7-27-2018 1:30 PM

The Statistical Exploration in the G-expectation Framework: The Pseudo Simulation and Estimation of Variance Uncertainty

Yifan Li, *The University of Western Ontario*

Supervisor: Kulperger, Reg, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Statistics and Actuarial Sciences

Abstract

The G-expectation framework, motivated by problems with *uncertainty*, is a new generalization of the classical probability framework. Similar to the Choquet expectation, the G-expectation can be represented as the supremum of a class of linear expectations. In the past two decades, it has developed into a complete stochastic structure connected with a large family of nonlinear PDEs. Nonetheless, to apply it to real-world problems with uncertainty, it is fundamentally necessary to build up the associated statistical methodology.

This thesis explores the *computation, simulation, and estimation* of the G-normal distribution (a typical distribution with variance uncertainty) by constructing a new substructure called the *Semi-G-normal distribution* which provides the transition from classical normal to G-normal distribution. Interestingly, it also gives an efficient iterative scheme to stochastically solve the nonlinear *Black-Scholes-Barenblatt equation with volatility uncertainty*. This thesis is the theoretical and technical preparation for the future industrial application of G-framework.

Keywords: Uncertainty, G-expectation framework, Semi-G-normal distribution, Sublinear expectation, Statistical theory, Black-Scholes-Barenblatt equation

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background of the *G*-framework

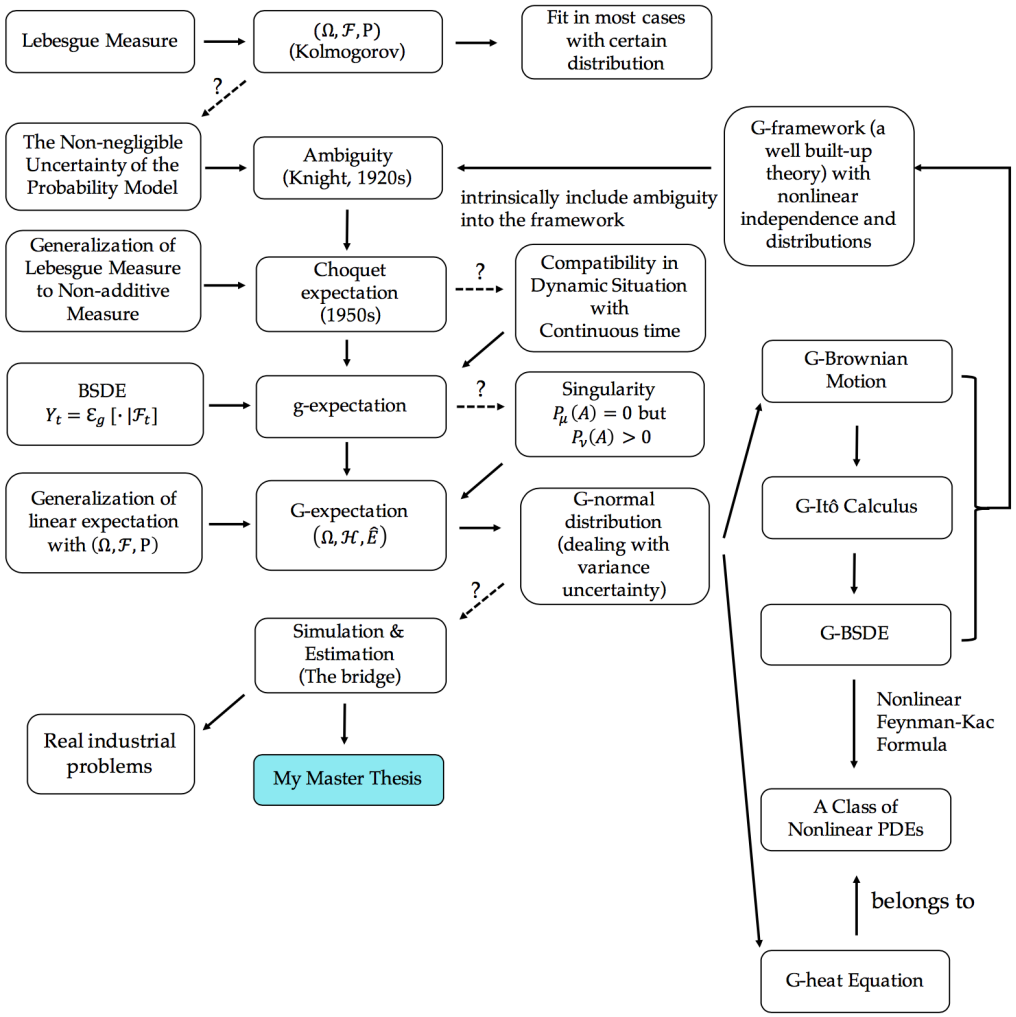


Figure 1.1: Overview of the Background and the Position of My Current Research

1.1.1 A simple story: What is *Uncertainty*?

For general readers, let us start from a very simple example (which is taken from Ellsberg (1961)). Here we have an urn (called Urn I) containing 30 red balls and 60 black balls. One ball will be randomly drawn from the urn, then the colour of the ball will decide the money you get. Give you two gambles to consider (Table 1.1).

Table 1.1: The two gambles (Urn I)

Gamble	Red (30)	Black (60)
A	\$100	\$0
B	\$0	\$100

Gamble A or B, which one do you prefer? I bet it should not take you even a second to choose Gamble B or "a bet on black" (as long as you do not hate money). I also believe most readers (who prefer \$100 to \$0) will give me this obvious reason: Since the ball is more likely to be black, why not bet on black to get that \$100 with higher chance?

In fact, this is one of the common senses in gambling strategies, which appeared even far before the notion called *probability*. The requirements of making better strategies in gambling practice has actually motivated the early study of probability. To systematically describe stochastic phenomena in a more rigorous way, Kolmogorov (1956) constructed the axiomatic system of probability theory (which we call the classical probability framework) based on the additive Lebesgue measure. As we know, this classical framework gives a complete construction of the random variables and expectation. We can actually use the language of the classical probability framework (which should not be hard for any students with elementary probability knowledge) to explain and describe people's preference in the scenario of Urn I (which is shown as follows).

In the classical probability space $(\Omega, \mathcal{F}, P)$, let $X : \Omega \to \{0, 100\}$ be a random variable representing your income after the experiment. Since each ball in the urn has equal chance to be drawn, in Gamble A, the distribution of X can be described by the following probability law mapping P_A:

$$\begin{pmatrix} X & 100 & 0 \\ P_A & 1/3 & 2/3 \end{pmatrix}. \tag{1.1}$$

Hence, the expected income in Gamble A can be expressed by the classical expectation of X under the law P_A, namely, $E_A[X] = 100 \times P_A(X = 100) + 0 \times P_A(X = 0) = 100/3$. Similarly, for Gamble B, we have the probability law P_B:

$$\begin{pmatrix} X & 0 & 100 \\ P_B & 1/3 & 2/3 \end{pmatrix}. \tag{1.2}$$

Then the expected income from Gamble B is $E_B[X] = 200/3$. In most people's mind, in order to maximize their "expected income", they prefer Gamble B to A because of the underlying quantified relation $E_B[X] > E_A[X]$. This is also a simplified version of the *expected utility theory* (established by Von Neumann and Morgenstern (1945)) through defining a utility function $U : \{0, 100\} \to \mathbb{R}$ to show people's preference on the income (in our case, we can define a U

such that $U(100) > U(0)$). In general, people choose one strategy if and only if it can maximize their expected utility. In the setting of Urn I, people prefer B to A if and only if B has larger expected utility or explicitly, $E_B[U(X)] > E_A[U(X)]$ (if they are equal, people should not have any preference over these two gambles), which is equivalent to $P_B(X = 100) > P_A(X = 100)$. Considering the probability regarding the colors the ball we draw might be and the rules in Table 1.2 (Gamble B is "a bet on black" and Gamble A is "a bet on red"), people prefer B to A if and only if P(draw a black ball) > P(draw a red ball) (they believe it is more likely to draw a black ball than a red one), which is consistent with the common gambling intuition.

Let us do one step further and turn to another urn (called Urn II) containing 30 red balls and 60 balls that are black and yellow (in some fixed but *unknown proportion*). Suppose we have x black balls and $60 - x$ yellow balls where x is some unknown integer in $[0, 60]$. Again, one ball is to be randomly drawn from the urn, the colour of which determines the money you get. You are provided with two gambles to reflect on, illustrated by Table 1.2.

Table 1.2: The first pair of gambles (Urn II)

Gamble	Red (30)	Black (x)	Yellow ($60 - x$)
A	\$100	\$0	\$0
B	\$0	\$100	\$0

Gamble A is "if the ball is red, you get \$100, otherwise you get nothing" where we know there are 30 red balls. Gamble B is "if the ball is black, you get \$100, otherwise nothing" where we only know there are $[0, 60]$ black ones. Do you prefer Gamble A or B? In other words, do you prefer "a bet on red" or "a bet on black"?

Now consider another pair of gambles C and D (Table 1.3).

Table 1.3: The second pair of gambles (Urn II)

Gamble	Red (30)	Black (x)	Yellow ($60 - x$)
C	\$100	\$0	\$100
D	\$0	\$100	\$100

Gamble C is a bet on "not black" and Gamble D is a bet on "not red". Which of them do you prefer? (Take your time!)

According to the results from Ellsberg (1961), we have two common responses:

1. **Response i** (very frequent): Gamble A is preferred to B and Gamble D is preferred to C;
2. **Response ii** (less frequent): Gamble B is preferred to A and Gamble C is preferred to D.

Usually you may fall into one of these two types. Before thinking about what was in your mind several minutes ago, let us try some probabilistic ways to explain these common responses.

Since the third column in Table 1.2 has equal two entries (so is Table 1.3), "whether the ball is yellow or not" is irrelevant in both comparisons ("A versus B" and "C versus D"). Based on the intuition of the *classical expected utility* which works perfectly well for Urn I, from

Response i, people prefer Gamble A to B if and only if they believe drawing a red ball is more likely than drawing a black one ("Reds are more than Blacks"). However, they also prefer Gamble D to C which is equivalent to their belief that it is more likely to draw a black ball than a red one ("Blacks are more than Reds"). This leads to a contradiction.

To better show the contradiction, if we apply the theory of *classical expected utility* theory, again, we first need to set a utility function $U : 0, 100 \to \mathbb{R}$ such that $U(100) > U(0)$ intuitively meaning people prefer \$100 to \$0. For instance, let

$$U(x) = \mathbb{I}_{\{x=100\}} := \begin{cases} 1 & x = 100 \\ 0 & x = 0 \end{cases}.$$

We still use X to denote your income after the experiment, based on the unknown x and letting $y := x/90 (\in [0, 2/3])$ which is the proportion of black balls. The distribution of X and $U(X)$, under different gambles, can be summarized by the following probability laws:

$$\begin{pmatrix} X & 0 & 100 \\ U(X) & 0 & 1 \\ P_A & 2/3 & 1/3 \\ P_B & 1-y & y \\ P_C & y & 1-y \\ P_D & 1/3 & 2/3 \end{pmatrix}. \tag{1.3}$$

Associated with the law (1.3) we also have, under probability P, $E_P[U(X)] = E_P[\mathbb{I}_{\{X=100\}}] = 1 \times P(\mathbb{I}_{\{X=100\}} = 1) = P(X = 100)$. Then according to the theory of expected utility, people's preferences are characterized by maximizing the expected utility. **Response i** showing Gamble A is preferred to B indicates $E_A[U(X)] > E_B[U(X)]$, which, by the probability law (1.3), can be expressed as $P_A(X = 100) > P_B(X = 100)$ which is equivalent to $1/3 > y$. We know y can be treated as *the unknown proportion of black balls* so "A is preferred to B" can be explained by

$$\text{the proportion of black balls} < 1/3.$$

Meanwhile, **Response i** also says Gamble D is preferred to C then $E_D[U(X)] > E_C[U(X)]$, which means $P_D(X = 100) > P_C(X = 100)$ implying $2/3 > 1 - y$ or

$$\text{the proportion of black balls} > 1/3.$$

This leads to an obvious contradiction! (Noting that Urn II is not in some quantum world, although the number of black balls in Urn II is unknown, it should be some fixed one.) Readers can check **Response ii** also gives a logically similar contradiction. Both of them *violate* the theory of classical expected utility.

Wait a moment. I know perhaps you were not really doing this kind of computation, since the proportion of black balls is unknown, how can we "pre-define" the threshold of y to compare the expected utility of two gambles? If not this, what kind of struggling was happening in your mind just now to make the strategy?

Actually, people (who give **Response i and ii**) usually tend to treat the proportion of black balls not a certain number y, but an *uncertain* one varying in a range ($[0, 2/3]$), which is *the uncertainty interval* they do care about, especially the two *bounds* (corresponding to the worst

and best cases). The unknown proportion of black balls is the *uncertainty* here since we do not have information about this. In this spirit, the probability law (1.3) should be changed to a new version with "uncertainty":

$$\begin{pmatrix} \text{Gamble} & P(X=100) \\ \text{A} & 1/3 \\ \text{B} & [0,2/3] \\ \text{C} & [1/3,1] \\ \text{D} & 2/3 \end{pmatrix}. \tag{1.4}$$

In Gamble A versus B, some people prefers Gamble A because they worry about the lower bound of the uncertainty interval given by B; ("what if there is no black balls, then I have no chance to get the money in B, at least Gamble A has $1/3$ win rate.") This kind of worry makes them they hate the uncertainty in B. This type of people have the so-called *uncertainty aversion*, then they will be in favour of D in the comparison between Gamble C and D to avoid the uncertainty in Gamble C (especially the lower bound).

Explicitly speaking, people with uncertainty aversion actually think about the "worst case" of all possible scenarios. For instance, in Gamble B, their "expected utility" in the worst case is actually the *minimum* of expected income in all possible settings of y, motivating us to reflect on a set of distributions or probability laws, namely, $\mathcal{Q}_B := \{P_B : P_B(X=0) = 1-y, P_B(X=100) = y, y \in [0,2/3]\}$, rather than the single distribution of X in Urn I governed by the law (1.1) or (1.2). Then the "expected utility" can be written as

$$\hat{\mathbb{E}}_B[U(X)] := \min_{P_B \in \mathcal{Q}_B} E_{P_B}[U(X)] = \min_{y \in [0,2/3]} y = 0. \tag{1.5}$$

For Gamble A, since we only have one probability law for X (with no uncertainty), the set of distributions $\mathcal{Q}_A := \{P_A : P_A(X=0) = 2/3, P_B(X=100) = 1/3\}$ degenerates to a singleton, so the expected income is

$$\hat{\mathbb{E}}_A[U(X)] := E_{P_A}[U(X)] = 1/3.$$

Therefore, we have

$$\hat{\mathbb{E}}_A[U(X)] > \hat{\mathbb{E}}_B[U(X)], \tag{1.6}$$

which describes the preference of people with uncertainty aversion in Gamble A versus B. Similarly, when people considering "the worst case" meet with Gamble C and D, since $\mathcal{Q}_C := \{P_C : P_C(X=0) = y, P_B(X=100) = 1-y, y \in [0,2/3]\}$, the expected income is

$$\hat{\mathbb{E}}_C[U(X)] := \min_{P_C \in \mathcal{Q}_C} E_{P_C}[U(X)] = \min_{y \in [0,2/3]} 1-y = 1/3,$$

while the expected gain of Gamble D is

$$\hat{\mathbb{E}}_D[U(X)] := E_{P_D}[U(X)] = 2/3,$$

so we have

$$\hat{\mathbb{E}}_D[U(X)] > \hat{\mathbb{E}}_C[U(X)]. \tag{1.7}$$

The inequalities (1.6) and (1.7) are consistent with the preference of people with uncertainty aversion and able to describe their strong worry about the worst case under uncertainty.

Meanwhile, there are another type of people (perhaps less frequent) preferring Gamble B to *seek the uncertainty* in it especially its upper bound; ("what if there is 60 black balls, then I have 2/3 chance, doubling the rate I get from A.") This inclination will also drive them to choose C when comparing Gamble C and D. These type of people have the feature called *uncertainty seeking*, whose expected utility can be expressed by the maximum of all possible cases since they are looking for the "best income", so we only need to change the minimum to maximum in Equation (1.5) to mathematically describe their preference.

Actually, the expectation $\hat{\mathbb{E}}$ in Equation (1.5) can be treated as the *Choquet expectation* (so is the one replacing min with max), which is a nonlinear expectation able to be represented as the supremum (or equivalently, infimum, by adding a minus sign before the random variable) of a class of linear expectations under additive probability measures. Choquet (1954) generalizes the Lebesgue integral to non-additive measures so as to get the Choquet expectation.

1.1.2 Literature review: How do we deal with *Uncertainty*?

I will briefly review the historic development of the methodology to deal with uncertainty (also shown in Figure 1.1), in which the order and meaning of events partially refers to the survey by Peng (2017).

Knight (1921) gave the notion of *Knightian Uncertainty* (also known as *Ambiguity* in finance) to distinguish it from *risk* in his work Risk, Uncertainty, and Profit by saying:

"Uncertainty must be taken in a sense radically distinct from the familiar notion of Risk, from which it has never been properly separated.... The essential fact is that 'risk' means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character."

After the journey we have in Section 1.1.1, you may notice that the unknown proportion of black balls in Urn II exactly brings us the *uncertainty* in the distribution of the income X, forcing us to consider a *set* of distributions to characterize or cover this uncertainty. In fact, what we played with Urn II is the famous *Ellsberg Paradox* which is a mind experiment proposed by Ellsberg (1961), showing the violation of von Neumann-Morgenstern expected utility theory in the scenario with uncertainty and strongly motivating the construction of the new theory of expected utility under the Choquet expectation by Schmeidler (1989) later on.

However, the methods based on Choquet expectation cannot deal with *the uncertainty in dynamic situations* (especially with continuously changing time) because it is difficult to define the conditional Choquet expectation (conditional on the filtration until time t), but the real world we are facing has fundamental dynamic features. Fortunately, Chen and Epstein (2002) efficiently made progress in this problem in the setting of a sublinear expectation defined through the BSDE (Backward Stochastic Differential Equation) called g-expectation (initially developed by Peng (1997)), which nicely borrows the dynamic property of BSDEs to define its conditional expectation.

In principle, g-expectation can deal with any set of probability measures $\{P_\theta\}_{\theta\in\Theta}$ dominated by a reference probability P. Nonetheless, g-expectation will fail when stepping into the *singular* scenario (that is, there exists an event A such that $P(A) = 0$ but $P_\theta(A) > 0$), which is common in practice like the problem with *volatility uncertainty*.

It looks like in classical probability framework, we cannot handle the uncertainty both with singularity and in dynamic situations (no matter using the Choquet's design or the g-

expectation governed by BSDEs). It took researchers many years to realize that it is necessary to jump out of the Kolmogorov's system $(\Omega, \mathcal{F}, P)$, start from scratch and directly construct a new generalized probability framework to describe the uncertainty, which was established by Peng (2004, 2007a, 2008) and further developed by the academic group led by him, called the G-expectation framework $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$.

In the past two decades (since its establishment in 2000s), the G-expectation framework has developed into a complete probabilistic structure with its own stochastic calculus and connection with a large family of nonlinear PDEs. It starts from the G-expectation (or sublinear expectation) $\hat{\mathbb{E}}$ to redefine what are independence and identical distribution, from which it induces the nonlinear version of "constants" (Maximal distribution with mean-uncertainty) and "normal distribution" (G-normal distribution with variance uncertainty). The independence in this framework has the *sequential order* which usually can not be reversed. Intuitively, this kind of design partially comes from the spirit of backward analysis similar to the analysis in BSDEs. From the starting axioms of $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$, we are able to completely and systematically construct all the basic and important results in probability, stochastic analysis and statistical theory, which actually give a brand-new understanding of those stochastic concepts under the uncertainty setting and also generalize the classical results in a non-trivial way with much weaker and more general conditions. Specially speaking, the G-framework has its own probabilistic inequalities, Central Limit Theorem (CLT) (Peng (2007b), Hu and Zhou (2015)) with convergence rate by Song (2017), Law of Large Numbers (LLN) (Chen et al. (2013)) and also G-Itô-stochastic calculus based on the construction of G-Brownian motion (Denis et al. (2011)), in which there are diffusion processes driven by G-Brownian motion including the BSDEs, the so called G-BSDEs. More complete collection of results can be found in Peng (2007a, 2008). Furthermore, similar to the counterparts in classical framework, Hu et al. (2014) shows that the G-BSDEs are connected by the Feynman-Kac formula in G-framework with a large class of fully nonlinear PDEs (able to be applied to control problems with uncertainty, especially in financial background). If we can numerically solve the G-BSDEs, it is very promising that we will be able to stochastically solve the high-dimensional fully nonlinear PDEs (whose nonlinearity and curse of dimensionality make most classical numerical PDE methods fail).

However, problems are arising from both the academy and industry:

1. If the G-framework intends to deal with the problems with uncertainty, given a real problem with dataset, how can we do the *estimation* of those parameters to fit into the setting of G-framework?
2. If the G-Brownian motion and the G-BSDE are able to cover or capture the volatility uncertainty (especially solving those nonlinear PDEs), is it possible to simulate the G-Brownian motion (or the G-normal distribution) as well as the G-BSDE, and how do we compute the corresponding G-expectation of those G-itô integral?

Before answer the questions above, we need to consider a crucial one:

How do we statistically and numerically deal with the distributions in the G-framework?

Typically, how can we better understand and handle the G-normal distribution? Actually, another key problem hidden here is the *sequential independence* attached with the distributions. We may also notice that, in general, for this new framework, the associated computation and statistics is the fundamentally unavoidable and necessary bridge connecting the G-framework with those

industrial problems with uncertainty and the implementation of the new stochastic nonlinear PDE methods based on the diffusion processes in the G-framework. This thesis is exactly trying to explore the computational and statistical methodology in the G-framework, especially starting from the G-normal distribution (a typical one with variance uncertainty).

The computation of the G-expectation based on a given parameter setting has been developed for several years like the numerical schemes by Dolinsky (2012) to approximate the G-expectation. In fact, due to the nonlinearity of the expectation and the uncertainty intrinsically included in the distributions, the statistical theory in the G-framework is not easy or trivial to develop. So far there are already some attempts like the Max-mean estimation by Jin and Peng (2016) aimed at estimating the parameters or the sublinear expectations from dataset but it still requires the context of real data to offer more information about how to do the grouping and decide the test function. (In this thesis, sometimes we stress "real" data to distinguish from the artificial data generated by ourselves.)

To summarize, Table 1.4 shows the existing objects and results in the G-expectation (or **sublinear** expectation) framework compared with the counterparts in classical probability framework (with **linear** expectation), where the concepts in G-frame will be further explained in Section 2.1.

Table 1.4: Summary of G-frame compared with Classical frame

	Linear	Sublinear
Space	$(\Omega, \mathcal{F}, P)$ with linear E	$(\Omega, \mathcal{H}, \hat{\mathbb{E}})$ with sublinear $\hat{\mathbb{E}}$
$X \overset{d}{=} Y$	$\forall \varphi, E[\varphi(X)] = E[\varphi(Y)]$	$\forall \varphi, \hat{\mathbb{E}}[\varphi(X)] = \hat{\mathbb{E}}[\varphi(Y)]$
Indep.	Symmetric	Asymmetric
LLN	$\frac{1}{n}\sum X_i \overset{d}{\to} \mu$	$\frac{1}{n}\sum X_i \overset{d}{\to} M[\underline{\mu}, \overline{\mu}]$
CLT	$\frac{1}{\sqrt{n}}\sum X_i \overset{d}{\to} N(0, \sigma^2)$	$\frac{1}{\sqrt{n}}\sum X_i \overset{d}{\to} N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$
BM	$B_t \overset{d}{=} N(0, \sigma^2 t)$	$B_t \overset{d}{=} N(0, [\underline{\sigma}^2 t, \overline{\sigma}^2 t])$
	$d\langle B\rangle_t = dt$	$d\langle B\rangle_t \neq dt$
Itô	Itô formula	G-Itô formula
Est.	$E[X] \approx \frac{1}{N}\sum_{i=1}^{N} X_i$	$\hat{\mathbb{E}}[X] \approx \max_{1\le i\le m} \sum_{j=1}^{m} X_{ij}$
		$-\hat{\mathbb{E}}[-X] \approx \min_{1\le i\le m} \sum_{j=1}^{m} X_{ij}$

Let us start our exploration and adventure beginning with the G-normal distribution $N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$, one of the basic objects in the G-framework, based on some existing theoretical work by the pioneers, to design and construct a new basic substructure, the semi-G-normal distribution, in order to provide statistical tools for the whole framework and also the larger community to better understand and study the variance uncertainty. Hope you will see some interesting, inspiring and delicate designs (such as the semi-G-normal distribution with its nice properties and extension) and enjoy the scenery during this journey.

1.2 Overview of my research

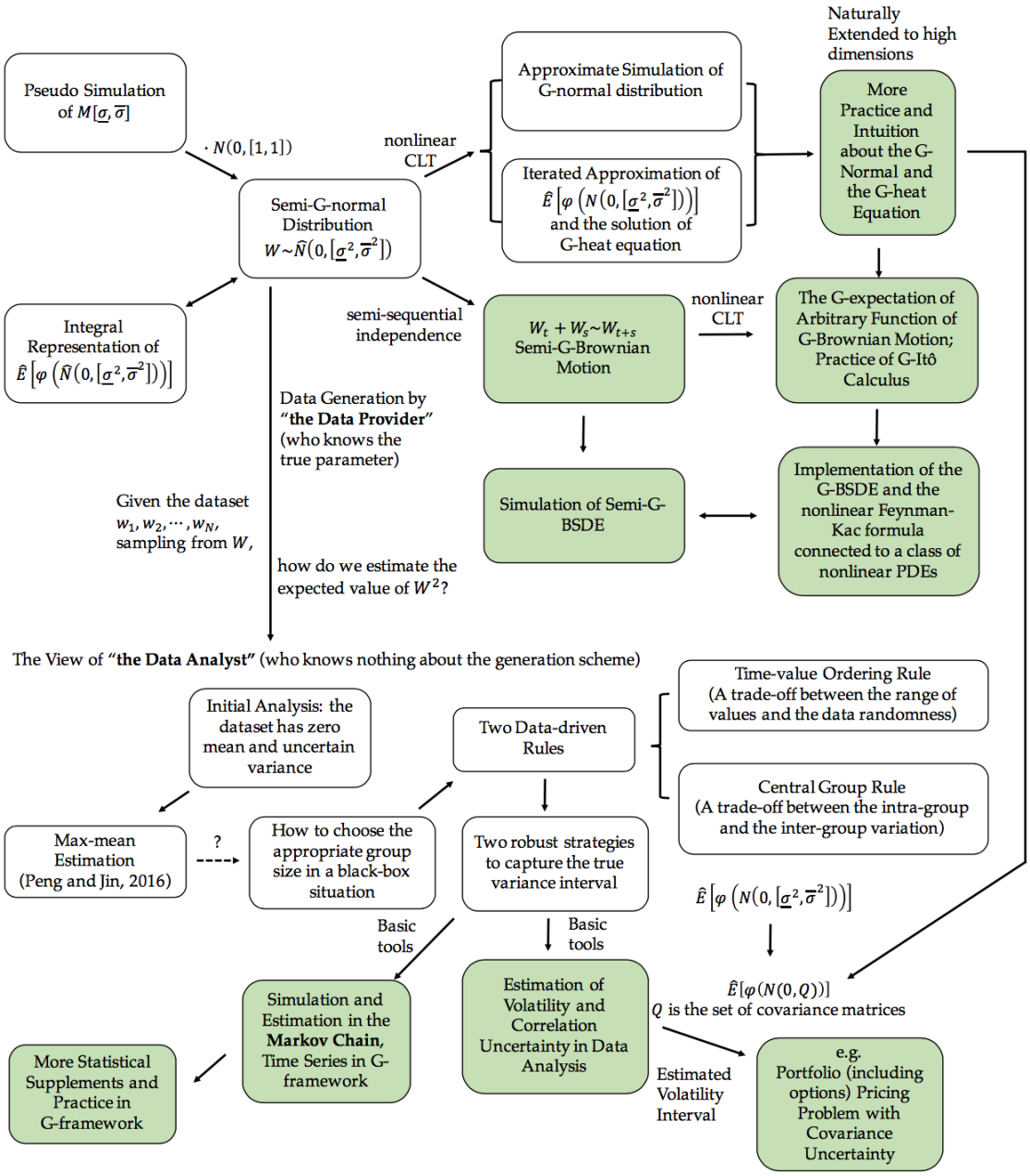


Figure 1.2: My Current Research and Future Development

The main content of this thesis, aimed at exploring the computation, simulation and estimation of G-normal distribution, consists of four chapters:

- Chapter 2: the distributions and independence in the G-framework;
- Chapter 3: the estimation of variance uncertainty;
- Chapter 4: the pseudo simulation of variance uncertainty;
- Chapter 5: concluding remarks and future development.

Although the G-framework has strong potential to theoretically deal with ambiguity, because of the sub-additivity of G-expectation $\hat{\mathbb{E}}[\cdot]$, it is hard to develop the statistical or computational version of G-framework, which is the bridge to real data analyses and industrial problems. This is especially true for the G-normal distribution $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$, as its G-expectation $\hat{\mathbb{E}}[\varphi(X)]$ does not have an explicit expression if φ is neither convex nor concave. In order to compute the $\hat{\mathbb{E}}[X^{2n+1}]$, previously we needed to use special PDE and ODE techniques to solve the G-heat equation (Hu (2012)).

Chapter 2 mainly investigates the distributions and independence in the G-framework and establishes a substructure (Section 2.3) based on a new concept called the *semi-G-normal distribution* like a transition from classical normal to G-normal distribution to fill in the thinking gap between these two objects to get better intuition and do the computation and simulation of G-normal distribution. It actually gives a probabilistic method (based on the G-framework) to compute the G-expectation of G-normal distribution. This is a great extension allowing us to deal with the G-expectation of a larger class of functions of G-normal distribution. Interestingly, this substructure gives an efficient iterative scheme to stochastically solve, by Monte-Carlo method, the G-heat equation (which is fully nonlinear PDE, also known as the *Black-Scholes-Barenblatt equation with volatility uncertainty*). Section 2.3 has also been written as a preprint paper (Li and Kulperger (2018), which is mainly done by myself under the supervision of Prof. Kulperger) we intend to publish in the near future. We also explore the independence (Section 2.4) regarding G-normal and semi-G-normal distribution, showing that one annoying property of G-normal distribution is that it not easy to construct *multivariate* G-normal distribution from *univariate* objects. Can we find a path to construct the multivariate G-normal distribution from univariate objects? We will give a positive answer by showing how to start from the univariate semi-G-normal objects, with its special design of independence, to approach the multivariate G-normal distribution.

Then we come to the side of dataset including estimation (Chapter 3) and simulation (Chapter 4) which require and apply both the theory and intuition we have learned in Chapter 2. One important question regarding the real data analysis with variance uncertainty is, if treated as G-normal distribution, how to estimate the variance interval. Although the established Max-mean method by Jin and Peng (2016) gives a theory of this kind of estimation, it still relies on the information from the practical background to decide the group size and test function.

Chapter 3 provides two heuristic data-driven rules (Section 3.2) for Max-mean estimation to appropriately select the group size so as to robustly capture the true variance interval, which turns out to numerically work well. We will further work on more practice of this improved Max-mean estimation and more theory to support it. Meanwhile, we also try to put the estimation method into practice so as to get more designing ideas from the application background.

Chapter 4 designs a simulation procedure centering around the semi-G-normal distribution, and starting from the pseudo simulation of maximal distribution to the approximate simulation

of the G-normal distribution. The simulation is essentially important for the numerical testing the estimation methods in Chapter 3.

Finally, Chapter 5 summarizes the whole thesis and discuss about the future development especially about the future industrial practice of the G-framework.

Throughout the thesis, we have many statements or sentences labelled as a "Remark" and "Comment". For readers' information, a "Remark" refers to the statements including some technical details explained by mathematical expressions (especially displayed ones) while a "Comment" refers to those sentences only consisting of words to give a general description or explanation.

Chapter 2

The Distributions and Independence in the G-framework

2.1 Preliminaries

The G-expectation or sublinear expectation framework (also called the G-framework), motivated by the problems with *ambiguity or uncertainty*, is a generalization of the linear probability framework. Similar to the Choquet expectation, the sublinear expectation can be represented as the supreme of a class of linear expectations. Intuitively, the linear expectation mainly considers the "average", while the sublinear expectation focuses on the "bound" to create a interval to cover the uncertainty which is hard to be described by a certain distribution. Please turn to Peng (2004, 2007a, 2008, 2010) for more details.

Definition 2.1.1. A *sublinear expectaion space* is defined as a triple $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$. Ω is a given set (also known as a sample space). $\mathcal{H}$ is a linear space of real valued functions defined on Ω satisfying $c \in \mathcal{H}$ for each constant c and $|X| \in \mathcal{H}$ if $X \in \mathcal{H}$, which can be regarded as the space of random variables. $\hat{\mathbb{E}}$ is a *sublinear expectation* which is a functional $\hat{\mathbb{E}} : \mathcal{H} \to \mathbb{R}^d$ satisfying:

1. Monotonicity: $\hat{\mathbb{E}}[X] \geq \hat{\mathbb{E}}[Y]$ if $X \geq Y$;
2. Constant preserving: $\hat{\mathbb{E}}[c] = c$ for $c \in \mathbb{R}^d$;
3. Sub-additivity: For each $X, Y \in \mathcal{H}$, $\hat{\mathbb{E}}[X + Y] \leq \hat{\mathbb{E}}[X] + \hat{\mathbb{E}}[Y]$;
4. Positive homogeneity: $\hat{\mathbb{E}}[\lambda X] = \lambda \hat{\mathbb{E}}[X]$ for $\lambda \geq 0$.

If only monotonicity and constant preserving are satisfied, $\hat{\mathbb{E}}$ is called a nonlinear expectation and $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$ is called a nonlinear expectation space.

In this thesis, since we will not deal with nonlinear expectation (without sub-additivity and positive homogeneity), we will not strictly distinguish between "sublinear" and "nonlinear." Therefore, in some casual context of the description of the G-framework, we may use "nonlinear" and"sublinear" interchangably, as they are both distinguished from the word "linear", which is the difference we mostly care about.

In fact, similar to Choquet expectation (for readers familiar with its setting), the sublinear expectation defined in Definition 2.1.1 can also be expressed as the supreme of a class of linear

expectations corresponding to a family of probability measures (Theorem 2.1.2). Choquet expectation is a special case of the sublinear expectation.

Theorem 2.1.2 (Representation of sublinear expectation)**.** *Let $\hat{\mathbb{E}}$ denote a sublinear expectation on $\mathcal{H}$. Then there exists a family of probability measures $\mathcal{Q}$ which correspondingly induces a collection of linear expectation $\{E_P,\ P \in \mathcal{Q}\}$ on $\mathcal{H}$ such that for each $X \in \mathcal{H}$,*

$$\hat{\mathbb{E}}[X] = \sup_{P \in \mathcal{Q}} E_P[X].$$

Remark 2.1.2.1. Meanwhile, for each $X \in \mathcal{H}$, there exists $\theta_X \in \mathcal{Q}$ satisfying $\hat{\mathbb{E}}[X] = E_{\theta_X}[X]$.

In the following context, we often use capital letters like $X := (X_1, X_2, \ldots, X_d), d \in \mathbb{N}_+$ to denote the random variables (or vectors) in $\mathcal{H}$. Meanwhile, if $X \in \mathcal{H}$, we also have $\varphi(X) \in \mathcal{H}$ for every φ in $C_{l.Lip}(\mathbb{R}^d)$ which is the linear space of functions satisfying the locally Lipchistz property:

$$|\varphi(x) - \varphi(y)| \le C_\varphi |1 + |x|^k + |y|^k||x - y|,$$

for $x, y \in \mathbb{R}^d$, some $k \in \mathbb{N}$ and $C_\varphi > 0$ depending on φ. If not specified, we will always stay in the *sublinear expectation space* $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$ and the function space $C_{l.Lip}(\mathbb{R}^d)$ (or $C_{l.Lip}$ in short, which can be replaced by other spaces). Our computation in this space is usually different from the linear expectation E mainly because of the sub-additivity and positive homogeneity of $\hat{\mathbb{E}}$. Here are some useful tools to understand and deal with $\hat{\mathbb{E}}$.

In general, for any $X \in \mathcal{H}$, we must have $-\hat{\mathbb{E}}[-X] \le \hat{\mathbb{E}}[X]$ because $0 = \hat{\mathbb{E}}[X + (-X)] \le \hat{\mathbb{E}}[X] + \hat{\mathbb{E}}[-X]$. Whether it is a strict inequality tells us whether X has "uncertainty" or not, which is better illustrated by Definition 2.1.3.

Definition 2.1.3 (Moments-uncertainty)**.** For each $X \in \mathcal{H}$, we say X has *the k-th moment-uncertainty* if $-\hat{\mathbb{E}}[-X^k] < \hat{\mathbb{E}}[X^k] < \infty$ for $k = 1, 2, \ldots, n$. X has *the k-th moment-certainty* if $-\hat{\mathbb{E}}[-X^k] = \hat{\mathbb{E}}[X^k] < \infty$. Specifically, we have

1. X has *mean-uncertainty* (or mean-certainty, respectively) if it has the 1-st moment-uncertainty (or 1-st moment-certainty, respectively);
2. When X satisfies $0 = -\hat{\mathbb{E}}[-X] = \hat{\mathbb{E}}[X]$, X has the *variance-uncertainty* (or variance-certainty, respectively) if it has the 2-nd moment-uncertainty (or 2-nd moment-certainty, respectively).

Proposition 2.1.4. *If X has the mean-uncertainty: $\underline{\mu} < \overline{\mu}$ with $\underline{\mu} := -\hat{\mathbb{E}}[-X]$ and $\overline{\mu} := \hat{\mathbb{E}}[X]$, for $\lambda \in \mathbb{R}$, we will have*

$$\hat{\mathbb{E}}[\lambda X] = \begin{cases} \lambda \hat{\mathbb{E}}[X] & \lambda \ge 0 \\ -\lambda \hat{\mathbb{E}}[-X] & \lambda < 0 \end{cases} = \lambda^+ \overline{\mu} - \lambda^- \underline{\mu},$$

where $\lambda^+ := \max\{\lambda, 0\}$ and $\lambda^- := \max\{-\lambda, 0\}$.

Proposition 2.1.5. *If X has the mean-certainty: $\underline{\mu} = \overline{\mu} =: \mu$ namely $-\hat{\mathbb{E}}[-X] = \hat{\mathbb{E}}[X] = \mu$, for $\lambda \in \mathbb{R}$, we will have*

$$\hat{\mathbb{E}}[\lambda X] = \lambda \hat{\mathbb{E}}[X] (= \lambda \mu),$$

and furthermore,

$$\hat{\mathbb{E}}[Y + \lambda X] = \hat{\mathbb{E}}[Y] + \lambda \hat{\mathbb{E}}[X].$$

Proof. The first one is directly from the mean-certain-1. For the second one, firstly we have

$$\hat{\mathbb{E}}[Y + \lambda X] \leq \hat{\mathbb{E}}[Y] + \hat{\mathbb{E}}[\lambda X] = \hat{\mathbb{E}}[Y] + \lambda\hat{\mathbb{E}}[X].$$

Secondly since $\hat{\mathbb{E}}[Y] = \hat{\mathbb{E}}[Y - X + X] \leq \hat{\mathbb{E}}[Y - X] + \hat{\mathbb{E}}[X]$ implies $\hat{\mathbb{E}}[Y - X] \geq \hat{\mathbb{E}}[Y] - \hat{\mathbb{E}}[X]$, we have

$$\hat{\mathbb{E}}[Y + \lambda X] = \hat{\mathbb{E}}[Y - \lambda(-X)] \geq \hat{\mathbb{E}}[Y] - \hat{\mathbb{E}}[\lambda(-X)] = \hat{\mathbb{E}}[Y] + \lambda\hat{\mathbb{E}}[X].$$

Combining the two inequalities, we get $\hat{\mathbb{E}}[Y + \lambda X] = \hat{\mathbb{E}}[Y] + \lambda\hat{\mathbb{E}}[X]$. □

In general, without any requirements like the mean-certainty of the random variables, we have the following probability inequalities (Proposition 2.1.6 and Proposition 2.1.7).

Proposition 2.1.6 (Hölder inequality)**.** *For $p, q > 0$, $\frac{1}{p} + \frac{1}{q} = 1$, we have*

$$\hat{\mathbb{E}}[|XY|] \leq (\hat{\mathbb{E}}[|X|^p])^{1/p}(\hat{\mathbb{E}}[|Y|^q])^{1/q}.$$

Proposition 2.1.7 (Minkowski inequality)**.** *For $p \geq 1$, we have*

$$(\hat{\mathbb{E}}[|X + Y|^p])^{1/p} \leq (\hat{\mathbb{E}}[|X|^p])^{1/p} + (\hat{\mathbb{E}}[|Y|^p])^{1/p}.$$

We will use $\hat{\mathbb{E}}$ to redefine distributions and independence.

Definition 2.1.8 (Distributions)**.** We give the notions of distribution, identical distribution and convergence in distribution as follow.

1. $\mathbb{F}_X$ is called the *distribution* of X, which is a functional: $\mathbb{F}_X[\varphi] := \hat{\mathbb{E}}[\varphi(X)] : \varphi \in C_{l.Lip}(\mathbb{R}^d) \to \mathbb{R}$.
2. X and Y are *identically distributed*, denoted by $X \overset{\mathrm{d}}{=} Y$, if for any $\varphi \in C_{l.Lip}$,

   $$\hat{\mathbb{E}}[\varphi(X)] = \hat{\mathbb{E}}[\varphi(Y)],$$

   namely, $\mathbb{F}_X[\varphi] = \mathbb{F}_Y[\varphi]$.
3. A sequence $\{X_n\}_n^\infty$ *converges in distribution* to X, denoted as $X_n \overset{\mathrm{d}}{\to} X$, if for any $\varphi \in C_{l.Lip}$,

   $$\lim_{n\to\infty} \hat{\mathbb{E}}[\varphi(X_n)] = \hat{\mathbb{E}}[\varphi(X)].$$

Definition 2.1.9 (Independence)**.** *Y is independent from X*, denoted by $X \dashrightarrow Y$, if for any $\varphi \in C_{l.Lip}$,

$$\hat{\mathbb{E}}[\varphi(X, Y)] = \hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi(x, Y)]_{x=X}].$$

Remark 2.1.9.1. For readers' convenience, the notation $\hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi(x, Y)]_{x=X}]$ means two steps of computation:

1. for any fixed x, compute $\hat{\mathbb{E}}[\varphi(x, Y)]$ which becomes a function of x denoted as $H(x)$;
2. replace x with X to compute $\hat{\mathbb{E}}[H(X)]$ where $H(X)$ is a transformation of X.

Comment 2.1.9.1. Intuitively, $X \dashrightarrow Y$ means that any realization of X will have no effect on Y's uncertainty set of distributions. $X \dashrightarrow Y$ does not mean $Y \dashrightarrow X$. In other words, this independence has its *order*. In the scenario of linear expectation, the notion of independence in Definition 2.1.9 becomes the classical one. The sequential independence is one of the most important notions in the G-framework, about which, more exploration can be found in Section 2.4.

Definition 2.1.10 (i.i.d.)**.** $\{X_i\}_{i=1}^{\infty}$ is *i.i.d.* if $X_{i+1} \overset{\mathrm{d}}{=} X_i$ and $(X_1, X_2, \ldots, X_i) \dashrightarrow X_{i+1}$ for each $i \in \mathbb{N}_+$.

Let $\bar{X}$ be an independent copy of X, which means $\bar{X} \overset{\mathrm{d}}{=} X$ and $X \dashrightarrow \bar{X}$.

Definition 2.1.11 (Maximal distribution)**.** X follows *Maximal Distribution* if, for any independent copy $\bar{X}$, we have

$$aX + b\bar{X} \overset{\mathrm{d}}{=} (a+b)X \ \forall a, b \geq 0,$$

which is equivalent to

$$X + \bar{X} \overset{\mathrm{d}}{=} 2X.$$

This is the sublinear version of a constant. A more specific definition with representation is given by Definition 2.1.12.

Definition 2.1.12 (Maximal distribution with representation)**.** X follows the maximal distribution $M(\Gamma)$ if there exists a bounded, closed and convex set $\Gamma \subset \mathbb{R}^d$ such that for any $\varphi \in C_{l.Lip}(\mathbb{R}^d)$,

$$\mathbb{F}_X[\varphi] = \hat{\mathbb{E}}[\varphi(X)] = \max_{v \in \Gamma} E_{\delta_v}[\varphi(X)] = \max_{v \in \Gamma} \varphi(v),$$

where δ_v is the Dirac measure with respect to $v \in \mathbb{R}^d$. For $d = 1$, we have $X \sim M[\underline{\mu}, \overline{\mu}]$ with *mean-uncertainty*: $\underline{\mu} := -\hat{\mathbb{E}}[-X]$ and $\overline{\mu} := \hat{\mathbb{E}}[X]$.

Definition 2.1.13 (G-normal distribution)**.** X follows *G-normal Distribution* if, for any independent copy $\bar{X}$, we have

$$aX + b\bar{X} \overset{\mathrm{d}}{=} \sqrt{a^2 + b^2}X, \ \forall a, b \geq 0,$$

which is equivalent to

$$X + \bar{X} \overset{\mathrm{d}}{=} \sqrt{2}X.$$

For $d = 1$, we have $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ $(0 \leq \underline{\sigma} \leq \overline{\sigma})$ with *variance-uncertainty*: $\underline{\sigma}^2 := -\hat{\mathbb{E}}[-X^2]$ and $\overline{\sigma}^2 := \hat{\mathbb{E}}[X^2]$.

The Proposition 2.1.14 is a good practice for general readers to work on the new operator $\hat{\mathbb{E}}$.

Proposition 2.1.14 (The scaling property of G-normal distribution)**.** *Let $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$, then for any given constant $c \in \mathbb{R}$, we have $cX \sim N(0, [c^2\underline{\sigma}^2, c^2\overline{\sigma}^2])$ which, by using the simplified notation $c[\underline{\sigma}, \overline{\sigma}] := [ca, cb]$, can be written as*

$$cX \sim N(0, c^2[\underline{\sigma}^2, \overline{\sigma}^2]).$$

Proof. Firstly, we need to show $Y \coloneqq cX$ is also G-normal distributed. For any independence copy $\bar{Y}$, it can be equivalently written as $c\bar{X}$ (readers can prove the two deductive directions), where $\bar{X}$ is also an independent copy of X. Then we have

$$Y + \bar{Y} = cX + c\bar{X} = c(X + \bar{X}) \overset{\mathrm{d}}{=} c\sqrt{2}X = \sqrt{2}Y.$$

Secondly, compute the two bounds of variance interval of Y:

$$\hat{\mathbb{E}}[Y^2] = \hat{\mathbb{E}}[(cX)^2] = \hat{\mathbb{E}}[c^2X^2] = c^2\hat{\mathbb{E}}[X^2] = c^2\overline{\sigma}^2,$$

and

$$-\hat{\mathbb{E}}[-Y^2] = -\hat{\mathbb{E}}[-(cX)^2] = -\hat{\mathbb{E}}[c^2(-X^2)] = -c^2\hat{\mathbb{E}}[-X^2] = c^2\underline{\sigma}^2.$$

Therefore, $Y = cX \sim N(0, [c^2\underline{\sigma}^2, c^2\overline{\sigma}^2])$. □

Let $\mathbb{S}_d$ denote the set of all real-valued $d \times d$ symmetric matrices and $\mathbb{S}_d^+ (\subset \mathbb{S}_d)$ represent the set of non-negative definite symmetric matrices.

Theorem 2.1.15 (G-normal distribution characterized by the G-heat Equation)**.** *X follows the d-dimensional G-normal distribution, iff $u(t,x) \coloneqq \hat{\mathbb{E}}[\varphi(x + \sqrt{1-t}X)]$ is the (unique viscosity) solution to* the G-heat Equation *defined on* $[0,\infty) \times \mathbb{R}^d$*:*

$$u_t + G(D_x^2 u) = 0,\ u|_{t=1} = \varphi,$$

where $G(\mathbf{A}) \coloneqq \frac{1}{2}\hat{\mathbb{E}}[\langle \mathbf{A}X, X\rangle] : \mathbb{S}_d \to \mathbb{R}$, which is a sublinear function characterizing the distribution of X. For $d = 1$, we have $G(a) = \frac{1}{2}(\overline{\sigma}^2 a^+ - \underline{\sigma}^2 a^-)$ and when $\underline{\sigma}^2 > 0$, this is also called the Black-Scholes-Barenblatt equation *with volatility uncertainty.*

Remark 2.1.15.1. We can use the function $G(\mathbf{A}) \coloneqq \frac{1}{2}\hat{\mathbb{E}}[\langle \mathbf{A}X, X\rangle]$ to characterize the definition of G-normal distribution. In fact, $G(\mathbf{A})$ can be further expressed as

$$G(\mathbf{A}) = \frac{1}{2}\sup_{\mathbf{V}\in\mathcal{V}} \mathrm{tr}[\mathbf{A}\mathbf{V}],$$

where $\mathcal{V} = \{\mathbf{B}\mathbf{B}^T : \mathbf{B} \in \mathbb{S}_d\}$ is a collection of non-negative definite symmetric matrices which can be treated as the uncertainty set of the covariance matrices.

Comment 2.1.15.1. In Theorem 2.1.15, we use the notion of *viscosity* solution as a replacement of the classical one when lacking smoothness because of the nonlinearity of the G-heat equation. The definitions and more details of viscosity solution can be found in Crandall et al. (1992). From Peng (2010), when G is non-degenerate, the viscosity solution becomes a classical $C^{1,2}$ solution. Since this thesis has not touched or stressed a lot on the differences between the viscosity and classical solutions yet, readers who are not familiar with viscosity solutions can simply treat this notion as the classical ones in the following context.

Remark 2.1.15.2 (A direct origin and application of the G-heat equation, Avellaneda and Paras (1996)). In fact, the form of the G-heat equation is not some brand-new design but was actually well-established in 1990s to consider the volatility uncertainty in many financial scenarios such as trading, risk management, option pricing, portfolio design and so on. The famous Uncertain

Volatility Model (UVM) is well-worth mentioning here. Consider a model valuing a contingent claim based on an underlying asset with volatility uncertainty. Suppose the spot price of the underlying asset follows a stochastic differential equation:

$$P : \frac{\mathrm{d}S_t}{S_t} = \mu_t \,\mathrm{d}t + \sigma_t \,\mathrm{d}Z_t,$$

with $\mu_t = r_t - d_t$ (where r_t is the spot domestic riskless rate and d_t is the dividend rate). The volatility process is assumed to fluctuate within a band

$$\underline{\sigma} \le \sigma_t \le \overline{\sigma}, 0 \le t \le T.$$

Consider an agent that must deliver a stream of cash-flows $F_1(S_{t_i}), i = 1, 2, \ldots, N$ where $F_j(\cdot)$ are payoffs due at settlement dates $t_1 < t_2 < \ldots < t_N$. The worst case scenario present value is

$$V(S_t, t) = \sup_{P \in Q} E_P \left\{ \sum_{j=1}^{N} \exp(-r(t_j - t)) F_j(S_{t_j}) \right\}.$$

Actually, $V(S, t)$ satisfies the nonlinear programming equation (similar to the G-heat equation):

$$V_t + \frac{1}{2}\sigma^2(V_{SS}) \cdot V_{SS} + \mu S V_S - rV = 0,$$

where

$$\sigma^2(V_{SS}) = \begin{cases} \overline{\sigma}^2 & \text{if } V_{SS} \ge 0 \\ \underline{\sigma}^2 & \text{if } V_{SS} < 0 \end{cases}.$$

Another equivalent form is easier for us to treat it as a PDE in control theory (HJB equation):

$$V_t + \sup_{\sigma^2 \in [\underline{\sigma}^2, \overline{\sigma}^2]} (\frac{1}{2}\sigma^2 V_{SS}) + \mu S V_S - rV = 0.$$

Definition 2.1.16 (G-normal distribution with characterization)**.** Let X be any d-dimensional G-normal distributed random vector. To be specific, we say $X \sim N(\mathbf{0}, \mathcal{V})$ with the set of covariance matrices $\mathcal{V}$ if its distribution is characterized by the G-heat equation with

$$G(\mathbf{A}) := \frac{1}{2}\hat{\mathbb{E}}[\langle \mathbf{A}X, X\rangle] = \frac{1}{2}\sup_{\mathbf{V} \in \mathcal{V}} \mathrm{tr}[\mathbf{A}\mathbf{V}].$$

In other words, $\mathcal{V}$ is the set corresponding to the G function characterizing the distribution of X. In order to show the *covariance-uncertainty* of $N(\mathbf{0}, \mathcal{V})$, we can expand the details of $\mathcal{V}$ as

$$\mathcal{V} := \Bigg\{ \mathbf{V} = \left(\rho_{ij}\sigma_i\sigma_j\right)_{d\times d} : \sigma_i^2 \in [\underline{\sigma}_i^2, \overline{\sigma}_i^2],$$

$$\rho_{ij} = \rho_{ji} = \begin{cases} 1 & i = j \\ \in [\underline{\rho}_{ij}, \overline{\rho}_{ij}] & i \ne j \end{cases}, \text{ such that } \mathbf{V} \in \mathbb{S}_d^+ \Bigg\}.$$

When $d = 1$, we say $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ with the variance interval $[\underline{\sigma}^2, \overline{\sigma}^2]$ if its distribution is characterized by the G-heat equation with

$$G(a) := \frac{1}{2}\hat{\mathbb{E}}[aX^2] = \frac{1}{2}(\overline{\sigma}^2 a^+ - \underline{\sigma}^2 a^-).$$

Based on the G-version of "constant" and "normal" distribution, we have the respective Law of Large Numbers (LLN) and Central Limit Theorem (CLT) in the G-framework.

Theorem 2.1.17 (Law of Large Numbers, Peng (2010))**.** *Consider a sequence of i.i.d.* $\{Z_i\}_{i=1}^{\infty}$ *with* mean-uncertainty *characterized by a (unique) bounded, closed and convex set* $\Gamma \subset \mathbb{R}^d$ *in the sense that*

$$\max_{v\in\Gamma}\langle p, v\rangle = \hat{\mathbb{E}}[\langle p, Z_1\rangle], p \in \mathbb{R}^d.$$

Then for any $\varphi \in C_{l.Lip}$*,*

$$\lim_{n\to\infty} \hat{\mathbb{E}}[\varphi(\frac{1}{n}\sum_{i=1}^{n} Z_i)] = \hat{\mathbb{E}}[\varphi(Z)],$$

where Z *is a maximal distributed random variable characterized by the set* Γ*:*

$$\hat{\mathbb{E}}[\varphi(Z)] = \max_{v\in\Gamma} \varphi(v).$$

For $d = 1$*,* Γ *becomes a closed interval* $[\underline{\mu}, \overline{\mu}]$ *with* $\underline{\mu} := -\hat{\mathbb{E}}[-Z_1]$ *and* $\overline{\mu} := \hat{\mathbb{E}}[Z_1]$*. Then we have* $\frac{1}{n}\sum_{i=1}^{n} Z_i$*) converges in distribution to* $Z \sim M[\underline{\mu}, \overline{\mu}]$*.*

Theorem 2.1.18 (Central Limit Theorem, Peng (2010))**.** *Consider a sequence of i.i.d.* $\{X_i\}_{i=1}^{\infty}$ *with mean-certainty* $\hat{\mathbb{E}}[X_1] = -\hat{\mathbb{E}}[-X_1] = \mathbf{0}$*. Let* X *be a* G*-normal distributed random variable characterized by the function* $G(\mathbf{A}) := \frac{1}{2}\hat{\mathbb{E}}[\langle \mathbf{A}X_1, X_1\rangle]$*. Then for any* $\varphi \in C_{l.Lip}$*,*

$$\lim_{n\to\infty} \hat{\mathbb{E}}[\varphi(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i)] = \hat{\mathbb{E}}[\varphi(X)].$$

For $d = 1$*, let* $\underline{\sigma}^2 := -\hat{\mathbb{E}}[-X_1^2]$ *and* $\overline{\sigma}^2 := \hat{\mathbb{E}}[X_1^2]$*. Then we have* $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$ *converges in distribution to* $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ *.*

We will call Theorem 2.1.18 the nonlinear CLT. We also have the convergence rate of nonlinear CLT.

Theorem 2.1.19 (The convergence rate of nonlinear CLT by Song (2017))**.** *Under the setting of Theorem 2.1.18 when* $d = 1$*, for bounded and Lipschitz continuous* φ *(i.e. for any* $x, y \in \mathbb{R}$*,* $|\varphi(x)| \le M$ *and* $|\varphi(x) - \varphi(y)| \le C_\varphi|x - y|$*), there exist* $\alpha \in (0, 1)$ *depending on* $\underline{\sigma}$ *and* $\overline{\sigma}$*, and* $C_{\alpha,G} > 0$ *depending on* α*,* $\underline{\sigma}$ *and* $\overline{\sigma}$ *such that*

$$\sup_{C_\varphi\le 1}\left|\hat{\mathbb{E}}[\varphi(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i)] - \hat{\mathbb{E}}[\varphi(X)]\right| \le C_{\alpha,G}\frac{\hat{\mathbb{E}}[|X_1|^{2+\alpha}]}{n^{\frac{\alpha}{2}}} = O(\frac{1}{n^{\frac{\alpha}{2}}}).$$

2.2 Motivation: How can we better understand the G-normal distribution?

For the academic community concerning the G-framework, there is a long existing thinking gap between classical normal and G-normal distribution.

For instance, according to the comparison theorem of parabolic PDEs given by Crandall et al. (1992), it is not hard to show

$$\hat{\mathbb{E}}[\varphi(N(0,[\underline{\sigma}^2,\overline{\sigma}^2]))] \geq \sup_{\sigma\in[\underline{\sigma},\overline{\sigma}]} E[\varphi(N(0,\sigma^2))],$$

which indicates that the uncertainty set of G-normal distribution is much larger that a class of linear normal distributions with $\sigma \in [\underline{\sigma},\overline{\sigma}]$ (so what else is in the G-normal set of distributions?). Especially, Hu (2012) shows the strict inequality that when $\varphi(x) = x^3$,

$$\hat{\mathbb{E}}[(N(0,[\underline{\sigma}^2,\overline{\sigma}^2]))^3] > 0.$$

According to Proposition 2.1.14, let $X \overset{\mathrm{d}}{=} N(0,[\underline{\sigma}^2,\overline{\sigma}^2])$ then we have $-X \overset{\mathrm{d}}{=} N(0,[\underline{\sigma}^2,\overline{\sigma}^2])$, namely,

$$X \overset{\mathrm{d}}{=} -X, \tag{2.1}$$

which indicates that the G-normal distribution should have some "symmetry". However, exactly due to this identity in distribution shown in Equation (2.1), from Definition 2.1.8, we have

$$\hat{\mathbb{E}}[-X^3] = \hat{\mathbb{E}}[(-X)^3)] = \hat{\mathbb{E}}[X^3](>0),$$

which directly implies, (letting E and $N(0,\sigma^2)$ respectively denote the classical expectation and normal distribution,)

$$\hat{\mathbb{E}}[(N(0,[\underline{\sigma}^2,\overline{\sigma}^2]))^3] > 0 = E[(N(0,\sigma^2))^3] > -\hat{\mathbb{E}}[-(N(0,[\underline{\sigma}^2,\overline{\sigma}^2]))^3].$$

telling us that the "skewness" of G-normal distribution becomes uncertain so its "symmetry" is uncertain, which somehow looks like a "contradiction" with Equation (2.1) and is quite counter-intuitive for a "normal" distribution.

Is it possible to understand the G-normal distribution from our familiar classical normal distribution? In other words, is it possible to use the linear expectation of linear normal distribution (or the heat equation) to approach the sublinear expectation of G-normal distribution (or the G-heat equation)? This section gives positive answers to both of these questions by providing a ladder from the ground of $N(0,\sigma^2)$ to approach the peak of $N(0,[\underline{\sigma}^2,\overline{\sigma}^2])$: the *Semi-G-normal* distribution $\hat{N}(0,[\underline{\sigma}^2,\overline{\sigma}^2])$, which is a classical normal distribution scaled by a sublinear "constant" (the maximal distribution).

2.3 The semi-G-normal distribution and G-normal distribution

2.3.1 The 1-dimensional situation

Definition 2.3.1 (The Semi-G-normal Distribution)**.** W follows the *Semi-G-normal distribution*, denoted as, $W \sim \hat{N}(0,[\underline{\sigma}^2,\overline{\sigma}^2])$, if there exist $Y \sim N(0,[1,1])$ and $Z \sim M[\underline{\sigma},\overline{\sigma}]$, $\overline{\sigma} \geq \underline{\sigma} \geq 0$ with independent relation $Z \dashrightarrow Y$, such that

$$W = Z \cdot Y,$$

where "$\cdot$" is the number multiplication (which can be omitted) and the direction of independence here cannot be reversed.

Remark 2.3.1.1. $Y \sim N(0,[1,1])$ can be regarded as the classical standard normal distribution $N(0,1)$ since the corresponding G-heat equation will reduce to the classical heat equation when $\underline{\sigma}$ and $\overline{\sigma}$ coincide.

Remark 2.3.1.2 (the mean and variance of W). It is not hard to show that it has a certain zero mean:

$$\hat{\mathbb{E}}[W] = \hat{\mathbb{E}}[ZY] = \hat{\mathbb{E}}[\hat{\mathbb{E}}[zY]_{z=Z}] = \hat{\mathbb{E}}[E[zY]_{z=Z}] = 0,$$

and $-\hat{\mathbb{E}}[-W] = 0$. For the variance, we have

$$\hat{\mathbb{E}}[W^2] = \hat{\mathbb{E}}[Z^2Y^2] = \hat{\mathbb{E}}[\hat{\mathbb{E}}[z^2Y^2]_{z=Z}] = \hat{\mathbb{E}}[E[z^2Y^2]_{z=Z}] = \hat{\mathbb{E}}[(z^2\cdot 1)_{z=Z}] = \max_{z\in[\underline{\sigma},\overline{\sigma}]} z^2 = \overline{\sigma}^2,$$

and similarly,

$$-\hat{\mathbb{E}}[-W^2] = \hat{\mathbb{E}}[-Z^2Y^2] = -\hat{\mathbb{E}}[E[-z^2Y^2]_{z=Z}] = -\max_{z\in[\underline{\sigma},\overline{\sigma}]}(-z^2) = \min_{z\in[\underline{\sigma},\overline{\sigma}]} z^2 = \underline{\sigma}^2.$$

Theorem 2.3.2 (The Integral Representation of the Semi-G-normal Distribution)**.** *Let* $W \sim \hat{N}(0,[\underline{\sigma}^2,\overline{\sigma}^2])$ *then for any* $\varphi \in C_{l.Lip}(\mathbb{R})$*, we have*

$$\hat{\mathbb{E}}[\varphi(W)] = \max_{z\in[\underline{\sigma},\overline{\sigma}]} E[\varphi(N(0,z^2))] = \max_{z\in[\underline{\sigma},\overline{\sigma}]} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\varphi(zy)\,\mathrm{d}y.$$

Proof. This is quite straightforward because:

$$\hat{\mathbb{E}}[\varphi(W)] = \hat{\mathbb{E}}[\varphi(YZ)] = \hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi(Yz)]_{z=Z}] =: \hat{\mathbb{E}}[G(Z)],$$

where $G(z) := \hat{\mathbb{E}}[\varphi(Yz)] = E[\varphi(N(0,z^2))]$ can be proved to be in $C_{l.Lip}$ based on $\varphi \in C_{l.Lip}$. Specifically, we have

$$\begin{aligned}
|G(x)-G(y)| &= |E[\varphi(x\cdot Z)-\varphi(y\cdot Z)]| \\
&\le E[C_\varphi(1+|xZ|^k+|yZ|^k)|Z|\cdot|x-y|] \\
&= C_\varphi\cdot E[|Z|+|Z|^{k+1}|x|^k+|Z|^{k+1}|y|^k]|x-y| \\
&\le C_G(1+|x|^k+|y|^k)|x-y|,
\end{aligned}$$

where $C_G = C_\varphi \max\{E[|Z|], E[|Z|^{k+1}]\}$. Therefore,

$$\begin{aligned}
\hat{\mathbb{E}}[\varphi(W)] = \hat{\mathbb{E}}[G(Z)] &= \max_{z\in[\underline{\sigma},\overline{\sigma}]} G(z) \\
&= \max_{z\in[\underline{\sigma},\overline{\sigma}]} E[\varphi(N(0,z^2))] \\
&= \max_{z\in[\underline{\sigma},\overline{\sigma}]} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\varphi(zy)\,dy.
\end{aligned}$$

□

Remark 2.3.2.1 (Why is it called a "semi" one?). The comparison theorem of parabolic PDEs (in Crandall et al. (1992)) tells us that

$$\hat{\mathbb{E}}[\varphi(N(0,[\underline{\sigma}^2,\overline{\sigma}^2]))] \ge \max_{v\in[\underline{\sigma},\overline{\sigma}]} E[\varphi(N(0,v^2))] = \hat{\mathbb{E}}[\varphi(\hat{N}(0,[\underline{\sigma}^2,\overline{\sigma}^2]))],$$

whose inequality is mostly strict (like $\varphi(x) = x^3$) and becomes equal when φ is convex or concave. From the representation theorem of $\hat{\mathbb{E}}$: $\hat{\mathbb{E}}[\varphi(X)] = \sup_{P\in\mathcal{Q}} E_P[\varphi(X)]$, we have the intuition that $\mathcal{Q}_{\text{Semi-}G\text{-normal}} \subset \mathcal{Q}_{G\text{-normal}}$ where $\mathcal{Q}_{\text{Semi-}G\text{-normal}}$ consists of measures corresponding to classical normal distributions with $\sigma \in [\underline{\sigma},\overline{\sigma}]$.

Remark 2.3.2.2. Intuitively, the G-normal distribution is more "uncertain" than the semi-G-normal distribution. Explicitly speaking, we already have the following representation for the G-normal distributed $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ (Denis et al. (2011) and Li (2015)) which is consistent with the spirit of the uncertain volatility model mentioned in Remark 2.1.15.2:

$$\hat{\mathbb{E}}[\varphi(X)] = \sup_{\theta \in A^\Theta} E_P[\varphi(\int_0^1 \theta_s dB_s)],$$

where $\Theta = [\underline{\sigma}, \overline{\sigma}]$, $A^\Theta := \{\theta : \theta_t \in \Theta, \text{for } t \in [0,1]\}$, the set of all processes valuing in $[\underline{\sigma}, \overline{\sigma}]$ in the time range $[0,1]$ and B is the classical Brownian motion in $(\Omega, \mathcal{F}, P)$. Meanwhile, from the integral representation (Theorem 2.3.2), we can show that for $W \in \hat{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$,

$$\hat{\mathbb{E}}[\varphi(W)] = \sup_{\theta \in A^\Theta} E_P[\varphi(\int_0^1 \bar{\theta} dB_s)],$$

where $\bar{\theta} = \int_0^1 \theta_s ds (\in [\underline{\sigma}, \overline{\sigma}])$, the average of the process θ over the time interval $[0,1]$. To summarize, the G-normal distribution has the uncertainty set consisting of *all processes* valuing in Θ, while the semi-G-normal distribution only has the set made up of *all constant processes* valuing in Θ.

Corollary 2.3.2.1 (the connection with the G-normal distribution)**.** *When φ is convex or concave and $\varphi \in C^2(\mathbb{R})$, for $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ and $W \sim \hat{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$, we have*

$$\hat{\mathbb{E}}[\varphi(X)] = \hat{\mathbb{E}}[\varphi(W)].$$

Proof. The proof consists of two parts: first show the sublinear expectation of G-normal distribution and then gives the expectation of semi-G-normal distribution to prove their coincidence.

The sublinear expectation of G-normal distribution: Under convexity (or concavity), the integral representation of G-normal distribution directly comes from the solution of the classical heat equation because $u(t,x)$ will be convex (or concave, respectively) to x then $u_{xx} \geq 0$ (or ≤ 0, respectively) (more details can be found in Peng (2010)), giving us

$$\hat{\mathbb{E}}[\varphi(X)] = \begin{cases} E[\varphi(N(0, \overline{\sigma}^2))] & \varphi \text{ is convex} \\ E[\varphi(N(0, \underline{\sigma}^2))] & \varphi \text{ is concave} \end{cases},$$

for which we give a scratch proof here to general readers with interests. According to Theorem 2.1.15, the sublinear expectation of G-normal distribution can be treated as the solution of a nonlinear PDE called the G-heat equation (or sometimes in one dimension called *the Black-Scholes-Barenblatt equation with volatility uncertainty*). Explicitly speaking, after a time transformation, for $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$, $u(t,x) := \hat{\mathbb{E}}[\varphi(x + \sqrt{t}X)]$ becomes the (unique viscosity) solution of the G-heat equation defined on $[0, \infty) \times \mathbb{R}^d$:

$$u_t - \frac{1}{2}(\overline{\sigma}^2 u_{xx}^+ - \underline{\sigma}^2 u_{xx}^-) = 0, u|_{t=0} = \varphi,$$

or equivalently,

$$u_t - \sup_{\sigma \in [\underline{\sigma}, \overline{\sigma}]} \{\frac{1}{2}\sigma^2 u_{xx}\} = 0, u|_{t=0} = \varphi. \tag{2.2}$$

When φ is convex, (that is, for any $x, y \in \mathbb{R}$ and $a \in [0,1]$, $\varphi(ax + (1-a)y) \leq a\varphi(x) + (1-a)\varphi(y)$), we can check that $u(t,x)$ is also convex with respect to x: for any $x, y \in \mathbb{R}$ and $a \in [0,1]$,

$$\begin{aligned}
u(t, ax + (1-a)y) &= \hat{\mathbb{E}}[\varphi(ax + (1-a)y + \sqrt{t}X] \\
&= \hat{\mathbb{E}}[[\varphi(ax + (1-a)y + \sqrt{t}c]_{c=X}] \\
&= \hat{\mathbb{E}}[[\varphi(a(x + \sqrt{t}c) + (1-a)(y + \sqrt{t}c)]_{c=X}] \\
&\leq \hat{\mathbb{E}}[[a\varphi(x + \sqrt{t}c) + (1-a)\varphi(y + \sqrt{t}c)]_{c=X}] \\
&= \hat{\mathbb{E}}[a\varphi(x + \sqrt{t}X) + (1-a)\varphi(y + \sqrt{t}X)] \\
&\leq a\hat{\mathbb{E}}[\varphi(x + \sqrt{t}X)] + (1-a)\hat{\mathbb{E}}[\varphi(y + \sqrt{t}X)] \\
&= au(t,x) + (1-a)u(t,y).
\end{aligned}$$

As mentioned in Comment 2.1.15.1, under some non-degenerate conditions, the viscosity solution $u(t,x)$ is $C^{1,2}$. Then we have $u_{xx} \geq 0$ so that $u_{xx}^- = 0$ and Equation (2.2) degenerates to a classical one:

$$u_t - \frac{1}{2}\overline{\sigma}^2 u_{xx} = 0, u|_{t=0} = \varphi,$$

which is a Cauchy problem of the *classical heat equation* with a nice explicit unique solution in terms of the Gaussian (or normal) kernel:

$$u(t,x) = \frac{1}{\sqrt{2\pi\overline{\sigma}^2 t}} \int_{\mathbb{R}} \varphi(y) \exp(-\frac{(y-x)^2}{2\overline{\sigma}^2 t}) \, \mathrm{d}y = E[\varphi(x + \sqrt{t}N(0, \overline{\sigma}^2))], \tag{2.3}$$

where E represents the linear expectation and $N(0, \overline{\sigma}^2)$ is the classical normal distribution. Hence, in terms of the connection between $u(t,x)$ and the G-normal distribution, we have

$$\hat{\mathbb{E}}[\varphi(X)] = u(1,0) = \frac{1}{\sqrt{2\pi\overline{\sigma}^2}} \int_{\mathbb{R}} \varphi(y) \exp(-\frac{(y-x)^2}{2\overline{\sigma}^2}) \, \mathrm{d}y = E[\varphi(N(0, \overline{\sigma}^2))].$$

Since we already have the result under convexity, namely, Equation (2.3) with variance taken as $\overline{\sigma}^2$, when φ is concave, symmetrically speaking, we can guess the solution may be the one with the same form taking the other extreme $\underline{\sigma}^2$:

$$\hat{u}(t,x) = \frac{1}{\sqrt{2\pi\underline{\sigma}^2 t}} \int_{\mathbb{R}} \varphi(y) \exp(-\frac{(y-x)^2}{2\underline{\sigma}^2 t}) \, \mathrm{d}y = E[\varphi(x + \sqrt{t}N(0, \underline{\sigma}^2))]. \tag{2.4}$$

We will prove Equation (2.4) is the (unique) solution of the G-heat Equation (2.2). In the convex case, $u(t,x)$ expressed by Equation (2.3) is the unique solution. Meanwhile, it is $C^{1,2}$ and convex with respect to x. Note that

$$-\hat{u}(t,x) = \frac{1}{\sqrt{2\pi\underline{\sigma}^2 t}} \int_{\mathbb{R}} (-\varphi(y)) \exp(-\frac{(y-x)^2}{2\underline{\sigma}^2 t}) \, \mathrm{d}y,$$

where $-\varphi$ is convex, then, borrowing the properties of the solution with the same form as Equation (2.3), $-\hat{u}(t,x)$ must also be a $C^{1,2}$ function convex with respect to x, that is, $\hat{u}(t,x)$

is a concave $C^{1,2}$ function. Hence, $\hat{u}_{xx} \leq 0$ or $\hat{u}^+_{xx} = 0$. When we plug $\hat{u}(t,x)$ into the G-heat Equation (2.2), for the left hand side, we have

$$\hat{u}_t - \sup_{\sigma\in[\underline{\sigma},\overline{\sigma}]}\{\frac{1}{2}\sigma^2\hat{u}_{xx}\} = \hat{u}_t - \frac{1}{2}\underline{\sigma}^2\hat{u}_{xx},$$

which belongs to a form of classical heat equation. Meanwhile, we know for sure $\hat{u}(t,x)$ shown in Equation (2.4) is the exact solution of the classical heat equation with a Cauchy condition:

$$v_t - \frac{1}{2}\underline{\sigma}^2 v_{xx} = 0, v|_{t=0} = \varphi.$$

Therefore, $\hat{u}(t,x)$ must solve the G-heat Equation (2.2):

$$\hat{u}_t - \sup_{\sigma\in[\underline{\sigma},\overline{\sigma}]}\{\frac{1}{2}\sigma^2\hat{u}_{xx}\}(= \hat{u}_t - \frac{1}{2}\underline{\sigma}^2\hat{u}_{xx}) = 0, \hat{u}|_{t=0} = \varphi.$$

Actually, due to the uniqueness of the solution of G-heat equation (Crandall et al. (1992)), $\hat{u}(t,x)$ becomes the unique solution of the G-heat Equation (2.2). Thus we have

$$\hat{\mathbb{E}}[\varphi(X)] = \hat{u}(1,0) = \frac{1}{\sqrt{2\pi\underline{\sigma}^2}}\int_{\mathbb{R}}\varphi(y)\exp(-\frac{(y-x)^2}{2\underline{\sigma}^2})\,\mathrm{d}y = E[\varphi(N(0,\underline{\sigma}^2))].$$

The sublinear expectation of semi-G-normal distribution: For the semi-G-normal distribution, by using its representation with $G(z) \coloneqq E[\varphi(zY)](z \in [\underline{\sigma},\overline{\sigma}])$ and $Y \sim N(0,1)$, we only need to prove that

$$\hat{\mathbb{E}}[\varphi(W)] = \max_{z\in[\underline{\sigma},\overline{\sigma}]} G(z) = \begin{cases} G(\overline{\sigma}) & \varphi \text{ is convex} \\ G(\underline{\sigma}) & \varphi \text{ is concave} \end{cases}.$$

First of all, φ has the Taylor expansion

$$\varphi(x) = \varphi(0) + \varphi^{(1)}(0)x + \varphi^{(2)}(\xi_x)\frac{x^2}{2},$$

where $\xi_x \in (0,x)$.

1. When φ is convex, $\varphi^{(2)}(\xi_x) \geq 0$. The Taylor expansion tells us that:

$$\begin{aligned} G(z) &= E[\varphi(zY)] \\ &= E[\varphi(0) + \varphi^{(1)}(0)zY + \varphi^{(2)}(\xi_{zY})\frac{z^2}{2}Y^2] \\ &= \varphi(0) + \frac{1}{2}E[\varphi^{(2)}(\xi_{zY})(zY)^2], \end{aligned}$$

where $\xi_{zY} \in (0,zY)$ is a random variable depending on Y. Let $M \coloneqq zY \sim N(0,z^2)$, then

$$K(z) \coloneqq E[\varphi^{(2)}(\xi_{zY})(zY)^2] = E[\varphi^{(2)}(\xi_M)M^2] = \int \frac{1}{\sqrt{2\pi}}\exp(-\frac{m^2}{2z^2})\varphi^{(2)}(\xi_m)m^2\,\mathrm{d}m.$$

In order to consider the monotone property of $K(z)$, work on its derivative:

$$\begin{aligned} K'(z) &= \frac{d}{dz}\int \frac{1}{\sqrt{2\pi}}\exp(-\frac{m^2}{2z^2})\varphi^{(2)}(\xi_m)m^2\,\mathrm{d}m \\ &= \int \frac{1}{\sqrt{2\pi}}\left[\frac{d}{dz}\exp(-\frac{m^2}{2z^2})\right]\varphi^{(2)}(\xi_m)m^2\,\mathrm{d}m \\ &= \int \frac{1}{\sqrt{2\pi}}\underbrace{\left[\frac{m^2}{z^3}\exp\left(-\frac{m^2}{2z^2}\right)\right]}_{\geq 0 \text{ for } z\in[\underline{\sigma},\overline{\sigma}]}\underbrace{\varphi^{(2)}(\xi_m)\,m^2}_{\geq 0}\,\mathrm{d}m \geq 0. \end{aligned}$$

This tells us $K(z)$ is increasing with respect to $z \in [\underline{\sigma}, \overline{\sigma}]$, then $K(z)$ reaches its maximum at $z = \overline{\sigma}$. Hence,

$$\hat{\mathbb{E}}[\varphi(W)] = \max_{z\in[\underline{\sigma},\overline{\sigma}]} G(z) = \max_{z\in[\underline{\sigma},\overline{\sigma}]} (\varphi(0) + \frac{K(z)}{2}) = G(\overline{\sigma}).$$

2. When φ is concave, then $-\varphi$ is convex. replace the φ above with $-\varphi$ and repeat the same procedure, we have

$$\begin{aligned} -G(z) &= -E[\varphi(zY)] \\ &= E[(-\varphi)(zY)] \\ &= (-\varphi)(0) + \frac{z^2}{2}\underbrace{E[(-\varphi)^{(2)}(\xi_{zY})Y^2]}_{K(z)\geq 0}, \end{aligned}$$

and

$$K'(z) = \int \frac{1}{\sqrt{2\pi}}\underbrace{\left[\frac{m^2}{z^3}\exp\left(-\frac{m^2}{2z^2}\right)\right]}_{\geq 0 \text{ for } z\in[\underline{\sigma},\overline{\sigma}]}\underbrace{(-\varphi)^{(2)}(\xi_m)\,m^2}_{\geq 0}\,\mathrm{d}m \geq 0,$$

Hence, $-G(z)$ is increasing with respect to z, that is, $G(z)$ is decreasing according to z. Therefore,

$$\hat{\mathbb{E}}[\varphi(X)] = \max_{z\in[\underline{\sigma},\overline{\sigma}]} G(z) = G(\underline{\sigma}). \qquad \square$$

The initial motivation of the *semi-G-normal* distribution is that we want to create a tool or ladder to help us better understand and handle the G-normal distribution, especially based on what we already know about the classical normal distribution, which turns out to be feasible from the nice properties of semi-G-normal distribution and thanks to the constructed theory in G-framework (like the nonlinear CLT by Peng). In this thesis, we will further explore the semi-G-normal distribution by considering its independence, its multivariate version (with the construction from univariate objects) and the semi-G-Brownian motion (with its connection to the G-Brownian motion. For the statistical side, we also provide a pseudo approach to simulate the semi-G-normal distribution and the estimation method for the variance interval of $\hat{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$. The semi-$G$-normal distribution $\hat{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ behaves like the transition from the linear normal distribution $N(0, \sigma^2)$ to the sublinear G-normal distribution $N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$. The following result is one of the exciting results from the semi-G-normal distribution to better understand and compute the expectation of the G-normal distribution.

Lemma 2.3.3 (General connection between the Semi-G-normal and the G-normal distribution)**.** *In a sublinear expectation space $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$, consider a sequence of nonlinear i.i.d. $\{W_i\}_{i=1} \sim \hat{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ and $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$, then for any $\varphi \in C(\mathbb{R})$ satisfying linear growth condition, we have*

$$\hat{\mathbb{E}}[\varphi(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i)] \to \hat{\mathbb{E}}[\varphi(X)],$$

as $n \to \infty$. In other words, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i$ converges in distribution to the G-normal distributed X.

Proof. Since it is easy to check that W_1 has the zero mean and variance uncertainty: $\hat{\mathbb{E}}[W_1] = -\hat{\mathbb{E}}[-W_1] = 0$, $\hat{\mathbb{E}}[W_1^2] = \overline{\sigma}^2$, and $-\hat{\mathbb{E}}[-W_1^2] = \underline{\sigma}^2$, applying the nonlinear CLT with zero-mean, we get the required result. □

Comment 2.3.3.1. For a large N, the distribution of $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} W_i$ will be approximately identical with $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ (they have the same distribution uncertainty). In some sense, this actually provides us with one way to use W to approximately generate the G-normal distribution. At least, we can use $\hat{\mathbb{E}}[\varphi(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} W_i)]$ to approximate $\hat{\mathbb{E}}[\varphi(X)]$.

Theorem 2.3.4 (The Iterative Approximation of the G-normal Distribution)**.** *Consider a G-normal distributed random variable $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$. For any $\varphi \in C_{l.Lip}(\mathbb{R})$ and integer $n \geq 1$, consider the series of iteration functions $\{\varphi_{i,n}\}_{i=1}^{n}$ with initial function $\varphi_{0,n}(x) := \varphi(x)$ and iterative relation:*

$$\varphi_{i+1,n}(x) := \max_{v \in [\underline{\sigma}, \overline{\sigma}]} E[\varphi_{i,n}(N(x, v^2/n))], i = 0, 1, \ldots, n-1.$$

The final iteration function for a given n is $\varphi_{n,n}$. As $n \to \infty$, we have $\varphi_{n,n}(0) \to \hat{\mathbb{E}}[\varphi(X)]$.

Proof. Set the initial function $\varphi_{0,n}(x) := \varphi(x)$ and the iteration

$$\varphi_{i,n}(x) := \max_{v \in [\underline{\sigma}, \overline{\sigma}]} E[\varphi_{i-1,n}(N(x, \frac{v^2}{n}))].$$

In order to use the integral representation of the Semi-G-normal distribution (Theorem 2.3.2) in the next stage, we want each iteration function to be in the function space $C_{l.Lip}$. For convenience, we omit the subscript n for a while and let $\varphi_i := \varphi_{i,n}$. We also know that the optimal v will be some value in $[\underline{\sigma}, \overline{\sigma}]$ depending on x, i.e.

$$\varphi_i(x) = E[\varphi_{i-1}(N(x, \frac{v_x^2}{n})], v_x \in [\underline{\sigma}, \overline{\sigma}].$$

By induction, we only need to show that given $\varphi_{i-1} \in C_{l.Lip}$, we also have $\varphi_i \in C_{l.Lip}$, for $i = 1, 2, \ldots, n$. Suppose for φ_{i-1} we have a constant C_{i-1} and an positive integer k such that

$$|\varphi_{i-1}(x) - \varphi_{i-1}(y)| \leq C_{i-1}(1 + |x|^k + |y|^k)|x - y|.$$

Let $Z \sim N(0,1)$ and consider

$$\begin{aligned}
|\varphi_i(x) - \varphi_i(y)| &= |E[\varphi_{i-1}(N(x, \frac{v_x^2}{n})] - E[\varphi_{i-1}(N(y, \frac{v_y^2}{n})]| \\
&= |E[\varphi_{i-1}(x + \frac{v_x}{\sqrt{n}}Z)] - E[\varphi_{i-1}(y + \frac{v_y}{\sqrt{n}}Z)]| \\
&\leq E|\varphi_{i-1}(x + \frac{v_x}{\sqrt{n}}Z) - \varphi_{i-1}(y + \frac{v_y}{\sqrt{n}}Z)| \\
&\leq E[C_{i-1}(1 + |x + \frac{v_x}{\sqrt{n}}Z|^k + |y + \frac{v_y}{\sqrt{n}}Z|^k)|(x + Z) - (y + Z)|] \\
&= C_{i-1}|x - y|(1 + E|x + \frac{v_x}{\sqrt{n}}Z|^k + E|y + \frac{v_y}{\sqrt{n}}Z|^k).
\end{aligned}$$

For given $\omega \in \Omega$, let $z \coloneqq Z(\omega)$, we have

$$\begin{aligned}
|x + \frac{v_x}{\sqrt{n}}z|^k = |(x + \frac{v_x}{\sqrt{n}}z)^k| &\leq \sum_{j=1}^{k} \binom{k}{j} |x|^j |\frac{v_x}{\sqrt{n}}z|^{k-j} \\
&\leq \sum_{j=1}^{k} \binom{k}{j} \max\{|x|^k, |\frac{v_x}{\sqrt{n}}z|^k\} \leq 2^k(|x|^k + |\frac{v_x}{\sqrt{n}}|^k|z|^k) \leq 2^k(|x|^k + |\frac{\overline{\sigma}}{\sqrt{n}}|^k|z|^k).
\end{aligned}$$

Then $|x + \frac{v_x}{\sqrt{n}}Z(\omega)|^k \leq 2^k(|x|^k + |\frac{\overline{\sigma}}{\sqrt{n}}|^k|Z(\omega)|^k)$; taking expectation on both sides, we have

$$E|x + \frac{v_x}{\sqrt{n}}Z|^k \leq 2^k(|x|^k + |\frac{\overline{\sigma}}{\sqrt{n}}|^k E|Z|^k).$$

Therefore,

$$\begin{aligned}
|\varphi_i(x) - \varphi_i(y)| &\leq C_{i-1}|x - y|(1 + E|x + \frac{v_x}{\sqrt{n}}Z|^k + E|y + \frac{v_y}{\sqrt{n}}Z|^k) \\
&\leq C_{i-1}|x - y|(1 + 2^{k+1}|\frac{\overline{\sigma}}{\sqrt{n}}|^k E|Z|^k + 2^k|x|^k + 2^k|y|^k) \\
&\leq C_i(1 + |x|^k + |y|^k)|x - y|,
\end{aligned}$$

where $C_i \coloneqq C_{i-1} \cdot \max\{1 + 2^{k+1}|\frac{\overline{\sigma}}{\sqrt{n}}|^k E|Z|^k, 2^k\}$.

Considering a sequence of nonlinear i.i.d. $\{W_i\}_{i=1} \sim \hat{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ and $W_{i,n} \coloneqq \frac{1}{\sqrt{n}}W_i$, since $\varphi_i \in C_{l.Lip}$, we can apply the integral representation of the Semi-G-normal distribution at each

step. Then we have

$$\begin{aligned}\hat{\mathbb{E}}[\varphi(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_i)] &= \hat{\mathbb{E}}[\varphi_{0,n}(\sum_{i=1}^{n}W_{i,n})]\\ &= \hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi_{0,n}(\sum_{i=1}^{n-1}w_{i,n}+W_{n,n})]_{w_{i,n}=W_{i,n}\ i=1,2,\ldots,n-1}]\\ &= \hat{\mathbb{E}}\Big[[\max_{v_n\in[\underline{\sigma},\overline{\sigma}]}E[\varphi_{0,n}(\sum_{i=1}^{n-1}w_{i,n}+N(0,\frac{v_n^2}{n}))]]_{w_{i,n}=W_{i,n},\,i=1,2,\ldots,n-1}\Big]\\ &= \hat{\mathbb{E}}\Big[[\max_{v_n\in[\underline{\sigma},\overline{\sigma}]}E[\varphi_{0,n}(N(\sum_{i=1}^{n-1}w_{i,n},\frac{v_n^2}{n}))]]_{w_{i,n}=W_{i,n},\,i=1,2,\ldots,n-1}\Big]\\ &= \hat{\mathbb{E}}[\varphi_{1,n}(\sum_{i=1}^{n-1}W_{i,n})].\end{aligned}$$

By continue extracting the last component and doing the iteration, we have

$$\begin{aligned}\hat{\mathbb{E}}[\varphi_{1,n}(\sum_{i=1}^{n-1}W_{i,n})] &= \hat{\mathbb{E}}\Big[[\max_{v_{n-1}\in[\underline{\sigma},\overline{\sigma}]}E[\varphi_{1,n}(N(\sum_{i=1}^{n-2}w_{i,n},\frac{v_{n-1}^2}{n}))]]_{w_{i,n}=W_{i,n}\ i=1,2,\ldots,n-2}\Big]\\ &= \cdots = \hat{\mathbb{E}}\Big[[\max_{v_2\in[\underline{\sigma},\overline{\sigma}]}E[\varphi_{n-2,n}(N(w_{1,n},\frac{v_2^2}{n}))]]_{w_{1,n}=W_{1,n}}\Big]\\ &= \hat{\mathbb{E}}[\varphi_{n-1,n}(W_{1,n})] = \max_{v_1\in[\underline{\sigma},\overline{\sigma}]}E[\varphi_{n-1,n}(N(0,\frac{v_1^2}{n}))] = \varphi_{n,n}(0).\end{aligned}$$

According to Lemma 2.3.3, we have

$$\hat{\mathbb{E}}[\varphi(X)] = \lim_{n\to\infty}\hat{\mathbb{E}}[\varphi(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_i)] = \lim_{n\to\infty}\varphi_{n,n}(0). \qquad \square$$

Remark 2.3.4.1. From the proof, we note that the iteration function can also be expressed as the sublinear expectation of the semi-G-normal distribution (letting $W_0 \coloneqq 0$):

$$\varphi_{i,n}(x) = \hat{\mathbb{E}}[\varphi(x+\sum_{j=0}^{i}\frac{W_{n-j}}{\sqrt{n}})] = \hat{\mathbb{E}}[\varphi(x+\sum_{j=0}^{i}\frac{W_j}{\sqrt{n}})],$$

for $i = 0, 1, \ldots, n$.

Corollary 2.3.4.1. *Consider the G-heat equation defined on* $[0,\infty)\times\mathbb{R}$*:*

$$u_t + G(u_{xx}) = 0,\ u|_{t=1} = \varphi,$$

where $G(a) \coloneqq \frac{1}{2}\hat{\mathbb{E}}[aX^2] = \frac{1}{2}(\overline{\sigma}^2a^+ - \underline{\sigma}^2a^-)$ *and* $\varphi \in C_{l.Lip}(\mathbb{R})$*. Then for the iterations* $\{\varphi_{i,n}\}_{i=0}^{n}$ *in Theorem 2.3.4, for each* $p \in (0,1]$*, there exists* $\alpha \in (0,1)$ *such that,*

$$|u(1-p,x) - \varphi_{\lfloor np\rfloor,n}(x)| = |\hat{\mathbb{E}}[\varphi(x+\sqrt{p}X)] - \hat{\mathbb{E}}[\varphi(x+\sum_{i=0}^{\lfloor np\rfloor}\frac{W_i}{\sqrt{n}})]| = C_\varphi(1+|x|^k)O(\frac{1}{(np)^{\alpha/2}}).$$

where $\lfloor np\rfloor$ *is the floor (or integer) part of np. When* $p = 0$*, we have* $u(1,x) = \varphi(x) = \varphi_{0,n}(x)$*.*

Proof. For each $p \in (0,1)$, consider the error when approximating $u(1-p,x)$, which can be approximated by

$$u(1-\frac{\lfloor np\rfloor}{n},x)=\hat{\mathbb{E}}[\varphi(x+\sum_{j=0}^{\lfloor np\rfloor}\frac{X_j}{\sqrt{n}})]\approx\hat{\mathbb{E}}[\varphi(x+\sum_{i=0}^{\lfloor np\rfloor}\frac{W_i}{\sqrt{n}})]=\varphi_{\lfloor np\rfloor,n}(x).$$

Specifically speaking, we intend to work on the error

$$|u(1-p,x)-\varphi_{\lfloor np\rfloor,n}(x)|\leq\underbrace{|u(1-p,x)-u(1-\frac{\lfloor np\rfloor}{n},x)|}_{(1)}+\underbrace{|u(1-\frac{\lfloor np\rfloor}{n},x)-\varphi_{\lfloor np\rfloor,n}(x)|}_{(2)}.$$

Before diving into these two parts, we can prepare the converging property of $p_n:=\frac{\lfloor np\rfloor}{n}$. Actually, the inequality

$$p_n=\frac{\lfloor np\rfloor}{n}\leq\frac{np}{n}\leq\frac{\lfloor np\rfloor+1}{n}=p_n+\frac{1}{n},$$

tells us that

$$|p_n-p|<\frac{1}{n}.$$

Part (1) involves the continuity of u on the time dimension specified by doing the Taylor expansion:

$$(1)\leq|u_t(1-p,x)||p_n-p|+\underbrace{O(|p_n-p|^2)}_{O(\frac{1}{n^2})}.$$

We are looking for bound of $|u_t|$. Fortunately, according to the properties of the solution to the G-heat equation (ref), there exist constant $C>0$ and $\beta>0$, such that

$$|u_t(t,x)-u_t(s,x)|\leq C|t-s|^{\frac{\beta}{2}}.$$

By letting $s=0$ and $c_x:=|u_t(0,x)|$, we have

$$|u_t(t,x)|\leq|u_t(t,x)-u_t(0,x)|+|u_t(0,x)|\leq C|t|^{\frac{\beta}{2}}+c_x.$$

Therefore, for a fixed x, we have

$$(1)\leq(C|1-p|^{\frac{\beta}{2}}+c_x)|p_n-p|+O(\frac{1}{n^2})=c_xO(\frac{1}{n}).$$

Part (2) can be rewritten as follows:

$$\begin{aligned}(2)&=|\hat{\mathbb{E}}[\varphi(x+\frac{1}{\sqrt{n}}\sum_{j=0}^{\lfloor np\rfloor}X_j)]-\hat{\mathbb{E}}[\varphi(x+\frac{1}{\sqrt{n}}\sum_{j=0}^{\lfloor np\rfloor}W_j)]|\\&=|\hat{\mathbb{E}}[\varphi(x+\sqrt{\frac{\lfloor np\rfloor}{n}}\frac{1}{\sqrt{\lfloor np\rfloor}}\sum_{j=0}^{\lfloor np\rfloor}X_j)]-\hat{\mathbb{E}}[\varphi(x+\sqrt{\frac{\lfloor np\rfloor}{n}}\frac{1}{\sqrt{\lfloor np\rfloor}}\sum_{j=0}^{\lfloor np\rfloor}W_j)]|.\end{aligned}$$

By letting $\mathbf{X}_{np} := \frac{1}{\sqrt{\lfloor np \rfloor}} \sum_{j=0}^{\lfloor np \rfloor} X_j$ and $\mathbf{W}_{np} := \frac{1}{\sqrt{\lfloor np \rfloor}} \sum_{j=0}^{\lfloor np \rfloor} W_j$, we have

$$\begin{aligned}(2) &= |\hat{\mathbb{E}}[\varphi(x + \sqrt{p_n}\mathbf{X}_{np})] - \hat{\mathbb{E}}[\varphi(x + \sqrt{p_n}\mathbf{W}_{np})]| \\ &\leq |\hat{\mathbb{E}}[\varphi(x + \sqrt{p_n}\mathbf{X}_{np})] - \hat{\mathbb{E}}[\varphi(x + \sqrt{p}\mathbf{X}_{np})]| \\ &+ |\hat{\mathbb{E}}[\varphi(x + \sqrt{p_n}\mathbf{W}_{np})] - \hat{\mathbb{E}}[\varphi(x + \sqrt{p}\mathbf{W}_{np})]| \\ &+ |\hat{\mathbb{E}}[\varphi(x + \sqrt{p}\mathbf{X}_{np})] - \hat{\mathbb{E}}[\varphi(x + \sqrt{p}\mathbf{W}_{np})]| \\ &:= (2)_1 + (2)_2 + (2)_3.\end{aligned}$$

where $(2)_1 + (2)_2$ involves the continuity of φ and the shrinking speed of $|p_n - p|$ and $(2)_3$ is exactly fitted into the convergence rate of nonlinear CLT. In $(2)_1$ or $(2)_2$, we do not need to worry about the random variables since $\mathbf{X}_{np} \overset{\mathrm{d}}{=} X \overset{\mathrm{d}}{=} N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ and $\mathbf{W}_{np} \overset{\mathrm{d}}{\to} X$ as $n \to \infty$ from nonlinear CLT. Hence, let us work on a general expression by replacing $\mathbf{X}_{np}$ and $\mathbf{W}_{np}$ by Z_n satisfying $Z_n \overset{\mathrm{d}}{\to} X$. We know $\varphi \in C_{l.Lip}$ satisfying, for $x, y \in \mathbb{R}$,

$$|\varphi(x) - \varphi(y)| \leq C_\varphi(1 + |x|^k + |y|^k)|x - y|,$$

with $C_\varphi > 0$ and $k \in \mathbb{N}$. When $k = 0$, φ is uniformly Lipschitz, then

$$\begin{aligned}(2)_1 \text{or} (2)_2 &= |\hat{\mathbb{E}}[\varphi(x + \sqrt{p_n}Z_n)] - \hat{\mathbb{E}}[\varphi(x + \sqrt{p}Z_n)]| \\ &\leq \hat{\mathbb{E}}[|\varphi(x + \sqrt{p_n}Z_n) - \varphi(x + \sqrt{p}Z_n)|] \\ &\leq \hat{\mathbb{E}}[C_\varphi|\sqrt{p_n} - \sqrt{p}||Z_n|] = C_\varphi|\sqrt{p_n} - \sqrt{p}|\hat{\mathbb{E}}[|Z_n|].\end{aligned}$$

For the expectation part in the last line, since $Z_n \overset{\mathrm{d}}{\to} X$, we can make n large enough so that

$$\hat{\mathbb{E}}[|Z_n|] \leq \hat{\mathbb{E}}[|X|] + 1 =: K_0.$$

For the p_n part, the Taylor expansion of $\sqrt{x}$ at $x = p$ tells us that

$$|\sqrt{p_n} - \sqrt{p}| \leq \frac{1}{2\sqrt{p}}|p_n - p| + O(|p_n - p|^2) = O(\frac{1}{n\sqrt{p}}).$$

Hence, when $k = 0$,

$$(2)_1 \text{or} (2)_2 \leq C_1 O(\frac{1}{n\sqrt{p}}),$$

with $C_1 = C_\varphi K_0$. When $k \geq 1$, we have

$$\begin{aligned}(2)_1 \text{or} (2)_2 &\leq \hat{\mathbb{E}}[|\varphi(x + \sqrt{p_n}Z_n) - \varphi(x + \sqrt{p}Z_n)|] \\ &\leq \hat{\mathbb{E}}[C_\varphi(1 + |x + \sqrt{p_n}Z_n|^k + |x + \sqrt{p}Z_n|^k)|\sqrt{p_n} - \sqrt{p}||Z_n|] \\ &\leq C_\varphi\hat{\mathbb{E}}\Big[(1 + 2^{k+1}|x|^k + 2^k|Z_n|^k(\underbrace{|\sqrt{p_n}|^k + |\sqrt{p}|^k}_{\leq 2}))|Z_n|\Big]|\sqrt{p_n} - \sqrt{p}|.\end{aligned}$$

For the expectation part in the last line, by letting n be large enough so that

$$(\hat{\mathbb{E}}[|Z_n|^2])^{1/2} \leq (\hat{\mathbb{E}}[|X|^2])^{1/2} + 1 = (E[|N(0, \overline{\sigma}^2)|^2)^{1/2} + 1 = \overline{\sigma} + 1 =: K_1,$$

and

$$(\hat{\mathbb{E}}[|Z_n|^{2k}])^{1/2} \le (\hat{\mathbb{E}}[|X|^{2k}])^{1/2} + 1 = (E[|N(0,\overline{\sigma}^2)|^{2k})^{1/2} + 1 = \overline{\sigma}^k((2k-1)!!)^{1/2} + 1 =: K_2,$$

we have

$$\begin{aligned}\hat{\mathbb{E}}\Big[(1 + 2^{k+1}|x|^k + 2^{k+1}|Z_n|^k)|Z_n|\Big] &\le (\hat{\mathbb{E}}[(1 + 2^{k+1}|x|^k + 2^{k+1}|Z_n|^k)^2])^{1/2}(\hat{\mathbb{E}}[|Z_n|^2])^{1/2} \\ &\le \Big(1 + 2^{k+1}|x|^k + 2^{k+1}(\hat{\mathbb{E}}[|Z_n|^{2k}])^{1/2}\Big)(\hat{\mathbb{E}}[|Z_n|^2])^{1/2} \\ &\le C_2(1 + |x|^k),\end{aligned}$$

where $C_2 = K_1 \cdot \max\{1 + 2^{k+1}K_2, 2^{k+1}\}$. Again, for the p_n part, we have $|p_n - p| = O(\frac{1}{n\sqrt{p}})$. Therefore, when $k \ge 1$, for a fixed x,

$$(2)_1 \text{or} (2)_2 \le C_3(1 + |x|^k)O(\frac{1}{n\sqrt{p}}),$$

where $C_3 = C_\varphi C_2$. In a word, for $k \in \mathbb{N}$,

$$(2)_1 \text{or} (2)_2 \le C_4(1 + |x|^k)O(\frac{1}{n\sqrt{p}}),$$

where $C_4 = \max\{C_1, C_3\}$.

In $(2)_3$, we can directly apply the convergence rate of nonlinear CLT (Theorem 2.1.19):

$$\begin{aligned}(2)_3 &= |\hat{\mathbb{E}}[\varphi(x + \sqrt{p}\frac{1}{\sqrt{\lfloor np\rfloor}}\sum_{j=0}^{\lfloor np\rfloor} X_j)] - \hat{\mathbb{E}}[\varphi(x + \sqrt{p}\frac{1}{\sqrt{\lfloor np\rfloor}}\sum_{j=0}^{\lfloor np\rfloor} W_j)]| \\ &= |\hat{\mathbb{E}}[\tilde{\varphi}(\frac{1}{\sqrt{\lfloor np\rfloor}}\sum_{j=0}^{\lfloor np\rfloor} X_j)] - \hat{\mathbb{E}}[\tilde{\varphi}(\frac{1}{\sqrt{\lfloor np\rfloor}}\sum_{j=0}^{\lfloor np\rfloor} W_j)] = C_{\tilde{\varphi}}O(\frac{1}{(np)^{\alpha/2}}),\end{aligned}$$

where $\tilde{\varphi}(a) \coloneqq \varphi(x + \sqrt{p}a)$, satisfying

$$\begin{aligned}|\tilde{\varphi}(a) - \tilde{\varphi}(b)| &\le C_\varphi(1 + |x + \sqrt{p}a|^k + |x + \sqrt{p}b|^k)\sqrt{p}|a - b| \\ &\le C_\varphi(1 + 2^k(|x|^k + |\sqrt{p}a|^k) + 2^k(|x|^k + |\sqrt{p}b|^k)|a - b| \\ &\le C_\varphi(1 + 2^{k+1}|x|^k + 2^k(|a|^k + |b|^k))|a - b| \\ &\le C_{\tilde{\varphi}}(1 + |a|^k + |b|^k)|a - b|,\end{aligned}$$

in which $C_{\tilde{\varphi}} = C_\varphi \max\{1 + 2^{k+1}|x|^k, 2^k\} \le C_\varphi(1 + 2^k + 2^{k+1}|x|^k)$. Hence,

$$(2)_3 \le C_5(1 + |x|^k)O(\frac{1}{(np)^{\alpha/2}}),$$

where $C_5 = C_\varphi 2^{k+1}$. To summarize, with $\alpha \in (0,1)$, for a given p and fixed x, the error is

$$\begin{aligned}|u(1-p,x) - \varphi_{\lfloor np\rfloor,n}(x)| &\le (1) + (2) \\ &\le c_x O(\frac{1}{n}) + 2C_4(1 + |x|^k)O(\frac{1}{n\sqrt{p}}) + C_5(1 + |x|^k)O(\frac{1}{(np)^{\alpha/2}}) \\ &= M(1 + |x|^k)O(\frac{1}{(np)^{\alpha/2}}).\end{aligned}$$

Since $p \in (0,1)$, we only need to consider what happens when p approaches to 0, in order to get a bound similar with a uniform bound, by letting $q_n := \frac{1}{\sqrt{n}}$, we have

$$\sup_{p>q_n} |u(1-p,x) - \varphi_{\lfloor np \rfloor, n}(x)| \leq M(1+|x|^k)O(\frac{1}{(\sqrt{n})^{\alpha/2}}).$$

□

2.3.2 The d-dimensional situation

The definition of semi-G-normal distribution can be naturally extended to multi-dimensional situation. Intuitively speaking, the multivariate semi-G-normal distribution can be treated as an analogue of the linear multivariate normal distribution: $N(\mathbf{0}, \mathbf{V}) = \mathbf{V}^{1/2} N(\mathbf{0}, \mathbf{I}_d)$, where $\mathbf{I}_d$ is a $d \times d$ identity matrix.

Definition 2.3.5 (the Semi-G-normal distribution in d dimension)**.** Let a bounded, closed and convex subset $\mathcal{V} \subset \mathbb{S}_d^+$ be the uncertainty set of covariance matrices, i.e.

$$\mathcal{V} := \Bigg\{ \mathbf{V} = \left(\rho_{ij}\sigma_i\sigma_j\right)_{d\times d} : \sigma_i^2 \in [\underline{\sigma}_i^2, \overline{\sigma}_i^2],$$
$$\rho_{ij} = \rho_{ji} = \begin{cases} 1 & i = j \\ \in [\underline{\rho}_{ij}, \overline{\rho}_{ij}] & i \neq j \end{cases}, \text{ such that } \mathbf{V} \in \mathbb{S}_d^+ \Bigg\},$$

and $\mathcal{V}^{1/2} := \{\mathbf{V}^{1/2} : \mathbf{V} \in \mathcal{V}\}$ where $\mathbf{V}^{1/2}$ is the symmetric square root of $\mathbf{V}$. We say a d-dimensional random vector $W : \Omega \to \mathbb{R}^d$ will follow the *Semi-G-normal distribution*, denoted as, $W \sim \hat{N}(\mathbf{0}, \mathcal{V})$, if there exist a d-dimensional G-normal distributed random vector

$$Y \sim N(\mathbf{0}, \mathbf{I}_d) : \Omega \to \mathbb{R}^d,$$

and a $d \times d$-dimensional maximal distributed random matrix

$$Z \sim M(\mathcal{V}^{1/2}) : \Omega \to \mathbb{R}^{d\times d},$$

as well as Y is independent from Z, such that

$$W = Z \cdot Y,$$

where "$\cdot$" is the matrix multiplication (which can be omitted) and the direction of independence here cannot be reversed.

Remark 2.3.5.1. Y can be regarded as the classical multivariate normal distribution with identity covariance matrix.

Corollary 2.3.5.1 (The Integral Representation of the Semi-G-normal distribution in d dimension)**.** *Consider a d-dimensional random vector $W \sim \hat{N}(\mathbf{0}, \mathcal{V})$, where $\mathcal{V}$ is the uncertainty set of covariance matrices. Then for any $\varphi \in C_{l.Lip}(\mathbb{R}^d)$, we have*

$$\hat{\mathbb{E}}[\varphi(W)] = \max_{\mathbf{V}\in\mathcal{V}} E[\varphi(N(\mathbf{0},\mathbf{V}))] = \max_{\mathbf{V}^{1/2}\in\mathcal{V}^{1/2}} E[\varphi(\mathbf{V}^{1/2}N(\mathbf{0},\mathbb{I}_d))]$$
$$= \max_{\mathbf{V}^{1/2}\in\mathcal{V}^{1/2}} \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{\frac{d}{2}}} \exp(-\frac{1}{2}\boldsymbol{y}'\boldsymbol{y})\varphi(\mathbf{V}^{1/2}\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}.$$

Similarly, we can obtain the iterative approximation in multi-dimensional case.

Theorem 2.3.6 (The Iterative Approximation of the G-normal Distribution in d dimension)**.** *Consider a G-normal distributed random variable $X \sim N(\mathbf{0}, \mathcal{V})$. For any $\varphi \in C_{l.Lip}(\mathbb{R}^d)$ and integer $n \geq 1$, consider the series of iteration functions $\{\varphi_{i,n}\}_{i=1}^n$ with initial function $\varphi_{0,n}(\boldsymbol{x}) := \varphi(\boldsymbol{x})$ and iterative relation:*

$$\varphi_{i+1,n}(\boldsymbol{x}) := \max_{\mathbf{V}\in\mathcal{V}} E[\varphi_{i,n}(N(\boldsymbol{x}, \mathbf{V}/n))], i = 0, 1, \ldots, n-1.$$

The final iteration function for a given n is $\varphi_{n,n}$. As $n \to \infty$, we have $\varphi_{n,n}(\mathbf{0}) \to \hat{\mathbb{E}}[\varphi(X)]$.

2.3.3 Implementation

The 1-dimensional case

Consider a G-normal distributed random variable $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ with $[\underline{\sigma}, \overline{\sigma}] = [0.5, 1]$. For any $\varphi \in C_{l.Lip}(\mathbb{R})$, in order to compute $\hat{\mathbb{E}}[\varphi(X)]$, setting a fixed large n as the total number of iterations, we implement the following procedure:

1. Start from
$$\varphi_{0,n}(x) := \varphi(x);$$
2. Since we are iterating the functions on the infinite domain $\mathbb{R}$, in practice, we need to set up a finite grid to do interpolation at each step. Choose a large constant K to decide the range of the numerical domain of x then set up the spatial grid:
$$-K = x_0 < x_1 < x_2 < \ldots < x_L = K;$$
3. At the iteration step $i(= 1, 2, \ldots, n-1)$, for each $x = x_j$, $j = 0, 1, \ldots, L$, evaluate
$$\varphi_{i+1,n}(x_j) := \max_{v\in[\underline{\sigma},\overline{\sigma}]} E[\varphi_{i,n}(N(x_j, \frac{v^2}{n}))],$$
where the linear expectation can be computed from integration or MC (Monte Carlo) method (by generating a linearly i.i.d. standard normal sample: $Z_1, Z_2, \ldots, Z_M \sim N(0, 1)$):
$$\begin{aligned} E[\varphi_{i,n}(N(x, \frac{v^2}{n}))] &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \frac{\sqrt{n}}{v} \exp(-(\frac{\sqrt{n}m}{v})^2/2)\varphi_{i,n}(x+m)\,\mathrm{d}m \\ &\approx \frac{1}{M}\sum_{k=1}^{M} \varphi_{i,n}(x + \frac{v}{\sqrt{n}}Z_k). \end{aligned}$$
Then take maximum of $E[\varphi_{i,n}(N(x, \frac{v^2}{n}))]$ over $v \in [\underline{\sigma}, \overline{\sigma}]$ to get $\varphi_{i+1,n}(x)$ by doing appropriate optimization: here we use a quasi-Newton method called the "L-BFGS-B" by Byrd et al. (1995) which "allows box constraints, that is each variable can be given a lower and/or upper bound". Then use $\varphi_{i,n}(x_j)$, $j = 0, 1, \ldots, L$, to fit the function $\varphi_{i,n}$ by choosing proper splines to do both interpolation and extrapolation based on the type of φ. For instance, for $\varphi(x) = x^3$, we can use the "fmm" method by Forsythe et al. (1977): "an exact cubic spline is fitted through the four points at each end of the data, and this is used to determine the end conditions. "

Finally, we have

$$\hat{\mathbb{E}}[\varphi(X)] \approx \max_{v\in[\underline{\sigma},\overline{\sigma}]} E[\varphi_{n-1,n}(N(0,v^2))] = \varphi_{n,n}(0).$$

Comment 2.3.6.1. In the step 3, we notice that it is necessary to use these values $\varphi_{i,n}(x+\frac{v}{\sqrt{n}}Z_k)$, $k = 1, 2, \ldots, M$, with M points in the neighbourhood of one points x, to estimate the $\varphi_{i,n}(x)$. If we use a larger grid (with M^i points) for $\varphi_{i,n}$ to preserve the precision of $\varphi_{i,n}$ on a smaller grid (with M^{i-1} points), we will be stuck into the so-called "nested situation" unless we are dealing with functions with bounded domain. As we increase n, namely, the number of iterations, even if we only want to compute one point of the last iterative function with certain precision, in the previous iterations, we will still need to prepare a series of grids which is enlarged exponentially with respect to the iteration step. Therefore, one crucial step here is the interpolation (and extrapolation) to avoid the nested dilemma, which can help us get a function with continuous domain from a fixed discrete grid. The splines should be chosen based on the type of φ (polynomial, periodic and so on). Meanwhile, we need to appropriately choose the constant K to determine the numerical domain, to make the spline model able to *capture* the pattern of the function so as to achieve the best extrapolation and interpolation performance.

Remark 2.3.6.1. When evaluating the linear expectation $E[\varphi_i(N(x,\frac{v^2}{n}))]$ (omitting n for a while), to preserve the precision regardless of x, we can apply the MC with control variables. One common problem of Monte Carlo method is that given a normal sample, the error of the MC estimation will increase when x is farther from zero. Specifically, let $h \coloneqq \frac{1}{n}$ and $Z \sim N(0,1)$, then

$$E[\varphi_i(N(x,\tfrac{v^2}{n}))] \approx \tfrac{1}{M}\textstyle\sum_{m=1}^{M} \varphi_i(x+\sqrt{h}vZ_i) =$$
$$\tfrac{1}{M}\textstyle\sum_{m=1}^{M}\left(\varphi_i(x) + \frac{\varphi_i^{(1)}(x)}{1}\sqrt{h}vZ_i + \frac{\varphi_i^{(2)}(x)}{2}hv^2Z_i^2 + \frac{\varphi_i^{(3)}(\xi_{x,Z_i})}{6}h^{\frac{3}{2}}v^3Z_i^3\right).$$

Hence, the error can be expressed as,

$$\tfrac{1}{M}\textstyle\sum_{m=1}^{M} \varphi_i(x+\sqrt{h}vZ_i) - E[\varphi_i(x+\sqrt{h}vZ)] \approx$$
$$\frac{\varphi_i^{(1)}(x)}{1}\sqrt{h}v(\tfrac{1}{M}\textstyle\sum_{m=1}^{M} Z_i - 0) + \frac{\varphi_i^{(2)}(x)}{2}hv^2(\tfrac{1}{M}\textstyle\sum_{m=1}^{M} Z_i^2 - 1).$$

Suppose $\epsilon_1 \coloneqq \frac{1}{M}\sum_{m=1}^{M} Z_i - 0$ and $\epsilon_2 \coloneqq \frac{1}{M}\sum_{m=1}^{M} Z_i^2 - 1$, for a given sample $\{Z_i\}_{i=1}^{M}$, ϵ_1 and ϵ_2 are fixed number, and even if we regenerate the $\{Z_i\}_{i=1}^{M}$ each time, the random variable ϵ_1 and ϵ_2 should also not have so much variation because we know

$$\epsilon_1 \sim N(0,\frac{1}{M}),\ \epsilon_2 \sim \frac{1}{M}\chi^2_{(M)},$$

then $\mathrm{Var}[\epsilon_1] = \frac{1}{M}$ and $\mathrm{Var}[\epsilon_2] = \frac{2}{M}$. However, for the case $\varphi_1(x) = x^3$, when x goes farther away from zero, the values of $|\varphi_1^{(1)}(x)| = 3|x|^2$ and $|\varphi_1^{(2)}(x)| = 6|x|$ will become larger. Therefore, without loss of generality, for a fixed normal sample, the error of the MC will increase as x goes away from zero. This problem can not be overcome by simply increase the M (since the $|\varphi_1^{(1)}(x)|$ and $|\varphi_1^{(2)}(x)|$ will always expand the small error anyway and these enlarged errors will be cumulated as time goes further backward). Fortunately, we may try a type of variance reduction method for Monte Carlo method, called the **control variables**. We can use this method

to preserve the consistent precision of MC method outside the neighbourhood of x. The idea is that, after getting the $\varphi_1^{(i)}(x), i = 1, 2$, approximate the $E[\varphi(x + \sqrt{h}vZ)]$ by

$$\begin{aligned} &E[\varphi(x + \sqrt{h}vZ) - \varphi(x) - \varphi(x) - \varphi^{(1)}(x)\sqrt{h}vZ - \tfrac{\varphi^{(2)}(x)}{2}hv^2Z^2] + \varphi(x) + \tfrac{\varphi^{(2)}(x)}{2}hv^2 \\ \approx\ &\tfrac{1}{M}\textstyle\sum_{i=1}^{M}[\varphi(x + \sqrt{h}vZ_i) - \varphi(x) - \varphi^{(1)}(x)\sqrt{h}vZ_i - \tfrac{\varphi^{(2)}(x)}{2}hv^2Z_i] + \varphi(x) + \tfrac{\varphi^{(2)}(x)}{2}hv^2 \\ &(= \tfrac{1}{M}\textstyle\sum_{i=1}^{M}[\tfrac{\varphi_i^{(3)}(\xi_{x,Z_i})}{6}h^{\frac{3}{2}}v^3Z_i^3] + \varphi(x) + \tfrac{\varphi^{(2)}(x)}{2}hv^2). \end{aligned}$$

In this way, because of the boundedness of $\varphi^{(3)}(x)(= 6)$ for $\varphi(x) = x^3$, the error will not be enlarged by $\varphi_i^{(3)}(\xi_{x,Z_i})$ when x moves farther away zero.

Remark 2.3.6.2. Let $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$. In the context of the 1-dimensional G-heat equation:

$$u_t + G(\partial_x^2 u) = 0,\ u|_{t=1} = \varphi,$$

where $G(a) := \frac{1}{2}\hat{\mathbb{E}}[aX^2] = \frac{1}{2}(\overline{\sigma}^2 a^+ - \underline{\sigma}^2 a^-)$. We know there is a natural connection between the G-heat equation and G-normal distribution:

$$u(t, x) = \hat{\mathbb{E}}[\varphi(x + \sqrt{1 - t}X)].$$

Then we can make a collection of the iterative functions with properties:

1. $\varphi_{n,n}(0) \approx \hat{\mathbb{E}}[\varphi(X)] = u(0, 0)$;
2. $\varphi_{n,n}(x) \approx u(0, x)$;
3. In general, $\varphi_{k,n}(x) \approx u(1 - \frac{k}{n}, x)$, for $k = 0, 1, \ldots, n$.

This will actually give us a surface of the solution $u(t, x)$ with $t = 1 - \frac{k}{n}, k = 0, 1, \ldots, n$.

In the 2×2-layout Figure 2.1, for the first row ($n = 50, K = 5$), panel-$(1, 1)$ is the approximated solution paths of the G-heat equation with $\varphi(x) = x^3 = \varphi_{0,n}(x)$ (the black solid curve, with y-axis showing the value of $\varphi_{i,n}(x)$) and the curves whose right branches are moving up are the $\varphi_{i,n}$'s as the iteration proceeds. The panel-$(1, 2)$ shows the approximated paths with $\varphi(x) = (1 - |x|)\mathbb{I}_{(-1,1)}(x)$ (which belongs to a class of functions that, previously, are hard to deal with if not applying special PDE methods). The panel-$(2, 1)$ is the pathwise comparison plot ($n = 100, K = 50, \varphi(x) = x^3$) with the explicit solution provided in Hu (2012) where we can only notice the error (with absolute value ≤ 0.004) around $x = 0$ and we can see the stable accuracy beyond the spatial grid $[-K, K]$ (the black dashed lines).The panel-$(2, 2)$ is the approximation of $\hat{\mathbb{E}}[X^3]$ as n increases ($K = 5$) (where the horizontal line labels the true value; here we use integration to compute E, if using MC, we will assign a variance for each point).

For curiosity, we can play with this algorithm ($n = 50, K = 5$) by changing the terminal function to other $\varphi \in C_{l.Lip}$ to produce the Figure 2.2: panel-$(1, 1)$ ($\varphi(x) = \sin x$), panel-$(2, 1)$ ($\varphi(x) = \cos x$), panel-$(1, 2)$ ($\varphi(x) = \sin x + \cos x$), and panel-$(1, 1)$ ($\varphi(x) = 1/(1 + \exp(-x^2))$). From these plots, we can see that our method works well for these terminal function which is neither convex nor concave. In fact, Corollary 2.3.4.1 guarantees the convergence of our iterative functions to the respective solution paths in theory although for the sake of numerical comparison, we still need to compare them with the solutions from numeric PDE methods. Interestingly, from panel-$(1, 1)$ of Figure 2.2, we can notice that the terminal function $u(1, x) = \varphi(x) = \sin(x)$ is centrosymmetric around origin point but as the time goes backward, solution paths gradually lose this symmetry. We can find the similar feature for the results of $\varphi(x) = x^3$.

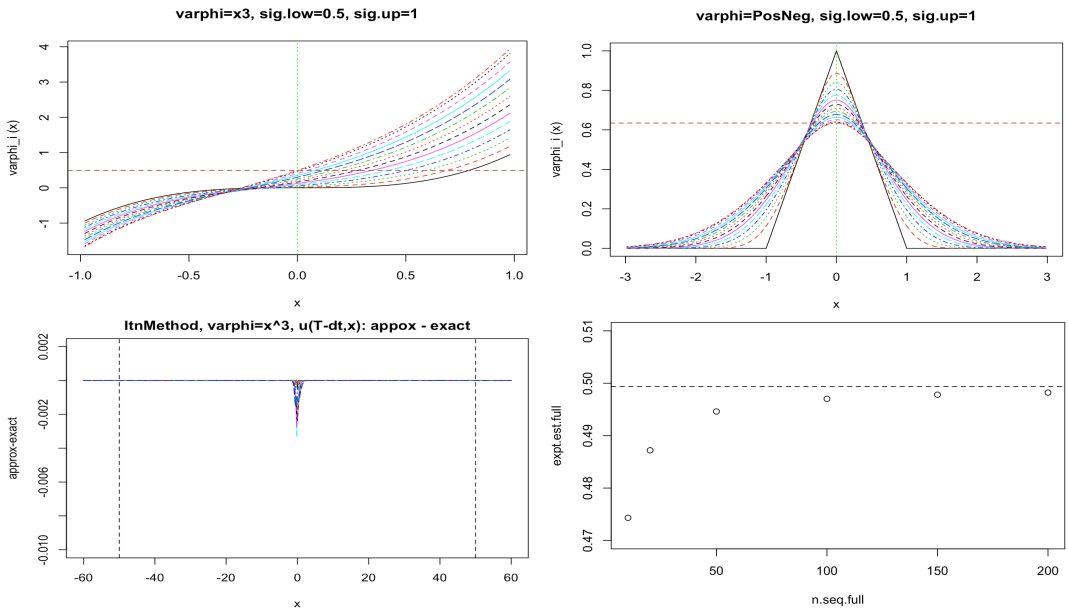


Figure 2.1: Numerical solution paths of G-heat equation from our iterative method (Set I)

The d-dimensional case

For the multi-dimensional G-heat equation (with covariance uncertainty), the MC setting may relieve the iterative algorithm from the curse of dimensionality.

Let $\boldsymbol{x} := (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$. Consider a d-dimensional G-normal distributed random vector $X \sim N(\mathbf{0}, \mathcal{V})$, where $\mathcal{V}$ is the uncertainty set of covariance matrices. For any $\varphi \in C_{l.Lip}(\mathbb{R}^d)$, in order to compute $\hat{\mathbb{E}}[\varphi(X)]$, similarly, set a large n as the total number of iterations, then we need to do the procedure:

1. Start from
$$\varphi_{0,n}(\boldsymbol{x}) := \varphi(\boldsymbol{x});$$

2. We need to set up a grid in the domain $\mathbb{R}^d$, to avoid the curse of dimensionality in the sense that the rectangular grid points become sparser in higher dimension and much more with exponential rate (L points in each dimension mean L^d points in total). Therefore, here we use the Monte Carlo grid points sampling from a d-dimensional multivariate normal distribution:
$$\{\boldsymbol{x}_j\}_{j=0}^{L} \sim N(\mathbf{0}, \overline{\sigma}^2 \mathbf{I}_d).$$

3. For $i = 1, 2, \ldots, n$, and for each $\boldsymbol{x} = \boldsymbol{x}_j$, $j = 0, 1, \ldots, L$, let
$$\varphi_{i+1,n}(\boldsymbol{x}) := \max_{\mathbf{V} \in \mathcal{V}} E[\varphi_{i,n}(N(\boldsymbol{x}, \frac{\mathbf{V}}{n}))],$$
where, again, to deal with the curse of dimensionality, the expectation can be computed by Monte Carlo method which maintains its convergence rate regardless of the dimension by generating a linearly i.i.d. sample from standard multivariate normal $N(0, \mathbb{I}_d)$: $Z_1, Z_2, \ldots, Z_M$ (and we can also apply the control variable method to protect the precision

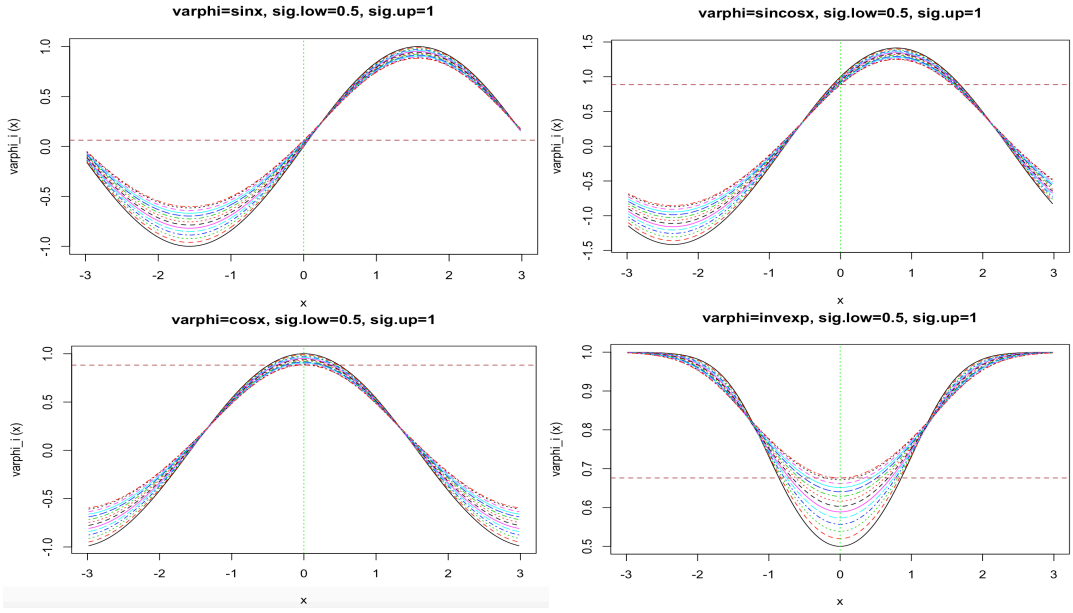


Figure 2.2: Numerical solution paths of G-heat equation from our iterative method (Set II)

from the variation of $\boldsymbol{x}$):

$$E[\varphi_{i,n}(N(\boldsymbol{x}, \frac{\mathbf{V}}{n}))] \approx \frac{1}{M}\sum_{k=1}^{M} \varphi_{i,n}(\boldsymbol{x} + \frac{\mathbf{V}^{1/2}}{\sqrt{n}} Z_k).$$

Take maximum of $E[\varphi_{i-1,n}(N(\boldsymbol{x}, \frac{\mathbf{V}}{n}))]$ over $\mathbf{V} \in \mathcal{V}$ by appropriate optimization (here we still use the "L-BFGS-B" method). Then use $\varphi_{i+1,n}(x_j)$, $j = 0, 1, \ldots, L$, to fit the function $\varphi_{i+1,n}$ by applying a proper spline model to do interpolation and extrapolation (here we work on the setting of *Generalized Additive Model* after testing and design the structure of splines based on the properties of φ).

Finally, we have

$$\begin{aligned}\hat{\mathbb{E}}[\varphi(X)] &\approx \max_{\mathbf{V}^{1/2} \in \mathcal{V}^{1/2}} E[\varphi_{n-1,n}(N(\mathbf{0}, \frac{\mathbf{V}}{n}))] \\ &= \varphi_{n,n}(\mathbf{0}).\end{aligned}$$

Remark 2.3.6.3. Consider a 2-dimensional G-normal distributed random variable $X \sim N(0, \mathcal{V})$ where

$$\mathcal{V} := \left\{ \mathbf{V} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} : \sigma_i \in [0.5, 1], i = 1, 2; \rho \in [-0.5, 0.5], \text{ such that } \mathbf{V} \in \mathbb{S}_d^+ \right\}.$$

In the context of 2-dimensional G-heat equation:

$$u_t + G(D_x^2 u) = 0,\ u|_{t=1} = \varphi,$$

where $G(\mathbf{A}) := \frac{1}{2}\hat{\mathbb{E}}[\langle \mathbf{A}X, X\rangle] : \mathbb{S}_d \to \mathbb{R}$, we also have the natural connection between the G-heat equation and G-normal distribution,

$$u(t, x) = \hat{\mathbb{E}}[\varphi(\boldsymbol{x} + \sqrt{1-t}X)],\ (t, \boldsymbol{x}) \in [0, \infty) \times \mathbb{R}^2.$$

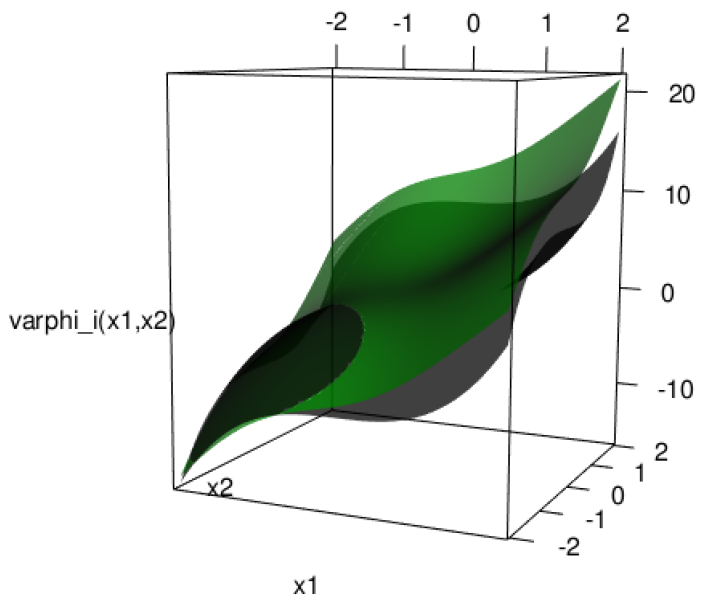


Figure 2.3: The numerical solution surface of the 2-dimensional G-heat equation with initial function $\varphi(x_1, x_2) = x_1^3 + x_2^3$

Then these results have been verified:

1. we have $\varphi_{n,n}(0) \approx \hat{\mathbb{E}}[\varphi(X)] = u(0, 0)$ which repeats the above remark;
2. by replacing $\varphi(X)$ with a shifted version $\varphi(\boldsymbol{x} + X)$, we directly get $\varphi_{n,n}(\boldsymbol{x}) \approx u(0, \boldsymbol{x})$;
3. In general, we have $\varphi_{k,n}(\boldsymbol{x}) \approx u(1 - \frac{k}{n}, \boldsymbol{x})$, for $k = 0, 1, \ldots, n$. In other words, the function in each iteration step has its connection with the solution of G-heat equation.

Figure 2.3 is the solution surface (where the black surface is $u(1, \boldsymbol{x}) = \varphi(\boldsymbol{x})$ and the green one is $u(0, \boldsymbol{x})$,) with $\varphi(\boldsymbol{x}) = x_1^3 + x_2^3$, the number of iteration steps $n = 10$, and

$$\mathcal{V} := \left\{ \mathbf{V} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} : \sigma_i \in [0.5, 1], i = 1, 2; \rho \in [-0.5, 0.5], \text{ such that } \mathbf{V} \in \mathbb{S}_d^+ \right\}.$$

I am still working on more details for this and also the Generalized Additive Model (GAM) with control variable method. Since in general, we don't have the exact solution for multi-dimensional G-heat equation, we will try to compare with the solutions produced by numerical PDE methods.

2.3.4 Assessment of the iterative method and future exploration

The strengths of this iterative method are well-worth mentioning:

1. It will work for any φ applicable in the nonlinear central limit theorem, including some irregular φ which makes $\hat{\mathbb{E}}[\varphi(X)]$ difficult to compute even by using classical numerical

PDE methods. This property has been verified by Theorem 2.3.4 and has been checked numerically for one dimension.

2. It will numerically solve the corresponding G-heat equation by not just giving one point or one path but actually directly providing the whole surface of $u(t, x)$, because the function in each iteration step has its connection with the $u(t, x)$ at one grid point t, as shown in Corollary 2.3.4.1;

3. It will give us a great visualization and intuition about how the solution surface of G-heat equation is evolved from the terminal function $\varphi(x)$ by looking at the procedure of iteration, and how the sublinear expectation of G-normal distribution is approached by iteratively maximizing the linear expectation of classical normal distribution. The inherent bridge or ladder here is the Semi-G-normal distribution which helps us climb from the stage of classical normal distribution to reach the stage of G-normal distribution. It partially fills in the long-existing thinking gap between them, since we have the inequality: $\hat{\mathbb{E}}[\varphi(X)] \geq \sup_{v\in[\underline{\sigma},\overline{\sigma}]} E[\varphi(N(0, v^2))]$ (for various φ, especially when φ is neither convex nor concave, the strict greater relation is much more frequent than equality), which, in some sense, strictly separated the G-normal distribution with the classical one and made us feel a little risky to connect them for a long time;

4. It can be naturally extended to higher dimension both in theory and algorithm, since we already have the established multi-dimensional distributions in G-framework, as well as the algorithm for computation of linear expectation of classical multivariate normal distribution. Then it can solve the corresponding multi-dimensional G-heat equation attached with covariance uncertainty;

However, for numerical practice, we still need to reflect on how to properly choose the splines to achieve better interpolation and extrapolation performance. There are several points well worth further exploring:

1. We may use design a set of nested grids for different iteration steps to avoid the extrapolation;

2. For the infinite domain, consider the truncated terminal functions first, then enlarge the truncation range;

3. For high dimension, start from bounded terminal functions (like the indicator functions);

4. We also need more delicate fitting model or spline design for different terminal functions (especially according to the properties of the viscosity solutions of G-heat equation);

5. Investigate this result from the PDE side (the viscosity solution of the G-heat equations can be expressed by a series of classical heat solution patches);

6. So far the numerical errors are better the theoretical errors (the CLT rate considers the general variance-uncertain sequence but here we use the semi-G-normal sequence), which inspires us to do more careful theoretical error approximation;

7. We may attempt to numerically adapt the method to high dimensional situation then we need to improve the optimization, grid setting and so on.

2.4 The sequential independence in the G-framework

In order to study the *simulation* and more statistical models (like the time series) in the G-framework, it is necessary for us to explore more about the (sequential) independence in the G-framework (see Definition 2.1.9).

Since the independence in the G-framework has the sequential order (the latter will be independent from the former), for convenience, we globally use the notation $X \dashrightarrow Y$, to denote the relation that "Y is independent from X". The convenience of this notation is to write $X \dashrightarrow Y \dashrightarrow Z$ to briefly express this kind of sequential independence: $X \dashrightarrow Y$ and $(X, Y) \dashrightarrow Z$. Recall the notion of i.i.d.(Definition 2.1.10),

$$X_1 \dashrightarrow X_2 \dashrightarrow \ldots \dashrightarrow X_n,$$

is defined as

$$(X_1, X_2, \ldots, X_i) \dashrightarrow X_{i+1},$$

for any $i = 1, 2, \ldots, n-1$.

Theorem 2.4.1 coming from Hu (2011) tells us the special properties brought by the mutual independence between two random variables in the G-framework.

Theorem 2.4.1 (Mutual independence, Hu (2011))**.** *$X \dashrightarrow Y$ and $Y \dashrightarrow X$ imply that X and Y are either linearly or maximally distributed.*

How to numerically understand the phenomenon shown in Theorem 2.4.1 is worth more exploration.

Question 2.4.1.1. Will $X \dashrightarrow Y$ and $Y \dashrightarrow Z$ imply $X \dashrightarrow Z$?

Remark 2.4.1.1. The answer of question Question 2.4.1.1 is NO. The counterexample is quite trivial and tedious. As shown in Theorem 2.4.1, it is possible that we have $X \dashrightarrow Y$ and $Y \dashrightarrow X$. In this case, if the answer is yes, we must have $X \dashrightarrow X$ which is an obvious contradiction. However, if we force Z and X to be non-identical (also excluding the scaled version like $Z = cX$ with a real constant c), from the knowledge of author, the answer is still unknown.

To prepare for the statistical theory regarding multivariate objects, we can explore the sequential independence for random vectors.

Theorem 2.4.2. *For any increasing subsequences $\{i_p\}_{p=1}^{k}$ and $\{j_q\}_{q=1}^{l}$ satisfying $1 \le i_1 < i_2 < \ldots < i_k \le n$ and $1 \le j_1 < j_2 < \ldots < j_l \le m$, we have the general result that*

$$(X_1, X_2, \ldots, X_n) \dashrightarrow (Y_1, Y_2, \ldots, Y_m) \implies (X_{i_1}, X_{i_2}, \ldots, X_{i_k}) \dashrightarrow (Y_{j_1}, Y_{j_2}, \ldots, Y_{j_l}).$$

Proof. For any test function $\varphi \in C_{l.Lip}(\mathbb{R}^{k+l})$, define another function $\psi \in C_{l.Lip}(\mathbb{R}^{n+m})$ which is a larger space by

$$\psi(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m) := \varphi(x_{i_1}, x_{i_2}, \ldots, x_{i_k}, y_{j_1}, y_{j_2}, \ldots, y_{j_l}),$$

then

$$\begin{aligned}
\hat{\mathbb{E}}[\varphi(X_{i_1}, X_{i_2}, \ldots, X_{i_k}, Y_{j_1}, Y_{j_2}, \ldots, Y_{j_l})] &= \hat{\mathbb{E}}[\psi(X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m)] \\
&= \hat{\mathbb{E}}[\hat{\mathbb{E}}[\psi(x_1, x_2, \ldots, x_n, Y_1, Y_2, \ldots, Y_m)]_{x_i = X_i, i=1,\ldots n}] \\
&= \hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi(x_{i_1}, x_{i_2}, \ldots, x_{i_k}, Y_{j_1}, Y_{j_2}, \ldots, Y_{j_l})]_{x_{i_p} = X_{i_p}, p=1,\ldots,k}].
\end{aligned}$$

□

Theorem 2.4.2 gives a very general and useful result rigorously telling us the sequential independence between two random vectors implies the independence between their sub-vectors.

Corollary 2.4.2.1. $(X, Y) \dashrightarrow Z \Longrightarrow X \dashrightarrow Z$ *and* $Y \dashrightarrow Z$.

Corollary 2.4.2.2. $(X_1, X_2) \dashrightarrow (X_3, X_4) \Longrightarrow (X_1, X_2) \dashrightarrow X_3$ *and* $(X_1, X_2) \dashrightarrow X_4$.

Corollary 2.4.2.3. *For any increasing subsequences* $\{i_p\}_{p=1}^k$ *satisfying* $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, *we have the result that*

$$X_1 \dashrightarrow X_2 \dashrightarrow \ldots \dashrightarrow X_n \implies X_{i_1} \dashrightarrow X_{i_2} \dashrightarrow \ldots \dashrightarrow X_{i_k}.$$

Proof. Our goal is to prove

$$(X_{i_1}, X_{i_2}, \ldots, X_{i_j-1}) \dashrightarrow X_{i_j},$$

holds for any $j = 2, \ldots, k$. For any $j = 2, \ldots, k$, since the definition of i.i.d. of the full sequence $\{X_i\}_{i=1}^n$ tells us that,

$$(X_1, X_2, \ldots, X_{i_j-1}) \dashrightarrow X_{i_j}.$$

From Theorem 2.4.2, we directly have the sequential independence for the sub-vectors:

$$(X_{i_1}, X_{i_2}, \ldots, X_{i_j-1}) \dashrightarrow X_{i_j}.$$

□

2.4.1 Independence regarding G-normal distributions

Theorem 2.4.3 (Peng (2010))**.** *Suppose* $X_1 \dashrightarrow X_2$ *and* $X_1 \overset{d}{=} X_2 \overset{d}{=} N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$, *for* $X := (X_1, X_2)$, *we have*

1. $\langle \boldsymbol{a}, X \rangle$ *is G-normal distributed for any* $\boldsymbol{a} \in \mathbb{R}^2$;
2. X *is not G-normal distributed.*

Proof. 1. $\langle \boldsymbol{a}, X \rangle = a_1 X_1 + a_2 X_2 \overset{\mathrm{d}}{=} \sqrt{a_1^2 + a_2^2} X_1 \sim N(0, (a_1^2 + a_2^2)[\underline{\sigma}^2, \overline{\sigma}^2])$.

2. The rigorous proof will use the theory of G-heat equation (more details can be found in Bayraktar and Munk (2015)). Here we give the intuitive reasoning to show why "we will fail to show X is G-normal". Consider $\bar{X} := (\bar{X}_1, \bar{X}_2)$ a independent copy of X, satisfying $\bar{X}_1 \dashrightarrow \bar{X}_2$ and $X \dashrightarrow \bar{X}$, namely, $(X_1, X_2) \dashrightarrow (\bar{X}_1, \bar{X}_2)$. This will imply $(X_1, X_2) \dashrightarrow \bar{X}_1$ (obviously from the above section) and $(X_1, X_2, \bar{X}_1) \dashrightarrow \bar{X}_2$: $\forall \varphi \in C_{l.Lip}(\mathbb{R}^4)$,

$$\begin{aligned}
\hat{\mathbb{E}}[\varphi(X_1, X_2, \bar{X}_1, \bar{X}_2)] &= \hat{\mathbb{E}}\left[\hat{\mathbb{E}}[\varphi(x_1, x_2, \bar{X}_1, \bar{X}_2)]_{x=X}\right] \\
&= \hat{\mathbb{E}}\left[\hat{\mathbb{E}}\left[\hat{\mathbb{E}}[\varphi(x_1, x_2, \bar{x}_1, \bar{X}_2)]_{\bar{x}_1=\bar{X}_1}\right]_{x=X}\right] \\
&\overset{(1)}{=} \hat{\mathbb{E}}\left[\hat{\mathbb{E}}[\varphi(x_1, x_2, \bar{x}_1, \bar{X}_2)]_{(x_1,x_2,\bar{x}_1)=(X_1,X_2,\bar{X}_1)}\right],
\end{aligned}$$

where (1) is due to $(X_1, X_2) \dashrightarrow \bar{X}_1$. In summary, we will have

$$X_1 \dashrightarrow X_2 \dashrightarrow \bar{X}_1 \dashrightarrow \bar{X}_2.$$

Consider $\forall \varphi \in C_{l.Lip}(\mathbb{R}^2)$,

$$\begin{aligned}\hat{\mathbb{E}}[\varphi(X+\bar{X})] &= \hat{\mathbb{E}}[\varphi(X_1+\bar{X}_1, X_2+\bar{X}_2)] \\ &= \hat{\mathbb{E}}\left[\hat{\mathbb{E}}[\varphi(x_1+\bar{X}_1, X_2+\bar{X}_2)]_{x_1=X_1}\right] \\ &= \begin{cases} \hat{\mathbb{E}}\left[\hat{\mathbb{E}}[\varphi(x_1+\bar{X}_1, x_2+\bar{X}_2)]_{x=X}\right] & \text{correct} \\ \hat{\mathbb{E}}\left[\hat{\mathbb{E}}[\varphi(x_1+\bar{x}_1, X_2+\bar{X}_2)]_{(x_1,\bar{x}_1)=(X_1,\bar{X}_1)}\right] & ? \end{cases},\end{aligned}$$

where the question mark comes from the fact that we do not have $(X_1, \bar{X}_1) \dashrightarrow (X_2, \bar{X}_2)$ which will indicate $\bar{X}_1 \dashrightarrow X_2$, reversing the existed relation $X_2 \dashrightarrow \bar{X}_1$.

□

Remark 2.4.3.1. The reasoning in the proof of Theorem 2.4.3 also leads us to a conclusion that

$$X_1 \dashrightarrow X_2 \dashrightarrow \bar{X}_1 \dashrightarrow \bar{X}_2,$$

is equivalent to

$$(X_1, X_2) \dashrightarrow (\bar{X}_1, \bar{X}_2), \text{ and } X_1 \dashrightarrow X_2, \bar{X}_1 \dashrightarrow \bar{X}_2.$$

Comment 2.4.3.1. Theorem 2.4.3 also tells us that it is hard to construct bivariate G-normal distribution from two univariate G-normal distributed random variables. Even under a straight forward sequential independence setting (mutual independence is impossible here according to Theorem 2.4.1), we are still strictly unable to achieve this, not to mention other linear transformations of the random vector (X_1, X_2) (more details about this difficulty can be found in Bayraktar and Munk (2015)).

Can we find a path to construct the multivariate G-normal distributed random vector from univariate objects? The answer is YES, given by the new concept Semi-G-normal distribution with its own independence setting (please see the following Section 2.4.2).

2.4.2 Independence regarding semi-G-normal distributions

In the following context, we will mainly consider two identically semi-G-normal distributed random variables $W(= ZY)$ and $\bar{W}(= \bar{Z}\bar{Y})$, with $Z \overset{\mathrm{d}}{=} \bar{Z} \sim M[\underline{\sigma}, \overline{\sigma}]$ and $Y \overset{\mathrm{d}}{=} \bar{Y} \sim N(0, [1, 1])$. Their relation will be more delicate because of the complexity involving the four objects: $Z, Y, \bar{Z}, \bar{Y}$.

Definition 2.4.4. There are several typical relations between W and $\bar{W}$:

1. $\bar{W}$ is *semi-sequentially independent* from W (denoted as $W \overset{\mathrm{S}}{\dashrightarrow} \bar{W}$) if :

$$Z \dashrightarrow \bar{Z} \dashrightarrow Y \dashrightarrow \bar{Y}; \tag{2.5}$$

2. $\bar{W}$ is *sequentially independent* from W (denoted as $W \dashrightarrow \bar{W}$) if:

$$Z \cdot Y \dashrightarrow \bar{Z} \cdot \bar{Y}; \tag{2.6}$$

3. $\bar{W}$ is *fully-sequentially independent* (denoted as $W \overset{\mathrm{F}}{\dashrightarrow} \bar{W}$) from W if :

$$Z \dashrightarrow Y \dashrightarrow \bar{Z} \dashrightarrow \bar{Y}. \tag{2.7}$$

Comment 2.4.4.1. Intuitively, the semi-sequential independence can be interpreted as "first setting all the parameters then drawing from the standard normal distributions to create the semi-G-normal sequence" (or "drawing from the linear normal distribution with preset parameters").

Remark 2.4.4.1. The semi-sequential independence ($W \overset{\text{S}}{\dashrightarrow} \bar{W}$) is equivalent with

$$(Z, \bar{Z}) \dashrightarrow (Y, \bar{Y}) \text{ and } Z \dashrightarrow \bar{Z} \text{ and } Y \dashrightarrow \bar{Y}.$$

Remark 2.4.4.2. Since we already have $Z \dashrightarrow Y$ and $\bar{Z} \dashrightarrow \bar{Y}$ by definition, the full-sequential independence ($W \overset{\text{F}}{\dashrightarrow} \bar{W}$) is equivalent with

$$(Z, Y) \dashrightarrow (\bar{Z}, \bar{Y}).$$

Remark 2.4.4.3. It is quite straightforward that

$$W \overset{\text{F}}{\dashrightarrow} \bar{W} \implies W \dashrightarrow \bar{W}.$$

However, $W \overset{\text{F}}{\dashrightarrow} \bar{W}$ (or $W \dashrightarrow \bar{W}$) does not necessarily imply $W \overset{\text{S}}{\dashrightarrow} \bar{W}$ since Equation (2.5) actually reverses the order of independence for Y and $\bar{Z}$ in Equation (2.7).

Theorem 2.4.5. *For any $\bar{W}$ satisfying $\bar{W} \overset{d}{=} W$ and $W \overset{S}{\dashrightarrow} \bar{W}$, we have*

$$W + \bar{W} \overset{d}{=} \sqrt{2}W.$$

Proof. Since $W = ZY$ and $\bar{W} = \bar{Z}\bar{Y}$, for any $\varphi \in C_{l.Lip}$, consider

$$\begin{aligned}
\hat{\mathbb{E}}[\varphi(ZY + \bar{Z}\bar{Y})] &= \hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi(zY + \bar{z}\bar{Y})]_{z=Z,\,\bar{z}=\bar{Z}}] \\
&= \hat{\mathbb{E}}[\underbrace{\hat{\mathbb{E}}[\varphi(\sqrt{z^2 + \bar{z}^2}Y)]}_{H\left(\sqrt{z^2+\bar{z}^2}\right)}{}_{z=Z,\,\bar{z}=\bar{Z}}] \\
&= \hat{\mathbb{E}}\left[H\left(\sqrt{Z^2 + \bar{Z}^2}\right)\right] \\
&= \hat{\mathbb{E}}\left[\hat{\mathbb{E}}[H(\sqrt{z + \bar{Z}})]_{z=Z}\right] \\
&= \max_{w\in[\underline{\sigma},\overline{\sigma}]} \max_{v\in[\underline{\sigma},\overline{\sigma}]} H(\sqrt{v^2 + w^2}),
\end{aligned}$$

where $H(v) := \hat{\mathbb{E}}[\varphi(vY)]$ can be proved to be in $C_{l.Lip}$ based on $\varphi \in C_{l.Lip}$.

Since $\{\sqrt{v^2 + w^2};\ v, w \in [\underline{\sigma}, \overline{\sigma}]\} = [\sqrt{2}\underline{\sigma}, \sqrt{2}\overline{\sigma}] = \{\sqrt{2}z;\ z \in [\underline{\sigma}, \overline{\sigma}]\}$ (for this step, the independence may not be necessary since we only need z and $\bar{z}$ will have mass point at (a, a) and (b, b)), we have

$$\begin{aligned}
\hat{\mathbb{E}}[\varphi(ZY + \bar{Z}\bar{Y})] &= \max_{w\in[\underline{\sigma},\overline{\sigma}]} \max_{v\in[\underline{\sigma},\overline{\sigma}]} H(\sqrt{v^2 + w^2}) \\
&= \max_{z\in[\underline{\sigma},\overline{\sigma}]} H(\sqrt{2}z) \\
&= \hat{\mathbb{E}}[H(\sqrt{2}Z)] \\
&= \hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi(\sqrt{2}zY)]_{z=Z}] \\
&= \hat{\mathbb{E}}[\varphi(\sqrt{2}ZY)].
\end{aligned}$$

Hence, $W + \bar{W} \overset{\text{d}}{=} \sqrt{2}W$. □

The following definition extends the relations in Definition 2.4.4 to n variables.

Definition 2.4.6. For a sequence of semi-G-normal distributed random variables $\{W_i\}_{i=1}^n (= \{Z_iY_i\}_{i=1}^n)$, we also have the corresponding typical relations:

1. $\{W_i\}_{i=1}^n$ are *semi-sequentially independent* (denoted as $W_1 \overset{\mathrm{S}}{\dashrightarrow} W_2 \overset{\mathrm{S}}{\dashrightarrow} \ldots \overset{\mathrm{S}}{\dashrightarrow} W_n$) if :

$$Z_1 \dashrightarrow Z_2 \dashrightarrow \ldots \dashrightarrow Z_n \dashrightarrow Y_1 \dashrightarrow Y_2 \dashrightarrow \ldots \dashrightarrow Y_n; \tag{2.8}$$

2. $\{W_i\}_{i=1}^n$ are *sequentially independent* (denoted as $W_1 \dashrightarrow W_2 \dashrightarrow \ldots \dashrightarrow W_n$) if:

$$Z_1Y_1 \dashrightarrow Z_2Y_2 \dashrightarrow \ldots \dashrightarrow Z_nY_n; \tag{2.9}$$

3. $\{W_i\}_{i=1}^n$ are *fully-sequentially independent* (denoted as $W_1 \overset{\mathrm{F}}{\dashrightarrow} W_2 \overset{\mathrm{F}}{\dashrightarrow} \ldots \overset{\mathrm{F}}{\dashrightarrow} W_n$) if:

$$Z_1 \dashrightarrow Y_1 \dashrightarrow Z_2 \dashrightarrow Y_2 \dashrightarrow \ldots \dashrightarrow Z_n \dashrightarrow Y_n. \tag{2.10}$$

The semi-sequential independence will preserve some of the intuitive properties we have in classical situation.

Theorem 2.4.7. *For a sequence of semi-G-normal distributed random variables* $\{W_i\}_{i=1}^n$ *satisfying* $W_i \overset{d}{=} W_{i+1}$, $i = 1, 2, \ldots, n-1$ *and*

$$W_1 \overset{S}{\dashrightarrow} W_2 \overset{S}{\dashrightarrow} \ldots \overset{S}{\dashrightarrow} W_n,$$

we have

$$\sum_{i=1}^n W_i \overset{d}{=} \sqrt{n} W_1.$$

Lemma 2.4.8. *For a sequence of (nonlinearly) i.i.d. random variables* $\{Y_i\}_{i=1}^n \sim N(0, [1, 1])$ *(namely,* $Y_1 \dashrightarrow Y_2 \dashrightarrow \ldots \dashrightarrow Y_n$*), we have*

$$(Y_1, Y_2, \ldots, Y_n)^T \sim N(\mathbf{0}, \mathbf{I}_n^2),$$

where $\mathbf{I}_n$ *is the* $n \times n$ *identity matrix. It also further implies that* $\{Y_i\}_{i=1}^n$ *are actually linearly independent.*

Proof. Denote the probability density function of classical multivariate normal distribution $N(\mathbf{0}, \mathbf{I}_n^2)$ as ϕ_n. Inductively speaking, starting from $n = 1$, it is not hard to show that $Y_1 \overset{\mathrm{d}}{=} N(0, 1)$ since the G-heat equation is reduced to the classical one when $\underline{\sigma}$ and $\overline{\sigma}$ coincide. Suppose for $n = k(\geq 1)$, we already have $\boldsymbol{Y}_k := (Y_1, Y_2, \ldots, Y_k)^T \sim N(\mathbf{0}, \mathbf{I}_k^2)$. When $n = k + 1$, first of all, $\boldsymbol{Y}_k \dashrightarrow Y_{k+1}$, then

$$\hat{\mathbb{E}}[\varphi(Y_1, Y_2, \ldots, Y_{k+1})] = \hat{\mathbb{E}}[\varphi(\boldsymbol{Y}_k, Y_{k+1})] = \hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi(\boldsymbol{y}_k, Y_{k+1})]_{\boldsymbol{y}_k = \boldsymbol{Y}_k}] =: \hat{\mathbb{E}}[G(\boldsymbol{Y}_k)],$$

where

$$G(\boldsymbol{y}_k) = E[\varphi(\boldsymbol{y}_k, Y_{k+1})] = \int \varphi(\boldsymbol{y}_k, y_{k+1}) f_{Y_{k+1}}(y_{k+1})\, \mathrm{d}y_{k+1},$$

with $f_{Y_{k+1}} = \phi_1$. Then

$$\begin{aligned}\hat{\mathbb{E}}[\varphi(\boldsymbol{Y}_k, Y_{k+1})] &= \hat{\mathbb{E}}[G(\boldsymbol{Y}_k)] = E[G(\boldsymbol{Y}_k)] \\ &= \int_{\mathbb{R}^k} G(\boldsymbol{y}_k) f_{\boldsymbol{Y}_k}(\boldsymbol{y}_k)\, \mathrm{d}\boldsymbol{y}_k \\ &= \int_{\mathbb{R}^{k+1}} \varphi(\boldsymbol{y}_k, y_{k+1}) f_{Y_{k+1}}(y_{k+1}) f_{\boldsymbol{Y}_k}(\boldsymbol{y}_k)\, \mathrm{d}y_{k+1}\, \mathrm{d}\boldsymbol{y}_k \\ &= \int_{\mathbb{R}^{k+1}} \varphi(\boldsymbol{y}_k, y_{k+1}) \phi_1(y_{k+1}) \phi_k(\boldsymbol{y}_k)\, \mathrm{d}y_{k+1}\, \mathrm{d}\boldsymbol{y}_k \\ &= E[\varphi(N(0, \mathbf{I}_{k+1}^2))] = \hat{\mathbb{E}}[\varphi(N(0, \mathbf{I}_{k+1}^2))],\end{aligned}$$

where we apply the independence between $N(0,1)$ and $N(\mathbf{0}, \mathbf{I}_k^2)$ to produce $N(\mathbf{0}, \mathbf{I}_{k+1}^2)$ (namely, $\phi_1 \phi_k = \phi_{k+1}$). Therefore, $(Y_1, Y_2, \ldots, Y_{k+1}) \overset{\mathrm{d}}{=} N(\mathbf{0}, \mathbf{I}_{k+1}^2)$. Finally, we have

$$(Y_1, Y_2, \ldots, Y_n)^T \sim N(0, \mathbf{I}_n^2),$$

in which the covariance matrix with zero off-diagonal entries indicates that $\{Y_i\}_{i=1}^n$ are linearly independent. □

Theorem 2.4.9. *For a sequence of semi-G-normal distributed random variables $\{W_i\}_{i=1}^n$, satisfying $W_i \sim \hat{N}(0, [\underline{\sigma}_i^2, \overline{\sigma}_i^2])$ for $i = 1, 2, \ldots, n$, and*

$$W_1 \overset{S}{\dashrightarrow} W_2 \overset{S}{\dashrightarrow} \ldots \overset{S}{\dashrightarrow} W_n,$$

we have

$$(W_1, W_2, \ldots, W_n)^T \sim \hat{N}(\mathbf{0}, \mathcal{V}),$$

where $\mathcal{V} \subset \mathbb{S}_d^+$ is the uncertainty set of covariance matrices defined as

$$\mathcal{V} := \left\{ \mathbf{V} = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix} : \sigma_i^2 \in [\underline{\sigma}_i^2, \overline{\sigma}_i^2], i = 1, 2, \ldots, n, \text{ such that } \mathbf{V} \in \mathbb{S}_d^+ \right\},$$

and

$$\mathcal{V}^{1/2} := \left\{ \mathbf{V}^{1/2} : \mathbf{V} \in \mathcal{V} \right\}.$$

Proof. Let $\boldsymbol{W} := (W_1, W_2, \ldots, W_n)^T$. Then

$$\boldsymbol{W} = (Z_1 Y_1, Z_2 Y_2, \ldots, Z_n Y_n)^T = \mathbf{Z}\boldsymbol{Y},$$

where $\mathbf{Z} = \mathbf{Z}_{n\times n} := \mathrm{diag}(Z_1, Z_2, \ldots, Z_n)$ and $\boldsymbol{Y} = \boldsymbol{Y}_{n\times 1} = (Y_1, Y_2, \ldots, Y_n)^T$. The semi-sequential independence implies that $\mathbf{Z}_{n\times n} \dashrightarrow \boldsymbol{Y}_{n\times 1}$. Also from the sequential independence of Z_i part, we know that for any $H \in C_{l.Lip}(\mathbb{R}^{n\times n})$:

$$\hat{\mathbb{E}}[H(\mathbf{Z})] = \max_{z_i \in [\underline{\sigma}_i, \overline{\sigma}_i], i=1,\ldots,n} \hat{\mathbb{E}}[H(\mathrm{diag}(z_1, \ldots, z_n))] = \max_{\mathbf{V}^{1/2} \in \mathcal{V}^{1/2}} \hat{\mathbb{E}}[H(\mathcal{V}^{1/2})].$$

In other words, $\mathbf{Z} \sim M(\mathcal{V}^{1/2})$. Meanwhile, the Lemma 2.4.8 shows that $\boldsymbol{Y} \sim N(\mathbf{0}, \mathbf{I}_n^2)$. Therefore, by the definition of d-dimensional semi-G-normal distribution,

$$\boldsymbol{W} = \mathbf{Z}_{n\times n}\boldsymbol{Y} \sim \hat{N}(\mathbf{0}, \mathcal{V}).$$

□

Then we can use a sequence of $\boldsymbol{W} \sim \hat{N}(\mathbf{0}, \mathcal{V})$ to approach the multivariate G-normal $N(\mathbf{0}, \mathcal{V})$ (defined in Definition 2.1.16) by nonlinear CLT. First, we give the definition of the G-normal distribution $N(\mathbf{0}, \mathcal{V})$. Let $\mathbb{S}_d$ denote the set of all real-valued $d \times d$ symmetric matrices.

Theorem 2.4.10. *Consider a sequence of i.i.d.* $\{W_i\}_{i=1}^{\infty} \sim \hat{N}(0, \mathcal{V})$ *. Let X be a G-normal distributed random vector following $N(0, \mathcal{V})$. Then for any $\varphi \in C_{l.Lip}$,*

$$\lim_{n\to\infty} \hat{\mathbb{E}}[\varphi(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i)] = \hat{\mathbb{E}}[\varphi(X)].$$

Proof. This is a direct result from the nonlinear central limit theorem (Theorem 2.1.18). We only need to check the validity of the conditions.

First of all, The sequence $\{W_i\}_{i=1}^{\infty}$ has the mean-certainty $\hat{\mathbb{E}}[X_1] = -\hat{\mathbb{E}}[-X_1] = \mathbf{0}$.

Secondly, the distribution of X is characterized by the function $G(\mathbf{A}) = \frac{1}{2} \sup_{\mathbf{V}\in\mathcal{V}} \mathrm{tr}[\mathbf{AV}]$. We only need to prove that $G(\mathbf{A}) = \frac{1}{2}\hat{\mathbb{E}}[\langle \mathbf{A}W_1, W_1\rangle]$ for any $\mathbf{A} \in \mathbb{S}_d$. By the representation of semi-G-normal distribution, letting $Z \sim N(0, \mathbf{V})$, we have

$$\begin{aligned}
\hat{\mathbb{E}}[\langle \mathbf{A}W_1, W_1\rangle] &= \sup_{\mathbf{V}\in\mathcal{V}} E[\langle \mathbf{A}Z, Z\rangle] = \sup_{\mathbf{V}\in\mathcal{V}} E[(\mathbf{A}Z)^T Z] \\
&= \sup_{\mathbf{V}\in\mathcal{V}} E[Z^T \mathbf{A} Z] = \sup_{\mathbf{V}\in\mathcal{V}} E[\mathrm{tr}[Z^T \mathbf{A} Z]] \\
&= \sup_{\mathbf{V}\in\mathcal{V}} E[\mathrm{tr}[\mathbf{A}ZZ^T]] = \sup_{\mathbf{V}\in\mathcal{V}} \mathrm{tr}[E[\mathbf{A}ZZ^T]] \\
&= \sup_{\mathbf{V}\in\mathcal{V}} \mathrm{tr}[\mathbf{A}E[ZZ^T]] = \sup_{\mathbf{V}\in\mathcal{V}} \mathrm{tr}[\mathbf{AV}].
\end{aligned}$$

Therefore, by applying the nonlinear CLT, we prove the limiting result. □

Since the semi-G-normal distribution preserves the multivariate property of the linear normal distribution, we can further play with it and produce more counterpart properties like the linear transformation of $\hat{N}(\mathbf{0}, \mathcal{V})$.

Corollary 2.4.10.1. *Let $\boldsymbol{W}_{n\times 1} \sim \hat{N}(\mathbf{0}, \mathcal{V})$. For any constant matrix $\mathbf{A} \in \mathbb{R}^{r\times n}$ with $r \le n$, we have*

$$\mathbf{A}\boldsymbol{W} \sim \hat{N}(\mathbf{0}, \mathbf{A}\mathcal{V}\mathbf{A}^T),$$

where

$$\mathbf{A}\mathcal{V}\mathbf{A}^T := \left\{\mathbf{A}\mathbf{V}\mathbf{A}^T : \mathbf{V} \in \mathcal{V}\right\} \subset \mathbb{R}^{r\times r}.$$

Proof. First of all, $\mathbf{A}_{r\times n}\boldsymbol{W}_{n\times 1} = \mathbf{A}_{r\times n}\mathbf{Z}_{n\times n}\boldsymbol{Y}_{n\times 1}$. with $\mathbf{Z} \sim M(\mathcal{V}^{1/2})$. For any $H \in C_{l.Lip}(\mathbb{R}^{r\times n})$, we have

$$\hat{\mathbb{E}}[H(\mathbf{AZ})] = \max_{\mathbf{V}^{1/2}\in\mathcal{V}^{1/2}} E[H(\mathbf{A}\mathbf{V}^{1/2})] = \max_{\mathbf{B}\in\mathbf{A}\mathcal{V}^{1/2}} E[H(\mathbf{B})],$$

so $\mathbf{AZ} \sim M(\mathbf{A}\mathcal{V}^{1/2})$, which can be treated as the scaling property for the $n \times n$-dimensional maximal distribution. Therefore,

$$\mathbf{A}\boldsymbol{W} \overset{\mathrm{d}}{=} M(\mathbf{A}\mathcal{V}^{1/2})N(\mathbf{0}, \mathbf{I}_n^2) \sim \hat{N}(\mathbf{0}, \mathcal{V}'),$$

where

$$\begin{aligned}\mathcal{V}' &:= \left\{\mathbf{B}\mathbf{B}^T : \mathbf{B} \in \mathbf{A}\mathcal{V}^{1/2}\right\} \\ &= \left\{(\mathbf{A}\mathbf{V}^{1/2})(\mathbf{A}\mathbf{V}^{1/2})^T : \mathbf{V}^{1/2} \in \mathcal{V}^{1/2}\right\} \\ &= \left\{\mathbf{A}\mathbf{V}\mathbf{A}^T : \mathbf{V} \in \mathcal{V}\right\}.\end{aligned}$$

In other words,

$$\mathbf{A}\boldsymbol{W} \sim \hat{N}(\mathbf{0}, \mathbf{A}\mathcal{V}\mathbf{A}^T).$$

□

Next diagram shows the relations among linear, semi-G- and G-normal distributions.

$$\begin{array}{lccccc}
\text{multiple dimension:} & N(\mathbf{0}, \mathbf{I}_n^2) & \xrightarrow{M(\mathcal{V}^{\frac{1}{2}})} & \hat{N}(\mathbf{0}, \mathcal{V}) & \xrightarrow[\dashrightarrow]{\text{nonlinear CLT}} & N(\mathbf{0}, \mathcal{V}) \\
& \uparrow \text{linearly Ind.} & & \uparrow S & & \updownarrow \\
\text{one dimension:} & N(0,1) & \xrightarrow{M[\underline{\sigma}, \overline{\sigma}]} & \hat{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2]) & \xrightarrow[\dashrightarrow]{\text{nonlinear CLT}} & N(0, [\underline{\sigma}^2, \overline{\sigma}^2])
\end{array}$$

As we know in Theorem 2.4.3, it is not clear how to construct *multivariate* G-normal distribution from *univariate* objects. For two identically one-dimensional G-normal distributed random variables X_1 and X_2 with $X_1 \dashrightarrow X_2$, we know that (X_1, X_2) does not follow the bivariate G-normal distribution. Since the semi-G-normal distribution has smaller uncertainty set, it partially preserves the properties of the classical normal distribution and with the variance uncertainty, it is also able to approach the multivariate G-normal distribution. Therefore, as shown in the diagram above, we can start from the *univariate* objects (semi-G-normal distribution), and construct its multivariate version under semi-sequential independence, then approach the *multivariate* G-normal distribution, which gives us a feasible way from "univariate" to "multivariate" and becomes another reason we want to study the semi-G-normal distribution.

2.4.3 Future attempts regarding G-Brownian motion

For identically distributed, sequentially independent semi-G-normal sequence $\{W_i\}_{i=1}^n$ each of which follows $\hat{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$, in the iterative approximation part, we have already proved that, letting $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$, for any $p \in (0, 1]$, we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor np \rfloor} W_i \xrightarrow{\mathrm{d}} \sqrt{p} X \overset{\mathrm{d}}{=} B_p,$$

where B is the G-Brownian motion (to be explained in the preliminaries of my future work). Furthermore, for $0 < p_1 < p_2 \leq 1$, let $\overline{W}_m := \frac{1}{\sqrt{n}} \sum_{i=1}^m W_i$, then we should have $\overline{W}_{\lfloor np_1 \rfloor} \dashrightarrow \overline{W}_{\lfloor np_2 \rfloor} - \overline{W}_{\lfloor np_1 \rfloor}$ and

$$(\overline{W}_{\lfloor np_1 \rfloor}, \overline{W}_{\lfloor np_2 \rfloor} - \overline{W}_{\lfloor np_1 \rfloor}) \xrightarrow{\mathrm{d}} (B_{p_1}, B_{p_2} - B_{p_1}).$$

In Chapter 2, we have explored the theory of the distributions and independence in G-framework and built up the substructure to better deal with these objects with uncertainty. Then we will come to the side of dataset including estimation (Chapter 3) and simulation (Chapter 4) which will apply both the theory and intuition we have learned in Chapter 2.

Chapter 3

The Estimation of Variance Uncertainty

3.1 The necessity of the statistical methods in the G-framework

In order to give readers some taste about the background, we will start from an inspiring practical example of the so-called *black-box situation*:

Suppose a company (the data provider) provides us (the data analysts) with a time-index dataset $W_1, W_2, \ldots, W_N$ ($N = 10^4$), sampling from the random variable W (with *known* mean parameter equal to zero), without offering any further information about the structure behind the dataset, the generation procedure is a *black box* for us. (This situation might be unlikely in practice but data analytics often fail to obtain critical information about how the data were generated.) Simply providing the dataset, they ask us to estimate the expected value of W^2.

Since we do not have any background information about the structure of the dataset (except the mean parameter is known as zero), we can only start from some initial visualization of this dataset. In addition, at this stage, it is not sensible to directly apply statistical tests usually requiring assumptions on the probabilistic structure (like the distribution or independence) of the dataset which still need investigation. Therefore, the following analysis is not a complete statistical investigation and is only aimed at giving readers some feelings about the dataset. Most statements can only be treated as "guesses" rather than strict conclusions, which are definitely required to have further testing.

Figure 3.1 (the time-index plot of the dataset) and Figure 3.2 (the values of the autocorrelation function against different lags) indicate that the data points are stationary and uncorrelated, which is consistent with the pattern of white noise.

We have mentioned the notion called the *variance-uncertainty* several times in previous chapters. We intend to show that this dataset actually has variance-uncertainty. To better see this, one graphical tool is to group the dataset into smaller groups then draw the box-plot of each successive group. In Figure 3.3, the box-plot of the dataset, we can see that the means of groups are relatively stable, which is consistent with the known information that the mean parameter of W is zero (if we generate a dataset with certain zero mean, its box-plot should have a similar pattern), but there exists variability in the spread-out of different groups, which might come from the inconsistent variance parameter of the dataset. This observation can be further illustrated by Figure 3.4, the box-plot of the square of the dataset, where the variability is enlarged by the square transformation. However, we try not to read too much into these initial

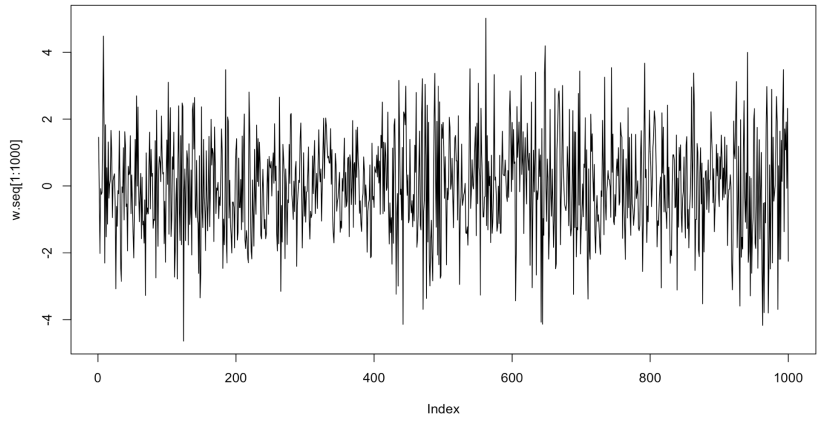


Figure 3.1: The time-index plot of the first 1000 points of the dataset

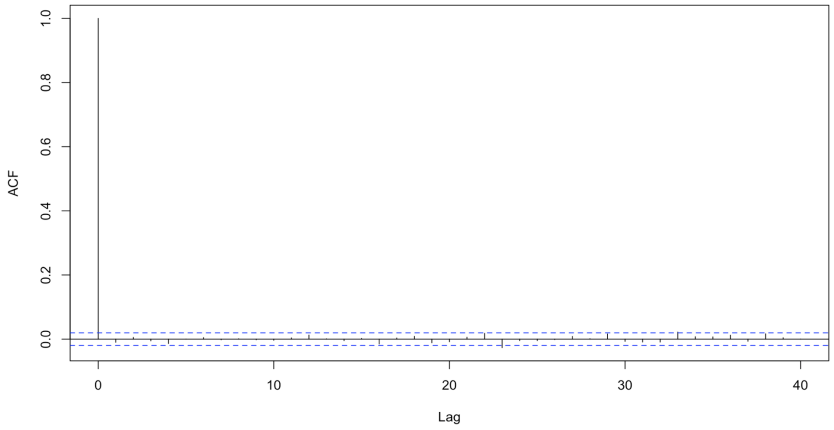


Figure 3.2: The autocorrelation function plot of the dataset to investigate possible linear autocorrelation

visualization plots before doing relatively strict statistical tests on the dataset.

At least, the observation (or guess) of the existence of inconsistent variance tells us the possibility that the data have the variance-uncertainty. The reason we call it "uncertainty"' is that we do not have any (prior) information about how the variance parameter changes as time goes along; otherwise, we may have a sensible reason to use a classical stochastic model (corresponding to one single probability measure) to describe the changing variance; in this sense, there is almost no uncertainty (corresponding to a set of more than one probability measures) for us.

Another potential evidence of the existence of variance-uncertainty is to show the failure of the classical law of large number under the i.i.d. assumption. To estimate the expected value of W^2, one natural way is to compute the linear expectation $E[W^2]$, which can be estimated by the sample variance $\frac{1}{n}\sum_{i=1}^{n} W_i^2$, since according to the classical law of large numbers, if its assumptions hold here, the sample mean should converge to $E[W^2]$ as n increases. However, Figure 3.5 tells us a different story.

When plotting the sample mean of W^2 against different sample size, we can see that as

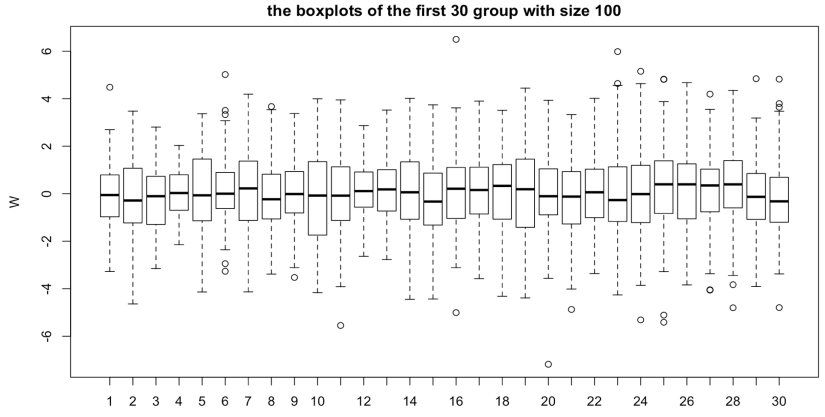


Figure 3.3: The boxplot of the first 30 groups with size 100 for the dataset

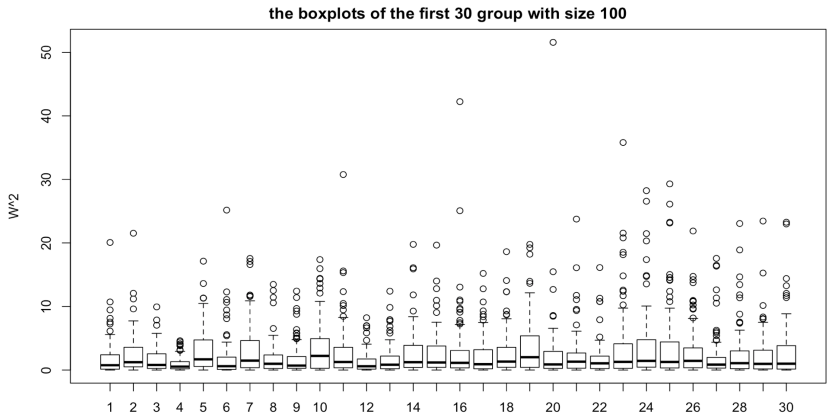


Figure 3.4: The boxplot of the first 30 groups with size 100 for the square of the dataset

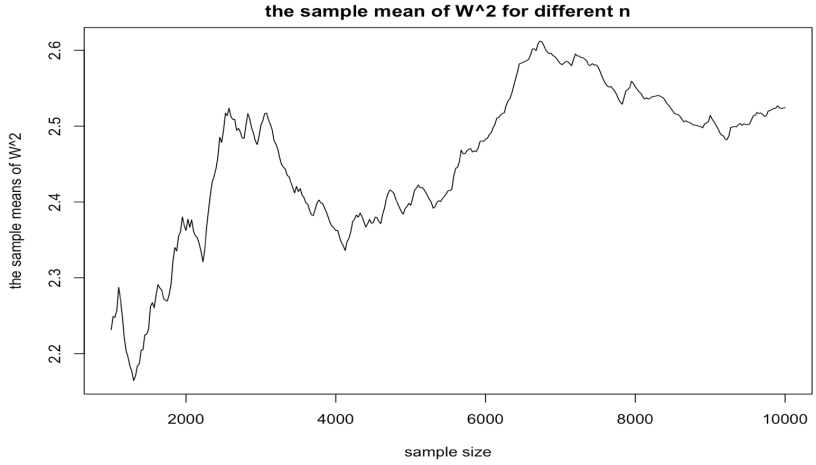


Figure 3.5: The sample variances against increasing sample size n

the sample size increases from 5×10^3 to 10^4 (which already reaches the size of the whole dataset), the sample mean seems not to converge to some value (at least, the sample mean does not really fluctuate around one certain value), instead, it is still fluctuating within some range (rises up to 2.8 at $n = 7 \times 10^3$, then drops down to 2.5 at $n = 9 \times 10^3$, then tends to increase again). From this draft plot, we can hardly find a converging pattern for large n. If this sequence of random variables is independent, one possible explanation of the failure of LLN is that random variables are not identically distributed and the dataset has distribution uncertainty, in particular, the variance-uncertainty.

Meanwhile, we do not know any information about the structure behind the data generation so we can hardly apply a certain distribution model to describe W.

For readers' interest, the above-mentioned dataset $W_1, W_2, \ldots, W_N$ actually comes from the pseudo simulation of $\hat{N}(0, [1, 4])$ (more details can be found in Chapter 4).

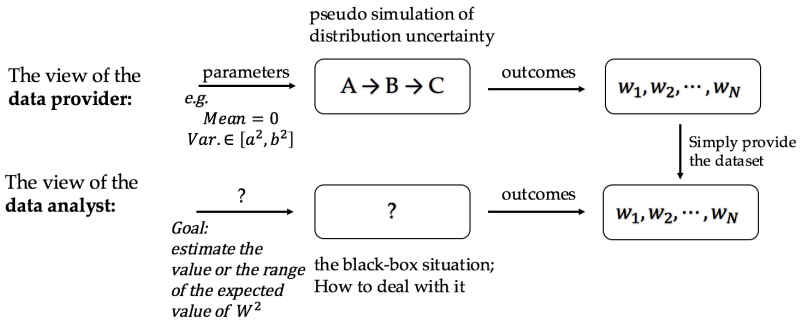


Figure 3.6: Black-box situation: the data provider vs the data analyst

Figure 3.6 is the big picture of the current situation and also the structure of our remaining content. From the view of the data provider, this dataset only has *the pseudo distribution uncertainty* but from the prospective of the data analyst, the dataset has *the true distribution uncertainty*.

3.2 Improved max-mean estimation in practice

If we directly apply the traditional statistical methods based on linear expectation, the assumptions are usually invalid here. What we know about the dataset is that it has zero mean and variance uncertainty, reminding us of the conditions attached to the CLT (Theorem 2.1.18 for $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i$) and LLN (Theorem 2.1.17 for $\frac{1}{n} \sum_{i=1}^{n} W_i^2$) in sublinear expectation space. We already have the max-mean estimation for maximal distribution (see Jin and Peng (2016) for more details).

First choose a group size n. Without loss of generality, suppose n is a divisor of N such that we have an integer $m := N/n$ as the number of groups. Let $Y := W^2$ and $Y_l := W_l^2$ with $l = 1, 2, \ldots, N$ and group the dataset as follows:

$$\left(Y_{ij}\right)_{m\times n}=\begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1n} \\ Y_{21} & Y_{22} & \cdots & Y_{2n} \\ \vdots & \vdots & & \vdots \\ Y_{m1} & Y_{m2} & \cdots & Y_{mn} \end{pmatrix},$$

where $Y_{ij} \coloneqq Y_{l(i,j)}$ with $l(i,j)=(i-1)n+j$ for $i=1,2,\ldots,m$ and $j=1,2,\ldots,n$.

For each group, $\bar{Y}_i \coloneqq \frac{1}{n}\sum_{j=1}^{n} Y_{ij}$, $i=1,2,\ldots,m$ will converge to $M[\underline{\sigma}^2,\overline{\sigma}^2]$ as n increases; meanwhile, the minimum (maximum) of a sequence of $M[\underline{\sigma}^2,\overline{\sigma}^2]$ will approach its bounds as the sample size increases.

The idea can be explained in this way

$$\min_{1\le i\le m}\frac{1}{n}\sum_{j=1}^{n} Y_{ij} \xrightarrow[\text{LLN}]{n\to\infty} \min_{1\le i\le m} M[\underline{\sigma}^2,\overline{\sigma}^2] \xrightarrow{m\to\infty} \underline{\sigma}^2,$$

and

$$\max_{1\le i\le m}\frac{1}{n}\sum_{j=1}^{n} Y_{ij} \xrightarrow[\text{LLN}]{n\to\infty} \max_{1\le i\le m} M[\underline{\sigma}^2,\overline{\sigma}^2] \xrightarrow{m\to\infty} \overline{\sigma}^2.$$

The problem is that we have the constraint $N=m\cdot n$ from the limited data. If the group size n is too small, the group average cannot really achieve its convergence to the maximal distribution; conversely, if the group size is too large, although we have the group averages close to $M[\underline{\sigma}^2,\overline{\sigma}^2]$, they are only a small sample of the maximal distribution and we can hardly expect them to reach the boundaries.

3.2.1 How to decide the pair (m,n)

If we force m and n to increase at the same time, for instance, let $m=n$, then make n increase from 1 to $[\sqrt{N}]$ and for each n, get the pair of max-mean estimations, shown in Figure 3.7 (to better show the further trend, we ask the data provider to artificially enlarge the dataset using the same simulation scheme), from which we can see that the estimated interval is shrinking as n increases and appears significant *undercoverage* (compared with the true one drawn as the two horizontal dashed lines) when n is too large (greater than around 400), indicating the poor performance of this decision rule of n, although it seems reasonable in theory (the two-stage convergence in Section Section 3.2). In addition, when n is far from reaching $\sqrt{N}$, this rule actually wastes a lot of information of the dataset.

If we fix a big n, to make the average have the distribution close enough to the $M[\underline{\sigma}^2,\overline{\sigma}^2]$, and let $m=[N/n]$. We hope that this m will also be generous so that the averages can vary much in the interval to touch the extremes. However, we do not know how big n should be at first so this plan does not really push forward in the decision (e.g. when n is close to N, the pair of estimations will almost meet each other at some point in $[\underline{\sigma}^2,\overline{\sigma}^2]$).

In order to choose the optimal pair (m,n) (or design a data-driven stopping rule for n) to do the trade-off and make the most of information then narrow down the variance interval, let us think more about what really happens when we increase the group size n. This is only *the intuition behind our methods*, not a rigorous proof (for which we are still struggling on it and will need more theories to back up).

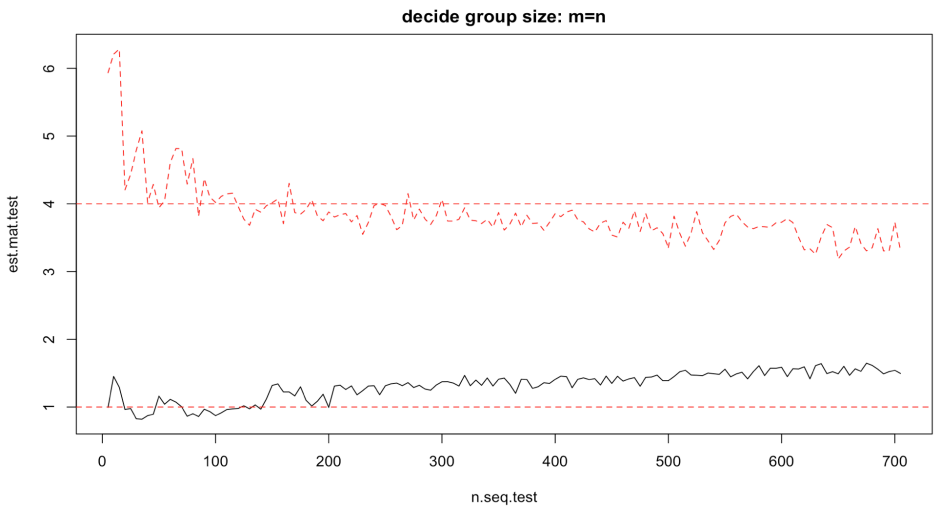


Figure 3.7: The change of max-mean estimations when $n = 1, 2, \ldots, \sqrt{N}$.

For a given group size n, define the range between the groups $[L(n), R(n)]$ as follows:

$$[L(n), R(n)] := [\min_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n} Y_{ij},\ \max_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n} Y_{ij}].$$

Meanwhile, we know the range within each group is, for $i = 1, 2, \ldots, m$,

$$[\min_{1\le j\le n} Y_{ij},\ \max_{1\le j\le n} Y_{ij}].$$

Let $Y_{i(1)} := \min_{1\le j\le n} Y_{ij}$ and $Y_{i(n)} := \max_{1\le j\le n} Y_{ij}$, we will need some functions $\underline{H} : \mathbb{R}^m \to \mathbb{R}$ and $\overline{H} : \mathbb{R}^m \to \mathbb{R}$, to describe the overall level of the range within each group $[l(n), r(n)]$ in this way:

$$[l(n), r(n)] := [\underline{H}(Y_{1(1)}, Y_{2(1)}, \ldots, Y_{m(1)}),\ \overline{H}(Y_{1(n)}, Y_{2(n)}, \ldots, Y_{m(n)})].$$

For the range between groups, when $n = 1$, we have

$$[L(1), R(1)] = [\min_{1\le i\le N} Y_i,\ \max_{1\le i\le N} Y_i],$$

which is the largest range and should cover the $[\underline{\sigma}^2, \overline{\sigma}^2]$. As n increases, since we have the inequality like

$$\min\{c, d, e, f\} \le \min\{\frac{c+d}{2}, \frac{e+f}{2}\} \le \max\{\frac{c+d}{2}, \frac{e+f}{2}\} \le \max\{c, d, e, f\},$$

the range between the groups will shrink, i.e. $R(n) - L(n)$ will decrease.

For the range within each group, for $n = 1$, since we have $Y_{i(1)} = Y_{i(n)}$, we want to choose proper functions to make $l(1) = r(1)$ and they also falls around the center of $[\underline{\sigma}^2, \overline{\sigma}^2]$.When n increases, since the group size is enlarged, the length of each range $Y_{i(n)} - Y_{i(1)}$ is expected

to increase for most i. Hence, for the overall level of the range within the group, $r(n) - l(n)$ is expected to increase. By setting a proper function H, for n not very large, usually we want to achieve the following subset relation,

$$[l(n), r(n)] \subset [\underline{\sigma}^2, \overline{\sigma}^2] \subset [L(n), R(n)], \tag{3.1}$$

which implies that we want the $[l(n),\ r(n)]$ and $[L(n),\ R(n)]$ to squeeze the $[\underline{\sigma}^2, \overline{\sigma}^2]$ as n increases like a sandwich.

Our intuitive idea is that the optimal n_0 will be reached when

$$[l(n_0),\ r(n_0)] \approx [L(n_0),\ R(n_0)],$$

which means the shrinking range of the group averages meets with the expanding range within each group. Meanwhile, when we do the group average under the optimal n_0 (close to the true "blocking size"), we should have some groups average able to estimate the bounds of $[\underline{\sigma}^2, \overline{\sigma}^2]$ which implies $[L(n_0),\ R(n_0)] \approx [\underline{\sigma}^2, \overline{\sigma}^2]$. We also want to construct appropriate H to make $[l(n), r(n)]$ identical with (or approximately equal to) some group averages, such that when the test group size equal to n_0, we have $[l(n_0),\ r(n_0)] \approx [\underline{\sigma}^2, \overline{\sigma}^2]$. Therefore, we expect to have the approximate equality

$$[l(n_0),\ r(n_0)] \approx [\underline{\sigma}^2, \overline{\sigma}^2] \approx [L(n_0),\ R(n_0)].$$

In other words, the n_0 is big enough to make the group means converge to $M[\underline{\sigma}^2, \overline{\sigma}^2]$ and the number of these maximal samples (that is m) is not so small that $[L(n),\ R(n)] \approx [\underline{\sigma}^2, \overline{\sigma}^2]$, while each group has enough variation within it so that it can absorb the extra variation caused by randomness to prevent the group average from falling out of the $[\underline{\sigma}^2, \overline{\sigma}^2]$. This approximated equality seems to imply a kind of self-similarity here.

The next step is about how to choose an appropriate H to make the relation (Equation (3.1)) happen (I do not expect H to have a explicit form but I do want to make the relation hold, which is the guideline).

We will provide two data-driven rules to achieve the trade-off so as to improve the max-mean estimation in practice.

3.2.2 Rule 1: Choosing the *central* group

For $n = 1$, $m = N$, since $Y_{i(1)} = Y_{i(n)}$ for $i = 1, 2, \ldots, m$, we can let the overall level $l(1) = r(1)$, we want to make them fall into $[\underline{\sigma}^2, \overline{\sigma}^2]$, a possible choice is the $l(1) = r(1) = \overline{Y} \coloneqq \frac{1}{N}\sum_{i,j} Y_{ij}$. For $n = 2$, $m = [N/2]$, we only have two points for each group, then in ideal situation, $\{\{Y_{i(1)}\}_{i=1}^m,\ \{Y_{i(2)}\}_{i=1}^m\}$ should divide the dataset approximately uniformly into two groups then the difference between these two values should not be too large, then we should set up a pair of values, which should also fall in $[\underline{\sigma}^2, \overline{\sigma}^2]$. A possible choice is to make a "central group" centering around the $\overline{Y}$ with size $n = 2$, (as long as the $\overline{Y}$ is not so far away from the midpoint of $[\underline{\sigma}^2, \overline{\sigma}^2]$), then select its two bounds. Let

$$l(2) = \underline{H}(Y_{1(1)}, Y_{2(1)}, \ldots, Y_{m(1)}) \coloneqq \max\{Y_{i(1)} :\ Y_{i(1)} \le \overline{Y},\ i = 1, 2, \ldots, m\},$$

and

$$r(2) = \overline{H}(Y_{1(2)}, Y_{2(2)}, \ldots, Y_{m(2)}) := \min\{Y_{i(2)} : Y_{i(2)} \geq \overline{Y},\ i = 1, 2, \ldots, m\}.$$

Then we should have $l(2) \leq \overline{Y} \leq r(2)$. Furthermore, since the division

$$\{\{Y_{i(1)}\}_{i=1}^{m},\ \{Y_{i(2)}\}_{i=1}^{m}\},$$

actually provides enough many points in the neighborhood of $\overline{Y}$, we should have

$$\underline{\sigma}^2 \leq l(2) \leq \overline{Y} \leq r(2) \leq \overline{\sigma}^2,$$

As n increases, for $n = k$, we follow the same design before. By letting

$$l(k) = \underline{H}(Y_{1(1)}, Y_{2(1)}, \ldots, Y_{m(1)}) := \max\{Y_{i(1)} : Y_{i(1)} \leq \overline{Y},\ i = 1, 2, \ldots, m\},$$

and

$$r(k) = \overline{H}(Y_{1(k)}, Y_{2(k)}, \ldots, Y_{m(k)}) := \min\{Y_{i(k)} : Y_{i(k)} \geq \overline{Y},\ i = 1, 2, \ldots, m\}.$$

we will have $[l(k), r(k)]$ expanding gradually as k increases. Explicitly, let $I := \{i : \overline{Y} \in [Y_{i(1)}, \infty) \cup (-\infty, Y_{i(k)}]\}$, then

$$[l(k), r(k)] = \bigcap_{i \in I} [Y_{i(1)}, Y_{i(k)}].$$

We can apply this method to the data problem raised up at the beginning of this chapter (Section Section 3.1) with the true variance interval $[\underline{\sigma}^2, \overline{\sigma}^2] = [1, 4]$.

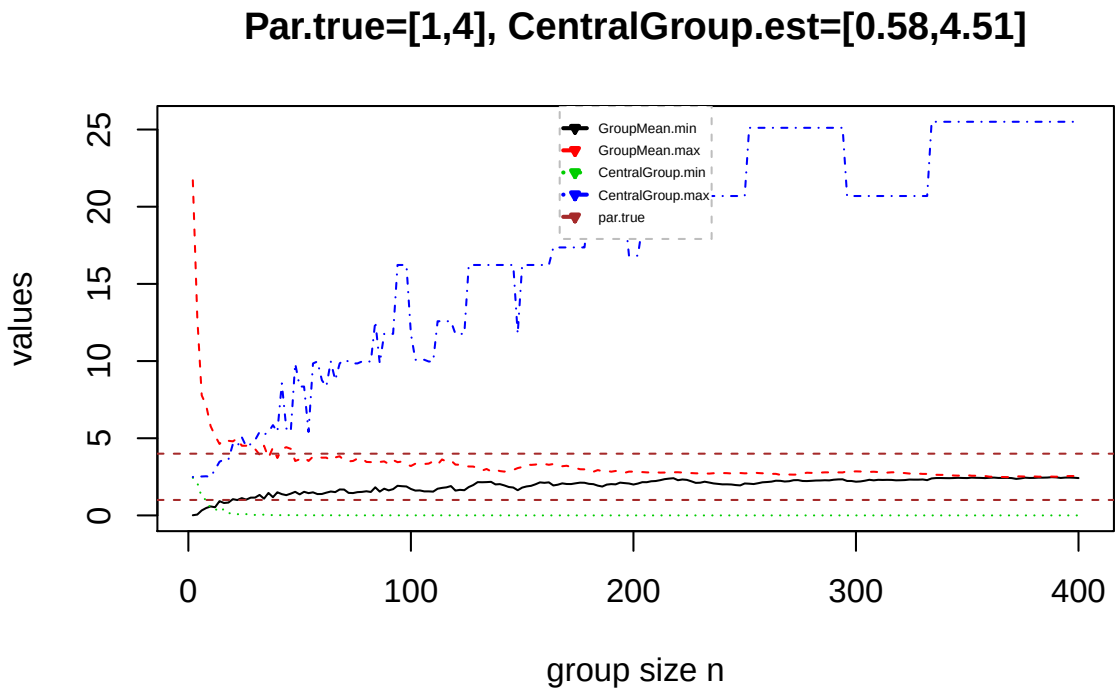


Figure 3.8: Central group estimation

In Figure 3.8, the (black, red) lines are the $(L(n), R(n))$ and the (blue, green) lines are the $(l(n), r(n))$. There are two intersections here ($R(n_1) \approx r(n_1)$ and $l(n_2) \approx L(n_2)$). The brown lines are the true values $[1, 4]$ of $[\underline{\sigma}^2, \overline{\sigma}^2]$ offered by the data providers. We can see that the estimated interval $[0.58, 4.51]$ (got from the y-axis of two intersections) can capture the true interval $[1, 4]$, indicating that our method provides a feasible rule to choose the group size n based on the dataset itself without relying on any distributional information.

In order to check whether our result is a fortunate accident, we can replicate this estimation procedure to get the sampling distribution of the pair of estimators of $[\underline{\sigma}^2, \overline{\sigma}^2]$.

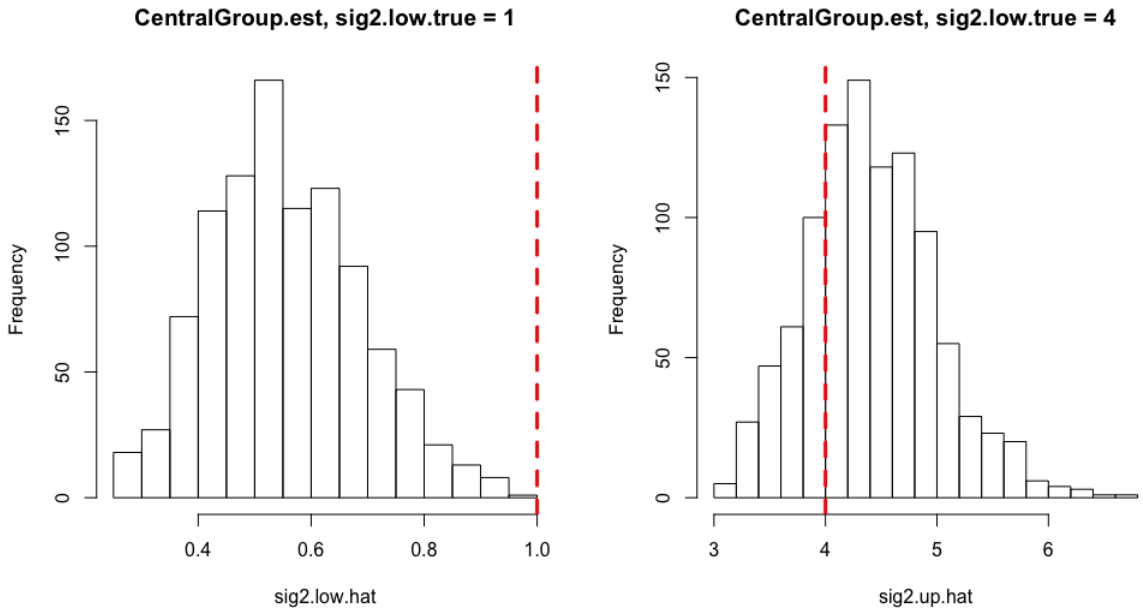


Figure 3.9: Sampling distribution of the central group estimators

With the true interval $[\underline{\sigma}^2, \overline{\sigma}^2] = [1, 4]$, in these 1000 replications, Figure 3.9 shows most of the estimated interval can cover the true one (which is good enough for our purpose: to cover the uncertainty): in fact, all of the estimations of $\underline{\sigma}^2$ are below the true value (and not very far away from it), and only 24% estimations of the right points "underestimate" the $\overline{\sigma}^2$. The average and standard deviations of the two estimated estimators can be expressed in this interval form: $[0.56 \pm 0.13, 4.41 \pm 0.58]$, so we can see that the estimated intervals can roughly cover the true interval and they are not very "wide".

3.2.3 Rule 2: Combining time and value ordering

Another approach is simply designed to efficiently achieve the squeezing relation (Equation (3.1)), but it seems not so relevant with the idea about "the variation between groups versus within groups".

Let us order the $Y_1, \ldots, Y_N$ by values to get $Y_{(1)}, \ldots, Y_{(N)}$ then similarly, group it into get $\{Y_{(ij)}\}_{m \times n}$.

Let n increases from 1 to $N/2$, (in our example, for convenience, N is a large even number)

$(L(n),\ R(n))$ uses the value ordering group methods,

$$[L(n),\ R(n)] := [\min_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n} Y_{(ij)},\ \max_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n} Y_{(ij)}].$$

$(l(n),\ r(n))$ uses the time ordering group methods,

$$[l(n),\ r(n)] := [\min_{1\le i\le m} \frac{1}{h(n)}\sum_{j=1}^{h(n)} Y_{ij},\ \max_{1\le i\le m} \frac{1}{h(n)}\sum_{j=1}^{h(n)} Y_{ij}],$$

where $h(n)$ is a reverse version of the n sequence, which is defined as $h(n) := \frac{N+2}{2} - n$ so that $\{h(n) : n = 1, 2, \ldots, N/2\} = \{N/2,\ N/2-1, \ldots, 1\}$, only to force the $[l(n),\ r(n)]$ to expand pointwise as $[L(n),\ R(n)]$ shrinks.

Set true parameter interval $[\underline{\sigma}^2, \overline{\sigma}^2] = [1, 4]$, then by implementing this method, we get Figure 3.10.

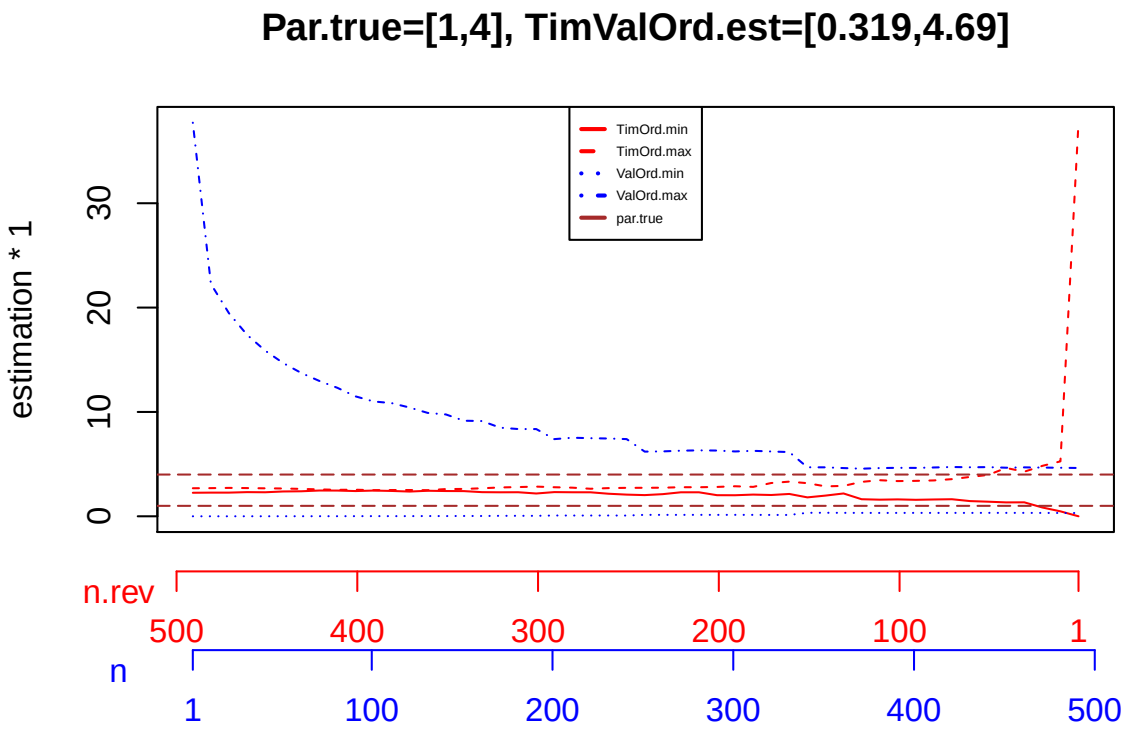


Figure 3.10: Time-value ordering estimation

In Figure 3.10, the brown lines are the true values ($[1, 4]$) and the method gives an estimation at the intersection, which is $[0.32, 4.69]$. We can see that as n increases until to the very end, the blues lines ($[L(n),\ R(n)]$) can cover the $[\underline{\sigma}^2, \overline{\sigma}^2]$ and the red lines ($[l(n),\ r(n)]$) can stay within the $[\underline{\sigma}^2, \overline{\sigma}^2]$. The relation (Equation (3.1)) holds until meeting the intersections, which capture the variance interval and provide us with a good pair of estimation.

Replicate this estimation procedure to get the sampling distribution shown in Figure 3.11.

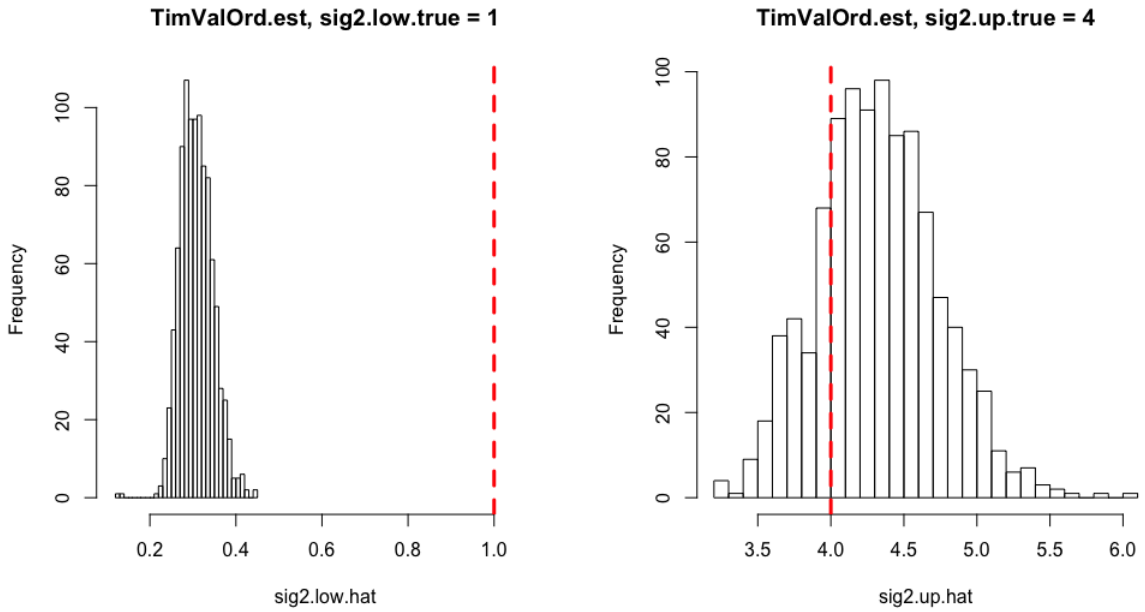


Figure 3.11: Sampling distribution of the time-value ordering estimators

Figure 3.11 shows that the estimations of $\underline{\sigma}^2$ are far smaller than the true ones, implying the time-value ordering estimation has too strong "underestimation" on the left points, which is not good for classical statistical purpose but it does not violate our purpose (to cover the uncertainty interval). 79% of the right-point estimations are greater than the true $\overline{\sigma}^2$. The average estimated interval (with standard deviation) $[0.31 \pm 0.04, 4.32 \pm 0.41]$ shows that the intervals from the time-value ordering rule have been quite narrowed down and can roughly cover the true interval.

From the results in Figures 3.9 and 3.11, you may notice that the central group estimation methods *outperforms* the time-value ordering approach especially on the left point $\underline{\sigma}^2$. In fact, for the estimations of $\underline{\sigma}^2$, 98.2% of the time-value ordering are smaller than those from the central group rule and their average difference is -0.24 ± 0.13. Their performances on the estimation of $\overline{\sigma}^2$ are similar.

Since so far we still do not have a rigorous theoretical backup for our heuristic data-driven rules, we definitely need to do more testing on these two estimation methods. Please turn to Section A.1 for current testing results.

3.2.4 More discussion about the theory of estimation

Let us reflect on why we have challenges to develop the statistical *theory* in the G-framework. From our knowledge, one typical reason is that in general, it is not very easy to directly apply the CLT and LLN in the G-framework to study the property of the estimators. For instance, similar to the results in Jin and Peng (2016), we can show that the order statistics of the maximal distribution $M[\underline{\mu}, \overline{\mu}]$ are actually identically distributed as $M[\underline{\mu}, \overline{\mu}]$, which exactly comes from the uncertainty intrinsically included in the G-framework's distributions whose parameters only describes the *extremes* rather than the *center*. For readers' interest, this idea can be better illustrated by Theorem 3.2.1, 3.2.2 and 3.2.3.

Theorem 3.2.1 (Jin and Peng (2016))**.** *Let $X_1, X_2, \ldots, X_n$ be a (nonlinearly) i.i.d. sample from maximal distribution $M[\underline{\mu}, \overline{\mu}]$. Then we have*

$$\max\{X_1, X_2, \ldots, X_n\} \overset{d}{=} \min\{X_1, X_2, \ldots, X_n\} \overset{d}{=} M[\underline{\mu}, \overline{\mu}].$$

Proof. For each $\varphi \in C_{l.Lip}$, we have

$$\begin{aligned}\hat{\mathbb{E}}[\varphi(\max\{X_1, X_2, \ldots, X_n\})] &= \max_{(x_1,x_2,\ldots,x_n)\in[\underline{\mu},\overline{\mu}]^m} \varphi(\max\{x_1, x_2, \ldots, x_n\}) \\ &= \max_{x\in[\underline{\mu},\overline{\mu}]} \varphi(x) = \hat{\mathbb{E}}[\varphi(X_1)],\end{aligned}$$

which means $\max\{X_1, X_2, \ldots, X_n\} \overset{\mathrm{d}}{=} M[\underline{\mu}, \overline{\mu}]$. Similarly, we can prove that $\min\{X_1, X_2, \ldots, X_n\} \overset{\mathrm{d}}{=} M[\underline{\mu}, \overline{\mu}]$ □

Remark 3.2.1.1 (Jin and Peng (2016)). In general, we can actually prove that for continuous function we have $f \in C(\mathbb{R}^m)$, we have

$$f(X_1, X_2, \ldots, X_n) \overset{\mathrm{d}}{=} M[\underline{\mu}_f, \overline{\mu}_f]$$

where

$$\underline{\mu}_f := \min_{(x_1,x_2,\ldots,x_n)\in[\underline{\mu},\overline{\mu}]^m} f(x_1, x_2, \ldots, x_n), \overline{\mu}_f := \max_{(x_1,x_2,\ldots,x_n)\in[\underline{\mu},\overline{\mu}]^m} f(x_1, x_2, \ldots, x_n).$$

To further explore properties of the order statistics, in the following context, we will use the common notation for order statistics. Let $X_{(k)}$ denote the k-th order statistic of the sequence $X_1, X_2, \ldots, X_n$.

Theorem 3.2.2. *Let X_1, X_2 be an i.i.d sample from $M[\underline{\mu}, \overline{\mu}]$ with independence $X_1 \dashrightarrow X_2$. Then for the random vector of order statistics $(X_{(1)}, X_{(2)})$, we have*

$$(X_{(1)}, X_{(2)}) \overset{d}{=} M(S_2)$$

where S_2 is the simplex in the rectangular $[\underline{\mu}, \overline{\mu}]^2$, or explicitly, the set of points

$$\{(v, w) : \underline{\mu} \le v \le w \le \overline{\mu}\}.$$

Proof. Since $(X_{(1)}, X_{(2)})$ has an intrinsic constraint that $X_{(1)} \le X_{(2)}$, in order to consider a function $\varphi \in C_{l.Lip}(\mathbb{R}^2)$ of this random vector, we actually need to restrict its domain on the simplex S_2. Introduce the bivariate function $g(x_1, x_2) := (\min\{x, y\}, \max\{x, y\}) = (x_{(1)}, x_{(2)})$. Then for any function $\varphi \in C_{l.Lip}(S_2)$, consider

$$\begin{aligned}\hat{\mathbb{E}}[\varphi(X_{(1)}, X_{(2)})] &= \hat{\mathbb{E}}[\varphi(g(X_1, X_2))] \\ &= \max_{(x_1,x_2)\in[\underline{\mu},\overline{\mu}]^2} \varphi(g(X_1, X_2)),\end{aligned}$$

where we use the definition of Maximal distribution for the random vector (X_1, X_2) following $M([\underline{\mu}, \overline{\mu}]^2)$ and $\varphi \cdot g \in C_{l.Lip}$. Meanwhile, we can naturally prove that

$$\{(x_{(1)}, x_{(2)}) : (x_1, x_2) \in [\underline{\mu}, \overline{\mu}]^2\} = \{(v, w) \in [\underline{\mu}, \overline{\mu}]^2 : v \le w\},$$

due to the inequality $\underline{\mu} \leq \min\{x, y\} \leq \max\{x, y\} \leq \overline{\mu}$ if $x,\ y \in [\underline{\mu}, \overline{\mu}]$. In other words, we have $\{g(x_1, x_2) : (x_1, x_2) \in [\underline{\mu}, \overline{\mu}]^2\} = S_2$. Hence,

$$\begin{aligned}\hat{\mathbb{E}}[\varphi(X_{(1)}, X_{(2)})] &= \max_{(x_1,x_2)\in[\underline{\mu},\overline{\mu}]^2} \varphi(g(X_1, X_2)) \\ &= \max_{(v,w)\in S_2} \varphi(v, w),\end{aligned}$$

which holds for any $\varphi \in C_{l.Lip}(S_2)$. In conclusion, we have

$$(X_{(1)}, X_{(2)}) \overset{d}{=} M(S_2).$$

□

Remark 3.2.2.1. This result tells us that the maximal distribution shows nothing more feature about the random variable but *the range of its variation*. $(X_{(1)}, X_{(2)})$ follows $M(S_2)$ because its range is exactly S_2. It will be not hard to generalize our results to $n-$ dimensional case by applying the similar technique.

Theorem 3.2.3. *Let X_i, $i = 1, \cdots, n$ be nonlinear i.i.d random variables in sublinear expectation space $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$ following $M[\underline{\mu}, \overline{\mu}]$. Then for the random vector of order statistics $(X_{(1)}, \cdots, X_{(n)})$, we have*

$$(X_{(1)}, \cdots, X_{(n)}) \overset{d}{=} M(S_n),$$

where S_n is the simplex in the $n-$ dimensional rectangular $[\underline{\mu}, \overline{\mu}]^n$, specifically, written as

$$S_n := \{(x_1, \cdots, x_n) : \underline{\mu} \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq \overline{\mu}\}.$$

Remark 3.2.3.1. Actually, to be consistent, we can use this general result about the joint distribution to retrieve the known properties for the marginal distribution (Theorem 3.2.1). In $n-$ dim case, given the joint distribution, we can show that $X_{(1)} \overset{d}{=} M[\underline{\mu}, \overline{\mu}]$ since

$$\begin{aligned}\hat{\mathbb{E}}[\varphi(X_{(1)})] &= \hat{\mathbb{E}}[\varphi(\min\{X_{(1)}, \cdots, X_{(n)}\})] \\ &= \max_{(x_1,\cdots,x_n)\in S_n} \varphi \cdot \min(x_1, \cdots, x_n) \\ &= \max_{(x_1,\cdots,x_n)\in S_n} \varphi(x_1) \\ &= \max_{x_1\in[\underline{\mu},\overline{\mu}]} \varphi(x_1),\end{aligned}$$

where $[\underline{\mu}, \overline{\mu}]$ is exactly the trajectory on the first dimension when the $(X_{(1)}, \cdots, X_{(n)})$ varying all over the simplex S_n. Furthermore, we can show the joint distribution for lower-dimensional random vector. For example,

$$\begin{aligned}\hat{\mathbb{E}}[\varphi(X_{(1)}, X_{(n)})] &= \hat{\mathbb{E}}[\varphi(h(X_{(1)}, \cdots, X_{(n)})] \\ &= \max_{(x_1,\cdots,x_n)\in S_n} \varphi \cdot h(x_1, \cdots, x_n) \\ &= \max_{(x_1,x_n)\in S_2} \varphi(x_1, x_n)\end{aligned}$$

where $h : (x_1, \cdots, x_n) \in S_n \to (x_1, x_n)$ and the range is nothing mysterious but the $S_2 = \{(x_1, x_n) : \underline{\mu} \leq x_1 \leq x_n \leq \overline{\mu}\}$. Therefore, we prove that $(X_{(1)}, X_{(n)}) \overset{d}{=} M(S_2)$. Generally speaking, for any subsequence $1 \leq i_1 \leq \cdots \leq i_k \leq n$, we actually have $(X_{(i_1)}, \cdots, X_{(i_k)}) \overset{d}{=} M(S_k)$, by using the same idea in the proof of $(X_{(1)}, X_{(n)})$. This reminds us again that the Maximal distribution is exactly attached with the range of variation of the random vector.

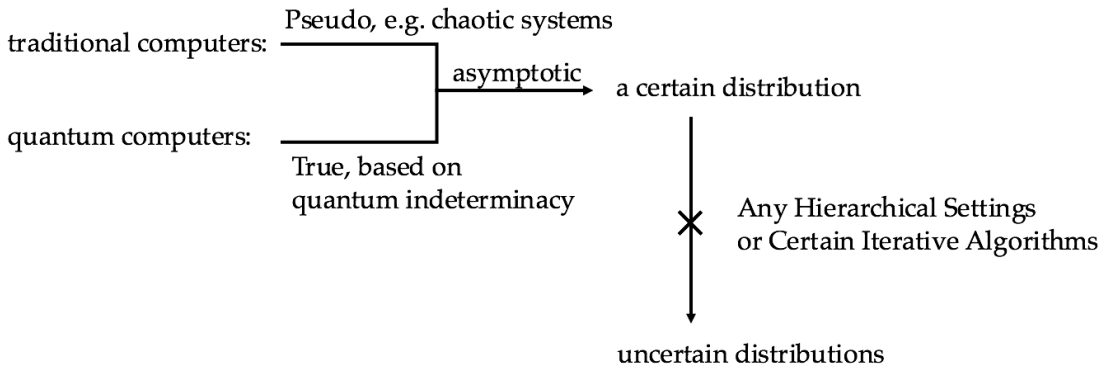
Here we give some *further attempts and ideas* about the estimation method in the G-framework:

1. We need to think about this trade-off: the linear version of converging or distributional properties of the G-estimators may require more information of the (linear) distribution, which may come from the application background;
2. Nonlinear assumptions may only lead us to the nonlinear-version properties of G-estimators, but we will not rule out the possibilities to get some surprisingly interesting results (which may be useful in a brand-new sense) by redefining a new notion of convergence which is more tractable in statistical sense.
3. For instance, we might apply *the conditional expectation or distribution* in the G-framework. In the context of order statistic, we can show that $Y_{(k)}|Y_{(k-1)} = y \sim M[y, \overline{\mu}]$ then we can use the length of the uncertainty interval $(\overline{\mu} - y)$ as the "distance" from the estimator to the true value. It will give a guideline or frame of the description of convergence for general theoretical analysis which will become more delicate once added more distributional information.
4. Another way to think about the theory of estimation is to start from a simplified version of a model describing both the simulation and estimation scheme. A recent progress is that we can look at the histograms of all max-mean estimators (at one side) to find actually the true value is very close to the tail part (tending to reach a relatively stable quantile regardless of the blocking size in the simulation), which motivates us to reflect on a theoretical model trying to explain this pattern. If we can do that, this model will actually be capable of explaining more similar patterns in future real data analysis involving the max-mean estimation.
5. Actually, coming back to the basic ideas of the max-mean estimation, if we take into account the convergence rate of the group average to the Maximal distribution, when n is large enough ($n \geq n_0$), we should we the group average approximately becomes a sample point from the $M[\underline{\sigma}^2, \overline{\sigma}^2]$. In this way, for the group averages with even larger n, we should also preserve this property. Then all we need to do is to collect these maximal sample points, for which we can directly consider the maximum and minimum because of the Dirac-type uncertainty of maximal distribution (we only need one sample point for a Dirac distribution). The threshold n_0 may come from the exploration on the dataset or some previous numerical experiments.

Chapter 4

The Pseudo Simulation of Variance Uncertainty

First of all, when it comes to the simulation in the G-framework, readers need to understand why we call it the *pseudo* simulation. Frankly speaking, from the view of data providers, it is almost impossible for us to truly generate the distribution uncertainty based on a certain algorithm (no matter how chaotic the system is, or how quantized the generator is). From my perspective, we can clearly see the barriers from the diagram below.



(In the diagram above, the traditional computers exclude the machines that have chipsets with a hardware random number generator that can produce true random sequence based on some physical phenomena.)

For the hierarchical settings, if we know all the prior distribution and the relations, then after some efforts of deduction, we can get the posterior distribution or prove its existence and uniqueness. For the iterative algorithms, in general, suppose it is written as $X_n = f(X_{n-1}, \epsilon_{n-1})$

where f is the iterative function and ϵ_{n-1} represents the random part, then

$$\begin{aligned} X_n &= f(X_{n-1}, \epsilon_{n-1}) \\ &= f\big(f(X_{n-2}, \epsilon_{n-2}), \epsilon_{n-1}\big) \\ &= f_1(X_{n-2}, \epsilon_{n-2}, \epsilon_{n-1}) = \cdots = f_n(X_0, \epsilon_0, \ldots, \epsilon_{n-1}). \end{aligned}$$

For each X_n, actually we can get a certain f_n which is a function of X_0 (usually a constant) and $(\epsilon_0, \ldots, \epsilon_{n-1})$ with known distributions. Therefore, we do not really have any distribution uncertainty here.

I would like to clarify that I did not deny the possibility that future technology advancement can make it happen. My point is so far it is hard to break the barriers to step from the stage of truly sampling from a certain distribution to the stage of generating from uncertain distributions based on a certain procedure.

However, we can do the pseudo simulation of distribution uncertainty, that is, generating some data which can hardly be described by the analysts (who do not know the generating scheme) using a certain distribution model. Under this spirit, by borrowing the notations from the G-framework, we will follow the following routine to do the simulation:

$$M[0,1] \xrightarrow{\cdot(\overline{\sigma}-\underline{\sigma})+\underline{\sigma}} M[\underline{\sigma}, \overline{\sigma}] \xrightarrow{\cdot N(0,1)} \hat{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2]) \xrightarrow{\text{NL. CLT}} N(0, [\underline{\sigma}^2, \overline{\sigma}^2]).$$

The $\hat{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ is the new concept *semi-G-normal distribution* which is formally defined in Section 4.2.

4.1 Pseudo simulation of maximal distribution

The motivation of $M[0,1]$ is to cover the distribution uncertainty (especially the mean uncertainty) within $[0,1]$, regardless of any bizarre fluctuation in that interval. It will based on the least conditions: as long as the data points is bounded in $[0,1]$, they can be regarded as a sample from $M[0,1]$. It will be clearer by looking at the definition: $X \sim M[0,1]$ iff

$$\hat{\mathbb{E}}[\varphi(X)] = \max_{v \in [0,1]} \varphi(v) = \max_{v \in [0,1]} E_{\delta_v}[\varphi(X)].$$

where δ_v is the dirac measure: $\delta_v(X = v) = 1$. This means that when we have no idea about the random patterns in $[0,1]$, we are making a concession that each data point is treated as a sample point from the dirac distribution (i.e. almost surely equal to a constant), which seems totally degenerated and boring but this will be an almost undoubtedly valid setting. In fact, because of the concession, $M[0,1]$ can cover all linear distributions with support on $[0,1]$. In other words, considering the set

$$\mathbb{F} := \{F : \text{ the CDF of distribution with support } [0,1]\},$$

then for any $F \in \mathbb{F}$, we have the inequality that

$$E_F[\varphi(X)] = \int_0^1 \varphi(x)\, \mathrm{d}F(x) \le \int_0^1 \max_{x \in [0,1]} \varphi(x)\, \mathrm{d}F(x) = \max_{x \in [0,1]} \varphi(x),$$

which implies $E_F[\varphi(x)] \le \hat{\mathbb{E}}[\varphi(x)]$.

The data provider can do the pseudo simulation of $M[0,1]$ by doing general procedure:

1. Generate the block length $n_i \in \mathbb{N}_+$ from some discrete random rule A;
2. randomly choose one distribution F_i from $\mathbb{F}$ based on some random rule B;
3. simulate a block $Z_1, Z_2, \ldots, Z_{n_i}$ from F_i;
4. put the block into the current dataset, then go back to step 1, until the data size reaches N.

For the second step, in practice, since we cannot really run through all the elements in $\mathbb{F}$, we can choose F_i from a subset $\{F_i : i \in I\} \subset \mathbb{F}$. There are many candidates of subsets we can choose. Rule A can be any discrete random generator, some special sequences like the digit numbers of some irrational number (e.g. π and e) or integer outcomes provided by quantum computers. Rule B can be some Markov models or hierarchical models.

For the example used at the beginning of this chapter, we actually generate $M[0,1]$ from this simple scheme:

1. Generate the block length $n_i \sim \text{Poisson}(N/100)$, where i is the index of loop;
2. generate $\alpha_i, \beta_i \sim \text{Unif}(0, 50)$; $c_i \sim \text{Beta}(\alpha_i, \beta_i)$ (the support of Beta distribution is $[0,1]$);
3. generate a block of data $Z_1, Z_2, \ldots, Z_{n_i} \sim F_{\delta_{c_i}}$ (which belongs to a family of dirac measures; in other words, they are all equal to c_i)
4. put the block into the current dataset, then go back to step 1, until the total data size reaches N.

The following theorem shows we can use the simulated $M[0,1]$ sequence to produce $M[\underline{\sigma}, \overline{\sigma}]$ for any $\underline{\sigma}, \overline{\sigma} \in \mathbb{R}$.

Theorem 4.1.1 (Scaling and translating properties of $M[0,1]$)**.** *In the sublinear expectation space* $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$*, let* $X \sim M[0,1]$*,* $c \in \mathbb{R}$ *and* $a \in \mathbb{R}_+$ *, then we have*

$$aX + c \overset{d}{=} M[c, a + c].$$

Proof. First we will show that $aX + c$ will also follow the Maximal distribution. Let $\bar{X}$ be an independent copy of X (then accordingly, $a\bar{X} + c$ will be an independent copy of $aX + c$), given $X + \bar{X} \overset{\text{d}}{=} 2X$, we have

$$(aX + c) + (a\bar{X} + c) \overset{\text{d}}{=} 2aX + 2c \overset{\text{d}}{=} 2(aX + c).$$

Next we only need to check the $\hat{\mathbb{E}}[aX + c]$ and $-\hat{\mathbb{E}}[-(aX + c)]$, which can be directly obtained from the positive homogeneity and cash translatability:

$$\hat{\mathbb{E}}[aX + c] = a\hat{\mathbb{E}}[X] + c = a + c,$$

and

$$-\hat{\mathbb{E}}[-(aX + c)] = -(\hat{\mathbb{E}}[a(-X)] - c) = a(-\hat{\mathbb{E}}[-X]) + c = c. \qquad \square$$

Remark 4.1.1.1. This result reminds us that $(\overline{\sigma} - \underline{\sigma}) \cdot M[0,1] + \underline{\sigma} \overset{\text{d}}{=} M[\underline{\sigma}, \overline{\sigma}]$.

Remark 4.1.1.2 (how to standardize $M[\underline{\mu}, \overline{\mu}]$). Given $Y \sim M[\underline{\mu}, \overline{\mu}]$, using the above theorem, we want $aX + c \overset{\mathrm{d}}{=} Y \overset{\mathrm{d}}{=} M[\underline{\mu}, \bar{\mu}]$, set $c = \underline{\mu}$, $a + c = \bar{\mu}$, then $a = \bar{\mu} - \underline{\mu}$ and we have

$$(\bar{\mu} - \underline{\mu})M[0, 1] + \underline{\mu} \overset{\mathrm{d}}{=} M[\underline{\mu}, \bar{\mu}],$$

equivalently, (since it is safe to move the constants around the identical distributed relation)

$$\frac{1}{\bar{\mu} - \underline{\mu}} M[\underline{\mu}, \bar{\mu}] - \frac{\underline{\mu}}{\bar{\mu} - \underline{\mu}} \overset{\mathrm{d}}{=} M[0, 1].$$

To summarize this section, we can implement the pseudo simulation to produce a sample sequence from $M[1, 2]$. Figures 4.1 and 4.2 visualize two simulated artificial datasets, in which we can see that when the blocking size (in the simulation algorithm) is very small, the generated sample actually change its "constant" value nearly in a pointwise way.

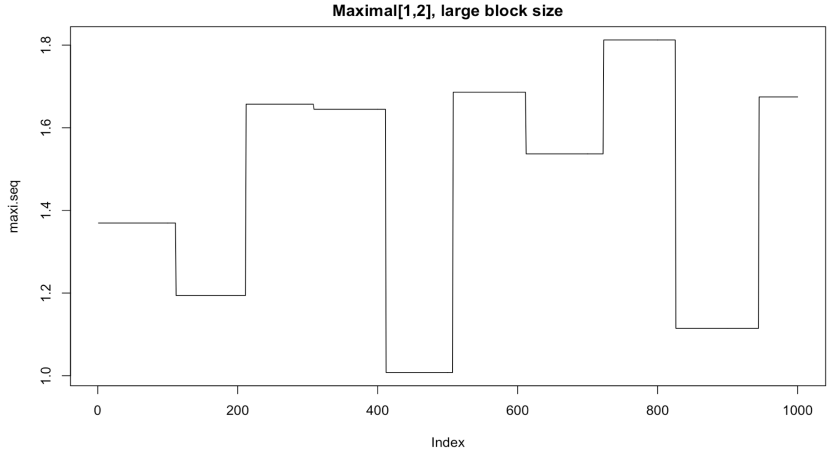


Figure 4.1: Simulated sample of $M[1, 2]$ with large block length

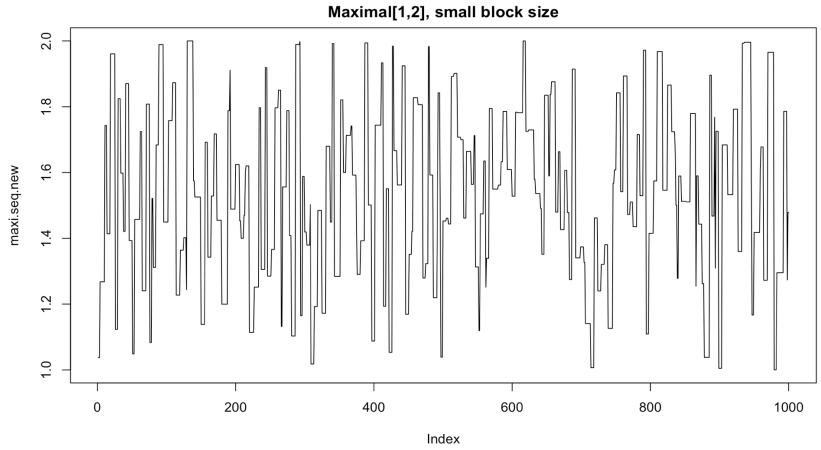


Figure 4.2: Simulated sample of $M[1, 2]$ with small block length

4.2 Pseudo simulation of semi-G-normal distribution

First we can review the new concept of semi-G-normal distribution and its properties.

Definition 4.2.1 (the Semi-G-normal distribution in one dimension)**.** In the sublinear expectation space $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$, we say W will follow the *Semi-G-normal distribution*, denoted as, $W \sim \hat{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$, if there exist $Y \sim N(0, [1, 1])$, $Z \sim M[\underline{\sigma}, \overline{\sigma}]$, $\overline{\sigma} \geq \underline{\sigma} \geq 0$ and Y is independent from Z, such that

$$W = Z \cdot Y$$

where the direction of independence here cannot be reversed and Y can be regarded as the classical standard normal distribution.

Remark 4.2.1.1 (the mean and variance of W). It is not hard to show that it has a certain zero mean:

$$\begin{aligned}\hat{\mathbb{E}}[W] &= \hat{\mathbb{E}}[ZY] \\ &= \hat{\mathbb{E}}[\hat{\mathbb{E}}[zY]_{z=Z}] \\ &= \hat{\mathbb{E}}[E[zY]_{z=Z}] = \hat{\mathbb{E}}[(z \cdot 0)_{z=Z}] = 0\end{aligned}$$

similarly, $-\hat{\mathbb{E}}[-W] = 0$. For the variance, we have

$$\begin{aligned}\hat{\mathbb{E}}[W^2] &= \hat{\mathbb{E}}[Z^2Y^2] \\ &= \hat{\mathbb{E}}[\hat{\mathbb{E}}[z^2Y^2]_{z=Z}] \\ &= \hat{\mathbb{E}}[E[z^2Y^2]_{z=Z}] \\ &= \hat{\mathbb{E}}[(z^2 \cdot 1)_{z=Z}] = \max_{z \in [\underline{\sigma}, \overline{\sigma}]} z^2 = \overline{\sigma}^2\end{aligned}$$

and similarly,

$$\begin{aligned}-\hat{\mathbb{E}}[-W^2] &= \hat{\mathbb{E}}[-Z^2Y^2] \\ &= -\hat{\mathbb{E}}[E[-z^2Y^2]_{z=Z}] \\ &= -\max_{z \in [\underline{\sigma}, \overline{\sigma}]} (-z^2) = \min_{z \in [\underline{\sigma}, \overline{\sigma}]} z^2 = \underline{\sigma}^2.\end{aligned}$$

We are still working on the simulation of the sequential independence in this framework and have some incomplete results, which show that to simulate $Z \dashrightarrow Y$, as long as the realization of Z have no effect on the distribution (or generation) of Y, we can achieve the sequential independence (which is much weaker than the classical independence). Then we can simulate the semi-G-normal sequence shown by Figure 4.3, the initial analysis of which is exactly the work we have done at the beginning of this chapter in Section 3.1.

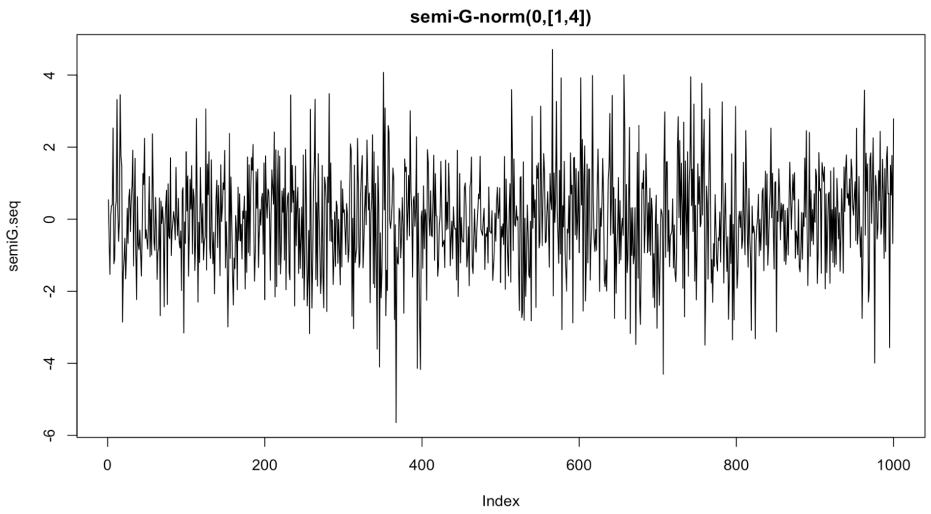


Figure 4.3: A sample sequence from $\hat{N}(0,[1,4])$

What we can do after getting the simulated semi-G-normal sequence? One attempt is to directly estimate its sublinear expectation (like its variance interval). Figure 4.4 shows that the estimated interval is able to describe the variance uncertainty of the sequence, indicating that we actually have built up a procedure from simulation of the distribution to estimation of its parameter interval. This procedure is beneficial to the validation of future candidates of the estimation and also inspires us to think about how to apply the estimation methods in G-framework into real dataset.

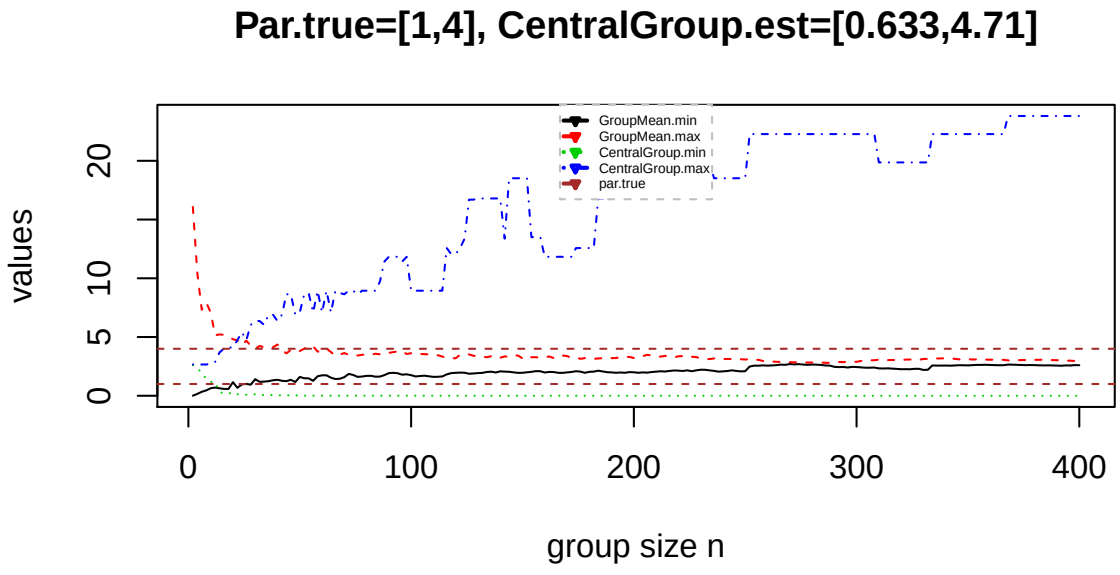


Figure 4.4: Central group estimation of the variance uncertainty interval of $\hat{N}(0,[1,4])$

4.3 Approximate simulation of G-normal distribution

Review the central limit theorem in the G-framework.

Theorem 4.3.1 (General connection between the Semi-G-normal and the G-normal distribution)**.** *In a sublinear expectation space* $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$*, consider a sequence of nonlinear i.i.d.* $\{W_i\}_{i=1} \sim \hat{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ *and* $X \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$*, then for any* $\varphi \in C(\mathbb{R})$ *satisfying linear growth condition, we have*

$$\hat{\mathbb{E}}[\varphi(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i)] \to \hat{\mathbb{E}}[\varphi(X)],$$

as $n \to \infty$*. In other words,* $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i$ *converges in distribution to the G-normal distributed* X*.*

After simulating the sequence of semi-G-normal random points (with the sequential nonlinear i.i.d.), we can group the sequence (with size n) to approximately simulate the G-normal distribution.

Figure 4.5 shows a sample sequence from $N(0, [1, 4])$ which looks like independent sample points and we need to know that, different from the semi-G-normal sequence, this is no longer simply a sample from a mixture of samples of normal distributions with different variances.

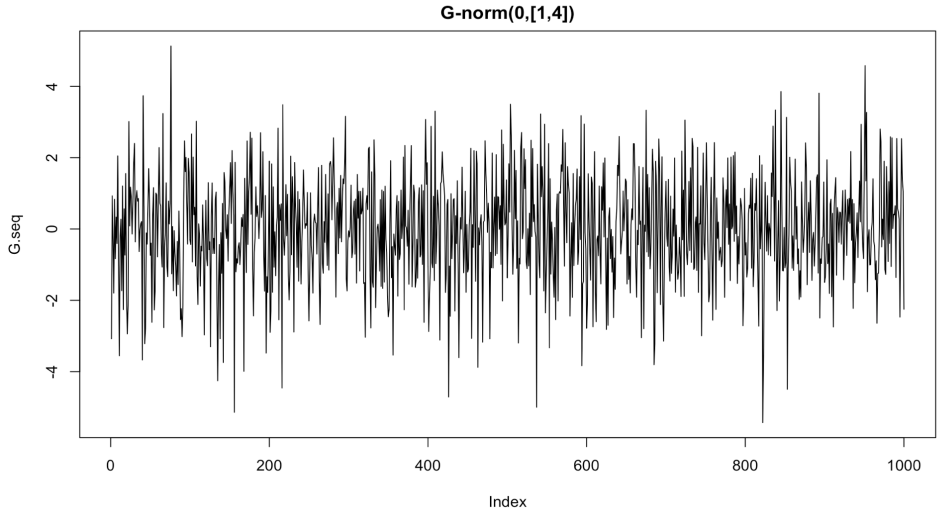


Figure 4.5: A sample sequence from $N(0, [1, 4])$

Figure 4.6 implies that the central-group max-mean estimation is able to capture the variance interval regardless of the degree of uncertainty of the nonlinear distribution (since $N(0, [1, 4])$ has much more uncertainty than $\hat{N}(0, [1, 4])$, as long as the sequence has certain zero mean and variance uncertainty.

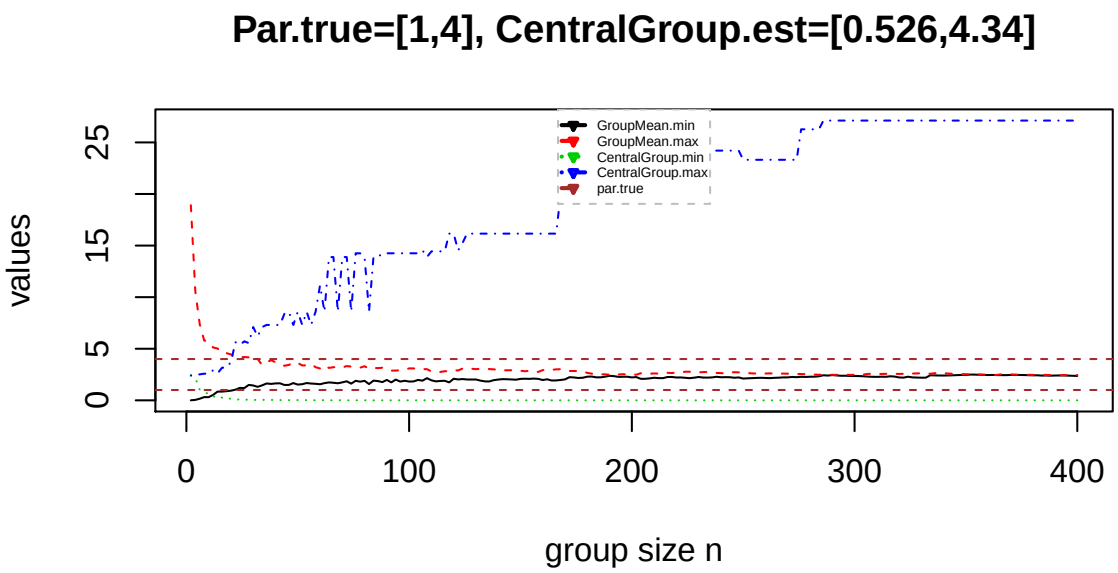


Figure 4.6: Central group estimation of the variance uncertainty interval of $N(0,[1,4])$

Chapter 5

Concluding Remarks and Future Development

To begin with, I would like to share one aphorism I learned from Prof. Peng:

"Having some precaution before data analysis is even more important than mathematics."

Actually, this thesis is the theoretical and technical *preparation* for future application of the G-expectation framework. In general, I plan to examine whether we can build up *a complete data analysis procedure* under ambiguity enhanced by the G-framework during my Ph.D. career. Given a real dataset from an industrial problem, we will try to test the distribution uncertainty, estimate the parameter uncertainty interval, compute the corresponding G-expectation using the estimated parameters and finally give a robust strategy to control or cover the uncertainty. In order to examine our data analysis procedure, we need to know the "true parameter intervals" so it requires us to design a simulation scheme to provide the artificial dataset with known details of its underlying structure. However, these details are only known by *the data provider* (or . Without awareness of any information of these details, *the data analyst* will use the statistical methods in the G-framework to analyze the dataset.

This thesis is also a *summary* of my thoughts on nonlinear expectation over the past two years, in which readers may notice that there are many points and ideas still waiting for more suggestions and discussion. For the remainder of this chapter, I would like to share with readers some hopefully useful and heuristic comments to make some conclusions, plans, ambition and also some dreams.

First of all, the newly-defined concept (*Semi-G-normal Distribution*) plays a central role in the thesis. This is also an *original* substructure which provides a transition from classical to G-normal distribution to fill in the thinking gap between these two basic concepts and, in a more general sense, also between the linear and nonlinear expectation worlds. Meanwhile, It offers a more feasible and intuitive way to think about, compute and simulate the G-normal distribution. For the PDE side, the iterative method based on the semi-G-normal distribution (Theorem 2.3.4) gives an new and interesting stochastic link inspiring more attempts to solve the high-dimensional G-heat equation using a Monte-Carlo method.

As we can see in Chapter 2, the set of measures of the semi-G-normal distribution, consisting of a class of linear measures of $N(0, \sigma^2)$ with $\sigma \in [\underline{\sigma}, \overline{\sigma}]$, is smaller than that of the G-normal distribution. Hence, on the one hand, it has less unusual properties than the G-normal

distribution and more similar properties to the linear normal distributions (e.g. its sublinear "skewness" is equal to zero); on the other hand, it has strong connections with G-normal distribution (e.g. the sublinear expectation with convexity and the nonlinear CLT). As a crucial concept in this thesis, the semi-G-normal distribution is aimed for constructing a bridge between the linear and the sublinear world so that in the future, more tools on the land of linear expectation framework can be transformed to the island of the sublinear expectation framework to help the larger community in both the academy and industry to understand the intuition, learn the theory or algorithms and apply them to the real world problems with uncertainty.

As a future continuation of Chapter 2, we intend to do more theoretical and numerical practice about the G-normal distribution (such as the simulation of multivariate G-normal distribution). Since we have also defined and studied the *semi-sequential independence*, in the future, under this semi-version nonlinear independence, we can make the semi-G-normal distribution into the *Semi-G-Brownian motion* which may also play as the semi-version of the G-Brownian motion (as a transition from classical to G-Brownian motion). By approximately simulating the G-Brownian Motion using the semi-G-Brownian Motion, we might be able to simulate the backward stochastic differential equations (BSDEs) driven by G-Brownian Motion which are also connected with a class of nonlinear PDEs, by appying the Feynman-Kac Formula in the G-framework in Hu et al. (2014), in which by computing the G-expectation of G-Brownian Motion using the corresponding *iterative algorithm*. In this way we may be able to numerically solve the multi-dimensional nonlinear PDEs (which often make classical numerical PDE methods fail) (similar ambition can be found in Beck et al. (2017)) and those PDEs are connected with a large class of financial pricing formulas under ambiguity.

Based on the distributions and independence studied in Chapter 2 (especially the semi-G-normal distribution $\hat{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ with its connection with $N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$ and $N(0, \sigma^2)$), Chapters 3 and 4 explore the statistical scheme in the G-framework, consisting of estimation and simulation, specially under variance uncertainty.

As shown in Chapter 3, the improved max-mean estimation methods with two data-driven rules (Central group rule and Time-value ordering rule) are able to capture the true parameter interval regardless of the underlying distributional setting on the uncertain parameter (especially the central group rule). When we design these two heuristic rules, we do not really rely on or assume any (explicit) prior distributional information on the varying parameter (although some implicit requirements are waiting to be found out in the future). Meanwhile, since usually we are unaware of the blocking design in the generation scheme of dataset (except in real application, we may make some reasonable assumptions based on past experiences), we should not expect the dataset is generated based on some changing parameters with equal blocking lengths. In this sense, our data-driven rules are not really "guessing" some "true" blocking size but "choosing" an optimal group size n to do the trade-off between the $[L(n), R(n)]$ and $[l(n), r(n)]$ because we know that the true parameter interval should approximately appear at some places on the max-mean and min-mean curves as n changes.

In the near future, we will need to do more examinations on the two data driven rules by changing the data generation scheme, comparing with some existing classical statistical models (such as those models based on Bayesian or Markov setting) and so on. By doing so, we want to reflect on the reasons why our methods have these interesting and promising numerical performances and also think about whether we can put these ideas into the context of data analysis in our real world. Both of these directions may push us further to a more general

statistical methodology associated with the G-framework. If possible, we intend to further think about the theoretical backup (or at least, the existence of more complete statistical theory) behind this estimation method, which will not only give more solid support on the method but more chances for us to design the estimation method in general (e.g. the estimation for time series in the G-framework). Meanwhile, we also need to put the max-mean estimation into real-world scenario where we can do the trade-off on the grouping design based on some application background, which is actually not only more sensible and reasonable but also more practically useful. The estimation in the G-framework will also be connected with some classical statistical theory with similar spirits such as the *Robust Statistics*. The former will raise up more interesting questions in situations under ambiguity for the latter as well as the larger statistical world.

For the simulation, Chapter 4 has already worked on the pseudo simulation of the maximal distribution then the simulation of semi-G-normal and G-normal distribution. The pseudo simulation scheme of the maximal distribution is based on the idea about the "black-box" situation involving two parties (the data analyst and the data provider), which is a quite general and flexible design. It is a usually valid concession when, without making assumptions, the pattern of dataset is hard to describe by a certain probabilistic model if the analyst is not offered any crucial information about the generation scheme (in the simulated world) or the data structure (in the real world). It is important to point out that the pseudo simulated sequence can be treated as the maximal distribution only by the data analysts (which is a conservative concession for them due to lack of prior information) but when provided with underlying information about the random pattern, these pseudo generated sequence does not need to be considered as the maximal sequence. Therefore, one cannot expect us to *truly* reproduce and visualize the "mysterious" uncertain pattern within the interval of $M[\underline{\mu}, \overline{\mu}]$. In other words, if one day I show you the exact changing rule to describe the stochastic pattern of $M[\underline{\mu}, \overline{\mu}]$, then without loss of generality, it can be expressed by a certain probability measure P which means its uncertainty set has degenerated to a singleton and it becomes a distribution in classical framework. From my point of view, the simulation of uncertainty (or those nonlinear distributions in the G-framework) should be discussed under the "provider-analyst" setting and the black-box situation. The simulated artificial datasets are mostly used in the exploration and design of the statistical methods in the G-framework, which will be further examined in real data analysis.

In addition to the simulation of the *distributions* in the G-framework, another aspect which is also an important and challenging question I plan to study in the future is how to understand and simulate the asymmetric *sequential independence*: in the G-framework, usually "X is independent from Y" does not imply "Y is independent from X". Do we have this kind of relations in real world and how can we further quantify them? Hu (2011) shows in the non-degenerate case (staying in the G-framework but not in the linear framework), if X and Y are mutually nonlinearly independent, they will be both maximally distributed. How can we intuitively understand or numerically check this astonishing result? The numerical study of sequential independence is also crucial to the future simulation of the G-Brownian motion.

For the statistics in the G-framework, besides the estimation and simulation, *doing hypothesis tests* is always the additional crucial part I will reflect on during the future study of the previous two parts, because I will need to give a reasonable test of the distribution uncertainty or the asymmetric independence for the dataset before applying any tools in the G-framework.

Meanwhile, the ideas of simulation and estimation are highly related with the intuition of hypothesis tests in the G-framework. For instance, the idea of improved max-mean estimation, to appropriately group the data, actually gives us a hint about how to test the parameter uncertainty in the dataset.

Coming back to the aphorism at the beginning of chapter, the statistical methods and practice of G-framework is aimed at encouraging the data analysts and statisticians to have the notion of precaution in order to get closer to the reality (usually with uncertainty), and providing them with tools in a generalized framework intrinsically including uncertainty so as to not simply be restricted in the classical probability framework. The methods in the G-framework is not a replacement of the classical ones but the supplement or pre-strategies before diving into the delicate analysis of dataset with more underlying information (which usually will apply the stochastic models in classical framework). With this spirit, I intend to explore the feasibility and make the industrial attempts of the statistical methods of G-framework, which become one of the important future research directions following this thesis.

Meanwhile, from the discussion with my supervisor, we plan to do data analysis in a typical industrial world, which is a relatively independent project so far. During this new journey, if there is any places (showing some uncertainty) potentially requiring the techniques in G-framework, it will be really fantastic to combine these two directions and see what we can achieve and how the precaution strategy from G-framework will be beneficial in application.

As the statistician George Box said, "all models are wrong but some are useful". All models are wrong but we do care about how wrong the model may be as long as it is useful. Nonetheless, if the assumptions of the model are too strong, we should not take "this wrong" as granted and assess it only based on its "usefulness", because there is another aspect called the "danger" (perhaps caused by the uncertainty brought by the lack of information). We need to worry about not only the "risk control" but also the "*uncertainty* control". We should never stop trying to improve the model to make it closer to reality (of course, before it becomes useless). The G-expectation framework actually gives us more space to explore to enhance the models with more precaution to minimize the danger brought by the classical ones.

Acknowledgement: I would like to express my sincere gratitude to Prof. Kulperger and Prof. Peng for their beneficial guidance and strong support all the time. Since I stepped into this area in 2016, especially during the one-year thesis research, thanks to the fruitful discussion with my supervisor, Prof. Peng, Prof. Davison, Prof. Yu, Prof. Huang and my fellow students, as well as the valuable communication in the meetings and workshops (during the summer of 2017 and 2018), which is very necessary and important when studying the G-framework (a brand-new comprehensive framework requiring the strengths from different areas), I have been able to gradually build up a basic system of perspectives on the statistical and practical methodology of G-framework. I treat this experience as an exploration or adventure full of curiosity, struggling and excitements, which will be definitely further pushed forward and expanded in my future career.

Bibliography

Avellaneda, M. and A. Paras (1996). Managing the volatility risk of portfolios of derivative securities: the Lagrangian uncertain volatility model. *Applied Mathematical Finance 3*(1), 21–52.

Bayraktar, E. and A. Munk (2015). Comparing the G-normal distribution to its classical counterpart. *Communications on Stochastic Analysis 9*(1), 1–18.

Beck, C., E. Weinan, and A. Jentzen (2017). Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations. *arXiv preprint arXiv:1709.05963*.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing 16*(5), 1190–1208.

Chen, Z. and L. Epstein (2002). Ambiguity, risk, and asset returns in continuous time. *Econometrica 70*(4), 1403–1443.

Chen, Z., P. Wu, and B. Li (2013). A strong law of large numbers for non-additive probabilities. *International Journal of Approximate Reasoning 54*(3), 365–377.

Choquet, G. (1954). Theory of capacities. In *Annales de l'Institut Fourier*, Volume 5, pp. 131–295.

Crandall, M. G., H. Ishii, and P.-L. Lions (1992). User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Mathematical Society 27*(1), 1–67.

Denis, L., M. Hu, and S. Peng (2011). Function spaces and capacity related to a sublinear expectation: application to G-Brownian motion paths. *Potential Analysis 34*(2), 139–161.

Dolinsky, Y. (2012). Numerical schemes for G-expectations. *Electronic Journal of Probability* (98), 1–15.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics 75*(4), 643–669.

Forsythe, G. E., C. B. Moler, and M. A. Malcolm (1977). *Computer methods for mathematical computations*. Prentice-Hall.

Hu, M. (2011). The independence under sublinear expectations. *arXiv preprint arXiv:1107.0361*.

Hu, M. (2012). Explicit solutions of G-heat equation with a class of initial conditions by G-Brownian motion. *Nonlinear Analysis 75*(18), 6588–6595.

Hu, M., S. Ji, S. Peng, and Y. Song (2014). Comparison theorem, Feynman–Kac formula and Girsanov transformation for BSDEs driven by G-Brownian motion. *Stochastic Processes and their Applications 124*(2), 1170–1195.

Hu, Z. C. and L. Zhou (2015). Multi-dimensional central limit theorems and laws of large numbers under sublinear expectations. *Acta Mathematica Sinica, English Series 31*(2), 305–318.

Jin, H. and S. Peng (2016). Optimal unbiased estimation for maximal distribution. *arXiv preprint arXiv:1611.07994*.

Knight, F. H. (1921). *Risk, uncertainty and profit, First edition*. Boston, New York, Houghton Mifflin Company.

Kolmogorov, A. N. (1956). *Foundations of the theory of probability*. Chelsea.

Li, X.-p. (2015). A central limit theorem for m-dependent random variables under sublinear expectations. *Acta Mathematicae Applicatae Sinica, English Series 31*(2), 435–444.

Li, Y. and R. Kulperger (2018). An iterative approximation of the sublinear expectation of an arbitrary function of G-normal distribution and the solution to the corresponding G-heat equation. *arXiv preprint arXiv:1804.10737*.

Peng, S. (1997). Backward sde and related g-expectation. backward stochastic differential equations (paris, 1995-1996), 141-159. *Pitman Res. Notes Math. Ser 364*.

Peng, S. (2004). Nonlinear expectations, nonlinear evaluations and risk measures. In *Stochastic Methods in Finance*, pp. 165–253. Springer Berlin Heidelberg.

Peng, S. (2007a). G-expectation, G-Brownian motion and related stochastic calculus of Itô type. In *Stochastic analysis and applications*, pp. 541–567. Springer.

Peng, S. (2007b). Law of large numbers and central limit theorem under nonlinear expectations. *arXiv preprint math/0702358*.

Peng, S. (2008). Multi-dimensional G-Brownian motion and related stochastic calculus under G-expectation. *Stochastic Processes and their Applications 118*(12), 2223–2253.

Peng, S. (2010). Nonlinear expectations and stochastic calculus under uncertainty. *arXiv preprint arXiv:1002.4546*.

Peng, S. (2017). Theory, methods and meaning of nonlinear expectation theory. *SCIENTIA SINICA Mathematica 47*(10), 1223–1254.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica 57*(3), 571–587.

Song, Y. (2017). Normal approximation by Stein's method under sublinear expectations. *arXiv preprint arXiv:1711.05384*.

Von Neumann, J. and O. Morgenstern (1945). Theory of games and economic behavior. *Bull. Amer. Math. Soc 51*(7), 498–504.

Appendix A

More Technical Details

A.1 More testing of the estimation methods

A.1.1 Small true parameters

Set a relatively small true parameter interval as $[\underline{\sigma}, \overline{\sigma}] = [0.1, 0.2]$ ($[\underline{\sigma}^2, \overline{\sigma}^2] = [0.01, 0.04]$), then implement the central group rule, by replicating the experiment 1000 times, to get the following results.

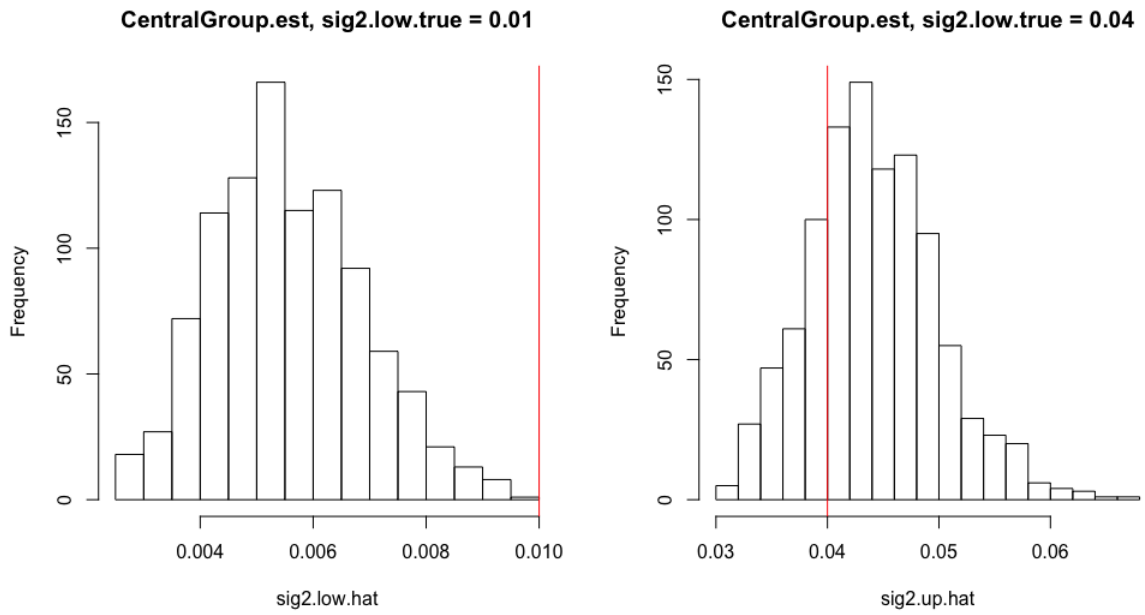


Figure A.1: Sampling distribution of the central group estimators with small true parameters

From Figure A.1, we can see that the central group method can still approximately capture and cover the true parameters, indicating that the results from our estimation method do have the potentials to move after the true parameters, rather than create some illusion (like randomly staying around some values irrelevant with the true parameters). Similar to Figure 3.9, most of the estimations of the left point are below and not too far away from the true one.

Similarly, we have the following histograms (Figure A.2) for the estimation method based on the time-value ordering.

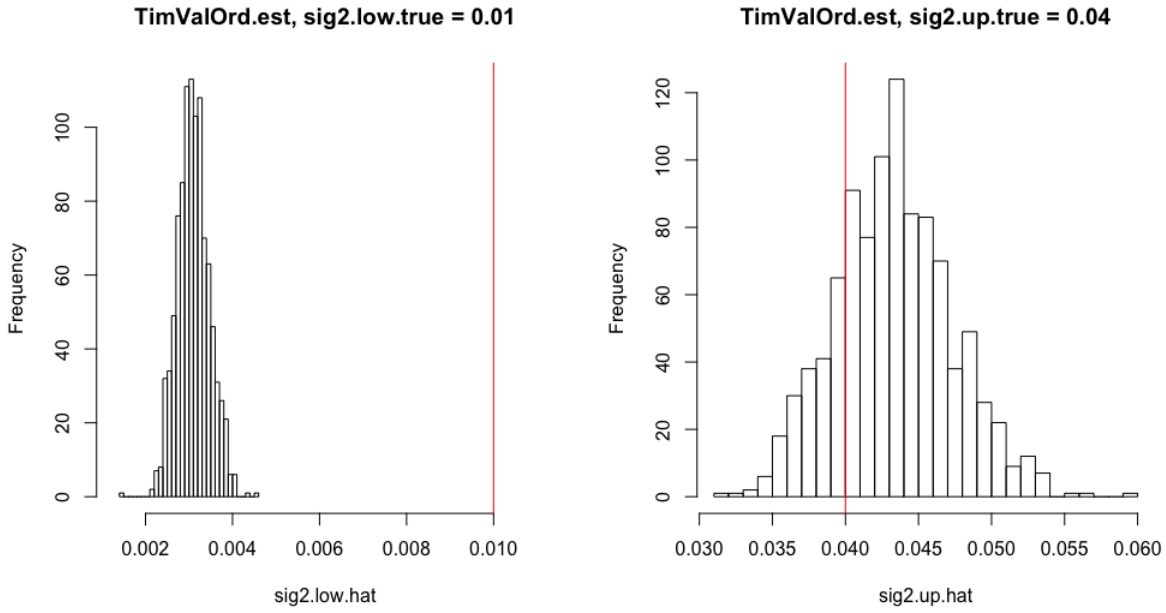
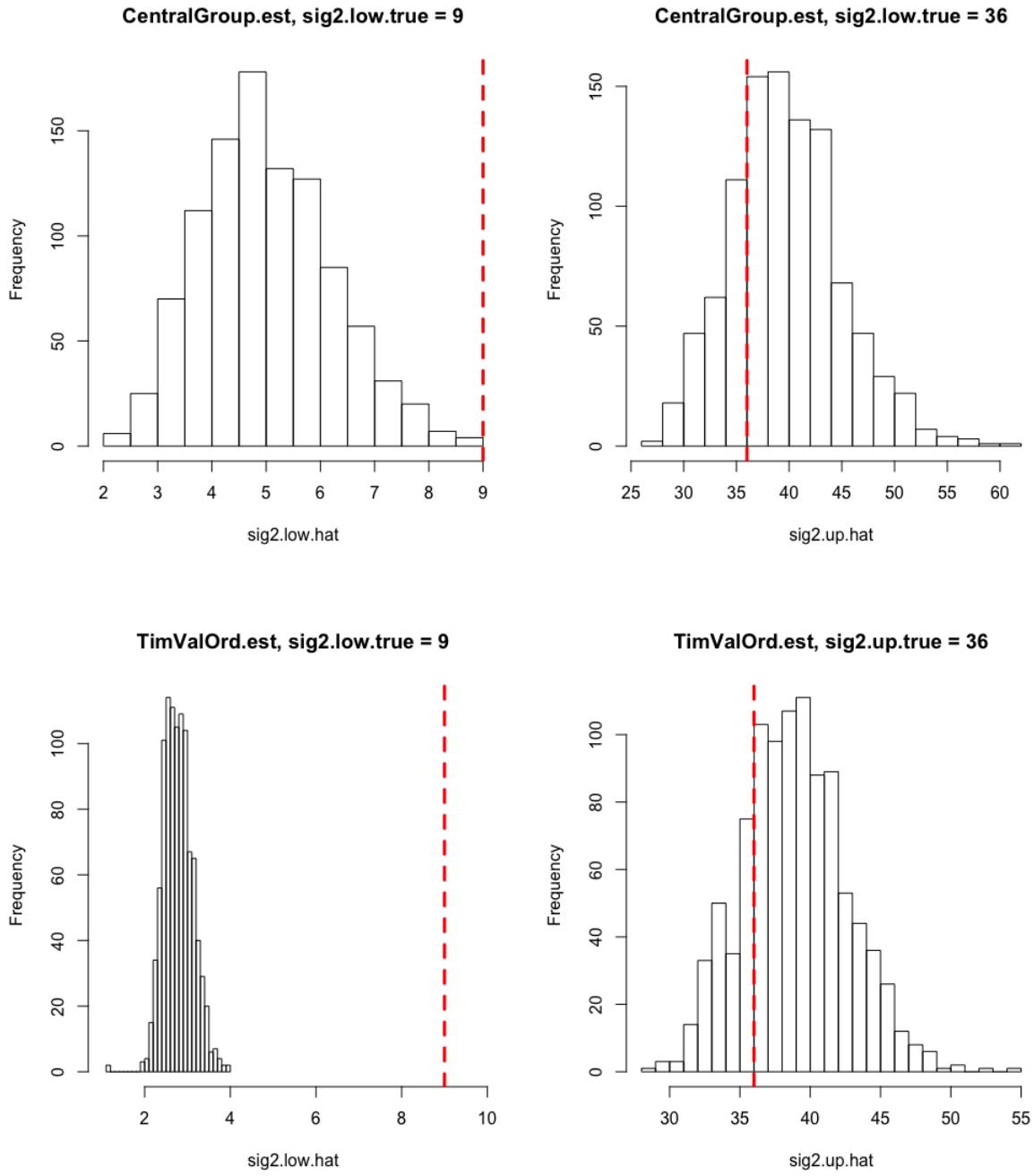


Figure A.2: Sampling distribution of the time-value ordering estimators with small true parameters

Figure A.2 shows that when the left-point estimations are much smaller than the true value, which is the similar pattern to Figure 3.11, indicating that when the true parameter is small, the time-value ordering method still has the "biased" feature when estimating $\underline{\sigma}^2$.

A.1.2 Large true parameters

Consider a relatively large parameter interval $[\underline{\sigma}, \overline{\sigma}] = [2, 3]$ ($[\underline{\sigma}^2, \overline{\sigma}^2] = [4, 9]$). We implement the central group and time-value ordering rule, by replicating the experiment 1000 times, to get the following 'sampling distribution" of estimators. We can see that for large variance, since the data points generally have much more "variation", our estimators have larger variance around the true parameters, but they still achieve our goal which is simply to let the estimators cover the uncertainty interval with a relatively narrow range (compared with the most points on the original max-mean curves).



A.2 Some R codes we use in the thesis

This section will provide some useful R codes (mainly written as functions) as an inspiring practical example for readers who are interested to apply them to reproduce and check some results in this thesis even further explore and play with these algorithms.

```
#Chapter1: Iterate method based on the semi-G-normal distribution
```

```
##1-dim
##function: Gheat.itn.int: using integration
##(can also be equivalently changed to MC method)
##gives the complete solution for u(t,x)
Gheat.itn.sol <- function(varphi.org, n=20,
                          T1=1, a=1, b=2,
                          x.step = .02,
                          x.range.scale = 2,
                          v.opt.mat=NULL, varphi.list=NULL,
                          splinefun.my = splinefun.my,
                          opt.my = opt.my.tri,
                          csv.name="test", save.ind=FALSE){
  #MC.size = 2e3,
  #"fmm" may not be good at extrapolation; but I checked that it works for x^n
  varphi.list <- vector("list", n+1)
  varphi.list[[1]] <- varphi.org
  x.min <- -x.range.scale*b; x.max <- x.range.scale*b;
  x.seq <- seq(x.min, x.max, x.step)
  y.seq <- numeric(length(x.seq))
  varphi.mat<- matrix(NA,nrow = length(x.seq), ncol = n+1)
  varphi.mat[,1] <- varphi.list[[1]](x.seq)
  #save each v.opt
  v.opt.mat <- matrix(NA, nrow = (n-1)+2, ncol = length(x.seq)+2)
  #also save the optimal v for the last x and last iteration,
  #although it may not be very useful
  v.opt.mat[,1] <- b
    for (i in seq_len(n)){
      #the iteration step i
      temp <- varphi.list[[i]]
      for (j in seq_along(x.seq)){
        #the jth value of x.seq
        #v.init <- v.opt.mat[i,j]
        x1 <- x.seq[j]
        #define the expt of N(x,v^2); x is fixed and v is the arg.
        expt.v.fn <- function(v){
          if(v==0){
            return(x1)
          } else {
            f <- function(m) temp(x1+m)*(sqrt(n)/v)*dnorm(sqrt(n)*m/v)
          integrate(Vectorize(f), -Inf, Inf)[[1]]
          }
        }
        #opt.list <- opt.my(v.init, expt.v.fn, method = "L-BFGS-B",
        lower = a, upper = b)
        #y.seq[j] <- opt.list$value * (-1)
```

```
        #v.opt.mat[i,j]<- v.opt.mat[i,(j+1)] <- opt.list$par
        opt.list <- opt.my(expt.v.fn)
        y.seq[j] <- opt.list$value
        v.opt.mat[i,j] <- opt.list$par
        #update the v.opt at (i,j), guess the same for (i,j+1)
      }
      varphi.mat[,(i+1)] <- y.seq
      varphi.list[[i+1]]<- splinefun.my(x.seq, y.seq)
    }
    if(save.ind){
      #file.name <- paste0(csv.name, format(Sys.time(), "%Y-%m%d-%H%M%S"), ".csv")
      file.name <- paste0(csv.name, ".csv")
      write.csv(cbind(x.seq, varphi.mat), file = file.name)
    }
    u.tx <- varphi.list[[n+1]]
    value <- varphi.list[[n+1]](0)
    return(list(value=value,
                varphi.mat = varphi.mat,
                varphi.list = varphi.list,
                v.opt.mat = v.opt.mat,
                x.seq=x.seq,
                u.tx = u.tx))
}

##function:summary.Gheat
##this will give all the visualization and numerical results we need.
##avoid assign the variable within the function environment
##t,x,n,a,b can be assigned in the arg list
summary.Gheat <- function(x.range.scale = 5, x.step = .02,
                          save.ind=FALSE,
                          save.global.ind = FALSE,
                          plot.ind=TRUE, plot.step=4, plot.x.range=3,
                          read.data = TRUE){
  #if we do not have the result, create it
  sd.name <- paste0("a",a.name,"b",b.name)
  time.name <- paste0("time.",varphi.name,".",sd.name,".",as.character(n))
  re.name <- paste0("re.",varphi.name,".",sd.name,".",as.character(n))
  csv.name <- paste0(varphi.name,"-",sd.name,"-",as.character(n))
  #delete the last -, we don't need add the time for now.
  if (!read.data){
  #for many arguments, the order may be different,
  so write x=x, to avoid the order problems
  time <- system.time({re <- Gheat.itn.sol(varphi.org=varphi.org,
                       n=n, a=a, b=b,
                       x.range.scale=x.range.scale, x.step=x.step,
```

```
                        save.ind=save.ind, csv.name=csv.name
                        )})
  #method.spline = method.spline
  if(save.global.ind){
    assign(time.name,time, envir = .GlobalEnv)
    assign(re.name, re, envir = .GlobalEnv)
  }

  } else {
    file.name <- paste0(csv.name, ".csv")
    re.data <- read.csv(file.name)
    x.seq <- re.data$x.seq
    varphi.mat <- re.data[,-c(1,2)]
    n1 <- n+1
    varphi.n1 <- splinefun(x.seq, varphi.mat[,n1],
                           method = method.spline)
    re <- list(x.seq=x.seq, varphi.mat=re.data[,-c(1,2)],
            value=varphi.n1(0))
  }
  #if we want to plot, plot it
  if (plot.ind){
    n1 <- n+1
    x.seq <- re$x.seq
    ind <- which(abs(x.seq)<plot.x.range)
    ind.y <- unique(c(seq(1,n1,plot.step),n1))
    matplot(x.seq[ind], re$varphi.mat[ind,ind.y], type = "l",
            main=paste0("varphi=", varphi.name,
                        ", sig.low=",a,", sig.up=",b),
            xlab = "x", ylab = "varphi_i (x)")
    abline(h=re$value, col="brown", lty=2)
    abline(v=0, col="green",lty=3)
  }
  return(re$value)
}


#parameter setting
Nt=1e2;
dt <- T1/Nt
#grid points for t
t.seq <- seq(0,T1,dt)

#the range of x is range.x
#grid points for x
#Nx=5e3;
```

```
#dx <- (range.x[2]-range.x[1])/Nx
#dt*b^2 #we want to keep it stable
dx <- 0.2
Nx <- (range.x[2]-range.x[1])/dx
x.seq <- seq(range.x[1], range.x[2], dx)

##An example
a <- 0.5; b <- 1;
n <- Nt
varphi.org <- function(x) x^3
varphi.name <- "x3"
a.name <- "05"; b.name <- "1"
summary.Gheat(x.range.scale = range.x[2]/b, x.step = dx,
              save.ind = TRUE, plot.x.range = 2, plot.step = 8,
              save.global.ind = TRUE,
              plot.ind = TRUE, read.data = FALSE)

##2-dim
###parameters
dim = 2;
#correlation uncertainty, rou12
rou1 <- -1/2; rou2 <- 1/2;
#variance uncertainty sigx, sigy
sig1 <- c(1/2, 1/2); #(sigx.low, sigy.low)
sig2 <- c(1,1) #(sigx.up, sigy.up)
#e.g. rou = 0.5, sig = c(1,1) (sigx, sigy)
x.step = .1; v.step=.1
x.range.scale = c(5,5); MC.size = 1e4;
x.size = 1e3;
x.range.scale.test = c(8,8);
n=10 #number of iterations
T1=1 #terminal time point T
rou <- 1/10
sig <- c(1,1)/2

###basic functions

varphi.org <- function(x) sum(x^3)

make.cov.mat <- function(sig, rou){
  #use rou, sig to create a cov matrix
  dim <- length(sig)
  corr.mat <- diag(rep(1,dim))
  corr.mat[upper.tri(corr.mat)] <- rou
  corr.mat[lower.tri(corr.mat)] <- t(corr.mat)[lower.tri(corr.mat)]
```

```
  diag(sig) %*% corr.mat %*% diag(sig)
}

#cov.mat <- diag(sig) %*% corr.mat %*% diag(sig)
sqrt.mat <- function(cov.mat){
  e <- eigen(cov.mat)
  V <- e$vectors
  V %*% diag(sqrt(e$values)) %*% t(V)
}

create.x.mat <- function(x.range.sc, sig, x.step){
  x.min <- -x.range.sc*sig; x.max <- x.range.sc*sig;
#we have x1,x2, ..., xd
#let d=2
  x.seq1 <- seq(x.min[1], x.max[1], x.step)
  x.seq2 <- seq(x.min[2], x.max[2], x.step)
  len.x1 <- length(x.seq1)
  len.x2 <- length(x.seq2)
  x.mat <- cbind(rep(rep(x.seq1,each=len.x2), 1),
              rep(rep(x.seq2,each=1), len.x1))
  list(x.mat=x.mat, x.seq1=x.seq1, x.seq2=x.seq2)
}

##function: splinefun.d

splinefun.d <- function(x.mat, y){
#dim = 2
#Y.seq <- apply(x.mat,1,varphi.org)
xy.dat <- as.data.frame(cbind(x.mat, y))
#choose the spline function
spline.test <- gam(y ~te(V1,V2, bs = "tp"), data = xy.dat)
##spline.test <- gam(y ~s(V1,V2, bs = "tp"), data = xy.dat)
##spline.test02 <- gam(formula = paste0("Y.seq","~","s(V1,V2)"), data = X)
##spline.test1 <- gam(Y.seq ~s(V1, bs = "cr")+s(V2, bs = "cr"), data = X)
varphi.pred <- function(x){
  x <- as.data.frame(t(x))
  predict(spline.test, x)
}
return(list(varphi.pred=varphi.pred, values.fit=spline.test$fitted.values))
}


##write into a function
###function:Gheat.itn.sol.d
```

```
Gheat.itn.sol.d <- function(dim = 2, rou1 = -1/2, rou2 = 1/2,
                            sig1 = c(1/2,1/2), sig2 = c(1,1),
                            varphi.org = function(x) sum(x^3),
                            x.step =.1,
                            x.range.scale = c(5,5),
                            MC.size = 1e3, x.size = 1e3,
                            n=10, T1=1,
                            xmat.input = FALSE, x.mat.user = NULL,
                            save.ind = TRUE, opt.method = "L-BFGS-B",
                            varphi.name = "x3", sd.name = "sig0501rou05"){
#Preparation
if(xmat.input){
  #if x.mat is input by user, then use x.mat.user
  #xmat.input = TRUE, user must specify the x.mat.user
  x.mat <- x.mat.user
} else {
#  x.mat <- rbind(rmvnorm(x.size, mean = rep(0,dim), sigma = diag(sig1)),
rmvnorm(x.size, mean = rep(0,dim), sigma = diag(2*sig2)))
#it may be too spread out
  x.mat <- rmvnorm(x.size, mean = rep(0,dim), sigma = diag(sig2))
}

#y.seq
y.seq <- numeric(nrow(x.mat))

#list of varphi's (for each iteration)
#varphi.list <- vector("list", n+1)
#varphi.list[[1]] <- varphi.org

#list of gam models
gam.list <- vector("list", n+1)

#arg-value matrix
varphi.mat<- matrix(NA,nrow = length(y.seq), ncol = n+1)
varphi.mat[,1] <- apply(x.mat, 1, varphi.org)

#save each v.opt
v.opt.mat <- array(NA, dim = c((n-1)+2, nrow(x.mat) +2,
length(sig1)+length(rou1)))

#main procedure
mc.sample <- rmvnorm(MC.size,
                     mean = rep(0,dim), sigma = diag(rep(1,dim)))
v.init1 <- c(sig2, rou2)
#set MC grid points
```

```
#initialize the guess for the first x point at each iteration.
v.opt.mat[,1,] <- matrix(rep(v.init1, nrow(v.opt.mat[,1,])),
nrow = nrow(v.opt.mat[,1,]),  byrow = TRUE)
i=1;
for (j in seq_len(nrow(x.mat))){
  cat("i=", i," ", "j=", j,"\n")
  #the jth value of x.seq
  v.init <- v.opt.mat[i,j,]
  x1 <- x.mat[j,]
  #define the expt of N(x,v^2); x is fixed and v is the arg.
  #mvnorm
  expt.v.fn <- function(v){
    sig <- v[1:dim]; rou <- v[-(1:dim)]
    cov.mat <- make.cov.mat(sig=sig, rou = rou)
    -mean(apply(x1 + mc.sample %*% (sqrt.mat(cov.mat=cov.mat)/sqrt(n)),
                1, varphi.org))
        }
  opt.list <- optim(v.init, expt.v.fn, method = opt.method,
                    lower = c(sig1,rou1), upper = c(sig2,rou2))
  y.seq[j] <- opt.list$value * (-1)
  #update the v.opt at (i,j), guess the same for (i,j+1)
  v.opt.mat[i,j,] <- v.opt.mat[i,(j+1),] <- opt.list$par
  }
varphi.mat[,(i+1)] <- y.seq
xy.dat <- as.data.frame(cbind(x.mat, y.seq))
gam.list[[i+1]] <- gam(y.seq ~te(V1,V2, bs = "tp"), data = xy.dat)

for (i in seq_len(n)[-1]){
    #the iteration step i
    gam.temp <- gam.list[[i]]
    for (j in seq_len(nrow(x.mat))){
      #j=1
      cat("i=", i," ", "j=", j,"\n")
      #the jth value of x.seq
      v.init <- v.opt.mat[i,j,]
      x1 <- x.mat[j,]
      #define the expt of N(x,v^2); x is fixed and v is the arg.
      #mvnorm
      expt.v.fn <- function(v){
          sig <- v[1:dim]; rou <- v[-(1:dim)]
          cov.mat <- make.cov.mat(sig=sig, rou = rou)
          new.dat <- as.data.frame(x1 + mc.sample %*%
           (sqrt.mat(cov.mat=cov.mat)/sqrt(n)))
          -mean(predict.gam(gam.temp, new.dat))
```

```
        }
      opt.list <- optim(v.init, expt.v.fn, method = opt.method,
                        lower = c(sig1,rou1), upper = c(sig2,rou2))
      y.seq[j] <- opt.list$value * (-1)
      #update the v.opt at (i,j), guess the same for (i,j+1)
      v.opt.mat[i,j,] <- v.opt.mat[i,(j+1),] <- opt.list$par
      }
    varphi.mat[,(i+1)] <- y.seq
    #store the spline.test object
    xy.dat <- as.data.frame(cbind(x.mat, y.seq))
    gam.list[[i+1]]<- gam(y.seq ~te(V1,V2, bs = "tp"), data = xy.dat)
}
    result <- list(varphi.mat = varphi.mat,
                gam.list = gam.list,
                v.opt.mat = v.opt.mat,
                x.mat=x.mat)
    if(save.ind){
  #re.name <- paste0("re.",varphi.name,".",sd.name,".",as.character(n))
  list.name <- paste0(varphi.name,"-",sd.name,"-",as.character(n),"-")
  file.name <- paste0(list.name, format(Sys.time(), "%Y-%m%d-%H%M%S"), ".Rdata")
  list.save(result, file = file.name)
}
    return(result)

   }


###function:summary.Gheat.d

#run the whole function then test the time
#consider the time of each iteration if needed
summary.Gheat.d <- function(result = NULL,
                            x.sc.new = c(2,2), k.plot.seq = NULL,
                            sig.new = c(1,1),
                            x.step.new =.1,
                            varphi.org = function(x) sum(x^3),
                            result.plot = NULL,
                            plot.ind = TRUE, alpha = .75){
if(is.null(result.plot)){
  gam.list <- result$gam.list
  varphi.mat <- result$varphi.mat
  x.mat <- result$x.mat

  re.x <- create.x.mat(x.range.sc=x.sc.new, sig=sig2,
                       x.step=x.step.new)
```

```
  x.mat.new <- re.x$x.mat
  x.seq1.new <- re.x$x.seq1
  x.seq2.new <- re.x$x.seq2

  varphi.mat.new <- matrix(NA,nrow = nrow(x.mat.new),
                           ncol = ncol(varphi.mat))
  varphi.mat.new[,1] <- apply(x.mat.new, 1, varphi.org)
  x.mat.new.dat <- as.data.frame(x.mat.new)
  for (k in seq_len(ncol(varphi.mat))[-1]){
  varphi.mat.new[,k] <- predict.gam(gam.list[[k]], x.mat.new.dat)
  }
  if(plot.ind){
k = 1;
alpha1 = alpha;
persp3d(x.seq1.new, x.seq2.new, varphi.mat.new[,k], col=k,
alpha = alpha1, xlab = "x1", ylab = "x2", zlab = "varphi_i(x1,x2)")
if(is.null(k.plot.seq)) k.plot.seq <- ncol(varphi.mat)
for (k in k.plot.seq){
persp3d(x.seq1.new, x.seq2.new, varphi.mat.new[,k], col=k,
alpha = alpha1, add = TRUE,xlab = "x1", ylab = "x2", zlab = "varphi_i(x1,x2)")
}
}
  return(list(x.mat.new = x.mat.new,
              varphi.mat.new = varphi.mat.new))
} else {
  x.mat.new <- result.plot$x.mat.new
  varphi.mat.new <- result.plot$varphi.mat.new
  if(plot.ind){
k = 1;
alpha1 = alpha;
persp3d(x.seq1.new, x.seq2.new, varphi.mat.new[,k], col=k,
alpha = alpha1, xlab = "x1", ylab = "x2", zlab = "varphi_i(x1,x2)")
if(is.null(k.plot.seq)) k.plot.seq <- ncol(varphi.mat)
for (k in k.plot.seq){
persp3d(x.seq1.new, x.seq2.new, varphi.mat.new[,k], col=k,
alpha = alpha1, add = TRUE,xlab = "x1", ylab = "x2", zlab = "varphi_i(x1,x2)")
}
}
}

                            }


#Chapter2: Estimation of variance uncertainty
```

```
##function: max.mean (time order)
max.mean <- function(y, n.guess){
m <- floor(length(y)/n.guess)
y <- y[1:(m*n.guess)]
y.mat <- matrix(y, ncol = n.guess, byrow = TRUE)
mean.seq <- apply(y.mat, 1, mean)
mu.low.est <- min(mean.seq)
mu.up.est <- max(mean.seq)
c(mu.low.est, mu.up.est)
}

##function: max.mean2 (value order)
max.mean2 <- function(y, n.guess){
m <- floor(length(y)/n.guess)
y <- y[1:(m*n.guess)]
y.mat <- matrix(y[order(y)], ncol = n.guess, byrow = TRUE)
mean.seq <- apply(y.mat, 1, mean)
mu.low.est <- min(mean.seq)
mu.up.est <- max(mean.seq)
c(mu.low.est, mu.up.est)
}

##Improvement: how to choose the group size
###function: CentralGroup.est
CentralGroup.est <- function(y, n.seq = NULL,
                             plot.ind=TRUE, par.true=c(1,4)){
center <- mean(y)
N <- length(y)
#y.ord <- y[order(y)]
#y.ord[N/2]
#y.ord[N/2+1]
#median(y.ord)
if(is.null(n.seq)){
  n.seq <- seq(2,floor(N/2),2)
}
est.all.mat <- matrix(nrow = length(n.seq), ncol = 4)
#i <- 100
for (i in seq_along(n.seq)){
n <- n.seq[i]
m <- floor(N/n)
y.mat <- matrix(y[1:(m*n)], ncol = n, byrow = TRUE)
mean.seq <- apply(y.mat, 1, mean)
min.seq <- apply(y.mat, 1, min)
max.seq <- apply(y.mat, 1, max)
mu.low.est <- min(mean.seq)
```

```
mu.up.est <- max(mean.seq)
est.all.mat[i,1:2] <- c(mu.low.est, mu.up.est)

l.gr <- max(min.seq[min.seq<=center])
r.gr <- min(max.seq[max.seq>=center])
est.all.mat[i,3:4] <- c(l.gr, r.gr)
}
L <- est.all.mat[,1]
R <- est.all.mat[,2]
l <- est.all.mat[,3]
r <- est.all.mat[,4]
ind1 <- min(which(R-r <= 0))
ind2 <- min(which(l-L <= 0))
n1 <- n.seq[ind1]; n2 <- n.seq[ind2]
a2.est <- L[ind2]
b2.est <- R[ind1]
if(plot.ind){
  matplot(n.seq, est.all.mat, type = "l",
          ylab = "values", xlab = "group size n",
          main = "CentralGroup Estimation")
  abline(h=par.true, col="brown", lty=2)
  legend("top",
         c("GroupMean.min", "GroupMean.max",
           "CentralGroup.min", "CentralGroup.max",
           "par.true"),
         col=c(1:4,"brown"), lty=c(1:4,2),
         lwd=2, cex=0.4, box.lty = 2, box.col = "grey",
         pch = 25)
}
return(list(est.all.mat = est.all.mat, ab.est = c(a2.est, b2.est),
     n.seq = n.seq))
}

###function: TimValOrd.est
TimValOrd.est <- function(y, step=10, plot.ind=TRUE,
                          min.n=1, min.m=2,
                          choprule="both.n",
                          par.true=c(0,5),
                          start=1, len.frac=1,
                          y.scale=1){
  #time ordering
  #c("both.n", "min.n", "max.n","fix.n")
  #min.n
  #min.m = min number of groups
y <- y*y.scale
```

```
N <- length(y)
n.seq <- seq(min.n, N/min.m, step)
est.mat <- matrix(0, ncol=2, nrow=length(n.seq))
for (i in seq_along(n.seq)){
  est.mat[i,] <- max.mean(y, n.seq[i])
}
est.mat2 <- matrix(0, ncol=2, nrow=length(n.seq))
for (i in seq_along(n.seq)){
  est.mat2[i,] <- max.mean2(y, n.seq[i])
}
if(plot.ind){
   len <- floor(length(n.seq)*len.frac)
  ind <- start+seq_len(len)-1
  n.seq1 <- n.seq[ind]
  sum2 <- n.seq1[1]+n.seq1[length(n.seq1)]
  n.lab <- seq(0, ceiling(N/(min.m*100))*100, step*10)
  n.lab[1] <- 1
  n.lab1 <- unique(floor(N/seq(min.m, N, 1)))
  m.lab <- floor(N/n.lab1)
  matplot(n.seq[ind],cbind(est.mat[rev(ind),], est.mat2[ind,]),
          type = "l",
          col = c(2,2,4,4),
        ylab = paste("estimation *", y.scale), xlab = "",
        main="Est for different group sizes",
        xaxt='n')
  #axis for the n.seq[rev(ind)]
  legend("top",
         c("TimOrd.min  ", "TimOrd.max  ",
           "ValOrd.min  ", "ValOrd.max  ",
           "par.true"),
         col=c(2,2,4,4,"brown"), lty=1:5,
         lwd=2, cex=0.4)
  axis(1, at = sum2 - n.lab, labels = n.lab, line = 1,
       col = 2, col.ticks = 2, col.axis = 2)
  mtext("n.rev", 1, line = 1, at = -40, col = 2)
  #axis for the n.seq[ind]
  axis(1, at = n.lab, line = 3,
       col=4, col.ticks=4, col.axis=4)
  mtext("n", 1, line=3, at=-40, col=4)
  #par.true
  abline(h=par.true[1], col="brown", lty = 5)
  abline(h=par.true[2], col="brown", lty = 5)
}
n1 <- n.seq[sum(est.mat2[,2] - est.mat[rev(seq_along(n.seq)),2] > 0)]
n2 <- n.seq[sum(est.mat[rev(seq_along(n.seq)),1] - est.mat2[,1] > 0)]
```

```
n12 <- c(n1,n2); n12.ord <- n12[order(n12)]
min.n <- n12.ord[1]; max.n <- n12.ord[2]
n.fix <- n.seq[length(n.seq)]
if(choprule=="min.n"){
  return(max.mean2(y, min.n)/y.scale)
} else if(choprule == "max.n"){
  return(max.mean2(y, max.n)/y.scale)
} else if(choprule == "both.n"){
  return(rbind(c(max.mean2(y, n1)/y.scale, n1),
               c(max.mean2(y, n2)/y.scale, n2)))
} else if(choprule == "fix.n"){
  return(max.mean2(y, n.fix)/y.scale)
}
}
```

```
#Chapter3: Pseudo simulation of variance uncertainty

##function: rmaximal
##(simulation of maximal distribution, several candidates)
rmaximal <- function(N, n=1e2, a = 1, b = 2, len.sd=3) {
  result <- numeric(N)
  count <- 0
  while (count < N) {
    #random decide the length
    #len <- rpois(1, n)
    len <- floor(rnorm(1, mean = n, sd = len.sd))
    alpha <- runif(1, 0,50)
    beta <- runif(1, 0,50)
    c <- rbeta(1, alpha, beta) *(b-a) + a
    #dirac distributions (constant)
    newvals <- numeric(len)+c
    result[count + seq_along(newvals)] <- newvals
    #make some space for the newvals
    count <- count + length(newvals)
  }
  result[seq_len(N)]
}

rmaximal1 <- function(n, a = 1, b = 2) {
  result <- numeric(n)
  count <- 0
  while (count < n) {
```

```
    #random decide the length
    lambda <- sample(1:1e3, 1, prob = rep(1,1e3)/1e3)
    len <- rpois(1, lambda)
    alpha <- runif(1, 0,50)
    beta <- runif(1, 0,50)
    c <- rbeta(1, alpha, beta) *(b-a) + a
    #dirac distributions (constant)
    newvals <- numeric(len)+c
    result[count + seq_along(newvals)] <- newvals
    #make some space for the newvals
    count <- count + length(newvals)
  }
  result[seq_len(n)]
}

rmaximal2 <- function(n, a = 1, b = 2) {
  result <- numeric(n)
  count <- 0
  while (count < n) {
    #random decide the length
    lambda <- sample(1:1e3, 1, prob = rep(1,1e3)/1e3)
    len <- rpois(1, lambda)
    alpha <- runif(1, 0,50)
    beta <- runif(1, 0,50)
    newvals <- rbeta(len, alpha, beta) *(b-a) + a
    result[count + seq_along(newvals)] <- newvals
    #make some space for the newvals
    count <- count + length(newvals)
  }
  result[seq_len(n)]
}

rmaximal3 <- function(n, a = 1, b = 2) {
  result <- numeric(n)
  count <- 0
  while (count < n) {
    #random decide the length
    lambda <- floor(1/runif(1))
    #lambda has no upper bound
    len <- rpois(1, lambda = lambda)
    alpha <- runif(1, 0,50)
    beta <- runif(1, 0,50)
    c <- rbeta(1, alpha, beta) *(b-a) + a
    #dirac distributions (constant)
    newvals <- numeric(len)+c
```

```
    result[count + seq_along(newvals)] <- newvals
    #make some space for the newvals
    count <- count + length(newvals)
  }
  result[seq_len(n)]
}

rmaximal4 <- function(n, a = 1, b = 2) {
  result <- numeric(n)
  count <- 0
  while (count < n) {
    #random decide the length
    lambda <- floor(1/runif(1))
    #lambda has no upper bound
    len <- rpois(1, lambda = lambda)
    alpha <- runif(1, 0,50)
    beta <- runif(1, 0,50)
    newvals <- rbeta(len, alpha, beta) *(b-a) + a
    result[count + seq_along(newvals)] <- newvals
    #make some space for the newvals
    count <- count + length(newvals)
  }
  result[seq_len(n)]
}

rf1 <- function(n) rnorm(n, mean = b2, sd = 2)
rf2 <- function(n) rexp(n, rate = 2/(a2+b2))
rf3 <- function(n) runif(n, min = a2, max = b2)
rmaximal.list <- function(n, a = 1, b = 4,
                      density.set = list(rf1, rf2, rf3)) {
  result <- numeric(n)
  count <- 0
  L <- length(density.set)
  #size of density.set
  while (count < n) {
    ind <- sample(1:L, 1)
    #random decide the length
    lambda <- floor(1/runif(1))
    #lambda has no upper bound
    len <- rpois(1, lambda = lambda)
    newvals <- density.set[[ind]](len)
    newvals <- newvals[a < newvals & newvals < b]
    result[count + seq_along(newvals)] <- newvals
    #make some space for the newvals
    count <- count + length(newvals)
```

```
  }
  result[seq_len(n)]
}
##function: rsemiGnorm
rsemiGnorm <- function(N, n, sig.low=1, sig.up=2,
                       rmaximal.k=rmaximal){
  #choose the rmaximal function
  #(pseudo sim of nl.iid maximal distn)
  rnorm(n) * rmaximal.k(N=N, n=n, a=sig.low, b=sig.up)
}

##function: rGnorm
##from nl.CLT, generate in approximated sense
rGnorm <- function(n, sig.low=1, sig.up=2, group.size = 50,
                       rmaximal.k=rmaximal){
  N <- group.size
  replicate(n, {
    w.seq <- rsemiGnorm(N, sig.low = sig.low, sig.up = sig.up,
                      rmaximal.k = rmaximal.k)
    1/sqrt(N) * sum(w.seq)
  })
}
```

Curriculum Vitae

Name:	Yifan Li
Post-Secondary Education and Degrees:	Shandong University Jinan, Shandong, China 2012 - 2016 Bachelor of Science, Pured and Applied Mathematics University of Western Ontario London, ON, Canada 2016 - 2018 Thesis-based Master of Science, Statistical Theory (to be defenced in July, 2018)
Honours and Awards:	Loo-Keng Hua Undergraduate Scholarship from the Chinese Academy of Sciences The First Prize in 2014 China Undergraduate Mathematical Contest
Related Work Experience:	Teaching Assistant University of Western Ontario 2016 - 2018

Publication:

An Iterative Approximation for the Sublinear Expectation of G-normal Distribution and the Solution of the Corresponding G-heat Equation
Yifan Li, Reg Kulperger
Preprinted at arXiv:1804.10737 (2018), to be published

Working Papers:

The Improved Max-mean Estimation of Variance Uncertainty in the G-expectation Framework
Yifan Li, Reg Kulperger

The Localized Method and 2BSDE (Second Order BSDE) Method for G-heat Equations with Covariance Uncertainty
Yifan Li, Nathan Gold, Huaxiong Huang

Linear Processes from G-normal Distribution under Sublinear Expectation
Reg Kulperger, Yifan Li, Hao Yu, Defei Zhang

Invited Presentations:

An Iterative Approximation of the G-expectation of an Arbitrary Function of G-normal Distribution and the Solution to the Corresponding G-heat Equation
Shanghai Jiao Tong University, Shanghai, China April 2018
The Fourth Young Researchers Meeting on BSDEs, Nonlinear Expectations and Mathematical Finance

Functional Data Analysis of Dynamic Changes in Air Quality with Spatial Factors
Wuhan University, Wuhan, China May 2015
The Annual Meeting and Mathematical Research Seminar held by NSFC (National Natural Science Foundation of China)

Dynamic Monitoring System around Incineration Plants and Financial Compensation Design
Shenzhen University, Shenzhen, China August 2014
Modeling Contest Summer Seminar held by CSIAM (China Society for Industrial and Applied Mathematics)

Contributed Presentations:

An Iterative Approximation for the Sublinear Expectation of G-normal Distribution and the Solution of the Corresponding G-heat Equation (Poster)
McGill University, Montreal, Canada June 2018
The 46th Annual Meeting of the Statistical Society of Canada