presence of correlated input variables. In *Proc. 5th Int. Conf. Reliable Engineering Computing (REC)*, Brno, 2012.
- [35] W. Hao, Z. Lu, and L. Li. A new interpretation and validation of variance based importance measures for models with correlated inputs. *Comput. Phys. Comm.*, 184(5):1401–1413, 2013.
- [36] L. Li et al. General validation and decomposition of the variance-based measures for models with correlated inputs. *Aerosp. Sci. Technol.*, 62:75–86, 2017.
- [37] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. In *Proc. R. Soc. Lond. A*, volume 115, pages 700–721. The Royal Society, 1927.
- [38] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate analysis (probability and mathematical statistics)*. Academic Press, London, 1 edition, 1980.
- [39] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81(2):591, 2009.

- [40] M. Kitsak et al. Identification of influential spreaders in complex networks. *Nature Phys.*, 6(11):888–893, 2010.
- [41] S. Pei and H. A. Makse. Spreading dynamics in complex networks. *J. Stat. Mech.*, 2013(12):P12002, 2013.
- [42] Z. Hui et al. Structure properties of evolutionary spatially embedded networks. *Physica A*, 392(8):1909–1919, 2013.
- [43] D. Sterratt et al. *Principles of computational modelling in neuroscience*. Cambridge University Press, Cambridge, 2011.
- [44] P. Vaughn et al. Representation of two-phase flow in the vicinity of the repository in the 1996 performance assessment for the waste isolation pilot plant. *Reliab. Eng. Syst. Safe.*, 69:205–226, 2000.
- [45] F. Campolongo et al. Hitchhiker’s guide to sensitivity analysis. In *Sensitivity analysis*. IEEE Computer Society Press, 2000.
- [46] J. C. Helton. Uncertainty and sensitivity analysis for models of complex systems. In *Computational Methods in Transport: Verification and Validation*, pages 207–228. Springer, Berlin, 2008.
- [47] A. Saltelli et al. *Global sensitivity analysis: the primer*. Wiley, 2008.
- [48] R. Rosen. *Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life*. Columbia University Press, Columbia, 1991.
- [49] A. Saltelli et al. *Sensitivity analysis in practice: a guide to assessing scientific models*. Wiley, 2004.
- [50] European commission impact assessment guidelines. http://ec.europa.eu/smart-regulation/impact/commission_guidelines/docs/iag_2009_2009. Accessed: 2009-01-15.
- [51] European commission commission staff working document. http://ec.europa.eu/smart-regulation/guidelines/docs/swd_br_guidelines_en.pdf.pdf. 2015. Accessed: 2015-05-19.
- [52] I. Farrance and R. Frenkel. Uncertainty of measurement: a review of the rules for calculating uncertainty components through functional relationships. *Clin. Biochem. Rev.*, 33(2):49, 2012.
- [53] Intergovernmental Panel on Climate Change. *Climate Change 2014–Impacts, Adaptation and Vulnerability: Regional Aspects*. Cambridge University Press, 2014.

- [54] N. Gaber et al. Guidance on the development, evaluation, and application of environmental models. *Tech. Rep. EPA/100/K-09/003, US Environmental Protection Agency*, 2009.
- [55] P. Gustafson, C. Srinivasan, and L. Wasserman. Local sensitivity analysis. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian statistics 5*, Proceedings of the Fifth Valencia International Meeting on Bayesian Statistics, pages 197–210. Oxford University Press, Oxford, 1996.
- [56] H. M. Wagner. Global sensitivity analysis. *Oper. Res.*, 43(6):948–969, 1995.
- [57] S. K. Maity et al. Emergence of fast agreement in an overhearing population: The case of the naming game. *EPL*, 101(6):68004, 2013.
- [58] K. Beuls and L. Steels. Agent-based models of strategies for the emergence and evolution of grammatical agreement. *PloS one*, 8(3):e58960, 2013.
- [59] D. Centola and A. Baronchelli. The spontaneous emergence of conventions: An experimental study of cultural evolution. *PNAS*, 112(7):1989–1994, 2015.
- [60] T. G. Cummings and C. G. Worley. *Organization development and change*. South-Western College Pub., 10 edition, 2014.
- [61] K. Lewin. *Field Theory in Social Science: Selected Theoretical Papers*. Harper, 1 edition, 1964.
- [62] R. E. Nelson and K. M. Mathews. Cause maps and social network analysis in organizational diagnosis. *J. Appl. Behav. Sci.*, 27(3):379–397, 1991.
- [63] F. E. Harrell, K. L. Lee, and D. B. Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.*, 15(4):361–387, 1996.
- [64] D. G. Altman and P. Royston. What do we mean by validating a prognostic model? *Stat. Med.*, 19(4):453–473, 2000.
- [65] M. Egger et al. Prognosis of hiv-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. *Lancet*, 360(9327):119–129, 2002.
- [66] K. G. M. Moons, D. G. Altman, Y. Vergouwe, and P. Royston. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*, 338:b606, 2009.

- [67] G. M. Williams, R. Kroes, and I. C. Munro. Safety evaluation and risk assessment of the herbicide roundup and its active ingredient, glyphosate, for humans. *Regul. Toxicol. Pharmacol.*, 31(2):117–165, 2000.
- [68] A. Klinké and O. Renn. A new approach to risk evaluation and management: risk-based, precaution-based, and discourse-based strategies. *Risk Anal.*, 22(6): 1071–1094, 2002.
- [69] V. P. Kochergin. Three-dimensional prognostic models. In N. S. Heaps, editor, *Three-dimensional coastal ocean models*, volume 4 of *Coastal and Estuarine Sciences*, pages 201–208. American Geophysical Union, Washington, 1987.
- [70] S. Riley et al. Transmission dynamics of the etiological agent of sars in hong kong: impact of public health interventions. *Science*, 300(5627):1961–1966, 2003.
- [71] M. Lipsitch et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science*, 300(5627):1966–1970, 2003.
- [72] S. Alam, R. Meyer, and E. Norling. A model for hiv spread in a south african village. In David N. and Sichman J. S., editors, *Multi-Agent-Based Simulation IX. MABS 2008*, volume 5269 of *Lecture Notes in Computer Science*, pages 33–45. Springer, Berlin, 2009.
- [73] E. C. Beck et al. A data-driven simulation of hiv spread among young men who have sex with men: the role of age and race mixing, and stis. *J. Acquir. Immune. Defic. Syndr.*, 70(2):186–194, 2015.
- [74] B. L. Smith, B. M. Williams, and R. K. Oswald. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transp. Res. Part C Emerg. Technol.*, 10(4):303–321, 2002.
- [75] G. Orosz, R. E. Wilson, and G. Stépán. Traffic jams: dynamics and control. *Phil. Trans. R. Soc. A*, 368:4455–4479, 2010.
- [76] M. R. Hassan and B. Nath. Stock market forecasting using hidden markov model: a new approach. In *Proc. 5th Int. Conf. Intelligent Systems Design and Applications (ISDA '05)*, pages 192–196. IEEE, 2005.
- [77] S. Dablemont, S. Van Bellegem, and M. Verleysen. Modelling and forecasting financial time series of” tick data” by functional analysis and neural networks. In *Proc. Forecasting Financial Markets 2007*, pages 1–18, Aix-en-Provence, May 2007. Conference Proceedings Cd-Rom.

- [78] S. M. Hassanizadeh. Derivation of basic equations of mass transport in porous media, part 2. generalized darcy's and fick's laws. *Adv. Water Resour.*, 9(4): 207–222, 1986.
- [79] S. Whitaker. The equations of motion in porous media. *Chem. Eng. Sci.*, 21(3): 291–300, 1966.
- [80] D. Coakley, P. Raftery, and M. Keane. A review of methods to match building energy simulation models to measured data. *Renew. Sustainable Energy Rev.*, 37: 123–141, 2014.
- [81] K. Happe et al. Agent-based modelling and sensitivity analysis by experimental design and metamodeling: an application to modelling regional structural change. In *XIth Int. Congr. European Association of Agricultural Economists*. European Association of Agricultural Economists, 2005.
- [82] L. Perez and S. Dragicevic. An agent-based approach for modeling dynamics of contagious disease spread. *Int. J. Health Geogr.*, 8(1):50, 2009.
- [83] M. Fonoberova, V. A. Fonoberov, and I. Mezić. Global sensitivity/uncertainty analysis for agent-based models. *Reliab. Eng. Syst. Safety*, 118:8–17, 2013.
- [84] I. G. Usman et al. Sensitivity analysis of the parameters of a model of hiv-1 transmission. (*GARJPAS*, 5:011–7, 2016.
- [85] R. McGill, J. W. Tukey, and W. A. Larsen. Variations of box plots. *Am. Stat.*, 32 (1):12–16, 1978.
- [86] J. R. M. Hosking and J. R. Wallis. Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29(3):339–349, 1987.
- [87] D. F. Williamson, R. A. Parker, and J. S. Kendrick. The box plot: a simple visual method to interpret data. *Ann. Intern. Med.*, 110(11):916–921, 1989.
- [88] M. Osterwald-Lenum. A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. *Oxford Bulletin of Economics and statistics*, 54(3):461–472, 1992.
- [89] J. C. Helton. Probability, conditional probability and complementary cumulative distribution functions in performance assessment for radioactive waste disposal. *Reliab. Eng. Syst. Safe.*, 54(2-3):145–163, 1996.
- [90] M. Frigge, D. C. Hoaglin, and B. Iglewicz. Some implementations of the boxplot. *Am. Stat.*, 43(1):50–54, 1989.

- [91] M. Hubert and E. Vandervieren. An adjusted boxplot for skewed distributions. *Comput. Stat. Data Anal.*, 52(12):5186–5201, 2008.
- [92] M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66(1):016128, 2002.
- [93] C. Castellano and R. Pastor-Satorras. Thresholds for epidemic spreading in networks. *Phys. Rev. Lett.*, 105(21):218701, 2010.
- [94] A. Saltelli, S. Tarantola, and KP-S. Chan. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1):39–56, 1999.
- [95] H. Christopher Frey and S. R. Patil. Identification and review of sensitivity analysis methods. *Risk Anal.*, 22(3):553–578, 2002.
- [96] K. K. Benke, K. E. Lowell, and A. J. Hamilton. Parameter uncertainty, sensitivity analysis and prediction error in a water-balance hydrological model. *Math. Comput. Model.*, 47(11):1134–1149, 2008.
- [97] A. Saltelli et al., editors. *Sensitivity analysis*, volume 1. Wiley, New York, 2000.
- [98] R. Bailis, M. Ezzati, and D. M. Kammen. Mortality and greenhouse gas impacts of biomass and petroleum energy futures in africa. *Science*, 308(5718):98–103, 2005.
- [99] J. E. Campbell et al. Photosynthetic control of atmospheric carbonyl sulfide during the growing season. *Science*, 322(5904):1085–1088, 2008.
- [100] A. D. Lopez et al. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet*, 367(9524):1747–1757, 2006.
- [101] J. M. Murphy et al. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, 430(7001):768–772, 2004.
- [102] M. D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.
- [103] F. Campolongo, J. Cariboni, and A. Saltelli. Sensitivity analysis: the morris method versus the variance based measures. *Technometrics*, 2003.
- [104] F. Campolongo, J. Cariboni, A. Saltelli, and W. Schoutens. Enhancing the morris method. In *Sensitivity Analysis of Model Output. Proceedings of the 4th International Conference on Sensitivity Analysis of Model Output (SAMO 2004)*, pages 369–379, 2005.

- [105] R. B. Darlington and A. F. Hayes. *Regression analysis and linear models: Concepts, applications, and implementation*. Methodology in the Social Sciences. The Guilford Publications, 1 edition, 2016.
- [106] A. M. Legendre. *Nouvelles méthodes pour la détermination des orbites des comètes*. F. Didot, 1805.
- [107] C. F. Gauss. *Theory of the motion of the heavenly bodies moving about the sun in conic sections: a translation of Carl Frdr. Gauss" Theoria motus": With an appendix. By Ch. H. Davis*. Little, Brown and Company, 1857.
- [108] F. Galton. Typical laws of heredity. *Nature*, 15:492–5, 512–4, 532–3, 1877.
- [109] G. U. Yule. On the theory of correlation. *J. R. Stat. Soc.*, 60(4):812–854, 1897.
- [110] R. Peck and J. L. Devore. *Statistics: The exploration & analysis of data*. Available Titles Aplia. Brooks/Cole, 7 edition, 2011.
- [111] D. A. Freedman. *Statistical models: theory and practice*. Cambridge University Press, 2 edition, 2009.
- [112] N. R. Draper and H. Smith. *Applied regression analysis*. Probability and Statistics. Wiley, 3 edition, 2014.
- [113] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to linear regression analysis*, volume 821 of *Probability and Statistics*. Wiley, 5 edition, 2012.
- [114] J. Neter et al. *Applied linear statistical models*, volume 1 of *Irwin series in statistics*. McGraw-Hill/Irwin, 4 edition, 1996.
- [115] D. M. Bates and D. G. Watts. *Nonlinear regression analysis and its applications*, volume 2. Wiley, 1988.
- [116] H. Motulsky and A. Christopoulos. *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. GraphPad Software Inc., San Diego, 2003.
- [117] R. L. Prentice and L. A. Gloeckler. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, pages 57–67, 1978.
- [118] B. C. Taylor et al. Alteration in lymphocyte subpopulations in bovine leukosis virus-infected cattle. *Vet. Immunol. Immunop.*, 31(1-2):35–47, 1992.
- [119] W. W. Daniel and C. L. Cross. *Biostatistics: a foundation for analysis in the health sciences*. Wiley New York, 1995.

- [120] K. Happe, K. Kellermann, and A. Balmann. Agent-based analysis of agricultural policies: an illustration of the agricultural policy simulator agripolis, its adaptation and behavior. *Ecol. Soc.*, 11(1), 2006.
- [121] H. Torjusen et al. Food system orientation and quality perception among consumers and producers of organic food in hedmark county, norway. *Food Qual. Prefer.*, 12(3):207–216, 2001.
- [122] J. C. Helton et al. Uncertainty and sensitivity analysis of food pathway results with the maccs reactor accident consequence model. *Reliab. Eng. Syst. Safe.*, 49(2):109–144, 1995.
- [123] G. E. P. Box and K. B. Wilson. On the experimental attainment of optimum conditions. In Kotz S. and Johnson N. L., editors, *Breakthroughs in Statistics*, Springer Series in Statistics, pages 270–310. Springer, New York, 1992.
- [124] A. I. Khuri and S. Mukhopadhyay. Response surface methodology. *WIREs Comput. Stat.*, 2(2):128–149, 2010.
- [125] G. E. P. Box, J. S. Hunter, and W. G. Hunter. *Statistics for experimenters: design, innovation, and discovery*. Wiley Series in Probability and Statistics. Wiley-Interscience, New York, 2 edition, 2005.
- [126] G. E. P. Box, W. G. Hunter, and J. S. Hunter. *Statistics for experimenters: an introduction to design, data analysis, and model building*, volume 154 of *Wiley series in probability and mathematical statistics*. Wiley, 1978.
- [127] D. C. Montgomery. *Design and analysis of experiments*. Wiley, 8 edition, 2012.
- [128] R. L. Plackett and J. P. Burman. The design of optimum multifactorial experiments. *Biometrika*, 33(4):305–325, 1946.
- [129] S. K. Ahuja, G. M. Ferreira, and A. R. Moreira. Application of plackett-burman design and response surface methodology to achieve exponential growth for aggregated shipworm bacterium. *Biotechnol. Bioeng.*, 85(6):666–675, 2004.
- [130] A. I. Khuri and J. A. Cornell. *Response surfaces: designs and analyses*. Statistics: A Series of Textbooks and Monographs. CRC press, 2 edition, 1996.
- [131] G. E. P. Box. Multi-factor designs of first order. *Biometrika*, 39(1-2):49–57, 1952.
- [132] M. R. Rajashekhar and B. R. Ellingwood. A new look at the response surface approach for reliability analysis. *Struct. safety*, 12(3):205–220, 1993.
- [133] G. E. P. Box and D. W. Behnken. Some new three level designs for the study of quantitative variables. *Technometrics*, 2(4):455–475, 1960.

- [134] R. A. McLean and V. L. Anderson. *Applied factorial and fractional designs*. Statistics: a Series of Textbooks and Monographs. Marcel Dekker Inc., New York, 1984.
- [135] A. I. Khuri. A measure of rotatability for response-surface designs. *Technometrics*, 30(1):95–104, 1988.
- [136] R. H. Myers. *Response surface methodology*. Allyn and Bacon, Boston, 1971.
- [137] G. E. P. Box and J. S. Hunter. Multi-factor experimental designs for exploring response surfaces. *Ann. Math. Stat.*, pages 195–241, 1957.
- [138] G. E. P. Box and N. R. Draper. *Response surfaces, mixtures, and ridge analyses*. Wiley, 2 edition, 2007.
- [139] R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook. *Response surface methodology: process and product optimization using designed experiments*. Wiley, 3 edition, 2016.
- [140] D. H. Doehlert. Uniform shell designs. *J. R. Stat. Soc. C*, pages 231–239, 1970.
- [141] M. J. Box and N. R. Draper. Factorial designs, the $-\bar{x}'x-$ criterion, and some related matters. *Technometrics*, 13(4):731–742, 1971.
- [142] A. T. Hoke. Economical second-order designs based on irregular fractions of the 3^n factorial. *Technometrics*, 16(3):375–384, 1974.
- [143] K. G. Roquemore. Hybrid designs for quadratic response surfaces. *Technometrics*, 18(4):419–423, 1976.
- [144] T. J. Vidmar and J. W. Mckean. A monte carlo study of robust and least squares response surface methods. *J. Stat. Comput. Simul.*, 54(1-3):1–18, 1996.
- [145] W. P. Gardiner and G. Gettinby. *Experimental design techniques in statistical practice: A practical software-based approach*. Ellis Horwood series in mathematics and its applications. Horwood Publishing, 1998.
- [146] J. C. Helton, R. L. Iman, and J. B. Brown. Sensitivity analysis of the asymptotic behavior of a model for the environmental movement of radionuclides. *Ecol. Model.*, 28(4):243–278, 1985.
- [147] M. Koda, A. H. Dogru, and J. H. Seinfeld. Sensitivity analysis of partial differential equations with application to reaction and diffusion processes. *J. Comput. Phys.*, 30(2):259–282, 1979.

- [148] D. J. Downing, R. H. Gardner, and F. O. Hoffman. An examination of response-surface methodologies for uncertainty analysis in assessment models. *Technometrics*, 27(2):151–163, 1985.
- [149] D. M. Hamby. A review of techniques for parameter sensitivity analysis of environmental models. *Environ. Model. Assess.*, 32(2):135–154, 1994.
- [150] I. M. Sobol and S. Kucherenko. Derivative based global sensitivity measures and their link with global sensitivity indices. *Math. Comput. Simulat.*, 79(10):3009–3017, 2009.
- [151] I. M. Sobol and S. Kucherenko. Derivative based global sensitivity measures. *Procedia Soc. Behav. Sci.*, 2(6):7745–7746, 2010.
- [152] M. Lamboni, B. Iooss, A-L. Popelin, and F. Gamboa. Derivative-based global sensitivity measures: general links with sobolindices and numerical tests. *Math. Comput. Simulat.*, 87:45–54, 2013.
- [153] R. I. Cukier, H. B. Levine, and K. E. Shuler. Nonlinear sensitivity analysis of multiparameter model systems. *J. Comput. Phys.*, 26(1):1–42, 1978.
- [154] M. J. W. Jansen, W. A. H. Rossing, and R. A. Daamen. Monte carlo estimation of uncertainty contributions from several independent multivariate sources. In J. Grasman and G. Van Straten, editors, *Predictability and nonlinear modelling in natural sciences and economics*, pages 334–343. Springer, Dordrecht, 1994.
- [155] T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Safe.*, 52(1):1–17, 1996.
- [156] A. Saltelli et al. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Comput. Phys. Commun.*, 181(2):259–270, 2010.
- [157] M. Gerken et al. Models for retrospective quantification of indoor radon exposure in case-control studies. *Health Phys.*, 78(3):268–278, 2000.
- [158] A. Saltelli. Sensitivity analysis for importance assessment. *Risk Anal.*, 22(3):579–590, 2002.
- [159] X. Song et al. Variance-based sensitivity analysis of a forest growth model. *Ecol. Model.*, 247:135–143, 2012.
- [160] B. Efron and C. Stein. The jackknife estimate of variance. *Ann. Stat.*, pages 586–596, 1981.

- [161] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Stat. Sci.*, pages 409–423, 1989.
- [162] G. E. B. Archer, A. Saltelli, and I. M. Sobol. Sensitivity measures, anova-like techniques and the use of bootstrap. *J. Stat. Comput. Simul.*, 58(2):99–120, 1997.
- [163] H. Rabitz, O. F. Aliş, J. Shorter, and K. Shim. Efficient inputoutput model representations. *Comput. Phys. Commun.*, 117(1-2):11–20, 1999.
- [164] J. Jacques, C. Lavergne, and N. Devictor. Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliab. Eng. Syst. Safe.*, 91(10):1126–1134, 2006.
- [165] S. Da Veiga, F. Wahl, and F. Gamboa. Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics*, 51(4):452–463, 2009.
- [166] K. Zhang et al. A new framework of variance based global sensitivity analysis for models with correlated inputs. *Struct. Safety*, 55:1–9, 2015.
- [167] M. Chun, S. Han, and N. Tak. An uncertainty importance measure using a distance metric for the change in a cumulative distribution function. *Reliab. Eng. Syst. Safe.*, 70(3):313–321, 2000.
- [168] E. Borgonovo and L. Peccati. Global sensitivity analysis in inventory management. *Int. J. Prod. Econ.*, 108(1):302–313, 2007.
- [169] E. Borgonovo. A new uncertainty importance measure. *Reliab. Eng. Syst. Safe.*, 92(6):771–784, 2007.
- [170] F. Pianosi and T. Wagener. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environ. Model. Softw.*, 67: 1–11, 2015.
- [171] A. Kolmogorov. Sulla determinazione empirica di una leggi di distribuzione. *G. Dell’istituto Ital. Degli Attuari*, 4:83–91, 1933.
- [172] F. K. Zadeh et al. Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the swat model. *Environ. Model. Softw.*, 91:210–222, 2017.
- [173] D. S. Starnes, D. Yates, and D. S. Moore. *The practice of statistics*. W. H. Freeman, 4 edition, 2010.
- [174] M. D. McKay, R. J. Beckman, and W. J. Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

- [175] I. A. Macdonald. Comparison of sampling techniques on the performance of monte-carlo based sensitivity analysis. In *11th Int. IBPSA Conf., Glasgow, Scotland*, pages 992–999, 2009.
- [176] T. H. Andres. Statistical sampling strategies. In *Uncertainty analysis for performance assessments of radioactive waste disposal systems*. 1987.
- [177] T. H. Andres and W. C. Hajas. Using iterated fractional factorial design to screen parameters in sensitivity analysis of a probabilistic risk assessment model. 1993.
- [178] A. Saltelli, T. H. Andres, and T. Homma. Sensitivity analysis of model output. performance of the iterated fractional factorial design method. *Comput. Stat. Data Anal.*, 20(4):387–407, 1995.
- [179] J. Morio, R. Pastel, and F. Le Gland. An overview of importance splitting for rare event simulation. *Eur. J. Phys.*, 31(5):1295, 2010.
- [180] J. Morio. Importance sampling: how to approach the optimal density? *Eur. J. Phys.*, 31(2):L41, 2010.
- [181] J. M. Hammersley and D. C. Handscomb. General principles of the monte carlo method. In *Monte Carlo Methods*, pages 50–75. Springer, 1964.
- [182] C. Anderson. Monte carlo methods and importance sampling. 1999.
- [183] R. L. Iman, J. E. Campbell, and J. C. Helton. An approach to sensitivity analysis of computer models. i- introduction, input, variable selection and preliminary variable assessment. *J. Qual. Technol.*, 13:174–183, 1981.
- [184] M. E. Johnson, L. M. Moore, and D. Ylvisaker. Minimax and maximin distance designs. *J. Stat. Plan. Inference*, 26(2):131–148, 1990.
- [185] M. D. Morris and T. J. Mitchell. Exploratory designs for computational experiments. *J. Stat. Plan. Inference*, 43(3):381–402, 1995.
- [186] C. E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.
- [187] A. B. Owen. Orthogonal arrays for computer experiments, integration and visualization. *Stat. Sin.*, pages 439–452, 1992.
- [188] K. Fang et al. Uniform design: theory and application. *Technometrics*, 42(3):237–248, 2000.
- [189] J. Park. Optimal latin-hypercube designs for computer experiments. *J. Stat. Plan. Inference*, 39(1):95–111, 1994.

- [190] Q. Y. Kenny, W. Li, and A. Sudjianto. Algorithmic construction of optimal symmetric latin hypercube designs. *J. Stat. Plan. Inference*, 90(1):145–159, 2000.
- [191] K. Fang, C. Ma, and P. Winker. Centered l2-discrepancy of random sampling and latin hypercube design, and construction of uniform designs. *Math. Comput.*, 71(237):275–296, 2002.
- [192] R. Jin, W. Chen, and A. Sudjianto. An efficient algorithm for constructing optimal design of computer experiments. *J. Stat. Plan. Inference*, 134(1):268–287, 2005.
- [193] V. R. Joseph and Y. Hung. Orthogonal-maximin latin hypercube designs. *Stat. Sin.*, pages 171–186, 2008.
- [194] F. A. C. Viana, G. Venter, and V. Balabanov. An algorithm for fast optimal latin hypercube design of experiments. *Int. J. Numer. Mech. Eng.*, 82(2):135–156, 2010.
- [195] N. Pholdee and S. Bureerat. An efficient optimum latin hypercube sampling technique based on sequencing optimisation using simulated annealing. *Int. J. Syst. Sci.*, 46(10):1780–1789, 2015.
- [196] J. H. Halton. Algorithm 247: Radical-inverse quasi-random point sequence. *Communications of the ACM*, 7(12):701–702, 1964.
- [197] S. Joe and F. Y. Kuo. Remark on algorithm 659: Implementing sobol’s quasirandom sequence generator. *ACM Trans. Math. Softw.*, 29(1):49–57, 2003.
- [198] I. M. Sobol. On quasi-monte carlo integrations. *Math. Comput. Simul.*, 47(2):103–112, 1998.
- [199] I. M. Sobol’. On the distribution of points in a cube and the approximate evaluation of integrals. *U.S.S.R Comput. Maths. Math. Phys.*, 7:86–112, 1967.
- [200] I. M. Sobol. Uniformly distributed sequences with an additional uniform property. *U.S.S.R Comput. Maths. Math. Phys.*, 16(5):236–242, 1976.
- [201] P. LEcuyer. Quasi-monte carlo methods with applications in finance. *Financ. Stoch.*, 13(3):307–349, 2009.
- [202] P. Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer, New York, 1 edition, 2003.
- [203] P. Bratley and B. L. Fox. Algorithm 659: Implementing sobol’s quasirandom sequence generator. *ACM Trans. Math. Softw.*, 14(1):88–100, 1988.
- [204] F. Gamboa et al. Statistical inference for sobol pick-freeze monte carlo method. *Statistics*, 50(4):881–902, 2016.

- [205] W. Yun et al. An efficient sampling method for variance-based sensitivity analysis. *Struct. Safety*, 65:74–83, 2017.
- [206] F. Galton. Opening address to the anthropological section of the british association by the president of the section. *Nature*, 32:507–510, 1885.
- [207] F. Galton. Regression towards mediocrity in hereditary stature. *J. Roy. Anthropol. Inst.*, 15:246–263, 1886.
- [208] K. Pearson. Note on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.*, 58:240–242, 1895.
- [209] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30:81–93, 1938.
- [210] H. Abdi. Multiple correlation coefficient. *Encyclopedia of measurement and statistics*, pages 648–651, 2007.
- [211] M. J. W. Jansen. Analysis of variance designs for model output. *Comput. Phys. Comm.*, 117(1-2):35–43, 1999.
- [212] B. Taylor. "Direct and Reverse Methods of Incrementation" Translated into English in "A Source Book in Mathematics 1200C1800". Harvard University Press, Cambridge, 1969.
- [213] J. V. Grabiner. Was newton's calculus a dead end? the continental influence of maclaurin's treatise of fluxions. *Am. Math. monthly*, 104(5):393–410, 1997.
- [214] L. Hörmander. *The analysis of linear partial differential operators I: Distribution theory and Fourier analysis*. Springer, Berlin, 2 edition, 2003.
- [215] R. Billinton and W. Li. *Reliability Assessment of Electric Power Systems Using Monte Carlo Methods*. Springer, New York, 1 edition, 1994.
- [216] V. Fthenakis and H. C. Kim. Land use and electricity generation: A life-cycle analysis. *Renew. Sustainable Energy Rev.*, 13(6):1465–1474, 2009.
- [217] T. Thiringer and J. Linders. Control by variable rotor speed of a fixed-pitch wind turbine operating in a wide speed range. *IEEE Transactions on Energy Conversion*, 8(3):520–526, 1993.
- [218] J. G. Slootweg and W. L. Kling. Modeling of large wind farms in power system simulations. In *2002 Power Engineering Society Summer Meeting Proceedings*, volume 3 of *Conference Proceedings*, pages 503–508. IEEE, 2002.
- [219] G. L. Park. Planning manual for utility application of wecs. Technical report, Michigan State Univ., East Lansing (USA). Div. of Engineering Research, 1979.

- [220] P. Giorsetto and K. F. Utsurogi. Development of a new procedure for reliability modeling of wind turbine generators. *IEEE transactions on power apparatus and systems*, PAS-102(1):134–143, 1983.
- [221] T. Jin and Z. Tian. Uncertainty analysis for wind energy production with dynamic power curves. In *11th Int. Conf. Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 745–750, San Marcos, June 2010. IEEE.
- [222] G. Sansavini et al. A stochastic framework for uncertainty analysis in electric power transmission systems with wind generation. *Renew. Energ.*, 64:71, 2014.
- [223] G. L. Johnson. *Wind energy systems*. Prentice hall, 1985.
- [224] F. W. Harris. How many parts to make at once. *Oper. Res.*, 38(6):947, 1990.
- [225] L. B. Schwarz. The economic order-quantity (eoq) model. In D. Chhajed and T. J. Lowe, editors, *Building Intuition*, volume 115 of *International Series in Operations Research & Management Science*, pages 135–154. Springer, 2008.
- [226] G. Chastaing et al. Generalized hoeffding-sobol decomposition for dependent variables-application to sensitivity analysis. *Electron. J. Stat.*, 6:2420–2448, 2012.
- [227] G. Chastaing, F. Gamboa, and C. Prieur. Generalized sobol sensitivity indices for dependent variables: numerical methods. *J. Stat. Comput. Simul.*, 85(7):1306–1333, 2015.
- [228] T. Ishigami and T. Homma. An importance quantification technique in uncertainty analysis for computer models. In *Proc. 1st Int. Symp. Uncertainty Modeling and Analysis*, pages 398–403. IEEE, 1990.
- [229] J. Morio. Global and local sensitivity analysis methods for a physical system. *Eur. J. Phys.*, 32(6):1577, 2011.
- [230] I. Lira. Beyond the gum: variance-based sensitivity analysis in metrology. *Meas. Sci. Technol.*, 27(7):075006, 2016.
- [231] C. Fraser et al. Pandemic potential of a strain of influenza a (h1n1): early findings. *Science*, 324(5934):1557–1561, 2009.
- [232] P. Holme and N. Masuda. The basic reproduction number as a predictor for epidemic outbreaks in temporal networks. *PloS one*, 10(3):e0120567, 2015.
- [233] R. M. Anderson, R. M. May, and B. Anderson. *Infectious diseases of humans: dynamics and control*, volume 28. Wiley Online Library, 1992.

- [234] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz. On the definition and the computation of the basic reproduction ratio r_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.*, 28(4):365–382, 1990.
- [235] O. Diekmann, J. A. P. Heesterbeek, and M. G. Roberts. The construction of next-generation matrices for compartmental epidemic models. *J. R. Soc. Interface*, 7(47):873–885, 2010.
- [236] G. Macdonald. *The epidemiology and control of malaria*. Oxford University Press, 1957.
- [237] P. Van den Driessche and J. Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.*, 180(1):29–48, 2002.
- [238] Z. Mukandavire and W. Garira. Hiv/aids model for assessing the effects of prophylactic sterilizing vaccines, condoms and treatment with amelioration. *J. Biol. Syst.*, 14(03):323–355, 2006.
- [239] R. Safiel, E. S. Massawe, and D. O. Makinde. Modelling the effect of screening and treatment on transmission of hiv/aids infection in a population. *Amer. J. Math. Stat.*, 2(4):75–88, 2012.
- [240] J. C. Helton and F. J. Davis. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliab. Eng. Syst. Safe.*, 81(1):23–69, 2003.
- [241] S. Kucherenko, S. Tarantola, and P. Annoni. Estimation of global sensitivity indices for models with dependent variables. *Comput. Phys. Comm.*, 183(4):937–946, 2012.
- [242] D. J. Daley and J. Gani. *Epidemic Modeling: An Introduction*. Cambridge Studies in Mathematical Biology. Cambridge University Press, 1999.
- [243] D. Bernoulli. Essai d’une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l’inoculation pour la prévenir. *Histoire de l’Acad. Roy. Sci. (Paris) avec Mém. des Math. et Phys. and Mém.*, 1:1–45, 1760.
- [244] R. Ross. *The prevention of malaria*. E.P. Dutton & company, New York, 1910.
- [245] W. Wang et al. Epidemic spreading on complex networks with general degree and weight distributions. *Phys. Rev. E*, 90(4):042803, 2014.
- [246] X. Wei et al. A unified framework of interplay between two spreading processes in multiplex networks. *EPL*, 114(2):26006, 2016.

- [247] S. Galam. Contrarian deterministic effects on opinion dynamics: the hung elections scenario. *Physica A*, 333:453–460, 2004.
- [248] A-L. Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005.
- [249] E. Ben-Naim. Opinion dynamics: rise and fall of political parties. *EPL*, 69(5):671, 2005.
- [250] W. Weidlich. The statistical description of polarization phenomena in society. *Br. J. Math. Stat. Psychol.*, 24:251, 1971.
- [251] S. Galam, Y. Gefen, and Y. Shapir. Sociophysics: A new approach of sociological collective behaviour. i. mean-behaviour description of a strike. *J. Math. Sociol.*, 9(1):1–13, 1982.
- [252] L. F. C. Pereira and F. G. B. Moreira. Majority-vote model on random graphs. *Phys. Rev. E*, 71(1):016123, 2005.
- [253] K. Sznajd-Weron and J. Sznajd. Opinion evolution in closed community. *Int. J. Mod. Phys. C*, 11(06):1157–1165, 2000.
- [254] P. L. Krapivsky and S. Redner. Dynamics of majority rule in two-state interacting spin systems. *Phys. Rev. Lett.*, 90(23):238701, 2003.
- [255] V. Sood and S. Redner. Voter model on heterogeneous graphs. *Phys. Rev. Lett.*, 94(17):178701, 2005.
- [256] G. Deffuant et al. Mixing beliefs among interacting agents. *Adv. Complex Syst.*, 3(01n04):87–98, 2000.
- [257] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *J. Artif. Soc. Soc. Simul.*, 5(3), 2002.
- [258] J. Lorenz. Continuous opinion dynamics under bounded confidence: A survey. *Int. J. Mod. Phys. C*, 18(12):1819–1838, 2007.
- [259] M. H. DeGroot. Reaching a consensus. *J. Amer. Statist. Assoc.*, 69:118, 1974.
- [260] S. Chatterjee and E. Seneta. Towards consensus: Some convergence theorems on repeated averaging. *J. Appl. Probab.*, 14:89, 1977.
- [261] Y. Zhao et al. Bounded confidence opinion dynamics with opinion leaders and environmental noises. *Comput. Oper. Res.*, 74:205–213, 2016.

- [262] L. Boudin and F. Salvarani. Opinion dynamics: Kinetic modelling with mass media, application to the scottish independence referendum. *Physica A*, 444:448–457, 2016.
- [263] A. R. Vieira, C. Anteneodo, and N. Crokidakis. Consequences of nonconformist behaviors in a continuous opinion model. *J. Stat. Mech.*, 2016(2):023204, 2016.
- [264] K. Fan and W. Pedrycz. Opinion evolution influenced by informed agents. *Physica A*, 462:431–441, 2016.
- [265] R. Albert and A-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47, 2002.
- [266] J. P. Onnela et al. Structure and tie strengths in mobile communication networks. *PNAS*, 104(18):7332–7336, 2007.
- [267] K. Choromański, M. Matuszak, and J. Mięgisz. Scale-free graph with preferential attachment and evolving internal vertex structure. *J. Stat. Phys.*, 151(6):1175–1183, 2013.
- [268] C. Brigatti. Consequence of reputation in an open-ended naming game. *Phys. Rev. E*, 78(4):046108, 2008.
- [269] N. Crokidakis and P. M. C. de Oliveira. The first shall be last: Selection-driven minority becomes majority. *Physica A*, 409:48–52, 2014.
- [270] M. A. Nowak and K. Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*, 393(6685):573–577, 1998.
- [271] N. Crokidakis and F. L. Forgerini. Consequence of reputation in the sznajd consensus model. *Phys. Lett. A*, 374(34):3380–3383, 2010.
- [272] N. Crokidakis. The influence of local majority opinions on the dynamics of the sznajd model. *J. Phys. Conf. Ser.*, 487:012016, 2014.
- [273] G. Weisbuch et al. Interacting agents and continuous opinions dynamics. In R. Cowan and N. Jonard, editors, *Heterogenous agents, interactions and economic performance*, volume 521 of *Lecture Notes in Economics and Mathematical Systems*, pages 225–242. Springer, 2003.
- [274] C. Anteneodo and N. Crokidakis. Symmetry breaking by heating in a continuous opinion model. *Phys. Rev. E*, 95(4):042308, 2017.
- [275] B. Derrida and H. Flyvbjerg. Multivalley structure in kauffman’s model: Analogy with spin glasses. *J. Phys. A*, 19(16):L1003, 1986.

Publications and pre-prints during PhD

1. **Y. Zhu**, Q. A. Wang, W. Li, and X. Cai, Analytic uncertainty and sensitivity analysis of models with input correlations, *submitted to Physica A*.
2. **Y. Zhu**, Q. A. Wang, W. Li, and X. Cai, The formation of continuous opinion dynamics based on a gambling mechanism and its sensitivity analysis, *Journal of Statistical Mechanics: Theory and Experiment (JSTAT)* 2017 (2017) 093401.
3. **Y. Zhu**, Q. A. Wang, W. Li, and X. Cai, Uncertainty and sensitivity analysis to complex systems, *International Journal of Modern Physics C* 28 (2017) 1750109.
4. **Y. Zhu**, Q. A. Wang, W. Li, and X. Cai, An analytic method for sensitivity analysis of complex systems, *Physica A* 469 (2017) 52.
5. **Y. Zhu**, L. Zhao, W. Li, Q. A. Wang, and X. Cai, Random walks on real metro systems, *International Journal of Modern Physics C* 27 (2016) 1650122.
6. **Y. Zhu**, W. Li, Q. A. Wang, and X. Cai, The dependence of graph energy on network structure, *Journal of Physics: Conference Series* 604 (2015) 012024.
7. J. Gu, **Y. Zhu**, L. Guo, et al., Recent Progress in Some Active Topics on Complex Networks, *Journal of Physics: Conference Series* 604 (2015) 012007.

Thèse de Doctorat

Yueying ZHU

Investigation on uncertainty and sensitivity analysis of complex systems

Résumé

Par un développement en série de Taylor, une relation analytique générale est établie pour calculer l'incertitude de la réponse du modèle, en assumant l'indépendance des entrées. En utilisant des relations de puissances et exponentielles, il est démontré que l'approximation souvent utilisée permet d'évaluer de manière satisfaisante l'incertitude sur la réponse du modèle pourvu que l'incertitude d'entrée soit négligeable ou que le modèle soit presque linéaire. La méthode est appliquée à l'étude d'un réseau de distribution électrique et à un modèle d'ordre économique.

La méthode est étendue aux cas où les variables d'entrée sont corrélées. Avec la méthode généralisée, on peut déterminer si les corrélations d'entrée doivent ou non être considérées pour des applications pratiques. Des exemples numériques montrent l'efficacité et la validation de notre méthode dans l'analyse des modèles tant généraux que spécifiques tels que le modèle déterministe du VIH.

La méthode est ensuite comparée à celle de Sobol. Les résultats montrent que la méthode de Sobol peut surévaluer l'incidence des divers facteurs, mais sous-estimer ceux de leurs interactions dans le cas d'interactions non linéaires entre les paramètres d'entrée. Une modification est alors introduite, aidant à comprendre la différence entre notre méthode et celle de Sobol.

Enfin, un modèle numérique est établi dans le cas d'un jeu virtuel prenant en compte la formation de la dynamique de l'opinion publique. L'analyse théorique à l'aide de la méthode de modification d'un paramètre à la fois. La méthode basée sur l'échantillonnage fournit une analyse globale de l'incertitude et de la sensibilité des observations.

Mots clés

Analyse d'incertitude, Analyse de sensibilité, Décomposition de variance, Échantillonnage, Corrélation, Mesure de sensibilité, Dynamique d'opinion

Abstract

By means of Taylor series expansion, a general analytic formula is derived to characterise the uncertainty propagation from input variables to the model response, in assuming input independence. By using power-law and exponential functions, it is shown that the widely used approximation considering only the first order contribution of input uncertainty is sufficiently good only when the input uncertainty is negligible or the underlying model is almost linear. This method is then applied to a power grid system and the eoq model.

The method is also extended to correlated case. With the extended method, it is straightforward to identify the importance of input correlations in the model response. This allows one to determine whether or not the input correlations should be considered in practical applications. Numerical examples suggest the effectiveness and validation of our method for general models, as well as specific ones such as the deterministic HIV model.

The method is then compared to Sobol's one which is implemented with sampling based strategy. Results show that, compared to our method, it may overvalue the roles of individual input factors but underestimate those of their interaction effects when there are nonlinear coupling terms of input factors. A modification is then introduced, helping understand the difference between our method and Sobol's one.

Finally, a numerical model is designed based on a virtual gambling mechanism, regarding the formation of opinion dynamics. Theoretical analysis is proposed by the use of one-at-a-time method. Sampling-based method provides a global analysis of output uncertainty and sensitivity.

Key Words

Uncertainty analysis, Sensitivity analysis, Variance decomposition, Sampling, Correlation, Sensitivity measure, Opinion dynamics