Sensitivity analysis and its role in expert judgement

Budhi Handoko

Submitted for the degree of Doctor of Philosophy

School of Mathematics and Statistics

March 2022

Supervisor: Professor Jeremy Oakley

The University Of Sheffield

To my
parents, family, and Universitas Padjadjaran

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This thesis is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and the following acknowledgements.

18 March 2022

Budhi Handoko

Acknowledgements

Firstly, I would like to acknowledge my Supervisor, Professor Jeremy Oakley, for his excellent supervision and patience for discussions and guidance during my study at the School of Mathematics and Statistics, The University of Sheffield.

I would like to express my gratitude to Universitas Padjadjaran, Indonesia as my sponsor, who has supported me and given me a great opportunity to study in The United Kingdom. Special thanks to Mr Gatot Riwi Setyanto as the former director of human resources and his staff: Mrs Rina and Mr Roby, and also Mrs Nur and Mr Nandang from the University Finance Department, Prof. Dr. Iman Rahayu as the Dean of Faculty of Mathematics and Natural Science and his staff, Dr Toni Toharudin as my former Head of Statistics department, Dr Irlandia Ginanjar as the Head of Statistics department and all colleagues in the department.

I would like to thank my parents for their support from the beginning of my study and motivate me at all time. A big thank you to my family, who give fantastic support and joy while living here in Sheffield. They are my lovely wife Yeny and my children: Danta, Mahran, and Shada.

Finally, I would thank all my friends in Sheffield for their exquisite attention and interaction. They are Indonesian friends in the Indonesian Student Union (PPI) and the Indonesian Muslim family community in The United Kingdom (KIBAR). Thanks to Jonathan Sykes, Laminu, Abdul Aziz, and all my friends in academia who fulfils my life with togetherness when studying at The University of Sheffield for the last four years.

Abstract

Sensitivity analysis has become an essential tool for assessing the importance of inputs in mathematical models or computer models. The models could be simple or complex and might not be formulated transparently. Model input uncertainty is often quantified using expert judgement, but obtaining or eliciting expert judgement can be time-consuming. It may not be practical to elicit distributions for all inputs in the model, especially when the number of inputs is large. Eliciting inputs that have a small effect on the output may not be worth it.

In this thesis, we explore sensitivity analysis methods and develop criteria for establishing when inputs are unimportant. We also study how to make elicitation for model inputs efficient, by employing sensitivity analysis to select which inputs are a priority for elicitation. However, we find that 'standard' application of sensitivity analysis for prioritising elicitation may not give the right results. We propose a novel model for expert judgement which we call the Beta Model. The Beta model is easy to implement compared with other previous models for expert opinion. Sensitivity analysis is proposed on the parameters of this model, and a grid-based computational method is proposed to compute suitable sensitivity indices.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this thesis, we investigate global sensitivity analysis and its application in eliciting probability distribution from experts. The context of the investigation is in computer modelling with uncertain inputs, where there has been much development of sensitivity analysis methods. The term global sensitivity analysis involving understanding how model outputs are sensitive to changes in model inputs over some range. Whereas, elicitation is a process of obtaining the expert beliefs about uncertain quantities and representing them with probability distributions. Elicitation is time consuming, and a modelling problem may require elicitation of a large number of input variables. Global sensitivity analysis concepts are explored deeply to determine whether they can support and prioritise elicitation. We first give a short overview of computer modelling and statistical issues, including elicitation and then outline the content of the thesis.

§ 1.1 Computer models and their issues

Mathematical modelling is essential in science and is an integral part of implementing scientific research and discovery. The models might involve non-linear or complex forms that are challenging to work with. Complex mathematical models are implemented in computer codes and are also called computer models. Computer models are used for many purposes. For instance, computer models are used to design a product, model the movement of the drug to various parts of a patient's body, and understand the behaviour of large-scale physical systems in astronomy. Computer models conduct 'virtual experiments' when physical experiments may be impractical or too costly.

In the following sections, we describe some general analysis problems involving computer models and understanding how the model output responds to changes in model inputs. Some problems that arise when using computer models or mathematical models are discussed, such as optimisation, prediction, and calibration. We also comment on the role of sensitivity analysis in the problems.

1.1.1 Optimisation of computer models

Computer models play an essential role in designing a product, especially when involving expensive materials. For example, a computer model is used to simulate car crashes to a solid target and determine the responsiveness of the airbag system in the car. However, performing such simulation using computer models mathematically will take a while as some complex functions need to be solved. Therefore, it is a great challenge to do optimisation for the models.

More specifically, let us consider the design of physical system in which we wish to maximise or minimise some target quantities. This system is represented by a function f that maps an input domain $\mathcal{X}$ onto the real sets $\mathbb{R}$, expressed as $f : \mathcal{X} \to \mathbb{R}$. Let us suppose a vector of input is $\mathbf{x} = (x_1, ..., x_p) \in \mathcal{X}$, where p is the number of inputs in the model, while y is the output. The relationship between inputs and output through a mathematical model is represented as $y = f(\mathbf{x})$.

The goal of optimisation is to maximise or minimise $f(\mathbf{x})$ with respect to the inputs $\mathbf{x}$. The optimisation process will depend on the model under study. If the model is simple and the function is not complex, it takes less time to find the solution. On the other hand, if the model consists of many inputs and has complex forms or non-closed forms, the optimisation will be time-consuming or computationally expensive.

An illustration of the optimisation of a mathematical model is provided by Tohsato et al. (2013). They considered a large number of inputs when studying kinetic model optimisation. Some studies on the optimisation of expensive functions or solution evaluations are conducted by Jones et al. (1998), and Knowles (2006). Jones et al. (1998) proposed the use of response surface methodology to approximate f and employed the approximation to construct an algorithm for efficient global optimisation. Another algorithm to perform an efficient global optimisation (EGO) was developed by Knowles (2006). He developed an algorithm ParEGO which stands for Pareto Efficient Global Optimisation which is an extension of the single-objective algorithm introduced by Jones et al. (1998). The ParEGO algorithm can be used for multiobjective optimisation problems.

What is the role of sensitivity analysis here? It may be helpful to understand the model better before conducting the optimisation. In particular, if we discover that only some inputs make significant changes to the output, this would reduce the complexity of the optimisation problem. We might only optimise the prioritised inputs and let the other input be constant.

In order to support the optimisation of computer models using this perspective, it is reasonable to ask a question: "how to detect inactive inputs that are not necessary to be involved in the optimisation"?

1.1.2 Prediction with uncertain inputs

Another role of a computer or mathematical model is to predict the output based on a set of inputs. The output of the computer model is the value that the analyst or the product designer wants to investigate. The model's output is used to adjust the system if it is unsuitable for the initial target or goal. However, if there is uncertainty on the true values of the inputs, the prediction would also lead to uncertainty. The problem of making a good prediction with a large number of inputs in the biochemical networks model was discussed by Vanlier et al. (2012). Another study on the ability of chemical kinetics models to provide a reliable prediction was addressed by Russi et al. (2008).

In order to understand the idea of prediction issue in the mathematical or computer model technically, let us consider a model $y = f(\mathbf{x})$. The true and uncertain inputs are denoted by $\mathbf{X} = (X_1, \ldots, X_p)$. The uncertainty about these inputs is represented by a probability distribution $P(X_1, \ldots, X_p)$ and the corresponding uncertain output is $Y = f(X_1, \ldots, X_p)$. The issue arises when the input uncertainty is high, which is indicated by a wide range of input space or a big variance. The high uncertainty in the inputs leads to the poor prediction of Y. Even though uncertainty in the inputs is unavoidable, the prediction of Y should also be accurate.

Therefore, prediction in computer models aims to predict Y with as little uncertainty as possible. One possible way to deal with this situation is the idea that it could be helpful to consider how the model output responds to changes in model inputs by considering sensitivity over the region of the input space described by $P(X_1, \ldots, X_p)$. It might also be helpful to consider if the uncertainty in Y depends only on some subsets of inputs. One might ask: "Can uncertainty about Y be reduced effectively by learning a subset of inputs $(X_1, \ldots, X_p)$?".

1.1.3 Calibration

A mathematical or computer model is intended to represent a real physical system. Hence, it is crucial to check the model and fit it based on physical observations. The process of fitting the model to the observed data by adjusting the parameters in the model is known as calibration (Kennedy & O'Hagan, 2001). Nevertheless, undertaking calibration for complex computer models is challenging as it is highly computationally expensive, and it cannot run all possible simulations to check with the available observations. A complex computer model may take hours or days to run a single simulation.

An example of a computer model by which calibration is challenging to carry out is the Long-term Generation Investment (LTGI) model (Xu et al., 2016). Policymakers use the LTGI model for predicting real-world generation projections to make decisions on investment and energy design policy. Klepper (1997) raised an issue of conducting calibration for environmental models that have high order characteristics. In this case, calibrating a non-linear model with a medium number of inputs (e.g. 10) is an intractable problem.

Performing calibration involves collecting the data related to the outputs of the model. We want to study how the observed data affects model inputs' uncertainty. Suppose we have a prior probability distribution $P(\mathbf{X})$. If we collected some observations from real physical system Z with noise ϵ, its relationship with the model $f(\mathbf{X})$ would be $Z = f(\mathbf{X}) + \epsilon$. The question regarding the calibration are: Would Z be informative for any elements of $\mathbf{X}$? Is the likelihood for Z responsive to changes in elements of $\mathbf{X}$?

§ 1.2 Problems in Elicitation

Having considered some problems in computer models, we also study the problem of elicitation. Elicitation is defined as the process of expressing expert judgment in the form of statistical distributions. We will discuss elicitation in detail in Chapter 4. The elicitation can be used in Section 1.1.2 to obtain input distribution and Section 1.1.3 as a prior for calibrating inputs. Nevertheless, the process is difficult and time-consuming. It is also possible that not all inputs are important. In particular, important means that output is not sensitive to all inputs. Therefore, it motivates the author to use sensitivity analysis methods to help prioritise expert judgment.

§ 1.3 Chapter Overview

This thesis consists of seven chapters, and the following chapters are outlined below.

Chapter 2 presents the concept of sensitivity analysis and some sensitivity measures or importance measures used to determine the important inputs in the computer or mathematical models. The importance measures methods based on different approaches such as derivative-based methods, regression-based methods, variance-based methods, and value of information-based methods. In this chapter, we replicate and extend an example presented of a flood model (see Iooss & Lemaître, 2014) to compare different sensitivity analyses.

Chapter 3 investigates the concept of 'unimportant' model inputs. We formalise this concept by developing two scenarios where the model user considers specific actions. Subsequently, we evaluate some methods proposed in the literature against these scenarios.

Chapter 4 reviews elicitation methods such as SHELF, IDEA, and classical methods. This chapter provides an example of elicitation using the SHELF method to determine the distribution of regular smokers in Indonesia.

Chapter 5 sets out a framework for using sensitivity analysis to prioritise expert elicitation. This chapter also reviews robust global sensitivity analysis methods and establish the need for a novel approach.

Chapter 6 develops a novel solution to the problem set out in Chapter 5 involving a model for expert judgment. We use some relevant sensitivity measures, as we found as findings in Chapter 3 and implement the measures. We also develop and test the computational methods.

Chapter 7 concludes the thesis and discusses the findings.

§ 1.4 The novelty

This thesis contains some novelties as a result of the research findings. These novelties are expected to contribute to the sensitivity analysis field's development and elicitation. Three novelties are achieved from the research and presented in the specified chapters below.

The first novelty is in Chapter 3, which introduces criteria for identifying unimportant inputs in computer models via two scenarios. These criteria are explicit rules in deter-

mining unessential inputs, and we find that none of the methods considered meets the requirements set out in both scenarios.

The second novelty is in Chapter 5. Through a counter-example, we show that the problem of prioritising expert judgement is not a 'standard' sensitivity analysis problem. Thus, a new framework is needed.

Chapter 6 provides novel methods of implementing sensitivity analysis in the context of elicitation. We propose the Beta model, which enables the analyst to capture the possibility of expert distributions by providing prior distributions for the parameters in the Beta model. We use the sensitivity analysis to select the necessary inputs and keep the rest elicited by analysts instead of experts. This chapter provides the algorithm and computational method for computing the measure of importance for assessing both inputs and parameters in the beta models.

Chapter 2

Sensitivity Analysis

§ 2.1 Introduction

As discussed in Chapter 1, some issues in computer models need to be tackled. Most of the issues are related with the relationship between uncertain inputs in the models, which affect the outputs. In this chapter, we consider how the concept sensitivity analysis may be used to help with these issues. Some settings and existing methods in sensitivity analysis are discussed, focusing on the factor prioritisation setting. We illustrate the methods by implementing the sensitivity measures on the flood model and comparing the results.

§ 2.2 Sensitivity analysis and importance measures

According to Saltelli et al. (2004), sensitivity analysis is defined as the study of how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input. Uncertainty in the model output is affected by uncertainty in inputs via a function or system. Using the sensitivity analysis, we can measure how sensitive the change in output of a model as a result of changes in the input. The next section introduces some basic notations of the functional relationship between input and output.

2.2.1 Notation

Let us suppose we have a vector of inputs $\mathbf{x} = (x_1, ..., x_p)$, where p is the number of inputs and the output is denoted by y. The relationship between inputs and output $f : \mathcal{R}^p \to \mathcal{R}$

can be represented in the form of a mathematical function

$$y = f(\mathbf{x}). \tag{2.1}$$

A "local" sensitivity analysis may involve considering some point $\mathbf{x}_0$ in input space and investigating how the output responds to small changes to the input in the neighbourhood of $\mathbf{x}_0$.

In "global" sensitivity analysis, we consider how the output responds to changes of the input over some region. This region can be defined by treating the inputs as random variables, represented in a random vector $\mathbf{X} = (X_1, ..., X_p)$. In particular $\mathbf{x} = (x_1, ..., x_p)$ denotes one of the possible input realizations of $\mathbf{X}$. The probability distribution for $\mathbf{X}$ describes the input region of interest. Global sensitivity analysis then involves an investigation of the random variable $Y = f(\mathbf{X})$.

2.2.2 The Flood model

In this section, we replicate and extend an example presented of a flood model (see Iooss & Lemaître, 2014) to compare different sensitivity measures. The notation used in the flood model has been changed from the original notation to variabel x_i, where $i = 1, ..., 8$

$$y = x_1 - x_2 - x_3 + \left(\frac{x_4}{x_5 x_6 \sqrt{\frac{x_7 - x_1}{x_8}}} \right)^{0.6}. \tag{2.2}$$

The model describes factors that influence maximal annual overflow (in meters). In particular, maximal annual overflow is the model's output and is denoted by y. Some factors that affect y are river downstream level (in meters) denoted by x_1, dyke height (in meters) denoted by x_2, bank-level (in meters) denoted by x_3, and the maximal annual height of the river (in meters) which is expressed by the last term of the model.

The maximal annual height of the river is computed from some other factors such as maximum annual flow rate (x_4) in m^3/s, river width (x_5) in m, Strickler coefficient (x_6), river upstream level (x_7) in m, and the length of the river stretch (x_8) in m. The uncertainty in the factors/inputs is represented in the form of probability distributions. The distributions for inputs are provided in Table 2.1. The flood model will be used as a model example throughout this thesis.

Table 2.1: Inputs of the flood model and their probability distributions

Input	Description	Probability Distribution
X_1	River downstream level	Triangular(49, 50, 51)
X_2	Dyke height	Uniform(7,9)
X_3	Bank level	Triangular(55, 55.5, 56)
X_4	Maximum annual flow rate	Truncated Gumbel (1013, 558) on [500, 3000]
X_5	River width	Triangular(295, 300, 305)
X_6	Strickler coefficient	Truncated Normal(30, 8) on [15, ∞]
X_7	River upstream level	Triangular(54, 55, 56)
X_8	Length of the river stretch	Triangular(4990, 5000, 5010)

2.2.3 Objectives for Sensitivity Analysis

There might be different motivations for analysts in performing a sensitivity analysis. Different research questions and goals could lead to distinct sensitivity analyses regarding how the sensitivity analysis was conducted. Factor settings were introduced to provide a direction for analysts on what appropriate types of sensitivity analysis should be used. According to Saltelli et al. (2008), a setting is a way of framing the sensitivity quest in such a way that the answer can be confidently entrusted to a well-identified measure. They proposed four different settings in the sensitivity analysis: factor prioritisation, factor fixing, variance cutting, and factor mapping. Each setting is used for different tests of sensitivity analysis, and it is important to choose the appropriate setting to solve the problem at hand. The settings are briefly described below. Note that "factor" here means a random input X_i.

a) **Factor Prioritisation**

Factor prioritisation is used to identify a factor (in this context is a model input) X_i in a mathematical model or computer model which gives the greatest reduction in the variance of Y, when the input is fixed to its true value. The factor that has been identified by this setting is called the most important factor. First-order sensitivity or main effect index is used as a measure in this setting which will be discussed in Section 2.2.7.

b) **Factor Fixing**

Factor fixing is used to identify factors that make no significant contribution to $Var(Y)$. Factors that are identified with this setting are considered non-influential factors. This setting is very useful when our purpose is to do a model simplification. Commonly, the "total effect index" is used in this setting to identify non-influential

factors in the model. The detail of the factor fixing and its development will be provided in the next chapter.

c) **Variance Cutting**

In variance cutting, the aim is to reduce $Var(Y)$ to below a given tolerance by fixing the factors. The factors are fixed one at a time until $Var(Y)$ is below some specified bound. The goal is to fix the smallest possible number of factors; hence we know which factors can be set into constants to obtain the target value on $Var(Y)$.

d) **Factor Mapping**

Factor mapping is used to identify factors responsible for the output realisation Y in a certain region. The region of the output realisation could be split into two areas based on, for example, its percentile. The goal is to find factors that contribute to the realisation of output in each area or factors that are influencing the splitting of output realisation. This setting is useful for determining a good combination of factors that are used to minimise the loss or risk.

In this chapter, the scope of factor setting is limited to factor prioritisation. The following section provides a detailed explanation of the factor prioritisation by providing various types of importance measures used in the sensitivity analysis that are available in the literature. The measures are based on different approaches such as derivative-based methods, regression-based methods, and variance-based methods. Each of these methods has its unique use and benefit.

Derivative-based methods are the simplest method of sensitivity analysis which consider only the partial derivatives of a model concerning an input. The method helps undertake sensitivity analysis in a fast way as long as the model is differentiable without taking into account the probabilistic analysis. The method is a mathematical approach and does not need any numerical computations. On the other hand, regression-based methods are used when the relationship between input and output becomes the analyst's concern. The regression-based methods are suitable for performing sensitivity analysis if the relationship pattern between inputs and output is linear. Some measures are based on quantities computed from the regression model, such as parameter estimates and model prediction. This method also considers other statistical methods such as correlation and association analyses. Unlike the regression-based methods, the variance-based method does not require a linear relationship between inputs and output; thus, it can be used for any type

of mathematical model, both linear and non-linear. It is widely used in practice as it accommodates global sensitivity analysis and has desirable properties, which were discussed in (Saltelli et al. 2000, pages 123-124). One of the properties is that the variance-based methods do not depend on the form of mathematical models, making them independent in performing sensitivity analysis.

2.2.4 Sensitivity analysis given data approaches

In this thesis, we will use the "given data" approaches in undertaking sensitivity analysis with different methods and models. Given data approaches are commonly used in sensitivity analysis when the data about inputs are available or the distribution of inputs that indicates their uncertainty is provided. When data are available, statistical approaches for modelling such as regression and correlation to investigate the relationship between variables are convenient to perform. The form of relationship could be linear, quadratic, cubic, or quartic depending on the pattern plot between variables. However, when the amount of data is too small or sparse, then the correlation is skewed. To overcome this issue, Pearson (1905) introduced a general method when dealing with non-linear regression and the variability of arrays in the case of skew correlation. In this case, one can calculate second and third-order moments to describe the deviation from the linearity and skewness of the regression line.

Another work that uses "given data" or data-driven to conduct sensitivity analysis is presented by Strong et al. (2012). The data is sampled using Monte Carlo sampling from the probability distribution of inputs. This is one way of quantifying input uncertainty by specifying a probability distribution for the true value of the inputs. Using this approach, the probabilistic sensitivity analysis focuses on quantifying uncertainty in the model output based on the uncertainty in the model input. If uncertainty arises in the model structure, Strong et al. (2012) proposed to use the discrepancy method. The Monte Carlo procedure is also used to estimate the Expected Value of Perfect Information (EVPI). The EVPI quantifies the value of learning input in the decision model. The computation of EVPI requires a two-level monte Carlo approach and is estimated using MCMC. Strong & Oakley (2013) proposed a faster and straightforward method using a single set method.

A new strategy for estimating global sensitivity measures from given data is proposed by Plischke et al. (2013). Using this strategy, a notable reduction in computational burden is achieved which makes the estimation cost independent of factors. This strategy is appropriate to implement in factor prioritisation and factor settings for a model with a

large number of inputs. According to Plischke et al. (2013), there are several ways of generating random inputs from a joint probability distribution: crude Monte Carlo, Latin hypercube sampling, and quasi-random low-discrepancy sequences. The input sample is then fed into the model to obtain the output distribution.

We implement the sensitivity analysis using given data to importance measure methods in the following section. The details of each method will be provided in the following subsections, along with some other approaches. Note that there are other sensitivity measures (Shapley values and distribution-based indices) that can be applied in the factor fixing setting, and we will discuss them in Chapter 3.

2.2.5 Derivative-based method

The derivative-based method is frequently named local sensitivity analysis (LSA). According to Saltelli et al. (2004), a local sensitivity measure that looks at the relation between input and output at a specified point in the space of the inputs. Let us consider a reference point

$$y^0 = f(\mathbf{x}_0) \tag{2.3}$$

with $\mathbf{x}_0 = (x_1^0, \ldots, x_k^0)$. One derivative-based sensitivity measure would be

$$S_i^u = \frac{\partial y}{\partial x_i} \tag{2.4}$$

evaluated at $\mathbf{x} = \mathbf{x}_0$. Local sensitivity measures are usually normalised by some reference values to remove the effect of measurement unit of X_i (e.g. meter, pounds, kg, etc.). A normalised measure would be

$$S_i^n = \frac{x_i^0}{y^0} \frac{\partial y}{\partial y_i}. \tag{2.5}$$

Both measures, S_i^u and S_i^n, are local sensitivity measures as they do not require any assumptions on the range of input variation.

LSA is efficient in terms of computation time but has two limitations. Firstly, the derivative-based method is only informative at the base point, yet it does not provide information about the other points in the input spaces. That makes the method is unable to capture the uncertainty in the input. Secondly, it is very challenging to use the derivative approach when the model is not linear, has a complex form, or the input-output model is not in a closed form. Such that derivatives are not easily obtained.

2.2.6 Regression-based methods

The regression-based method can be used when both inputs and output samples are available. In this case, the relationship between inputs and output should be linear. To investigate whether there is a linear relationship between inputs and output, we can use the coefficient of determination R^2 of the model or Nash-Sutcliffe model efficiency, which is recommended by Iooss & Lemaître (2014). We collect n samples $(x_{1i}, ..., x_{pi}, y_i)$, where $i = 1, ..., n$ and $n > p$, and each input vector is sampled from the probability distribution for $\mathbf{X}$. The relationship between inputs and output are fitted by a linear model or approximated by a regression model

$$y = f(x_1, ..., x_p) \approx \hat{\beta}_0 + \sum_{i=1}^{p} \hat{\beta}_i x_i.$$

According to Borgonovo (2017) some important sensitivity measures can be obtained from linear models.The measures are as following.

(i) Standardized regression coefficient (SRC). The sensitivity formula for SRC is

$$SRC_i = \hat{\beta}_i \sqrt{\frac{Var(X_i)}{Var(Y)}} \tag{2.6}$$

where X_i is the sample data for input ke-i in the linear model and Y is the sample data for output.

(ii) Pearson's correlation coefficient (ρ). The correlation coefficient is also can be used as a sensitivity measure with the formula

$$\rho(Y, X_i) = \frac{Cov(Y, X_i)}{\sqrt{Var(X_i)}\sqrt{Var(Y)}} \tag{2.7}$$

(iii) Partial Correlation Coefficient (PCC). The PCC can be expressed as

$$PCC_i = \rho(X_i - \hat{X}_{-i}, Y - \hat{Y}_{-i}) \tag{2.8}$$

where $\hat{X}_{-i}$ is the prediction of linear model when X_i is treated as an output while the rest of the inputs remain in the model. $\hat{Y}_{-i}$ is the prediction of the linear model when X_i is not involved in the model.

When inputs and output has non-linear relationship, Iman & Conover (1979) suggested to transform the samples $(x_{1i}, ..., x_{pi}, Y_i)$ into ranks $(R(x_{1i}), ..., R(x_{pi}), R(y_i))$. Sensitivity

indices are measured using the previous measures on the rank values producing the Standardized Rank Regression Coefficient (SRRC), Spearman correlation coefficient (Spearman 1987), and the Partial Rank Correlation Coefficient (PRCC).

There are some advantages and disadvantages of the use of regression-based methods. The advantages are: (i) Sensitivity indices can be estimated straightforward using a Monte Carlo sample, (ii) Non-parametric sensitivity measures such as SRRC and PRCC can be computed without a specific design. The disadvantages are (i) If R^2 (determination coefficient) is too small, then the regression is poor fitted and the importance measures become unrealistic; (ii) If the model is non linear or there is interaction between inputs, then the regression models is not fit and the sensitivity measure derived from the regression model is not correct.

2.2.7 Variance-based methods

The variance-based approach to sensitivity analysis involves investigating how uncertain input in $\mathbf{X}$ contribute to the variance of $Y = f(\mathbf{X})$. In particular, a decomposition of the function f is used to decompose $Var(Y)$ into terms associated with different inputs.

In mathematical modelling, we always deal with a function f of inputs $\mathbf{x} = (x_1, ..., x_p)$ where p is the number of input. The question is how to represent the function $f(\mathbf{x})$?

Sobol (1993) describes an expansion of function $f(\mathbf{x})$ into summands (summation) of different dimensions. The function $f(\mathbf{x})$ must be an integrable function on input space $x_i \in [0,1]^p$. This corresponds to supposing that each $X_i \sim U[0,1]$ and independent; other distributions can be assumed, as long as independence is maintained. An expansion of $f(\mathbf{x})$ is represented as

$$f(\mathbf{x}) = f_0 + \sum_{i}^{p} f_i(x_i) + \sum_{i} \sum_{j>i} f_{ij}(x_i, x_j) + ... + f_{12...p}(x_1, ..., x_p) \tag{2.9}$$

where f_0 is a constant. This expression is unique under conditions

$$\int_0^1 f_{i_1,...,i_s}(x_{i_1}, ..., x_{i_s}) dx_{i_k} = 0, \tag{2.10}$$

where $1 \leq k \leq s$ and $\{i_1, ..., i_s\} \subseteq \{1, ..., p\}$. According to Sobol (1993) based on the conditions, the constant term is

$$f_0 = \int_x f(\mathbf{x}) d\mathbf{x} \tag{2.11}$$

Now, let us consider a function

$$g_i(x_i) = \int_0^1 \dots \int_0^1 f(\mathbf{x}) d\mathbf{x}/dx_i \tag{2.12}$$

where $d\mathbf{x}/dx_i$ is the product of all $d\mathbf{x}$ except dx_i. By integrating (2.9) with respect to all inputs except x_i, the function $g_i(x_i)$ can be simplified into

$$g_i(x_i) = f_0 + f_i(x_i). \tag{2.13}$$

Therefore for a single input function, $f_i(x_i)$ can be expressed as

$$f_i(x_i) = g_i(x_i) - f_0. \tag{2.14}$$

Let us consider another functions with two inputs $g_{ij}(x_i, x_j)$ which is obtained from the integration

$$g_{ij}(x_i, x_j) = \int_0^1 ... \int_0^1 f(\mathbf{x}) d\mathbf{x}/dx_i dx_j \tag{2.15}$$

where $i < j$. We obtain the following formula

$$g_{ij}(x_i, x_j) = f_0 + f_i(x_i) + f_j(x_i) + f_{ij}(x_i, x_j). \tag{2.16}$$

Obviously, from (2.16) the two-input function is

$$f_{ij}(x_i, x_j) = g_{ij}(x_i, x_j) - f_i(x_i) - f_j(x_i) - f_0. \tag{2.17}$$

In the global sensitivity analysis, input vector $\mathbf{x}$ is uncertain and treated as random vector $\mathbf{X} = (X_1, ..., X_p)$ which are mutually independent. Consequently, the output $Y = f(\mathbf{X})$ turns into random variable as well. We can rewrite (2.9) as an expansion of function in term of random variables

$$Y = f(\mathbf{X}) = f_0 + \sum_i^p f_i(X_i) + \sum_i \sum_{j>i} f_{ij}(X_i, X_j) + \dots + f_{12\dots p}(X_1, \dots, X_p). \tag{2.18}$$

In the variance-based method, we aim to evaluate the variance of Y in (2.18). As a result, we obtain the following variance decomposition

$$Var(Y) = \sum_i^p Var\{f_i(X_i)\} + \sum_i \sum_{j>i} Var\{f_{ij}(X_i, X_j)\} + ... + Var\{f_{12\dots p}(X_1, \dots, X_p)\} \tag{2.19}$$

In the context of random variables, the expression of (2.11), (2.14), and (2.17) turn into

$$f_0 = \mathbb{E}[f(\boldsymbol{X})] \tag{2.20}$$

$$f_i(x_i) = \mathbb{E}[f(\boldsymbol{X})|X_i = x_i] - f_0 \tag{2.21}$$

$$f_{ij}(x_i, x_j) = \mathbb{E}[f(\boldsymbol{X})|X_i = x_i, X_j = x_j] - f_i(x_i) - f_j(x_j) - f_0 \tag{2.22}$$

The functions for more than two inputs are derived using the similar fashion. Based on the variance decomposition in (2.19) and the component function in (2.21), we can derive the expression for variance of a single input function as

$$\begin{aligned} Var\{f_i(X_i)\} &= Var_{X_i}[E\{f(\mathbf{X})|X_i\} - f_0] \\ &= Var_{X_i}[E(Y|X_i)] \end{aligned} \tag{2.23}$$

and the variance for two input functions using (2.22) is

$$\begin{aligned} Var\{f_{ij}(X_i, X_j)\} &= Var_{X_iX_j}[E\{f(\mathbf{X})|X_i, X_j\} - f_i(X_i) - f_j(X_j) - f_0] \\ &= Var_{X_iX_j}[E(Y|X_i, X_j)] - Var_{X_i}[E(Y|X_i)] - Var_{X_j}[E(Y|X_j)]. \end{aligned} \tag{2.24}$$

We can also derive the variance decomposition for the sum of two component functions, a single input functions and two inputs function

$$Var\big(f_i(X_i)+f_{i,j}(X_i, X_j)\big) = Var_{X_i,X_j}\big(\mathbb{E}(Y|X_i, X_j)\big)+2Var_{X_i}\big(\mathbb{E}(Y|X_i)\big)+Var_{X_j}\big(\mathbb{E}(Y|X_j)\big). \tag{2.25}$$

The proof for equation (2.25) is in the Appendix.

If we divide equation (2.23) by the unconditional variance $Var(Y)$, we obtain the variance-based measure for input X_i as following

$$S_i = \frac{Var_{X_i}[E_{X_{-i}}(Y|X_i)]}{Var(Y)}. \tag{2.26}$$

S_i is called the main effect index or first-order effect index or importance measure and is used as an instrument to measure important inputs in the factor prioritisation setting. We can write Equation (2.26) as the expected reduction in variance and scaled it with the variance of Y as

$$S_i = \frac{Var(Y) - E_{X_i}[Var(Y|X_i)]}{Var(Y)}. \tag{2.27}$$

As we can see that the main effect index formulae in (2.26) and (2.27) consider the difference between the unconditional variance of Y and the expected conditional variance of $Y|X_i$. This difference is derived from the general principle of global sensitivity analysis which has been discussed in Section 2.2.12. This also applies to the total effect index. According to Borgonovo et al. (2014), the variance-based measures are derived from the global sensitivity measure for input X_i

$$\gamma^Y(X_i) = \mathbb{E}\{T(X_i)\} = \mathbb{E}\{t(\mathbb{P}, \mathbb{P}_{Y|X_i})\} \tag{2.28}$$

where $t(\cdot, \cdot)$ is a function that measures the discrepancy between $\mathbb{P}_Y$ and $\mathbb{P}_{Y|X_i}$. The main effect index is obtained by setting

$$T(x_i) = \frac{Var(Y) - Var(Y|X_i = x_i)}{Var(Y)}, \tag{2.29}$$

and the total effect index is obtained by setting

$$T(x_i) = \frac{Var(Y|X_{-i} = x_{-i})}{Var(Y)}. \tag{2.30}$$

Nevertheless, according to Borgonovo et al. (2014), not all of function $t(\cdot, \cdot)$ can produce the appropriate main effect index or total effect index formula. This is because the expectation of conditional variance might be zero which is unwanted. Some families of functions may cause the null expectation if we apply statistic $T(x_i)$ in (2.29) and (2.30). The general family of functions is presented by Plischke et al. (2013) as follows

$$\begin{gathered} y = z(x_1, \cdots, x_p) = a(x_J)g(x_j) + b(x_{J'}), \\ J \oplus \{j\} \oplus J' = \{1, \cdots, p\} \end{gathered} \tag{2.31}$$

where J and J' indicates two subsets of indices such that $j \in J \cup J'$, $J \cap J' = \emptyset$. Whereas $g(x_j)$, $a(\cdot)$, and $\underline{(\cdot)}$ depend on x_j. In the case that random vector X_j is independent with $X_{J'}$, it is shown that if $\mathbb{E}\{g(X_j)\} = 0$ then the global sensitivity measure for X_i, $\eta_1^Y(X_i) = 0$ for all j. In addition, if the probabilities $\mathbb{P}_y$ and $\mathbb{P}_{Y|X_i}$ follow the Pareto law with a shape parameter smaller than two, the variance cannot be used to measure the gap between these probabilities (Soofi 1994).

The advantages of the variance-based method are: (i) The method works well for measuring the sensitivity of the output for non-linear models or models with interaction; (ii) The method is able to compute some terms such as the main effect, interaction effect, and "total effect" for an input (which we will discuss in Chapter 3).

The variance-based method has some drawbacks: (i) This measure is intertwined with the functional ANOVA decomposition. There is no longer linkage between functional

decomposition and variance decomposition when inputs are correlated. As a result, the integration measure is not a product measure, variance decomposition cannot be derived from the functional ANOVA decomposition; (ii) Variance-based method is based on one moment, which cannot be used to fully describe the uncertainty of an input.

A substantial application of variance-based sensitivity analysis is given in Lee et al. (2011). In this example, the researchers investigate the effect of parameter uncertainty in a 3-D global aerosol model. The aerosol model, which represents the climate model, has a complex form and is not in a closed form. Therefore, they used a complex computer model in conducting the GSA. This process employs a Bayesian approach that uses prior distribution and data from model runs to predict the uncertainty on model output. Lee et al. (2011) used a subgroup of aerosol particles, which is named cloud condensation nuclei (CCN), as the output of interest. The CCN concentration (Y) is depend on eight parameters: diameter of oxidation activation (X_1), coefficient of mass accommodation (X_2), threshold for H_2SO_4 nucleation (X_3), crucial cluster size of nucleation (X_4), particulate emissions associated with anthropogenic SO_2 (X_5), cloud nucleation scavenging diameter (X_6), sulphur emissions (X_7) and sea spray emissions (X_8).

2.2.8 Value of information - based method

In the value of information approach, sensitivity measures are related to a decision problem and how inputs may influence a decision. Let y_d be the output of the model for a quantity of interest such as payoff, loss, utility, or other criteria, under a decision option d, where $d = 1, ..., D$ and D is a set of possible choices. The relationship between input and output is $Y_d = g_d(\mathbf{X})$, where g_d would be a function of the model output $f(\mathbf{X})$ and other quantities related to the choice of decision d. Subsequently, we assume that $g_d(\mathbf{x})$ is the decision-maker's utility for decision d and input $\mathbf{x}$. A decision-maker is then needed to solve the problem below.

$$\max_{d=1,...,D}\{E[Y_d]\} = \max_{d=1,...,D}\{E[g_d(\mathbf{X})]\} \tag{2.32}$$

Value of information was introduced by Howard (1966) and defined as a gap between the expected utilities. The first expected utility is someone who knows the true value of X_i and the expected utility someone would obtain if the decision is made without this knowledge. The following formula expresses it

$$\epsilon_i = E[\max_j\{Y_j\}|X_i] - \max_j\{E[Y_j(\mathbf{X})]\} \tag{2.33}$$

Examples of using this for sensitivity analysis are given in Oakley (2009) and Strong et al. (2012) where it is referred to as the (partial) expected value of perfect information (EVPI).

The advantage of the information-based method is that it uses another perspective to determine sensitivity measures that can be directly implemented to solve decision theory problems. The disadvantage of the method is that it requires decision problems before the sensitivity measure calculation, which is a particular case. Many mathematical models do not have a problem definition for calculating the sensitivity index.

We implement the EVPI method to compute the sensitivity indices for the flood model. To compute the EVPI, we start by defining two possible decisions according to the criteria. Let us suppose that we are interested in the output Y will exceed a threshold c. For illustration, we choose c to be -10.8. Two decisions are defined as follows.

Decision d_1: state that $Y \leq c$

Decision d_2: state that $Y > c$

If we choose the decision 'correctly', then there is no loss, but if we choose 'incorrectly', the loss is 1. The expected loss of d_1 is

$$0 \times P(Y \leq c) + 1 \times P(Y > c) = 0.446$$

and the expected loss of d_2 is

$$1 \times P(Y \leq c) + 0 \times P(Y > c) = 0.554$$

d_1 has lower expected loss than d_2 and the baseline expected loss is $L^* = 0.446$. The EVPI for input X_i is

$$\text{EVPI}_i = L^* - E_{X_i}[g(X_i)]$$

where

$$g(X_i) = \min\big[E\big(L(d_1)|X_i\big),\, E(L(d_2)|X_i\big)\big].$$

and

$$E(L(d_1)|X_i) = 0 \times P(Y \leq c|X_i) + 1 \times P(Y > c|X_i)$$

and

$$E(L(d_2)|X_i) = 1 \times P(Y \leq c|X_i) + 0 \times P(Y > c|X_i).$$

We normalise by the baseline expected loss L^* to give an EVPI index.

Some importance measure methods that have been discussed in the previous section are applied for analysing the Flood model in Equation (2.2). The computation of sensitivity indices is performed using a simulation of 100,000 samples.

The results of the importance measure index computation are provided in Table 2.2. The table indicates that SRC and Pearson correlation provide relatively close results. The variance-based method (S_i) produced relatively close indices with the EVPI. On the other hand, PCC tends to produce larger values for most inputs. The largest sensitivity index for all importance measures is X_4, implying that X_4 is the most important input in the Flood model. In contrast, X_8 is the least important input with the lowest index.

Table 2.2: Sensitivity indices using different importance measures

Input	SRC_i	Pearson ρ_i	PCC_i	S_i	EVPI_i
X_1	0.4663	0.4649	0.9881	0.2196	0.2451
X_2	-0.5743	-0.5719	-0.9921	0.3323	0.3622
X_3	-0.2032	-0.2046	-0.9417	0.0425	0.0908
X_4	0.6203	0.6191	0.9932	0.3894	0.3912
X_5	-0.0099	-0.0070	-0.1350	0.0000	0.0621
X_6	-0.1431	-0.1384	-0.8919	0.0195	0.0717
X_7	-0.0621	-0.0651	-0.6499	0.0043	0.0621
X_8	0.0008	-0.0028	0.0115	0.0000	0.0621

2.2.9 The screening method

Often, mathematical models have so many inputs which cause the computational time to evaluate the model takes quite a while. The model might be only has a few important inputs and some other inputs might be negligible. To screen a bunch of inputs and identify which input is important and which ones can be negligible, Morris (1991) proposed a method called the elementary effect. The elementary effects is computed using discretised approach. Let us suppose a model has p inputs $x_1, \ldots, x_p$ that varies in the p dimensional unit cube accross k, with k is the number of levels. These levels represents the number of grid used to discretise the input space. In particular, x_i may take values from $\{0, 1/(k-1), 2/(k-1), \ldots, 1\}$ and $\mathbf{x} = (x_1, \ldots, x_p)$. The elementary effect of the i-th input is expressed by

$$d_i(\mathbf{x}) = \frac{(y(x_1, x_2, \ldots, x_{i-1}, x_i + \Delta, x_{i+1}, \ldots, x_p) - y(\mathbf{x}))}{\Delta} \tag{2.34}$$

where Δ is specified as the multiple of $1/(k-1)$. For convenient purposes, the number of input p is even and $\Delta = k/(2(k-1))$. The elementary effect for input x_i has a distribution F_i as it is randomly sampled from different $\mathbf{x}$. The Morris' sensitivity measures are obtained from the estimates of the mean, μ and the standard deviation σ of the distribution of F_i.

Campolongo et al. (2007) revised the method of elementary effect by proposing a better sampling technique. The sampling is performed by choosing r trajectories in the input region to maximise the dispersion. The sample chosen should have the highest spread or distance between a pair of trajectories. If there are two trajectories t_1 and t_2, the distance between the two is measured by

$$d_{t_1 t_2} = \begin{cases} \sum_{i=1}^{k+1} \sum_{j=1}^{k+1} \sqrt{\sum_{z=1}^{k} [X_i^{t_1}(z) - X_j^{t_2}(z)]^2} & , \quad \text{for} \quad t_1 \neq t_2 \\ 0 & , \quad \text{otherwise} \end{cases}$$

where k is the number of inputs and $X_i^{t_1}(z)$ denotes the zth coordinate of the ith point of the mth trajectory. Campolongo et al. (2007) had also refined the measure μ used in the screening method used by Morris (1991) and denoted by μ^*. This refinement is expected to overcome the inefficiency of the use of two measures μ and σ when the model is complex. The new measures μ^* is computed as the mean of the absolute value of the elementary effects. The elementary effect, $d_i(\mathbf{x})$, is computed for r grid points and the average of the effect for each input is

$$\mu_i^* = \frac{1}{r} \sum_{j=1}^{r} \mid d_{ij}(\mathbf{x}) \mid \tag{2.35}$$

and

$$\sigma_i^2 = \frac{1}{r-1} \sum_{j=1}^{r} \big(d_{ij}(\mathbf{x}) - \mu\big)^2. \tag{2.36}$$

Example 2.1. The Morris method will be used to screen the inputs in the flood model. The Morris' sensitivity measures are represented in a graphic of μ^* and σ as axes as follows. Based on Figure 2.1, the input X_4 has the biggest mean indicating it is the most important inputs. Following X_4, there are X_1, X_2, X_6 and X_3. While X_5 and X_8 are negligible as their means are around zero.

2.2.10 Active Subspace method

A slightly different sensitivity measure is developed by Constantine & Doostan (2017) and is named activity scores. The key idea is to utilise the active subspaces, which are used to determine critical directions in the parameter space of an input. Let us consider an input vector $\mathbf{x} = (x_1, \ldots, x_p)^T$ where $\mathbf{x} \in [-1, 1]^p$, the function of the inputs is denoted by $f(\mathbf{x})$ which is differentiable and square-integrable, and the density of $\mathbf{x}$ is uniform, $\rho(\mathbf{x}) = 2^{-p}$

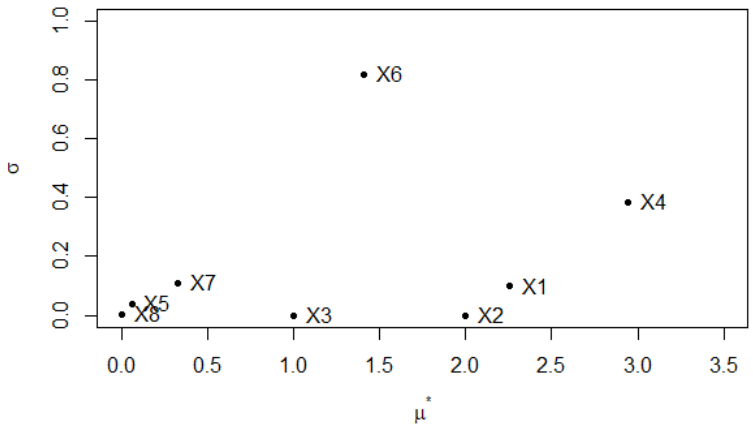


Figure 2.1: Screening of inputs in the flood model using the Morris method. The inputs are located in the μ^* and σ axis.

for $\mathbf{x} \in [-1, 1]^p$. The active subspace is represented by an eigenvector of the $p \times p$ matrix below

$$\mathbf{C} = \int \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T \rho(\mathbf{x}) d\mathbf{x} = \mathbf{W} \Lambda \mathbf{W}^T \tag{2.37}$$

where $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_p]$ is the eigenvectors matrix and $\Lambda = \text{diag}\lambda_1, \ldots, \lambda_p$ is the diagonal eigenvalues. The eigenpairs of $\mathbf{C}$ is defined by

$$\lambda_i = \mathbf{w}_i^T \mathbf{C} \mathbf{w}_i, \tag{2.38}$$

and it is similar to the total sensitivity indices. The matrix $\mathbf{C}$ in Equation 2.37 is estimated using Monte Carlo method. For M samples obtained from an input space, the estimate of $\mathbf{C}$ is given as

$$\hat{\mathbf{C}} = \frac{1}{M} \sum_{i=1}^{M} \nabla f(\mathbf{x}_i) \nabla f(\mathbf{x}_i)^T. \tag{2.39}$$

Example 2.2. The active subspace method is implemented in the Flood model. We have simplified the equation in the flood model as

$$y = f(\mathbf{x}) = x_1 - x_2 - x_3 + x_4^{0.6} \Big(x_5^{-0.6} \cdot x_6^{-0.6} \cdot (x_7 - x_1)^{-0.3} \cdot x_8^{0.3} \Big) \tag{2.40}$$

The derivative of function $f(x_i)$, $\nabla f(x_i)$, is the derivation of the function $f(\mathbf{x})$ that is evaluated at x_i, i.e.

$$\nabla f(\mathbf{x}_i) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \\ \frac{\partial f}{\partial x_4} \\ \frac{\partial f}{\partial x_5} \\ \frac{\partial f}{\partial x_6} \\ \frac{\partial f}{\partial x_7} \\ \frac{\partial f}{\partial x_8} \end{bmatrix} = \begin{bmatrix} 1 + 0.3(x_7 - x_1)^{-1.3} \cdot x_4^{0.6} \cdot x_5^{-0.6} \cdot x_8^{0.3} \\ -1 \\ -1 \\ 0.6x_4^{-0.4} \cdot x_5^{-0.6} \cdot x_6^{-0.6} \cdot (x_7 - x_1)^{-0.3} \cdot x_8^{0.3} \\ -0.6x_5^{-1.6} \cdot x_4^{-0.6} \cdot x_6^{-0.6} \cdot (x_7 - x_1)^{-0.3} \cdot x_8^{0.3} \\ -0.6x_6^{-1.6} \cdot x_4^{0.6} \cdot x_5^{-0.6} \cdot (x_7 - x_1)^{-0.3} \cdot x_8^{0.3} \\ -0.3(x_7 - x_1)^{-1.3} \cdot x_4^{0.6} \cdot x_5^{-0.6} \cdot x_6^{-0.6} \cdot x_8^{0.3} \\ 0.3x_8^{-0.7} \cdot x_4^{0.6} \cdot x_5^{-0.6} \cdot x_6^{-0.6} \cdot (x_7 - x_1)^{-0.3} \end{bmatrix}$$

The derivative matrix $\nabla f(\mathbf{x}_i)$ is evaluated at $\mathbf{x} = \mathbf{x}_i$, where $\mathbf{x}_i$ is obtained by sampling the value of input from their sample spaces. In this example, the sampling size is $M = 1000$ and the estimated matrix $\hat{\mathbf{C}}$ is

$$\hat{\mathbf{C}} = \begin{bmatrix} 4.877 & -2.180 & -2.180 & 0.003 & 0 & -0.115 & -0.352 & 0 \\ -2.180 & 1.000 & 1.000 & -0.001 & 0 & 0.051 & 0.154 & 0 \\ -2.180 & 1.000 & 1.000 & -0.001 & 0 & 0.051 & 0.154 & 0 \\ 0.003 & -0.001 & -0.001 & 0.000 & 0 & 0.000 & 0.000 & 0 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0 & 0.000 & 0.000 & 0 \\ -0.115 & 0.051 & 0.051 & 0.000 & 0 & 0.003 & 0.008 & 0 \\ -0.352 & 0.154 & 0.154 & 0.000 & 0 & 0.008 & 0.026 & 0 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0 & 0.000 & 0.000 & 0 \end{bmatrix}$$

Next, we compute the eigenvalues and the corresponding eigenvector matrix from the matrix $\hat{\mathbf{C}}$. We obtain $\mathbf{\Lambda}$ as the diagonal matrix of the eigenvalues in descending order.

$$\mathbf{\Lambda} = \begin{bmatrix} 6.96053 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0 & 0.03623 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.00048 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.00001 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.00023 \end{bmatrix}$$

and the orthogonal matrix $\mathbf{W}$ as the eigenvector

$$\mathbf{W} = \begin{bmatrix} 0.844 & -0.519 & 0.106 & 0 & 0 & 0.000 & -0.003 & 0.084 \\ -0.377 & -0.594 & 0.048 & 0 & 0 & 0.707 & -0.008 & 0.049 \\ -0.377 & -0.594 & 0.048 & 0 & 0 & -0.707 & -0.008 & 0.049 \\ 0.000 & -0.010 & 0.467 & 0 & 0 & 0.000 & -0.567 & -0.678 \\ 0.000 & 0.000 & 0.000 & 0 & -1 & 0.000 & 0.000 & 0.000 \\ -0.020 & 0.019 & 0.519 & 0 & 0 & 0.000 & 0.796 & -0.309 \\ -0.061 & 0.152 & 0.704 & 0 & 0 & 0.000 & -0.209 & 0.658 \\ 0.000 & 0.000 & 0.000 & 1 & 0 & 0.000 & 0.000 & 0.000 \end{bmatrix}$$

According to the eigenvalue $\mathbf{\Lambda}$, we can partition it into two pairs based on the cut-off of $\lambda_i > 0$ and $\lambda_i \leq 0$ in three decimal places.

$$\mathbf{\Lambda}_1 = \begin{bmatrix} 6.96053 & 0 \\ -0 & 0.03623 \end{bmatrix}; \mathbf{\Lambda}_2 = \begin{bmatrix} 0.00048 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.00001 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.00023 \end{bmatrix}$$

and the corresponding partition of matrix $\mathbf{W}$ are

$$\mathbf{W}_1 = \begin{bmatrix} 0.844 & -0.519 \\ -0.377 & -0.594 \\ -0.377 & -0.594 \\ 0.000 & -0.010 \\ 0.000 & 0.000 \\ -0.020 & 0.019 \\ -0.061 & 0.152 \\ 0.000 & 0.000 \end{bmatrix}; \mathbf{W}_2 = \begin{bmatrix} 0.106 & 0 & 0 & 0.000 & -0.003 & 0.084 \\ 0.048 & 0 & 0 & 0.707 & -0.008 & 0.049 \\ 0.048 & 0 & 0 & -0.707 & -0.008 & 0.049 \\ 0.467 & 0 & 0 & 0.000 & -0.567 & -0.678 \\ 0.000 & 0 & -1 & 0.000 & 0.000 & 0.000 \\ 0.519 & 0 & 0 & 0.000 & 0.796 & -0.309 \\ 0.704 & 0 & 0 & 0.000 & -0.209 & 0.658 \\ 0.000 & 1 & 0 & 0.000 & 0.000 & 0.000 \end{bmatrix}$$

Matrices $\mathbf{\Lambda}_1$ and $\mathbf{W}_1$ are called the first eigenvalues and the first eigenvectors, respectively. The active subspace is identified from $\mathbf{W}_1$, which inform us that input x_1, x_2, and x_3 are important inputs in the Flood model.

2.2.11 The PAWN method

Pianosi & Wagener (2015) introduced an efficient method to compute the sensitivity indices obtained using the density-based approach. The method uses the cumulative distribution function of the model's output instead of the probability density function as it is easier to derive. The definition of the PAWN method is presented below.

Let us consider a model $y = f(\mathbf{x})$ where $\mathbf{x} = (x_1, \ldots, x_p)$ where p is the number of input in the model. The PAWN sensitivity index is defined as

$$T_i = \underset{x_i}{\text{stat}}[KS(x_i)] \tag{2.41}$$

where

$$KS(x_i) = \underset{y}{max} \mid F_y(y) - F_{y|x_i}(y) \mid \tag{2.42}$$

and *stat* is a statistic that can be the maximum or the median value of x_i. The PAWN method is practical to compute when we have find out the cumulative distribution for unconditional Y and the conditional $Y|X_i$. The interpretation of the index is also practical as it does not depend on the measurement unit of y.

Based on sensitivity measure using The PAWN method in Figure 2.2, the most important input is X_4, the second important one is X_2 and followed by X_1. The least important inputs are X_5 and X_8.

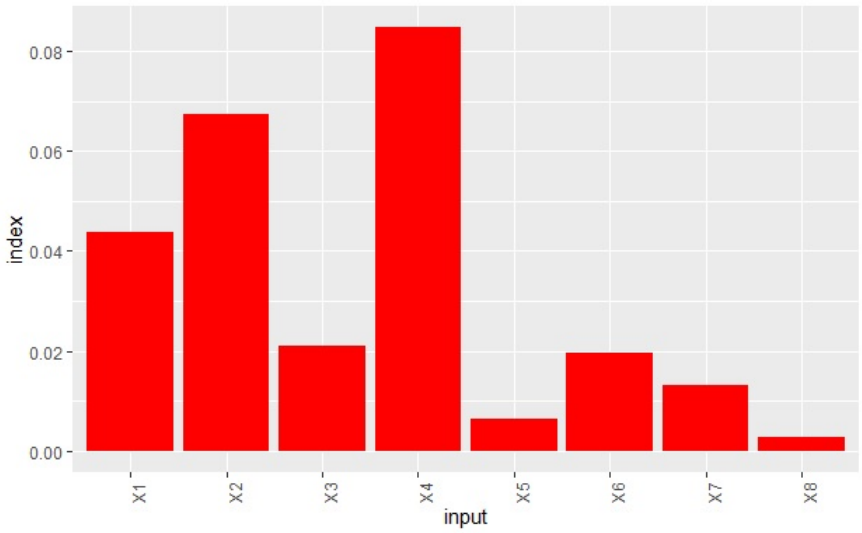


Figure 2.2: The PAWN sensitivity indices (T_i) for the Flood model.

2.2.12 General principle of Global Sensitivity Analysis

We have presented some sensitivity measures which mostly belong to the global sensitivity analysis scope and computed the sensitivity indices. Some measures such as variance-based methods, EVPI, and the PAWN method seem to have a similar principle. In the variance-based method, the sensitivity index is computed using the difference between the unconditional variance ($Var(Y)$) and the expected conditional variance ($Var(Y|X_i)$). EVPI computes the difference between the expected of the maximum of conditional expectation ($\max_j(Y_j)|X_i)$) and the maximum of expected unconditional expectation ($\max_j(E(Y_j(\mathbf{X}))$). While in the PAWN method, the index is computed using the difference between the unconditional cumulative distribution function ($F_Y(y)$) and the conditional cumulative distribution function ($F_{Y|X_i}(y)$). The latter uses the cumulative distribution function method which will be discussed in Chapter 3.

We will also investigate other methods of sensitivity analysis in Chapter 3 which is a distribution-based method, such as the quantile-based method (Chun et al. (2000)) and the density-based method (Borgonovo (2007)). These methods have also a common rationale as the previous ones which take into account the differences between two quantities. The quantile-based method considers the difference between the quantile of the base case and the sensitivity case. The base case is the quantile of the unconditional distribution of output Y, while the sensitivity case is the quantile of the conditional distribution $Y|X_i$.

On the other hand, the density-based method calculates the difference between the unconditional density function ($f_Y(y)$) and the conditional density function ($f_{Y|X_i}(y)$).

These common rationales for the GSA have been formalised by Borgonovo et al. (2016). Generally, the global sensitivity analysis measures the discrepancy between the unconditional probability $\mathbb{P}_Y(y)$ and the conditional probability $\mathbb{P}_{Y|X_i}(y)$. The discrepancy between the two probabilities is defined by an inner statistic $\gamma_i(x_i) = \zeta(\mathbb{P}_Y(y), \mathbb{P}_{Y|X_i}(y))$, where $\zeta(\cdot,\cdot)$ is an inner operator which fulfill the condition of $\zeta(\mathbb{P},\mathbb{P}) = 0$ for all distribution of $\mathbb{P}$. Using this definition, the global sensitivity analysis measures for input X_i is formulated as $\mathbb{E}\{\gamma_i(X_i)\}$.

The next section provides some methods for computing the sensitivity indices which are based on the Monte Carlo sampling algorithm as it uses the given-data approach.

2.2.13 Computational Methods

We consider three types of computational methods used in global sensitivity analysis literature and apply the methods for computing sensitivity indices for the flood model. The methods are Pick and freeze method (Sobol (2001)), Fourier Analysis Sensitivity Test (Cukier et al. (1977)), and The Generalized Additive Model (Hastie & Tibshirani (1990)).

Pick and freeze method

A computation method to compute the global sensitivity indices for moderate complex model based on Monte Carlo algorithm was proposed by Sobol (2001). This method considers a set of inputs which are split into two subsets. We can select which inputs as the first subset and which inputs belong to another subset and then compute the sensitivity index for the subset of interest. Let us suppose that $\mathbf{x} = (x_1, \ldots, x_p)$ is a vector of p dimensional input in a mathematical model $y = f(\mathbf{x})$. We consider an arbitrary m inputs as a subset of $\mathbf{x}$, where $m < p$ and denotes the subset as $\mathbf{u} = (x_{j_1}, \ldots, x_{j_m})$ where $1 \leq j_1 \leq \ldots \leq j_m \leq p$. We define the complementary subset of $\mathbf{v}$ which has $p - m$ inputs so that $\mathbf{x} = (\mathbf{u}, \mathbf{v})$.

The main effect index and the total effect index for the subset $\mathbf{u}$ are

$$S_{\mathbf{u}} = \frac{V_{\mathbf{u}}}{V} \tag{2.43}$$

and

$$S_{\mathbf{u}}^{tot} = \frac{V_{\mathbf{u}}^{tot}}{V} \tag{2.44}$$

respectively, where $S_{\mathbf{u}}^{tot} = 1 - S_{\mathbf{v}}$. According to Sobol (2001), the variance of subset $\mathbf{u}$, $V_{\mathbf{u}}$, is calculated by the following integration

$$V_{\mathbf{u}} = \int f(\mathbf{x}) f(\mathbf{u}, \mathbf{v}^{'}) d\mathbf{x} d\mathbf{v}^{'} - f_0^2 \tag{2.45}$$

where $f_0 = \int f(x)dx$ and the variance for subset $\mathbf{v}$, $V_{\mathbf{v}}$, is obtained using the similar fashion as Equation (2.45) as follows.

$$V_{\mathbf{v}} = \int f(\mathbf{x}) f(\mathbf{u}^{'}, \mathbf{v}) d\mathbf{x} d\mathbf{u}^{'} - f_0^2 \tag{2.46}$$

A Monte Carlo method is employed to solve the integration parts in Equation (2.45) and (2.46). The Monte Carlo estimates are computed using N iterations using the following formulae

$$f_0 \approx \frac{1}{N} \sum_{i=1}^{N} f(\tau_i) \tag{2.47}$$

$$V + f_0^2 \approx \frac{1}{N} \sum_{i=1}^{N} f^2(\tau_i) \tag{2.48}$$

$$V_{\mathbf{u}} + f_0^2 \approx \frac{1}{N} \sum_{i=1}^{N} f^2(\tau_i) f(\tau_{1i}, \tau_{2i}^{'}) \tag{2.49}$$

$$V_{\mathbf{v}} + f_0^2 \approx \frac{1}{N} \sum_{i=1}^{N} f^2(\tau_i) f(\tau_{1i}^{'}, \tau_{2i}) \tag{2.50}$$

In this case, two independent random points τ and $\tau^{'}$ are considered in the computation. In addition, each random point has its own subsets, $\tau = (\tau_1, \tau_2)$ and $\tau^{'} = (\tau_1^{'}, \tau_2^{'})$. The application of the pick and freeze method to the flood model produces the main effect indices for each input as shown in Figure 2.3. The Figure shows that input X_4 is the most important input, X_2 and X_3 are the second and the third most important inputs respectively.

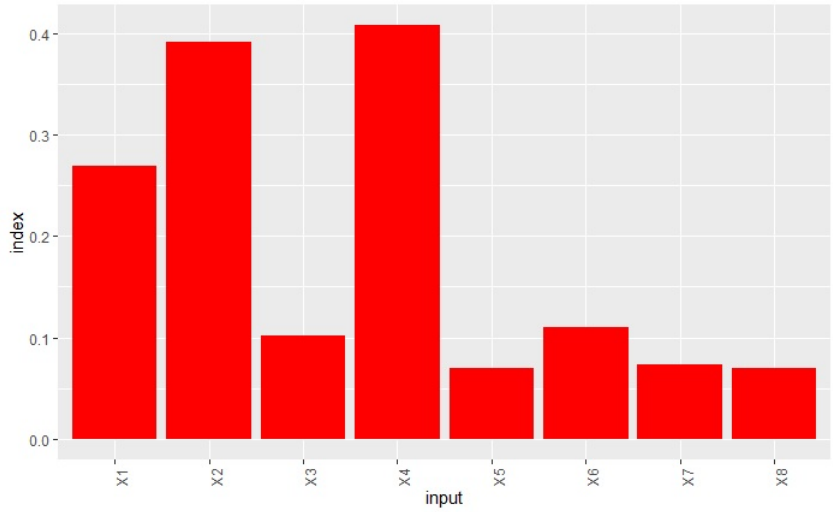


Figure 2.3: Main effect indices for inputs in the Flood Model obtained using the Pick and Freeze method

Fourier Analysis Sensitivity Test (FAST)

Cukier et al. (1977) introduced a method that relates the probability distribution for each input in the model to a frequency. This approach is called the Fourier Amplitude Sensitivity Test (FAST) method. In this method, the parameters k_l is defined in the range of $[-\infty, \infty]$. The parameter k_l is a function of u_l and is written as

$$k_l = g_l(u_l), \quad -\infty < u_l < \infty, \quad l = 1, 2, ..., n, \tag{2.51}$$

where g is a function. Equation 2.51 determines the range of parameter k_l is affected by variable u_l. Furthermore, variable u_l is assumed to be independent with their respective probability distribution $P_l(u_l)$. The probability density of $\mathbf{u}$ is expressed as

$$P(\mathbf{u}) = \prod_{l=1}^{n} P_l(u_l) \tag{2.52}$$

The average over the density is introduced by Cukier et al. (1977) using the following relation.

$$< f(\mathbf{u}) > \equiv \int d\mathbf{u} f(\mathbf{u}) P(\mathbf{u}) \tag{2.53}$$

Cukier et al. (1977) constructs a search curve in the $\mathbf{u}$-space which is parameterised by a search variable s. The translation of variable u_l into s is performed by a transformation

function G_l so that

$$u_l = G_l(\sin \omega_l s). \tag{2.54}$$

where ω_l are a set of frequencies which each frequency is assigned to each variable u_l. FAST is appropriate when the number of input variables is less than 10. Otherwise, it becomes costly, unstable, and biased. Saltelli et al. (1999) made an extension of FAST which is called the extended FAST (eFAST). To gain a more flexible sampling, they modified the transformation of nominal value inputs x_i from frequencies ω_i. This method can be used to estimate both first-order indices and total effect indices. An example of the eFAST method

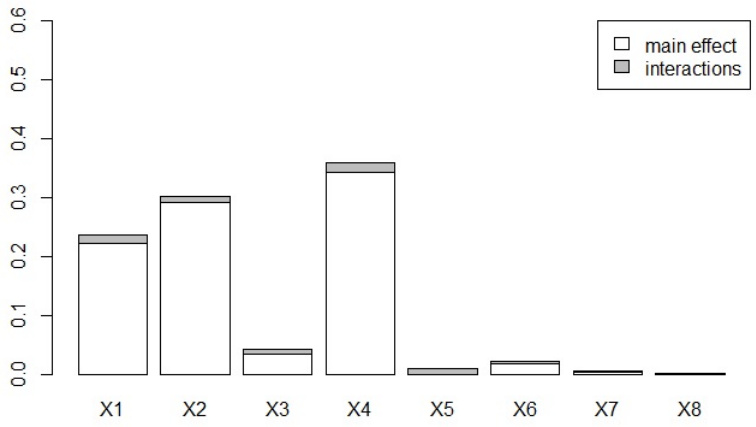


Figure 2.4: Main effects and interaction effects of the Flood Model computed using extended FAST method

is implemented to the Flood model, and the result is given in Figure 2.4. According to the figure, X_4 has the largest main effect indicating it is the most important input, X_2 is the second-largest and X_3 has the third-largest main effects. For most of inputs, the interaction effects are small and give a tiny contribution to the total effects.

Generalised Additive Models (GAM)

According to Hastie & Tibshirani (1990), Generalised Additive Models (GAMs) were introduced to overcome the non-linearity relationships between inputs and output variables. For some cases, it can also be used to solve the problem of modelling when the model specification is hard to determine. This circumstance might happen when the data

points obtained from the measurements have no specific pattern matching with any predetermined literature models. Therefore, it is hard to use any existing model such as linear regression, polynomial regression or generalised linear models.

Let us suppose that we have a model $y = f(x_1, \ldots, x_p)$ and the inputs are uncertain and denoted as $\mathbf{X} = (X_1, \ldots, X_p)$. We are interested in the model of output uncertainty $Y = f(\mathbf{X})$ with f is a function that connects input $\mathbf{X}$ to the output Y. The idea of GAMs is to approach each inputs in the model $Y = f(\mathbf{X})$ by a function and make the whole function in an additive form. Therefore the model $Y = f(\mathbf{X})$ can be represented in GAMs model in the following form

$$\mathbb{E}\left(Y|X_1, X_2, ..., X_k\right) = f_0 + f_1(X_1) + f_2(X_2) + ... + f_k(X_k) \tag{2.55}$$

where $f_i's$ are unspecified smooth(non-parametric) functions. In order to fit the functions $f_i's$, cubic smoothing splines are frequently used. The principle of this smoothing spline is to find one function among all functions $f(x_i)$ that minimize the penalized least square below

$$\sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2 + \lambda \int_a^b \left[f^{"}(x)\right]^2 dx \tag{2.56}$$

where λ is a non-negative smoothing parameter and $a \leq x_1 \leq ... \leq x_n \leq b$. In addition, (a, b) is a possible range of x_i's.

The next step is to estimate the additive models. A general algorithm to fit these models is called the back-fitting algorithm. In order to undertake the estimation using this algorithm, firstly, we need to define the partial residual

$$R_j = Y - f_0 - \sum_{k \neq j} f_k(X_k) \tag{2.57}$$

then $\mathbb{E}(R_j|X_j) = f_j(X_j)$. This conditional expectation produce a path to estimate each smoothing function $\hat{f}_j$.

Furthermore, suppose that there are set of observation (x_i, y_i), then the penalized least square in (2.56) has turn into

$$\sum_{j=1}^{p} \left(y_j - f_0 - f_j(x_j)\right)^2 + \sum_{j=1}^{p} \lambda_j \int \left[f_j^{"}(t_j)\right]^2 dt_j \tag{2.58}$$

Based on that penalized least square function, the procedure of the back-fitting algorithm are

1 Initialize:

$$\hat{f}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i, \hat{f_1^j} = 0, m = 1$$

2 Iterate:

$$m \Leftarrow m + 1, j = 1, 2, ..., p$$

$$f_j^m(x_j) \leftarrow S_j \left[\left\{ y_i - \hat{f}_0 - \sum_{k=1}^{j-1} \hat{f}_k^m(x_k) - \sum_{k=j+1}^{p} \hat{f}_k^{m-1}(x_k) \right\}_1^n \right] \tag{2.59}$$

3 Repeat until:

$$RSS = S\left(Y - \hat{f}_0 - \sum_{k=1}^{p} \hat{f}_k^m(x_k) \right)^2$$

fails to decrease, which means $\hat{f^j}$ is tends to have a stable form.

$f_j^m(.)$ denotes the estimate of $f_j(.)$ at the m^{th} iteration. Each of the functions f_j is a cubic spline in the component x_j, with knots at each of unique value of x_{ij}, $j = 1, 2, ..., p$.

A convergence criterion for the back-fitting algorithm is as following

$$\frac{\sum_{j=1}^{p}\sum_{i=1}^{n}(f_j^{m-1}(x_{ij}) - f_j^m(x_{ij})^2}{\sum_{j=1}^{p}\sum_{i=1}^{n} f_j^{m-1}(x_{ij})^2} \leq \epsilon \tag{2.60}$$

where ϵ is a very small number which usually less than 10^{-6}.

The example of computing the sensitivity indices using GAM method for The Flood Model will be discussed in Chapter 6.

2.2.14 Meta-modelling Based Methods

Some different ways aim at minimising cost to estimate global sensitivity indices. These approaches are called meta-models that use N model runs to fit a meta-model (an emulator). The emulator representing the relation between input and output based on a sample will be used as a surrogate model.

Gaussian Process meta-modelling

A Bayesian approach for estimating the sensitivity indices were developed by Oakley & O'Hagan (2004). In this approach, the main model is treated as an unknown function. The

prior model is a Gaussian process distribution, and the posterior distribution is obtained. Subsequently, the posterior mean of main effects or interaction effects will be determined. Let suppose the uncertainty of the output Y is the function of the uncertain input $\mathbf{X}$ such that $Y = f(\mathbf{X})$. The main effect and the interaction effects are derived from variance decomposition in Equation (2.18). The posterior mean of the main effect is

$$E^*\{f_i(x_i)\} = \{R_i(x_i) - R\}\hat{\beta} + \{T_i(x_i) - T\}\mathbf{e} \tag{2.61}$$

and the posterior mean for the interaction effect is

$$E^*\{f_{i,j}(\mathbf{x}_{i,j})\} = \{R_{i,j}(\mathbf{x}_{i,j}) - R_i(x_i) - R_j(x_j) - R\}\hat{\beta} + \{T_{i,j}(\mathbf{x}_{i,j}) - T_i(x_i) - T_j(x_j) - T\}\mathbf{e}. \tag{2.62}$$

where

$$\mathbf{e} = \boldsymbol{A}^{-1}(\mathbf{y} - \boldsymbol{H}\hat{\boldsymbol{\beta}})$$

$$\hat{\boldsymbol{\beta}} = \boldsymbol{V}^*(\boldsymbol{V}^{-1}\mathbf{z} + \boldsymbol{H}^T\boldsymbol{A}^{-1}\mathbf{y})$$

$$\boldsymbol{V}^* = (\boldsymbol{V}^{-1} + \boldsymbol{H}^T\boldsymbol{A}^{-1}\boldsymbol{H})^{-1}$$

$$\boldsymbol{H}^T = (\boldsymbol{h}(\mathbf{x}_1)^T, \ldots, \boldsymbol{h}(\mathbf{x}_n)^T)$$

$$\boldsymbol{A} = \begin{pmatrix} 1 & c(\mathbf{x}_1, \mathbf{x}_2) & \cdots & c(\mathbf{x}_1, \mathbf{x}_n) \\ c(\mathbf{x}_2, \mathbf{x}_1) & 1 & & \vdots \\ \vdots & & \ddots & \\ c(\mathbf{x}_n, \mathbf{x}_1) & \cdots & & 1 \end{pmatrix}$$

The notation $\mathbf{y}$ denotes the output of a model, $c(\mathbf{x}, \mathbf{x}')$ is a function which decreases as $|\mathbf{x} - \mathbf{x}'|$ increases and satisfies $c(\mathbf{x}, \mathbf{x}) = 1$ for all $\mathbf{x}$. In addition the other terms, $R_i(x_i)$ and $T_i(x_i)$ can be obtained using the following formulae

$$R_i(\mathbf{x}_i) = \int_{\mathcal{X}_{-i}} \mathbf{h}(\mathbf{x})^T dG_{-i|i}(\mathbf{x}_{-i}|\mathbf{x}_i) \tag{2.63}$$

$$T_i(\mathbf{x}_i) = \int_{\mathcal{X}_{-i}} \mathbf{t}(\mathbf{x})^T dG_{-i|i}(\mathbf{x}_{-i}|\mathbf{x}_i) \tag{2.64}$$

where $\mathbf{h(x)}$ is a vector which consist of regression functions of $\mathbf{x}$ and

$$t(\mathbf{x})^T = (c(\mathbf{x}, \mathbf{x}_1), \ldots, c(\mathbf{x}, \mathbf{x}_n)),$$

while R and T are special cases of $R_i(x_i)$ and $T_i(x_i)$ respectively when i is the null set.

Another work in this method was conducted by Srivastava et al. (2017). They made a comparison between the model which uses Gaussian Process from Kennedy & O'Hagan (2001) which is called Bayesian Hybrid Model, and with the analytical approach. The analytical approach calculates the sensitivity indices by Sobol's formulae. The comparison was also considered about the correlated inputs in the model.

The Bayesian model averaging framework was used by Shao et al. (2017) in determining plausible competing sparse PCE models to estimate the posterior probability for each model to be the best one among the competing models. A meta-model with correlated inputs was proposed by Gauchi et al. (2017). To build the meta model, they employed a multivariate Legendre truncated Polynomial Chaos Expansion (PCE). Furthermore, they used Partial Least Square Regression (PLSR) to estimate the regression coefficients of the PCE.

State Dependent Parameter meta-modelling

Ratto et al. (2007) proposed a method to estimate the terms in functional decomposition in (2.18). They used a recursive fixed interval smoothing algorithm to estimate the parameters in a State-Dependent Parameter (SDP) formulation. The formulation describes the relationship between input and output. They considered the first order of terms in Equation (2.18) and simplified it as

$$Y_t - f_0 = f_1(X_{1,t}) + f_2(X_{2,t}) + \ldots + f_k(X_{k,t}) + o(\mathbf{X}\mathbf{X}^T) \tag{2.65}$$

The model in Equation (2.65) is regarded as State Dependent Regression (SDR) and can be reconstruct as

$$Y_t - f_0 = \mathbf{X}_t^T \mathbf{p}_t + e_t \tag{2.66}$$

where $\mathbf{p}_t = p_{1,t}, \ldots, p_{2,t}$ are called the state dependent parameters.

We have discussed various methods in sensitivity analysis and provided examples. Next, we discuss sensitivity analysis status and how sensitivity can be developed to support recent issues.

§ 2.3 The Future of Sensitivity Analysis

Razavi et al. (2021) provide the state-of-the-art of SA initially started with Local Sensitivity Analysis and has developed into a Global Sensitivity Analysis (GSA). The categories of GSA described in the previous section are the regression-based approach and the variance-based approach. Another sensitivity measure will be discussed later, which is the distribution-based approach. A sensitivity measure that comes later is the variogram-based approach (VARS) proposed by Razavi & Gupta (2016a). VARS involves using a variogram function that computes the difference of response surface values of two points within the factor space. VARS has some advantages such as computationally efficient, robust, and stable estimates even using a small number of model runs. Some progression indicates the development of computation in sensitivity analysis in computing the sensitivity measure when SA has a connection with DACE. In particular, the sensitivity measure is approximated using response surface when the models are computationally expensive. The recent research of SA has also considered the correlated inputs in the mathematical model. Many packages or functions have been developed to assist the SA computation based on various computer programs such as R ('sensitivity' package, SAFE), Phyton (SobolGSA, SALib), MATLAB (UQLab, VARS-Tool), C (PSUADE), C++ (Dakota, OpenTURNS), and Julia (MADS.jl).

According to Razavi et al. (2021), there are six main themes of challenges and outlooks in SA: (1) SA should be structured and standardised as a discipline, (2)Introduce SA for system modelling, (3) overcome the computational challenges of SA, (4) Connect SA for supporting machine learning, (5) clearing up the connection between SA and what role can SA provide for uncertainty quantification, and (6) optimising the use of SA for decision making.

The first outlook is that SA should be structured and standardised as a discipline. The challenge to this outlook is that the terminology, fundamental definitions, and methodology is not consistent. This challenge needs to be tackled to create a standard perception and understanding. There is an opportunity for teaching SA as a discipline more broadly and consistently in the future because standardising SA is visible.

The second theme is that SA has the potential to be implemented in mathematical modelling for management uncertainty, model reduction, model verification and diagnostic testing and non-identifiability. SA can map the uncertainty in every model assumption in managing uncertainty and help answer a critical question: when and how does the uncertainty matter? In diagnostic and model verification, SA can diagnose a mathematical

model behaviour and assess how the model mimics the investigated system. For example, the current development of the given-data SA method will enable it to perform model diagnostic testing. SA can also identify a model component due to over parameterisation, which makes it non-identifiable.

The third challenge for SA is how to overcome the potential burden when performing the analysis for complex mathematical models. Mathematical models can have lots of factors or parameters which will increase the computation time and require more computing power. Therefore, SA needs a new generation of algorithms to handle the progression of model complexity. Furthermore, the SA algorithm must also be efficient and can reach convergence. In addition, the algorithm should be reliable and robust. Efficiency means the time required for the algorithm to undertake the computation of SA. The convergence of an SA algorithm is not easy to assess. It depends on some aspects such as the complexity of a model, the aim of performing SA, SA methods, and the convergence criteria. Reliability means the correctness of SA results indicated by a particular measure. Robustness refers to the consistency measure of an algorithm to produce a similar output when the parameters of a model are changed.

The fourth outlook is the possibility of implementing SA in machine learning. SA can offer the development and application of machine learning as they have a similar way of addressing the problem but with different approaches. Machine learning aims to construct a function that relates input space to those in an output space using a data-driven approach. On the other hand, SA investigates the relationship between inputs and outputs without developing a relationship function. Instead, it estimates the magnitude of the relationship using sensitivity indices. Nevertheless, the difference between SA and machine learning should also need to be considered. The major difference is in the source of data. Data for SA is obtained from computer experiments, while data for machine learning comes from more general experiments, such as laboratory or field experiments.

The fifth is to clarify the role of SA in uncertainty quantification. By definition, uncertainty quantification is the science of quantitative characterisation with aims to reduce the uncertainty regarding a particular output of a model. Even though SA has been used in uncertainty quantification, there are still some challenges. Some challenges which need to be considered are as follows. (1) potential misconception in framing an SA for an uncertainty quantification goal, (2) some of SA frameworks cannot be used for some model types, (3) It is complicated to use SA when inputs are correlated or in the case of multivariate inputs (4) sensitivity of uncertainty quantification to problem setup, and (5) uncertainty could also be investigated in the SA results. Razavi et al. (2021) provides

some explanations and suggestions to address the challenges.

The final review is the role of SA in decision making. SA has widely used in the decision-making process, such as in shadow prices concept and scenario analysis. Recently SA has been used to help in making a decision under uncertainty. In this case, SA can parse the outcome uncertainty of a decision option and assign it to different sources of uncertainty. There are challenges and opportunities in developing SA to make robust decisions under deep uncertainty. In this case, stakeholders do not know or disagree about the probability distributions of inputs in the model, so the best way to perturb the inputs can be developed. SA in the context of qualitative decision making has also been developed. The framework for sensitivity auditing is proposed to frame a decision analysis with the involvement of researchers, stakeholders and policymakers.

In conclusion, SA currently has many roles in various fields and research topics. Also, SA has lots of opportunities to get involved more in supporting other research areas. However, some existing challenges need to be handled to enable SA to play a more critical role in advancing scientific methods and model complexity.

§ 2.4 Summary

Analysing complex computer models involving optimisation, prediction, and calibration is challenging for a modeller, especially when the model has many inputs. We propose learning and investigating inputs utilising sensitivity analysis to detect inputs that should be considered or not in the model for those analyses. We consider some important measures in the sensitivity analysis using different methods such as regression-based methods, variance-based methods, and the value of the information-based method to determine the prioritisation of inputs from the most important input to the least important one. Based on the order of importance, we can only focus on the important input to perform further analysis.

The next chapter explores some limitations of the factor prioritisation setting, especially the main effect index, and proposes another approach based on the factor fixing setting to learn and investigate inputs in the computer model.

Chapter 3

Identifying Unimportant Inputs

§ 3.1 Introduction

In the previous chapter, we considered the factor prioritisation setting to identify unimportant inputs in a mathematical model. In particular, the main effect index is used to measure the expected reduction in variance when input is fixed on its true value. A relatively large main effect index implies that there is a value in learning an input. In other words, the input is important. Conversely, a small main effect does not imply that input is unimportant especially when the model has interaction terms. The main effect is not able to measure the joint effect of the input of interest with the other inputs. Consequently, a zero main effect does not imply that input is unimportant as interaction terms might not be zero.

As a result, identifying unimportant inputs using the main effect is not obvious. Furthermore, in factor prioritisation, 'choosing an input to learn' is a clearly defined action, but it is less clear what the consequences would be of declaring an input to be 'unimportant'. Therefore, two questions arise.

1 How can we define an input as unimportant with regard to actions that a modeller might take?

2 What are appropriate sensitivity measures to identify unimportant inputs?

This chapter aims to answer these questions. This chapter is organised in the following way. Section 3.2 describes examples of main effect index limitations, Section 3.3 proposes

two scenarios for identifying unimportant inputs, Section 3.4 discusses some existing importance measures with independent and dependent input cases, and Section 3.5 presents the conclusions.

§ 3.2 Main effect index limitations

As mentioned in the previous section, the main effect index does not provide a full picture of importance. This is because a large value of a main effect index indicates an important input but a small value of main effect index does not imply that an input is unimportant. The following are two different examples to show the limitations of main effect indices.

Example 3.1. Let us consider a function $f : \mathbb{R}^4 \to \mathbb{R}$ with

$$f(x_1, \ldots, x_4) = x_1 x_2 + x_3 + 0.01 x_4. \tag{3.1}$$

Define $Y = f(X_1, \ldots, X_4)$ with $X_1, \ldots, X_4$ identical, independent and normally distributed random variables with mean 0 and variance 1. The (unnormalised) main effect index V_1 of X_1 is calculated as follows.

$$\begin{aligned} V_1 &= Var_{X_1}[E(Y|X_1)] \\ &= Var_{X_1}[E(X_1X_2 + X_3 + 0.01X_4|X_1)] \\ &= Var_{X_1}[X_1E(X_2) + E(X_3) + 0.01E(X_4)] \\ &= 0. \end{aligned}$$

The main effect indices of other inputs are obtained using the similar way and the results are $V_2 = 0$, $V_3 = 1$, and $V_4 = 0.0001$. The interaction effect for X_1 and X_2 is calculated as below.

$$\begin{aligned} V_{12} &= Var_{X_1,X_2}[E(Y|X_1, X_2)] - V_1 - V_2 \\ &= Var_{X_1,X_2}[E(X_1X_2 + X_3 + 0.01X_4|X_1, X_2)] \\ &= Var_{X_1,X_2}[X_1X_2 + E(X_3) + 0.01E(X_4)] \\ &= Var_{X_1,X_2}(X_1X_2) \\ &= E[(X_1X_2)^2] - [E(X_1X_2)]^2 \\ &= E(X_1^2)E(X_2^2) - E(X_1)^2E(X_2)^2 \quad (\text{since } X_1 \text{ and } X_2 \text{ are independent}) \\ &= \left[Var(X_1) + E(X_1)^2\right]\left[Var(X_2) + E(X_2)^2\right] - E(X_1)^2E(X_2)^2 \\ &= 1. \end{aligned}$$

The main effect indices of X_1 and X_2 are zero but we could not conclude the inputs as unimportant because the interaction effect is not zero. The main effect indices of X_3 and X_4 do, however, better indicate these inputs' respective importance.

Example 3.2. Let us consider another function $f : \mathbb{R}^4 \to \mathbb{R}$, with

$$f(x_1, \ldots, x_4) = x_1 + x_2 + x_3 + 0.01x_4, \tag{3.2}$$

and define $Y = (X_1, \ldots, X_4)$. Now suppose X_1 and X_2 are correlated with $Cor(X_1, X_2) = \rho$ and have bivariate normal distribution with mean vector $\boldsymbol{\mu} = (0\ 0)^T$ and a covariance matrix

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

Suppose X_3 and X_4 are independent and normally distributed with mean 0 and variance 1. The main effect of X_1 is calculated by taking into account the conditional expectation of $X_1|X_2$ in the following expectation

$$\begin{aligned} E(Y|X_1) &= E[X_1 + X_2 + X_3 + 0.01X_4|X_1] \\ &= X_1 + E(X_2|X_1) + E(X_3) + 0.01E(X_4) \\ &= X_1 + \rho X_1 \end{aligned}$$

and the main effect of X_1 is attained by working out the variance of conditional expectation of $Y|X_1$

$$\begin{aligned} V_1 &= Var_{X_1}[E(Y|X_1)] \\ &= Var_{X_1}[X_1 + \rho X_1] \\ &= (1 + \rho)^2. \end{aligned}$$

The main effect index of X_1 (and of X_2) tends to 2 as $\rho \to 1$. Main effect indices of X_3 and X_4 are unchanged from those in Example 3.1. It seems that X_1 and X_2 are worth learning like X_3. However, since X_1 and X_2 are correlated it may not be worth learning *both* these inputs separately (we would not pay for two independent experiments to learn X_1 and to learn X_2).

If the correlation between those inputs is very high (ρ close to 1) then we could write

$$Y \simeq 2X_1 + X_3 + 0.01X_4. \tag{3.3}$$

Therefore, one of X_1 or X_2 could be considered unimportant, although the main effect is two. This contradictory result shows that input can have a large main effect yet still be 'removed' from a model.

The examples show that the main effect index cannot delineate unimportant inputs properly, mainly when the interaction between inputs exists, or inputs are correlated. Therefore, an unimportant input could not be determined from its main effect index. For this reason, it is helpful to establish definitions of unimportant inputs. We propose two scenarios as ways of formulating the definitions.

§ 3.3 Scenarios for identifying unimportant inputs

Two scenarios for unimportant input identification are proposed. The first scenario is based on a notion of the loss function used in statistical decision theory. The second scenario is motivated by model simplification through factor fixing or factor replacing. The details of each scenario are discussed below.

3.3.1 Scenario 1

In Example 3.1, the main effect of X_1 was zero, but the interaction effect was not, indicating that X_1 could not be neglected as it would have an effect when combined with X_2. In other words, it is worth learning X_1 if we also plan to learn X_2, whereas learning X_1 or X_2 individually is not worth it.

Hence an input could be classified as unimportant if it is not worth learning in any circumstances, individually or in combination with others. In this example, input X_1 is unimportant if the main and interaction effects are zero.

For this reason, we propose a scenario in which an unimportant input is defined. The notion of developing the first definition emerges from a decision theoretic approach to sensitivity analysis (see for example Oakley 2009). Formally, let us define a true and uncertain input vector $\boldsymbol{X} = (X_1, \ldots, X_p)$, where p is the number of inputs in the model f and the output of interest is $Y = f(\boldsymbol{X})$. Next, we consider choosing a decision d from a set of possible decisions $\mathcal{D}$, with loss function $L(d, Y) = L(d, f(\boldsymbol{X}))$. For example, if the decision problem was to supply an estimate d of Y, we might have a quadratic loss function $L(d, Y) = (d - Y)^2$. The aim here is to choose d which minimise the loss function $L(d, Y)$.

Let us define d^* to be the decision that minimises the expected loss given our current information and define

$$L^* = \mathbb{E}_Y[L(d^*, Y)],$$

which we refer to as the baseline minimised expected loss.

Furthermore, let u be a subset of $\{1, ..., p\}$ with $\boldsymbol{X}_u$ is the corresponding sub vector of $\boldsymbol{X}$. We define

$$L_u^* = \mathbb{E}_{\boldsymbol{X}_u}\left[\min_d \quad \mathbb{E}_{\boldsymbol{X}|\boldsymbol{X}_u}\left[L(d, Y)\right]\right] \tag{3.4}$$

to be the minimised expected loss, if we can observe $\boldsymbol{X}_u$ before making a decision d. If $u = \emptyset$ then we write $L_u^* = L_\emptyset^* = L^*$ as the baseline of the minimised expected loss. If $u = 1$ then we write $L_u^* = L_1^*$ as the minimised expected loss if we can observe X_1 before making a decision d. Using the expected loss concept and the formula in Equation (3.4), we propose the definition of an unimportant input as follows.

Definition 1. An input X_i is unimportant if, for any subset u of $\{1, \ldots, p\}$ where $i \notin u$,

$$L_{u\cup\{i\}}^* = L_u^* \tag{3.5}$$

As we can see in (3.5), an unimportant input X_i is determined by the equality of two expected loss functions. This equation means that there is no reduction in expected loss if we learn X_i in addition to $\boldsymbol{X}_u$ for any set u.

To illustrate Definition 1, let us suppose that we have three inputs in a model: X_1, X_2, and X_3. We want to identify whether X_1 is unimportant or not, so we set $i = \{1\}$ and we make several comparisons by including $u = \{\emptyset\}, \{2\}, \{3\}$, and $\{2, 3\}$ in the loss function in Equation (3.5). X_1 is identified as an unimportant input if all of the following comparisons are hold

(i) $L_{\{1,\emptyset\}}^* = L_{\{\emptyset\}}^*$

(ii) $L_{\{1,2\}}^* = L_{\{2\}}^*$

(iii) $L_{\{1,3\}}^* = L_{\{3\}}^*$

(iv) $L_{\{1,2,3\}}^* = L_{\{2,3\}}^*$

Example 3.3. Let us implement Definition 1 for the model in Example 3.1. Recall the model in Equation (3.1) in the form of $Y = f(\mathbf{X})$ as

$$Y = X_1X_2 + X_3 + 0.01X_4$$

where $X_i \sim N(0,1)$. We suppose the decision problem is to estimate Y, with a quadratic loss function $L(d,Y) = (d-Y)^2$. We would like to examine whether X_1 in the model is unimportant or not.

Firstly, we calculate the baseline of expected loss function L_u^* where $u = \emptyset$. In this case, we have obtained $d^* = \mathbb{E}_{\mathbf{X}}[f(\mathbf{X})]$ and $L^* = Var[f(\mathbf{X})] = 2.0001$. Secondly, we need to calculate L_1^*: the expected loss if we first learn X_1. The expression for the expected loss for X_1 derived from (3.4) using a quadratic loss function is

$$L_{\{1\}}^* = \mathbb{E}_{X_1}\left[\min_d \quad \mathbb{E}_{\boldsymbol{X}|X_1}\left[(d - f(\mathbf{X}))^2\right]\right] \tag{3.6}$$

Similarly, we would obtain the optimal $d^* = \mathbb{E}_{\mathbf{X}|X_1}[f(\mathbf{X})]$ so

$$\mathbb{E}_{\boldsymbol{X}|X_1}\left[(d^* - f(\mathbf{X}))^2\right] = \mathbb{E}_{\boldsymbol{X}|X_1}\left[(\mathbb{E}_{\mathbf{X}|X_1}[f(\mathbf{X})] - f(\mathbf{X}))^2\right]$$

consequently,

$$L_{\{1\}}^* = \mathbb{E}_{X_1}\left[Var(f(\mathbf{X})|X_1)\right]. \tag{3.7}$$

By evaluating the conditional variance, it would give $Var(f(\mathbf{X})|X_1) = X_1^2 + 1.0001$ and its expected value or the loss for X_1 is $L_{\{1\}}^* = \mathbb{E}_{X_1}[X_1^2 + 1.0001] = 2.0001$. Therefore, $L_{\{1\}}^* = L^*$ which means there is no reduction in the expected loss from learning X_1 on its own.

Next, we examine the effect of the combination of X_1 and X_2 in terms of their loss. We first calculate $L_{\{2\}}^* = \mathbb{E}_{X_2}\left[Var(f(\mathbf{X})|X_2)\right] = 2.0001$. Afterwards, we calculate the loss for combination of X_1 and X_2, $L_{\{1,2\}}^* = \mathbb{E}_{X_1X_2}\left[Var(f(\mathbf{X})|X_1,X_2)\right] = 1.0001$. We notice that $L_{\{1,2\}}^*$ is smaller than $L_{\{2\}}^*$ meaning that there is a reduction in the loss function if we learn both inputs X_1 and X_2. In conclusion, learning X_1 on its own does not have value but learning X_1 and X_2 does have value.

Example 3.4. Let us consider an example on alternative loss function using the following model

$$y = x_1x_2 + 0.01x_3 \tag{3.8}$$

where x_1, x_2, x_3 are uncertain inputs and are represented as random variables X_1, X_2, X_3 which are i.i.d. and normally distributed with mean 0 and variance 1. Thus, we have a vector of input $\boldsymbol{X} = (X_1, X_2, X_3)$ and we define $Y = f(\boldsymbol{X})$.

We consider two possible decisions d_1 and d_2. These decisions relate to the prediction of Y based on a threshold c and are defined as follows.

- d_1: predict that Y is greater than or equal to c;
- d_2: predict that Y is less than c

The decision will have a consequence because our prediction can be right or wrong. Table 3.1 shows the losses for the wrong prediction (a_1 and a_2) and no losses for the correct prediction.

Table 3.1: The loss of choosing a certain decision

Decision	Truth	
	$Y < c$	$Y \geq c$
d_1	a_1	0
d_2	0	a_2

Let us suppose we choose d_1. If $Y \geq c$ then the loss is 0 or we can write it as a loss function below

$$L(d_1, Y) = 0,$$

and if $Y < c$ then the loss is a_1 or we write

$$L(d_1, Y) = a_1.$$

Alternatively, if we choose d_2 then the loss is

$$L(d_2, Y) = 0,$$

if $Y < c$, and

$$L(d_2, Y) = a_2$$

otherwise. For illustration, suppose we choose $c = 2$, with $P(Y \leq c) = 0.931$ to three decimal places. Now, let us suppose we have the losses of wrong predictions $a_1 = 1$ and $a_2 = 10$. The values indicate a relatively small penalty if we predict $Y \geq c$ and we are wrong. In contrast, we set a larger penalty if we predict $Y < c$ and we are wrong. There is no penalty if the prediction is correct.

The next step is to determine the baseline decision. The baseline decision for d_1 is derived

from the expected loss function of choosing d_1 with respect to Y, $\mathbb{E}(L(d_1, Y))$, which is

$$\begin{aligned}
\mathbb{E}(L(d_1, Y)) &= \int_{-\infty}^{\infty} L(d_1, y) p_Y(y) dy \\
&= \int_{-\infty}^{c} L(d_1, y) p_Y(y) dy + \int_{c}^{\infty} L(d_1, y) p_Y(y) dy \\
&= \int_{-\infty}^{c} a_1 p_Y(y) dy + \int_{c}^{\infty} 0 p_Y(y) dy \\
&= a_1 P(Y \leq c) + 0 \times P(Y > c) \\
&= a_1 P(Y \leq c)
\end{aligned}$$

while the expected loss for d_2 is

$$\begin{aligned}
\mathbb{E}(L(d_2, Y)) &= 0 \times P(Y \leq c) + a_2 P(Y > c) \\
&= a_2 P(Y > c)
\end{aligned}$$

In case of Example 3.4, the baseline decision for d_1 is 0.969 while for d_2 is 0.311. d_2 has the lower loss, so we choose d_2 as our baseline decision with $L = 0.311$.

To identify whether X_1 is unimportant or not, we need to calculate the expected loss for each decision if we learn X_1. The expected loss if we learn X_1 by choosing decision d_1 is $\mathbb{E}(L(d_1|X_1 = x_1, Y))$ while the expected loss if we learn X_1 by choosing d_2 is $\mathbb{E}(L(d_2|X_1 = x_1, Y))$. For a given x_1, we choose a decision which has the minimum loss or

$$\min\Big(\mathbb{E}\big(L(d_1|X_1 = x_1, Y)\big), \mathbb{E}\big(L(d_2|X_1 = x_1, Y)\big)\Big).$$

As we do not know about the true value of X_1, we calculate the expected of the minimum loss and denote it as $L_{\{1\}}$ as below

$$L_{\{1\}} = \mathbb{E}_{X_1}\Big[\min\Big(\mathbb{E}\big(L(d_1|X_1 = x_1, Y)\big), \mathbb{E}\big(L(d_2|X_1 = x_1, Y)\big)\Big)\Big]. \tag{3.9}$$

Using simulation, we obtain $L_{\{1\}} = 0.222$ which is less than our baseline L. Since $L_{\{1\}} < L$ we identify that X_1 is not an unimportant input under Scenario 1. In other words, learning X_1 has value as it reduces the loss.

In general, the expected loss when we learn an input X_i under two possible decisions d_1 and d_2 is

$$L_{\{i\}} = \mathbb{E}_{X_i}\Big[\min\Big(\mathbb{E}\big(L(d_1|X_i = x_i, Y)\big), \mathbb{E}\big(L(d_2|X_i = x_i, Y)\big)\Big)\Big] \tag{3.10}$$

We can calculate the expected loss for other inputs in the model (3.8) using (3.10). We obtain the expected loss if we learn X_2 is $L_{\{2\}} = 0.220$. It is clear that X_2 is also not an unimportant input as $L_{\{2\}} < L$.

The expected loss for learning input X_3 is $L_{\{3\}} = 0.309$. It is relatively similar with the baseline yet we need to examine the expected loss for other possible set of inputs with X_3 to ensure that it is an unimportant input. Based on the simulation we obtain $L_{\{1,3\}} = 0.217$, $L_{\{2,3\}} = 0.212$, $L_{\{1,2\}} = 0.000$, and $L_{\{1,2,3\}} = 0$. If we compare the expected loss for the addition X_3 to other possible set of inputs we obtained the following comparisons

- $L_{\{3\}} \approx L$
- $L_{\{1,3\}} \approx L_{\{1\}}$
- $L_{\{2,3\}} \approx L_{\{2\}}$
- $L_{\{1,2,3\}} = L_{\{1,2\}}$

These results match Definition 1, so we can conclude that X_3 is an unimportant input.

3.3.2 Scenario 2

Let us recall Example 3.2 and focus our attention on X_1 and X_2. Learning either X_1 or X_2 would reduce uncertainty about Y; neither input would be classified as unimportant in Scenario 1. However, the output Y could be simplified, expressed as dependent on three inputs only due to the strong correlation. In that sense, we might wish to classify *one* of X_1 or X_2 as unimportant.

We proposed another scenario to tackle this issue, and we called it Scenario 2. In this scenario, the modeller wishes to simplify the model by reducing the number of model inputs. The idea is that the model can be simplified by fixing one or more inputs to be functions of the remaining inputs.

Definition 2. Let $\boldsymbol{X}_{-i}$ denote a set of all inputs which does not include X_i. An input X_i is unimportant if

$$P\left[f(\boldsymbol{X}_{-i}, X_i)\right] = P\left[f(\boldsymbol{X}_{-i}, X_i = g(\boldsymbol{X}_{-i})\right]. \tag{3.11}$$

where g is a known function, or it might be a constant. If the full probability distributions are not known, we might consider using variance instead as defined in Definition 3.

Definition 3. Let $\boldsymbol{X}_{-i}$ denote a set of all inputs which does not include X_i. An input X_i is unimportant if

$$Var\left[f(\boldsymbol{X}_{-i}, X_i)\right] = Var\left[f(\boldsymbol{X}_{-i}, X_i = g(\boldsymbol{X}_{-i})\right]. \tag{3.12}$$

Definition 2 or 3 could be used to identify an input with a small weight in a model. For instance, we want to identify input X_4 in the model in equation (3.3) which has a small weight, 0.01. Using Definition 3, we calculate

$$\begin{aligned} Var\left[f(X_1, X_2, X_3, X_4)\right] &= Var\left[2X_1 + X_3 + 0.01X_4\right] \\ &= 5.0001 \end{aligned}$$

Following Definition 3, let us suppose to set X_4 to a constant $k = 1$ or formally we can write it as $X_4 = g(\boldsymbol{X}_{-4}) = k$, then we calculate the associated variance

$$\begin{aligned} Var\left[f(X_1, X_2, X_3, X_4 = 1)\right] &= Var\left[2X_1 + X_3 + 0.01\right] \\ &= 5. \end{aligned}$$

As the variances are similar, we could classify X_4 is unimportant and replace it by a constant 1.
In Example 3.2, since X_1 is correlated with X_2, we could replace the input X_1 with a function $g(X_2) = E(X_1|X_2)$ and calculate the variance difference. The variance if we replace X_1 with $E(X_1|X_2)$ is

$$\begin{aligned} Var\left[f(E(X_1|X_2), X_2, X_3, X_4)\right] &= Var\left[2E(X_1|X_2) + X_3 + 0.01X_4\right] \\ &= 4Var(\rho X_2) + 1 + 0.0001 \\ &= 4\rho^2 + 1.0001. \end{aligned}$$

Thus, if $|\rho|$ is close to 1 we can conclude that X_1 is unimportant under Definition 3.

The following sections review some alternative importance measures to main effect indices and analyse their suitability for identifying unimportant inputs within each of these scenarios.

§ 3.4 Total effect indices

In Chapter 2, we described the functional decomposition of $f(\boldsymbol{x})$ and the corresponding variance partition of $Y = f(\boldsymbol{X})$ in the case of independent inputs:

$$f(\boldsymbol{x}) = f_0 + \sum_{i=1}^{p} f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) + \sum_{i<j<k} f_{ijk}(x_i, x_j, x_k) + \ldots + f_{1,\ldots,p}(x_1, ..., x_p) \quad (3.13)$$

where each function in the decomposition is defined as below

$$f_0 = \mathbb{E}[f(\boldsymbol{X})] \quad (3.14)$$

$$f_i(x_i) = \mathbb{E}[f(\boldsymbol{X})|X_i = x_i] - f_0 \quad (3.15)$$

$$f_{ij}(x_i, x_j) = \mathbb{E}[f(\boldsymbol{X})|X_i = x_i, X_j = x_j] - f_i(x_i) - f_j(x_j) - f_0 \quad (3.16)$$

$$\begin{aligned} f_{ijk}(x_i, x_j, x_k) = {} & \mathbb{E}[f(\boldsymbol{X})|X_i = x_i, X_j = x_j, X_k = x_k] - f_i(x_i) - f_j(x_j) - f_k(x_k) \\ & - f_{i,j}(x_i, x_j) - f_{i,k}(x_i, x_k) - f_{j,k}(x_j, x_k) - f_0 \end{aligned} \quad (3.17)$$

and

$$\begin{aligned} Var(Y) = {} & \sum_{i=1}^{p} Var[f_i(X_i)] + \sum_{i<j} Var[f_{ij}(X_i, X_j)] + \sum_{i<j<k} Var[f_{ijk}(X_i, X_j, X_k)] + \ldots \\ & + Var[f_{1,\ldots,p}(X_1, ..., X_p)] \end{aligned} \quad (3.18)$$

The (unnormalised) total effect index for X_i denoted by T_i is defined by summing all of the variance components in (3.18) with indices i. According to Saltelli et al. (2008), in general, total effect can be obtained from the following equation

$$T_i = Var[f_i(X_i)] + \sum_{j} Var[f_{ij}(X_i, X_j)] + \sum_{j} \sum_{k} Var[f_{ijk}(X_i, X_j, X_k)] + \ldots \quad (3.19)$$

Note that

$$T_i = Var(Y) - Var_{\boldsymbol{X}_{-i}}(E(Y|\boldsymbol{X}_{-i})) \quad (3.20)$$

The expression in (3.20) can be seen as an algebraic rule of variance

$$Var(Y) = Var_{\boldsymbol{X}_{-i}}(E(Y|\boldsymbol{X}_{-i})) + E_{\boldsymbol{X}_{-i}}(Var(Y|\boldsymbol{X}_{-i})). \quad (3.21)$$

Therefore the total effect can be obtained by the expectation of conditional variance of the remaining inputs

$$\underset{\sim}{T_i} = E_{\boldsymbol{X}_{-i}}(Var(Y|\boldsymbol{X}_{-i})). \tag{3.22}$$

3.4.1 Identifying unimportant input with total effects

If a total effect T_i is zero then the component of variance decomposition: $Var\Big(f_i(X_i)\Big) = 0$, $Var\Big(f_{i,j}(X_i, X_j)\Big) = 0$ for $\forall j$, $Var\Big(f_{i,j,k}(X_i, X_j, X_k)\Big) = 0$ for $\forall j, k$ and so forth. Each function has expectation zero, for example

$$Var\Big(f_i(X_i)\Big) = \int_{\mathcal{X}_i} f_i(x_i)^2 dp(x_i) = 0,$$

hence $f_i(X_i) = 0$ almost everywhere on $\mathcal{X}_i$. (If X_i has a continuous distribution, then $f_i(X_i) = 0\ \forall \mathcal{X}_i$). Consequently, all functions involving x_i in the decomposition of f are zero. Therefore, in the case of independent inputs, a total effect index of zero would tell us that an input is unimportant under either scenario.

3.4.2 Total effect indices and correlated inputs

If the inputs are correlated, the variance decomposition in (3.18) does not hold. To illustrate this, let us consider the following example.

Example 3.5. Let us suppose that we have a function $f(\cdot) : \mathbb{R}^2 \to \mathbb{R}$ with $f(x_1, x_2) = x_1 + x_2$. Define $Y = f(X_1, X_2)$ with

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

where ρ is the correlation coefficient. We can calculate the variance of output

$$\begin{aligned} Var(Y) &= Var(X_1) + Var(X_2) + 2\,\mathrm{Cov}\,(X_1, X_2) \\ &= 1 + 1 + 2\rho \\ &= 2 + 2\rho. \end{aligned}$$

Now, if we consider the terms in the functional decomposition which are used for the independent input case, we obtain

$$
\begin{aligned}
f_0 &= 0 \\
f_1(x_1) &= \mathbb{E}[Y|X_1] - f_0 = X_1 + \rho X_1 \\
f_2(x_2) &= \mathbb{E}[Y|X_2] = X_2 + \rho X_2 \\
f_{1,2}(x_1, x_2) &= \mathbb{E}[Y|X_1, X_2] - f_1(x_1) - f_2(x_2) = -(X_1 + X_2).
\end{aligned}
$$

If we sum the variance of the functional decomposition we will get

$$
\begin{aligned}
\sum_{i=1}^{2} Var(f_i(X_i)) + Var(f_{1,2}(X_1, X_2)) &= Var(X_1 + \rho X_1) + Var(X_2(1+\rho)) + Var[-(X_1 + X_2)] \\
&= (1+\rho)^2 + (1+\rho^2) + (2+2\rho) \\
&= (2+2\rho^2) + (2+2\rho),
\end{aligned}
$$

thus

$$
Var(Y) \neq Var(f_1(X_1)) + Var(f_2(X_2)) + Var(f_{1,2}(X_1, X_2)). \tag{3.23}
$$

The result confirmed that the variance decomposition does not hold when inputs are correlated. Additionally, we find that

$$
\underset{\sim}{T_1} = \mathbb{E}_{X_2}\Big(Var(Y|X_2)\Big) = 1 - \rho^2,
$$

and

$$
V_1 = Var_{X_1}\Big(\mathbb{E}[Y|X_1]\Big) = (1+\rho)^2.
$$

We can notice that the main effect of X_1 is greater than the total effect when the correlation exist, i.e. if $\rho > 0$ then $\underset{\sim}{T_1} < V_1$. This is not helpful to describe the variance contribution of X_1 to Y.

However, following Kucherenko et al. (2012), we could still define a 'total effect index' as

$$
\tilde{T}_i := E_{\boldsymbol{X}_{-i}}(Var(Y|\boldsymbol{X}_{-i})),
$$

and consider whether this sensitivity measure is appropriate for identifying unimportant inputs in either of our two scenarios. We can easily see that this is *not* appropriate in Scenario 1: in Example 3.2, with $\rho = 1$, we have $\tilde{T}_1 = \tilde{T}_2 = 0$, but there is clearly value in learning either input.

In Scenario 2, $\tilde{T}$ is potentially still valid for identifying unimportant inputs and is more appropriate to use. Consider Definition 2 in Scenario 2, and suppose we choose

$$g(\boldsymbol{X}_{-i}) = E(X_i|\boldsymbol{X}_{-i}).$$

In Definition 3, the requirement is that simplifying the model by setting $X_i = g(\mathbf{X}_{-i})$ does not change $Var(Y)$. Now define

$$D_i := Var(Y) - Var(f(X_i = g(\boldsymbol{X}_{-i}), \boldsymbol{X}_{-i})) \tag{3.24}$$

so that, in Definition 2, we classify X_i as unimportant if $D_i = 0$. If we compare this with

$$\tilde{T_i} = Var(Y) - Var(E[f(\boldsymbol{X})|\boldsymbol{X}_{-i}]), \tag{3.25}$$

we see that if f is linear in X_i, then $\tilde{T_i} = D_i$, as

$$f(E(X_i|\boldsymbol{X}_{-i}), \boldsymbol{X}_{-i}) = E[f(\boldsymbol{X})|\boldsymbol{X}_{-i}]. \tag{3.26}$$

Additionally, $\tilde{T_i}$ may still be close to D_i if f is 'locally linear' in x_i: linear over the main support of X_i conditional on $\boldsymbol{X}_{-i}$.

The illustration of locally linear is provided in Figure 3.1. The figure shows that the two curves (blue and black curves) are intersected with each other. The intersection of the blue curve (curve for $y = f(x_1, x_2)$ with x_2 is fixed) inside the black curve (curve for $P(X_1|X_2)$) is indicated by a red line. The red line looked linear inside the black curve, which indicates that it is locally linear. Locally linear means that it is linear on some parts of the blue curve, but it is no longer linear on the other parts.

Continuing Example 3.5, we can see that as $\rho \to 1$, both $\tilde{T_1}$ and $\tilde{T_2}$ tend to 0, and that we could indeed simplify the function without changing the variance. If for example we write replace X_1 with $E[X_1|X_2]$, where

$$E[X_1|X_2] = \rho X_2 \tag{3.27}$$

then

$$Var_{X_2}(\rho X_2 + X_2) = (1+\rho)^2 Var(X_2) = (1+\rho)^2 \tag{3.28}$$

and both this variance and $Var(Y)$ tend to 4 as $\rho \to 1$.

In conclusion, we can use Scenario 2 for correlated input by replacing input of interest, X_i, with a function $g(\mathbf{X}_{u\setminus\{i\}}) = E[X_i|\mathbf{X}_{u\setminus\{i\}}]$ if f is locally linear in X_i. This setting is also useful for independent input cases. If inputs are independent then X_i does not depend on $X_{u\setminus\{i\}}$ and we simply replace X_i with its expectation $E(X_i)$.

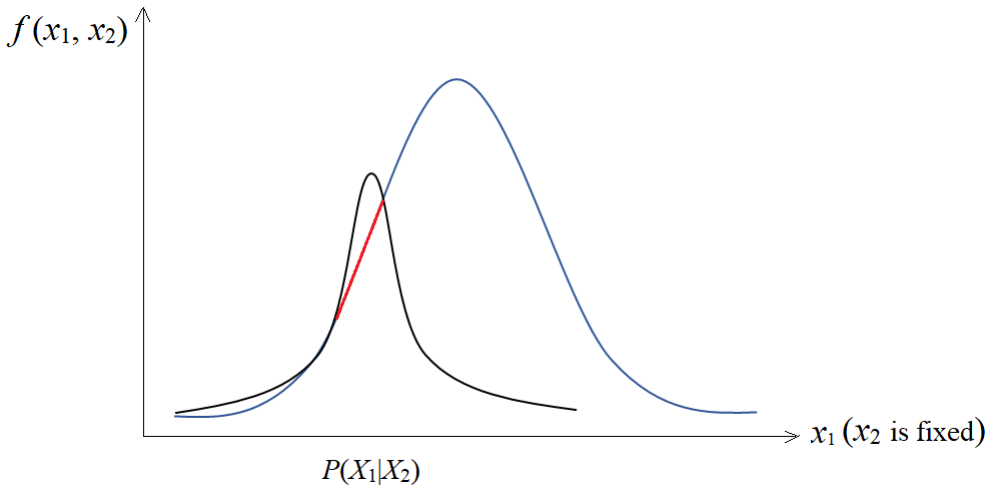


Figure 3.1: An illustration of locally linear. The red line in the small distribution curve of $P(X_1|X_2)$ indicates the function $y = f(x_1, x_2)$ is linear. Nevertheless, it is not linear on the other part of the blue curve.

§ 3.5 Shapley Effects

Owen (2014) proposed a new sensitivity measure based on the concept of the Shapley value in game theory. This new measure can be used to guarantee that the sums of input effects equal the output variance. Shapley value is used to evaluate the "fair share" of a player in a cooperative game. Let us suppose that there is a set of players $K = \{1, ..., k\}$, where k is the number of players. There is a cost of involving players in the game and it is denoted by $c(.)$. If a subset of players J joins the game where $J \subseteq K$ then the cost of involving subset J is $c(J)$.

The Shapley effect of player i with respect to cost, $c(.)$, is defined as

$$Sh_i = \sum_{J \subseteq K \setminus \{i\}} \frac{(k - |J| - 1)!|J|!}{k!} \left(c(J \cup \{i\}) - c(J)\right) \tag{3.29}$$

where $|J|$ indicates the size of J while $K \setminus \{i\}$ denotes the subset of players not involving player i in the game. In the context of global sensitivity analysis, the set of player K is regarded as the set of inputs in a mathematical model, and a cost function $c(.)$ involving variance can be chosen. In this case, $c(J)$ measures the variance of Y caused by the uncertainty of the inputs in J. The cost value of $\emptyset$ and K are $c(\emptyset) = 0$ and $c(K) = Var[Y]$ respectively. Owen (2014) suggested the cost function which suitable in term of global sensitivity analysis. The cost function is

$$c(J) = Var[E(Y|\mathbf{X}_J)] \tag{3.30}$$

Song et al. (2016) showed that using the cost function

$$\tilde{c}(J) = E[Var(Y|\boldsymbol{X}_J)] = Var(Y) - Var[E(Y|\boldsymbol{X}_J)], \tag{3.31}$$

where $\mathbf{X}_{-J} = \mathbf{X}_{K \setminus \{J\}}$, results in the same Shapley value Sh_i. The choice of this cost function will be interpreted similarly with total effect in the global sensitivity analysis.

Example 3.6.(*Correlated inputs*). Let us consider a model.

$$y = x_1 + x_2 + x_3$$

and we define $Y = f(X_1, X_2, X_3)$ where

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & \rho \\ 0 & 1 & 0 \\ \rho & 0 & 1 \end{bmatrix} \right)$$

where $-1 \leq \rho \leq 1$. In this case, we have $k = 3$ and $K = 1, 2, 3$ and the possible subsets of $K \setminus \{1\}$ are $\{\emptyset\}$, $\{2\}$, $\{3\}$, and $\{2,3\}$. The Shapley Effect formula for X_1 is

$$\begin{aligned} Sh_1 &= \sum_{J \subseteq K \setminus \{1\}} \frac{(3 - |J| - 1)!|J|!}{3!} \left(c(J \cup \{1\}) - c(J)\right) \\ &= \sum_{J \subseteq \{\emptyset, \{2\}, \{3\}, \{2,3\}\}} \frac{(2 - |J|)!|J|!}{3!} \left(c(J \cup \{1\}) - c(J)\right) \end{aligned} \tag{3.32}$$

The calculation of the Shapley effect for X_1 was carried out by subtracting the cost function involving X_1 with the cost function of other possible subsets of $K\backslash\{1\}$ as follows

$$\begin{aligned} Sh_1 &= \frac{2!0!}{3!}\left(c(\{1\}) - c(\{\emptyset\})\right) + \frac{1!1!}{3!}\left(c(\{1,2\}) - c(\{2\})\right) \\ &+ \frac{1!1!}{3!}\left(c(\{1,3\}) - c(\{3\})\right) + \frac{1!2!}{3!}\left(c(\{1,2,3\}) - c(\{2,3\})\right) \\ &= \frac{1}{3}\left(c(\{1\}) - c(\{\emptyset\})\right) + \frac{1}{6}\left(c(\{1,2\}) - c(\{2\})\right) \\ &+ \frac{1}{6}\left(c(\{1,3\}) - c(\{3\})\right) + \frac{1}{3}\left(c(\{1,2,3\}) - c(\{2,3\})\right) \end{aligned} \tag{3.33}$$

We choose the cost function in (3.30) for each element in (3.33). The cost for each input can be obtained as below.

$$\begin{aligned} c(\{\emptyset\}) &= Var[E(Y|X_{\{\emptyset\}})] = 0, \\ c(\{1\}) &= Var[E(Y|X_{\{1\}})] \\ &= Var[E(X_1 + X_2 + X_3|X_1)] \\ &= Var[X_1 + E(X_2) + E(X_3|X_1)] \\ &= Var[X_1 + 0 + \rho X_1] \\ &= (1+\rho)^2. \end{aligned}$$

Similarly, we can obtain the cost function for the other inputs and their combinations: $c(\{2\}) = 1$, $c(\{3\}) = (1+\rho)^2$, $c(\{1,2\}) = c(\{2,3\}) = 2+\rho^2$, $c(\{1,3\}) = 2\rho^2$ and $c(\{1,2,3\}) = 1+2\rho^2$. Therefore the Shapley effect for X_1 is

$$Sh_1 = \frac{1}{3}(1+\rho)^2 + \frac{1}{6}(2+\rho^2-1) + \frac{1}{6}(2\rho^2 - (1+\rho)^2) + \frac{1}{3}(1+2\rho^2-2-\rho^2) = \rho^2 + \frac{1}{3}\rho.$$

The Shapley effects for X_2 and X_3 are obtained using the same way, and the results are $Sh_2 = 1 - \frac{1}{3}\rho$ and $Sh_3 = \rho^2 + \frac{1}{3}\rho$.

3.5.1 Identifying unimportant inputs with Shapley Effects

As we notice in equation (3.29), the Shapley effect for X_i is calculated from the sum of the incremental cost of adding input X_i to every possible subset. If $Sh_i = 0$ then there is no incremental cost for adding input i to any subset J. This matches Scenario 1, where we consider the incremental reduction in loss from learning X_i in addition to a subset of inputs J. Therefore, if $Sh_i = 0$ we can conclude that X_i is unimportant within Scenario 1.

Shapley effects for correlated inputs have a similar principle with the independent inputs. When calculating the cost functions, the correlated inputs are treated as a conditional variance.

The Shapley effect would not identify an unimportant input under Scenario 2, but an individual term in the sum that defines Sh_i would. We can identify X_i as an unimportant input using the cost function in the Shapley effect which satisfies the following condition

$$c(K\backslash\{i\} \cup \{i\}) - c(K\backslash\{i\}) = 0. \tag{3.34}$$

§ 3.6 Distribution-based methods

According to Park & Ahn (1994) distribution-based sensitivity analysis (DSA) computes sensitivity of the input random concerning the change in the distribution of output response. These methods are also called moment-independent methods since we are not analysing particular moments of the output distribution, as would do with variance-based methods. Some of the methods using this approach are described in Chun et al. (2000), Borgonovo (2007) and Plischke et al. (2013).

3.6.1 Difficulties with distribution-based methods: an example

The distribution-based methods might have a potential problem, especially with the difficulty of its interpretation. To illustrate the issue, let us consider the following example.

Example 3.7. Let us suppose a model with the interaction of two inputs

$$Y = X_1 X_2$$

with $X_1 \sim N(0,1)$ and X_2 is a discrete distribution where $P(X_2 = -1) = 0.5$ and $P(X_2 = 1) = 0.5$. We can notice that the distribution of Y is also Normal (0,1) as, in theory, the standard normal distribution remains the same if we multiply it by -1. This means that learning the value of X_2 will not change the distribution of Y. If we identify X_2 using the distribution-based methods, we see that it is an unimportant input. Nevertheless, if we look at the model, we cannot say that X_2 is unimportant as it still has a value of learning the input because of its interaction with X_1.

3.6.2 Chun-Han-Tak Index

A sensitivity index that is based on moment independent and looks at the entire distribution of the model output is developed by Chun et al. (2000). The name of the index is Chun-Han-Tak (CHT) and is formulated as follows.

$$CHT_i = \frac{\left(\int_0^1 [y_\alpha^i - y_\alpha^o] d\alpha] \right)^{1/2}}{E(Y^o)} \tag{3.35}$$

where y_α^i is the α^{th} quantile of Y for the "sensitivity case" and y_α^o is the α^{th} quantile of Y for the base case. The base case is a case when we obtain the distribution of output with all inputs included in the model with their distributions: the base case is the distribution of $Y = f(X_1, \ldots, X_p)$. The sensitivity case is a case when the distribution of output is obtained by changing a particular input distribution in the model: we change the distribution of X_i in some way (which could mean holding X_i fixed).

If we implement the CHT measure to Example 3.1, we will get the value of CHTs for X_1 = 20.0371, X_2 = 20.0371, X_3 = 1.5338, and X_4 = 0.032. Based on this result, the CHT index correctly identifies that X_1 and X_2 has some effects.

3.6.3 Identifying unimportant inputs with the Chun-Han-Tak index

According to Example 3.1 above, a small CHT index does not guarantee unimportance under Scenario 1. A small value of CHT index occurs when $y_\alpha^i \approx y_\alpha^o$ for $\forall \alpha$ and this is only one part of the unimportance criteria in Scenario 1 which corresponds to $L^*_{\{i\}} = L^*_{\{\emptyset\}}$. The CHT index can be used for Scenario 2 if in the sensitivity case; we fix X_i as an appropriate function of X_{-i}.

3.6.4 Borgonovo's δ and The Kolmogorov-Smirnov Index

Another distribution-based index was introduced by Borgonovo (2007). He developed an index δ_i as following

$$\delta_i = \frac{1}{2} E_{X_i}[s(X_i)] \tag{3.36}$$

where

$$s(X_i) = \int_y |f_Y(y) - f_{Y|X_i}(y)| dy \tag{3.37}$$

We can see that $s(X_i)$ is the shift of two distributions, $f_Y(y)$ and $f_{Y|X_i}(y)$.

A similar importance measure was proposed in Baucells & Borgonovo (2013), based on the conditional distribution function rather than conditional density function:

$$\beta_i^d = E[d\{F_Y, F_{Y|X_i=x_i}\}] \tag{3.38}$$

with $d\{.,.\}$ is a probability metric between the distributions. If we choose the Kolmogorov-Smirnov distance

$$d^{KS}\{F_Y, F_{Y|X_i=x_i}\} = \sup_y |F_Y(y) - F_{Y|X_i=x_i}(y)| \tag{3.39}$$

we will obtain The Kolmogorov-Smirnov's (K-S) importance measure

$$\beta_i^{KS} = E[\sup_y |F_Y(y) - F_{Y|X_i=x_i}(y)|]. \tag{3.40}$$

Baucells & Borgonovo (2013) explains that an advantage of β_i^d is that it is transformation invariant: transforming the output (e.g. to a log scale) does not change the sensitivity measure.

Let us recall the problem in Example 3.1 and apply these two measures. If we use Borgonovo's δ we will get the δs for each input $\delta_1 = 0.0077$, $\delta_2 = 0.0077$, $\delta_3 = 0.0170$, and $\delta_4 = 0.0000$. The delta indices for X_1 and X_2 are very small but they still have some effects. X_3 has the largest index amongst other input indicating it gives the effect to the output. In contrast, X_4 has the smallest index which is likely not having any effect to the output Y.

Alternatively, if we use the K-S measures, we will obtain the value for each input $\beta_1^{KS} = 0.0768$, $\beta_2^{KS} = 0.0772$, $\beta_3^{KS} = 0.1580$, and $\beta_4^{KS} = 0.0061$. According to these results, inputs X_1 and X_2 have some effects to the output Y although their indices are very small.X_3 has the biggest effect on Y while X_4 only gives a very small effect.

3.6.5 Identifying unimportant inputs with Borgonovo's index and The Kolmogorov-Smirnov index

We review the use of Borgonovo's δ and K-S index for identifying unimportant inputs based on our definition in Scenario 1 and 2. We observed that these measures do not guarantee unimportance under Scenario 1, as they do not consider the comparison of all possible subsets which includes X_i. They only consider one comparison of the loss functions in Definition 1 which is $L^*_{\{i\}} = L^*_{\{\emptyset\}}$.

Under Scenario 2, the measures would work if X_i could be fixed at any constant without changing distribution of Y, but they would not detect the case if we set X_i to be some functions of X_{-i}. Therefore, if we fix $X_i = k$ then using Borgonovo's measure, input X_i is unimportant if it satisfies $f(y) = f_{Y|X_i=k}(y)$ and using K-S measure if it satisfies $F(y) = F_{Y|X_i=k}(y)$. Nevertheless, if we set $X_i = g(X_{-i})$ it is very unlikely that we would obtain the condition such that $f(y) = f_{Y|X_i=g(X_{-i})}(y)$ or $F(y) = F_{Y|X_i=g(X_{-i})}(y)$. This is because the density function and the distribution function will change given the value of $g(X_{-i})$.

§ 3.7 Conclusion

According to the review and comparison in this chapter, we can now tabulate the relationship between the proposed scenarios and importance measures. The tabulation of suitability between scenarios and importance measures are provided in Table 3.2.

Table 3.2: The suitability between proposed scenarios and importance measures

Importance Measures	Scenario 1		Scenario 2	
	Independent Inputs	Correlated Inputs	Independent Inputs	Correlated Inputs
Total Effect Indices	✓	×	✓	✓[a]
Shapley Effects	✓	✓	×	×
Chun-Han-Tak	×	×	✓	✓[b]
Borgonovo's δ	×	×	✓	✓[b]
Kolmogorov-Smirnov	×	×	✓	✓[b]

[a] If the function used is $g(x_{-i}) = E[X_i|X_{-i}]$ and f is locally linear in X_i.
[b] If the number of correlated inputs is small.

The table summarises that we would not expect a single importance measure to cover all scenarios/cases because each measure has a different approach to expressing the input unimportance. Total effect indices and Shapley effects identify unimportant inputs based on the variance comparisons, including the input of interest with other subsets of input. This type of comparison closely links with the Scenario 1 principle, which investigates whether there is a value in adding input of interest to other subsets of input. Therefore, if

the total effect and Shapley effects are zero, then all element comparisons in their formula regarding the input of interest are zero including the interaction. These indicate that the input is unimportant. Note that this does not apply to correlated input for the total effect indices as the variance decomposition is not held when the inputs are correlated.

However, it is interesting to note that the total effect index comes close, as it also matches Scenario 2. As discussed in Section 3.4.2, conditional expectation term, $E(f(\mathbf{X})|\mathbf{X}_{-i})$, in Equation (3.25) can be replaced by the function of $f(E(X_i|\mathbf{X}_{-i}), \mathbf{X}_{-i})$. For independent inputs, the function is in the form of $f(E(X_i), \mathbf{X}_{-i})$ and for dependent inputs, we can use the function with the condition of locally linear must be fulfilled.

On the other hand, some importance measures such as Chun-Han-Tak, Borgonovo's δ, and Kolmogorov-Smirnov consider the differences in the unconditional distribution (quantile, density, or cumulative) of the output with the conditional distribution of output given the input of interest. These approaches match with Scenario 2 Definition 2, which compares the probability distribution of the output with the probability distribution when an input of interest X_i is set to the function $g(\mathbf{X}_{-i})$. The choice of this function for instance $g(\mathbf{X}_{-i}) = k$ as discussed in Section 3.6.5, or $g(\mathbf{X}_{-i}) = E(X_i|\mathbf{X}_{-i})$. In the case of independent input, the later function simplifies to $g(\mathbf{X}_{-i}) = E(X_i)$. Therefore, if $\delta_i = 0$, $\beta_i^{KS} = 0$, or $CHT_i = 0$ indicates that input X_i is unimportant under Scenario 2. When inputs are correlated, then we can use $g(\mathbf{X}_{-i}) = E(X_i|\mathbf{X}_{-i})$ instead. Nevertheless, as discussed in Section 3.6.5, if we use this function, it is difficult to satisfy the condition, for example $f(y) = f_{Y|X_i=E(X_i|\mathbf{X}_{-i})}(y)$, as the conditional density function will always change and depend on the rest of inputs. It is possible to have a similar condition if there are only a few numbers of correlated inputs in the model.

The distribution-based methods are not suitable for Scenario 1, as we need to modify these methods for appropriate comparisons. For example, if we want to investigate input X_i to be an unimportant input, then there are some density comparisons for the input. For example, if we use δ_i then we need to compute some different comparisons of density functions. If we had three inputs in the model say X_i, X_j, and X_k where $i \neq j \neq k$, and our interest is in identifying whether input X_i is unimportant, then we need to compute some density differences as follows.

$$f_Y(y) - f_{Y|X_i}(y),$$

$$f_{Y|X_j}(y) - f_{Y|X_i,X_j}(y), \text{ where } i \neq j, \text{ and}$$

$$f_{Y|X_j,X_k}(y) - f_{Y|X_i,X_j,X_k}(y), \text{ where } i \neq j \neq k.$$

Thus, if we use one single function difference, it is not sufficient to conclude that an input X_i is unimportant under Scenario 1. In other words, if δ_i in (3.36) is zero, it does not imply that input X_i is unimportant under Scenario 1. A similar reason applies to Chun-Han-Tak and Kolmogorov-Smirnov methods.

Chapter 4

Elicitation

§ 4.1 Introduction

This chapter explains the elicitation concept, including the definition, an example, and some established elicitation protocols. The structure of this chapter is organised as follows. Section 4.1 provides the chapter overview, Section 4.2 defines elicitation and its connection with the computer models, Section 4.3 presents an example of elicitation with a single judgment, Section 4.4 discusses the general elicitation process, Section 4.5 introduces some existing elicitation protocols and Section 4.6 is the summary of this chapter.

§ 4.2 Elicitation and computer models

In the previous chapter, we presented the criteria for identifying unimportant inputs in the mathematical or computer models. In this chapter, we will be focusing on the elicitation of inputs in the computer models, which is essential before performing the sensitivity analysis.

Complex computer models are advantageous to simulate actual physical phenomena; however, they are computationally expensive. Lee et al. (2011) used the emulation technique to reduce the computational time for analysing the global model of aerosol process (GLOMAP). Using the emulation, the computer model is replaced by a statistical surrogate model to make the computation time more efficient. In the basic procedure of the emulation study carried out by Lee et al. (2011), an elicitation step is included before running the computer model. Although the elicitation of parameter uncertainties was not

carried out in the initial step in their study, this step will affect the model output and influence the sensitivity analysis results. The following describes the fundamental concept of elicitation.

Experts are often involved in supporting a decision-maker to make a decision in various fields, such as in public policy or risk analysis. Their opinions or judgments are frequently used as a fast and efficient way for assisting in making such decisions. Furthermore, the expert's judgment must be quantitatively expressed formally, i.e., statistical probability distribution. This leads to a concept called elicitation.

According to O'Hagan (2019), elicitation is defined as the process of expressing expert knowledge in the form of a probability distribution for uncertain quantities. Another definition of probability elicitation represents subjective uncertainty about a fixed quantity using a probability distribution. Therefore, the goal of the elicitation is to construct a probability distribution that properly represents the expert's knowledge/uncertainty (O'Hagan et al. 2006). The probability distribution produced from the elicitation is subjective to an expert. Morgan (2014) provided a guideline on how to conduct a sensible elicitation and how the best way to do the elicitation. Some experts involved in the elicitation might provide different probability distributions based on their own judgments. Therefore some elicitation protocols are required to combine the expert distribution. Some established elicitation protocols such as SHELF, IDEA, and the classical model. Hanea et al. (2021) made a review of the three established elicitation protocols regarding their characteristics.

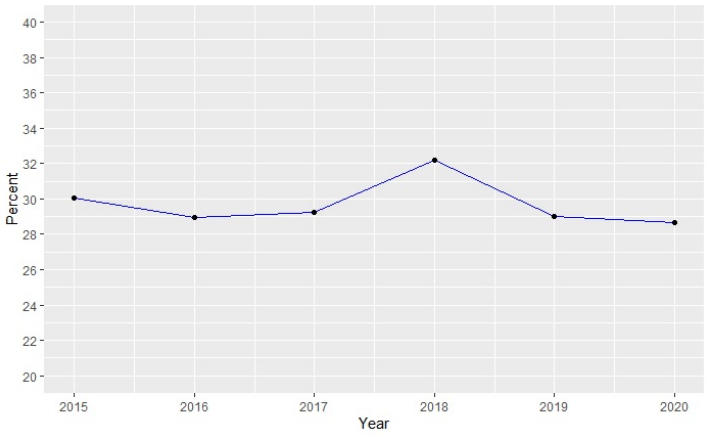
The following section illustrates an example of elicitation for an uncertain quantity using an individual judgment.

§ 4.3 An elicitation example

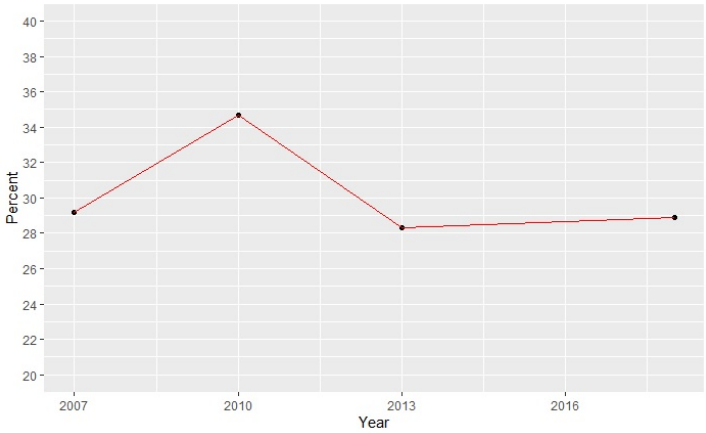
Let X be the proportion of all adults in Indonesia who will be regular smokers (at least one cigarette per day) in 2030. In this example, the author's own judgments will be elicited to provide a probability distribution for this uncertain quantity X. In order to make sensible judgments, it is worth searching for relevant time-series data available in the literature or websites.

There are two sources of data that can be used to make appropriate judgments. The first data set is from *Statistics Indonesia* which records the percentage of smokers aged ≥ 15 years from 2015 to 2020. The data was obtained from the annual Indonesian socio-economic survey (Indonesia 2021). The second data set is obtained from research con-

ducted by the Department of Research and Development for health, Indonesian Ministry of Health (Health 2018). The research aimed to capture the prevalence of smokers aged ≥ 10 collected in 2007, 2010, 2013, and 2018.



(a) Percentage of smokers *aged* $\geq$ 15 years from Statistics Indonesia



(b) Prevalence of smoker aged ≥ 10 years from reports of the research and development, Ministry of Health

Figure 4.1: The trend of percentage of smokers recorded by two different institutions

The available data do not directly estimate the uncertain quantity X. The first reason is that X should represent the adult smokers whose ages are ≥ 18 years. Secondly, the available data are insufficient to predict adult smokers in 2030. The author will make a judgment based on the data even if the range of ages includes the teenager, but it will not reduce the proportion of adult smokers as the author thought that adults would more likely smoke than teenagers.

The first step of making a judgment is to consider the plausible range of X. Two bounds need to be set up, the lower bound (L) and the upper bound (U). The plausible range of X should lie between L and U with the probability greater than 98%, and it is very unlikely that the value of X will fall below L or greater than U i.e., the probability of those occurrences are less than 1%. In terms of probabilistic judgments, we can represent them as $P(L < X < U) \geq 0.98$, $P(X < L) < 0.01$, and $P(X > U) < 0.01$. The next step is to implement the tertiles or quartiles method to make other judgments about the value of X, which should lie inside the plausible range $[L, U]$. In this example, the quartiles method will be used and the expert will be asked to make their judgments about the value of X based on Median (M), the first quartile (Q_1), and the third quartile (Q_3).

Based on the time series data in Figure 4.1a and Figure 4.1b, the author makes the following judgments. The author believes that the proportion of smokers will fall unlikely below 20% nor exceeds 40%; therefore, the author sets the $L = 20\%$ and $U = 40\%$. The next judgments on the quartiles will be bounded between 20 - 40%. According to the figures, the proportions of smokers in six years were around 30%. Thus, the author judges the median (M) to be 30% with $Q_1 = 26\%$ and $Q_3 = 31\%$.

We can represent the judgment values in the form of cumulative probability distributions as below.

$$P(X \leq 26) = 0.25, \;\; P(X \leq 30) = 0.5, \;\; P(X \leq 31) = 0.75.$$

Next, we can use the quantile values to fit a distribution using an R package, `SHELF`. A Beta distribution (a, b) was chosen as the general distribution of X whose values between zero and one. The parameters $\theta = (a, b)$ are estimated numerically by minimising the following function

$$H(\theta) = \Big(F(26; \theta) - 0.25\Big)^2 + \Big(F(29; \theta) - 0.5\Big)^2 + \Big(F(31; \theta) - 0.75\Big)^2, \qquad (4.1)$$

where $F(\cdot; \theta)$ is the cumulative distribution function of the Beta distribution with parameters $\theta = (a, b)$. In general, the estimates of parameter θ are achieved by

$$\hat{\theta} = \arg\min_{\theta} H(\theta).$$

The minimisation process in `SHELF` is done numerically using the Nelder-Mead optimisation method to get the best estimates. The fitted beta distribution is shown in Figure 4.2.

The graph in Figure 4.2 indicates that the proportion of Indonesian adult smokers in 2030 has a Beta distribution with parameters (2.67, 3.17). The beta distribution plot indicates that the proportion of adult smokers is likely to be around 29%.

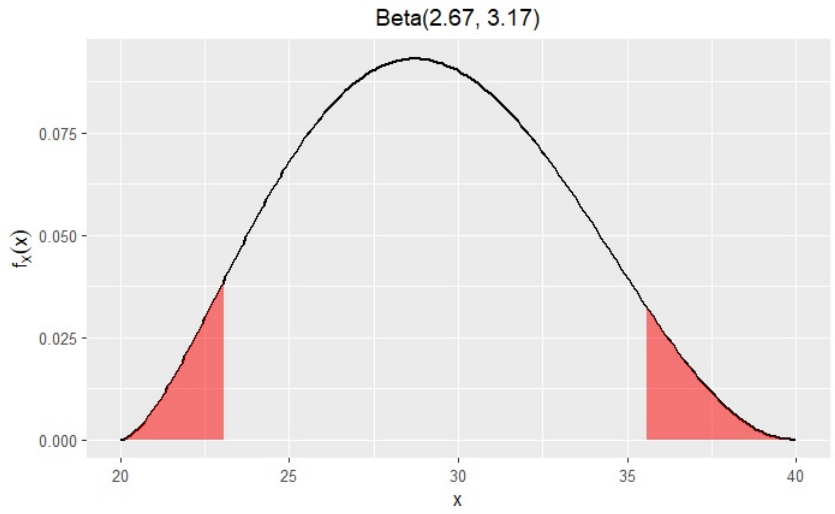


Figure 4.2: The fitted distribution Beta distribution of the initial judgment for X

Feedback is required to the fitted distribution to check whether the distribution satisfies our judgment or not. To do so, we can compute the 1%, 10%, 90%, and 99% quantiles of the fitted Beta distribution (2.67, 3.17) and observe the values. The values of the quantiles are 0.0807, 0.2073, 0.7176, and 0.8735, respectively. The 1%, 10%, and 99% quantiles seem to be sensible. However, it is very unlikely that the smoker proportion could be greater than 0.9 quantile, i.e. 71.76%, which means more than half of the adults will be regular smokers.

Therefore, another judgment to adjust X is made using the following values: $L = 20\%$, $U = 45\%$, $M = 28\%$, $Q_1 = 25\%$, and $Q_3 = 29\%$. Similarly, we can represent them as cumulative distribution functions

$$P(X \leq 25) = 0.25, \;\; P(X \leq 28) = 0.5, \;\; P(X \leq 29) = 0.75$$

and re-fit the distribution. The distribution of the later judgment fits with Beta (3.61, 8.26) as shown in Figure 4.3. The mode of the distribution indicates that the proportion of adult smoker is likely to be 27%. Again, we clarify the distribution by computing other quantiles. The results for quantiles 1%, 10%, 90%, 99% are 0.067, 0.144, 0.478, and 0.633 respectively. The proportion of adult smokers is very unlikely to be less than 6% (1% quantile) or greater than 63% (99% quantile). There is a genuine possibility that the proportion of smokers could be less than 14% (10% quantile) or greater than 48% (90% quantile). Now, all four quantiles have confirmed the sensible proportions of adult smokers.

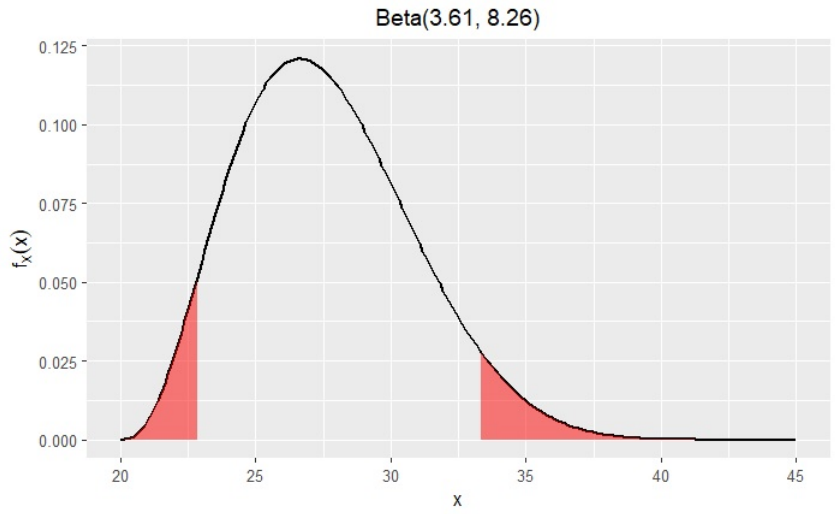


Figure 4.3: The fitted Beta distribution of the second judgment for X

§ 4.4 The general elicitation process

There are five steps of the elicitation process which are generally used for obtaining expert probability distribution (O'Hagan et al. 2006). The steps are discussed below.

1) Background and preparation
   The initial step of the elicitation process is to identify variables that are needed to be elicited. The elicitation user or the client commonly carries out this step with support from a statistician. Once the variables of interest have been identified, the statistician or the facilitator of elicitation is able to understand the scope of the field and the relevant expertise. Subsequently, a document is created based on the background of the study and the identified variables as prior information for the experts. The supporting documents such as the elicitation plan and a list of questions are also prepared.

2) Identify and recruit experts
   The next step is to identify suitable experts for eliciting the variables of interest. According to O'Hagan et al. (2006), an expert should fulfill some criteria: (i) tangible evidence of expertise, (ii) having a good reputation, (iii) availability and willingness to participate, (iv) understanding of the general problem area, (v) impartiality, and (vi) lack of economic or personal stake in the potential findings.

3) Motivating and training the experts

Experts might have different backgrounds, and possibly it is the first time they make judgments in the form of probability. Therefore, it is essential to provide sufficient information about the elicitation to the expert. The information which can motivate experts involves the need of their appropriate judgments for supporting the decision-making required by the client. In addition, the nature of uncertainty for their judgments required to be minimised, and the way of expressing their judgment into a probability distribution by the support of a facilitator. The elicitation training for experts is then carried out to equip them with the following knowledge: Probability and probability distributions, information about the most common judgment heuristics and biases, including advice about how to overcome them, and practising an elicitation using a given example.

4) Structuring and decomposition

An effort to ensure elicitation works well when the expert is ready to do their task elicits structure such as dependencies and functional relationships. The experts should agree upon such a relationship before they are used to providing a similar initial perception of the problem of interest. A precise definition of the quantity of interest is essential for the elicitation process. This definition includes the measurement unit of the variables. The expert will need to use some evidence or information associated with the variables to make their judgments.

5) The Elicitation

The elicitation step is an iterative process that involves the following actions: Elicit specific summaries of expert's distribution, fit a probability distribution to those summaries, and assess the adequacy: if the probability distribution is adequate then it stops otherwise the process is repeated by asking the expert to make adjustment.

§ 4.5 Elicitation protocols

The elicitation will produce a probability distribution based on the judgment made by an expert. In the previous example, in Section 1.3, the author has performed an elicitation to determine the proportion of Indonesian adult smokers in 2030, and the final judgment is fitted by Beta (3.61, 8,26). Multiple experts might get involved in the real elicitation and might arise additional problems: How to manage the interaction between experts? Furthermore, how to produce a single distribution from the variety of expert distributions?

An elicitation protocol is then required to accommodate the issues. There are some established elicitation protocols in the literature, such as IDEA, SHELF, and Cooke's protocol or classical model. The protocols are described as follows.

1) IDEA

   The IDEA protocol was introduced by Hemming et al. (2018) and is an acronym of Investigate, Discuss, Estimate and Aggregate, which represent the steps in undertaking the elicitation. The elicitation involves multiple experts with an extensive discussion between experts to obtain a group judgment.

   The protocol advises the expert judgment is conducted in two rounds. In the first round, the "Investigate" step, the experts are given a set of questions. The experts should answer the questions in the form of probabilities or quantities (e.g. mean) about the uncertain quantity X. These types of judgments (in the form of probabilities or quantities) depend on format is used in the elicitation. There are two elicitation formats in the IDEA protocols: the three-step elicitation and the four-step elicitation. In the three-step elicitation, experts will be asked to provide the lowest, the highest, and the best plausible probabilities of an event will occur. While in the four-step elicitation the experts will need to provide three plausible quantities: the lowest (L), the highest (U) and the best guess (B) for X and the expert also need to specify the level of confidence (C). The data obtained from the first round using the four-step elicitation method are then standardised using the following formulae:

   $$\text{Lower standardised level} = B - ((B - L)) \times (S/C))$$

   $$\text{Upper standardised level} = B + ((U - L)) \times (S/C))$$

   where S is the level of credible intervals (typically to 80% or 90% credible intervals).

   Next, the standardised values of the lower and upper levels and the best guess value are aggregated using mathematical aggregation. In particular, one type of

mathematical aggregation i.e., quantile aggregation, is used by taking the mean of each of those values. Graphical feedback containing the expert's responses and the quantile aggregate is created and is reported to the experts.

The expert are invited to discuss the graphical feedback from Round 1. This step is in the "Discuss" step by the guidance of a facilitator. The facilitator will clarify the judgments made by experts by asking for their reason and evidence.

In Round 2, the "Estimate" step, the experts are given similar questions and are asked to re-answer the questions. After attending the discussion session, they might change or revise their judgments by providing their reason and comments.

The updated judgment results in Round 2 are aggregated similarly to Round 1, producing the final aggregation estimate. This final aggregation is in the "Aggregate" step.

2) SHELF

The SHELF protocol was developed by Oakley & O'Hagan (2019) and stood for SHeffield ELicitation Framework. Multiple experts are also required to carry out the elicitation, and a facilitator will assist the discussion between experts. Unlike the IDEA, SHELF uses behavioural aggregation to obtain the expert's group judgment via a consensus. Each expert will be asked for their judgment about an uncertain quantity using quantiles or tertiles, and a distribution will be fitted according to their judgment's values. Afterwards, all experts with their probability distribution will be asked to make a consensus judgment from potential differences in their distributions. This protocol uses an R package `SHELF` to fit the expert's distribution which was developed by Oakley (2020).

3) Classical Model

The classical model is referred to the Cooke's protocol (Cooke 1991) which also involves multiple experts to obtain a group judgment. A facilitator of the elicitation must have a good understanding of the aggregation method but not necessarily have the ability to manage the expert's discussion. Likewise, in IDEA, the mathematical aggregation is used to obtain the group judgment. In the classical model, the expert's initial judgments can be used to formulate an aggregate result. The aggregate result is obtained by a pooling rule considering the weight of the expert's performance on making a judgment on seed variables (variables whose true values are known by the facilitator).

The number of experts involved in the structured elicitation, such as IDEA, SHELF, and the Classical Model, will affect the aggregate result or the final distribution. Therefore, an optimum number of experts recruited in the elicitation is essential to be predetermined. Hanea & Nane (2021), and O'Hagan (2019) advised that the number of experts for all protocols is anything between four and ten.

A formal elicitation framework requires a great effort as the experts are expected to gather in a particular place and time. It is challenging as genuine experts are usually busy and do not have much time to attend a long meeting. Additionally, some selected experts might live in remote places that are not practical if they come to the specified place for an elicitation. To overcome this issue, it is worth considering a remote elicitation. Grigore et al. (2017) developed a tool for supporting this type of elicitation which is based on Excel software called EXPert eliCItation Tool (EXPLICIT). The experts do not need to gather in a specific place; instead, they are given a question template in which they can fill up on their own based on their judgment, and the distribution will show up in the template afterwards. The experts might revise their judgment if they feel the resulting distribution is not appropriate.

Online training for experts is necessary to provide the background of elicitation to support a remote elicitation. Training materials also consist of probability concept and probability distributions, and also how to make a judgment in terms of probability. A very useful online training of probabilistic judgments was provided by O'Hagan (2018b).

§ 4.6 Summary

Elicitation is a way of obtaining the expert's qualitative judgment into quantitative information in the form of probability distributions about uncertain quantities of interest. The method is very useful in supporting a decision-maker to make a fast decision when the data or information is not available or insufficient. The data collection using standard research methodology is expensive, or experiments are impractical.

Nevertheless, a formal elicitation workshop using protocols such as IDEA, SHELF, or Classical Model is expensive to carry out in terms of time. In particular, the quantity of interest that needs to be elicited will affect the duration of the elicitation process. If the number of inputs in the mathematical models as the quantity of interest was very large, say more than ten. The elicitation would be very time-consuming, even if it uses a remote elicitation.

To overcome this issue, it is worth thinking about prioritising the quantity of interest that will be elicited. For instance, if we had ten inputs in the mathematical model as the quantity of interest, then we might order them from the most important input to the least important input. We might only choose the top five from the order of the inputs required to be elicited and keep the rest of the inputs as they were less important in the model or set them as constants. This strategy will save much time and make the elicitation more efficient.

In conclusion, we need to prioritise the uncertain quantities (in this case, inputs in the mathematical model) before the formal elicitation is carried out to choose based on the priority which inputs will be elicited and which ones will not. The next chapter discusses the methodology of prioritising expert judgments.

Chapter 5

Prioritising Elicitation and Robust Global Sensitivity Analysis

§ 5.1 Motivation

The previous chapter discussed elicitation and reviewed some elicitation protocols. The elicitation is a valuable technique for quantifying uncertain quantities when the data are unavailable or insufficient. However, the process of conducting an elicitation is time-consuming, especially when the number of uncertain quantities is large. In computer models, uncertain quantities are present in the model's inputs. If a model has many inputs, the number of uncertain quantities to be elicited will also increase. Eliciting an input means that the analyst should determine a suitable distribution, and it may not be feasible to elicit all of the input's distributions. Nevertheless, the analyst can elicit some of the inputs. The issue is that the analyst should choose some inputs from the model to be elicited. It means that the analyst needs to prioritise inputs before conducting elicitation. The elicitation will only be performed for inputs that have a high priority, while inputs with low priority will not be elicited. Based on this motivation, the research question is: "How might the analyst prioritise the elicitation?".

In this chapter, we will consider making prioritisation for input elicitation. We provide an example to show whether it is correct to use sensitivity measures in the standard sensitivity analysis for making such prioritisation. In addition, we also review some literature on robust global sensitivity analysis to study how sensitivity analysis results will be affected if the uncertainty in the input of a model changes. The following section discusses two

types of distributions considered in this chapter, the analyst's distribution and the expert's distribution.

§ 5.2 The analyst's distribution and expert's distribution

We now distinguish between two possible distributions for any input. The two distributions are the analyst's distribution and the expert's distribution. The analyst's distribution is a distribution specified by an analyst using his or her own belief about the uncertainty of an input. In comparison, the expert's distribution is the distribution provided by an expert about uncertainty in an input based on expertise. The analyst's distribution is quick to obtain but has high uncertainty. In contrast, the expert's distribution is slow to obtain but less uncertain.

To illustrate the two distributions, let us define a mathematical model $y = f(x_1, \ldots, x_p)$, where p is the number of inputs. Inputs $x_1, \ldots, x_p$ in the model are uncertain and denoted by $X_1, \ldots, X_p$. Because inputs are uncertain and the true value of input X_i is unknown, using the elicitation concept, we can elicit the distribution for X_i, $\pi(X_i)$. Let us suppose that $\pi_A(X_i)$ and $\pi_E(X_i)$ denote the analyst's distributions and the expert's distributions for input X_i. The analyst might specify a simple, crude and uninformative distribution $\pi_A(X_i)$ with wide support from a to b. On the other hand, experts might specify another distribution $\pi_E(X_i)$ that is more suitable, more informative, and might have narrower support than the analyst's distribution. The illustration of both distributions is shown in Figure 5.1.

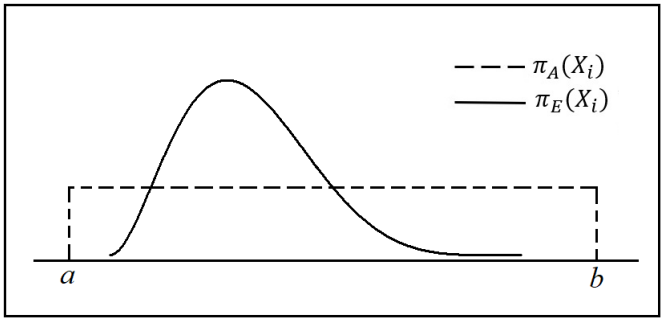


Figure 5.1: An illustration of an analyst's distribution ($\pi_A(X_i)$) and an expert's distribution ($\pi_E(X_i)$) for quantifying the uncertainty of an input. The distribution indicated by a dash line is the analyst's distribution and the distribution indicated by the solid line is the expert's distribution.

In a mathematical or a computer model, there are p uncertain inputs and the analyst or the expert should elicit all of those inputs. However, it is unlikely that all inputs are equally important. Thus, it appears that we have another sensitivity analysis problem. It is reasonable to rank inputs in the model based on their importance level, which is related to the factor prioritisation problem in the previous chapter. Therefore we have questions: Can we apply methods from previous chapters to decide how to prioritise expert elicitation? Can we consider sensitivity analysis methods where input uncertainty is described by $\pi_A(X_i)$?

As an illustration, let us consider two inputs in a model, X_1 and X_2, and we want to determine which inputs should be elicited by the expert. If the expert elicits X_1 then the distribution for X_1 from the expert is $\pi_E(X_1)$ and if expert elicits X_2 then we will obtain the distribution for X_2 instead of X_1, $\pi_E(X_2)$. To decide which input should be elicited, we might prioritise inputs using factor prioritisation based on their level of importance. For example, the analyst might use the main effect measures for X_1 and X_2 based on the analyst distribution $\pi_A(X_i)$. In this case, the analyst's distribution for X_1 and X_2 are $\pi_A(X_1)$ and $\pi_A(X_2)$ respectively. Applying the main effect formula, we will obtain the main effect for X_1 is

$$V_1 = Var_{\pi_A(X_1)}(E(Y|X_1))$$

and the main effect for X_2 is

$$V_2 = Var_{\pi_A(X_2)}(E(Y|X_2)).$$

Based on this measure, we might elicit $\pi_E(X_1)$ if $V_1 > V_2$. This is because the expected reduction in the output variance in learning X_1 is greater than X_2 if we know the true value of X_1 and X_2. However, another question will arise: Is it correct to prioritise inputs using this approach for choosing inputs for elicitation?

In particular, we will not learn the true value of X_i; instead, we just obtain a different distribution for X_i after elicitation from the expert where $\pi_E(X_i)$ is more informative than $\pi_A(X_i)$. Does this difference in the distributions matter? We provide a counter-example to show that main effects do not necessarily rank inputs in order of 'correct' priority for elicitation.

§ 5.3 A counter-example

In this example, we consider a non-linear function with two inputs x_1 and x_2. We created our own function as a model for the counter-example purpose. The model is non-linear by employing the sinus function and is additive where the first term involves x_1 while the second involves x_2. The model has the following form.

$$f(x_1, x_2) = 10 \sin\Big(\frac{10\pi x_1}{2}\Big) + \frac{1}{\sin(2\pi/5)} \sin\Big(- \pi x_2 + \frac{\pi}{2}\Big), \tag{5.1}$$

and we are interested in

$$Y = f(X_1, X_2)$$

with X_1 and X_2 are uncertain. For $i = 1, 2$ let

$$X_i = \theta_i + \epsilon_i \tag{5.2}$$

with ϵ_1 and ϵ_2 are independent, and

$$p(\epsilon_i = -0.1) = p(\epsilon_i = 0.1) = 0.5.$$

In particular, the expert will change our distribution for X_i by telling us the true value of θ_i but the expert will not change our belief about ϵ_i. For $i = 1, 2$, we think that expert will either tell us $\theta_i = 0$ or $\theta_i = 1$, with probabilities

$$p(\theta_i = 0) = p(\theta_i = 1) = 0.5.$$

We use the main effect measures to make the prioritisation of inputs. We consider two types of main effects calculation for the model in the counter-example. The first calculation calculates main effects for X_1 and X_2 using the standard sensitivity analysis, while the second calculates main effects for θ_1 and θ_2.

5.3.1 Main effects for X_1 and X_2

To calculate the main effects for X_1 and X_2, we need to determine the marginal distribution for X_i. We use the list of possible values of X_i, θ_i and ϵ_i in Table 5.1 for obtaining the marginal distribution for X_i.

Table 5.1: Possible values for X_i given θ_i and ϵ_i

X_i	θ_i	ϵ_i
-0.1	0	-0.1
0.1	0	0.1
0.9	1	-0.1
1.1	1	0.1

According to Table 5.1 the conditional probability distribution for $X_i = -0.1$ and $X_i = 0.1$ given $\theta_i = 0$ are 0.5 as there are only two possible outcomes of X_i when we know the true value $\theta_i = 0$. The conditional probability of $X_i = 0.9$ and $X_i = 1.1$ given $\theta_i = 1$ are zero as there are no possibilities of outcome X_i equals those values when we know the true value of θ_i is equal to one. Thus, we obtain the probability for each value for inputs X_i (which is the analyst's distribution for X_i) as

$$p(X_i = -0.1) = p(X_i = 0.1) = p(X_i = 0.9) = p(X_i = 1.1) = 0.25.$$

Now, we compute the main effect for X_1 and X_2 as below. The main effect for X_1 is computed using the formula

$$\begin{aligned} Var_{X_1}\big[E(Y|X_1)\big] &= Var_{X_1}\big[E(f_1(X_1)|X_1)\big] + Var_{X_1}\big[E(f_2(X_2)|X_1)\big] \\ &= Var_{X_1}\big[E(f_1(X_1)|X_1)\big] \\ &= Var_{X_1}\big[f_1(X_1)\big] \\ &= E\big[f_1(X_1)^2\big] - E\big[f_1(X_1)\big]^2. \end{aligned}$$

In this case,

$$\begin{aligned} E\big[f_1(X_1)^2\big] &= \sum_{X_i \in \mathrm{R}_{X_1}} f_1(X_1)^2 \cdot 0.25 \\ &= 0.25 \cdot \big[f_1(-0.1)^2 + f_1(0.1)^2 + f_1(0.9)^2 + f_1(1.1)^2\big] \\ &= 100, \end{aligned}$$

and $E\big[f_1(X_1)\big] = 0$. Therefore, the main effect for X_1 is

$$V_1 = Var_{X_1}\big[E(Y|X_1)\big] = 100$$

and the main effect for X_2 is

$$V_2 = Var_{X_2}\big[E(Y|X_2)\big] = 1.$$

The main effect for input X_1 is bigger than X_2 ($V_1 > V_2$), indicating the expected variance reduction in the output if we learn the true value of X_1 is much bigger than the one of learning X_2. Using this result, we will choose X_1 to be elicited by the expert rather than X_2.

5.3.2 Main effects for θ_1 and θ_2

The second way for making prioritisation is by calculating the main effects for the parameters θ_1 and θ_2.The main effects for θ_1 and θ_2 are calculated as below

$$\begin{aligned} Var_{\theta_1}\big[E(Y|\theta_1)\big] &= Var_{\theta_1}\big[E(f_1(X_1)|\theta_1)\big] + Var_{\theta_1}\big[E(f_2(X_2)|\theta_1)\big] \\ &= Var_{\theta_1}\big[E(f_1(X_1)|\theta_1)\big]. \end{aligned}$$

Let $\big[E(f_1(X_1)|\theta_1)\big] = h(\theta_1)$, and we need to calculate

$$Var_{\theta_1}\big[h(\theta_1)\big] = E[h(\theta_1)^2] - (E[h(\theta_1)])^2.$$

Since $p(\theta_1 = 0) = p(\theta_1 = 1) = 0.5$ then

$$\begin{aligned} E[h(\theta_1)^2] &= \sum_{\theta_1 \in R_{\theta_1}} h(\theta_1)^2 p(\theta_1) \\ &= h(\theta_1 = 0)^2 p(\theta_1 = 0) + h(\theta_1 = 1)^2 p(\theta_1 = 1) \\ &= 0.5 \cdot [h(\theta_1 = 0)^2 + h(\theta_1 = 1)^2] \\ &= 0.5 \cdot \Big[E\big(f_1(X_1)|\theta_1 = 0\big)^2 + E\big(f_1(X_1)|\theta_1 = 1\big)^2\Big] \\ &= 0.5 \cdot \Big[\Big(f_1(-0.1)p(\theta_1 = 0) + f_1(0.1)p(\theta_1 = 0)\Big)^2 + \Big(f_1(0.9)p(\theta_1 = 1) + f_1(1.1)p(\theta_1 = 1)\Big)^2\Big] \\ &= 0.5\big[(-10 \times 0.5 + 10 \times 0.5)^2 + (10 \times 0.5 - 10 \times 0.5)^2\big] \\ &= 0. \end{aligned}$$

and $E[h(\theta_1)] = 0$. Therefore, the main effect for θ_1 is

$$Var_{\theta_1}[E(Y|\theta_1)] = 0$$

and the main effect for θ_2 is

$$Var_{\theta_2}[E(Y|\theta_2)] = 1.$$

According to these results, the expected reduction in output variance from eliciting the expert's belief about θ_1 is zero. This means that the analyst will not ask the expert to elicit the distribution for X_1; instead, the expert will be asked to elicit the distribution for X_2.

The conclusion now is contradictory with the previous main effects measures, V_1 and V_2. We have shown through the counter-example that if $V_1 > V_2$ does not necessarily imply better to elicit $\pi_E(X_1)$. Therefore, the direct application of, e.g. main effects in the sensitivity analysis using analyst's distribution may not be appropriate.

Previously, we considered two distributions, the analyst's distribution and the expert's distribution. The analyst will specify a "crude" distribution for input X_i while the expert will give a "smooth" distribution. The expert's distribution for a particular input X_i is likely to differ from the analyst's distribution, so the expert elicitation changes the distribution for input X_i. There is also literature that considers changing the input distribution on sensitivity analysis results, and the concept is called robust sensitivity analysis. We will now review some of this literature to investigate if anything might be applied to our problem.

§ 5.4 A review on robust global sensitivity analysis methods

In general, we denote the distribution for input X_i as $P(X_i)$. Changing $P(X_i)$ might change $Var(E(Y|X_i))$ but not the $\frac{Var(E(Y|X_i))}{Var(Y)}$ because if $Var(E(Y|X_i))$ changes, then $Var(Y)$ can change too. Thus we would be more interested in $Var(E(Y|X_i))$ when using robust sensitivity analysis. How might Robust sensitivity analysis method helpful in terms of the following points?

(a) Useful methods for modifying $P(X_i)$, or describing range of distributions for X_i.

(b) Useful measures of robustness.

(c) Useful computational tools for doing GSA with lots of different input distributions.

The following are a review of some robust GSA methods in the literature.

5.4.1 Gao & Bryan method

Gao & Bryan (2016) proposed robust sensitivity indicators based on concepts from decision theory, such as maximax, weighted average, minimax regret, and the limited degree of

confidence. In the first stage of the analysis, GSA was carried out to analyse a Land-Use Trade-Offs (LUTO) model with 50 inputs. The GSA is conducted for a scenario with a specific setting for the input parameter distributions. The influential inputs are identified using the main effect index and the total effect index. The GSA is then repeated for the other scenarios with different input parameter distributions. The result of the first stage of the analysis is a small number of inputs influential to the output for each Scenario. The subsequent analysis stage is to apply the robust sensitivity indicators: maximax, weighted average, minimax regret and the limited degree of confidence. These indicators are calculated based on the total effect indices to determine which inputs are robust regarding their effect on the output for all scenarios.

Example 5.1. To illustrate minimax regret, let us consider the following model.

$$y = x_1 + x_2 + x_1 x_2$$

where input x_i are uncertain and have three scenarios as follows:

(a) $X_1 \sim N(0, 1)$ and $X_2 \sim U(0, 1)$

(b) $X_1 \sim N(0, 2)$ and $X_2 \sim U(0, 4)$

(c) $X_1 \sim N(0, 4)$ and $X_2 \sim U(0, 12)$

The objective of the minimax regret is to determine which choice of input distribution is the most robust. The regret is computed using the total effect index. The total effect indices for each Scenario is shown in Table 5.2.

Table 5.2: Total effect indices for each Scenario

Scenario	S_{T1}	S_{T2}
a	0.9656	0.0692
b	0.9691	0.1559
c	0.9883	0.2081

A robust sensitivity indicator, minimax regret, is applied to the example. The regret is calculated based on the absolute difference of the total effect indices from two scenarios (Scenario a and b) as below.

$$r_{a,b} = |S_{Ti,a} - S_{Ti,b}| \tag{5.3}$$

A scenario which has the lowest maximum regret value is chosen using the following argument

$$\min_{a}\Big(\max_{b}(r_{a,b})\Big) \tag{5.4}$$

and the calculation is indicated using Tables 5.3 and 5.4 below.

Table 5.3: Regrets for total effect index X_1

	Scenario a	Scenario b	Scenario c	max
Scenario a	0.0000	0.0035	0.0227	0.0227
Scenario b	0.0035	0.0000	0.0192	0.0192
Scenario c	0.0227	0.0192	0.0000	0.0227

Table 5.4: Regrets for total effect index X_2

	Scenario a	Scenario b	Scenario c	max
Scenario a	0.0000	0.0867	0.1389	0.1389
Scenario b	0.0867	0.0000	0.0522	0.0867
Scenario c	0.1389	0.0522	0.0000	0.1389

According to Table 5.3, the minimum of the maximum values of regrets for X_1 is 0.0192 which indicates that Scenario b satisfies the minimax regret indicator. Similarly, for input X_2 in Table 5.4, the minimax regret indicator chooses Scenario b. In other words, Scenario b is chosen for input X_1 and X_2 to minimise the regret (opportunity loss) of accepting total effect indices that does not occur. As a result, the total effect indices obtained from Scenario b, are selected to represent the robust sensitivity indicators and are written as $S^{reg}_{RT_1} = 96.91\%$ and $S^{reg}_{RT_2} = 15.15\%$.

If different scenarios are the minimax regrets for input X_1 and X_2, then we can use these scenarios to perform robust sensitivity analysis. There might be a case when we can choose more than one scenario because the minimax regrets for input X_1 and X_2 are similar for these scenarios. If the minimax regret for X_1 lead to a different choice of scenario than for X_2, one might set some other scenarios with different choices of the input distribution and repeat the analysis. If we cannot change the distribution of inputs and define other scenarios for some reason, then we do not have scenarios for robust sensitivity analysis. In this case, the scenario might be provided by experts or it may not realistic if the distribution of inputs is changed.

The similarity of Gao & Bryan method with the elicitation is it considers the different

possibilities of input distributions. In the elicitation, we also consider two different distributions, the analyst's distribution ($\pi_A(X_i)$) and the expert's distribution ($\pi_B(X_i)$).

The difference between Gao & Bryan method and the elicitation is, in the elicitation, if $\pi_A(X_i)$ and $\pi_E(X_i)$ are both known, we will choose $\pi_E(X_i)$. In the initial analysis, we begin with the $\pi_A(X_i)$ while $\pi_E(X_i)$ is unknown.

In conclusion, if we propose possibilities for $\pi_E(X_i)$, we could use regret (perhaps based on Var(Y) rather than total effect) to measure switching from $\pi_A(X_i)$ to $\pi_E(X_i)$. We would need our own method for proposing possibilities for $\pi_E(X_i)$. For instance, in Example 5.1, suppose the analyst's distribution is Scenario (a). The expert will either give us (b) or (c). Given the distribution possibilities, we can use maximax regret to choose either X_1 or X_2 as the most important elicitation input.

5.4.2 Lacirignola et al. method

Lacirignola et al. (2017) developed a methodology to inspect the sensitivity of global sensitivity analysis. They considered a model called the life cycle assessment (LCA) model. The inputs of the model were analysed by considering different types of distributions. The GSA was performed using a one factor at a-time approach for repeating the analysis. In the first iteration of GSA, the inputs are set into their baseline distributions, and the Sobol's indices were computed. The inputs are then ranked based on the largest index to the lowest. The second iteration of GSA was carried out by changing the first input by an alternative distribution while other inputs were set to their baseline distributions. The Sobol's indices were recalculated, and the inputs were re-ranked according to the new indices. This analysis is repeated for other inputs, and the sensitivity of the input ranking based on different input distribution sets is investigated. The key parameters are determined based on the ranking. The key parameters are inputs that have the highest-ranking positions. The number of key parameters is determined by summing the Sobol's indices of the inputs, and it exceeds the required percentage of the overall variability of the output, for instance, 60%. The inputs that are not included in the key parameter are called the non-key parameters. The effect of the modification in the input distributions is observed during the GSA iteration. Suppose the change of the input distribution affects the input status from key parameters to non-key parameters. In that case, the input influences the identification of the key parameters.

The focus of the Lacirignola et al. method was to investigate whether the importance of input will change as we change the input distribution. On the other hand, our situation

is that input could be consistently important under $\pi_A(X_i)$ or $\pi_E(X_i)$, but changing from $\pi_A(X_i)$ to $\pi_E(X_i)$ has little effect. Therefore, it does not necessarily give priorities for elicitation.

5.4.3 Borgonovo et al. method

Borgonovo et al. (2018) investigate the impact of relaxing the unique distribution assumption on the classical functional ANOVA expansion. As we know that functional ANOVA is used based on the unique probability distribution assumption for factors or inputs. However, this assumption does not always hold due to some situations, such as insufficient data, different expert opinions, or measurement errors. Therefore, there might be some possible set of the probability distribution for inputs that the analyst can apply. Suppose that analyst have w plausible set of probability distribution for inputs, $\mathcal{P} = (P^1_{\mathbf{X}}(x), ..., P^w_{\mathbf{X}}(x))$, where $\mathbf{X} = (X_1, ..., X_p)$. As a result, the associated functional ANOVA is potentially different, and the sensitivity measures produced from each functional ANOVA are evaluated for robustness. There are two ways used for evaluating the sensitivity measures based on the probability input set, called without-prior path and with-prior path. In the without-prior path, the model is evaluated for each $P^i_{\mathbf{X}}(x)$, and the sensitivity measures are obtained. Whereas, in the second path, the analyst assigns a prior for $P^i_{\mathbf{X}}(x)$. The robustness of sensitivity analysis is discussed in these two paths in the context of sensitivity analysis settings such as factor prioritisation. The setting for the robust SA in without-prior path is *We are asked to bet a model input if fixed to its true value, would lead to the greatest expected reduction in the variance of the output under all* p^m *in* $\mathcal{P}$. A robust factor prioritisation setting searches for model inputs that satisfy the following condition:

$$\underline{S}_i > \bar{S}_j, \text{ for all } j \neq i, j = 1, ..., n, \tag{5.5}$$

where

$$\underline{S}_i = \sup_{p^m \in \mathcal{P}} S_i^{p^m} \quad \text{and} \quad \bar{S}_j = \inf_{p^m \in \mathcal{P}} S_i^{p^m}. \tag{5.6}$$

In the with-prior path, the analyst can average the sensitivity indices over the probability measures in $\mathcal{P}$.

There is slight suitability in the perspective of the with-prior approach with our concept. The analyst has some sets of input distributions, and because one is uncertain about the inputs, then a prior is assigned. In our case, the posit distributions are obtained from experts, and as the analyst does not know what distribution that expert will provide, then the prior is assigned to the parameter of the input distribution. The difference would

be the effect of a different set of input distributions investigated through the resulting sensitivity indices in Borgonovo et al. (2018). While in our cases, the effect of changes in individual input distribution from the analyst distribution $\pi_A(x)$ to the expert distribution $\pi_E(X)$ becomes our concern.

5.4.4 Meynaoui et al. method

Meynaoui et al. (2019) employed the Hilbert-Schmidt Information Criterion (HSIC) sensitivity analysis measures for performing GSA and developed it to second-level GSA (GSA2). The HSIC is estimated using a sample generated from a probability distribution of inputs, not their prior distribution. This sampling distribution is called the alternative law or the modified law. The GSA is not repeated in this method, but inputs are generated using the importance sampling.

We do not provide details of this sensitivity as HSIC is too technical to be discussed in this section. We note that the Meynaoui et al. method measure can be re-computed for different input distributions using importance re-weighting. However, importance re-weighting needs two distribution functions to be similar, and it will not work if the two distributions are different in their supports. In our cases of elicitation, importance re-weighting is unlikely to help with the computation if $\pi_A(X_i)$ and $\pi_E(X_i)$ are very different.

5.4.5 Hart & Gremaud method

The method proposed by Hart & Gremaud (2019*a*) and Hart & Gremaud (2019*b*) aimed at quantifying the robustness of Sobol's indices to changes in the input distributions. Let us assume a baseline distribution $\varphi(x)$, and the alternative distributions are not directly specified. A Frechet derivative of the Sobol index is derived with the input distribution evaluated at $\varphi(x)$. Next, we find that the direction of perturbation maximises the derivative. Afterwards, compute the Sobol's index at a perturbed input distribution some distance along this direction. There is a restriction on the perturbation as follows. Suppose the baseline distribution for d inputs is represented as $\varphi = (\varphi_1, \ldots, \varphi_d)$ and there is another function $\psi = (\psi_1, \ldots, \psi_d) \in V$ with $||\psi - \varphi||_V \leq 1$. In addition, ψ_i, is a function with the following norm

$$\| \psi_i \|_{V_i} = \| \frac{\psi_i}{\varphi_i} \|_{L^\infty(\Omega_i)} \tag{5.7}$$

$$= \sup_x \Big\{ \frac{\psi_i(x)}{\varphi_i(x)} \mid x \in \Omega_i \Big\} \tag{5.8}$$

In other words, $\| \psi_i \|_{V_i}$ describes how far ψ_i from φ_i. If $\| \psi - \varphi \|_V \leq 1$ then

$$\sup_x \left\{ \frac{\psi_i(x)}{\varphi_i(x)} - \frac{\varphi_i(x)}{\varphi_i(x)} \mid x \in \Omega_i \right\} = 1$$
$$\sup_x \left\{ \frac{\psi_i(x)}{\varphi_i(x)} - 1 \mid x \in \Omega_i \right\} = 1$$
$$\sup_x \left\{ \frac{\psi_i(x)}{\varphi_i(x)} \right\} = 2,$$

so $\psi_i(x) \leq 2\varphi_i(x) \, \forall x$. The implication for the alternative distribution due to the restriction is

$$0 \leq \psi(x_i) \leq 2\varphi_i(x).$$

For an illustration, let the baseline distribution, $\varphi(x_i)$, is a uniform distribution and the alternative distribution, $\psi(x_i)$ is also uniform. The maximum possible values of $\psi(x_i)$ are twice as $\varphi(x_i)$ as shown in Figure 5.2.

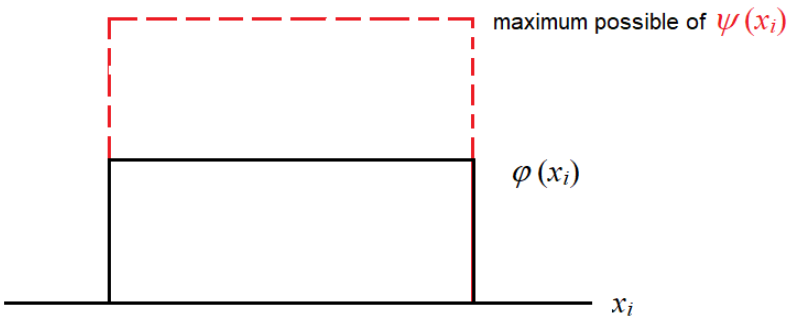


Figure 5.2: An illustration of alternative uniform distribution

The Hart & Gremaud method is useful for constructing perturbations of $\pi_A(X_i)$ but may not be appropriate for considering large changes from $\pi_A(X_i)$ to $\pi_E(X_i)$ due to the maximum constraint of distribution $\psi(x_i)$. The large changes are likely to happen as the analyst does not know about the expert's distribution, so our emphasis here is on measuring differences between distributions.

5.4.6 Gauchy et al. method

Another method was proposed by Gauchy et al. (2020). This method considered a general uncertainty quantification for $f(\mathbf{X})$, e.g. 95^{th} percentile of $f(X)$. The method started with the probability distribution $P(\mathbf{X})$ and we wish to consider perturbation of $P(\mathbf{X})$.

For example, if $P(\mathbf{X}) = P_1(X_1)P_2(X_2)$ with $X_1 \sim N(m, v)$ and $X_2 \sim Beta(a, b)$. What would equivalent perturbation of $P_1(X_1)$ and $P_2(X_2)$ be?

Gauchy et al. (2020) proposed to use a Fisher distance for comparing the distance between the two distributions, the initial distribution and the perturbed distribution. For the robustness measure for the uncertainty quantification or sensitivity measure of interest, they suggest finding the maximum and the minimum values overall distribution that is the same Fisher distance from $P(\mathbf{X})$.

The method of Gauchy et al. considers the small changes in the distribution of X_i by the perturbation parameter δ. The perturbation will not change the distribution of $P(\mathbf{X})$; instead, it will change the parameters of the input distribution. In elicitation, the analyst's distribution may change to another distribution when experts elicit inputs. Thus, the Gauchy et al. (2020) method is not suitable to describe the switching distribution between $\pi_A(X_i)$ to $\pi_E(X_i)$.

§ 5.5 Conclusion

To sum up, the methods presented above have their values in the robustness analysis. In most papers, the uncertainty in the input distributions was investigated using small changes in their probability distributions called perturbation. In our case of prioritising inputs for elicitation, we need a method that considers significant changes in the input distributions from the analyst's distribution to the expert's distribution.

Chapter 6

Sensitivity Analysis for Expert Judgment

§ 6.1 Introduction

In Chapter 5, we introduced the problem of prioritising elicitation. Now, we develop a novel model for expert judgment/opinion. The outline of this chapter is as follows. Section 6.2 provides the current context of our models for expert judgement. Section 6.3 reviews some models for the expert opinion. Section 6.4 proposes a new model for expert judgement called The Beta Model. Section 6.5 discusses an example of the importance of expert uncertainty in the Beta model. Section 6.6 provides the computational method for computing sensitivity measures in the Beta Model, and Section 6.7 explores factors involved in the computational method through some numerical experiments.

§ 6.2 Models for expert judgement

The previous chapter introduced two types of subjective probabilities: the analyst's distribution, π_A and the expert's distribution, π_E. The analyst's distribution is a distribution specified by an analyst for describing the uncertainty of an input in a mathematical model or a computer model. π_A might be non-informative as the analyst has limited knowledge about uncertain quantities in the model. In contrast, the expert's distribution is a distribution elicited by experts based on their knowledge and expertise. The expert's distribution is more informative as it is elicited based on the expert's experience and bet-

ter knowledge of uncertain inputs. We consider these two types of distributions as they have advantages and disadvantages, which has been discussed in Chapter 5. However, the expert's distribution is still superior to the analyst's distribution. If available, the analyst will use the expert's distribution instead of his or her own distribution for analysing the model.

As discussed in the previous chapter, eliciting all input distributions in a model is not feasible. Therefore, we proposed input prioritisation for elicitation. The input is a priority for elicitation if changing from $\pi_A(X_i)$ to $\pi_E(X_i)$ has a large effect on $P(Y)$. Inputs that are not the elicitation's priority will use the analyst's distribution to describe their uncertainties. Thus, to prioritise the input for elicitation, the analyst needs to know about $\pi_E(X_i)$. The issue here is that the analyst will not know about $\pi_E(X_i)$; hence, $\pi_E(X_i)$ is uncertain. The analyst needs to consider what $\pi_E(X_i)$ could be. The analyst needs a model for $\pi_E(X_i)$ which is called a model for expert opinion.

A few models for expert opinion have been proposed in the literature, and the development of such models was also specific for certain settings. In the next section, we present some available models for expert opinion.

§ 6.3 Models on expert opinion

This section reviews two models for expert opinion from the literature: The Lindley, Tversky, and Brown model (LTB model) and The Gelfand-Mallick-Dey (GMD) model. We want to investigate whether the models are practical compared with our expert judgement model.

6.3.1 The LTB model

Lindley et al. (1979) proposed a model to reconcile incoherent probability assessment: the case when the probabilities are not consistent with the probability laws; and their approach was developed in French (1980). We call the model of Lindley et al. (1979) as Lindley, Tversky, and Brown (LTB) model. The model is described as follows.

Suppose there is a sequence of uncertain events $A = (A_1, A_2, \ldots, A_m)$. A person has an underlying coherent probability distribution $\pi(A)$, but instead of reporting $\pi(A)$, the person reports a distribution $q(A)$, which may not be coherent. The model contains three elements: A, $\pi(A)$, and $q(A)$. In this case, π is the 'true' probability which is not directly

known, while q is observable or directly stated by the person.

Next, consider an investigator whose tasks are to reconcile the person's stated values q and assess π. From the investigator's point of view, A, π, and q are uncertain and can be represented as a joint probability distribution $p(A, \pi, q)$. The joint probability distribution can be viewed in three stages: First, $p(A)$ and then $p(\pi|A)$, lastly $p(q|\pi, A)$. Assuming that the conditional distribution q given π does not depend on A, the last expression of probability can be written as $p(q|\pi)$.

For simplicity, suppose A is a single event so that a single probability specifies $\pi(A)$. We now define π and q as the person's 'true' and stated log-odds for the event A.

Lindley et al. (1979) model consist of three elements: $p(A)$, $p(\pi|A)$ and $p(q|\pi)$. In this case, $p(A)$ is the investigator's probability about the event A, $p(\pi|A)$ is the investigator's opinion of the person's knowledge of the event, and $p(q|\pi)$ is the investigator's opinion of the person as a probability assessor.

The first probability, $p(A)$, is the standard probability for an event A will occur where $0 < p(A) < 1$. For the second probability, $p(\pi|A)$, is assumed to be normal in log-odds with mean μ_1 and variance τ^2. If event A does not occur then the probability $p(\pi|\bar{A})$ is assumed to be normal with log-odds μ_2 and variance τ^2. The last distribution, $p(q|\pi)$, is assumed to be normal $N(\pi, \sigma^2)$ in log-odds, with mean π and variance σ^2. Therefore the LTB model can be written as below.

i. $p(A) = \alpha$, where $0 < \alpha < 1$,

ii. $\log\big(\pi|A\big) \sim N(\mu_1, \tau^2)$, $\log\big(\pi|\bar{A}\big) \sim N(\mu_2, \tau^2)$,

iii. $\log\big(q|\pi\big) \sim N(\pi, \sigma^2)$.

The authors go on to consider two cases: one in which the investigator derives $p(\pi|q)$ and another in which the investigator derives $p(A|q)$. For example, using the Bayes' theorem we obtain

$$\begin{aligned} p(\pi|q) &= \frac{p(q|\pi)p(\pi)}{\sum_\pi p(q|\pi)p(\pi)} \\ &= \frac{p(q|\pi)\big(\alpha p(\pi|A) + (1-\alpha)p(\pi|\bar{A})\big)}{\alpha p_1(q) + (1-\alpha)p_2(q)} \end{aligned}$$

where $p_i(q) = \sum_\pi p(q|\pi)p_i(\pi)$. According to standard normal theory, we will obtain

$$\pi|q \sim N((wq + (1-w)\mu_i), w\sigma^2)$$

where $w = \tau^2/(\sigma^2 + \tau^2)$. We will also obtain $p_i(q) \sim N(\mu_i, \sigma^2 + \tau^2)$. Consequently, $p(\pi|q)$ is a mixture of normal distributions.

Applying the LTB model

In this section, we want to investigate how to use the LTB model to get a model for an expert distribution $\pi_E(X_i)$. According to the LTB model, we would need to make a partition of the sample space of input X_i, A, into m parts as

$$A = (A_1, \ldots, A_m),$$

where A_i is the i-th partition of the sample space of X_i.

Next, we might assume for simplicity that $\pi_E(X_i)$ is a piece-wise uniform. The analyst considers probabilities for each partition in A, $P(A_1), \ldots, P(A_m)$ and then needs to consider a model for

$$\pi_E(A_1), \cdots, \pi_E(A_m)|X_i, \quad \text{for} \quad i = 1, \ldots, p,$$

where p is the number of input. This is a difficult modelling exercise as the analyst should think and specify many probabilities for a single input. Therefore, we develop an alternative model which is more practical.

6.3.2 The GMD model

Gelfand et al. (1995) used a partial probabilistic specification to elicit a distribution for an uncertain quantity θ. In this approach, an expert could provide probabilities for a small collection of disjoint complete intervals in the domain of θ or a small set of quantiles for the distribution of θ. They proposed using a family of mixtures of beta distributions to model experts' opinions.

Suppose we are interested in an unknown quantity $\theta \in \Theta$. The domain Θ is partitioned onto k sets and determined by points such

$$\alpha_0 < \alpha_1 < \ldots < \alpha_k$$

where $\alpha_0 = \inf\{\theta \in \Theta\}$ and $\alpha_k = \sup\{\theta \in \Theta\}$ and define $I_j = (a_{j-1}, a_j)$.

The expert can provide either a set of probabilities $\mathbf{p}^T = (p_1, \ldots, p_k)$ or a set of quantiles $\mathbf{q}^T = (q_1, \ldots, q_k)$ for θ. We will focus on the former case. In this case, p_j is the expert's

probability for θ will be in I_j. Therefore, we suppose that $p_1 = P(\alpha_0 < X_i < \alpha_1), p_2 = P(\alpha_1 < X_i < \alpha_2), \ldots, p_k = P(\alpha_{k-1} < X_i < \alpha_k)$.

Gelfand et al. (1995) scaled θ_i into an interval (0, 1). A mixture of beta distributions can arbitrarily well approximate a distribution on (0, 1):

$$\theta_i | w_i \sim \sum_{l=1}^{r} w_l \text{Beta}(\alpha_l, \beta_l) \tag{6.1}$$

where w_l is a random weight for each Beta distribution, α_l and β_l are Beta distribution parameters for each mixand. The α_l and β_l parameters are fixed, and only the weights which are uncertain with density $f_w(r)$.

The advantage of the model is that it can model multiple distributions from experts in the domain of θ. The disadvantage of the model is, however, the number of mixands r required to induce $\mathbf{p}$ is quite large. Gelfand et al. (1995) found that $r = 10$ is sufficient to approximate the density of $f(\mathbf{p}|\theta)$. Consequently, we need to cope with high-dimensional mixture density.

This approach may be easier to apply than the LTB model to our setting. However, specifying the analyst's prior on the weights may be complex. Therefore, we will develop a more straightforward method in this chapter.

§ 6.4 The Beta Model

We propose a new model for the expert opinion, which we call the Beta Model. As mentioned previously, it is implausible that an expert will provide a true value for an input distribution. Instead, the expert will provide an informative distribution for X_i. What might the expert tell us?

Experts could be reasonably certain or somewhat uncertain about their beliefs. Eliciting an expert's input distribution X_i would result in $X_i \sim P(X_i|\theta_i)$, where θ_i is an unknown parameter of expert's distribution. The analyst is uncertain about θ_i because the analyst does not yet know what the expert believes about X_i. Thus, the analyst could specify a prior for the parameter θ_i, denoted by $P(\theta_i)$.

Now we present how the Beta model is developed. Initially, the analyst specifies lower and upper bounds, L and U, respectively. The analyst considers that experts will provide means anywhere between the lowest bound, L, and the highest bound, U. Experts might

also provide means within a lesser range $(L+\epsilon, U-\epsilon)$, with a small number $\epsilon > 0$. See Figure 6.1 for the illustration.

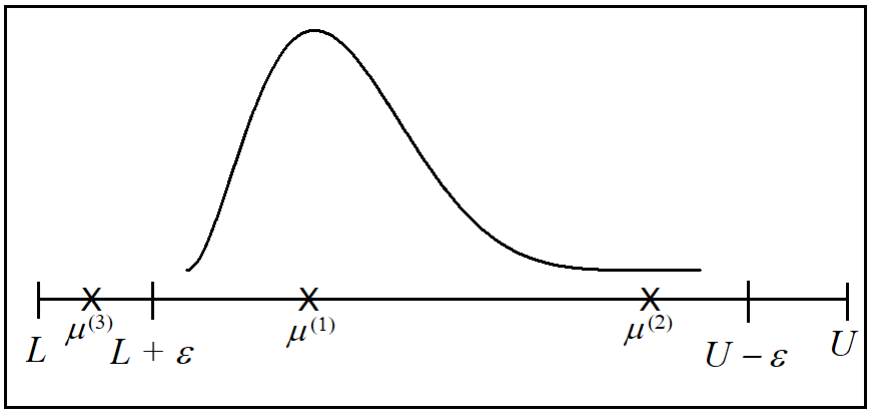


Figure 6.1: A possible range of expert's means for uncertain inputs X_i. Expert might give means anywhere between L and U. Analyst expects that expert's means will fall within the interval of $[L+\epsilon,\ U-\epsilon]$, for instance $\mu^{(1)}$ and $\mu^{(2)}$. Expert's means will have a small probability to be outside of the interval, for example $\mu^{(3)}$.

We propose The Beta model to be a distribution model for expert judgment. The Beta model assumes that an expert will provide a distribution for input $X_i \mid \alpha_i, \beta_i \sim \text{Beta}(\alpha_i, \beta_i)$ so that $\theta_i = (\alpha_i, \beta_i)$. The Beta distribution is chosen as an expert's model as it can reflect any possible curve shapes experts might provide. The scale and the shape parameters can control the density function of the distribution in a large variety of forms. Therefore, it can represent symmetric distributions, skewed distributions (positively skewed and negatively skewed distributions), exponential distributions, and heavy-tailed distributions. Other distributions, such as gamma or log-normal distribution have flexible curves as well. However, these distributions are only for positively-skewed distributions and can not represent expert opinion when it fits with a negatively-skewed distribution.

Now we introduce the beta model in detail as follows. The mean of the beta distribution is

$$\mu_i = \mathbb{E}[X_i|\alpha_i, \beta_i] = \frac{\alpha_i}{\alpha_i + \beta_i}. \tag{6.2}$$

Analyst specifies priors for μ_i as α_i and β_i are uncertain. Analyst might suppose that μ_i is scaled to uniform distribution as follows

$$\mu_i \sim U[L+\epsilon, U-\epsilon] \tag{6.3}$$

where ϵ is a small positive value.

Next, we define another parameter for the Beta model. The variance of Beta distribution, $\text{var}(X_i|\alpha_i, \beta_i)$ is related to the sum of $\alpha_i + \beta_i$. Let us suppose that analyst is uncertain about order of magnitude of $\alpha_i + \beta_i$. We define

$$\varphi_i = \alpha_i + \beta_i,$$

and suppose

$$log_{10}\varphi_i \sim U[a, b] \tag{6.4}$$

where a and b are the lower bound and the upper bound respectively.

If we want to assess the value of learning X_i, we need to obtain the predictive distribution of X_i from the Beta Model. The predictive distribution for X_i is

$$P(X_i) = \int_{\varphi_i,\mu_i} P(X_i|\varphi_i, \mu_i)P(\varphi_i, \mu_i)d\varphi_i\mu_i, \tag{6.5}$$

where $P(X_i|\varphi_i, \mu_i)$ is the prior distribution that the expert would provide and $P(\varphi_i, \mu_i)$ is the analyst's prior for the expert's prior distribution.

§ 6.5 Importance of expert's uncertainty

In the Beta model, we assess the value of eliciting beliefs about X_i by measuring the importance of μ_i and logφ_i. However, does uncertainty about the expert's uncertainty, in this case, log φ_i, matter? Would it be sufficient only to measure the importance of μ_i?

Example 6.1. To investigate the questions, we can illustrate how the Beta model works by comparing the effect of the parameter involved in the main effect computation. Let us consider the following simple function.

$$y = x_1 + x_2.$$

The inputs are uncertain and denoted by X_1 and X_2, and we are interested in

$$Y = X_1 + X_2$$

Using the Beta Model, we will elicit X_i by the Beta distribution approach as follows.

$$X_i|\alpha, \beta \sim \text{Beta}(\alpha, \beta)$$

with

$$\mu_i = E(X_i) = \frac{\alpha}{\alpha + \beta} \sim U[a, b]$$

and $\varphi_i = \alpha_i + \beta_i$, and $log\varphi_i \sim U[c, d]$. We note that the variance of Beta Distribution is

$$\mathrm{Var}(X_i) = \sigma_i^2 = \frac{\alpha\beta}{(\alpha+\beta)^2 + (\alpha+\beta+1)} = \frac{\alpha\beta}{\varphi^2 + \varphi + 1}$$

We use the variance-based approach to show whether there is a value of learning $\log\varphi_i$ or not. The expected reduction in the output variance for learning $\log\varphi_i$ is

$$\begin{aligned}
Var_{\log\varphi_i}[E(Y|\log\varphi_i)] &= E[(X_1 + X_2|\log\varphi_i)] \\
&= Var_{\log\varphi_i}[E(X_1 + X_2|\log\varphi_i)] \\
&= Var_{\log\varphi_i}[E_{\mu_1,\mu_2}[E(X_1 + X_2|\mu_1, \mu_2, \log\varphi_i)] \\
&= Var_{\log\varphi_i}[E_{\mu_1,\mu_2}(E(X_1) + E(X_2)] \\
&= Var_{\log\varphi_i}(\mu_1 + \mu_2) \\
&= 0
\end{aligned}$$

The result indicates that there is no value of learning $\log\varphi_i$ in this model. Let us consider another model which is non-linear in X_1 as below.

$$Y = X_1^2 + X_2$$

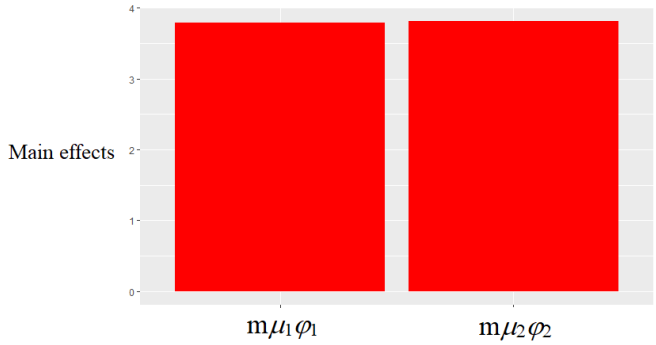
and we calculate the expected reduction in the variance of the output of learning $\log\varphi_i$ as

$$\begin{aligned}
Var_{\log\varphi_i}[E(Y|\log\varphi_i)] &= E[(X_1^2 + X_2|\log\varphi_i)]] \\
&= Var_{\log\varphi_i}[E(X_1^2 + X_2|\log\varphi_i)]] \\
&= Var_{\log\varphi_i}[E_{\mu_1,\mu_2}[E(X_1^2 + X_2|\mu_1, \mu_2, \log\varphi_i)]] \\
&= Var_{\log\varphi_i}[E_{\mu_1,\mu_2}[(E(X_1^2) + E(X_2)]] \\
&= Var_{\log\varphi_i}[E_{\mu_1,\mu_2}[Var(X_1) + E(X_1)^2 + E(X_2)]] \\
&= Var_{\log\varphi_i}(\sigma_1^2 + \mu_1^2 + \mu_2) \\
&= Var_{\log\varphi_i}\Big(\frac{\alpha\beta}{\varphi^2 + \varphi + 1}\Big) \\
&\neq 0.
\end{aligned}$$

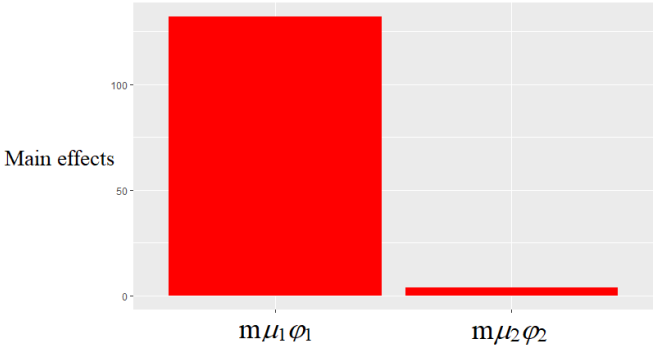
We can conclude that there is a value of learning $\log\varphi_i$ if the model is non-linear. Thus, the uncertainty in the $\log\varphi_i$ can matter if the model is non-linear. In this case, it would not be sufficient to measure the importance of μ_i only and it requires to measure the importance of $\log\varphi_i$.

We continue this example and now consider which distribution is better to be elicited, $\pi_E(X_1)$ or $\pi_E(X_2)$ for both models? In order to do so, we will focus on uncertain quantities

(μ_i, φ_i), where $i = 1, 2$. Eliciting input X_i corresponds to learning (μ_i, φ_i). We use main effects to evaluate the importance of (μ_i, φ_i). We use a Monte Carlo simulation for computing the main effects for $m_{\mu_i \varphi_i}$ to scrutinise the answer. For this purpose, we use the number of sample 100,000 and we made assumption for the input distributions. X_1 is normally distributed (2, 2) while X_2 has a Uniform distribution (1, 5). The main effects $m_{\mu_i \varphi_i}$ is computed for the model $y = x_1 + x_2$ and $y = x_1^2 + x_2$. The results are presented in Figure 6.2.



(a) Main effects for μ_1, φ_1 and μ_2, φ_2 for a linear model $y = x_1 + x_2$.



(b) Main effects for μ_1, φ_1 and μ_2, φ_2 for a non-linear model $y = x_1^2 + x_2$.

Figure 6.2: Main effects for learning whether the distribution for X_1 or X_2 that is better to be elicited. Figure (a) displays main effects for the linear model and Figure (b) Main effects for non-linear model.

According to Figure 6.2b, the distribution of X_1 is better to be elicited rather than X_2 as the main effects for μ_1, φ_1 is bigger than μ_2, φ_2.

We have considered a simple example to investigate whether uncertainty in learning log

φ_i matters in linear and non-linear models using main effects. We have also shown which inputs should be elicited from both models using the main effect measures.

Nevertheless, as shown in Chapter 3, a small main effect does not imply that input is unimportant. Let us look back to the conclusion from Scenario 1 in Chapter 3 in Table 3.2. We can say that input is unimportant in the case of independent input if total effects are zero instead of using main effects.

We now consider another measure used to investigate appropriate inputs to be elicited. We choose to consider Borgonovo's δ. The concept of δ has been discussed in Chapter 3, and now we consider the computation.

In order to assess the value of learning σ_i^2 using δ index, we need to compute the index for X_i, μ_i, σ_i, and $\mu_i\sigma_i$ and compare their values. If $\delta(\sigma_i^2)$ is relatively big compare with $\delta(\mu_i)$ and $\delta(\mu_i\sigma_i^2)$, then it is valuable to assess the value of σ_i^2. We want also to look at the comparison between $\delta(\sigma_i^2)$ with $\delta(X_i)$ to know whether assessing X_i directly is valuable compare with learning σ_i^2 in its prior distribution.

It is very challenging to calculate the δ mathematically for terms μ_i, σ_i^2, $\mu_i\sigma_i^2$, and X_i. Therefore, we propose to solve it numerically. The algorithm for computing the δ for σ_i^2 is presented as follows.

§ 6.6 Computational Methods

In this section, we consider Borgonovo's δ as a measure to determine which inputs are better elicited. Borgonovo's δ for input X_i is defined as

$$\delta(X_i) = \mathbb{E}_{X_i}\Big[\frac{1}{2}\int \mid f_Y(y) - f_{Y|X}(y|x_i) \mid dy\Big]. \tag{6.6}$$

In our elicitation problem, the elicitation is performed to input that has a large $\delta(\mu_i, \varphi_i)$. Therefore, we need to compute δ for the pair of parameters μ_i and φ_i. The extension of δ from an individual parameter to a group of the parameter is based on Definition 2 in Borgonovo (2007) pg. 775. For two parameters μ_i and φ_i, δ is formulated as

$$\delta(\mu_i, \varphi_i) = \mathbb{E}_{\mu_i\varphi_i}\Big[\frac{1}{2}\int \mid f_Y(y) - f_{Y|\mu_i\varphi_i}(y) \mid dy\Big]. \tag{6.7}$$

In practice, it is difficult to compute $\delta(\mu_i\varphi_i)$ as $f_y(y)$ is unknown and $f_{y|\mu_i\varphi_i}$ is also unknown. Consequently, in the later density, it is difficult to determine how distribution

of Y changes given any possible μ_i, φ_i. The task is now how to integrate over the possible μ_i, φ_i?

Borgonovo (2007) uses five steps to estimate $\delta(X_i)$ in Equation (6.6). The first step is the determination of unconditional density of Y; second step is sampling a value of x_i from $f_{X_i}(x_i)$; the third step is the sampling of conditional distribution of Y given X_i or $f_{Y|X_i}(y)$; the fourth step is the computation of $s(X_i)$, and the last step is the estimation of δ_i from the computed $s(x_i)$. We propose a Monte Carlo approach and the grid-based approach to compute $\delta(\cdot)$.

Monte Carlo algorithm for computing $\delta(\sigma_i^2)$

We notice from (6.6) that the term in the expectation is a function of X_i or $\mathbb{E}[g(X_i)]$. To compute $\delta(\sigma_i^2)$, we replace X_i with σ_i^2 and we obtain

$$\delta(\sigma_i^2) = \mathbb{E}\Big[\frac{1}{2}\int |f_y(y) - f_{y|\sigma_i^2}(y)dy\Big]. \tag{6.8}$$

Thus we can write it as a function $g(\sigma_i^2)$ and the Equation (6.8) reduces to $\delta(\sigma_i^2) = \mathbb{E}[g(\sigma_i^2)]$. Using a numerical simulation, we aim to estimate the expectation of $g(\sigma_i^2)$ using the approximation below.

$$\mathbb{E}[g(\sigma_i^2)] \approx \frac{1}{N}\sum_{j=1}^{N} g(\sigma_{i,j}^2) \tag{6.9}$$

where $\sigma_{i,1}^2, \ldots, \sigma_{i,N}^2$ is a sample from $p(\sigma_i^2)$. The algorithm for computing the estimate of $g(\sigma_i^2)$ is

(i) write a function that compute $g(\sigma_i^2)$

(ii) sample $\sigma_{i,1}^2, \ldots, \sigma_{i,N}^2$ from $p(\sigma_i^2)$

(iii) compute (i) and it will get $g(\sigma_{i,1}^2), \ldots, g(\sigma_{i,N}^2)$

(iv) use (6.9) to estimate $\delta(\sigma_i^2)$.

We now provide a detail explanation for Step (i) in the algorithm. To estimate $g(\sigma_i^2)$, we use the numerical integration using Simpson's rule as

$$g(\sigma_i^2) = \frac{1}{2}\sum_{k=1}^{K} w_k h(y) \tag{6.10}$$

where w_k is the Simpson rule's weight and $h(y_k) =\mid f_y(y_k) - f_{y|\sigma_i^2}(y_k) \mid$. Next, we need to estimate $f_y(y_k)$ and $f_{y|\sigma_i^2}(y_k)$. Since the form of model output distribution could be anything so we cannot use a statistical standard distributions to fit the distribution of the output. Instead we will estimate the density of $f_y(y)$ and $f_{y|\sigma_i^2}$ using the Kernel Density Estimate. We then perform sampling for two functions to get y_k where $k = 1, \ldots, R$. We implement the algorithm for two models in Example 5.1 below.

Example 5.1. Let us consider the following models, a linear and a non-linear model, as an illustrative example to learn σ_i^2

$$Y = X_1 + X_2 \quad \text{and} \quad Y = X_1^2 + X_2$$

where $X_i|\mu_i, \sigma_i^2 \sim N(\mu_i, \sigma_i^2)$, $\mu_i \sim N(2, 2)$, and $\sigma_i \sim U(1, 5)$.

To compute $\delta(\sigma_i)$, the sampling procedure for the model is carried out for the densities $f_y(y_k)$ and $f_{y|\sigma_i}(y_k)$. For the first density function, $f_y(y_k)$, we sample $y_1, \ldots, y_R$ by sampling $\mu_{1,r}$ from $p(\mu_1)$, $\sigma_{1,r}$ from $p(\sigma_1)$ and $X_{1,r}$ from $N(\mu_{1,r}, \sigma_{1,r}^2)$. The similar sampling is also applied for $i = 2$. We sample $\mu_{2,r}$ from $p(\mu_2)$, $\sigma_{2,r}$ from $p(\sigma_2)$, and $X_{2,r}$ from $N(\mu_{2,r}, \sigma_{2,r}^2)$.

Subsequently, we estimate the second function, $f_{y|\sigma_{i,j}^2}(y_k)$ using the kernel density estimate of $\tilde{y_1}, \ldots, \tilde{y_R}$ by sampling $\mu_{1,r}$ from $p(\mu_1)$ and $X_{1,r}$ from $N(\mu_{1,r}, \sigma_{i,j}^2)$. For $i = 2$, we sample $\mu_{2,r}$ from $p(\mu_2)$, $\sigma_{2,r}$ from $p(\sigma_2)$, and $X_{2,r}$ from $N(\mu_{2,r}, \sigma_{2,r}^2)$.

We can now compute $\delta(\sigma_i)$ using the algorithm with $R = 10,000$ and $N = 100$ and we obtain $\delta(\sigma_1) = \delta(\sigma_2) = 0.06$ for the linear model and for the non-linear model $\delta(\sigma_1) = 0.09$ and $\delta(\sigma_2) = 0.03$. From these results, we notice that the expected shift of the two densities $f_y(y)$ and $f_{y|\sigma_i}(y)$ is very small for both the linear and the non-linear model. Thus, the different model forms do not affect the expected shift of the densities. At this stage, we have evaluated the value of $\delta(\sigma_i)$, yet to learn about the value of assessing σ_i we need to compare δ for other parameters μ_i and $\mu_i\sigma_i$ and also δ for input X_i. There are computational issues if we use Monte Carlo simulation to compute all terms needed to learn σ_i. We require many different variable settings for the parameters, i.e. which parameters need to be fixed and which needs to be left vary. Another issue is that the computation time for producing the output for each term is quite a while. If we want to double the number of iterations from the initial value $N = 100$, then it will cause the computation time is also doubled.

We propose to use the grid-based method instead of the Monte Carlo simulation to perform the computation more efficiently. The detail of the Grid-Method is in Section 6.6. If we

use the Grid method, we need to include dummy variables in our computation as the indicator of input with a less important effect. In this case, we added two dummy inputs d_1 and d_2 which are normally distributed (0,1) to the model with weights 0.1. A normal distribution (0,1) is selected because it has a smaller mean and variance than the main inputs that is considered not to influence δ. Therefore, the linear model after the addition of dummy inputs becomes $y = x_1 + x_2 + 0.1d_1 + 0.1d_2$ and the nonlinear model after the addition of the dummy inputs become $y = x_1^2 + x_2 + 0.1d_1 + 0.1d_2$.

The result of δ for all terms and the dummy variables for the linear model $y = x_1 + x_2 + 0.1d_1 + 0.1d_2$ is shown in Figure 6.3.

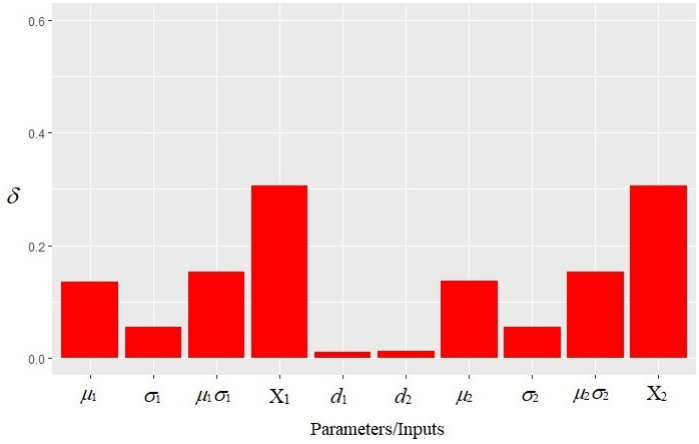


Figure 6.3: Delta indices for learning σ_i in the linear model $y = x_1 + x_2 + 0.1d_1 + 0.1d_2$

Based on the figure, the value of δ for each term is relatively similar for inputs X_1 and X_2. The $\delta(\sigma_i)$ is 0.06, and these results agree with the Monte Carlo simulation method. The δ for μ_i is greater than $\delta(\sigma_i)$ which is around 0.14 for both inputs. If we learn both parameters simultaneously we will get $\delta(\mu_i\sigma_i)$ is around 0.15. Thus learning μ_i will cover 93.33% while assessing δ will contribute 40%. On the other hand, if we assess the value of learning input X_i we will get $\delta(X_i) = 0.3$ and this index represents learning the inputs have a higher value and they are important as much greater than $\delta(d_i)$.

If we compare the above results with the δs for the non-linear model which is provided in Figure 6.4, we notice that the value of learning input X_1 indicated by $\delta(X_1)$ rises almost two-fold. The non-linearity in X_1 affects the increase of $\delta(\mu_1\sigma_1)$, but the $\delta(\mu_1)$ and $\delta(\sigma_1)$ are remain the same. On the other hand, the δ for input X_2 shrink to less than 0.2. The δs for other parameters are shrink as well. In particular, $\delta(\sigma_2)$ became very small and almost reach the δs for the dummy inputs so that learning σ_2 seems unnecessary.

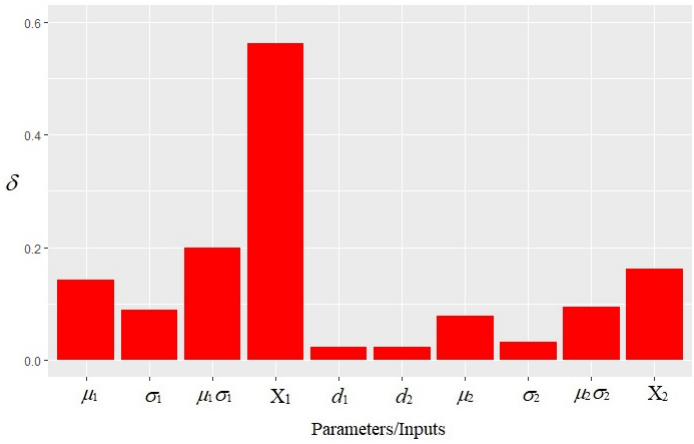
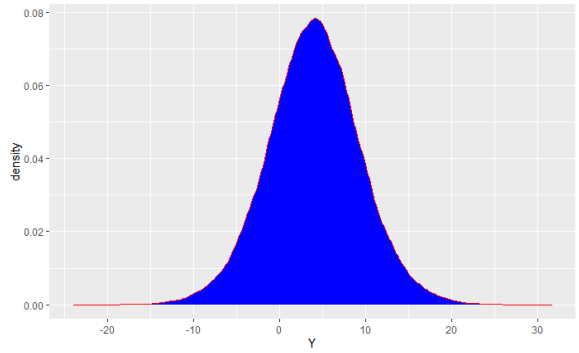


Figure 6.4: Delta indices for learning σ_i in the non-linear model $y = x_1^2 + x_2 + 0.1d_1 + 0.1d_2$

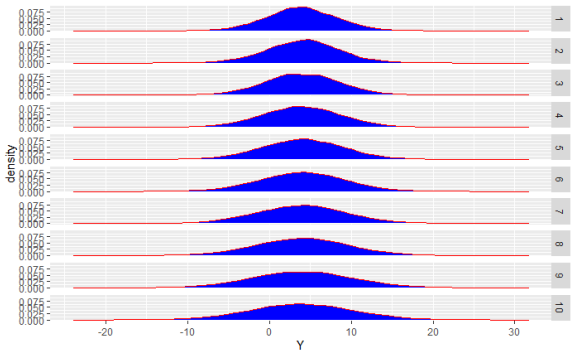
The density plots of Y and $Y|\sigma_1$ for the linear model are provided in Figure 6.5. The unconditional density plot for the linear model $Y = X_1 + X_2$ in Figure 6.5a looked symmetric with the mode of the output is around four. This is sensible as the output Y is the sum of two normal random variables with means two and variances four, which theoretically would be a normal distribution with mean four and variance eight. The conditional distributions of $Y|\sigma_1$ in Figure 6.5b are computed using the sample points of Y which fall in the bins. The bins are constructed by splitting the value of output Y into ten equal parts based on the quantiles of σ_1. The conditional densities seem to have similar modes or means with symmetrical forms, but the later conditional densities tend to have lower density values among ten densities. The Borgonovo's δ for σ_1 is obtained by comparing the unconditional density $f_y(Y)$ with each conditional density $f_{y|\sigma_1}(y)$ and finding the expected shifts between them.

We can compute δ for another parameter μ_1 and input X_1 using the bin partition for determining the conditional densities. We can use the grids for the conditional densities of $\mu_1\sigma_1$ as it involves two parameters. A similar way is used for computing δ for input X_2 and their parameters with results have been provided in Figure 6.3.

Following the linear model $y = x_1 + x_2$, we also provide graphical representation of the densities of output in the non-linear model $y = x_1^2 + x_2$. Similar to the previous model, the unconditional and the conditional densities for the output of the non-linear model are presented in Figure 6.6. The unconditional distribution for y is right-skewed with a long tail as the result of the squared input X_1 (See Figure 6.6a). The tail of the unconditional density $f_y(y)$ could reach more than 500 on the X-axis. The forms of conditional densities



(a) Unconditional density of $f_y(y)$ for model $Y = X_1 + X_2$

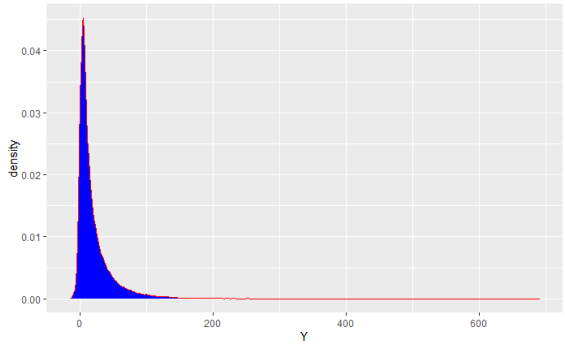


(b) Conditional densities of $f_{y|\sigma_1}(y)$ for model $Y = X_1 + X_2$ in ten bins size.

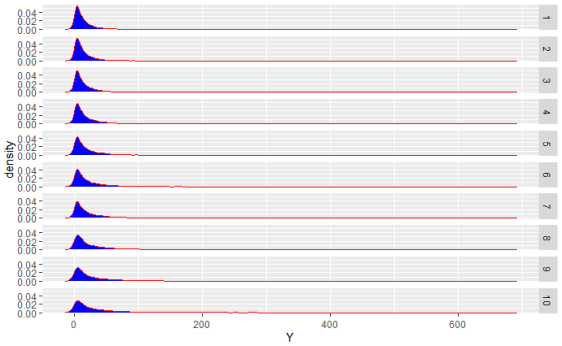
Figure 6.5: The comparison of unconditional density $f_y(y)$ and conditional densities $f_{y|\sigma_1}(y)$ for linear model $Y = X_1 + X_2$.

for $f_{y|\sigma_1}(y)$ are provided in Figure 6.6b. The conditional densities of $f_y|\sigma_1(y)$ are also right-skewed with long tails. Similarly, the conditional distributions are obtained from samples of the output model $Y = X_1^2 + X_1$ which are partitioned onto some parts. We called these parts bins.

We use these bins not only for computing conditional densities but also for conditional variances, conditional cumulative probabilities, and so forth with one conditional input. In this example, bins are constructed to approximate the conditional density of $f_{y|\sigma_1}(y)$. The partition of the output samples is based on some quantiles of σ_1 as the conditional input is σ_1. The number of bins for the conditional density approximation depends on the number of samples. The more samples we have for our model, the more bins we can create to better approximate the conditional density. For this example, we used 100,000



(a) Unconditional density of $f_y(y)$ for model $Y = X_1^2 + X_2$



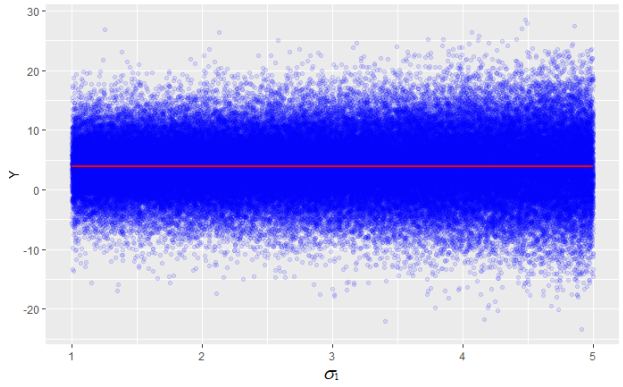
(b) Conditional densities of $f_{y|\sigma_1}(y)$ for model $Y = X_1^2 + X_2$ in ten bins size.

Figure 6.6: The comparison of unconditional density $f_y(y)$ and conditional densities $f_{y|\sigma_1}(y)$ for the non-linear model $Y = X_1^2 + X_2$.

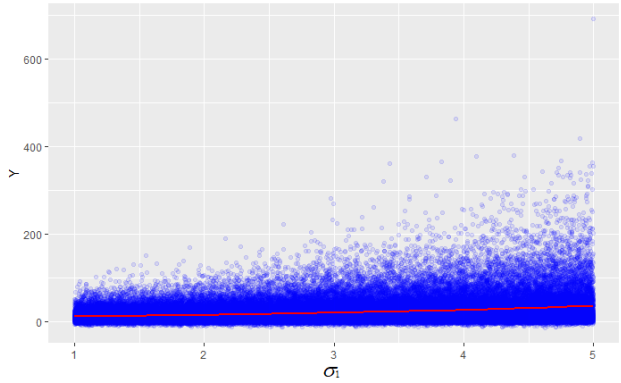
samples and partitioned the samples into ten bins. Ten bins are chosen for the sake of illustration. As we need to create ten bins, the number of quantiles required for the boundary for the bins is 11. The quantiles are 0%, 10%, 20%, ..., 100%. Bin 1 is the first two nearest quantile ranges of σ_1, i.e. between 0% and 10%. Bin 2 is the sample space within quantile 10% and 20%, and so forth. The bins discretise the samples of output Y into ten parts. Samples that stay in a bin are used to construct the density of Y, and the density approximates $f_{y|\sigma_1}(Y)$.

The Borgonovo's δ for σ_1 is computed by considering the shift or difference between the unconditional distribution $f(y)$ and the conditional distributions $f_{y|\sigma_1}(y)$. There are ten comparisons for unconditional distribution as there are ten conditional distributions, and $\delta(\sigma_1)$ is obtained from the mean of the shifts. From the previous results, we obtain $\delta(\sigma_1)$

for the linear and non-linear model are 0.06 and 0.09, respectively. We want to model the relationship between σ_1 and Y. The graphical representation of the relationship is indicated in Figure 6.7.



(a) The red line indicates the relationship between σ_1 and Y for model $Y = X_1 + X_2$.



(b) The red line indicates the relationship between σ_1 and Y for model $Y = X_1^2 + X_2$.

Figure 6.7: The relationship between σ_1 and Y for linear and non-linear models

According to the figure, the two models have a different relationship between σ_1 and Y. The constant line in Figure 6.7a indicates that there is no effect of σ_1 on the average value of Y. On the other hand, for the non-linear model, there is a trend of a slight rise in Y as σ_1 gets bigger. The trend is indicated by the slight slope of the red line in Figure 6.7b.

The grid-based estimation method

Let us consider again Example 5.1. We propose to compute $\delta(\mu_i\varphi_i)$ using a grid-based approach. The grid is an extension of the bin approach which has been used for approximate $f_{y|\sigma_1}(y)$ in the previous section. As there are two inputs as the condition of Y, which are μ_i and φ_i, we will make a two-way partition. The first step is we need to obtain a sample of model runs. For illustration, we suppose the function for linear and non-linear model in Example 6.1 is $Y = f(X_1, X_2)$. The second step is we need to sample $\mu_{i,j}$ and $\varphi_{i,j}$ from $P(\mu_i, \varphi_i)$. Next, we sample $x_{i,j}$ from $P(X_i|\mu_i\varphi_i)$, for $i = 1, 2$ and $j = 1, \cdots, N$ and evaluate $y_j = f(x_{1,j}, x_{2,j})$ for $j = 1, \cdots, N$. The third step is to display all values sampled from the previous step in Table 6.1.

Table 6.1: Model runs layout for computing sensitivity indices using the Grid-based method for models with two inputs.

μ_1	τ_1	μ_2	τ_2	X_1	X_2	Y
μ_{11}	τ_{11}	μ_{21}	τ_{21}	x_{11}	x_{21}	y_1
μ_{12}	τ_{12}	μ_{22}	τ_{22}	x_{12}	x_{22}	y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
μ_{1N}	τ_{1N}	μ_{2N}	τ_{2N}	x_{1N}	x_{2N}	y_N

The next step is to estimate the unconditional density $f_Y(y)$ using a Kernel density estimate $\hat{f}_Y(y)$ given $y_1, \cdots, y_N$. Now, we need to estimate another density function which is conditional $f_Y|\mu_i, \varphi_i(y)$. To do so, we make an assumption that if μ_{ij} and μ_{ik} are similar and so are φ_{ij} and φ_{ik}, then

$$f_{Y|\mu_i=\mu_{ij},\varphi_i=\varphi_{ij}}(y) \simeq f_{Y|\mu_i=\mu_{ik},\varphi_i=\varphi_{ik}}(y)$$

We discritise the distribution of μ_i and φ_i and then identify $S_{i,p,q}$. The explanation about $S_{i,p,q}$ will follow. After that we construct Kernel density estimate for $f_{Y|\mu_i\in S_{i,p,q},\varphi_i\in S_{i,p,q}}(y)$.

Next, we present the notation to describe the grid cells and indicate which output values y_j in the model run in each grid cell. An illustration of the grid is provided in Figure 6.8.

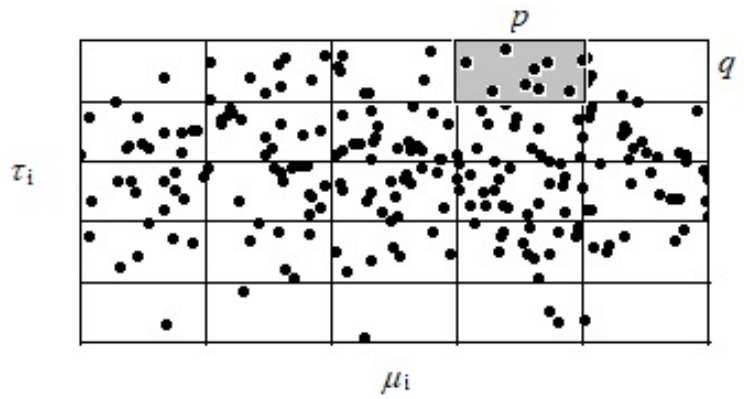


Figure 6.8: The Grid contains the output values y_j in each cell where they come from the samples of parameter μ_i and τ_i. The grey cell indicates the output values y_j in a particular cell in the intervals $A_{\mu_{i,p}}$ and $B_{\tau_{i,q}}$.

If we look at the grid, we notice that the grid is constructed by two parameters of the Beta model, μ_i on the horizontal side and $\tau_i = \log\varphi_i$ on the vertical side. For analysis convenient purpose, it is useful to construct model runs from values sampled from the Beta model for all elements such as X_i, Y, μ_i, and τ_i. Model runs for two inputs are presented in Table 6.1.

The grid is constructed from partitions on the sample space of the Beta model parameters. Firstly, we make partition on the sample space of μ_i into n_{μ_i} contiguous intervals $A_{\mu_{i,1}}, \ldots, A_{\mu_i, n_{\mu_i}}$. Secondly, we make partition on the sample space of τ_i into n_{τ_i} contiguous intervals $B_{\tau_{i,1}}, \ldots, B_{\tau_i, n_{\tau_i}}$, so that we will have a grid with dimension of n_{μ_i} x n_{τ_i}. Thirdly, we define a specific grid cell, p x q, so that

$$S_{i,p,q} \subset \{1, \ldots, N\}$$

with $j \in S_{i,p,q}$ if and only if both $\mu_{i,j} \in A_{\mu_i,p}$ and $\tau_{i,j} \in B_{\tau_i,q}$. For instance, an index j of the output is inside the grid cell $S_{i,1,1}$ if and only if both $\mu_{i,j} \in A_{\mu_{i,1}}$ and $\tau_{i,j} \in B_{\tau_{i,1}}$. By identifying an index that falls in a grid cell, we are able to determine the corresponding model run y_j belongs to the grid cell. Now, we can compute, for example, the model output mean for the particular grid cell (p, q) by the following expression

$$\bar{y}_{i,p,q} = \frac{1}{|S_{i,p,q}|} \sum_{j \in S_{i,p,q}} y_j$$

where $|S_{i,p,q}|$ is the number of y_j which falls in the cell (p, q). Using the grid-based method, we can compute the conditional expectation $E_{\mu_{-i},\tau_{-i}}(Y|\mu_i, \tau_i)$ for the purpose of computing

the sensitivity index using main effect using the variance-based method in Equation (2.27). We can also compute the conditional density function $f(Y|\mu_i, \tau_i)$ to calculate the density-based sensitivity index such as Borgonovo's δ for two variables or parameters. Therefore we can compute the generalisation of Borgonovo's δ for two conditioning variables in the Equation (6.7). If there is only one conditioning variable, we use bins to make partitions of the model outputs instead of the grids.

We can also use the grid method for computing main effects and compare it with the Generalised Additive Method (GAM) method to evaluate the accuracy of the sensitivity indices using the grid-based method. The following section reports numerical experiments for computing Borgonovo's δ and main effects using the grid-base method. We compare the result of main effects computed using the grid and using the GAM method afterwards.

§ 6.7 Numerical experiments

We now conduct some numerical experiments to test the grid. The numerical experiments aim at investigating the following questions: 1) How many model runs should be used? 2) How many grid cells are sufficient? 3) What is the number of quadrature points for the numerical integration? 4) How many model inputs can the method handle, especially in the relationship between questions number (1) and (2)?

In this numerical experiment, we will consider two models, a model with a small number of inputs and a model with a large number of inputs. The first model is the flood model with eight inputs (see equation 2.2), and the second model is the modified Ishigami function which will be presented in Section 6.7.2. Two sensitivity indices, Borgonovo's δ and main effect index, are computed for each model. The resulting indices will be compared with those computed using the GAM method.

6.7.1 The Flood Model experiment

The first model we would like to consider in the numerical experiment is the flood model in Equation (2.2). The numerical experiments were conducted using three factors: number of Grid cells (nG^2), number of Runs (nR), and the number of quadrature points/Simpsons's points (nS). The levels for nG are 5, 10, 20, 30, and 40. The level of $nG = 5$ means that we use $5 \times 5 = 25$ grid cells to compute the sensitivity indices, $nG = 10$ means we use 100 grid cells, and so forth. The second experiment factor is nR, which has three levels: 25000, 50000, and 100000, and the third factor is nS which has levels 101, 1001, and

10001. We use a factorial design to obtain the combination of factor levels for computing the sensitivity indices.

Borgonovo's δ computation using The grid-based method

In the first experiment, the grid-based method is used for computing Borgonovo's δ for input in the flood model. The computation of the δ is repeated two times for each configuration (combination of factor levels). The computation took around 28 minutes to complete and experiment outputs are shown in Figure 6.9 - Figure 6.13.

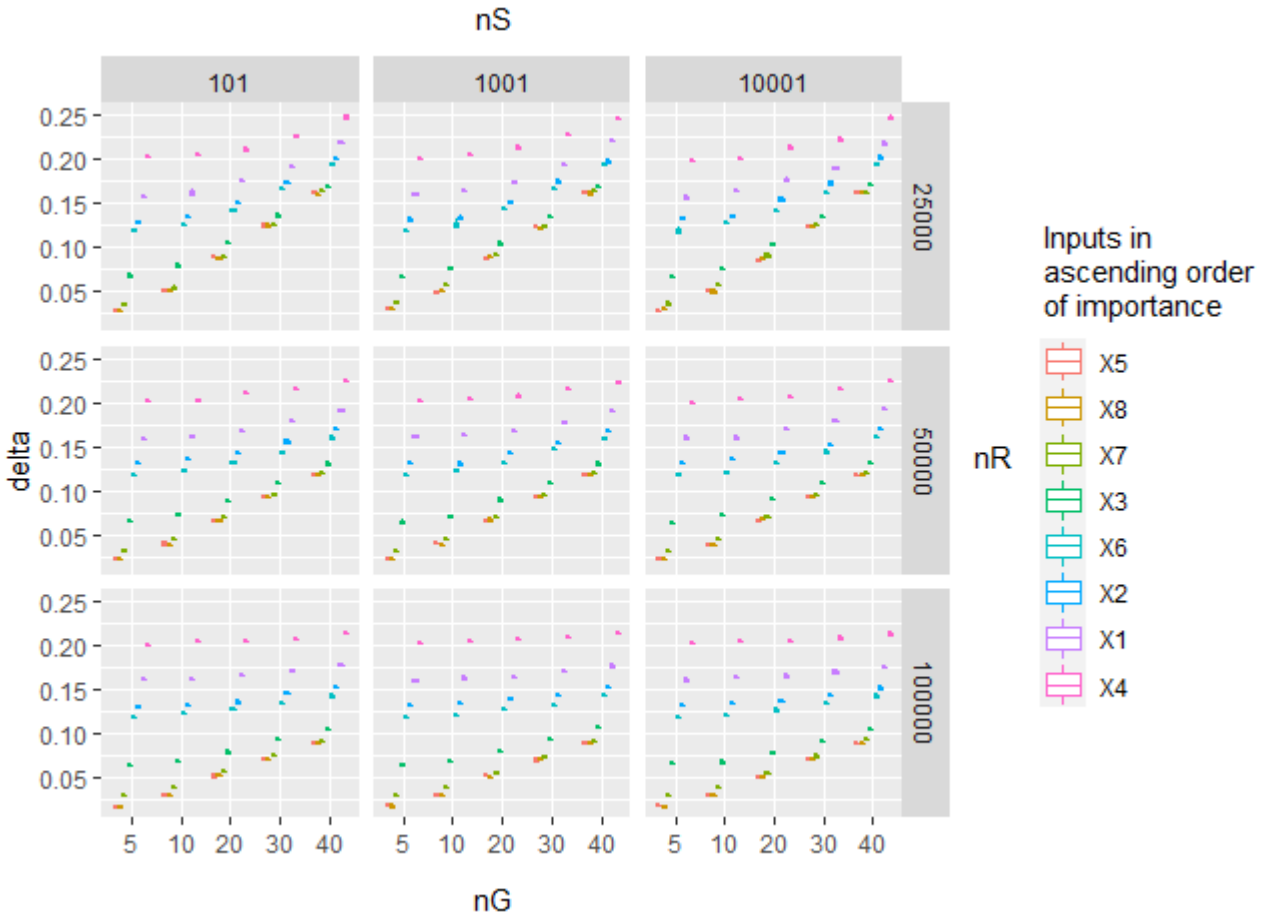


Figure 6.9: Borgonovo's δ computed using different combination settings of the number of grid cells (nG^2), the number of runs (nR), and number of Simpson's points (nS). We can conclude that if the number of runs get larger than the estimate for δ tend to steady.

Figure 6.9 displays boxplots of Borgonovo's δ indices computed using different combination

levels of factors. The figure shows that input X_4 has the largest δ, input X_1 is the second-largest δ index, followed by X_2, X_6, and so forth. On the other hand, input X_5 and X_8 have the smallest δ indices, and their values are relatively similar. These results indicate that X_4 is the most important input while X_5 and X_8 are the least important inputs. We notice that δ indices in Figure 6.9 tend to increase when the number of grid cells rises. δ indices looked stable if computed using a large number of runs, especially for important inputs. Simpson's points seem do not affect the δ indices.

Now, the behaviour of δ for some inputs will be observed. We will focus on two inputs, X_4 (which has the largest δ) and X_5 (which has the smallest δ). The detail plots for X_4 is shown in Figure 6.10. For input X_4, the number of runs seems to affect the stability of the indices. The largest number of runs, 100000, makes δ indices of input X_4 more stable, around 0.20 - 0.22. While the smallest number of runs, 25000, makes the indices of input X_4 range a little bit wider, i.e. between 0.20 - 0.25.

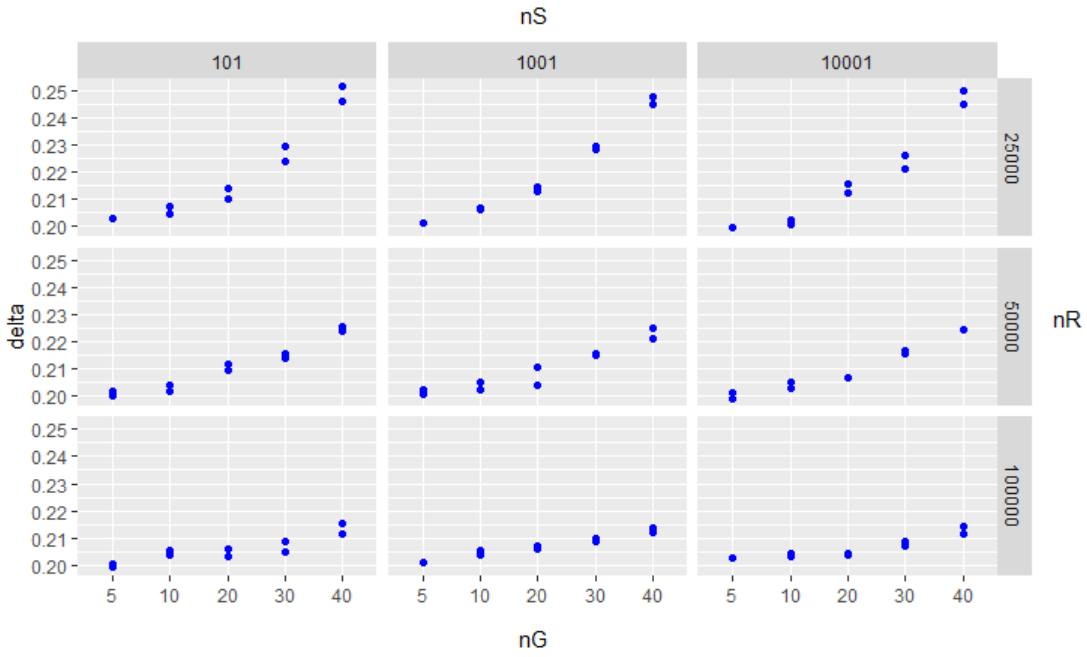


Figure 6.10: δs for input X_4 computed using different combination of levels nG, nR, and nS.

In addition, δs for X_4 increases exponentially when the number of runs is 25000. This increment is along with the addition of the number of the grid. It means that the choice of the grid using the current number of runs, i.e. 25000, will affect the addition of δ in two

decimal places. If the number of runs doubles to 50000, it reduces the range of δ in half, and If the number of runs is multiplied by four, then the range of δ will be more stable. These results suggest that the number of runs 100,000 is a good choice for computing the δ. The more runs, the more stable the δ, but the computation time is longer. The number of Simpson's points do not affect δs for X_4 for all combination of nR and nG^2. It means that the numerical integration for computing δ using different quadrature points produces relatively similar values.

The boxplot in Figure 6.11 confirms some of these results. According to Figure 6.11a, the choice of number of grid cells affects the variation of δ. For instance, the grid cells equal to 5^2 and 10^2 produce the smallest variation in δ. While the number of grid cells 40^2 produces a big variation in δ. This is because the number of grid cells, for example, 5^2, creates 25 cells which cause the number of sample points to be bigger than 1600 grid cells. The more sample points in the cells produce δ more accurately. Figure 6.11b shows that the number of runs affect the variation of δ. The larger number of runs, the smaller variation in δ. Even though the median of δs for each run do not differ so much, figure 6.11c confirms that the number of quadrature points used in the numerical integration for computing δ does not seem to have any effect.

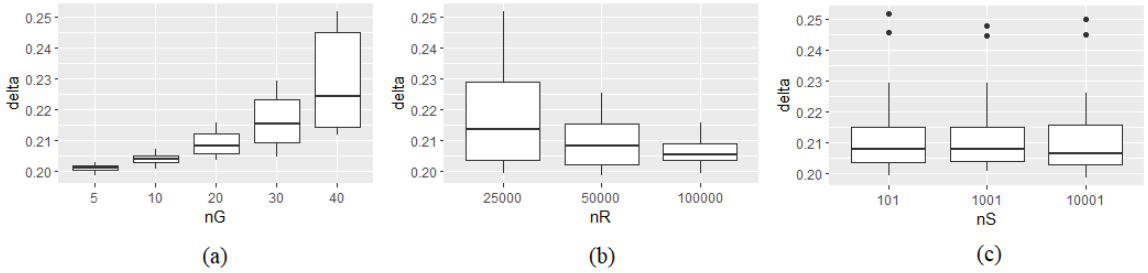


Figure 6.11: Boxplots of δ for X_4 viewed from separate factors: number of grids (nG), number of runs (nR), and the number of Simpson's points (nS).

The inputs having small δ indices are also investigated. We can observe from Figure 6.9 that small δ indices are also affected by the number of runs. When the number of runs used is 25000, the indices for inputs X_8 and X_5 are in 0.03 - 0.16. However, if we used 100000 runs, the indices are more stable in the range of 0.03 - 0.08.

We now focus on the inputs that have the smallest δ index, X_5. The points plot for the detail index for X_5 is shown in Figure 6.12. The figure shows δ indices for X_5 increase as the number of grids rise in linear forms. When the number of runs is 25000, the range of

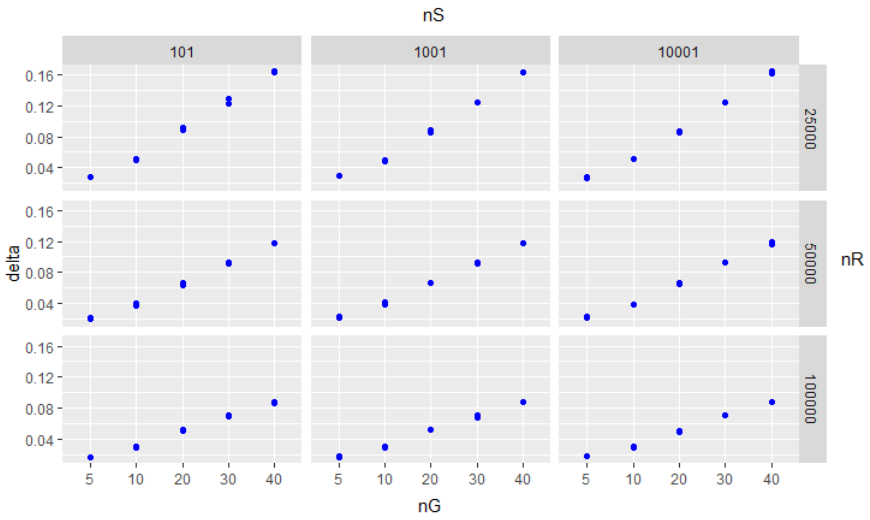


Figure 6.12: Point plots for Borgonovo's δ of X_5 using different number of grid cells (5^2, 10^2, 20^2, 30^2, 40^2), number of runs (25000, 50000, 100000) and number of Simpson's points (101, 1001, 10001).

the index for X_5 is between 0 - 0.16. If the number of runs rises by two times, the range of the index shrinks to 0 - 0.12. The index range continues to get smaller for the 100000 runs with the interval between 0 - 0.09.

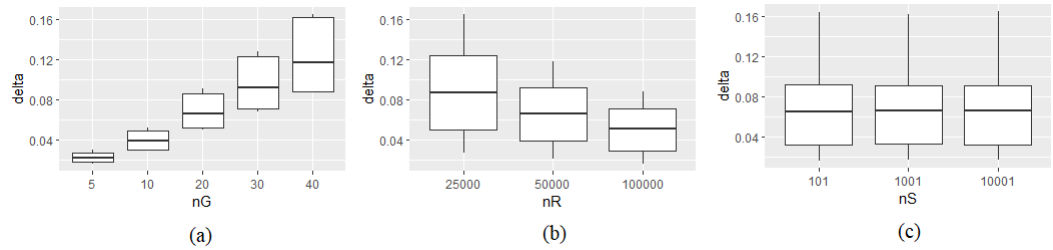


Figure 6.13: Boxplots of δs for input X_5 viewed separately based on each experimental factor: the number of grid cells (nG^2), the number of runs (nR), and the number of Simpson's points (nS).

We can also investigate the behaviour of δ for each factor separately as displayed in Figure 6.13. We can see clearly from the boxplot presented in Figure 6.13a that the value and the range of δ indices rise along with the addition of the number of grid cells. On the other hand, the median of the index declines smoothly along with the rise of runs (see Figure 6.13b). Figure 6.13 shows that the addition of Simpson's points does not change

the median of the index or the index range.

The next sensitivity measure to be computed using the grid-based method is the main effect index and is explored in the following section.

Main effect index computation Using the grid-based method

This experiment compares the main effect index computed with the grid-based method and the Generalised Additive Model (GAM) method. The main effects index will be computed using the grid-based method at various settings. The main effects will also be computed using the GAM method based on a different number of runs.

The grid-based method for computing the main effect index involves two factors: The number of grid cells (nG^2) and the number of runs (nR). We use the same levels for each factor as in the previous experiment: nG has five levels: 5, 10, 20, 30, and 40 and nR has three levels: 25000, 50000, and 100000. For efficiency, we used three replications for each combination of factor levels. The results of the experiments are presented in Figures 6.14 - 6.16.

Figure 6.14 shows boxplots of the main effect index for all configurations of factors. The main effect index shows similar results as Borgonovo's δ in determining important inputs. X_4 is the most important input while X_8 is the least important. Although X_5 is also the least important as the values of the main effect index is very close to X_8. The figure indicates that a different number of grid-cells does not differ in the resulting main effect index. Additionally, the number of runs does not affect the main effect index too much.

As before, we want to look over in detail the main effect experiments in some inputs and focus on the most important input X_4, and the least important input X_8. The results of the numerical experiment for input X_4 is provided in Figure 6.15. The figure shows that the main effect indices do not differ too much for almost all configurations, and the index value is stable at around 0.27.

The boxplots in Figure 6.16a suggest that the computation of the main effect for X_4 looks stable on 0.270 in three decimal places if we use at least 20^2 grid cells. In addition, the number of runs does not affect the main effect index too much, although the number of runs 100000 give the smaller variation of the index (See Figure 6.16b).

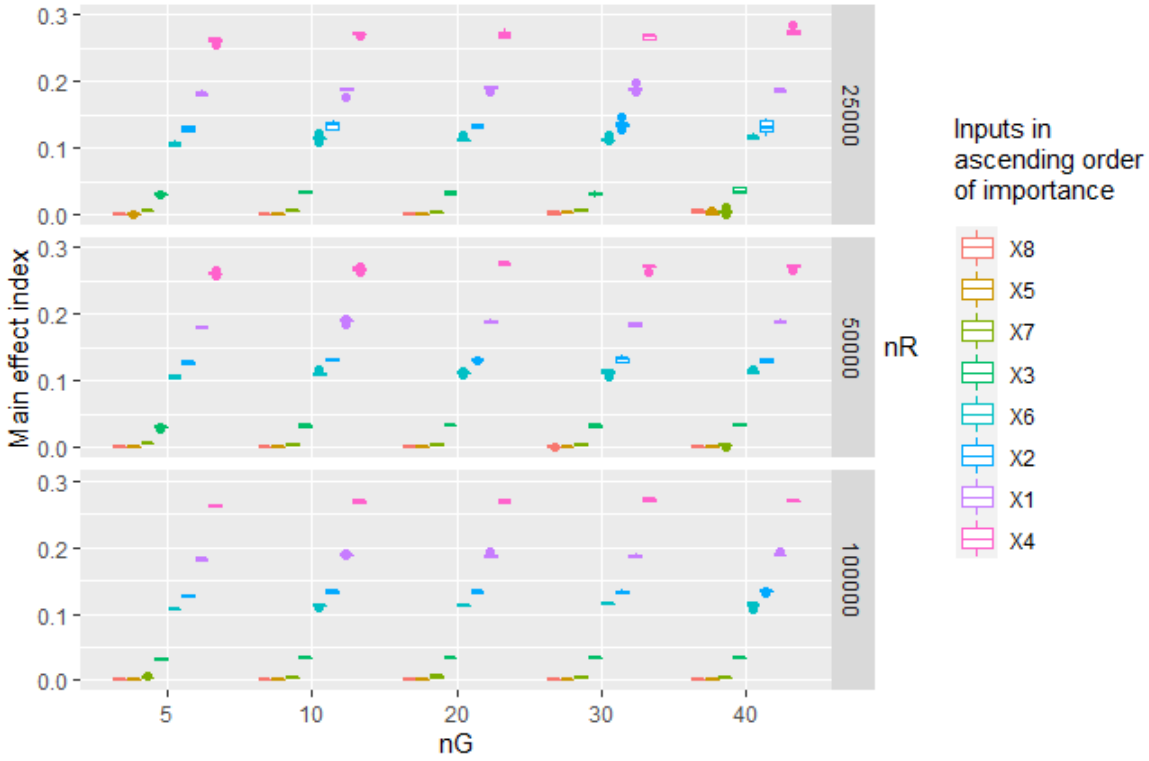


Figure 6.14: Boxplots for all input's main effect indices in the modified flood model computed using different number of grid cells (5^2, 10^2, 20^2, 30^2, 40^2) and different numbers of runs (25000, 50000, 100000). The figure indicate that different number of grid cells and number of runs do not affect the main effects for all input significantly.

Main effect index computed using The GAM method

The GAM method can be used as an alternative to computing the main effect index. GAM is considered in this section to be compared with our computational method, the grid-based method. This is because GAM has been established for modelling the non-linear relationship between input and output, as discussed in Chapter 2, and we want to use the method to compute the main effect based on the fitted values of the non-linear flood model. GAM method is the generalisation of the regression-based method which is used for a more flexible form of mathematical model relationships. The fitted values of the model obtained using GAM will be used to compute the main effect index. Meanwhile, in our method, the main effect is computed using the difference between unconditional variance and conditional variance calculated from the grids. Therefore, the GAM method

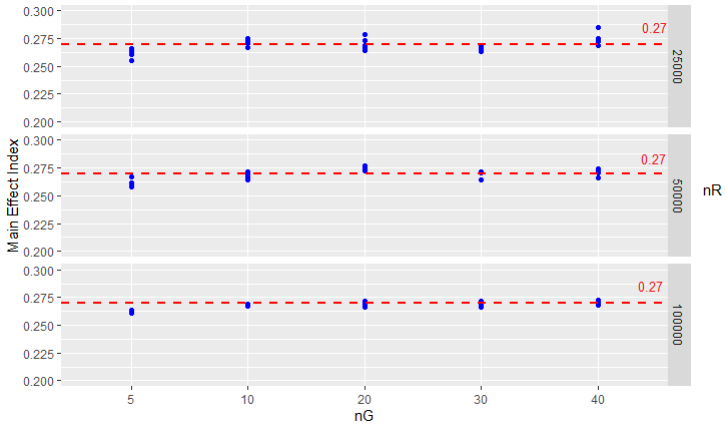


Figure 6.15: Point plots of the main effect for X_4 computed using two experimental factors: the number of grid cells (5^2, 10^2, 20^2, 30^2, 40^2) and the number of runs (25000, 50000, and 100000). Three replications are used in this computation.

is expected to be a good benchmark for our method.

If we compute the main effects using GAM method, we obtain Input X_4 is the largest main effect index with values around 0.255. Input X_1 is the second largest, with main effect index values around 0.3. On the other hand, X_8, X_5, X_7, and X_3 have very small main effect indices close to zero.

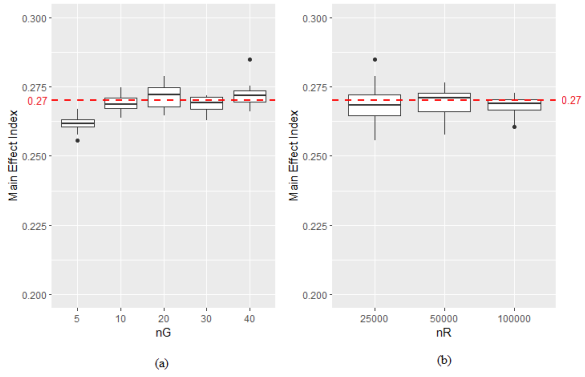


Figure 6.16: Boxplots of the main effect index for X_4 viewed separately according to the number of grid cells (nG^2) and the number of runs (nR).

The following section provides another numerical experiment of the well-known Ishigami function with the addition of many dummy inputs in the function.

6.7.2 The Modified Ishigami function experiment

The second numerical experiment is conducted on the modified Ishigami function. This function is an extension of the original Ishigami function introduced by Ishigami & Homma (1990) with the addition of dummy variables. There are 47 dummy variables added in the function starting from $x_4, \ldots, x_{50}$ with each input has weight w. This function is used to test the grid with many inputs and see how the grid differentiates the index between the main inputs and the dummy inputs.

The dummy inputs $x_4, \ldots, x_{50}$ should have small weights to indicate they are different from the main inputs. The smaller the weights, the smaller the effects of inputs on the model output. For example, if the weight is 0.1, the portion of each dummy input contributing to the model output y is only 10%. If we use a smaller weight of 0.01, then the portion of each dummy input contribution becomes only 1%, and zero weight means the dummy input has no contribution to the model output. The small weight for input has been identified for input unimportance using Scenario 2 Definition 3 in Chapter 3 via an example.

Now we have the full expression of the modified Ishigami model as follows.

$$f(x) = \sin(x_1) + a\sin(x_2)^2 + b\,x_3^4\sin(x_1) + w\sum_{i=4}^{50} x_i \tag{6.11}$$

where $a = 7, b = 0.1$, $w = 0.1$, and $x_i \sim U[-\pi, \pi]$, for $i = 1, \ldots, 50$. We chose the weight of 0.1 for the dummy inputs as it is sufficient to make differentiation of sensitivity indices with main inputs. A smaller weight for dummy inputs says 0.01, makes the dummy inputs seem not to do anything to the less realistic output. Using this test example, we aim for the dummy inputs to have around 90% of the weights and 10% for the main inputs. The sensitivity indices obtained using this weight can be observed for all 47 dummy inputs to ensure the tendency of small indices.

The Borgonovo's δ for the inputs of this function are shown in Figures 6.17. For the sake of illustration, we only plot δs for the main inputs and two dummy inputs. The figure indicates the Borgonovo's δ for the main inputs of the Ishigami function lie separately above the δs for the dummy inputs. Additionally, the δ for dummy inputs seem to be similar for each combination. The plots indicate that the main inputs X_1 have the biggest main effect index, so they are the most important inputs. In this case, the meaning of the most important input is that if X_1 is fixed to its true value, it will give the highest expected reduction in the output variance. X_2 has the second-largest main effect index, while X_3 has the lowest main effect index.

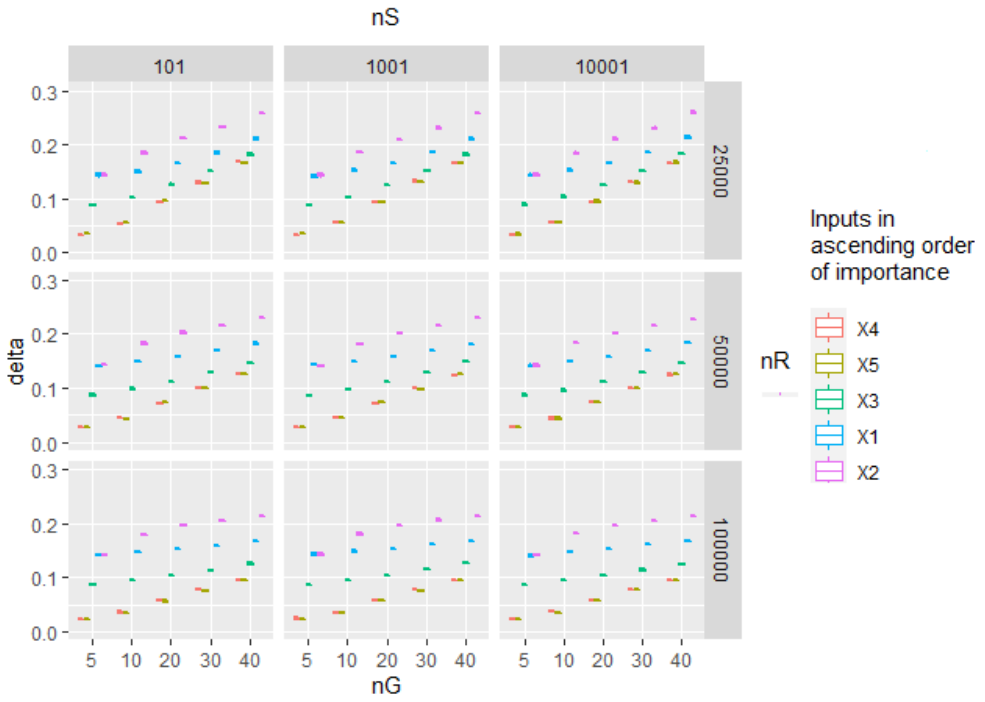


Figure 6.17: Boxplots of δ for main inputs and the first two dummy inputs in the Modified Ishigami function computed using different levels of factors. Three factors considered in the computation of δ are the number of grid cells (nG^2) 5^2 up to 40^2, the number of runs (nR) 25000, 50000, and 100000; and the number of Simpson's points (nS) 101, 1001, and 10001. Borgonovo's δ for the main inputs are separated clearly with the dummy inputs. Different numbers of Grids affect the δs with linear increment. A larger number of runs produce a more stable δ estimate. However, the number of Simpson's points do not affect δs.

Additionally, according to Figure 6.17, the value of δ rises when the grid size increases. When the number of runs is elevated to 100,000, the rise of δ along the gird size are not too sharp. The different settings of Simpson's points do not affect the δs. In addition, input X_2 has the largest δ with a range is between 0.15 to 0.27 when the number of samples is 25,000. The range of δ for X_2 shrinks to 0.15 - 0.22 for 100,000 number of runs. Input X_3 has the smallest δ with the range of 0.08 - 0.13 for the maximum number of runs 100,000.

The value of δs for dummy inputs shows the same trend in the rise as the main inputs and the addition of grid cells. The range of δ for dummy inputs are from 0.03 - 0.18 for the number of runs 25,000. The maximum value of δ for dummy input is below 0.10 when

the number of runs is 100,000.

We provide barplots for δ for all inputs in the modified Ishigami function, including the dummy inputs and presented in Figure 6.18. The barplots indicate δ for inputs produced using a selected combination level of factors. The combination are nR = 100000, nG = 20, and nS = 101. We select nR = 100000 because we want the maximum number of sample points for each cell in the grid. nG = 20 is chosen for the medium number of grid cells in our experiment. nS is selected to be minimal as it is not affected the δ from the previous experiment in Section 6.7.1.

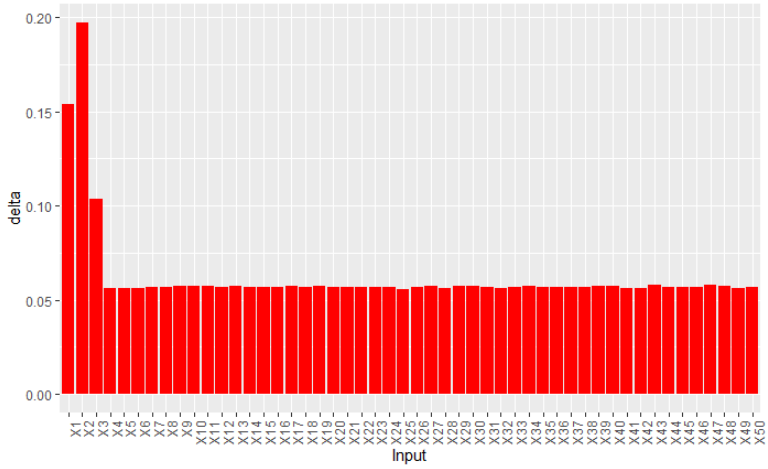


Figure 6.18: Barplots of δ for all inputs in the Modified Ishigami function for selected level of factors. The levels are: nR = 100000, nG = 20, and nS = 101.The δ for X_2 is the highest indicating it is the most important input while X_3 is the least important, while dummy inputs' delta are around 0.05.

According to barplots in Figure 6.18, input X_2 has the biggest δ so it is the most important input. X_1 is the second biggest δ with value around 0.16 while X_3 has the lowest δ. The dummy inputs have homogeneous values of δ around 0.05. We have explored the behaviour of the δ using different combination settings. We found that if we want to reduce δ for the dummy inputs, the grid cells should be minimum for a fixed nR.

Subsequently, we want to investigate the pattern of δ for different grid cells to discover the minimum grid cells used in the grid method for computing Borgonovo's δ. We choose the most important input X_2 to be investigated. Figure 6.19 shows the points plot of δ for x_2 using three replications. The plots demonstrate that δ increases linearly across the addition of grid cells. The surge in the δ lessens by the rise of the number of runs. The

range of δ for nR = 100,000 is 0.14 - 0.22, which will shrink if we use more runs. The number of Simpson's points do not affect the pattern of δ.

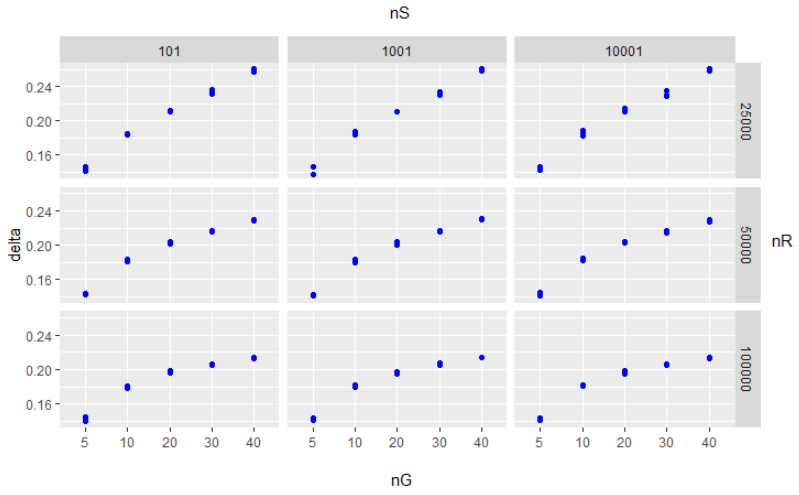


Figure 6.19: Point plots for δ_{X_2} of the modified Ishigami Function computed using the grid method. The factors considered in the Grid methods are number of grids (5 up to 40), the number of runs (25000, 50000, and 100000), and the number of Simpson's points (101, 1001, and 10001)

The boxplot in Figure 6.20a indicates that the addition of the grid cells will increase the δ and its variation, while the increase of the number of runs will decrease the δ and its range slightly which is pointed in Figure 6.20b. On the other hand, the number of Simpson's points for the numerical integration does not affect δ (See Figure 6.20c). Similar results also apply for input X_1 and X_3.

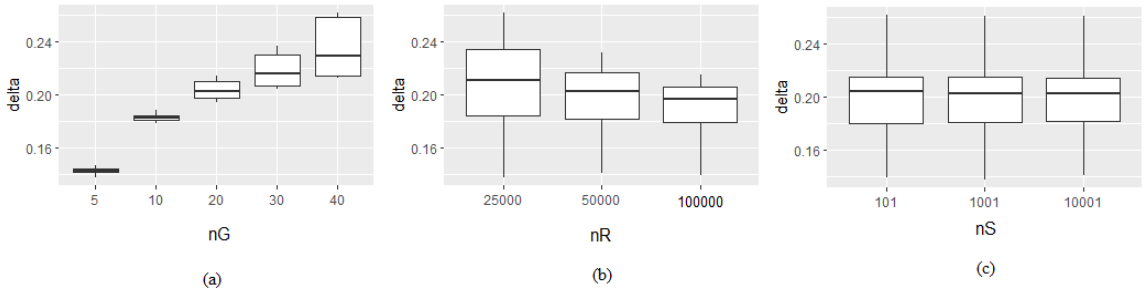


Figure 6.20: Boxplots for δ_{X_2} viewed for each factor involed in the grid method. The factors are the number of Grid (nG), the number of runs (nR), and the number of Simpson's points (nS).

Next, using a similar way of the numerical experiment as presented in Section 6.7.1, we compute the main effect index using the grid-based method and also compare the results with the GAM method. For illustration, we chose a combination level for the grid method and compared the result with the GAM method.

The main effect index for inputs in the modified Ishigami function are presented in Figure6.21. The main effects indices were computed using 400 grid cells and 25000 runs. We notice from the figure that X_2 is the most important input while X_1 is in the second rank. Input X_3 seems to have a main effect index close to zero, similar to the indices for the dummy inputs. The results for main effect indices computed using the grid-based method are quite similar with the ones computed using the GAM method in Figure 6.22. The barplots indicate that X_2 is the most important input and X_3 is the least important input. The dummy input indices seem to be more homogenous, with values close to zero.

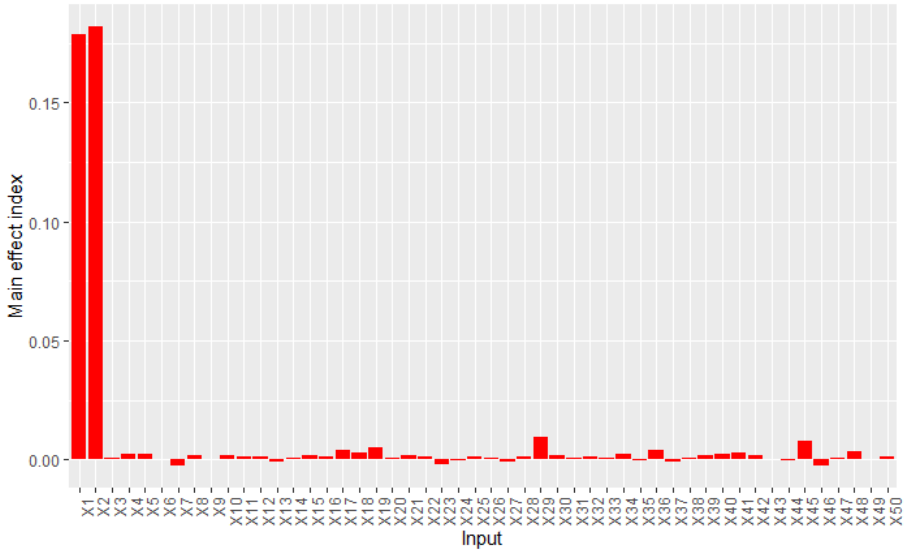


Figure 6.21: Main effect index for all inputs in the Modified Ishigami function computed using the grid method for selected number of grids and number of runs. For this plot, we choose number of grid cells = 20^2 and number of runs = 25000.

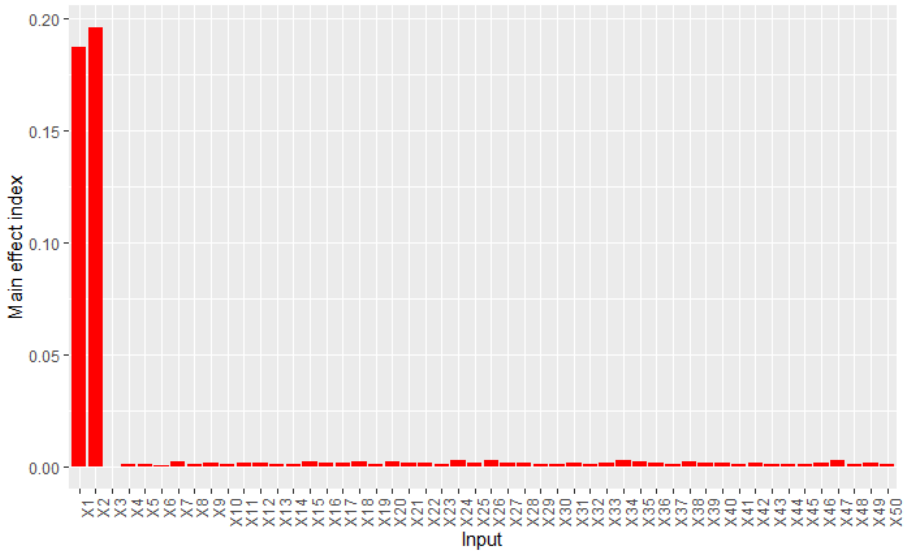


Figure 6.22: Main effect index for all inputs in the Modified Ishigami function computed using the GAM method for selected number of runs. For this plot, we choose number of runs = 25000.

§ 6.8 Conclusion

This chapter introduces a new model for expert judgment called The Beta Model. The model is more practical to use the ones proposed by Lindley et al. (1979) and Gelfand et al. (1995). We use sensitivity analysis to prioritise elicitation from the Beta model. The Grid-Based method is used to support the computation of sensitivity indices for model inputs and also parameters in the Beta model. We conduct numerical experiments to test the grid-based method for computing sensitivity indices δ and main effect index.

The results of the numerical experiment using the three methods, Borgonovo, the Main effect using the grid, and GAM, provide the mutually supportive conclusion. Input X_2 is identified to be the most important input in the modified Ishigami function, and X_3 is the least important one. The input X_2 has a nice convergence pattern by combining the number of grid cells and the number of runs. The use of grid cells at least 20^2 with several runs 100,000, and a small number of Simpson's points are recommended for computing the δ. If we want to compute the main effect index using the grid-based method, 20^2 grid cells and 25,000 runs are sufficient.

Numerical experiments indicate that the grid-based method is comparable to the GAM method for computing the main effect index. For computing Borgonovo's δ, it is important to have sufficient runs per grid cell, and a large number of cells with smaller runs seems to give an inflated estimation.

Chapter 7

Conclusion and Future Works

§ 7.1 Conclusion

The conclusions of the thesis are as the following points:

1) The criteria of defining unimportant inputs in the mathematical models are formulated formally in this thesis. The criteria are based on two scenarios related to the similarity of the loss functions and the similarity of two variances or probabilities. The criteria for identifying the unimportant inputs are investigated by implementing measures of importance such as total effect indices, Shapley effects, Chun-Han-Tak measures, Borgonovo's δ and the Kolmogorov-Smirnov measures. Most of the measures can be used to identify unimportant inputs under Scenario 2 in the case of independent inputs. Shapley effects is an exception as it is unable to detect the input under Scenario 2. Total effect indices can be used to identify the unimportant inputs for both Scenario except for correlated input for Scenario 1.

2) Prioritising elicitation does not mean just applying sensitivity analysis measures, e.g. main effects to rank inputs based on their importance. Using a counter-example, we have shown that input having a large main effect is not necessarily selected for elicitation.

3) Robust global sensitivity analysis methods that we have reviewed consider small changes in input distribution range (perturbation). Thus, robust global sensitivity analysis concepts are not quite appropriate for prioritising elicitation because the distribution of inputs might have a big change from the analyst distribution to the expert distribution.

4) The Beta model is a novel distribution model for modelling expert opinions. The model has two parameters α and β with two additional parameters: the function of α and β. Parameter μ_i is the expectation of the Beta distribution so that it has the form of $\alpha/(\alpha+\beta)$ and has a uniform distribution. Parameter φ_i is the precision of the model which is equal to $\alpha+\beta$ and $log\varphi_i$ is also uniformly distributed. If the model is linear, it is worth learning parameter μ_i, but it is not worth learning parameter $log\varphi_i$. However, if the model is non-linear, we need to consider learning μ_i and φ_i.

5) The Grid-Based method is effective for computing sensitivity measures such as Borgonovo's δ and main effect index. The numerical experiments suggest that the number of runs 100,000 provides stable indices, and the number of grid cells 20 x 20 is sufficient to compute indices. The number of Simpson's points for computing Borgonovo's δ does not affect the computation of indices, so we can use a small number of points to reduce the computation time. The Grid-Based method is more efficient than the Monte Carlo method as it will simultaneously provide the estimates of sensitivity index for model inputs and parameters μ_i and φ_i.

§ 7.2 Future Works

There are two ideas related to sensitivity analysis and expert judgment, which can be topics for further research. The topics are briefly explained as follows.

1) In this thesis, we have discussed prioritising elicitation for inputs in mathematical models. Inputs in the model may be correlated. Therefore, the author proposes that the following research prioritise elicitation for correlated or dependent inputs.

2) Another topic the author still does not continue to do is the sensitivity analysis for chained models. This topic is motivated by the phenomena in the real world that a mathematical model might connect to other models. Consider two models, and the output of the first model is an input for the second model. We can see this as an example of a chained model.

References

Baucells, M. & Borgonovo, E. (2013), 'Invariant Probabilistic Sensitivity Analysis', *Management Science* **59**(11), 2536–2549.

Borgonovo, E. (2007), 'A new uncertainty importance measure', *Reliability Engineering and System Safety* **92**(6), 771–784.

Borgonovo, E. (2017), *Sensitivity Analysis: An Introduction for the Management Scientist*, International Series in Operations Research & Management Science, Springer International Publishing.
URL: *https://books.google.co.uk/books?id=H0K4DgAAQBAJ*

Borgonovo, E., Hazen, G. B. & Plischke, E. (2016), 'A common rationale for global sensitivity measures and their estimation', *Risk Analysis* **36**(10), 1871–1895.

Borgonovo, E., Morris, M. D. & Plischke, E. (2018), 'Functional anova with multiple distributions: Implications for the sensitivity analysis of computer experiments', *SIAM/ASA Journal on Uncertainty Quantification* **6**(1), 397–427.

Borgonovo, E., Tarantola, S., Plischke, E. & Morris, M. D. (2014), 'Transformations and invariance in the sensitivity analysis of computer experiments', *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **76**(5), 925–947.

Campolongo, F., Cariboni, J. & Saltelli, A. (2007), 'An effective screening design for sensitivity analysis of large models', *Environmental Modelling and Software* **22**(10), 1509–1518.

Chun, M.-H., Han, S.-J. & Tak, N.-I. (2000), 'An uncertainty importance measure using a distance metric for the change in a cumulative distribution function', *Reliability Engineering & System Safety* **70**(3), 313–321.

Constantine, P. G. & Doostan, A. (2017), 'Time-dependent global sensitivity analysis with active subspaces for a lithium ion battery model'.

Cooke, R. M. (1991), *Experts in Uncertainty: Opinion and Subjective Probability in Science (Environmental Ethics and Science Policy)*, Vol. 44.

Cukier, R. I., Levine, H. B. & Shuler, K. E. (1977), 'Nonlinear sensitivity analysis of multiparameter model systems', *Journal of Physical Chemistry* **81**(25), 2365–2366.

French, S. (1980), 'Updating of Belief in the Light of Someone Else ' s Opinion', *Journal of Royal Statistical Society* **143**(1), 43–48.

Gao, L. & Bryan, B. A. (2016), 'Incorporating deep uncertainty into the elementary effects method for robust global sensitivity analysis', *Ecological Modelling* .

Gauchi, J. P., Bensadoun, A., Colas, F. & Colbach, N. (2017), 'Metamodeling and global sensitivity analysis for computer models with correlated inputs: A practical approach tested with a 3D light interception computer model', *Environmental Modelling and Software* **92**, 40–56.

Gauchy, C., Stenger, J., Sueur, R. & Iooss, B. (2020), 'An information geometry approach for robustness analysis in uncertainty quantification of computer codes'.
URL: *http://arxiv.org/abs/2008.03060*

Gelfand, A. E., Mallick, B. K. & Dey, D. K. (1995), 'Modeling expert opinion arising as a partial probabilistic specification', *Journal of the American Statistical Association* **90**(430), 598–604.

Grigore, B., Peters, J., Hyde, C. & Stein, K. (2017), 'EXPLICIT: A feasibility study of remote expert elicitation in health technology assessment', *BMC Medical Informatics and Decision Making* **17**(1), 1–10.

Hanea, A. M., Hemming, V. & Nane, G. F. (2021), 'Uncertainty Quantification with Experts: Present Status and Research Needs', *Risk Analysis* **0**(0).

Hanea, A. M. & Nane, G. F. (2021), An In-Depth Perspective on the Classical Model, *in* 'International Series in Operations Research and Management Science', Vol. 293, Springer, Cham, Switzerland, pp. 225–256.

Hart, J. & Gremaud, P. (2019*a*), 'Robustness of the Sobol' indices to distributional uncertainty', *International Journal for Uncertainty Quantification* **9**(5), 453–469.

Hart, J. & Gremaud, P. (2019*b*), 'Robustness of the Sobol' indices to marginal distributional uncertainty', *International Journal for Uncertainty Quantification* **9**(5), 453–469.

Hastie, T. & Tibshirani, R. (1990), *Generalized Additive Models*, Chapman and Hall, London.

Health, M. (2018), 'Reports of Fundamental Health Research'.
URL: *https://www.litbang.kemkes.go.id/laporan-riset-kesehatan-dasar-riskesdas/*

Hemming, V., Burgman, M. A., Hanea, A. M., McBride, M. F. & Wintle, B. C. (2018), 'A practical guide to structured expert elicitation using the IDEA protocol', *Methods in Ecology and Evolution* **9**(1), 169–180.

Howard, R. A. (1966), 'Information value theory', *IEEE transactions on systems science and cybernetics* **2**(1), 22–26.

Iman, R. L. & Conover, W. J. (1979), 'The use of the rank transform in regression', *Technometrics* **21**(4), 499–509.

Indonesia, S. (2021), 'Percentage of Smokers Age 15 years above by Province'.
URL: *https://www.bps.go.id/indicator/30/1435/1/percentage-of-smokers-age-15-years-by-province.html*

Iooss, B. & Lemaître, P. (2014), A review on global sensitivity analysis methods.
URL: *http://arxiv.org/abs/1404.2405*

Ishigami, T. & Homma, T. (1990), An Importance Quantification Technique in Uncertainty Analysis for Computer Models, *in* 'Proceedings., First International Symposium IEEE', pp. 398–403.

Jones, D. R., Schonlau, M. & W. J. Welch (1998), 'Efficient Global Optimization of Expensive Black-Box Functions," , vol. 13, no. 4, pp. 455-492, 1998.', *Journal of Global Optimization* **10**, 455–492.

Kennedy, M. C. & O'Hagan, A. (2001), 'Bayesian calibration of computer models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**(3), 425–464.
URL: *http://doi.wiley.com/10.1111/1467-9868.00294*

Klepper, O. (1997), 'Multivariate aspects of model uncertainty analysis: Tools for sensitivity analysis and calibration', *Ecological Modelling* **101**(1), 1–13.

Knowles, J. (2006), 'ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems', *IEEE Transactions on Evolutionary Computation* **10**(1), 50–66.

Kucherenko, S., Tarantola, S. & Annoni, P. (2012), 'Estimation of global sensitivity indices for models with dependent variables', *Computer Physics Communications* **183**(4), 937–946.
URL: *http://dx.doi.org/10.1016/j.cpc.2011.12.020*

Lacirignola, M., Blanc, P., Girard, R., Pérez-López, P. & Blanc, I. (2017), 'LCA of emerging technologies: addressing high uncertainty on inputs' variability when performing global sensitivity analysis', *Science of the Total Environment* **578**, 268–280.
URL: *http://dx.doi.org/10.1016/j.scitotenv.2016.10.066*

Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W. & Spracklen, D. V. (2011), 'Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters', *Atmospheric Chemistry and Physics* **11**(23), 12253–12273.

Lindley, D., Tversky, A. & Brown, R. (1979), 'On the Reconciliation of Probability Assessments', *Journal of Royal Statistical Society* **142**(2), 146–180.

Meynaoui, A., Marrel, A. & Laurent-bonneau, B. (2019), 'New statistical methodology for second level global sensitivity analysis', pp. 1–29.

Morgan, M. G. (2014), 'Use (and abuse) of expert elicitation in support of decision making for public policy', *Proceedings of the National Academy of Sciences of the United States of America* **111**(20), 7176–7184.

Morris, M. D. (1991), 'Factorial sampling plans for preliminary computational experiments', *Technometrics* **33**(2), 161–174.

Oakley, J. (2009), 'Decision-theoretic sensitivity analysis for complex computer models', *Technometrics* **51**(2), 121–129.

Oakley, J. (2020), *SHELF: Tools to Support the Sheffield Elicitation Framework*. R package version 1.7.0.
URL: *https://CRAN.R-project.org/package=SHELF*

Oakley, J. & O'Hagan, A. (2004), 'Probabilistic sensitivity analysis of complex models: a Bayesian approach', *Journal of The Royal Statistical Society. Series B* **66**(3), 751–769.

Oakley, J. & O'Hagan, A. (2019), SHELF: the Sheffield Elicitation Framework (version 4), Technical report, School of Mathematics and Statistics, University of Sheffield, UK.
URL: *http://tonyohagan.co.uk/shelf*

O'Hagan, A. (2018b), 'Probabilistic judgements e-learning'.
URL: *http://www.tonyohagan.co.uk/shelf/ecourse.html*

O'Hagan, A. (2019), 'Expert Knowledge Elicitation: Subjective but Scientific', *American Statistician* **73**(sup1), 69–81.

O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, J. R., Garthwhite, P., Jenkinson, D. J., Oakley, J.E & Rakow, T. (2006), *Uncertain Judgements: Eliciting Experts' Probabilities*, John Wiley and Sons Ltd, Chichester.

Owen, A. B. (2014), 'Sobol' Indices and Shapley Value', *SIAM/ASA Journal on Uncertainty Quantification* **2**(1), 245–251.

Park, C. K. & Ahn, K. I. (1994), 'A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment', *Reliability Engineering and System Safety* **46**(3), 253–261.

Pearson, K. (1905), *On the General Theory of Skew Correlation and Non-Linear Regression*, Vol. XIV. Mathematical Contributions to The Theory of Evolution. Drappers' Company Research Memoirs. Dulau and Co., London.

Pianosi, F. & Wagener, T. (2015), 'A simple and efficient method for global sensitivity analysis based oncumulative distribution functions', *Environmental Modelling and Software* **67**, 1–11.
URL: *http://dx.doi.org/10.1016/j.envsoft.2015.01.004*

Plischke, E., Borgonovo, E. & Smith, C. L. (2013), 'Global sensitivity measures from given data', *European Journal of Operational Research* **226**(3), 536–550.

Ratto, M., Pagano, A. & Young, P. (2007), 'State Dependent Parameter metamodelling and sensitivity analysis', *Computer Physics Communications* **177**(11), 863–876.

Razavi, S. & Gupta, H. V. (2016a), 'Water Resources Research', *Journal of the American Water Resources Association* **5**(3), 2–2.

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J. H., Jakeman, J.,

Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., Smith, S., Sheikholeslami, R., Hosseini, N., Asadzadeh, M., Puy, A., Kucherenko, S. & Maier, H. R. (2021), 'The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support', *Environ. Model. Softw.* **137**(December 2020).

Russi, T., Packard, A., Feeley, R. & Frenklach, M. (2008), 'Sensitivity analysis of uncertainty in model prediction', *Journal of Physical Chemistry A* **112**(12), 2579–2588.

Saltelli, A., Chan, K. & Scott, E. M. (2000), *Sensitivity Analysis*, John Wiley and Sons Ltd, West Sussex.

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

Saltelli, A., Tarantola, S., Campolongo, F. & Ratto, M. (2004), *Sensitivity analysis in practice: a guide to assessing scientific models (Google eBook)*, John Wiley and Sons Ltd.
URL: *http://books.google.com/books?id=NsAVmohPNpQC&pgis=1*

Saltelli, A., Tarantola, S. & Chan, K. (1999), 'A quantitative model-independent method for global sensitivity analysis of model output', *Technometrics* **41**(1), 39–56.

Shao, Q., Younes, A., Fahs, M. & Mara, T. A. (2017), 'Bayesian sparse polynomial chaos expansion for global sensitivity analysis', *Computer Methods in Applied Mechanics and Engineering* **318**, 474–496.

Sobol, I. M. (1993), 'Sensitivity analysis for nonlinear mathematical models', *Math. Model. Computer.Exp* **1**(4), 407–414.

Sobol, I. M. (2001), 'Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates', *Mathematics and Computers in Simulation* **55**(1), 271–280.
URL: *http://doi.wiley.com/10.1002/wilm.42820050114*

Song, E., Nelson, B. L. & Staum, J. (2016), 'Shapley effects for global sensitivity analysis: Theory and computation', *SIAM-ASA Journal on Uncertainty Quantification* **4**(1), 1060–1083.

Soofi, E. S. (1994), 'Capturing the intangible concept of information', *Journal of the American Statistical Association* **89**(428), 1243–1254.

Spearman, C. (1987), 'The proof and measurement of association between two things', *The American Journal of Psychology* **100**(3/4), 441–471.
URL: *http://www.jstor.org/stable/1422689*

Srivastava, A., Subramaniyan, A. K. & Wang, L. (2017), 'Analytical global sensitivity analysis with Gaussian processes', pp. 235–250.

Strong, M. & Oakley, J. E. (2013), 'An efficient method for computing single-parameter partial expected value of perfect information', *Medical Decision Making* **33**(6), 755–766.

Strong, M., Oakley, J. E. & Chilcott, J. (2012), 'Managing structural uncertainty in health economic decision models: A discrepancy approach', *Journal of the Royal Statistical Society. Series C: Applied Statistics* **61**(1), 25–45.

Tohsato, Y., Ikuta, K., Shionoya, A., Mazaki, Y. & Ito, M. (2013), 'Parameter optimization and sensitivity analysis for large kinetic models using a real-coded genetic algorithm', *Gene* **518**(1), 84–90.
URL: *http://dx.doi.org/10.1016/j.gene.2012.11.080*

Vanlier, J., Tiemann, C. A., Hilbers, P. A. & van Riel, N. A. (2012), 'An integrated strategy for prediction uncertainty analysis', *Bioinformatics* **28**(8), 1130–1135.

Xu, M., Wilson, A. & Dent, C. J. (2016), 'Calibration and sensitivity analysis of long-term generation investment models using Bayesian emulation', *Sustainable Energy, Grids and Networks* **5**, 58–69.

Appendix

Let us consider the similarity of two variance terms

$$Var\Big(\sum_{i=1}^{p} f_i(X_i)\Big) = \sum_{i=1}^{p} Var\big(f(X_i)\big) \tag{A.1}$$

as $X_1, \ldots, X_p$ are independent, and we have

$$Var\big(f(X_i)\big) = Var_{X_i}(\mathbb{E}[Y|X_i]). \tag{A.2}$$

Now we consider component functions that contain X_i

$$\begin{aligned} Var\Big(f_i(X_i) + f_{i,j}(X_i, X_j)\Big) = Var\Big(f_i(X_i)\Big) + Var\Big(f_{i,j}(X_i, X_j)\Big) \\ + 2\,\mathrm{Cov}\,\big(f_i(X_i), f_{i,j}(X_i, X_j)\big). \end{aligned} \tag{A.3}$$

Note that all terms in equation (2.19) have expectation zero. For example,

$$\begin{aligned} \mathrm{Cov}\,\Big(f_i(X_i), f_{i,j}(X_i, X_j)\Big) &= \mathbb{E}_{X_i X_j}\Big(\big(\mathbb{E}[Y|X_i] - f_0\big)\big(\mathbb{E}[Y|X_i, X_j] - \mathbb{E}[Y|X_i] - \mathbb{E}[Y|X_j] + f_0\big)\Big) \\ &= \mathbb{E}_{X_i}\Big(\big(\mathbb{E}[Y|X_i] - f_0\big)\big(\mathbb{E}_{X_j}(\mathbb{E}[Y|X_i, X_j] - \mathbb{E}[Y|X_i] - \mathbb{E}[Y|X_j] + f_0)\big)\Big) \\ &= \mathbb{E}_{X_i}\Big(\big(\mathbb{E}[Y|X_i] - f_0\big)\big(\mathbb{E}[Y|X_i] - \mathbb{E}[Y|X_i] - \mathbb{E}[Y] + f_0)\big)\Big) \\ &= 0. \end{aligned} \tag{A.4}$$

Now we expand

$$\begin{aligned} Var\Big(f_{i,j}\big(X_i, X_j\big)\Big) &= Var_{X_i,X_j}\big(\mathbb{E}(Y|X_i, X_j)\big) + Var_{X_i,X_j}\big(f_i(X_i)\big) + Var_{X_i,X_j}\big(f_j(X_j)\big) \\ &\quad - 2\,\mathrm{Cov}\,\big(\mathbb{E}(Y|X_i, X_j), \mathbb{E}(Y|X_i)\big) - 2\,\mathrm{Cov}\,\big(\mathbb{E}(Y|X_i, X_j), \mathbb{E}(Y|X_j)\big) \\ &\quad - 2\,\mathrm{Cov}\,\big(\mathbb{E}(Y|X_i), \mathbb{E}(Y|X_j)\big) \end{aligned} \tag{A.5}$$

and check the covariances

$$\begin{aligned}\text{Cov}\Big(\mathbb{E}\big(Y|X_i,X_j\big),\mathbb{E}\big(Y|X_i\big)\Big) &= \mathbb{E}_{X_iX_j}\Big(\mathbb{E}\big(Y|X_i,X_j\big)\mathbb{E}\big(Y|X_i\big)\Big)-\mathbb{E}(Y)^2\\ &= \mathbb{E}_{X_i}\Big(\mathbb{E}\big(Y|X_i\big)\mathbb{E}_{X_j}\big(\mathbb{E}[Y|X_i,X_j]\big)\Big)-\mathbb{E}(Y)^2\\ &= \mathbb{E}_{X_i}\Big(\mathbb{E}\big(Y|X_i\big)^2\Big)-\mathbb{E}(Y)^2\\ &= Var_{X_i}\Big(\mathbb{E}\big(Y|X_i\big)^2\Big),\end{aligned} \tag{A.6}$$

while another covariance is zero. Thus

$$Var\Big(f_{i,j}\big(X_i,X_j\big)\Big) = Var_{X_i,X_j}\big(\mathbb{E}(Y|X_i,X_j)\big)+Var_{X_i}\big(\mathbb{E}(Y|X_i)\big)+Var_{X_j}\big(\mathbb{E}(Y|X_j)\big). \tag{A.7}$$

According to (A.2) and (A.7) equation (A.3) has the final result as follows

$$Var\big(f_i(X_i)+f_{i,j}(X_i,X_j)\big) = Var_{X_i,X_j}\big(\mathbb{E}(Y|X_i,X_j)\big)+2Var_{X_i}\big(\mathbb{E}(Y|X_i)\big)+Var_{X_j}\big(\mathbb{E}(Y|X_j)\big). \tag{A.8}$$