MULTISCALE FRACTALITY WITH APPLICATION AND STATISTICAL MODELING AND ESTIMATION FOR COMPUTER EXPERIMENT OF NANO-PARTICLE FABRICATION

A Thesis
Presented to
The Academic Faculty

by

Hin Kyeol Woo

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2012

MULTISCALE FRACTALITY WITH APPLICATION AND STATISTICAL MODELING AND ESTIMATION FOR COMPUTER EXPERIMENT OF NANO-PARTICLE FABRICATION

Approved by:

Professor Jye-Chyi Lu, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Brani Vidakovic, Advisor
Department of Biomedical Engineering
Georgia Institute of Technology

Professor Martha Grover
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Professor Jianjun Shi
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Kamran Paynabar
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: 22 August 2012

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my appreciation to all who have influenced, stimulated, and warmly supported my work in various ways.

First and foremost, I would like to express my deep and sincere gratitude to my advisors, Professor Jye-Chyi Lu and Professor Brani Vidakovic, for their guidance, assistance, encouragement, and hearty support at all phase of my doctoral program. They took care of my both academically and personally in every conceivable way. Additionally, I would like to express my appreciation to my co-advisor, Professor Martha Grover, for her wonderful advice and comments. Besides my advisors, I would like to thank the rest of my thesis committee: Professor Jianjun Shi and Professor Kamran Paynabar for their encouragement, insightful comments, and hard questions

I would like to thank all the staff members of ISyE, Georgia Tech, especially Pamela Morrison, Mark Reese, and Harry Sharp, for their kind support and help at each and every stage of my graduate life at Georgia Tech. Special thanks goes to my friends, Sunyoung, Joo-Young, Seonghye, Haewon, Jeff, Milim, Chuljin, Donggu, Lynn, Seungbae, Margaret, and Moojoong for enriching my life at Georgia Tech. I would also like to thank Sungil, Nagesh, Justin, Andres, Sky, and Jongphil for their academic help.

Last, but by no means the least, my heartfelt appreciation and gratitude goes to my family, especially my father, Mr. Jongku Woo, and mother, Ms. Aesook Jeong. I sincerely appreciate their constant support and encouragement. Thank you and love you as always.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Chapter 1 proposes multifractal analysis to measure inhomogeneity of regularity of ^{1}H-NMR spectrum using wavelet-based multifractal tools. The geometric summaries of multifractal spectrum are informative summaries, and as such employed to discriminate ^{1}H-NMR spectra associated with different treatments. The novel summaries are based on the descriptors, originally introduced by Shi et al. (2005). The methodology is applied to evaluate the effect of sulfur amino acids. With only six univariate summaries, two statistical models capture the effect of sulfur amino acids.

Chapter 2 provides essential materials for understanding engineering background of a nano-particle fabrication process. In particular, certain physics of the engineering process are described and process outcomes are quantified. Several noise factors contributing to process uncertainties are identified. Preliminary analyses based on data obtained from computer simulations of physical experiment show the potential of further statistical modeling research opportunities.

Chapter 3 develops a two-part model for observations from nano-particle fabrication experiments. Since there are certain combinations of process variables resulting to unproductive process outcomes, a logistic model is used to characterize such a process behavior. For the cases with productive outcomes a normal regression serves the second part of the model. Because the data are obtained from computer experiments, random-effects are included in both logistics and normal regression models to describe the potential spatial correlation among data. The likelihood function for this two-part model is complicated and thus the maximum likelihood estimation is intractable. This chapter researches approximation techniques based on Taylor series

extension to simplify the likelihood. An algorithm is developed to find estimates for maximizing the approximated likelihood.

Chapter 4 presents a method to decide the sample size under multi-layer system. The multi-layer is a series of layers, which become smaller and smaller. Our focus is to decide the sample size in each layer. The sample size decision has several objectives, and the most important purpose is the sample size should be enough to give a right direction to the next layer. Specifically, the bottom layer, which is the smallest neighborhood around the optimum, should meet the tolerance requirement. Other objectives considered are budget limit and model improvement. Performing the hypothesis test of whether the next layer includes the optimum gives the required sample size. We demonstrate an illustrative example to evaluate the proposed methodology.

CHAPTER I

MULTISCALE FRACTALITY WITH APPLICATION IN SPECTRAL CLASSIFICATION

1.1 Introduction

Many areas, such as medicine and finance, use fractal and multifractal approaches to analyze a variety of signals. The implicit occurrence of irregularities and certain degree of self-similarity over a range of scales can characterize the signals. In our research, we proposes multifractal analysis to measure inhomogeneity of regularity of ^{1}H-NMR spectrum using wavelet-based multifractal tools.

In 1990, Goldberger et al. introduced the concept of fractal dynamics in biological system [24]. Monofractal analysis has the same scaling properties throughout the entire signal. Jung et al. [32] used the monofractal methodology to transform ^{1}H-NMR spectrum. However, the methodology cannot be applied in real life because most signals generated in physiological condition are not monofractal.

We use the methodology in a health system to see the effect of sulfur amino acids. Sulfur amino acids (SAA) are involved in important aspects of human health and cellular function. Traditional methodology focuses only on specific biomakers, which are chosen after a series of pre-processing procedure, while multifractal analysis does not require the pre-processing procedure. Therefore, multifractal analysis is useful in health system.

In Section 1.2, the basic background of wavelets and the Hurst exponent is explained. Section 1.3 describes six fractality measures which will characterize a spectrum and discriminate the SAA deficiency. Section 1.4 introduces the test to distinguish a mono-fractal and multi-fractal. Section 1.5 illustrates the data set that we

used in the research. The statistical analysis is summarized in Section 1.6. Section 1.7 contains concluding remarks and future research directions.

1.2 Background

1.2.1 Scaling process

British hydrologist Harold Edwin Hurst spent 62 years in Egypt and mostly worked on design and construction of reservoirs along the Nile River. By inspecting historical data on the Nile River flows, Hurst discovered phenomenon (now called Hurst effect). Hurst was trying to find an optimal reservoir capacity R such that it can accept the river flow in N units of time,$X_1, X_2, \dots X_N$, and have a constant withdrawal of $\bar{X}$ per unit time. The optimal volume of the reservoir was given by the so called adjusted range,

$$R = \max_{1\le k\le N}(X_1 + \cdots + X_k - k\bar{X}) - \min_{1\le k\le N}(X_1 + \cdots + X_k - k\bar{X}) \tag{1}$$

Since the records for the waterflow rarely exceeded 100 years Hurst inspected other geophysical data and in order to compare them, he standardized their adjusted ranges R, with sample standard deviation

$$S = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2}\ , \tag{2}$$

and obtained dimensionless ratio R/S - rescaled and adjusted range.

On basis of more that 800 records, he found (Hurst, 1951) that quantity R/S scales as N^H, for ranging from 0.46 to 0.93, with mean 0.73 and standard deviation of 0.09.

This result was is contrast with the fact that for independent normal random variables H is 1/2 in limit. Feller proved that the limit is 1/2 for independent identically distributed random variables with finite second moment, this limit was 1/2. It was

believed that strong Markovian dependence was responsible for this deviation untill Barnard (1956) proved that limit $H = 1/2$ holds for the Markovian dependence case.

It was the work of Mandelbrot (1975), Mandelbrot and Van Ness (1968), and Mandelbrot and Wallis (1968) who associated the Hurst (or Joseph) phenomenon on the presence of long-memory.

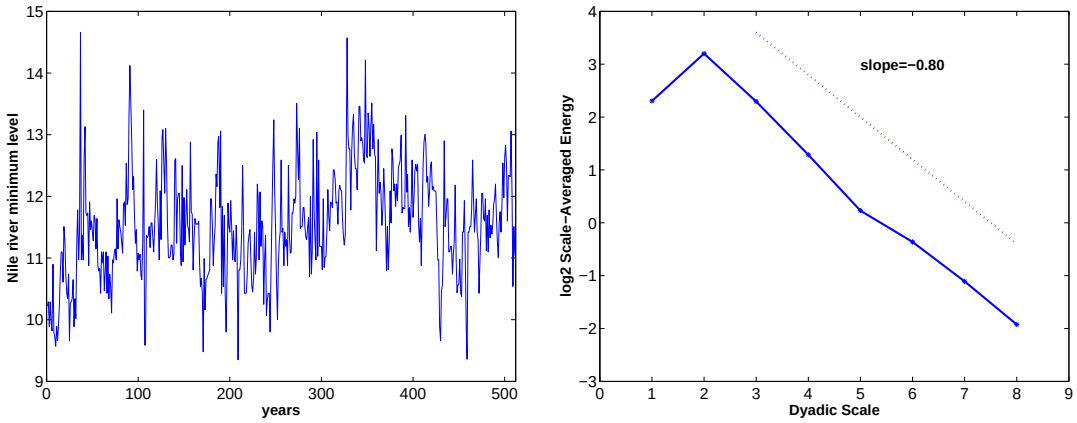

Figure 1: (a) Nile yearly minimal level data; (b) its Wavelet log spectra

Figure 1 (a) gives n=512 consecutive yearly measurements from the famous Nile River Data set for the years 62-1281 A.D. Panel (b) gives its wavelet spectra demonstrating the scaling law.

1.2.2 Wavelet

Wavelet domains are appropriate for area of scaling. Thus, we give a brief introduction to wavelet transformation. The discrete wavelet transform expresses a real signal $X(t)$ in terms of shifted and dilated versions of a wavelet (or mother) function $\psi(t)$ and shifted versions of a scaling (or father) function $\phi(t)$. For specific choices of the scaling functions and wavelets, an orthonormal basis can be formed from the atoms

$$\begin{aligned} \psi_{j,k}(t) &= 2^{j/2}\psi(2^j t - k) \\ \phi_{j,k}(t) &= 2^{j/2}\phi(2^j t - k), \quad j,k \in \mathbb{Z}. \end{aligned}$$

The signal $X(t)$ can be thus represented by wavelets as

$$X(t) = \sum_k c_{J_0,k}\phi_{J_0,k}(t) + \sum_{j=J_0}^{\infty}\sum_k d_{j,k}\psi_{j,k}(t),$$

where $d_{j,k} = \int X(t)\psi_{j,k}(t)dt$ and $c_{j,k} = \int X(t)\phi_{j,k}(t)dt$. $d_{j,k}$ and $c_{j,k}$ are detail and scaling coefficients, respect. Here, J_0 indicates the coarsest scale or lowest resolution of analysis, and a larger j corresponds to higher resolutions (detailed wavelets theories are explained in [68]). In practice, many signals are multi-dimensional. Examples include measurements in geophysics, medicine, astronomy, economics, and so on. The wavelet transform is readily generalized to the multi-dimensional case, and the generalization of wavelet to 2-D case are straightforward. The 2D wavelet bases functions are constructed via translations and dilations of a tensor product of univariate wavelets and scaling functions:

$$\begin{aligned}
\phi(t_1,t_2) &= \phi(t_1)\phi(t_2) \\
\psi^h(t_1,t_2) &= \phi(t_1)\psi(t_2) \\
\psi^v(t_1,t_2) &= \psi(t_1)\phi(t_2) \\
\psi^d(t_1,t_2) &= \psi(t_1)\psi(t_2).
\end{aligned} \tag{3}$$

The symbols h, v, d in Equation (3) stand for horizontal, vertical and diagonal directions, respectively. Consider the wavelet atoms

$$\begin{aligned}
\phi_{j,\mathbf{k}}(\mathbf{t}) &= 2^{2j}\phi(2^j t_1 - k_1, 2^j t_2 - k_2) \\
\psi^i_{j,\mathbf{k}}(\mathbf{t}) &= 2^{2j}\psi^i(2^j t_1 - k_1, 2^j t_2 - k_2),
\end{aligned}$$

for $i = h, v, d$ and where $\mathbf{t} = (t_1, t_2) \in \mathbb{R}^2$, and $\mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2$. Then, any function $X \in \mathbb{L}^2(\mathbb{R}^2)$ can be represented as

$$X(\mathbf{t}) = \sum_{\mathbf{k}} c_{J_0\mathbf{k}}\phi_{J_0,\mathbf{k}}(\mathbf{t}) + \sum_{j\geqslant J_0}\sum_{\mathbf{k}}\sum_i d^i_{j,\mathbf{k}}\psi^i_{j,\mathbf{k}}(\mathbf{t}), \tag{4}$$

where the wavelet coefficients are given by

$$d^i_{j,\mathbf{k}} = 2^{2j}\int X(\mathbf{t})\psi^i(2^j\mathbf{t} - \mathbf{k})d\mathbf{t}. \tag{5}$$

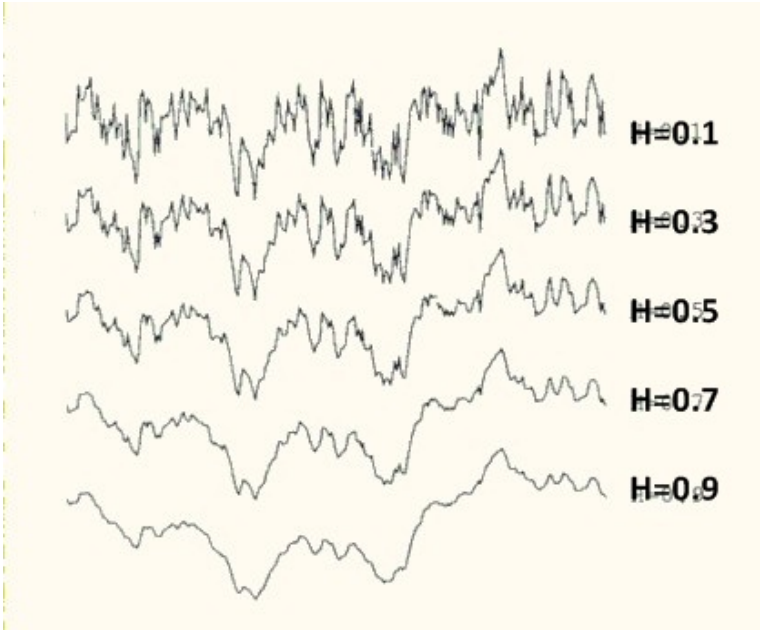

Figure 2: Irregularity defined by the Hurst exponent

1.2.3 Fractional Brownian motion (fBm)

We saw some examples of scaling in irregular and the importance of regular decay of spectra in Fourier and wavelet domains.

The Hurst exponent is used as a measure of the irregularity of time series. The value ranges from 0 to 1. The irregularity is separated by the mid point, $H = 1/2$. When the value is between 0 and 1/2, the time series has a negative autocorrelation. In this case, we say the series has "anti-persistent behavior". This means that an increase will tend to be followed by a decrease, or vice versa (irregular). When the value is between 1/2 and 1, the time series has a positive correlation (regular). Similarly, in this case, an increase is followed by an increase and a decrease is followed by a decrease. When the Hurst exponent is close to 1/2, the time series is a random walk (a Brownian time series). In a random walk, the future return values will go either up or down with 50% probability, respectively. Therefore, it is hard to predict. Figure 2 gives a good example to show that the irregularity differs by the Hurst exponent.

Mandelbrot and Van Ness [41] proposed the fractional Brownian motion(fBm). Now, it becomes one of the most important models in self-similar modeling. A random process $X(t), t > 0$ is called *self-similar* if for any $a > 0$, there exists $b > 0$ such that $X(at) \stackrel{d}{=} bX(t)$, where $\stackrel{d}{=}$ denotes the equality in distribution. Lamperti(1962) proved that a random process $X(t), t \geq 0$ which is nontrivial and stochastically continuous at 0 has an unique hurst exponent H such that $b = a^H$. Therefore, a fBm process is indexed by the self-similarity parameter, H and a *self-similar process*. In the one-dimensional case the fBm process, denoted by $\{B_H(t), t \in \mathbb{R}\}$, is characterized by the following correlation function

$$R_{B_H}(t,s) = E\{B_H(t)B_H(s)\} = \frac{\sigma_H^2}{2}\left[|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right], \tag{6}$$

where $\sigma_H^2 = \Gamma(1-2H) \cdot (\cos(\pi H))/(\pi H)$ and $0 < H < 1$. As can be seen from Equation (6), the fBm is a non-stationary process ($R_{B_H}(t,s)$ is a function of $|t-s|$), but it has stationary increments. Because of its non-stationarity, the spectrum does not exist in the strict sense. From the correlation function Equation (6) and the definition of a generalized power spectrum [51] we obtain the "power spectrum" of $B_H(t)$ as

$$S_{B_H}(\omega) = |\omega|^{-2H-1}. \tag{7}$$

These definitions can be extended to any dimension. Unlike the 1-D case the generalization of fBm to higher dimensions is not unique. A simple generalization to a 2-D surface is the *fractional Brownian field* (fBf). The fBf is a Gaussian, zero mean, random field $B_H(\mathbf{u})$, where $\mathbf{u}$ denotes the position in a selected domain, usually $[0,1] \times [0,1]$. Then, the autocorrelation function is

$$R_{B_H}(\mathbf{u},\mathbf{v}) = E\left[B_H(\mathbf{u})B_H(\mathbf{v})\right] = \frac{\sigma_H^2}{2}\left(\|\mathbf{u}\|^{2H} + \|\mathbf{v}\|^{2H} + \|\mathbf{u}-\mathbf{v}\|^{2H}\right), \tag{8}$$

where $0 < H < 1$, the variance σ_H^2 is

$$\sigma_H^2 = \frac{2^{-(1+2H)}\Gamma(1-H)}{\pi H \Gamma(1+H)} \tag{9}$$

and $\|\cdot\|$ is the usual Euclidean norm in $\mathbb{R}^2$. The increments of a fBf represent stationary, zero mean Gaussian random fields: the variance of such increments depends only on the distance $\|\mathbf{h}\|$ so that $E\left[B_H(\mathbf{u}+\mathbf{h}) - B_H(\mathbf{u})\right]^2 = \sigma_H^2 \|\mathbf{h}\|^{2H}$, where σ_H^2 is given in Equation (9). As in the one-dimensional case, the power spectrum of a fBf is obtained from the correlation function Equation (8) and the 2-D Generalized Power Spectrum and can be defined as $S_{B_H}(\omega) = \|\boldsymbol{\omega}\|^{-2H-2}$. Hurst exponent is a significance of monofractality and fBm are monofractal. Not all signals in nature are monofractals. For example, turbulence signals are categorized as multifractals, even though they have regular decay of spectra. Different tools are needed to assess the multifractality. One of these tools is multifractral spectrum.

1.2.4 Wavelet-based Multifractal Spectrum

Definition of multifractal spectrum involves limiting processes and geometric dimensions, and is calculationally involved. Wavelets provide an easy way to calculate an approximate to multifractal spectrum. The wavelet-based calculation of multifractal spectrum depends on the concepts of *partition function* and *Legendre transform*. The partition function, $T(q)$ is defined as

$$T(q) = \lim_{j \to -\infty} \log_2 E|d_{j,k}|^q \tag{10}$$

where $d_{j,k}$ is the wavelet coefficient at level j and location k, and q is the order of moments. We emphasize that q is a real number within a certain range covering the negative numbers as well. However, the interpretation of the negative moments is still not clear.

Even though (10) is very informative, the singularity measure is not explicit. It has been proposed in Gonçalvès et al (1998) that the local singularity strength could be measured in terms wavelet coefficients as :

$$\alpha(t) = \lim_{k2^j \to t} \frac{1}{j} \log_2 |d_{j,k}| \tag{11}$$

where $d_{j,k}$ is the normalized wavelet coefficient at scale j and location k. The local singularity strength measure (11) converges to the local Hölder index the process at time t. Small values of $\alpha(t)$ reflect the more irregular behavior at time t. Any inhomogeneous process has a collection of local singularity strength measures and their distribution $f(\alpha)$ forms the multifractal spectrum. A direct way to obtain this spectrum is to use the counting technique,

$$f(\alpha) = \lim_{\epsilon \to 0} \#\{\alpha(t) \,:\, \alpha - \epsilon < \alpha(t) < \alpha + \epsilon, \quad -\infty < t < \infty\}. \tag{12}$$

Although it is feasible to estimate the multifractal spectrum using (11) and (12), the method is not practicable due to the difficulty of approximating the limit as well as the large computational complexity. A useful tool to make estimation efficient is the Legendre transform. The Legendre transform of the partition function is defined as

$$f_L(\alpha) = \inf_q \{q\alpha - T(q)\}. \tag{13}$$

It can be shown that $f_L(\alpha)$ converges to the true multifractal spectrum using the theory of large deviations (Ellis, 1984).

From the practical point of view, we need a good estimator of the partition function. If we rearrange (10), it becomes

$$E|d_{j,k}|^q \sim 2^{jT(q)} \quad \text{as} \quad j \to -\infty. \tag{14}$$

On the other hand, it has been shown that the qth moment of the wavelet coefficients of the power law process, Arneodo (1998), satisfies the following equation:

$$E|d_{j,k}|^q = C_q 2^{jqH} \tag{15}$$

where H is the so-called self-similarity exponent and C_q is a constant depending only on q. Comparing (14) and (15), one can easily connect the partition estimation with the self-similarity exponent estimation problem. It has been a standard practice

to use linear regression to identify the self-similarity exponent H since the values $E|d_{j,k}|^q$ could be easily obtained by moment-matching method making estimation of the partition function $T(q)$ easy. Formally speaking,

$$\log_2 \widehat{S_j(q)} = jT(q) + \varepsilon_j, \tag{16}$$

where $\widehat{S_j(q)} = \frac{1}{2^j} \sum_{k=1}^{N2^{-j}} |d_{j,k}|^q$ is the empirical q^{th} moment of the wavelet coefficients (N is the length of the time series) and the error term ε_j is introduced from the moment matching method when replacing the true moments with the empirical ones. Simple ordinary least square (OLS) is the most convenient choice of estimating the partition function.

Once the $T(q)$ is estimated, the next step is to perform the Legendre transform. Since $\frac{\partial}{\partial q}(\alpha q - T(q)) = \alpha - T'(q)$ and $T''(q) < 0$ (Gonçalvès et al 1998), the maximum value of $\alpha q - T(q)$ is achieved at $q = T'^{(-1)}(\alpha)$. So performing the Legendre transform is divided into two steps: First, the numerical derivative of $T(q)$ is obtained using the finite difference. Then, the value of Legendre spectrum at $\alpha = \widehat{T'(q)}$ is evaluated. We point out that the Legendre transform is not able to estimate the multifractal spectrum value at arbitrary singularity strength α. The set of the multifractal spectrum values is determined by set of q values. The more q values adopted, the finer multifractal spectrum that obtained, i.e., the resolution of the spectrum is determined by the "(order) sampling frequency" of the moments.

1.3 Geometric descriptors of multifractal spectra

Several geometric descriptors are introduced to summarize the multifractal spectra. These descriptors have interpretation in terms of location of and definition from monofractality. Traditionally, three geometric parts [25]: the vertical line, the maximum point and the right slope summarize the multifractal spectrum.

Theoretically, the multifractal spectrum of fBm (a representative of monofractal) consists of three geometric parts: the vertical line, the maximum point and the right

slope [25]. The maximum point corresponds to the Hurst exponent and the vertical line and the right slope are thought to be inherent features, which distinguish fBm from the multifractal process. However, it is rare to obtain such a perfect spectrum in practice. Even for the well simulated fBm, due to error of estimation (most of them are due to the partition function estimation and derivative calculation as shown above), its spectrum may deviate from the theoretical form, as shown in Figure 3. Even with the lack of precise estimation of the spectrum, the extent of deviation from the vertical line could be still utilized in the discrimination between the monofractal and multifractal processes. For example, two type processes are presented in the multifractal spectra in Figure 3. One is the fBm and the other is the turbulence measurement, which is widely believed to be a multifractal process. Comparing with the turbulence measurement, the fBm is much closer to the vertical line and this closeness may be quantified by the left slope of the spectra. Another important difference between these two spectra is the width spread of the spectra. It is obvious that the width spread of the fBm is much smaller than that of the turbulence measurement indicating the richness of singularity indices.

Despite the existence of the estimation error, the multifractal spectrum can be approximately described by 3 canonical descriptors without loss of the discriminant information, which are (1) Spectral Mode (Hurst exponent, H or SM), (2) left slope (LS) or left tangent (LT) and (3) width spread (Broadness, B) or right slope (LS) or right tangent (RT). A typical multifractal spectrum can be quantitatively described as shown in Figure 4. Understanding the SM and LS (or LT) is straightforward. SM represents the apex of spectrum or most common Hölder regularity index α found within the signal, and LS (or LT) represents the slope of the distribution produced by the collection of Hölder regularity index α with smaller values of the mode (SM). However, broadness (B) is more intricate descriptor of the multifractal spectrum. Broadness (B) is believed to be a more meaningful than right slope (RS) or right

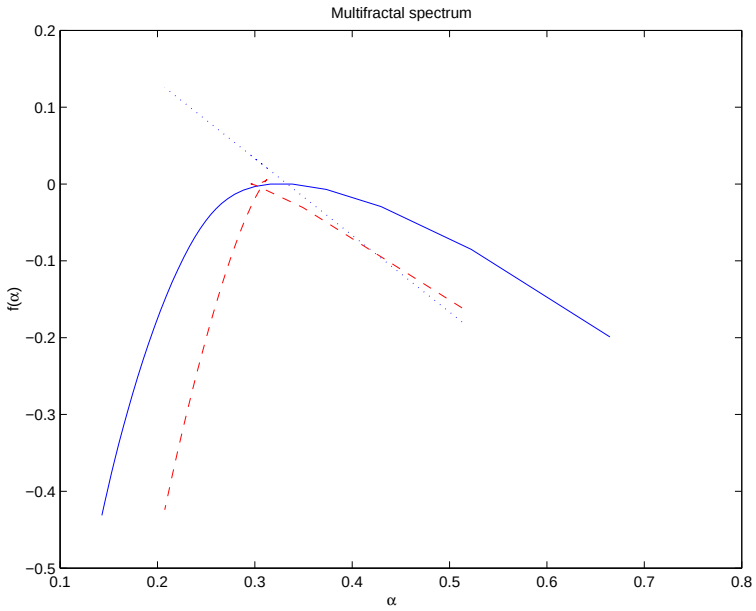

Figure 3: Multifractal spectra for monofractal (dash line) and multifractal (solid line) processes (The dotted line indicates the theoretical slope of the spectrum for the monofractal process)

tangent (RT) because it is a compound measure representing the overall nature of the multifractal spectra, taking into account the overall variability among the Hölder regularity index α. In addition, broadness (B) partially accounts for right slope (RS) or right tangent (RT) in calculation, as the resultant value of B is based on the relative values of RS and LS. In this study, we choose the left slope (LS), Spectral Mode (SM) and the broadness (B) as the spectral characteristics because we believe that the width spread has more power to discriminate the multifractality.

The both slopes (or both tangents) can be obtained easily using the interpolation technique, while it is not straightforward to define the broadness (B) automatically. The difficulties are related to two aspects - (i) difficulty to locate the start and end points of the width spread, and (ii) treatment of the discreteness of the spectrum. It is easy to see that the former is conceptually difficult, while the latter is computationally difficult. There are many ways to define the broadness (B). In this paper, we follows the definition, defined by Shi et al. [62]. When α_1 and α_2 are two roots which satisfy

the equation $f(\alpha)+0.2=0$ and $\alpha_1 < \alpha_2$, then broadness B is defined as $B=\alpha_2-\alpha_1$, where $f(\alpha)$ is the spectrum function in terms of Hölder regularity indices α's.

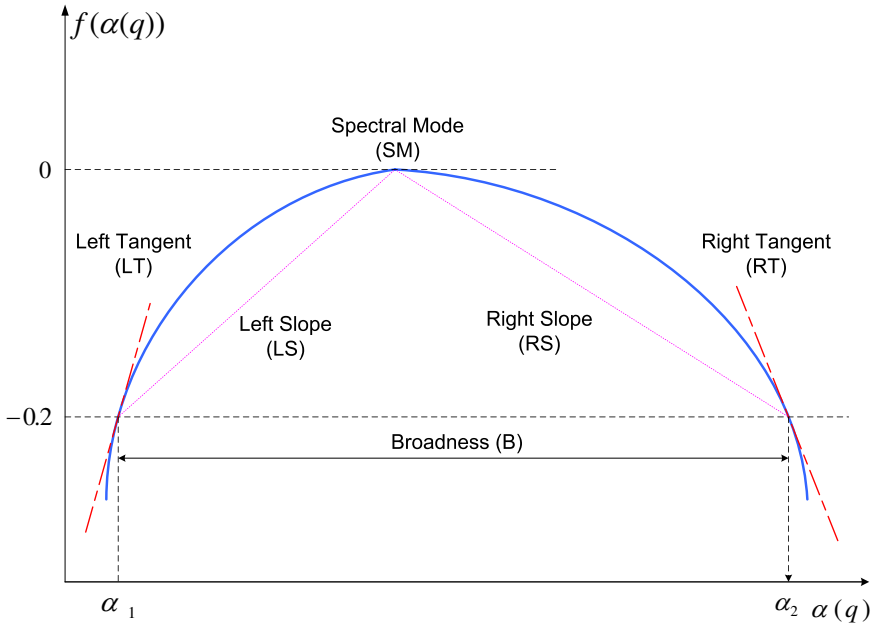

Figure 4: Illustration of geometric descriptors of multifractal spectra

This definition with two tangents and two slopes is also graphically presented in Figure 4. In Figure 4, the horizontal axis represents the value of Hölder regularity index $\alpha(q)$, and vertical axis represents values proportional to the relative frequency of these indices, $f(\alpha(q))$. The deviation from the monofractal could be fairly compared according to this broadness measure since it posts a universal standard on the width spread. It is worth to point out the threshold value 0.2 used in this definition could be adjusted empirically in the practice analysis to insure that this measure is well defined for all analyzed signals. The choice of threshold value is correlated back to the choice of q and the inherent data characteristics because these are the factors that affect the resolution of the spectrum.

As mentioned before, the discreteness may produce difficulties in the computation. The problem is that it may be hard to find the exact roots of the equation $f(\alpha)+0.2=$

0 among the discrete values of α's. To get around this, we first find two closest points $(\alpha_i^l,\ f(\alpha_i^l))$ and $(\alpha_i^u,\ f(\alpha_i^u))$ for each i such that

$$f(\alpha_i^l) < -0.2 \quad \text{and} \quad f(\alpha_i^u) > -0.2 \quad i = 1, 2$$

and then the two solutions α_1, α_2 can be easily obtained by interpolation. Thus two slopes LS, RS and two tangents LT, RT can be easily obtained by

$$LT = (f(\alpha_1^u) - f(\alpha_1^l))/(\alpha_1^u - \alpha_1^l) \quad \text{and} \quad RT = (f(\alpha_2^u) - f(\alpha_2^l))/(\alpha_2^u - \alpha_2^l)$$
$$LS = 0.2/(SM - \alpha_1) \quad \text{and} \quad RS = -0.2/(\alpha_2 - SM)$$

1.4 Test for distinguishing mono- and multi-fractal

In this section, we analyze the descriptors as potential statistics for distinguishing monofractal and multifractal objects. We propose a test with the left tangent as an informative index.

1.4.1 Parametric bootstrap test

Bootstrapping is a computer-based method for assigning measures of accuracy to statistical estimates with sampling from an approximating distribution. The advantage of bootstrapping is that it is straightforward to simulate empirical null-distribution of complex statistics such as percentile points, proportions, odds ratios, or correlation coefficients. The bootstrap method may also be used for constructing hypothesis tests as an alternative to inference based on parametric assumptions. In the case in which exact distributions are unknown or analytic procedures are too complex to obtain, even an approximation of the distribution, the bootstrap techniques could be successful. In our case, the distribution of an index for monofractality of fixed size, wavelet basis, and precision settings of a multifractal spectrum calculation over overly complex.

If a signal is monofractal, then the signal can be examined relatively easily compared to a multifractal signal. Thus, multifractal analysis is unnecessary for a monofractal signal, and monofractal analysis is useless for a multifractal signal. Therefore, before the analysis, we need to decide whether the signal is monofractal or multifractal. That is, the following hypothesis test is of interest.

Hypothesis Test	
$H_0:$ *the signal is monofractal*	$H_1:$ *the signal is not monofractal*

To illustrate the methodology, we compare a fractional brownian motion (fBm) and a multifractional Brownian motion (mBm) with time varying Hurst exponent. That is, the Hurst exponent for mBm is a function $H(t)$, where t is time. We consider a $mBm_{H(t)}$ with $H(t)$ given as, for $T = 2^{11}$,

$$H(t) = \frac{0.6}{T}t + 0.2, \qquad t \in [0, T].$$

Now we compare this $mBm_{H(t)}$ with a standard Brownian motion, $fBm_{0.5}$. Figure 5 depicts both signals and their multifractal spectrums.

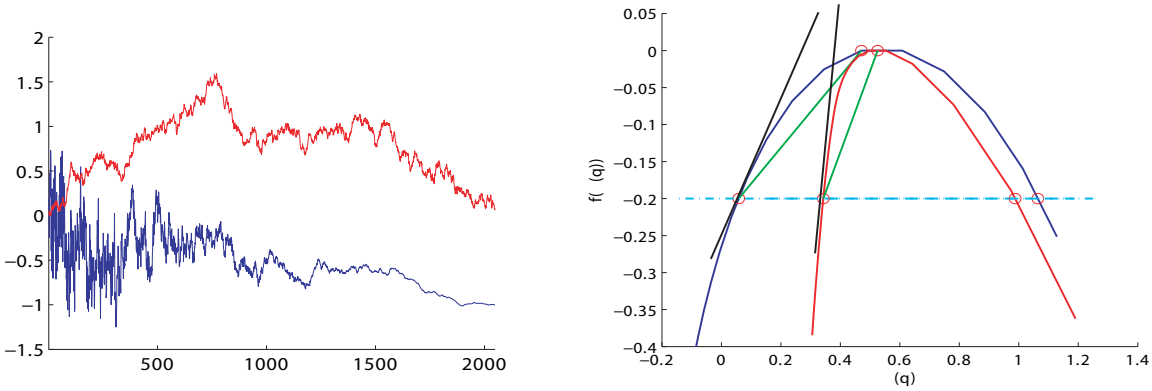

Figure 5: (a) Simulated signals of $fBm_{0.5}$ in red, mBm with straight line $H(t) = 0.6t/T + 0.2$ in blue; (b) MF spectrum for both signals

Figure 5(b) clearly shows that the multifractal spectrum for $fBm_{0.5}$ is steeper than that of mBm, especially the left side. This is obvious because the Hurst exponent

is a constant number for fBm, while it is a changing variable for mBm. Therefore, among six descriptors, left slope (LS), left tangent (LT), and broadness(B) is capable of discriminating two spectrums.

Table 1: Characters of monofractal and multifractal

	Monofractal	Multifractal
Left Slope	large	small
Left Tangent	large	small
Broadness	narrow	wide

The left slope is correlated to the Hurst exponent. However, the Hurst exponent could be unstable for a multifractal spectrum. Thus, left slope sometimes could be un-estimable. Therefore, left tangent as a more stable descriptor is used for the test. As seen in Table 1, multifractal spectrum have small left tangent, while multifractal spectrum of the monofractal signal have large number. Thus, when a signal is given, parametric bootstrap enables us to obtain the distribution of left tangent by simulated bootstrap surrogates. With the distribution, a critical value decides whether a signal is monofractal or not. The critical value can be 0.1 quantile, when we want to have 0.1 level of significance. The generating fractional brownian motion with the same H as given signals is performed by Wavelab, Matlab, developed by Buckheit and Donoho [9]. The procedure for the test has steps as follows:

Step 1: Given a signal, S_{obs}, in interest.

Step 2: Calculate $\widehat{H}_{obs}$ and $\widehat{LT}_{obs}$ for the signal.

Step 3: Generate B signals (fBm) which have hurst exponent $H = \widehat{H}_{obs}$ (Parametric bootstrapping with bootstrap number B).

Step 4: Estimate LT for B signals, i.e., $\widehat{LT}_1, \widehat{LT}_2, \cdots, \widehat{LT}_B$.

Step 5: Find the empirical distribution of LT from B simulations.

Step 6. Find the critical value with a pre-fixed level of significance, α.

Step 7. Reject H_0 if $\widehat{LT}_{obs}$ is in the rejection region.

1.5 Experimental data

This study was reviewed and approved by the Emory Investigational Review Board and was performed in the Emory University Hospital General Clinical Research Center (GCRC).

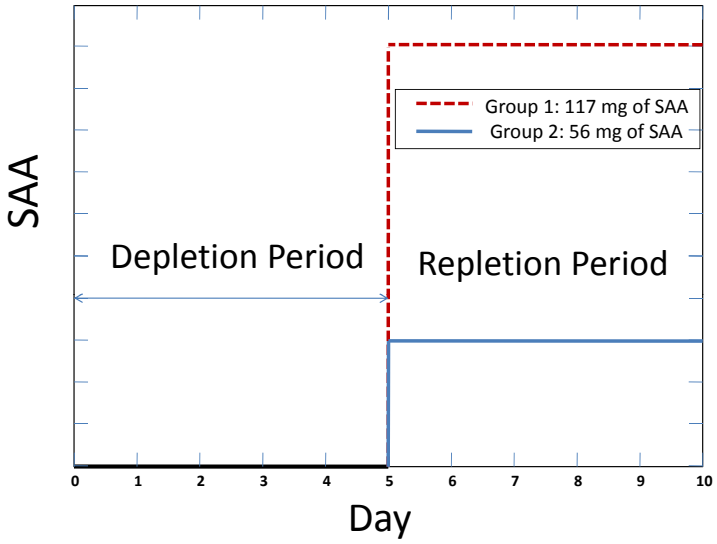

Figure 6: Experiment design: data were collected through 2 phases; depletion and repletion.

Motivated by the fact that a large number of biologic systems are functionally dependent upon sulfur amino acids (SAA) nutrition, researchers at Clinical Biomarkers Laboratory at Emory University started metabolomics research to identify metabolic changes associated with SAA deficiency. This research is of great importance in nutritional support because one of the SAA, cysteine, is routinely omitted due to its instability in nutritional support formulas. The ^{1}H-NMR spectroscopy was selected because of the speed and simplicity in sample processing and the possibility to obtain information on macronutrition that is quantitatively comparable over time.

To determine whether the variation in the amount of SAA intake affects metabolic processes, we conducted the study with two study groups (high and low dose of SAA group) and with three phases as shown in Figure 6. The first phase is a three day equilibrium period with meals prepared by the GCRC bionutrition unit. Following the equilibration period, subjects were placed on the 0 $mg.kg^{-1}d^{-1}$ SAA diet for the 5-day depletion period and then the diet SAA for the 5-day repletion period while remaining in the GCRC inpatient unit. The intake SAA amount is $56mg.kg^{-1}d^{-1}$ SAA for high dose of SAA group (High group, 4 subjects) and 117 $mg.kg^{-1}d^{-1}$ SAA for low dose of SAA group (Low group, 8 subjects.)

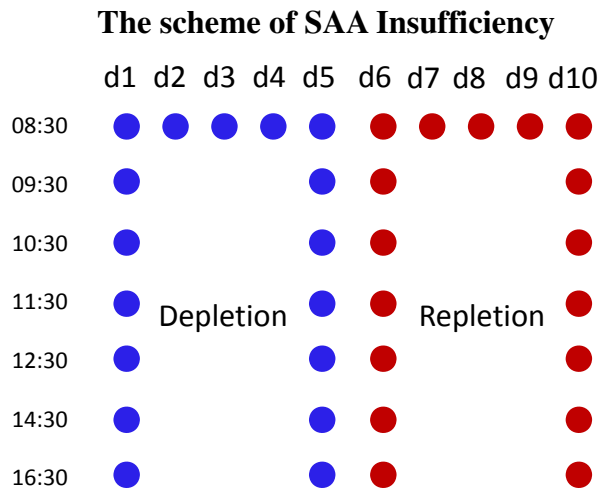

Figure 7: Experimental data collection times and days

Metabolic changes were investigated due to SAA intake detectable via ^{1}H-NMR spectroscopy. Blood samples were drawn on each day after the 5 day equilibration period at 08:30. On the first and last days of the depletion and repletion periods, draws occurred 7 times after the morning meal at 08:30, 09:30, 10:30, 11:30, 12:30, 14:30, and 16:30. On these days, breakfast and lunch were combined into the morning meal. On the other days, samples were drawn only one time a day after the morning

meal at 8:30, and 3 meals and snack were given. Figure 7 depicts the data collection time and day. Adequate hydration and vitamin, mineral and electrolyte requirements were provided to all subjects to meet or exceed recommended allowances [49], and body weights were determined daily and vital signs were obtained every 8 hours.

1.6 Statistical model analysis

1.6.1 Test for monofractality

After obtaining a data, we need to test for the monofractality. Data can generate multiple spectrums. If at least one of them reject the hypothesis, then we can conclude that the data have potential to be multifractal. For the test, I choose the signal of patient 1 from Day 1 at 8:30 am.

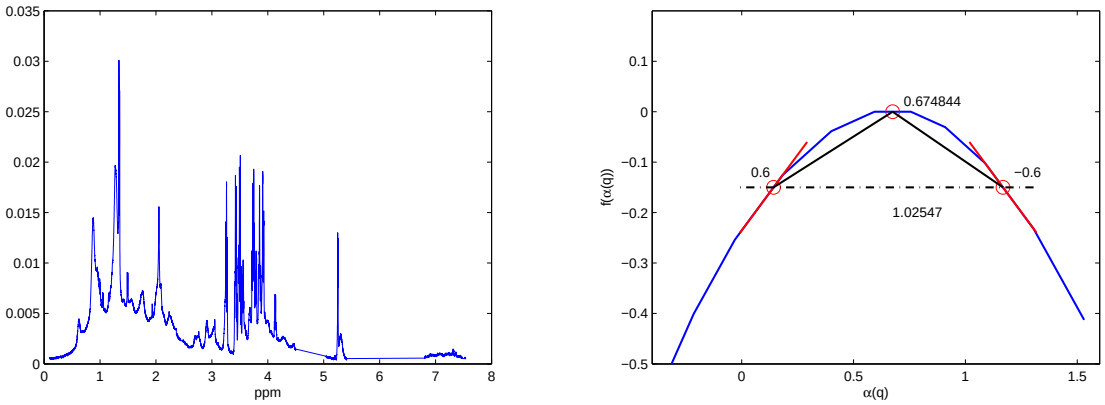

Figure 8: An example of showing (a) ^{1}H-NMR spectra example; (b) MF spectrum.

Figure 8 shows an example of ^{1}H-NMR spectra. The ppm varies from 0.1 to 7.5 ppm. The ^{1}H-NMR spectra is transformed by the wavelet transformation with a Daubechies wavelet function. The outcome of the transformation is a multifractal spectra. Then, geometric summaries of the spectra are calculated in Figure 8 (b).The hurst exponent for the signal is $H = 0.6748$. With bootstrap number $B = 2,000$, we can obtain the distribution of the left tangent. The distribution for the left slope is depicted in Figure 9 (a). The level of significance is 0.05.

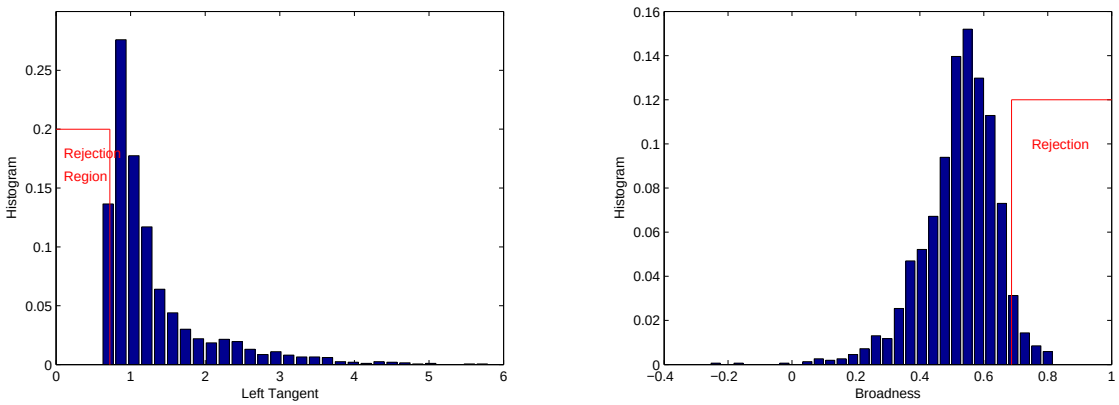

Figure 9: The distribution of (a) the left tangent; (b) the broadness.

The critical value for the left tangent is 0.6868, and the estimated left tangent for the given signal is $\widehat{LT}_{obs} = 0.6$. Thus, the test rejects the null hypothesis and concludes that the signal is not monofractal. In addition, Figure 9(b) is the distribution of the broadness. The critical value ($q_{0.95}$) is 0.7209 and the estimated broadness for the signal is $\widehat{B}_{obs} = 1.0255$. Therefore, the broadness confirms that the signal is not monofractal. Therefore, multifractal analysis is necessary to analyze the given data.

1.6.2 Statistical models

Depending upon whether variation over day in the morning or hourly variation in each day are of interest for time factor, two statistical models for each geometric feature in form of an unbalanced 3-way ANOVA model are appropriate because blood samples from each subject were drawn through depletion and either repletion of SAA intake. The outputs y^{1l}_{ijk} and y^{2l}_{ijk} are the values of each geometric feature of multifractal spectrum of ^{1}H-NMR spectrum for each model,$l = H, LS, \cdots$.

Model I. (Day Model) In this model, effect of day for each geometric feature is of interest. Therefore, the statistical model for each feature l contains the day effect as well as other effects.

$$y^{1l}_{ijkm} = \mu^{1l} + \alpha^{1l}_i + \beta^{1l}_j + \gamma^{1l}_k + \epsilon^{1l}_m \tag{17}$$

where μ^{1l} is the grand mean, α_i^{1l} is the effect of depletion ($i = 1$) and repletion ($i = 2$), β_j^{1l} is the effect of j^{th} day ($j = 1, \cdots, 5$), γ_k^{1l} is the effect of group k (High group $k = 1$, Low group $k = 2$), and ϵ_m^{1l} is assumed to be independent normal variable with mean 0 and unknown variance σ_{1l}^2 for eight geometric features

Model II. (Hour Model) Rather than effect of day as in Model I (Day Model), the Model II (Hour Model) assesses the hourly variation as well as group effect and depletion/repletion effect. Similarly to previous model, the corresponding statistical model for each feature l can be expressed as

$$y_{ijkm}^{2l} = \mu^{2l} + \alpha_i^{2l} + \beta_k^{2l} + \gamma_k^{2l} + \epsilon_m^{2l} \tag{18}$$

where μ^{2l} is grand mean, α_i^{2l} is effect of depletion ($i = 1$) and repletion ($i = 2$), β_j^{2l} is the effect of j^{th} hour slot ($k = 1, \cdots, 7$),γ_k^{2l} is the effect of group k (High group $k = 1$, Low group $k = 2$), and ϵ_m^{2l} are assumed to be independent normal variables with mean 0 and unknown variance σ_{2l}^2 for each feature, $k = 1, \cdots, 8$.

1.6.3 Model evaluation

Table 2: Descriptive statistics of geometric features of Model I (day model)

			Repletion		
		Depletion	Overall	High	Low
H	mean	0.7516	0.7325	0.7014	0.7481
	s.d	0.0057	0.0051	0.0078	0.0061
LS	mean	0.3685	0.3667	0.3951	0.3524
	s.d	0.0027	0.003	0.0052	0.0028
RS	mean	-0.4471	-0.4637	-0.4746	-0.4583
	s.d	0.0133	0.0209	0.0526	0.0172
LT	mean	0.6679	0.6595	0.7143	0.6321
	s.d	0.0073	0.0071	0.0133	0.007
RT	mean	-0.7429	-0.7548	-0.7071	-0.7786
	s.d	0.0124	0.0134	0.021	0.0167
B	mean	1.0227	1.0221	0.9903	1.0381
	s.d	0.0092	0.01	0.0176	0.0119

Table 2 provides the descriptive statistics for each geometric feature obtained from the multifractal analysis for each period and each group. 3-way ANOVA model for each geometric features concludes the group effect is significant for all features except B, because different amount of SAA was given to different group at repletion period, while the time effect is not significant for all features. According to the SAA intake effect, clearly the Group 1 (high dose) is discriminated with the depletion. Compared to the Group 1, the results of Group 2 (low dose) are close to the results in depletion. This is because the blood samples considered in this model were drawn at 8:30 a.m. on each day and only snack was provided at 21:30, which means the effect of SAA disappeared at every morning. And, there was no significant variation over day in the morning for all features. Thus, Table 2 implies that the amount of SAA significantly affect ^{1}H-NMR spectra which were evaluated by multifractal analysis, and that ^{1}H-NMR spectra of Group 1 has significantly more irregular pattern with larger variability when blood samples were drawn every in the morning over depletion/repletion period.

Table 3: Descriptive statistics of geometric features of Model II (hour model)

			Repletion		
		Depletion	Overall	High	Low
H	mean	0.7183	0.7228	0.6982	0.7351
	s.d	0.0084	0.0093	0.0124	0.0122
LS	mean	0.3621	0.371	0.3933	0.3598
	s.d	0.0038	0.0054	0.0106	0.0055
RS	mean	-0.536	-0.4737	-0.4437	-0.4886
	s.d	0.0936	0.0233	0.0215	0.0331
LT	mean	0.63	0.66	0.72	0.63
	s.d	0.0093	0.0119	0.0225	0.0114
RT	mean	-0.76	-0.76	-0.72	-0.78
	s.d	0.0241	0.0206	0.0268	0.0275
B	mean	1.0211	1.0053	0.9837	1.0162
	s.d	0.0165	0.0192	0.0312	0.0243

Table 3 provides the descriptive statistics of three geometric features of Model II. Similarly to Day Model, the mean values of H and B of Low group is larger than

those of High group, while the mean values of others of Low group are smaller. Only one different feature, which shows a different trend, is RS. Contrary to Day Model, all three effects for H feature are significant ($p-$values are <0.0001 for Group, 0.0205 for Depletion/Repletion, and 0.0268 for Hourly variation, respectively). That is, ^{1}H-NMR spectra of Low group has significantly more regular pattern, the difference between depletion and repletion for Hurst exponent is also significant, and hourly variation for Hurst exponent is significant as well. And all group effects for other features are also significant. Thus, group effect is important for all three features, while the difference between depletion and repletion and hourly variation are significant for Hurst exponent.

The aim of the research is to see the effect of SAA. We separate two phases, depletion and repletion, to give the difference of SAA. The results given in Tables 2, 3 conclude that the difference between depletion and repletion is larger in the High group than in Low group, which is the obvious result because the 117mg SAA intake should affect more than the 56mg SAA intake. The difference between the depletion and repletion in low group is so small that it can be considered as "no difference". Thus, we conclude that 56mg SAA intake does not affect the metabolic process. Because the intake SAA amount is not enough to give the difference between depletion and repletion, two groups of people, unhealthy ICU patients and healthy normal individuals, were introduced to compare.

Figure 10 depicts the Hurst exponent of all four groups. x and y-axis represents the Hurst exponent in (a) and (b), respectively. Figure 10 clearly show that the Hurst exponent of ICU people is much higher than the one of normal people. Also, no difference is detected between SAA-intakers and normal people. However, if you compare the blue circle (depletion) and red x (repletion) in (a), you can see repletion is more stable than depletion. Numerically, the variance of repletion is 0.0049 and the variance of depletion is 0.0091, which is the twice bigger than repletion variance.

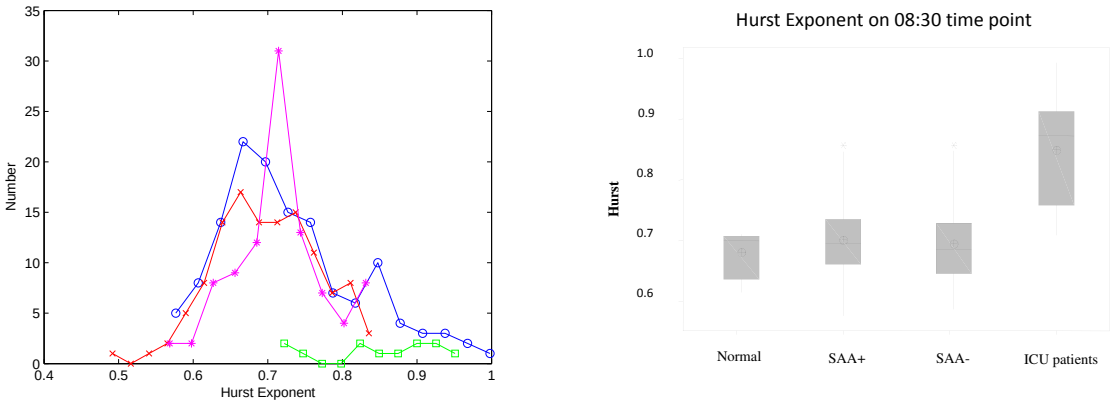

Figure 10: (a) Histograms (b) Boxplots of the Hurst exponents for depletion, repletion, ICU and normal people: blue circle is depletion; red x is repletion; green square is ICU; pink * is normal people

(b) shows the comparison in Box plots, which enable easier comparison of Hurst exponents. In (b), SAA+ represents the high group and SAA- represents the low group. ICU has the longest length of the Box plot, while SAA+ has the longest range.

1.7 Discussion

This research determines whether variation in sulfur amino acides (SAA) intake affects metabolic process by using the multifractal analysis. The multifractal analysis methodology uses functional data so that we can focus on the interplay of several biomarkers, while the traditional approaches focus on particular biomarkers. Moreover, this research improves the previous monofractal methodology which was studied by Jung et al [32]. The improvement lets the methodology applies to the real life example because most signals found in natureare multifractal, not monofractal.

For this study, two groups intake the different amount of SAA during equilibration and repletion period. Based upon three geometric features, H, LS, and B, of multifractal spectrum, two statistical models for each feature were employed to assess

underlying effects such as the group, depletion/repletion, and time (day or hour) effects. In Model I, the group effect is very significant for each feature, while variation over day and depletion/repletion effect are not significant for all three features at all. However, in Model II the group effect is significant for all features as well, and depletion/repletion effect and hourly variation in each day are significant only for H. This implies that the amount of SAA is important, and that SAA intake significantly affect hourly variation of metabolic process and the difference between depletion and repletion in each day and SAA effect disappears in the morning. Moreover, ^{1}H-NMR spectra of the Low group have significantly more regular patterns with more variability than those of High Group for both statistical models, which could be the evidence that high dose is recommended because irregularity of spectra could be considered as the sign of good health. But the relationship between the regularity of ^{1}H-NMR spectra and metabolic process would be important but left for the further research topic.

CHAPTER II

NANOPARTICLE SYNTHESIS IN A SUPERCRITICAL CO_2 PROCESS

2.1 Introduction

The cost-effective high-throughput generation of monodisperse nanoparticles of controlled size remains a challenge in nanomaterials processing. Due to the stochastic nature of nanoparticle nucleation, there will always be a distribution in nanoparticle size. Moreover the process is highly sensitive to process conditions such as temperature and pressure, which are useful for tailoring the size distribution but can also create robustness problems if these conditions and others are not tightly controlled. However, there is a significant need for nanoparticles of controlled size in a variety of applications including catalysis [26], medicine [4], and photonics [10].

In our research, we focus on energy applications in which transition metals are used as supported catalysts. However, the detailed chemical mechanisms in many catalysis processes are unknown due to the difficulty of measurements. Part of the uncertainty is due to the difficulty of synthesizing supported nanoparticle catalysts with well controlled size and structure. In particular, the size distribution of a nanoparticle is difficult to control, even though a monodisperse distribution is desirable for understanding the catalysis reactions. Extensive research (experimental and modeling) has been performed to understand the relationship between catalyst structure and the resulting catalytic reactions, with detailed studies on the colloidal synthesis of nanoparticles [13], [64]. However, there is still a significant need to understand how to produce controlled size distributions of supported nanoparticles in catalysts [59]. One of the reasons is that there are so many possible settings and choices for the

process, and fundamental yet accurate models of these processes are still lacking. As a result, the development of synthesis methods for these new catalysts is largely empirical, leaving the vast process space largely unexplored. The design of the temperature, pressure, concentrations of multiple species, choice of support, and its surface treatment is needed in order to optimize the process for a desired nanoparticle-based catalyst.

This chapter provides background about the supercritical CO_2 deposition of platinum nanoparticles. Computer experiments are substituted for real physical experiments because the physical experimentation is time consuming and expensive. In Section 2.2, chemical reactions in the nanoparticle synthesis are explained so that we can understand what input factors are included in the computer simulator, and what they are doing in the model, and why. In Section 2.3, we discuss the factors which give uncertainty. After running an experiment, several measures decide the performance of the experiment. In our research, four performance measure are considered and explained in Section 2.4. Section 2.5 shows preliminary computer experiment results. We finally conclude this chapter with some remarks in Section 2.6.

2.2 Mechanistic model

The nanoparticle synthesis consists of two sequential stages: The first stage is the adsorption stage, and the second stage is the thermal reduction stage. The reactions which happen in each stage are described later in this section. The synthesis uses two chemicals: a platinum precursor and a carbon nanotube (CNT) support. Experimenters can control the amounts of the chemicals. Moreover, experimental temperature, pressure, adsorption time, and growth time are controllable as well. In statistical terms, the process has six control factors. In addition, many uncontrollable factors exist in real experiment. Among many, we consider four, which are the most

intuitively important uncontrollable factors in our emulator: temperature error, pressure error, the impurity of a precursor, and the functionalization of the CNT. Figure 11 depicts the overall chemical reaction steps and factors. Because experimenters

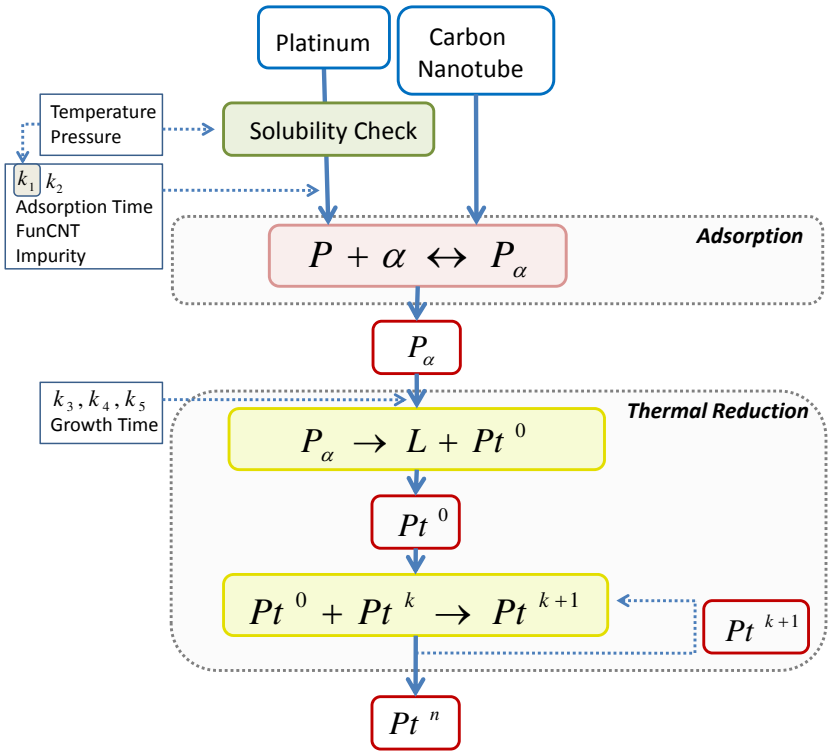

Figure 11: Reactions in a nanoparticle synthesis in a supercritical CO_2 process

control the amount of platinum, pressure, and temperature, some platinum is not soluble in the CO_2 under certain experimental conditions. In a computer experiment, we artificially have to check whether the input amount of platinum is all soluble or not. This checking step is called "Solubility Check" in Figure 11 and will be studied more in Chapter 3.

2.2.1 Adsorption reaction

Two chemical reactions in equations (19) and (20) formulate the adsorption reaction.

$$P + \alpha \xrightarrow{k_1} P_\alpha \tag{19}$$

$$P_\alpha \xrightarrow{k_2} P + \alpha \tag{20}$$

P represents a precursor molecule, which is the organometallic platinum precursor dimethyl (1,5-cyclooctadiene) platinum (II) ($Pt(COD)Me_2$) in this study. α represents the concentration of an active site on a CNT. k_1 is the precursor adsorption rate constant and k_2 is the precursor desorption rate constant. Equation (19) represents the adsorption of the precursor onto the CNT, and Equation (20) is the backward reaction for the adsorption, that is, the desorption of a precursor. Reactions (19) and (20) imply that P_α and $P + \alpha$ are interacting with the constant rates k_1 and k_2.

The adsorption time, the initial concentrations in the system, and the rate constant k_2 are pre-fixed (Initial concentrations of precursor, CNT, active sites on the CNT are $C_P^0, C_{CNT}^0, C_\alpha^0$, respectively). Only the adsorption rate constant k_1 depends on temperature.

$$k_1 \propto exp\{-E_a/RT\}. \tag{21}$$

T represents temperature, E_a is an active energy, and R is a gas constant. Equation (21) represents the effect of temperature in the adsorption process of the precursor in the CNT. In general, this dependency can be written in a form of a Arrhenius expression. Therefore, k_1 is changing by temperature, while k_2 is a fixed number, 3×10^{-4}. Now, the following two equations allow us to calculate the concentration of a precursor and an active site.

$$\frac{dC_P}{dt} = -k_1 C_P C_\alpha + k_2 C_{P_\alpha} \tag{22}$$

$$\frac{dC_\alpha}{dt} = -k_1 C_P C_\alpha + k_2 C_{P_\alpha}, \tag{23}$$

The concentration of precursor absorbed on the CNT is C_{P_α} and is the change from C_α^0 to C_α, $C_{P_\alpha} = C_\alpha^0 - C_\alpha$. As the time t changes, the concentration of a precursor C_P changes. Both equations (19) and (20) include precursor. In Equation (19), precursor is consumed and is produced in Equation (20). Therefore, the concentration is reduced by the rate constant k_1 and increased by the rate constant k_2. So, two terms $-k_1 C_P C_\alpha$ and $+k_2 C_{P_\alpha}$ comprise the differential equation (22). Because the precursor and the active site are consumed and produced in the same reactions with the same rate constants, the amount change in P is the same as that for α, i.e., Equation (23) equals to Equation (22).

2.2.2 Thermal reduction reaction

Once the precursor is absorbed on the active surface,the pressure is released, the vessel is heated up, and the following thermal reduction reactions happen.

$$P_\alpha \xrightarrow{k_3} L + Pt_1 \tag{24}$$

$$Pt_1 + Pt_1 \xrightarrow{k_4} Pt_2 \tag{25}$$

$$Pt_1 + Pt_n \xrightarrow{k_5} Pt_{n+1}, \tag{26}$$

where k_3 is the reduction rate constant, k_4 is the nucleation rate constant, k_5 is the growth rate constant. P_α is decomposed into two components; the organic ligand L and the elemental platinum Pt_1, which is an isolated platinum atom on the CNT surface. The ligand L causes the precursor to be soluble in the CO_2. However, we need pure platinum nanoparticles. The precursor can be separated by heat; thus, Equation (24) explains the separation of the precursor. After degrading the precursor, nucleating Pt_1 and Pt_1 with rate k_4 produces a platinum cluster with two atoms, Pt_2. This reaction is simplified in Equation (25). Reacting platinum with n atoms of Pt increases number of atoms by one. The growth rate k_5 is assumed to be constant

same for all n. Pt_n is a nanoparticle with n platinum atoms. The concentration of Pt_i can be derived by the following equations.

$$\frac{dC_{P_\alpha}}{dt} = -k_3 C_{P_\alpha} \tag{27}$$

$$\frac{dC_{Pt_i}}{dt} = k_5 C_{Pt_1}(C_{Pt_{i-1}} - C_{Pt_i}) \qquad i = 3, \cdots, n-1 \tag{28}$$

$$\frac{dC_{Pt_2}}{dt} = k_4 C_{Pt_1}^2 - k_5 C_{Pt_1} C_{Pt_2}, \qquad \frac{dC_{Pt_n}}{dt} = k_5 C_{Pt_1} C_{Pt_{n-1}}$$

$$\frac{dC_{Pt_1}}{dt} = k_3 C_\alpha - 2k_4 C_{Pt_1}^2 - k_5 C_{Pt_1} \sum_{i=2}^{n-1} C_{Pt_i} \tag{29}$$

The initial concentration of the intermediate P_α is always consumed with the rate constant k_3, not generated. Thus, we can get Equation (27) which comes from the reaction in Equation (24). The nucleation of a nanoparticle from two individual platinum atoms occurs with the rate constant k_4. The incorporation of additional platinum atoms results in growth of the platinum nanoparticles. Equations (28) and vv(29) are derived from equations (24), and (25),(26). The positive term means the generation of the chemical, and the negative term means the consumption of the chemical. The reactions are active until we obtain all nanoparticles having n atoms. After finishing the reactions, the final result of an experiment is a transmission electron microscopy (TEM) image. The following is one of our experimental results.

Figure 12 is a TEM image of platinum nanoparticles deposited on a CNT in a sc-CO_2 process. The dark circles in the figure are the final nanoparticles. The experimenter should manually see the image and count and measure the amount and the size. The typical nanoparticle size is between 5 to 10 nm with a visible distribution of sizes.

2.3 Noise factors

The most difficult part of simulating the physical experiment is uncertainty. Many unknown and uncontrollable factors cause uncertainty. Some factors are easier to

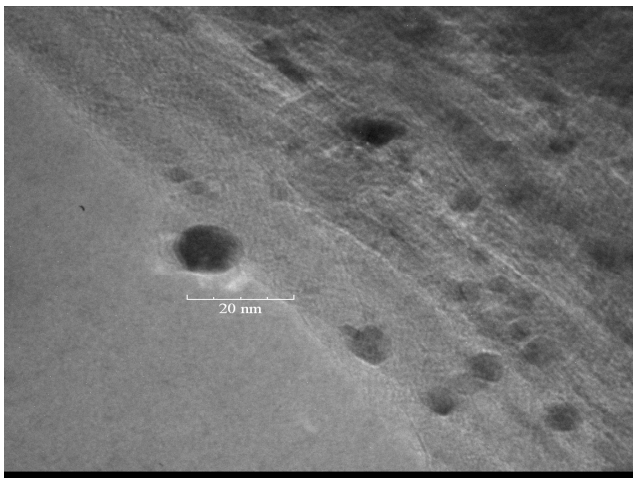

Figure 12: Transmission electron microscopy image from a sc-CO_2 process for Pt nanoparticles on carbon nanotubes (performed by Dr. Galit Levitin at Georgia Tech)

estimate than others. Our research includes the most basic four noise factors: temperature error, pressure error, impurity, functionalized CNT. Other possible noise factors are explained in Appendix A.

1. Temperature error

During the thermal reduction reaction (24), the vessel is heated in an oven for degrading. We use a MDSi8 series thermometer for the oven and the manufacturer states that the thermometer has an accuracy of $\pm 0.5\ K$ [1]. For example, even though we set the oven temperature as 300 K, the actual oven temperature varies from 299.5 to 300.5 K. To include temperature noise, "$Temperature$" is replaced by "$Temperature\ setting + Noise$". We assume that "$Noise$" follows a normal distribution with mean zero and variance 1.2^2 to obtain 90% of errors are in π 2 K.

2. Pressure error

Similar to temperature, we cannot control pressure perfectly. Therefore, the pressure value in an experiment is a combination of the pressure setting (the pressure we want to set) and the pressure noise, i.e., "$Pressure = Pressure\ Setting + Noise$", where

"$Noise$" follows a normal distribution with mean zero and variance 0.1^2 to obtain 90% of errors are in π 0.172 MPa.

3. Impurity

Based on the molecular weight of the precursor, the amount of platinum should be 49.6%, but the specifications state 48% minimum [3]. The calculation based on the molecular weight of platinum, which is explained in Appendix C, says the impurity is at most 3.22% and is obviously nonnegative. Because we know a precursor includes the impurity, we have to exclude the impurity weight in the model. For example, $x\ mg$ precursor is included in a certain simulation. Obviously, 0 – 3.22 % impurity is included in $x\ mg$, but we do not know the exact percentile. Thus, we can choose a uniform random number, y which is between 0 and 3.22 and consider the chosen number as impurity percentile. Then, the actual precursor we can use during an experiment is $x \times (1 - y/100)$. Therefore, from now on, the initial concentration of precursor is $x \times (1 - y/100)$ instead of x.

4. Functionalized Carbon Nanotube (FunCNT)

Performing a reflux in sulfuric/nitric acid to functionalize the surfaces of nanotubes results a large concentration of carboxyl (-COOH) groups on the nanotube surface. After functionalization, these carboxylated nanotubes have 2 – 7 $wt\%$ COOH by titration. This changes the number of surface sites α on the carbon nanotubes. Therefore, this value changes the initial concentration of α. When the initial concentration of α is x, the actual active site concentration is $x \times FunCNT$. "$FunCNT$" follows a normal distribution with mean = 4.5, and 99% varying between 2 and 7 %. Obviously FunCNT cannot have a negative value.

2.4 Performance measure

The output of an experiment is a distribution of final nanoparticles. The distribution gives two performance measure: the average size of the final nanoparticles, and the standard deviation of the distribution of sizes. In addition, we can calculate the yield of the precursor and the loading of the Pt on the CNT. Therefore, four performance measures are used to evaluate an experiment.

1. Average final nanoparticle size

Reactions generate nanoparticles which have different sizes. Our goal is to obtain the target size nanoparticle. The mean of the histogram will be compared to the target size. The following equation calculates the mean of the histogram.

$$m = average\ size = \frac{\sum_{i=2}^{n} i \cdot C_i}{\sum_{i=2}^{n} C_i}\ , \tag{30}$$

where C_i is the concentration of precursor when the number of atoms is i, and n is the maximum number of atoms in the simulation. The n is a pre-fixed number set to enable practical computations. Equation (30) is based on the number of atoms, but we can convert the *atom* number to a *nanometer* scale diameter. Appendix B explains the conversion.

2. Within-batch Variance

Within-batch variance represents how the final nanoparticle is spread out from the mean at the end of each experiment. Because we compare the target size to the mean which is explained above, we need to consider the variance of the histogram. The variance is calculated for each realization, and we use the equation proposed in [69].

$$Variance\ =\ \frac{\sum_{i=2}^{n} (C_i \cdot (i-m)^2)}{(n-1) \cdot \sum_{i=2}^{n} C_i}\ ,$$

where m is the calculated average. To get more confidence of the mean, the smaller variance is the better.

3. Yield

Yield represents how much of the pure precursor was deposited on the CNT surface. It can be expresed as the ratio of absorbed Pt(COD)Me_2 to initial Pt(COD)Me_2.

$$Yield = \frac{\text{final mass of final nanoparticle}}{\text{initial mass of platinum}} = \frac{\sum_{i=1}^{n} i \cdot C_i}{C_P}.$$

Both C_i and C_P are in mol/L. Because we do not want to waste the precursor, a larger yield is better.

4. Load

Loading is how much the weight of the CNT increases due to the platinum on the surface, relative to the initial amount of CNT in the system. The details of the calculation is explained in Appendix D.

$$Load = \frac{mw_{P_T} \sum_{i=1}^{n} i\ C_i}{mw_{P_T} \sum_{i=1}^{n} i\ C_i\ +\ massCNT}\ .$$

Same reason with yield, the larger load is the better.

2.5 Preliminary computer experiments

After running a simulation, the final nanoparticles have the distribution which has the form of a log-normal distribution. Figure 13 is one example of the distribution. As you see in Figure 13, no nanoparticles are generated at sizes larger than 1200 *atoms*. Thus, we need to set large enough number n not to miss possible nanoparticles of large size.

The histogram gives two performance measures: the average size of final nanoparticles and the standard deviation of final nanoparticles. The average size of final

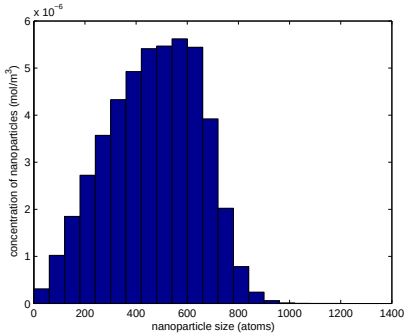

Figure 13: Final nanoparticle size distribution

nanoparticles is roughly in the middle of the histogram. Because the graph is right-skewed, the average is slightly left value of the mode. In Figure 13, we can guess the average size is around 450 atoms. The standard deviation is correlated to the width of the histogram. If the histogram is broadly spread, then the variance is large. If the histogram is narrowly spread, then the variance is small.

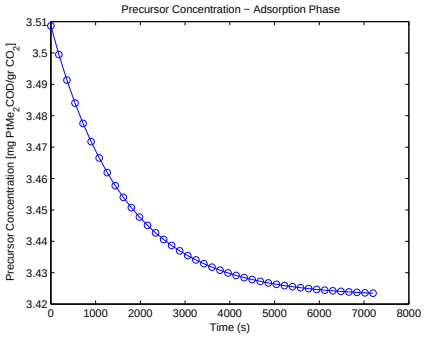

Figure 14: Precursor concentration change for adsorption step

Figure 14 shows the change of precursor concentration as time goes. The x-axis of the figure is adsorption time, and the y-axis is the precursor concentration. Thus, the graph gives the information that the precursor is exponentially reduced as time goes on. The figure gives the additional two performance measures: yield, and load. From the above graph, we can calculate the absorbed precursor amount. The left-most point in the graph is the initial amount of the precursor. The right-most point is the final amount of the precursor. Thus, the difference is the amount of absorbed

precursor. With the value, we can calculate the yield and the load.

Figures 13 and 14 are the results of a single experiment. The following results are based on a two-level full factorial design. Each of the ten factors (six control factors and four noise factors) has two levels, thus the total number of possible combinations is $2^{10} = 1024$.

Table 4: Variable setting for the two level full factorial design

	low	high	Unit
PreLoad	150	170	[mg]
CNTLoad	6.9	12.5	[mg]
Temperature	330	354	[K]
Pressure	18	30	[MPa]
Adsorp. Time	1.5	4.5	[Hr]
Growth Time	1.5	4.5	[Hr]
FunCNT	3.8	5.2	[%]
Temp. Noise	-1.3	1.3	[K]
Press. Noise	-0.11	0.11	[MPa]
Impurity	0.0064	0.024	[%]

Table 4 gives the settings of control and noise variables. Expert's knowledge is used to decide the low and high levels of each factors. After running 1024 simulations, four outcomes are produced in each simulation. The simulation results are depicted in Figure 15, which contains histograms of four measures. Clockwise from top left: final nanoparticle size, standard deviation, yield, and load. All performance measure are standardized. Even after standardizing, still histograms are left skewed. Therefore, the results show that neither a regular design nor a model can explain the synthesis, so further research should be required.

2.6 Discussion

This chapter presents the details of nanoparticle synthesis, including preliminary experimental results. After running an experiment, we obtain one histogram of final nanoparticles. We can also plot a precursor concentration change after finishing the adsorption stage. From the two figures, we compare four performance measures: the

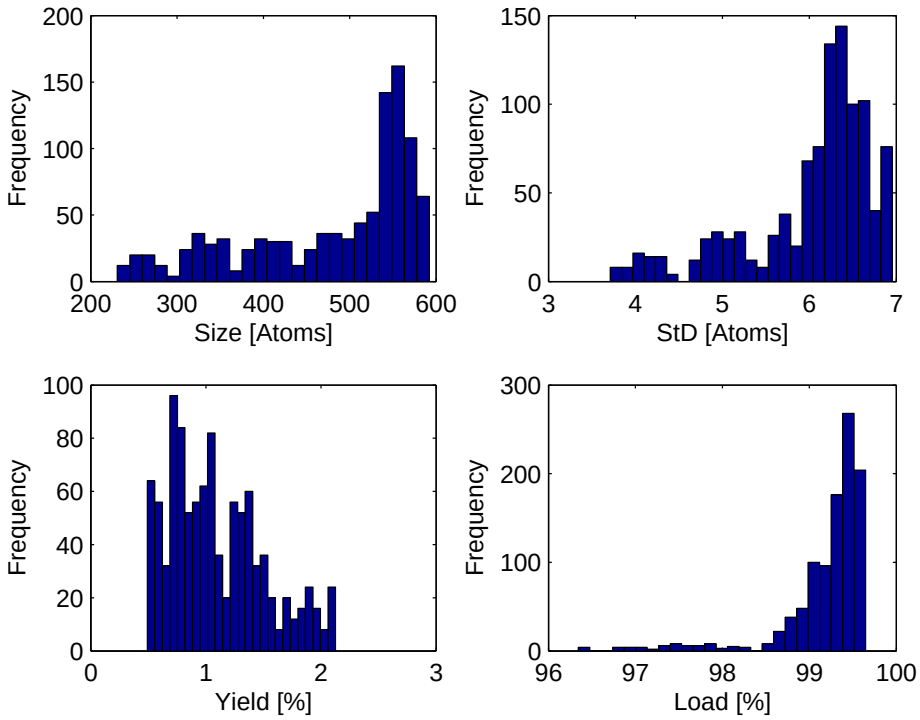

Figure 15: Four performance measure histograms for the two-level full factorial design

average of final nanoparticles, the standard deviations of final nanoparticles, yield, and load. The preliminary results, derived from two level full factorial design, show that the regular design is not suitable for the nanoparticle synthesis.

Currently, we use TEM images to find the final nanoparticle size distribution. From a statistical point of view, we need to consider the error, which comes from manual counting and measuring. Therefore, the model and design should consider the measurement error.

CHAPTER III

THE CONSTRAINED RANDOM EFFECT MODELS FOR NANOPARTICLE SYNTHESIS

3.1 Introduction

Real physical experiments have many constraints such as time and cost. Among the many constraints, some constraints are estimable and some are not. Because the solubility constraint has well-known equation and obviously affect the experimental results, it is used in this study. The solubility constraint represents the constraint of the precursor amount. Without the constraint, any amount of precursor can be used under any temperature and pressure. However, the equation says that the maximum soluble amount of the precursor depends on temperature and pressure. If the precursor amount does not exceed the maximum soluble amount, then all of the platinum precursor can be solubilized in the fluid phase; otherwise, the experimental setting is not appropriate for running an experiment. The solubility of the precursor in sc-CO_2 has been measured and then modeled using the Chrastil model, proposed in 1982 [14].

$$ln(S) = kln(\rho(T,P)) + \frac{a}{T+b}, \tag{31}$$

where S is the maximum soluble amount of a precursor [g/L], T the temperature [K], P the pressure [MPa], and ρ the density of the sc-CO_2[g/L]. k, a, b are the adjustment parameters. Equation (31) provides information about the dependency of the solubility of a precursor on the temperature and the pressure. That is, a combination of both the temperature and the pressure determines the solubility of a precursor.

Before starting an experiment, we want to check whether an input combination of temperature and pressure meets the solubility constraint or not. We examine

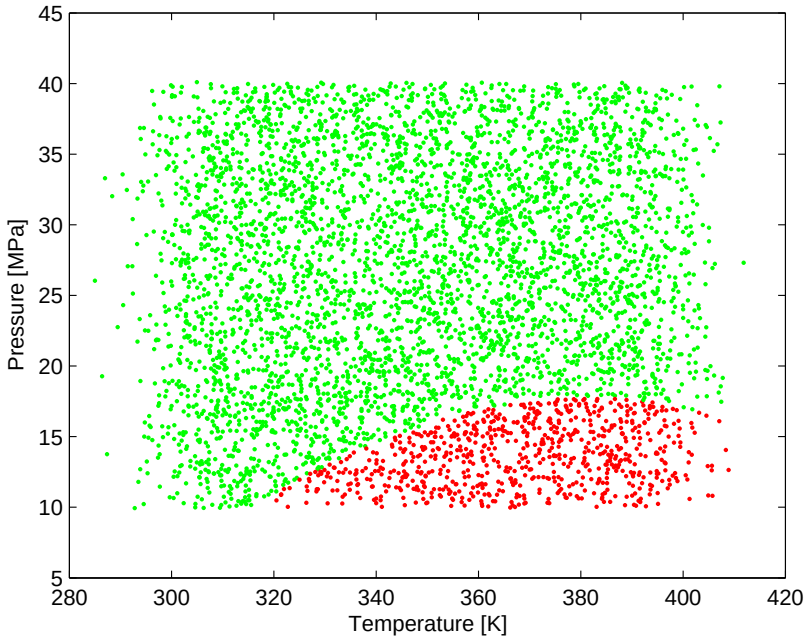

Figure 16: Nonlinear solubility constraint

over 5,000 combinations of temperature and pressure, the results of which are shown in Figure 16. Clearly, the regions are separated by a nonlinear constraint. In the figure, green dots (upper left part) represent the soluble region. That is, a precursor can be soluble under the combination of temperature and pressure. In contrast, red dots (lower right part) represent the insoluble region. If the input combinations of temperature and pressure are in the insoluble region, the amount of precursor exceeds the maximum soluble amount. Therefore, the experiment does not produce sensible outcomes. This research develops a regression model to model data in the soluble and insoluble regions.

Firstly, a logistic model is used to characterize the constraint. For the cases with productive outcomes a normal regression serves the second part of the model. Because the data are obtained from computer experiments, random-effects are included in both logistics and normal regression models to describe the potential spatial correlation among data. The likelihood function for this two-part model is complicated

and thus the maximum likelihood estimation is intractable. This research investigates approximation techniques based on Taylor series extension to simplify the likelihood. An algorithm is developed to find estimates for maximizing the approximated likelihood.

The chapter is organized as follows. In Section 3.2, we propose two models to explain the process. The likelihood functions of the models are approximated in Section 3.3. Based on the likelihood functions which are found in the previous section, we found the estimators of unknown parameters and its asymptotic distribution in Sections 3.5 and 3.6, respectively. All estimates are in Section 3.7, and the chapter concludes with a summary in Section 3.8.

3.1.1 Literature review

The generalization of computer experiments brought a lot of research pioneered by Sack [57, 58]. In these literatures, statistical methods in computer experiments are comprehensively reviewed for the first time. The statistical modeling with the computer-simulated data requires the uncertainty analysis. Kennedy and O'Hagan [34] introduced the Bayesian calibration technique which can handle the uncertainty. In 2011, Kim [35] develop a new methodology under a two-part model , but he focused on optimal designs to tackle the uncertainty, not the statistical modeling.

The Chrastil equation derived by Chrastil [14] motivates us to model a semi-continuous data separately. Olsen and Schafer proposed a two-part model, treating the outcome as continuous but adding a special modeling feature to take into account the strong floor effect [47]. However, authors in the literature did not consider the uncertainty coming from the computer simulation.

Our problem formulation incorporates the modeling of semi-comtinuous data and the uncertainty from computer experiment. Especially, our proposed model includes two different forms of uncertainty, within- and between-batch variation. In addition,

distributions of model parameters' estimates are derived.

3.2 Notations and models

Let $\mathbf{Y}$ be the response of the process. Without considering the solubility constraint, the process model is following:

$$\mathbf{Y}(x) = f(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}(x) \tag{32}$$

where $\mathbf{X}$ represents the input variables, β represents the unknown coefficient parameters, and $\boldsymbol{\varepsilon}$ is a measurement error. When $\mathbf{Y}$ contains many unproductive outcomes, the estimation of f based on $\boldsymbol{Y}$ is not accurate. Therefore, we will discriminate the zero values from others.

Let Y_{ij} denote a semi-continuous response and $\mathbf{X}_i$ a combination of explanatory variables for $i = 1, \cdots, m$. The subscript i is the index of a batch, the subscript j is the index of a particle in a batch, and m is the number of batches. The latter indicates that simulations are done by m different combinations of control factors. The simulation result, Y_{ij}, can have 0 or not. "$Y_{ij} = 0$" implies that the i^{th} experiment is not successful, which suggest that the input values do not meet the solubility constraint. If the experiment is successful, there is a continuous outcome value representing process quality like the nano-particle sizes. From the value Y_{ij}, we will propose two variables, U_i and V_{ij}. Both U_i and V_{ij} come from Y_{ij}. U_i is the indicator of the soluble region. Therefore, U_i is zero when the input values do not meet the solubility constraint, and U_i is one when the input values satisfy the constraint. V_{ij} is transformed Y_{ij}. The transformation is done by a function g, which is a monotone increasing function that makes V_{ij} approximately Gaussian. Therefore, the value V_{ij} exists only for the case $U_i = 1$.

$$U_i = \begin{cases} 1, & \text{if } Y_{ij} \neq 0 \\ 0, & \text{if } Y_{ij} = 0 \end{cases} \qquad V_{ij} = \begin{cases} g(Y_{ij}), & \text{if } Y_{ij} \neq 0 \\ \text{irrelevant}, & \text{if } Y_{ij} = 0 \end{cases}.$$

U_i follows a Bernoulli distribution and V_{ij} a Normal distribution.

3.2.1 Logistic regression

A logistic regression predicts the probability of an occurrence of an event. Because U_i follows a Bernoulli distribution, a logistic model can explain the trend of U_i. In this model, the response is a function of π_i which is $\pi_i = P(U_i = 1)$. $\mathbf{X}_{i,\eta}$ is the input data used in the logistic model.

$$\eta_i = \mathbf{X}_{i,\eta}\boldsymbol{\beta} + D_i, \tag{33}$$

where $\eta_i = log[\pi_i/(1-\pi_i)]$, β is a set of coefficients, and δ_i is the random effect in the logistic model. A random effect D_i is used in the logistic regression for capturing the possible spatial dependency between η_i and η_j $(i \neq j)$ at various process runs. The literature in computer experiments assumes that the data are spatially dependent [57]. The model includes the following structure of covariance between D_i and D_k at two different design points, $\mathbf{X}_{i,\eta}, \mathbf{X}_{k,\eta}$.

$$Cov(D_i, D_k) = \sigma_D^2 R(\mathbf{X}_{i,\eta} - \mathbf{X}_{k,\eta}).$$

Note that a common variance σ_D^2 is assumed for all D_i's, and $R(\cdot)$ is a distance function. The $\mathbf{D}$ is a vector of D_i, $\mathbf{D} = [D_1, D_2, \cdots, D_m]^T$, where the notation ^{T} represents the vector (or matrix) transpose. Assume $\mathbf{D}$ follows a normal distribution with mean zero and covariance matrix, Σ_D.

$$f_D(\boldsymbol{\delta}) \propto |\Sigma_D|^{-1/2} exp\{-\boldsymbol{\delta}^T \Sigma_D^{-1} \boldsymbol{\delta}/2\} \tag{34}$$

Note that the distribution of U_i is Bernoulli given the condition of the random effect D_i. See Section 3.4 for the way to estimate Σ_D.

3.2.2 Normal regression function

The transformed normal output V_{ij} follows a normal regression function. Equation (35) is valid only for the soluble region, $U_i = 1$. If the data come from the soluble

region, we can fit the data with the response.

$$V_{ij} = \mathbf{X}_{i,V}\boldsymbol{\gamma} + T_i + E_{ij}. \tag{35}$$

New subscript j represents an each particle size. From each batch of $(i = 1, 2, \cdots, m)$ reaction, n different particle sizes are generated. Thus, $j = 1, 2, \cdots, n$ represents that there are n different particle sizes. Each nanoparticle have the within batch variation E_{ij}. Suppose the total number of nanoparticle sizes does not change by batches. Since there are n nanoparticle sizes, similar to modeling repeated measurements taken from the same subject in the statistical literature, a random effect T_i is used to model the possible dependency between batches. The model has three assumptions regarding the random effects. First, the within-batch variation is independent each other. That is, E_{ij} and E_{il} are independent when $j \neq l$. Thus, the off-diagonal elements in Σ_{E_i} are zero, $Cov(E_{ij}, E_{il}) = 0$ when $j \neq l$. Second, between batch variation and within batch variation are independent, i.e., T_i and E_{ij} are independent. Lastly, the between batch variation have spatial Kriging correlation. That is, $Cov(T_i, T_j) = \sigma_T^2 R(\mathbf{X}_{i,V} - \mathbf{X}_{i,V})$, where $R(\cdot)$ is a distance function and σ_T^2 is a common variance for all i, $i = 1, 2, \cdots, m$. The distributions of between and within batch variations are estimated by Kriging methodology as described in Section 3.4.

Let $\mathbf{T}$ be a vector of T_i, i.e., $\mathbf{T} = [T_1, T_2, \cdots, T_m]^T$, and $\mathbf{E}_i$ be a vector of E_{ij}, i.e., $\mathbf{E}_i = [E_{i1}, E_{i2}, \cdots, E_{im}]^T$. Let covariance matrices of $\mathbf{T}$ and $\mathbf{E}_i$ be Σ_T, Σ_{E_i}. That is, $\mathbf{T}$ and $\mathbf{E}_i$ follow multivariate normal distributions with both zero means and covariance matrices Σ_T, Σ_{E_i}, respectively. The distribution functions of $\mathbf{T}$ and $\mathbf{E}_i$ are

$$f_T(\boldsymbol{\tau}) \propto |\Sigma_T|^{-1/2} exp\{-\boldsymbol{\tau}^T \Sigma_T^{-1} \boldsymbol{\tau}/2\}, \tag{36}$$

$$f_{E_i}(\boldsymbol{\varepsilon}_i) \propto |\Sigma_{E_i}|^{-1/2} exp\{-\boldsymbol{\varepsilon}_i^T \Sigma_{E_i}^{-1} \boldsymbol{\varepsilon}_i/2\}. \tag{37}$$

Equations (33) and (35) contain $\mathbf{X}_{i,\eta}$ and $\mathbf{X}_{i,V}$. Both matrices $\mathbf{X}$ imply the input data. However, $X_{i,\eta}$ and $\mathbf{X}_{i,V}$ are different in two ways. First, their coverage regions are different. $\mathbf{X}_{i,\eta}$ covers all regions while $\mathbf{X}_{i,V}$ covers only feasible regions $(U_i = 1)$.

For example, in Figure 16, $\mathbf{X}_{i,\eta}$ covers 280 to 420 K for temperature, but $\mathbf{X}_{i,V}$ covers only 318 to 420 K. Second, their dimensions are different. Since $\boldsymbol{\eta}$ is based on the constraint, $\mathbf{X}_{i,\eta}$ also is related to the constraint. If q out of p ($p \leq q$) factors are related to the constraint, the dimension of $\mathbf{X}_{i,\eta} = 1 \times q$. Obviously, the associated coefficient $\boldsymbol{\beta}$ has the dimension of $q \times 1$. For example, if the solubility constraint model contains temperature and pressure, $\mathbf{X}_{i,\eta}$ is a matrix with the dimension 1×2 when the number of columns (2) comes from the number of related parameters (temperature and pressure). However, $\mathbf{X}_{i,V}$ is related to overall process. The dimension of $\mathbf{X}_{i,V}$ is $1 \times p$ when p is the number of all control variables. Additionally, $\boldsymbol{\gamma}$ has the dimension $p \times 1$. For example, the process introduced in Chapter 2 includes six control parameters (temperature, pressure, platinum, carbon nanotube, adsorptin time, and growh time). Then the dimension of $\mathbf{X}_{i,V}$ is 1×6.

3.3 Likelihood function

With m batches and n particles in a batch, the outcome Y_{ij} becomes $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_m]^T_{mn\times 1}$, where $\mathbf{Y}_i = [Y_{i1}, Y_{i2}, \cdots, Y_{in}]$ for $i = 1, 2, \cdots, m$. Consequently $\mathbf{U} = [U_i]_{m\times 1}$ and $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_m]^T_{mn\times 1}$, where $\mathbf{V}_i = [V_{i1}, V_{i2}, \cdots, V_{in}]$ for $i = 1, 2, \cdots, m$. With $m \times n$ data, Equations (33) and (35) turn into

$$\boldsymbol{\eta}_{m\times 1} = \mathbf{X}_\eta \boldsymbol{\beta} + \mathbf{D}_{m\times 1}$$

$$\mathbf{V}_{i,n\times 1} = \mathbf{X}'_{i,V}\boldsymbol{\gamma} + \mathbf{T}_i + \mathbf{E}_{i,n\times 1}, \quad i = 1, 2, \cdots, m$$

with $\mathbf{X}_\eta = [\mathbf{X}^T_{1,\eta}, \mathbf{X}^T_{2,\eta}, \cdots, \mathbf{X}^T_{m,\eta}]^T$, $\boldsymbol{\eta} = [\eta_1, \eta_2, \cdots, \eta_m]^T$, $\mathbf{D} = [D_1, D_2, \cdots, D_m]^T$, $\mathbf{X}'_{i,V} = \mathbb{1}_{n\times 1}\mathbf{X}_{i,V}$, $\mathbf{T}_i = \mathbb{1}_{n\times 1}T_i$, and $mathbf{E}_i = [E_{i1}, E_{i2}, \cdots, E_{in}]^T_{n\times 1}$. D_i and T_i are between batch variations come from the batch-variation, and E_{ij} is a within batch variation for individuals with-in a batch. T_i and E_{ij} exist only when $U_i = 1$. Thus, we assumed that T_i and E_{ij} are independent conditioning on $U_i = 1$. The input data $\mathbf{X}$ represents the process input setting, which does not change within a batch i.

Accordingly, the coefficient $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ does not change within a batch. In addition, the between-batch variation T_i has only i index, while the within batch variation E_{ij} has i and j indexes. Therefore, the one vector ($\mathbb{1}_{n\times 1}$) takes place to adjust the dimension as repeating "$\mathbf{X}_{i,V}\boldsymbol{\gamma}$" and "$T_i$" n times. When p is the number of control variables, then $\mathbf{X}_{i,V}$ has a dimension $1 \times p$, and the dimension of $\boldsymbol{\gamma}$ is $p \times 1$. Among p variables, only q ($q \leq p$) can determine the process condition($U_i = 1$ or 0). Then, $\mathbf{X}_{\eta}$ has a dimension $m \times q$, and the dimension of $\boldsymbol{\beta}$ is $q \times 1$.

Let the likelihood function of unknown parameters, $\boldsymbol{\beta}, \boldsymbol{\gamma}$, from two models be L_Y. As the outcome $\mathbf{Y}$ is split into $\mathbf{U}$ and $\mathbf{V}$, the distribution of $\mathbf{Y}$ is the joint distribution of $\mathbf{U}$ and $\mathbf{V}$. That is, $L_Y = f_{\beta,\gamma}(\mathbf{U}, \mathbf{V})$. Without loss of generality, $f_{\beta,\gamma}(\mathbf{U}, \mathbf{V})$ can be $f_{\gamma}(\mathbf{V}|\mathbf{U})f_{\beta}(\mathbf{U})$. Based on the definition, the normal random variable $\mathbf{V}$ exists only when $\mathbf{U}$ is equal to 1. Thus, $f_{\gamma}(\mathbf{V}|\mathbf{U})$ refers $f_{\gamma}(\mathbf{V}|\mathbf{U} = 1)$

$$\begin{aligned} log\ L_Y &= log\ f_{\gamma}(\mathbf{V}|\mathbf{U} = 1) + log\ f_{\beta}(\mathbf{U}) \\ &= log\ L_{\mathbf{V},\gamma} + log\ L_{\mathbf{U},\beta}, \end{aligned}$$

where $L_{\mathbf{V},\gamma}$ is a likelihood function of $\mathbf{V}$ with the condition of $\mathbf{U} = 1$, and $L_{U,\beta}$ is a likelihood function of $\mathbf{U}$. The random variable which is a between batch variation, $\boldsymbol{\delta}$, is in the logistic model, thus the distribution of $\mathbf{U}$ will be integrated over the variable $\mathbf{D}$.

$$f_{\beta}(\mathbf{U}) = \int f(\mathbf{U}|\mathbf{D} = \boldsymbol{\delta}) f_D(\boldsymbol{\delta}) d\boldsymbol{\delta}$$

From the conditional Bernoulli distribution,

$$f(\mathbf{U}|\mathbf{D} = \boldsymbol{\delta}) = \prod_{i=1}^{m} exp\{U_i\eta_i - log[1 + exp(\eta_i)]\}, \tag{38}$$

where m is the number of batches, and $\eta_i = log[\pi_i/(1 - \pi_i)]$. The equation (38) can be derived as explained in Appendix E. The distribution of $\mathbf{D}$ is a standard normal distribution as in Equation (34).

Next, we discuss normal random variable $\mathbf{V}_i$. Conditioning on $\mathbf{U}_i = 1$, V_i follows a normal distribution. Recall that the between batch variation $\mathbf{T} \sim N(0, \Sigma_T)$ and the

within batch variation $\mathbf{E}_i \sim N(0, \Sigma_{E_i})$, where Σ_{E_i} is the variance of outcomes within a batch i. When two independent random variables follow normal distributions, their summation also follows a normal distribution with the mean and the variance as the summations of means and variances, respectively. For this reason, the mean and the variance of $\mathbf{V}$ are obtained. The mean of $\mathbf{V}$ is $E[\mathbf{V}] = \mathbf{X}'_V\gamma$, where $\mathbf{X}'_V = [\mathbf{X}'_{1,V}; \mathbf{X}'_{2,V}; \cdots; \mathbf{X}'_{m,V}]_{mn\times 1}$. The variance of $\mathbf{V}$ is the combination of Σ_τ and Σ_{ε_i}, $i = 1, 2, \cdots, m$. Let the covariance matrix be Σ_V. The covariance matrix can be expressed as m^2 block matrices, such as

$$\Sigma_V = \begin{bmatrix} \Sigma_V^{11} & \Sigma_V^{12} & \cdots & \Sigma_V^{1m} \\ \Sigma_V^{21} & \Sigma_V^{22} & \cdots & \Sigma_V^{2m} \\ \vdots & \vdots & \ddots & \cdots \\ \Sigma_V^{m1} & \Sigma_V^{m2} & \cdots & \Sigma_V^{mm} \end{bmatrix}, \tag{39}$$

where $\Sigma_V^{ij} = Cov(\mathbf{V}_i, \mathbf{V}_j)$ for $i, j = 1, 2, \cdots, m$. Because the dimension of $\mathbf{V}_i$ is $n \times 1$, each block in Σ_V has the dimension of $n \times n$. The diagonal matrix Σ_V^{ii} is the covariance of $\mathbf{V}_i$. In Σ_V^{ii}, (j,k)-th term is $Cov(V_{ij}, V_{ik}) = Cov(T_i + E_{ij}, T_i + E_{ik}) = Var(T_i) + Cov(E_{ij}, E_{ik}) = \sigma_T^2 + \Sigma_{E_i}^{jk}$, where $\Sigma_{E_i}^{jk}$ is the (j,k)-th term in the covariance matrix of $\mathbf{E}_i$, Σ_{E_i}. That is, Σ_V^{ii} becomes $\mathbb{1}_{n\times n}\sigma_T^2 + \Sigma_{E_i}$. For the (l,s)-th off-diagonal matrix, Σ_V^{ls}, the (j,k)-th term in the matrix is $Cov(V_{lj}, V_{sk}) = Cov(T_l + E_{lj}, T_s + E_{sk}) = Cov(T_l, T_s)$. That is, $\mathbf{V}|_{\mathbf{U}=1} \sim N(X'_V\boldsymbol{\gamma}, \Sigma_V)$. Then, the likelihood function of $\mathbf{V}$ is $L_{V,\gamma} = f(\mathbf{V}|\mathbf{U} = 1)$, and the distribution of $V|_{U=1}$ is

$$f(\mathbf{V}|\mathbf{U} = 1) \propto |\Sigma_V|^{-1/2} exp\{-(\mathbf{V} - X'_V\boldsymbol{\gamma})^T\Sigma_V^{-1}(\mathbf{V} - X'_V\boldsymbol{\gamma}))/2\}, \tag{40}$$

where Σ_V is the variance of $\mathbf{V}$ from Equation (39).

3.3.1 Approximation of logistics likelihood components

We will separately obtain the likelihood functions. First, we go over the likelihood function $L_{U,\beta}$ of the logistics model. If we substitute the Equations (34) and (38) into

$L_{U,\beta}$, it yields the following result.

$$L_{U,\beta} \propto \int \prod_{i=1}^{m} exp\{U_i\eta_i - log[1+exp(\eta_i)]\}exp\{-\boldsymbol{\delta}^T\Sigma_D^{-1}\boldsymbol{\delta}/2\}d\boldsymbol{\delta} \quad (41)$$

$$\propto \int exp\{h(\mathbf{U},\boldsymbol{\delta},\boldsymbol{\beta})\}d\boldsymbol{\delta}, \quad (42)$$

where $h(\mathbf{U},\boldsymbol{\delta},\boldsymbol{\beta}) = \sum_{i=1}^{m}\{U_i\eta_i - log[1+exp(\eta_i)]\} - \boldsymbol{\delta}^T\Sigma_D^{-1}\boldsymbol{\delta}/2$. If we set a vector $\mathbf{U}$ as $\mathbf{U} = [U_1, U_2, \cdots, U_m]^T$ and a vector $\boldsymbol{\eta}$ as $[\eta_1, \eta_2, \cdots, \eta_m]^T$, then easily $\mathbf{U}^T\boldsymbol{\eta}$ can replace $\sum_{i=1}^{m}(U_i\eta_i)$. That is,

$$h(U,\delta,\beta) = \mathbf{U}^T\boldsymbol{\eta} - log[\prod_{i=1}^{m}(1+exp(\eta_i))] - \boldsymbol{\delta}^T\Sigma_D^{-1}\boldsymbol{\delta}/2. \quad (43)$$

Our goal is to calculate the integral in Equation (42). However, $h(\mathbf{U},\boldsymbol{\delta},\boldsymbol{\beta})$ is too complicated to conduct the numerical integral calculation; therefore, we will use the Taylor series approximation. The Taylor series is a series expansion of a function about a point. A one-dimensional Taylor series around the point $\boldsymbol{\delta} = \tilde{\boldsymbol{\delta}}$ is $h(\mathbf{U},\boldsymbol{\delta},\boldsymbol{\beta}) = h(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta}) + h_\delta(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}) + 1/2(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})^T h_{\delta\delta}(\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}) + \cdots$, where

$$h_\delta(\cdot) = \frac{\partial\, h(\cdot)}{\partial\, \boldsymbol{\delta}}, \quad h_{\delta\delta}(\cdot) = \frac{\partial^2\, h(\cdot)}{\partial\, \boldsymbol{\delta}^2}.$$

Let the center point $\tilde{\boldsymbol{\delta}}$ be the maximum likelihood estimator of $h(\mathbf{U},\boldsymbol{\delta},\boldsymbol{\beta})$. The function h is positively proportional to the likelihood function $L_{U,\beta}$; consequently, the center point $\tilde{\boldsymbol{\delta}}$ maximizes the likelihood function as well.

A Taylor series is an infinite sum. If we take only a part of a series, then the sum of the chosen terms is an approximation of the original function. For the simplicity, we will take the first three terms.

$$\begin{aligned} h(\mathbf{U},\boldsymbol{\delta},\boldsymbol{\beta}) &= h(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta}) + h_\delta(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}) + \frac{1}{2!}(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})^T h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}) + \cdots \\ &\approx h(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta}) + h_\delta(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}) + \frac{1}{2!}(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})^T h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}). \quad (44) \end{aligned}$$

Let's substitute Equation (44) into Equation (42).

$$\int exp\{h(\mathbf{U},\boldsymbol{\delta},\boldsymbol{\beta})\}d\boldsymbol{\delta} \tag{45}$$

$$\approx \int exp\{h(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta}) + h_{\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}) + \frac{1}{2}(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})^T h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})\}d\boldsymbol{\delta} \tag{46}$$

$$= \int exp\{h(\tilde{\boldsymbol{\delta}})\}\cdot exp\{h_{\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})\}\cdot exp(\frac{1}{2}(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})^T h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}))d\boldsymbol{\delta} \tag{47}$$

$$= exp\{h(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})\}\cdot\int exp\{h_{\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})\}\cdot exp\{\frac{1}{2}(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})^T h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\delta-\tilde{\delta})\}d\boldsymbol{\delta}. \tag{48}$$

In Equation (48), $exp\{(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})^T h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})/2\}$ can be the density function of a normal distribution with mean $\tilde{\boldsymbol{\delta}}$ and variance $-h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})^{-1}$. Thus, $\int exp\{(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})^T h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})/2\}d\delta = \int(2\pi)^{m/2}|-h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})|^{-1/m}\cdot f(\boldsymbol{\delta})d\boldsymbol{\delta}$ when $\boldsymbol{\delta}\sim MN(\tilde{\boldsymbol{\delta}},-h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})^{-1})$ and $f(\boldsymbol{\delta})$ is the probability density function of $\boldsymbol{\delta}$. Therefore, the integral is equal to $(2\pi)^{m/2}|-h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})|^{-1/m}\cdot E[exp\{h_{\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})\}]$. $h_{\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})$ equals 0 due to the definition of $\tilde{\boldsymbol{\delta}}$ in Equation (44). Finally, Equation (48) is approximated to

$$\begin{aligned}\int exp\{h(\mathbf{U},\boldsymbol{\delta},\boldsymbol{\beta})\}d\boldsymbol{\delta} &\approx exp\{\frac{h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})^{-2}h_{\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})^2}{2}\}(2\pi)^{m/2}|-h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})|^{-1/m}\\ &= (2\pi)^{m/2}\cdot|-h_{\delta\delta}(\mathbf{U},\tilde{\boldsymbol{\delta}},\boldsymbol{\beta})|^{-1/m}.\end{aligned} \tag{49}$$

Now, we need to find $\tilde{\boldsymbol{\delta}}$ to get the solid form of $L_{U,\beta}$. Recall that $\tilde{\boldsymbol{\delta}}$ is a value such that $\partial h(\mathbf{U},\boldsymbol{\delta},\boldsymbol{\beta})/\partial\boldsymbol{\delta}=0$. The following equations show how the $h_{\delta}(\mathbf{U},\boldsymbol{\delta},\boldsymbol{\beta})$ is calculated.

$$\frac{\partial h(\mathbf{U},\boldsymbol{\delta},\boldsymbol{\beta})}{\partial\boldsymbol{\delta}} = \frac{\partial[\mathbf{U}^T\boldsymbol{\eta} - log[\prod_{i=1}^m(1+exp(\eta_i))] - \boldsymbol{\delta}^T\Sigma_D^{-1}\boldsymbol{\delta}/2]}{\partial\boldsymbol{\delta}} \tag{50}$$

$$= \frac{\partial[\mathbf{U}^T(\mathbf{X}_{\eta}\boldsymbol{\beta}+\boldsymbol{\delta}) - log[\prod_{i=1}^m(1+exp(X_{i,\eta}\boldsymbol{\beta}+\delta_i))] - \boldsymbol{\delta}^T\Sigma_D^{-1}\boldsymbol{\delta}/2]}{\partial\boldsymbol{\delta}}$$

$$= \mathbf{U} - \Sigma_D^{-1}\boldsymbol{\delta} - [\cdots, \frac{exp(X_{i,\eta}\boldsymbol{\beta}+\delta_i)}{1+exp(X_{i,\eta}\boldsymbol{\beta}+\delta_i)}, \cdots]^T \tag{51}$$

$$= \mathbf{U} - \Sigma_D^{-1}\boldsymbol{\delta} - [\cdots, B(\boldsymbol{\beta},\delta_i), \cdots]^T, \tag{52}$$

where $B(\boldsymbol{\beta},\delta_i) = exp(X_{i,\eta}\boldsymbol{\beta}+\delta_i)/(1+exp(X_{i,\eta}\boldsymbol{\beta}+\delta_i))$. By satisfying Equation (53),

we can find $\tilde{\boldsymbol{\delta}}$ iteratively, and do so using methods such as Newton-Raphson's procedute.

$$\tilde{\boldsymbol{\delta}} = \Sigma_D \ \mathbf{U} - \Sigma_D \ [\cdots, B(\boldsymbol{\beta}, \delta_i), \cdots]^T. \tag{53}$$

The second derivative of h comes from Equation (52).

$$\begin{aligned} h_{\delta\delta}(\mathbf{U}, \tilde{\boldsymbol{\delta}}, \boldsymbol{\beta}) &= \frac{\partial h_\delta(\mathbf{U}, \boldsymbol{\delta}, \boldsymbol{\beta})}{\partial \boldsymbol{\delta}}\Big|_{\boldsymbol{\delta}=\tilde{\boldsymbol{\delta}}} & (54) \\ &= -{\Sigma_D}^{-1} - diag[\cdots, B(\boldsymbol{\beta}, \tilde{\delta}_i), \cdots]. & (55) \end{aligned}$$

Once we find the $\tilde{\boldsymbol{\delta}}$, we can calculate $|-h_{\delta\delta}(\mathbf{U}, \tilde{\boldsymbol{\delta}}, \boldsymbol{\beta})|^{-1/m}$. As a result, the likelihood function $L_{U,\beta}$ is derived in a function of $\boldsymbol{\beta}$. Then, $\tilde{\boldsymbol{\beta}} = arg\ max_{\boldsymbol{\beta}} L_{\mathbf{U},\boldsymbol{\beta}}$ is an estimator such that $d\ log\ L_{U,\beta}/d\boldsymbol{\beta} = 0$. From Equation (49), $log\ L_{U,\beta} = (m/2)log(2\pi) - (1/m)log|-h_{\delta\delta}(\mathbf{U}, \tilde{\boldsymbol{\delta}}, \boldsymbol{\beta})|$. Thus, $\hat{\boldsymbol{\beta}}$ satisfies

$$\frac{d\ log\ L_{U,\beta}}{d\beta} = -\frac{1}{m} \cdot \frac{d\ log|-h_{\delta\delta}(\mathbf{U}, \tilde{\boldsymbol{\delta}}, \boldsymbol{\beta})|}{d\boldsymbol{\beta}} = 0 \tag{56}$$

From Equation (55), $log|-h_{\delta\delta}(\mathbf{U}, \tilde{\boldsymbol{\delta}}, \boldsymbol{\beta})| = log|{\Sigma_D}^{-1} + diag[\cdots, B(\boldsymbol{\beta}, \tilde{\delta}_i), \cdots]|$. For any matrix A, $d\ ln|A|/dx = tr(A^{-1} \cdot d\ A/dx)$. Thus, Equation (56) becomes

$$\frac{1}{m} tr(A(\boldsymbol{\beta})^{-1} \cdot \frac{d\ A(\boldsymbol{\beta})}{d\beta_i}) = 0, i = 1, 2, \cdots, q \tag{57}$$

where $A(\boldsymbol{\beta}) = {\Sigma_\delta}^{-1} + diag[\cdots, B(\boldsymbol{\beta}, \tilde{\delta}_i), \cdots]$, and $\boldsymbol{\beta}$ has a $q \times 1$ dimensions, i.e., $\boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_q]^T$. Section 3.5 estimates the estimator or $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, from the approximated likelihood from Equation (57).

3.3.2 normal likelihood components

The log-likelihood function of $\mathbf{V}$ is $log\ L_{V,\gamma} = log\ f(\mathbf{V}|\mathbf{U} = 1)$. From Equation (40), the likelihood function of $\mathbf{V}$ is

$$log\ f(\mathbf{V}|\mathbf{U} = 1) \ \propto \ log\ |\Sigma_V|^{-1/2} - \frac{(\mathbf{V} - \mathbf{X}_V\boldsymbol{\gamma})^T \Sigma_V^{-1} (\mathbf{V} - \mathbf{X}_V\boldsymbol{\gamma})}{2},$$

where Σ_V is a covariance matrix for $\mathbf{V}$. The derived likelihood function $L_{V,\gamma}$ is utilized to estimate the maximum likelihood function of $\boldsymbol{\gamma}$.

3.4 Distributions of with-in batch random effect

Note that $V_{ij} = \mathbf{X}_{i,V}^T \boldsymbol{\gamma} + T_i + E_{ij}$. The within batch variation E_i follows a multivariate normal distribution with zero mean and unknown covariance, that is, $\mathbf{E}_i \sim N(0, \Sigma_{E_i})$. The goal in this section is to estimate the variance of $\mathbf{E}_i$ with using the Kriging. The basic form of the kriging estimator is

$$\hat{\mathbf{E}}_i = \sum_{j=1}^{n} \lambda_j E_{ij},$$

where λ_j is the kriging weight. Variogram ($r_{E_i}(h)$) and covariogram ($C_{E_i}(h)$) functions from the Kriging are used to estimate the variance of E_i. Even though the underlying idea comes from the Kriging, the kriging weight λ_j does not need to be estimated. Let $C_{E_i}(\cdot)$ is the variance component from Σ_{E_i}. Let h be the distance between any two input points, i.e., $h = X_{ij} - X_{ik}, \forall i, j, k$. Then, the (j,k)-th element in Σ_{E_i} is $C_{E_i}(X_{ij} - X_{ik})$. The covariance matrix Σ_{E_i} is symmetric, thus $C_{E_i}(h)$ and $C_{E_i}(-h)$ are the same. The covariance term $C_{E_i}(h)$ has a covariance parameter $\sigma_{E_i}^2$ and the correlation parameter θ_{E_i}. That is, $C_{E_i}(h) = \rho(\sigma_{E_i}^2, \theta_{E_i}; h)$. Let $r_{E_i}(j-k)$ be $Var(E_{ij} - E_{ik})/2$. Then,

$$\begin{aligned} r_{E_i}(j-k) &= Var(E_{ij} - E_{ik})/2 \\ &= Var(E_{ij})/2 + Var(E_{ik})/2 - Cov(E_{ij}, E_{ik}) \\ &= C_{E_i}(0) - C_{E_i}(j-k) \\ &= \sigma_{E_i}^2 - C_{E_i}(j-k). \end{aligned}$$

The method-of-moments (MoM) estimates the value of $r_{E_i}(h)$ by

$$\hat{r}_{E_i}(h) = \frac{1}{2|S_h|} \sum_{(j,k) \in S_h} (V_{ij} - V_{ik})^2,$$

where $S_h = \{(j,k) : X_{ij} - X_{ik} = h\}$ and $|S_h|$ is the number of distinct pairs in S_h. Then, $\sigma_{E_i}^2$ and θ_{E_i} are estimated by minimizing $(\hat{r}_{E_i} - r_{E_i}(\sigma_{E_i}^2, \theta_{E_i}; j-k))^T (\hat{r}_{E_i} -$

$r_{E_i}(\sigma^2_{E_i}, \theta_{E_i}; j-k))$, and $\hat{r}_{E_i} - r_{E_i}(\sigma^2_{E_i}, \theta_{E_i}; j-k)$ is.

$$\hat{r}_{E_i} - r_{E_i}(\sigma^2_{E_i}, \theta_{E_i}; j-k) = \frac{1}{2|S_{j-k}|} \sum_{(j,k)\in S_{j-k}} (V_{ij} - V_{ik})^2 - \sigma^2_{E_i} + \rho(\sigma^2_{E_i}, \theta_{E_i}; j-k) \quad (58)$$

Once we estimate $\sigma^2_{E_i}$ and θ_{E_i}, any correlation function can structure the covariance matrix of E_i, that is, the (j,k)-th element in Σ_{E_i} is $\rho(\sigma^2_{\hat{E}_i}, \hat{\theta}_E; j-k)$. One example for the correlation function ρ is an exponential function which is the commonly used: $\rho(\sigma^2, \theta; h) = \hat{\sigma}^2 exp(-|h/\theta|),\ h > 0$ or a gaussian function: $\rho(\sigma^2, \theta; h) = \sigma^2 exp(-(h/\theta)^2),\ h > 0$. Similarly, we can estimate the variance of the between batch random effect δ and τ. Equation (58) becomes

$$\hat{r}_\delta - r_\delta(\sigma^2_\delta, \theta_\delta; j-k) = \frac{1}{2|S_{j-k}|} \sum_{(j,k)\in S_{j-k}} (\eta_j - \eta_k)^2 - \sigma^2_\delta + \rho(\sigma^2_\delta, \theta_\delta; j-k) \quad (59)$$

to estimate the covariance parameter σ^2_δ and the correlation parameter θ_δ. The correlation function $\rho(\cdot)$ can be different by random effects.

3.5 Estimation based on approximated likelihood function

In Section 3.3, we approximate the likelihood function for normal regression functions. Then, our ultimate goal is to estimate the unknown parameters, $\boldsymbol{\beta}, \boldsymbol{\gamma}$, in each model with approximated likelihood functions.

3.5.1 Estimator for logistic regression

We approximate the likelihood function $L_{U,\beta}$ for the logistic regression in Section 3.3.1. The estimator $\hat{\boldsymbol{\beta}}$ satisfies Equation (57). Since $A(\boldsymbol{\beta})$ includes $\tilde{\boldsymbol{\delta}}$, without having $\tilde{\boldsymbol{\delta}}$, $L_{U,\beta}$ does not have an explicit form and cannot get a closed form analytically. Newton-Raphson method is one of the commonly used methods for finding the root of a function in a numerical way. With the Newton-Raphson method, we can find the likelihood of the logistic regression. Then $\hat{\boldsymbol{\beta}}$ is a value that satisfies $d\ log\ L_{U,\beta}/d\boldsymbol{\beta} = 0$.

3.5.2 Estimator for normal regression

Recall that the log-likelihood function of $\mathbf{V}$ is

$$log\ L_{V,\gamma} = log\ |\Sigma_V|^{-1/2} - (\mathbf{V} - \mathbf{X}'_V\boldsymbol{\gamma})^T\Sigma_V^{-1}(\mathbf{V} - \mathbf{X}'_V\boldsymbol{\gamma})/2, \tag{60}$$

where Σ_V is the covariance matrix of $\mathbf{V}$. Maximum likelihood estimator $\hat{\boldsymbol{\gamma}}$ is a value that satisfies $d\ log\ L_{V,\gamma}/d\boldsymbol{\gamma} = 0$. In Equation (60), the first term $log\ |\Sigma_V|^{-1/2}$ is independent of $\boldsymbol{\gamma}$, so the term is canceled in the calculation of $d\ log\ L_{V,\gamma}/d\boldsymbol{\gamma} = 0$ for the simplicity.

$$\begin{aligned} \frac{d\ log\ L_{V,\gamma}}{d\boldsymbol{\gamma}} &= -\frac{d}{d\boldsymbol{\gamma}}\{(\mathbf{V} - \mathbf{X}'_V\boldsymbol{\gamma})^T\Sigma_V^{-1}(\mathbf{V} - \mathbf{X}'_V\boldsymbol{\gamma})\} &(61)\\ &= 0 &(62)\end{aligned}$$

Now $(\mathbf{V} - \mathbf{X}'_V\boldsymbol{\gamma})^T\Sigma_V^{-1}(\mathbf{V} - \mathbf{X}'_V\boldsymbol{\gamma})$ in Equation (61) will be fully expanded for the purpose of finding $\hat{\boldsymbol{\gamma}}$. Without any trouble, we can easily switch the order of $\mathbf{V}$ and $\mathbf{X}'_V\boldsymbol{\gamma}$. Thus, $(\mathbf{V}-\mathbf{X}'_V\boldsymbol{\gamma})^T\Sigma_V^{-1}(\mathbf{V}-\mathbf{X}'_V\boldsymbol{\gamma}) = (\mathbf{X}'_V\boldsymbol{\gamma}-\mathbf{V})^T\Sigma_V^{-1}(\mathbf{X}'_V\boldsymbol{\gamma}-\mathbf{V})$. The $mn\times 1$ matrix $(\mathbf{X}'_V\boldsymbol{\gamma} - \mathbf{V})$ is $[\mathbf{X}_{1,V}\gamma - V_{11}, \cdots, \mathbf{X}_{m,V}\gamma - V_{mn}]^T$. For the simple notation, let the inverse matrix of Σ_V be

$$\Sigma_V^{-1} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \ddots & \cdots \\ A_{m1} & A_{m2} & \cdots & A_{mm} \end{bmatrix}, \tag{63}$$

where

$$A_{ij} = \begin{bmatrix} \sigma_{11}^{ij} & \sigma_{12}^{ij} & \cdots & \sigma_{1n}^{ij} \\ \sigma_{21}^{ij} & \sigma_{22}^{ij} & \cdots & \sigma_{21n}^{ij} \\ \vdots & \vdots & \ddots & \cdots \\ \sigma_{n1}^{ij} & \sigma_{n2}^{ij} & \cdots & \sigma_{nn}^{ij} \end{bmatrix}. \tag{64}$$

When p is the dimension of $\boldsymbol{\gamma}$, $(\mathbf{X}'_V\boldsymbol{\gamma} - \mathbf{V})^T\Sigma_V^{-1}(\mathbf{X}'_V\boldsymbol{\gamma}\mathbf{V})$ becomes

$$(\mathbf{X}'_V\boldsymbol{\gamma} - \mathbf{V})^T\Sigma_V^{-1}(\mathbf{X}'_V\boldsymbol{\gamma}\mathbf{V}) = \sum_{j,l=1}^{m}\sum_{i,k=1}^{n}[\mathbf{X}_{l,V}\boldsymbol{\gamma} - V_{lk}][\mathbf{X}_{j,V}\boldsymbol{\gamma} - V_{ji}]\sigma_{ik}^{jl}$$

The s-th element in $\hat{\boldsymbol{\gamma}}$, $\hat{\gamma_s}$, satisfies $d\ log\ L_{V,\gamma}/d\gamma_s = 0, s = 1, 2, \cdots, p$, and p is the dimension of $\boldsymbol{\gamma}$. That is,

$$\begin{aligned}
\frac{d\ log\ L_{V,\gamma}}{d\gamma_s} &= \frac{d}{d\gamma_s}\sum_{j,l=1}^{m}\sum_{i,k=1}^{n}[\mathbf{X}_{l,V}\boldsymbol{\gamma} - V_{lk}][\mathbf{X}_{j,V}\boldsymbol{\gamma} - V_{ji}]\sigma_{ik}^{jl} && (65)\\
&= \sum_{j,l=1}^{m}\sum_{i,k=1}^{n}\frac{d}{d\gamma_s}[\mathbf{X}_{l,V}\boldsymbol{\gamma} - V_{lk}][\mathbf{X}_{j,V}\boldsymbol{\gamma} - V_{ji}]\sigma_{ik}^{jl} && (66)\\
&= \sum_{j,l=1}^{m}\sum_{i,k=1}^{n}\frac{d}{d\gamma_s}[\sum_{b=1}^{p}X_{lb}\gamma_b - V_{lk}][\sum_{b=1}^{p}X_{jb}\gamma_b - V_{ji}]\sigma_{ik}^{jl} && (67)\\
&= \sum_{j,l=1}^{m}\sum_{i,k=1}^{n}\left(2X_{ls}X_{js}\gamma_s + \sum_{b=1,\neq s}^{p}X_{lb}X_{js}\gamma_b + \sum_{b=1,\neq s}^{p}X_{jb}X_{ls}\gamma_b - X_{ls}V_{ji} - X_{js}V_{lk}\right)\sigma_{ik}^{jl}\\
&= \sum_{j,l=1}^{m}\sum_{i,k=1}^{n}\left(\mathbf{X}_{jl}^{s}\boldsymbol{\gamma} - X_{ls}V_{ji} - X_{js}V_{lk}\right)\sigma_{ik}^{jl} && (68)\\
&= 0 && (69)
\end{aligned}$$

In Equation (68),

$$\begin{aligned}
\mathbf{X}_{jl}^{s} &= [X_{j1}X_{ls} + X_{l1}X_{js}, X_{j2}X_{ls} + X_{l2}X_{js}, \cdots, 2X_{ls}X_{js}, \cdots, X_{jp}X_{ls} + X_{lp}X_{js}]\\
&= X_{ls}\mathbf{X}_{j,V} + X_{js}\mathbf{X}_{l,V} && (70)
\end{aligned}$$

That is, s-th element is $2X_{ls}X_{js}$, and k-th element ($k \neq s$) is $X_{jk}X_{ls} + X_{lk}X_{js}$. The last two lines above become

$$\sum_{j,l=1}^{m}\sum_{i,k=1}^{n}\mathbf{X}_{jl}^{s}\hat{\boldsymbol{\gamma}}\sigma_{ik}^{jl} = \sum_{j,l=1}^{m}\sum_{i,k=1}^{n}(X_{ls}V_{ji} + X_{js}V_{lk})\sigma_{ik}^{jl}. \quad (71)$$

Because the right hand side only depends on s, let's define $c_s = \sum_{j,l=1}^{m}\sum_{i,k=1}^{n}(X_{ls}V_{ji} + X_{js}V_{lk})\sigma_{ik}^{jl}$. In the left hand side, $\mathbf{X}_{jl}^{s}\hat{\boldsymbol{\gamma}}$ does not depend on i and k. Thus, it can be written as $\sum_{j,l=1}^{m}\mathbf{X}_{jl}^{s}\hat{\boldsymbol{\gamma}}d_{jl}$, where $d_{jl} = \sum_{i,k=1}^{n}\sigma_{ik}^{jl}$. Then, Equation (71) becomes $\sum_{j,l=1}^{m}\mathbf{X}_{jl}^{s}\hat{\boldsymbol{\gamma}}d_{jl} = c_s$. Because the dimension of $\boldsymbol{\gamma}$ is p, $\hat{\boldsymbol{\gamma}}$ should satisfy all p equations:

$d\ log\ L_{V,\gamma}/d\boldsymbol{\gamma}_s = 0, s = 1, 2, \cdots, p$. Let's sum all p equations:

$$\begin{aligned}
\sum_{j,l=1}^{m} \mathbf{X}_{jl}^{1} \hat{\gamma} d_{jl} &= c_1 \\
\sum_{j,l=1}^{m} \mathbf{X}_{jl}^{2} \hat{\gamma} d_{jl} &= c_2 \\
\vdots \qquad & \quad \vdots \\
\sum_{j,l=1}^{m} \mathbf{X}_{jl}^{p} \hat{\gamma} d_{jl} &= c_p
\end{aligned}$$

Then, $\sum_{s=1}^{p} \sum_{j,l=1}^{m} \mathbf{X}_{jl}^{s} \hat{\gamma} d_{jl} = (\sum_{s=1}^{p} \sum_{j,l=1}^{m} \mathbf{X}_{jl}^{s} d_{jl}) \hat{\boldsymbol{\gamma}} = \mathbf{X}\hat{\boldsymbol{\gamma}} = \sum_{s=1}^{p} c_s$, where $c_s = \sum_{j,l=1}^{m} \sum_{i,k=1}^{n} (X_{ls} V_{ji} + X_{js} V_{lk}) \sigma_{ik}^{jl}$, and the $1 \times p$ matrix $\mathbf{X}$ is

$$\begin{aligned}
\mathbf{X} &= \sum_{s=1}^{p} \sum_{j,l=1}^{m} \mathbf{X}_{jl}^{s} d_{jl} && (72) \\
&= \sum_{s=1}^{p} \sum_{j,l=1}^{m} (X_{ls} \mathbf{X}_{j,V} + X_{js} \mathbf{X}_{l,V}) d_{jl} && (73)
\end{aligned}$$

where d_{jl} is the sum of all elements in A_{jl}, which is the (j,l)-th element in Σ_V^{-1}. Finally, the estimator $\hat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ is derived by the linear least squares,

$$\begin{aligned}
\hat{\boldsymbol{\gamma}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\sum_{s=1}^{p} c_s) \\
&= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\sum_{s=1}^{p} \sum_{j,l=1}^{m} \sum_{i,k=1}^{n} (X_{ls} V_{ji} + X_{js} V_{lk}) \sigma_{ik}^{jl}),
\end{aligned}$$

with the matrix $\mathbf{X}$ from Equation (73).

The mean of $\hat{\boldsymbol{\gamma}}$ is

$$\begin{aligned}
E[\hat{\boldsymbol{\gamma}}] &= E[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\sum_{s=1}^{p}\sum_{j,l=1}^{m}\sum_{i,k=1}^{n}(X_{ls}V_{ji}+X_{js}V_{lk})\sigma_{ik}^{jl})] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\sum_{s=1}^{p}\sum_{j,l=1}^{m}\sum_{i,k=1}^{n}(X_{ls}E[V_{ji}]+X_{js}E[V_{lk}])\sigma_{ik}^{jl}) \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\sum_{s=1}^{p}\sum_{j,l=1}^{m}\sum_{i,k=1}^{n}(X_{ls}\mathbf{X}_{j,V}\boldsymbol{\gamma}+X_{js}\mathbf{X}_{l,V}\boldsymbol{\gamma})\sigma_{ik}^{jl}) \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\sum_{s=1}^{p}\sum_{j,l=1}^{m}(X_{ls}\mathbf{X}_{j,V}\boldsymbol{\gamma}+X_{js}\mathbf{X}_{l,V}\boldsymbol{\gamma})d_{jl}) \quad (\because \textit{the definition of } d_i) \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\sum_{s=1}^{p}\sum_{j,l=1}^{m}(X_{ls}\mathbf{X}_{j,V}+X_{js}\mathbf{X}_{l,V})d_{jl})\boldsymbol{\gamma} \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\gamma} \quad (\textit{From Equation } (73)) \\
&= \boldsymbol{\gamma}.
\end{aligned}$$

That is, $\hat{\boldsymbol{\gamma}}$ is an unbiased estimator.

3.6 Asymptotic distribution of the estimate $\hat{\boldsymbol{\beta}}$

Because $\hat{\boldsymbol{\beta}}$ is an estimator when the associated likelihood function is approximated. Therefore, in this section, we will show that $\hat{\boldsymbol{\beta}}$ which satisfying $1/m{\cdot}tr(A(\boldsymbol{\beta})^{-1}d\ A(\boldsymbol{\beta})/d\beta_i) = 0, i = 1, 2\cdots, q$ from Equation (57) converges to a true value, $\boldsymbol{\beta}_0$, and derive its asymptotic distribution. Let's define

$$H_i(\boldsymbol{\beta}) = \frac{1}{m}tr(A(\boldsymbol{\beta})^{-1}\frac{d\ A(\boldsymbol{\beta})}{d\beta_i}), \quad i = 1, 2, \cdots, q$$

The sample estimate $\hat{\boldsymbol{\beta}}$ can be obtained by solving $H_i(\boldsymbol{\beta}) = 0,\ i = 1, 2, \cdots, q$. To derive the asymptotic properties of $\hat{\boldsymbol{\beta}}$, we shall discuss the asymptotic properties of $H_i(\boldsymbol{\beta})$ first. In the following discussion, the technique of matrix characteristic decomposition will be used without trivial details. Let $\lambda_1(\boldsymbol{\beta}) \geq \lambda_2(\boldsymbol{\beta}) \geq \cdots \geq \lambda_m(\boldsymbol{\beta}) > 0$ be the eigenvalues of $A(\boldsymbol{\beta})$, and $q_j(\boldsymbol{\beta})$ satisfied $||q_j(\boldsymbol{\beta})|| = 1$ be the standardized eigenvectors with respect to $\lambda_j(\boldsymbol{\beta}), j = 1, 2, \cdots, m$, respectively. Then

we have

$$A(\boldsymbol{\beta}) = Q(\boldsymbol{\beta})\triangle(\boldsymbol{\beta})Q^T(\boldsymbol{\beta}) = \sum_{j=1}^{m} \lambda_j(\boldsymbol{\beta})q_j(\boldsymbol{\beta})q_j^T(\boldsymbol{\beta}),$$

where $Q(\boldsymbol{\beta}) = (q_1(\boldsymbol{\beta}), q_2(\boldsymbol{\beta}), \cdots, q_m(\boldsymbol{\beta}))$, an $m \times m$ orthogonal matrix, and $\triangle(\boldsymbol{\beta}) = diag(\lambda_1(\boldsymbol{\beta}), \lambda_2(\boldsymbol{\beta}), \cdots, \lambda_m(\boldsymbol{\beta}))$. Let

$$B_i(\boldsymbol{\beta}) = Q^T(\boldsymbol{\beta})A(\boldsymbol{\beta})^{-1}\frac{dA(\boldsymbol{\beta})}{d\beta_i}A(\boldsymbol{\beta})^{-1}Q(\boldsymbol{\beta}),$$

and the eigenvalues of $B_i(\boldsymbol{\beta})$ are $\lambda_{i1}(\boldsymbol{\beta}) \geq \lambda_{i2}(\boldsymbol{\beta}) \geq \cdots \geq \lambda_{im}(\boldsymbol{\beta})$. Their corresponding standardized eigenvectors are $r_{ij}(\boldsymbol{\beta})$. Let η is a random variable which converges to normal distribution with mean zero and variance $A(\boldsymbol{\beta})$ as $n \to \infty$, i.e., $\eta \to N(0, A(\boldsymbol{\beta}))$, then

$$\begin{aligned} E[\eta^T A(\boldsymbol{\beta})^{-1}\frac{d\ A(\boldsymbol{\beta})}{d\beta_i}A(\boldsymbol{\beta})^{-1}\eta] &\to tr(A(\boldsymbol{\beta})^{-1}\frac{d\ A(\boldsymbol{\beta})}{d\beta_i}A(\boldsymbol{\beta})^{-1}A(\boldsymbol{\beta})) && (74)\\ &= tr(A(\boldsymbol{\beta})^{-1}\frac{d\ A(\boldsymbol{\beta})}{d\beta_i}) && (75)\\ &= m \cdot H_i(\boldsymbol{\beta}) && (76) \end{aligned}$$

Because $E[\epsilon^T\Lambda\epsilon] = tr(\Lambda\Sigma)$ when $\epsilon \sim N(0, \Sigma)$. Therefore, for a large n, $H_i(\boldsymbol{\beta}) \approx \frac{1}{m}E[\eta^T A(\boldsymbol{\beta})^{-1}\frac{d\ A(\boldsymbol{\beta})}{d\beta_i}A(\boldsymbol{\beta})^{-1}\eta]$.

$$\begin{aligned} \eta^T A(\boldsymbol{\beta})^{-1}\frac{d\ A(\boldsymbol{\beta})}{d\beta_i}A(\boldsymbol{\beta})^{-1}\eta &= \eta^T Q(\boldsymbol{\beta})Q^T(\boldsymbol{\beta})A(\boldsymbol{\beta})^{-1}\frac{d\ A(\boldsymbol{\beta})}{d\beta_i}A(\boldsymbol{\beta})^{-1}Q(\boldsymbol{\beta})Q^T(\boldsymbol{\beta})\eta\\ &\quad (\because\ \ Q^T(\boldsymbol{\beta})Q(\boldsymbol{\beta}) = I) && (77)\\ &= \eta^T Q(\boldsymbol{\beta})B_i(\boldsymbol{\beta})Q^T(\boldsymbol{\beta})\eta && (78)\\ &= \sum_{j=1}^{m} \lambda_{ij}(\boldsymbol{\beta})[\eta^T Q(\boldsymbol{\beta})r_{ij}(\boldsymbol{\beta})r_{ij}^T(\boldsymbol{\beta})Q^T(\boldsymbol{\beta})\eta] && (79)\\ &= \sum_{j=1}^{m} \lambda_{ij}(\boldsymbol{\beta})\xi_{ij}^2 && (80) \end{aligned}$$

where $\xi_{ij} = r_{ij}^T(\boldsymbol{\beta})Q^T(\boldsymbol{\beta})\eta$. $\lambda_{ij}(\boldsymbol{\beta})$ and $r_{ij}(\boldsymbol{\beta})$ are eigenvalue and eigenvector of $B_i(\boldsymbol{\beta})$, and $B_i(\boldsymbol{\beta})$ is a symmetric matrix with a full rank. Eigenvalue decomposition theorem decomposes $B_i(\boldsymbol{\beta})$ by $R_i(\boldsymbol{\beta})$ and $\triangle_i(\boldsymbol{\beta})$, i.e., $B_i(\boldsymbol{\beta}) = R_i(\boldsymbol{\beta})\triangle_i(\boldsymbol{\beta})R_i^T(\boldsymbol{\beta})$ where $R_i(\boldsymbol{\beta}) = [r_{i1}(\boldsymbol{\beta}), r_{i2}(\boldsymbol{\beta}), \cdots, r_{im}(\boldsymbol{\beta})]$, an orthogonal matrix. The diagonal matrix

$\triangle_i(\boldsymbol{\beta})$ is $diag(\lambda_{i1}(\boldsymbol{\beta}), \lambda_{i2}(\boldsymbol{\beta}), \cdots, \lambda_{im}(\boldsymbol{\beta}))$. Then, $B_i(\boldsymbol{\beta}) = \sum_{j=1}^{m} \lambda_{ij}(\boldsymbol{\beta}) r_{ij}(\boldsymbol{\beta}) r_{ij}^T(\boldsymbol{\beta})$ derives Equation (78) to (79). Because $\eta \rightarrow N(0, A(\boldsymbol{\beta}))$, ξ_{ij} is asymptotically normal with mean zero and variance $v_{ij}^2(\boldsymbol{\beta})$. In addition, the eigenvectors are orthogonal each other. Thus, $\xi_{ij},\ j = 1, 2, \cdots, m$ are independent each other. The variance ,$v_{ij}^2(\boldsymbol{\beta})$, is

$$
\begin{aligned}
Var(\xi_{ij}) &= Var(r_{ij}^T(\boldsymbol{\beta}) Q^T(\boldsymbol{\beta}) \eta) && (81)\\
&= r_{ij}^T(\boldsymbol{\beta}) Q^T(\boldsymbol{\beta}) Var(\eta) Q(\boldsymbol{\beta}) r_{ij}(\boldsymbol{\beta}) && (82)\\
&= r_{ij}^T(\boldsymbol{\beta}) Q^T(\boldsymbol{\beta}) A(\boldsymbol{\beta}) Q(\boldsymbol{\beta}) r_{ij}(\boldsymbol{\beta}) && (83)\\
&= r_{ij}^T(\boldsymbol{\beta}) \triangle(\boldsymbol{\beta}) r_{ij}(\boldsymbol{\beta}) && (84)\\
&\quad (\because A(\boldsymbol{\beta}) = Q(\boldsymbol{\beta}) \triangle(\boldsymbol{\beta}) Q(\boldsymbol{\beta}) \ \&\ Q(\boldsymbol{\beta}) \textit{ is an orthogonal matrix})\\
&= v_{ij}^2(\boldsymbol{\beta}). && (85)
\end{aligned}
$$

To use Equation (80) in Equation (74), the expectation and the variance of Equation (80) are

$$
\begin{aligned}
E[\eta^T A(\boldsymbol{\beta})^{-1} \frac{d\ A(\boldsymbol{\beta})}{d\beta_i} A(\boldsymbol{\beta})^{-1} \eta] &= E[\sum_{j=1}^{m} \lambda_{ij}(\boldsymbol{\beta}) \xi_{ij}^2]\\
&= \sum_{j=1}^{m} \lambda_{ij}(\boldsymbol{\beta}) E[\xi_{ij}^2]\\
&= \sum_{j=1}^{m} \lambda_{ij}(\boldsymbol{\beta}) Var[\xi_{i,j}]\\
&= \sum_{j=1}^{m} \lambda_{ij}(\boldsymbol{\beta}) v_{ij}^2(\boldsymbol{\beta})\\
&= \sum_{j=1}^{m} \lambda_{ij}(\boldsymbol{\beta}) r_{ij}^T(\boldsymbol{\beta}) \triangle(\boldsymbol{\beta}) r_{ij}(\boldsymbol{\beta})
\end{aligned}
$$

$$
\begin{aligned}
Var[\eta^T A(\boldsymbol{\beta})^{-1} \frac{d\ A(\boldsymbol{\beta})}{d\beta_i} A(\boldsymbol{\beta})^{-1} \eta] &= Var[\sum_{j=1}^{m} \lambda_{ij}(\boldsymbol{\beta}) \xi_{ij}^2]\\
&= \sum_{j=1}^{m} \lambda_{ij}^2(\boldsymbol{\beta}) Var[\xi_{ij}^2]
\end{aligned}
$$

Because $\xi_{ij} \sim N(0, v_{ij}^2(\boldsymbol{\beta}))$, $\xi_{ij}/v_{ij}(\boldsymbol{\beta})$ follows a standard normal distribution. Then, $(\xi_{ij}/v_{ij}(\boldsymbol{\beta}))^2$ follows a chi-square distribution with degree of freedom $\lambda = 1$. Then $Var[(\xi_{ij}/v_{ij}(\boldsymbol{\beta}))^2] = Var[\xi_{ij}^2]/v_{ij}(\boldsymbol{\beta}))^4 = 2$. Thus, $Var[\eta^T A(\boldsymbol{\beta})^{-1} \frac{d\ A(\boldsymbol{\beta})}{d\beta_i} A(\boldsymbol{\beta})^{-1}\eta] = \sum_{j=1}^{m} \lambda_{ij}^2(\boldsymbol{\beta}) \cdot 2 \cdot v_{ij}^4(\boldsymbol{\beta})$. Now, for a large n, $H_i(\boldsymbol{\beta})$ can be equivalently written as

$$H_i(\boldsymbol{\beta}) = \frac{1}{m}\sum_{j=1}^{m} \lambda_{ij}(\boldsymbol{\beta}) v_{ij}^2(\boldsymbol{\beta}) = \frac{1}{m}\sum_{j=1}^{m} \lambda_{ij}(\boldsymbol{\beta}) r_{ij}^T(\boldsymbol{\beta}) \triangle(\boldsymbol{\beta}) r_{ij}(\boldsymbol{\beta})$$

where $\lambda_{ij}(\boldsymbol{\beta})$ is an eigenvalue of $B_i(\boldsymbol{\beta})$ and $r_{ij}(\boldsymbol{\beta})$ is a corresponding eigenvector. $\triangle(\boldsymbol{\beta})$ is a diagonal matrix, $\triangle(\boldsymbol{\beta}) = diag(\lambda_1(\boldsymbol{\beta}), \lambda_2(\boldsymbol{\beta}), \cdots, \lambda_m(\boldsymbol{\beta}))$, where $\lambda_i(\boldsymbol{\beta})$ is an eigenvalue of $A(\boldsymbol{\beta})$. Because neither $\lambda_{ij}(\boldsymbol{\beta})^2$ nor $v_{ij}^2(\boldsymbol{\beta})$ is a random variable, $Var[H_i(\boldsymbol{\beta})]$ goes to zero as $m \to \infty$. The estimate $\hat{\boldsymbol{\beta}}$ estimates a parameter which makes $H_i(\hat{\boldsymbol{\beta}}) = 0$ by using m samples. Therefore, $H_i(\boldsymbol{\beta}_0) \xrightarrow{wp1} 0$ as $m \to \infty$.
Denoted by $H(\boldsymbol{\beta}) = (H_1(\boldsymbol{\beta}), H_2(\boldsymbol{\beta}), \cdots, H_k(\boldsymbol{\beta}))^T$. Taylor expansion of $H(\boldsymbol{\beta})$ around $\boldsymbol{\beta}_0$ leads

$$H(\boldsymbol{\beta}) = H(\boldsymbol{\beta}_0) + \frac{d\ H}{d\ \boldsymbol{\beta}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + O((\boldsymbol{\beta} - \boldsymbol{\beta}_0)^2).$$

The estimator $\hat{\boldsymbol{\beta}}$ makes $H(\hat{\boldsymbol{\beta}}) = 0$. Thus,

$$\begin{aligned} H(\hat{\boldsymbol{\beta}}) &= H(\boldsymbol{\beta}_0) + \frac{d\ H}{d\ \boldsymbol{\beta}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + O((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^2) = 0 \\ H(\boldsymbol{\beta}_0) &= -\frac{d\ H}{d\ \boldsymbol{\beta}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - O((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^2) \\ &= (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \times (-\frac{d\ H}{d\ \boldsymbol{\beta}} + O(|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|)) \end{aligned}$$

We know that $H(\boldsymbol{\beta}_0) \xrightarrow{wp1} 0$ as $m \to \infty$. Thus,$\hat{\boldsymbol{\beta}} \xrightarrow{wp1} \boldsymbol{\beta}_0$. That is, $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}_0$.

The variance of $H_i(\boldsymbol{\beta})$ is

$$\begin{aligned} Var[H_i(\boldsymbol{\beta})] &= \frac{1}{m^2}\sum_{j=1}^{m}\lambda_{ij}^2(\boldsymbol{\beta})Var[v_{ij}^2(\boldsymbol{\beta})] && (86)\\ &= \frac{1}{m}\sum_{j=1}^{m^2}\lambda_{ij}^2(\boldsymbol{\beta})Var[r_{ij}^T(\boldsymbol{\beta})\triangle(\boldsymbol{\beta})r_{ij}(\boldsymbol{\beta})] && (87)\\ &= \frac{1}{m^2}\sum_{j=1}^{m}\lambda_{ij}^2(\boldsymbol{\beta})Var[r_{ij}^T(\boldsymbol{\beta})Q^T(\boldsymbol{\beta})A(\boldsymbol{\beta})Q(\boldsymbol{\beta})r_{ij}(\boldsymbol{\beta})] && (88)\\ &= \frac{1}{m^2}\sum_{j=1}^{m}\lambda_{ij}^2(\boldsymbol{\beta})w_{ij}^2(\boldsymbol{\beta}) && (89)\\ & && (90) \end{aligned}$$

Then the limit of $Var[\sqrt{m}H_i(\boldsymbol{\beta})]$ exists as a function of $\boldsymbol{\beta}$:

$$\begin{aligned} lim_{n,m\to\infty}Var[\sqrt{m}H_i(\boldsymbol{\beta})] &= lim_{m\to\infty}mVar[H_i(\boldsymbol{\beta})] && (91)\\ &= lim_{m\to\infty}m\cdot\frac{1}{m^2}\sum_{j=1}^{m}\lambda_{ij}^2(\boldsymbol{\beta})w_{ij}^2(\boldsymbol{\beta}) && (92)\\ &= lim_{m\to\infty}\frac{1}{m}\sum_{j=1}^{m}\lambda_{ij}^2(\boldsymbol{\beta})w_{ij}^2(\boldsymbol{\beta}) && (93)\\ &= \mathbf{U}(\boldsymbol{\beta}) && (94) \end{aligned}$$

In addition, $Var[H_i(\boldsymbol{\beta})]$ converges to zero as m goes to infinity. The Lindeberg-Feller condition can be applied to prove the asymptotic normality of $H(\boldsymbol{\beta})$, i.e., $\sqrt{m}H(\boldsymbol{\beta}) \xrightarrow{d} N(0, \mathbf{U}(\boldsymbol{\beta}))$. From $H(\hat{\boldsymbol{\beta}}) = 0$,

$$\sqrt{m}[H(\boldsymbol{\beta}) - H(\hat{\boldsymbol{\beta}})] = \sqrt{m}[H(\hat{\boldsymbol{\beta}}) - H(\boldsymbol{\beta})] \xrightarrow{d} N(0, \mathbf{U}(\boldsymbol{\beta})).$$

By delta theorem, the above equation goes to

$$\sqrt{m}[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0] \xrightarrow{d} N(0, \Sigma(\boldsymbol{\beta}_0)),$$

where

$$\Sigma(\boldsymbol{\beta}_0) = \left(\frac{dH(\boldsymbol{\beta}_0)}{d\boldsymbol{\beta}}\right)\mathbf{U}(\boldsymbol{\beta}_0)\left(\frac{dH(\boldsymbol{\beta}_0)}{d\boldsymbol{\beta}}\right)^T.$$

3.7 Summary of estimators

When the output Y_{ij} is semi-continuous and has with-in and between variations, we propose the constrained random effect models. Conditioning on $U_i = 1$, the output decomposed into two random variables, U_i and V_{ij} having two models.

$$Logistic\ regression: \quad \eta_i = \mathbf{X}_{i,\eta}\boldsymbol{\beta} + D_i$$

$$Normal\ regression: \quad V_{ij} = \mathbf{X}_{i,V}\boldsymbol{\gamma} + T_i + E_{ij}$$

where $\eta_i = log[P(U_i = 1)/(1 - P(U_i = 1))]$. Both models have between variations, E_i and T_i, and the normal regression function has within variation, E_{ij}. All variations are random variables. The distribution of between variations are derived by Kriging, as described in Section 3.4. Then, both $\mathbf{D}$ and $\mathbf{T}$ follows a normal distribution, and their distribution notates as $N(0, \Sigma_D)$ and $N(0, \Sigma_T)$, respectively. In addition, the noise is assumed as $\mathbf{E}_i \sim N(0, \Sigma_{E_i})$.

The estimate of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, in logistic regression model is derived from the approximated likelihood function, while the estimate of $\boldsymbol{\gamma}$, $\hat{\boldsymbol{\gamma}}$, in normal regression model is directly derived from its likelihood function. From the exact likelihood function of the normal regression model, the estimate of $\boldsymbol{\gamma}$, $\hat{\boldsymbol{\gamma}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\sum_{s=1}^{p}\sum_{j,l=1}^{m}\sum_{i,k=1}^{n}(X_{ls}V_{ji} + X_{js}V_{lk})\sigma_{ik}^{jl})$, where $\mathbf{X} = \sum_{s=1}^{p}\sum_{j,l=1}^{m}(X_{ls}\mathbf{X}_{j,V} + X_{js}\mathbf{X}_{l,V})\sum_{i,k=1}^{n}\sigma_{ik}^{jl}$, and σ_{ik}^{jl} is an (i,k)-th element in A_{jl} from Σ_V, a covariance matrix of $\mathbf{V}$.

The estimate $\hat{\boldsymbol{\beta}}$ which is explained in Section 3.5 satisfies the following q equations when q is the dimension of $\boldsymbol{\beta}$.

$$\frac{1}{m}tr(A(\boldsymbol{\beta})^{-1} \cdot \frac{dA(\boldsymbol{\beta})}{d\beta_i}) = 0,\ i = 1, 2, \cdots, q$$

where $A(\boldsymbol{\beta}) = \Sigma_\delta^{-1} + diag[\cdots, exp(X_{i,\eta}\beta + \tilde{\delta}_i)/((1 + exp(X_{i,\eta}\beta + \tilde{\delta}_i))^2), \cdots]$. The $\tilde{\boldsymbol{\delta}}$ is a value such that $\partial h(\mathbf{U}, \boldsymbol{\delta}, \boldsymbol{\beta})/d\boldsymbol{\delta} = 0$ and $h(\mathbf{U}, \boldsymbol{\delta}, \boldsymbol{\beta}) = \mathbf{U}^T\boldsymbol{\eta} - log[\prod_{i=1}^{m}(1 + exp(\eta_i))] - \boldsymbol{\delta}^T\Sigma_\delta^1\boldsymbol{\delta}/2$. Section 3.6 derives the asymptotic distribution of $\hat{\boldsymbol{\beta}}$:

$$\sqrt{m}[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0] \xrightarrow{d} N(0, \Sigma(\boldsymbol{\beta}_0)),$$

where

$$\Sigma(\boldsymbol{\beta}_0) = \left(\frac{dH(\boldsymbol{\beta}_0)}{d\boldsymbol{\beta}}\right)\mathbf{U}(\boldsymbol{\beta}_0)\left(\frac{dH(\boldsymbol{\beta}_0)}{d\boldsymbol{\beta}}\right)^T,$$

$$H(\boldsymbol{\beta}) = [H_1(\boldsymbol{\beta}), H_2(\boldsymbol{\beta}), \cdots, H_q(\boldsymbol{\beta})]^T$$

$$H_i(\boldsymbol{\beta}) = \frac{1}{m}\sum_{j=1}^{m}\lambda_{ij}(\boldsymbol{\beta})v_{ij}^2(\boldsymbol{\beta}) = \frac{1}{m}\sum_{j=1}^{m}\lambda_{ij}(\boldsymbol{\beta})r_{ij}^T(\boldsymbol{\beta})\triangle(\boldsymbol{\beta})r_{ij}(\boldsymbol{\beta})$$

$$A(\boldsymbol{\beta}) = {\Sigma_\delta}^{-1} + diag[\cdots, \frac{exp(\mathbf{X}_{i,\eta}\boldsymbol{\beta} + \tilde{\delta}_i)}{(1 + exp(\mathbf{X}_{i,\eta}\boldsymbol{\beta} + \tilde{\delta}_i))^2}, \cdots]$$

$$\triangle(\boldsymbol{\beta}) = diag(\lambda_1(\boldsymbol{\beta}), \lambda_2(\boldsymbol{\beta}), \cdots, \lambda_m(\boldsymbol{\beta}))$$

$$B_i(\boldsymbol{\beta}) = Q^T(\boldsymbol{\beta})A(\boldsymbol{\beta})^{-1}\frac{dA(\boldsymbol{\beta})}{d\beta_i}A(\boldsymbol{\beta})^{-1}Q(\boldsymbol{\beta}),$$

$\lambda_{ij}(\boldsymbol{\beta})$ and $r_{ij}(\boldsymbol{\beta})$ are the eigenvalue and the eigenvector of $B_i(\boldsymbol{\beta})$, respectively. $\lambda_i(\boldsymbol{\beta})$ is the eigenvalue of $A(\boldsymbol{\beta})$, and $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$. Therefore, the estimate $\hat{\boldsymbol{\beta}}$ based on the approximated likelihood function converges to the true value $\boldsymbol{\beta}_0$.

3.8 Discussion

Semi-continuous data arise in many applications in which naturally-continuous data become contaminated by the data-generating mechanism. This chapter proposed an approach for modeling semi-continuous data with within- and between-batch variations. Similar to many papers dealing with semi-continuous data, this chapter has two models: a logistic model and a normal model. Both models include the between-batch variation, while only the normal model includes the within-batch variation. Between-batch variation is correlated for the different batches (e.g., $cov(D_i, D_j) \neq 0$), while the within-batch variation is not correlated for different particles in a batch (e.g., $cov(E_{i,j}, E_{i,k}) = 0$). Under these circumstances, we estimated the coefficients in two models.

A Taylor approximation estimates likelihood functions. Because true likelihood functions are too complicated to obtain, we use an approximation of the likelihood functions. Because the coefficient estimates are based on an approximation of the

likelihood functions, we have to be careful whether the coefficient estimates based on the approximated likelihood functions ($\hat{\beta}_{\tilde{L}}$) are close to those based on the real likelihood functions ($\hat{\beta}_L$). Therefore, we derive the asymptotic distribution of $\hat{\beta}$ which converges to the true value, β_0.

A Newton-Raphson method, which is used to find $\tilde{\delta}$ in the logistic model, is the best known method for finding a better approximation of the roots of a real-value function. The method converges quickly when the iteration begins near the desired root, which can be both a strength and a weakness of the method. When the iteration begins far from the desired root, the method can fail to converge. Therefore, choosing a starting point is the key to success. Currently, this chapter assumed that a Newton-Raphson method can successfully find $\tilde{\delta}$.

In addition, several works could be added in this chapter to validate and strengthen the methodology. Future work includes illustrative examples to prove whether the research result is correct or not. Thus, finding a good example and collecting a set of data are necessary.

CHAPTER IV

DETERMINING A SAMPLE SIZE TO ESTIMATE AN OPTIMUM AND A PROCESS WITH A MULTI-LAYER SYSTEM

4.1 Introduction

High quality modeling has been a priority goal for statisticians ever since statisticians began modeling processes. Particularly, with complicated processes, such as nanofabrication, modeling is challenging but critical. Currently, there is enormous widespread interest in the field of nanotechnology, and we can expect that nanotechnology will impact every sector of our economy and daily lives in the near future. However, processes in nanotechnology are extremely sensitive so that even minor changes to a processes settings may lead to different results [39]. Therefore, extra attention and effort is necessary for complicated processes. However, in most existing literature, the synthesis of nanomaterials lack theoretical guidance for achieving high quality and reproducible nanomaterials [16]. Consequently, enormous amounts of data are required to model the unclear and complicated nanofabrication process.

For the purpose of delicate process modeling, we propose a *multi-layer* system which focuses on the process optimum. In addition, model estimation around the optimum should return better quality results than models that are not focused around the optimum. The main idea is to zoom the explanatory domain around the process optimum. The first, initial layer is the overall explanatory domain. The second layer is the neighborhood of the optimum which is estimated from the model built from the first layer, and is thus a sub-region of the first layer. Therefore, as the layers go down, the domain size becomes smaller around the true optimum. See Section 4.2

for a more detailed explanation regarding the multi-layer system. When resources are limited, allocating the resources around the optimum is necessary in order to attain the best estimation. Therefore, samples are collected sequentially from each layer. However, sample collection for multi-layered systems have never been explored before. In the multi-layer system, samples are used to develop a model for the next layer until the bottom layer is reached. At the bottom layer, a model developed by the collected sample should meet the tolerance requirement. Therefore, existing sample size calculation methodologies cannot be applied in the multi-layer system, and we need to develop a unique sample size calculation procedure for the multi-layer system. The challenges in developing the new procedure are (1) to define different goals for the sample collections by layers and (2) to aggregate the information from previous layers into the current layer sample size calculation.

In the past, researchers have attempted to develop the sample size calculation for multi-stage system. In 1945, Stein [66] investigated a two-stage procedure for the first time. The paper provides a sample size for two stages in which the second result depends on the first result. Based on the Stein's method, many follow-up methods were proposed. Lehmacher [38] updated the method for a group of sequential trials that is based on the inverse normal method. Other than the number of stages, the major difference between the existing methodologies and our proposed method is the explanatory domain range. Mainly, the sizes of the stages are the same, but our method assumes that the latter layer is included in the previous layer. In 2011, Wang and Mei [72] suggested an asymptotically optimal scheme under a Bayesian setting. The paper used "two-stage" tests to find the optimum from discrete candidates. The restriction of discreteness prevents the application of this paper to the nano-fabrication process. Recently, Kim [35] and Casciato et al. [11] proposed the layers of experiments with adaptive combined design. The methodologies in these papers are suitable for the delicate process modeling, however they focused on the

location for the data collection with the assumption of a known sample size. In addition, Soltanian [63] shows how large of an effect the sample size has on nano-scale data. Therefore, this chapter proposes a novel method that selects a sub-region for the purpose of accurate model estimation around the optimum when the process is too complicated to meet the tolerance requirement easily. Consequently, the sample size calculation associated to the multi-layer system is suggested.

In the present chapter, we have three objectives. First, we introduce the multi-layer system to to explain the nano-frabrication process more effectively. Second, we set goals of the sample collection. To evaluate the goals, we define hypotheses to test regarding the goals. Finally, we provide a procedure for determining the minimum sample size necessary to achieve the goals.

The rest of the chapter is organized as follows. Section 4.2 explains why the multi-layer system is important and how it works in detail. Section 4.3 introduces the methodology of determining the sample size. Section 4.4 provides technical algorithms to present the methodology explained in Section 4.3. Section 4.5 derives the distribution of a test static, and an example is illustrated in Section 4.6. Section 4.7 presents the methodology properties, and Section 4.8 gives some concluding remarks.

4.2 The Multi-layer system

The ultimate goal of this research is to build a model with accuracy meeting the tolerance requirement near the process optimum. However, because the true optimum is unknown and the model requires collecting and analyzing a vast amount of data due to the complexity of the nanofabrication process, we instead focus on collecting the minimum number of samples needed to construct a precise model around the optimum. This chapter introduces a multi-layer system that identifies the optimum after samples are collected. Figure 17 illustrates the concept of the proposed multi-layer system.

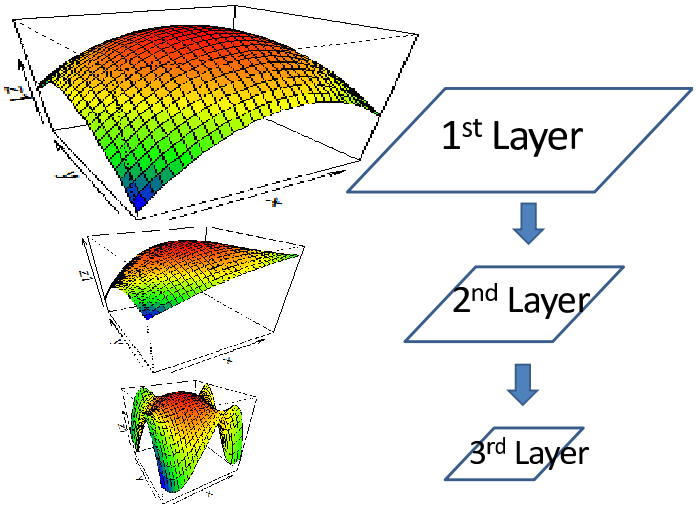

Figure 17: The concept of the multi-layer. The left column figures show statistical models for each layer, and the right column figures represent the underlying explanatory domains. For both columns, top figures represent the initial layer, the middle layer represents the second layer, and the bottom layer is the third/last layer.

In the multi-layer system, the first layer, Layer 1, is the overall explanatory domain. Some samples are collected from the first layer. The statistical model constructed with the samples roughly shows the global trend of the underlying process. Suppose we want to maximize the outcome; the vertex of the concave curve in Figure 17 is the optimum from Layer 1, and Layer 2 is in the neighborhood of the vertex. In this chapter assume that the neighborhood is a sphere with a center point as the estimated optimum and a pre-specified radius m, i.e. when $x \in$ Layer 2, the distance between x and the estimated optimum from the Layer 1 is at most m. From Layer 2, we collect additional samples to build another model over it. Each model produces an estimation of the optimum, and the optimum decides the location of the following layer. Therefore, each model should be confident, or reliable enough, to determine an accurate estimation of the optimum, i.e. by providing the correct direction to the subsequent layer (Objective 1: model assessment) until any model in any layer meets the tolerance requirement (Objective 2: tolerance requirement). The "zoom-in" procedure stops when the statistical model from a layer meets the

tolerance requirement.

Another factor, budgetary constraints, plays a key role in the sample size decision. Clearly, the process budget has to cover a number of samples in the multi-layer methodology. However, once the layers go into the wrong direction, the optimum cannot be detected, and the budget is exceeded. Thus, the model assessment must determine the minimum sample size and the budget constraint must determine the maximum sample size. When the sample size is n, then $n \leq c, n \geq m$, where c is the boundary of the budget constraint and m is the boundary of the first constraint. Therefore, we need to find the sample size to satisfy both constraints.

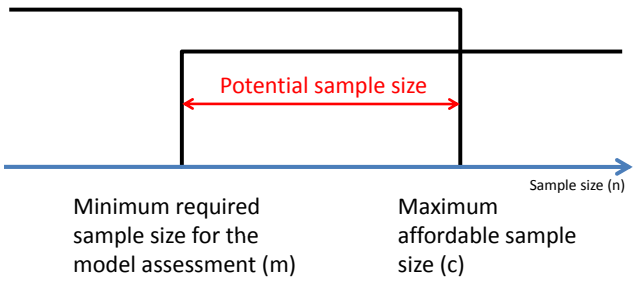

Figure 18: Constraints for Sample Size Decision. The model assessment gives a lower bound and the budget gives an upper bound.

Figure 18 denotes the potential sample size that satisfies both constraints. The two constraints should be strictly followed. If the maximum affordable sample size is smaller than the minimum required sample size for the model assessment, increasing the budget or changing the model structure allows us to determine the optimal sample size. The range of the potential sample size (n) is $m \leq n \leq c$, where m is the minimum required sample size for the model assessment objective, and c is the maximum number of sample size within a budget constraint.

4.3 Methodology

For the sample size calculation, we assume the following conditions:

1. Layer size structure

   The formulation of the multi-layer system tells us that a lower layer size is always smaller than the sizes of all upper layers. However, the number of explored layers is unknown in advance. Thus, the layer size will be a relative number rather than an absolute number. That is, the lower layer size depends on the previous layer size. For example, Layer $i+1$ is half the size of Layer i.

2. Tolerance requirement

   The tolerance requirement represents the maximum acceptable error for the target value. The engineers who use the proposed multi-layer system will need to determine the tolerance requirement for guiding the sample size decision procedure.

Sequential sample collection requires a consistent space-filling design for every layer. We examine six designs: optimal packing of sphere inside a cube, uniform, Latin hyper cube with optimal spacing, minimum energy design in a spherical region, maximum entropy design for a Gaussian process, and an integrated mean square error (MSE) optimal design for a Gaussian process. We observe that the minimum energy design in a spherical region spreads points around the center of a sphere. Therefore, we select the minimum energy design due to the robustness. We use the design using statistical software JMP in this research.

Consider n points on a k-dimensional sphere around the zero vector. Each point can have values between $[-1, 1]$ for k continuous factors. Let d_{ij} be the distance between i-th and j-th points. Next, the optimization technique finds the $n \times k$ values

(n points from k-dimensional) that minimize total energy E, where

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(d_{ij}^2 + \frac{1}{d_{ij}} \right).$$

Thus, the minimum energy design decides the n design points by *minimizing* the energy.

Recall that the decision of a sample size has two objectives: the model assessment and the tolerance requirement. Because the number of layers is undefined, every layer requires the determination of the final layer. The evaluation metric under the tolerance requirement determines the final layer. Therefore, at every layer, the required sample size based on the evaluation metric is calculated and compared to the budget constraint. If the sample size for the tolerance requirement does not meet the budget constraint, then the sample size for the model assessment is calculated. The following two subsections explain the procedures to find the minimum sample size for each objective.

4.3.1 Objective 1: Model assessment

At any layer, say Layer i (L_i), n_i samples collected from L_i estimate the process within the layer, and let $\hat{f}_i$ be the estimated model. If $\hat{f}_i$ meets a tolerance requirement, then Layer i becomes the bottom layer. Otherwise, the purpose of the model $\hat{f}_i$ is to identify the next layer location. Then, the analysis of $\hat{f}_i$ will estimate the optimum at L_i and locate the next layer, L_{i+1}, around the optimum. Therefore, the fitted model in L_i, $\hat{f}_i$ highly affects the location of L_{i+1}. However, once the layers go into a wrong direction, the multi-layer system cannot find a true optimum. That is, the quality of $\hat{f}_i$ is highly significant, and the sample size n_i has to be large enough to give a correct direction at any layer.

In summary, our strategy is to estimate a potential optimum point and then to identify the layer which is the neighborhood around the point with a radius r. In this chapter, we call the radius, *layer size*. Then, the window with center point x and

layer size (radius) r notes $B(x;r)$, i.e. the window is a sphere with a radius r. Let the true optimum point be μ, and the estimated optimum (the center point of the next layer) be $\tilde{X}$. When $d(\cdot,\cdot)$ is a distance function between the two points, such as Euclidean norm, then $d(\mu,\tilde{X})$ serves the distance between μ and $\tilde{X}$. The estimated center point $\tilde{X}$ is the local maximum or minimum. Therefore, the first derivative of the function $\hat{f}$ derives $\tilde{X}$ such that $(\nabla\hat{f})_{\tilde{X}}=0$. In addition, the second derivative, or Hessian matrix, at $\mathbf{x}=\tilde{X}$ should be considered to avoid the saddle point. In other words, a hessian matrix of $\hat{f}$, $H(\hat{f})$ should not be indefinite to be a local extreme. The confidence interval of $\tilde{X}$ is a function of the sample size n. For example, when $\tilde{X}$ follows a t-distribution, the confidence interval of $\tilde{X}$ is

$$\tilde{X}\pm t_{\alpha/2,n-p-1}\sqrt{var(\tilde{X})}, \tag{95}$$

where n is the sample size, p is the number of variables, α is the level of significance, and $t_{\alpha/2,n-p-1}$ is the t-distribution with $1-\alpha/2$ percentile and $n-p-1$ degree of freedom.

After the $\tilde{X}$ estimation, the result can be categorized into three cases.

Case 1: The estimated window $B(\tilde{X};r)$ contains the true mean, i.e., $d(\mu,\tilde{X})\leq r$.

Case 2: The estimated window $B(\tilde{X};r)$ does not contain the true mean, but $B(\tilde{X};r)$ and $B(\mu;r)$ have a nonempty intersection, i.e., $r<d(\mu,\tilde{X})\leq 2r$.

Case 3: both $B(\tilde{X};r)$ and $B(\mu;r)$ are disjoint, i.e., $d(\mu,\tilde{X})>2r$.

Note that Case 1 is desirable, yet Case 2 is also acceptable, because there is a chance to detect the true optimum in the next layer. For example, if the estimated optimum in the layer, $B(\tilde{X};r)$, is located on the boundary of $B(\tilde{X};r)$ to the direction of μ, then the following layer can contain the true optimum. However, Case 3 encounters an undesirable situation. When $\tilde{X}$ and μ are distanced apart more than $2r$, the estimated

optimum even on the boundary of $B(\tilde{X}; r)$ cannot capture the true optimum in the following layer. That is, the direction of the important "zoom-in" region is incorrect, which is a situation we wish to avoid. Therefore, we develop and conduct a hypothesis test to determine whether $\tilde{X}$ yields a correct or an incorrect direction to the true optimum.

Hypothesis Test	
$H_0 : d(\mu, \tilde{X}) = 2r$	$H_1 : d(\mu, \tilde{X}) \leq 2r$

The $\tilde{X}$ is derived from a process model estimation ($\hat{f}$) by a set of sample data. Let the size of the set be n, and the true coefficients and their estimate for the process model, β and $\hat{\beta}$, respectively. The estimated optimum $\tilde{X}$ depends on the sample size n and the corresponding coefficient estimates $\hat{\beta}$, i.e., $\tilde{X} = \tilde{X}_n(\hat{\beta})$. The test statistic is

$$T_{\mu,n} = \frac{\hat{D} - 2r}{\sigma_{\hat{D}}} \tag{96}$$

where $\hat{D} = d(\mu, \tilde{X}_n(\hat{\beta}))$ is an estimator of $d(\mu, \tilde{X})$ and $\sigma_{\hat{D}}$ is a standard deviation of the estimator. The estimator $\hat{D}$ estimates the distance between the true optimum and the estimated optimum, i.e., $\hat{D}$ = effect size. We perform the bootstrap resampling to estimate the distribution of the estimate. When m is the bootstrapping number, the m outcomes, $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_m$, formulate the estimator of the effect size. Then, the estimator is defined as $\hat{D} = max|\tilde{X}_i - \tilde{X}_j|$, $i, j \in [1, m]$. The standard deviation of $\hat{D}$, $\sigma_{\hat{D}}$, is unknown. Thus its estimate, $\hat{\sigma}_{\hat{D}}$, will substitute the denominator. $\hat{\sigma}_{\hat{D}}$ will be derived from the distribution of $\tilde{X}$. The detail regarding the derivation is explained in Section 4.5. Both $\hat{D}$ and $\sigma_{\hat{D}}$ are random variables that depend on the sample size n. Thus the test statistic is a function of n as well.

Because our methodology is based on the multi-layer system, the standard deviation of $\hat{D}$, which is the denominator of the test statistic, needs to be calibrated by the standard deviations of $\hat{D}$ from the previous layers. We include the calibration

step in the bootstrap resampling (see Section 4.4.1 for the explanation). The decision rule for the hypothesis test is that H_0 is rejected when the test statistic is smaller than a critical value, a function of a sample size n. If the test statistic follows a t-distribution, the critical value is $-t_{\alpha,n-p-1}$, i.e., the decision rule is to reject H_0 when $TestStatistic < -t_{\alpha,n-p-1}$. The rejection of the hypothesis test refers that the distance between the true optimum and the estimated optimum $\tilde{X}$ is smaller than the radius of the next layer, that is, there is a chance to estimate the optimum in the next layer. When a sample size for a model estimation increases, the estimation result becomes more accurate. Thus, the goal is to find the minimum required sample size for the next layer estimation (See Section 4.5 for the explanation of how to derive the distribution of the test statistic.) For the simplicity, we introduce the procedure by assuming that the test statistic follows a t-distribution. However, at the end, the t-distribution will be replaced by the actual distribution of the test statistic derived in Section 4.5.

4.3.2 Objective 2: Tolerance requirement

The tolerance requirement depends upon the confidence of the estimator. If a model estimation is less reliable, it is highly possible that the estimated optimum based on the model will be incorrect. Therefore, the quality of the model estimation is critical. Generally, the confidence interval of a function can be a measurement of the model estimation accuracy in a region R. Clearly, a wider confidence interval indicates that a model is less reliable than one with the a smaller confidence interval. In a certain region R, we can calculate the confidence interval of parameter estimates ($\hat{\beta}$), and it derives the confidence interval of the output estimate ($\hat{y} = X\hat{\beta}$) for any explanatory input X in R ($\forall X \in R$). Since the tolerance requirement specifies the upper bound, we want to define the upper bound of the confidence interval to compare the quality of various functions. The evaluation metric defines the maximum confidence interval

among every X in R, i.e., the evaluation metric $L(R)$ is

$$L(R) \;=\; \max_{X_p \in R} \; (2t_{\alpha/2,n-p+1}\sqrt{MSE \times (X_p(X_D^T X_D)^{-1}X_p^T)}), \tag{97}$$

where p is the order of the model, MSE represents the mean squared error, and it estimates the model variance. X_D is a design matrix that estimates the model parameters, and X_p is a model structure at X, such as $X_p = [1,\ X,\ X^2]$ when $p = 2$. When $f(x) = X\beta + \varepsilon$, the variance of $\hat{f}(x)$ is $Var(\hat{f}(x)) = Var(X\hat{\beta}) = XVar(\hat{\beta})X^T$. The variance of a parameter estimation is $\sigma_\varepsilon^2(X_D^T X_D)^{-1}$. MSE substitutes σ_ε^2, because MSE is an unbiased estimator of σ_ε^2. Therefore, X_D is a fixed matrix in the region R, but X_p changes by X. For example, let $p = 2$, $R = [0, 25]$, and four samples $(5, 10, 15, 20)$ are used for the model estimation. Then the design matrix X_D is

$$X_D = \begin{bmatrix} 1 & 5 & 5^2 \\ 1 & 10 & 10^2 \\ 1 & 15 & 15^2 \\ 1 & 20 & 20^2 \end{bmatrix}.$$

The estimate of σ^2, MSE, also depend on those four points. Now, the confidence interval for any X in R is $2t_{\alpha/2,n-p+1}\sqrt{MSE \times (X_p(X_D^T X_D)^{-1}X_p^T)}$, where X_p depends on X. For $X = 2$, X_p is $[1, 2, 2^2]$, and the confidence interval for $X = 2$ can be calculated with X_p and X_D. Note that the confidence interval for $X = 2$ is different to the one for $X \neq 2$, because X_p is different by X, while X_D stays the same. Evaluation metric $L(R)$ represents the maximum confidence interval within R model.

As we collect more samples, the estimated model becomes more stable, which means the model variance σ_ϵ^2 becomes smaller. Because the variance of parameter estimate is $Var(\hat{\beta}) = (X^T X)^{-1}\sigma_\epsilon^2$, there is less variation of the parameter estimate $\hat{\beta}$, and the confidence interval of $\hat{\beta}$ narrows. Thus, the sample size increases until the evaluation metric $L(R)$ is smaller than the tolerance requirement δ.

4.3.3 Stopping criteria

Let n_1^i and n_2^i be the sample size for the model assessment (Objective 1) and the tolerance requirement (Objective 2), respectively. Assume that a single experiment costs a constant c amount. Let the overall experimental cost be C_k when the multi-layer system has k layers. The objective of the sample collection in a layer is to perform a model assessment until the layer becomes the final(bottom) layer. At the bottom layer, the objective of the sample collection is to meet the tolerance requirement. Therefore, from Layers $1, 2, \cdots, k-1$, samples with the size of $n_1^1, n_1^2, \cdots, n_1^{k-1}$ are collected, and the bottom layer (L_k) has n_2^k samples. The cumulative collected sample size is $n_1^1 + n_1^2 + \cdots + n_1^{k-1} + n_2^k$. Thus, the overall experimental cost for the system is

$$C_k = c \times (\sum_{j=1}^{k-1} n_1^j + n_2^k).$$

If C_k is smaller than the budget, stop the investigation at Layer k; otherwise, explore the next layer. If the remaining budget of the total sample size cannot cover the needed sample size to meet the goal in this step, a warning to the user should be provided. Then, the current estimate of the process recipe is given at this layer. No subsequent layer will be studied unless more data budget is allocated.

4.4 Algorithm

In order to assist practitioners in applying the proposed methodology, we provide a step-by-step guideline to determine the number of layers and the sample size in each layer. Until the minimum required sample sizes for both objectives are decided, no experiment in either computer simulation or physical experiment is performed. The initial sample size n will be decided by the underlying model structure. Figure 19 provides the overall procedure for the sample size decision when entering a new layer.

Algorithm 1: Model assessment

Step 0. Given condition: layer size (r)

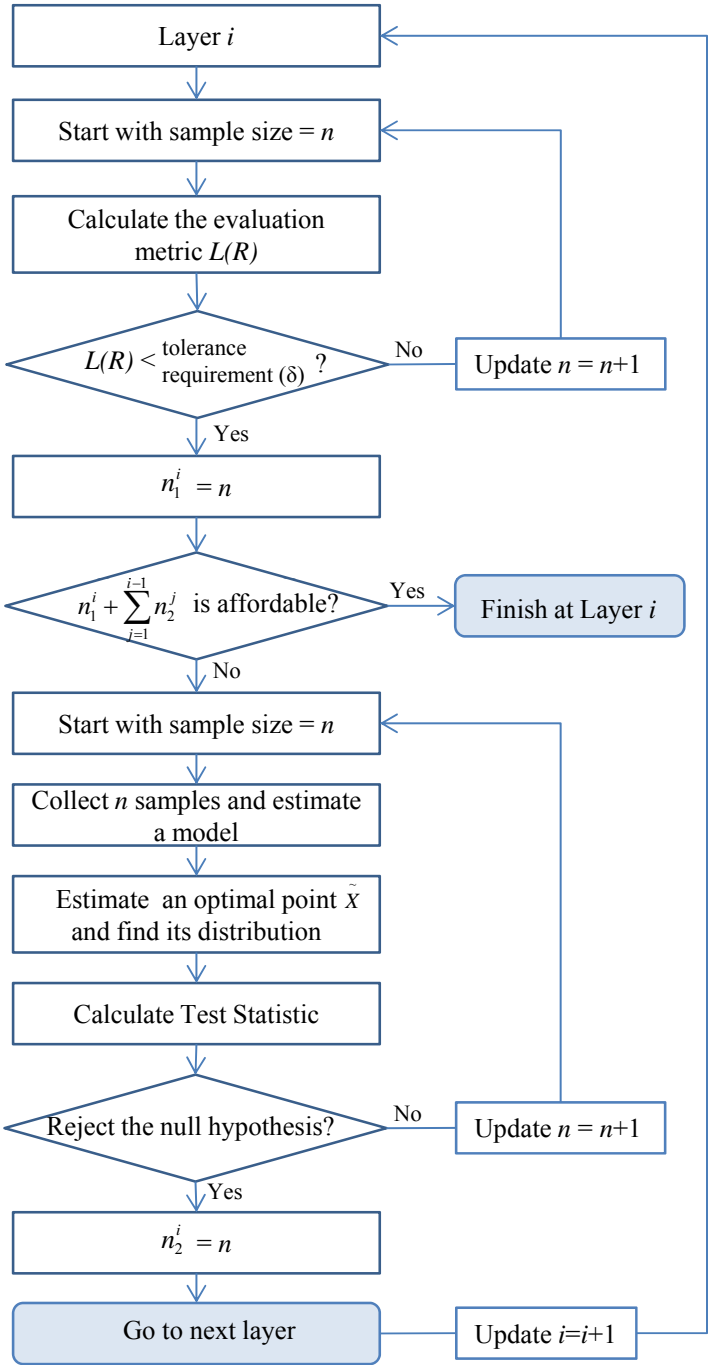

Figure 19: Flowchart for experimental developers: Comprehensive algorithm combined with Algorithms 1 and 2.

Step 1. Assume a fitting model structure (e.g., 2-nd order or 3-rd order)

Step 2. Choose an initial sample size n.

Step 3. Collect n samples by the minimum energy design from the layer

Step 4. Estimate a statistical model $\hat{f}(x) = X\hat{\beta}$

Step 5. Estimate the distribution of $\tilde{X}$ through the bootstrap resampling

(Step 5-1). If $\tilde{X}$ distributes skewed, then a certain transformation makes it normal. Find the distribution of $\tilde{X}$ as well as the mean and the variance.

Step 6. Calculate the estimator $\hat{D} = max|X_i - X_j|$

Step 7. Estimate the distribution of $\hat{D}$

Step 8. Calculate a test statistic $T = (\hat{D} - 2r)/\hat{\sigma}_{\hat{D}}$ to perform a hypothesis test $H_0 : d(\mu, \tilde{X}) = 2r$.

Step 9. Compare the test statistic to t-value, $t_{\alpha,n-p-1}$

Step 10. Repeat Steps 2 to Step 10 by increasing one sample size ($n = n + 1$) until the hypothesis test is rejected

Step 11. (Decision) : If the hypothesis test is rejected for n^* samples, then n^* is the minimum sample size requirement in the layer.

Algorithm 2: Tolerance requirement

Step 0. Given condition: layer size (r)

Step 1. Assume a fitting model structure (e.g., 2-nd order, 3-rd order)

Step 2. Choose an initial sample size n

Step 3. Collect n samples by minimum energy design from the layer

Step 4. Estimate a statistical model $\hat{f}(x) = X\hat{\beta}$ by linear regression

Step 5. Calculate the evaluation metric $L(R)$ in the layer

Step 6. Compare the metric to the tolerance requirement

Step 7. Repeat Steps 2 to Step 6 by increasing one sample size ($n = n + 1$) until the evaluation metric is smaller than the tolerance requirement

Step 8. (Decision) : If the evaluation metric exceeds the tolerance requirement for n^* samples, then n^* is the minimum sample size requirement in the layer.

4.4.1 Bootstrap resampling

Bootstrapping is commonly used to simulate empirical null-distribution of complex statistics such as percentile points, proportions, odds ratios, or correlation coefficients. It can also be utilized in constructing hypothesis tests as an alternative to inference based on parametric assumptions as well as to approximate distributions when the exact distributions are unknown, or the analytic procedures are too complex to obtain. In our case, the distribution of $\tilde{X}$ is difficult to estimate analytically. While we can derive the distribution of $\tilde{X}$ analytically, the approximation is inevitable. Therefore, we use bootstrap resampling to find the distribution of $\tilde{X}$. The distribution of $\tilde{X}$ from a second-order polynomial follows a normal distribution as shown in Appendix F. However, when the model is not restricted to the 2nd-order polynomial model, e.g., exponential model, the following bootstrap procedure will be used to find the sampling distribution of $\tilde{X}$ and the distribution of the test statistics in Objective 1.

The model fitting with the given sample, $\hat{y} = X\hat{\beta}$, yields the distribution of the estimated parameters, i.e., $\hat{\beta} \sim N(\beta, (X'X)^{-1}\hat{\sigma}_\epsilon^2)$. The bootstrap resamples $\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_m$ from the distribution of $\hat{\beta}$, $N(\beta, (X'X)^{-1}\sigma_\epsilon^2)$, when m is the bootstrap number. Then, $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_m$ are the estimated optima from models $X\hat{\beta}_1, X\hat{\beta}_2, \cdots$, $X\hat{\beta}_m$. However, instead of $\hat{\sigma}_\epsilon^2$, we will use the aggregated MSE from previous layers including the current layer. The proposed aggregated MSE is $\bar{\hat{\sigma}}_\epsilon^2 = 1/L \times \sum_{i=1}^{L} \hat{\sigma}_{\epsilon,i}^2$, where $\hat{\sigma}_{\epsilon,i}^2$ is the MSE in Layer i, $1 \leq i \leq L$ and L is the index for the current layer. Recall that the MSE estimates the variance of the model noise and measures the accuracy of the model prediction. By averaging previous MSEs, $\bar{\hat{\sigma}}_\epsilon^2$ contains all the information gathered from the previous models. For example, if any previous model

is inaccurate and is likely to lead to the wrong direction, then the variance of $\hat{\beta}$ is large, and $\tilde{X}$ varies greatly. However, if all previous models are accurate and guide to the right direction, then $\sigma^2_{\hat{\beta}}$ are small, and consequently $\tilde{X}$ has less variability. The following steps explain the bootstrap resampling method to estimate the distribution of $\tilde{X}$.

Step 1. Fit the model, $\hat{y} = X\hat{\beta}$

Step 2. Calculate $\bar{\hat{\sigma}}^2_\epsilon$ from previous layers and the current layer

Step 3. Estimate the distribution of $\hat{\beta}$,i.e., $\hat{\beta} \sim N(\beta, (X'X)^{-1}\bar{\hat{\sigma}}^2_\epsilon)$

Step 4. Collect B samples $(\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_m)$ from the estimated distribution

Step 5. Find optima $(\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_m)$ from each model, $X\hat{\beta}_1, X\hat{\beta}_2, \cdots, X\hat{\beta}_m$

Step 6. Estimate the distribution of $\tilde{X}$

4.5 Finding the distribution of a test statistic

When we encounter a new layer, we need to find a test statistic and its distribution. Let the estimated model in the layer be $\hat{f}_n(\mathbf{X}) = \mathbf{X}\hat{\boldsymbol{\beta}}_n$ when n is the sample size. Assume that the noise variance, σ^2_ϵ, is a constant. The coefficient estimate $\hat{\boldsymbol{\beta}}_n$ by linear regression follows a normal distribution, i.e., $\hat{\boldsymbol{\beta}}_n \sim MN\left(\mu_{\hat{\beta}_n}, \Sigma_{\hat{\beta}_n}\right)$.

The estimated optimum $\tilde{\mathbf{X}}_n$ is the optimal point on model $\hat{f}_n(\mathbf{X})$. Therefore, the distribution of $\tilde{\mathbf{X}}_n$ depends on the parameter estimate $\hat{\boldsymbol{\beta}}_n$. Hinkley [28] proposed that a vertex $\tilde{\mathbf{X}}_n$ has a normal distribution when $\hat{f}_n(\mathbf{X})$ is a second-order polynomial. If $\hat{f}_n(\mathbf{X})$ is not a second-order polynomial, however, the general form of the distribution of $\tilde{X}$ is unknown. Thus, we use the parametric bootstrap to find the distribution of $\tilde{\mathbf{X}}_n$ as explained in Section 4.4.1. Even though $\tilde{\mathbf{X}}_n$ comes from a general model other than second-order polynomial, the model can be approximated to a second-order polynomial near $\tilde{\mathbf{X}}_n$ as depicted in Figure 20 as an example.

Two plots in Figure 20 are plotted by the same model but with different explanatory regions. Left figure shows the high level view, and right figure shows the lower

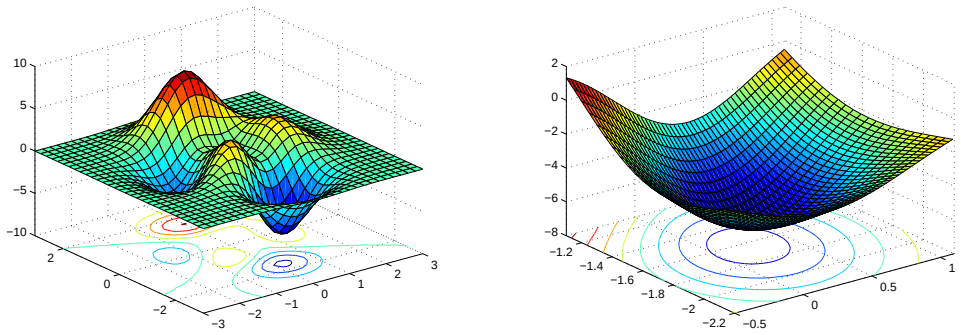

Figure 20: An example of showing (a) the overall view and (b) the close-up view.

level view. The explanatory region for the left figure includes the one for the right figure. Even though the overall model is complex, the close-up model becomes simple, and it can be estimated by a second-order model. Because both are the same models, the vertex in the left plot is equivalent to that in the right plot. As mentioned, the vertex in a second-order polynomial has a normal distribution. Consequently, we can approximate the vertex in any type of model to have a normal distribution (See Appendix F for analytical proof of normality).

By increasing the number of bootstraps, the mean and the variance of $\tilde{\mathbf{X}}_n$ converge after exceeding a certain amount of bootstrap number (m). Completing the bootstrapping, we obtain $\tilde{\mathbf{X}}_{n,1}, \tilde{\mathbf{X}}_{n,2}, \cdots, \tilde{\mathbf{X}}_{n,m}$. The estimator of the effect size $D_n = |\tilde{\mathbf{X}}_n - \boldsymbol{\mu}|$ is $\hat{D}_n = max|\tilde{\mathbf{X}}_{n,i} - \tilde{\mathbf{X}}_{n,j}|, \quad i,j \in [1,m]$. In one dimension, $\hat{D}_n$ equals $\hat{D}_n = max(\tilde{X}_{n,i}) - min(\tilde{X}_{n,i}), \quad 1 \leq i \leq m$. Then $\hat{D}_n$ represents the distance between the two farthest apart $\tilde{X}_{n,i}$'s. Because $\hat{D}_n$ represents the range of order statistics, the probability density function of $\hat{D}_n$ is

$$f_{\hat{D}_n}(d) = \int_{\infty}^{-\infty} m(m-1) f_{\tilde{X}_n}(x+d) f_{\tilde{X}_n}(x) [F_{\tilde{X}_n}(x+d) - F_{\tilde{X}_n}(x)]^{m-2} dx, \qquad (98)$$

where $f_{\tilde{X}_n}(x)$ is the probability density function of $\tilde{X}_n$, and $F_{\tilde{X}_n}(x)$ is the cumulative density function of $\tilde{X}_n$. Equation (98) enables us to obtain the distribution of $\hat{D}$.

The procedure for estimating the variance of $\hat{D}$, $\hat{\sigma}^2_{\hat{D}}$,is as follows. Let $min(\tilde{X}) = U$

and $max(\tilde{X}) = V$. Then, the variance of $\hat{D}$ is

$$\begin{aligned}\sigma^2_{\hat{D}} = Var(\hat{D}) &= Var(V - U) \\ &= Var(V) + Var(U) - 2 \cdot Cov(U, V).\end{aligned}$$

The marginal distributions of U and V are $f_U(u) = mf_{\tilde{X}}(u)[1 - F_{\tilde{X}}(u)]^{m-2}$ and $f_V(v) = mf_{\tilde{X}}(v)[F_{\tilde{X}}(v)]^{m-2}$, respectively. The joint distribution of U and V is $f_{U,V}(u, v) = m(m-1)f_{\tilde{X}}(u)f_{\tilde{X}}(v)[F_{\tilde{X}}(v) - F_{\tilde{X}}(u)]^{m-2}$. Knowing the distribution of $\tilde{X}$ allows us to estimate $\hat{\sigma}^2_{\hat{D}}$. Distribution $f_{\tilde{X}}(x)$ depends on $\mu_{\hat{\beta}}$ and $\Sigma_{\hat{\beta}}$. Therefore, $\hat{\sigma}^2_{\hat{D}}$ depends on $\mu_{\hat{\beta}}$ and $\Sigma_{\hat{\beta}}$ as well. The variance of $\hat{\beta}$, $\Sigma_{\hat{\beta}}$, is a function of $\hat{\sigma}^2_{\epsilon}(= MSE)$, and MSE is a random variable. Consequently, $\hat{\sigma}^2_{\hat{D}}$ is a random variable, and we can estimate its distribution. In the same manner, the estimator of the variance of $\tilde{X}$, $\hat{\sigma}^2_{\tilde{X}}$, is also a random variable.

Proposition 1. The variance of $\hat{D}$ has the same distribution as the variance of $\tilde{X}$.

Proof Let the mean and the variance of $\tilde{X}$ be $\mu_{\tilde{X}}$ and $\sigma^2_{\tilde{X}}$, i.e., $\tilde{X} \sim N(\mu_{\tilde{X}}, \sigma^2_{\tilde{X}})$. From Equation (98), the dummy variable x is changed to $k = (x - \mu_{\tilde{X}})/\sigma_{\tilde{X}}$. Then, Equation (98) can be re-expressed as

$$\begin{aligned}f_{\hat{D_n}}(d) &= \int_{\infty}^{-\infty} m(m-1)f_{\tilde{X_n}}(\sigma_{\tilde{X}}k + \mu_{\tilde{X}} + d)f_{\tilde{X_n}}(\sigma_{\tilde{X}}k + \mu_{\tilde{X}}) \\ &\quad \times [F_{\tilde{X_n}}(\sigma_{\tilde{X}}k + \mu_{\tilde{X}} + d) - F_{\tilde{X_n}}(\sigma_{\tilde{X}}k + \mu_{\tilde{X}})]^{m-2} dk \cdot \sigma_{\tilde{X}}. \qquad (99)\end{aligned}$$

For example, $f_{\tilde{X_n}}(\sigma_{\tilde{X}}k+\mu_{\tilde{X}}+d)$ is $1/\sqrt{2\pi\sigma^2_{\tilde{X}}} \cdot exp\ [-(\sigma_{\tilde{X}}k+\mu_{\tilde{X}}+d-\mu_{\tilde{X}})^2/2\sigma^2_{\tilde{X}}] = 1/\sqrt{2\pi\sigma^2_{\tilde{X}}} \cdot exp\ [-(k+d/\sigma_{\tilde{X}})^2/2] = 1/\sigma_{\tilde{X}} \cdot \phi(k+d/\sigma_{\tilde{X}})$, where $\phi(\cdot)$ is the probability density function of a standard normal distribution. Similarly, $f_{\tilde{X_n}}(\sigma_{\tilde{X}}k+\mu_{\tilde{X}})$ becomes $1/\sigma_{\tilde{X}} \cdot \phi(k)$. Then, Equation (99) is reexpressed as

$$\begin{aligned}f_{\hat{D_n}}(d) &= \int_{-\infty}^{\infty} \frac{m(m-1)}{\sigma_{\tilde{X}}}\ \phi(k + \frac{d}{\sigma_{\tilde{X}}})\ \phi(k)\ [\Phi(k + \frac{d}{\sigma_{\tilde{X}}}) - \Phi(k)]^{m-2} dk \\ &= \frac{1}{\sigma_{\tilde{X}}}\ f_S(\frac{d}{\sigma_{\tilde{X}}}),\end{aligned}$$

where $f_S(s)$ is the probability density function of $S = Z_{(m)} - Z_{(1)}$, when Z_i is a standard normal random variable, and $Z_{(i)}$ represents the i-th order statistic. Therefore, we can easily obtain that $E[\hat{D}] = \sigma_{\tilde{X}} \cdot E[S]$ and $Var[\hat{D}] = \sigma^2_{\tilde{X}} \cdot Var[S]$. Because S is a '$\mu_{\tilde{X}}$ and $\sigma'_{\tilde{X}}$- free random variable, $Var[\hat{D}]$ has a linear relationship to $Var[\tilde{X}]$. Thus, we can conclude that the variance of $\hat{D}$ has the same distribution as the variance of $\tilde{X}$. □

Using the algorithm of estimating the covariance matrix introduced by Davis [18], we determine $Var[S]$, and $\hat{\sigma}^2_{\hat{D}}$ is $\hat{\sigma}^2_{\tilde{X}} \cdot Var[S]$. The variance estimate of $\tilde{X}$ is the sample variance of $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_m$. Because $\tilde{X}_i, 1 \leq i \leq m$ follows a normal distribution, their sample variance has a chi-squared distribution. As a result, $\hat{\sigma}^2_{\hat{D}}$ also has a chi-squared distribution. From Equation (96), we estimate the test statistic with $\hat{D}$ and $\hat{\sigma}^2_{\hat{D}}$ as well as their distributions.

4.6 Illustrative example

This section illustrates simulation investigations of the sample size calculation using the proposed methodology, particularly of the model assessment. Denoted Steps of the following match the steps in Algorithm 1 in Section 4.4. **[Step 0]** Let f be a true model defined by

$$f(x) = -7 \cdot 10^{-5} x(x-5)(x-7)(x-10)(x-1) + 2 \cdot 10^{-2}(x-5)^3 + 10^{-8} \cdot exp(x)) + \varepsilon,$$

where noise $\varepsilon \sim N(0, 5^2)$. The explanatory domain is $x \in [-18, 26]$, and the tolerance requirement is $\delta = 1$. The true parametric model is a fifth-order polynomial with an exponential term, thus simple parametric regression modeling can never capture the true model. Figure 21 depicts the true function with different x domains.

The left plot shows the overall trend, and the right plot shows the closer view with a smaller design space ($x \in [-12, 8]$). Figure 21 proves that the upper(wider) layer needs a more complex model structure than the lower(narrower) layer does. **[Step 1]** Assume that the underlying model for the estimation is a fourth-order polynomial

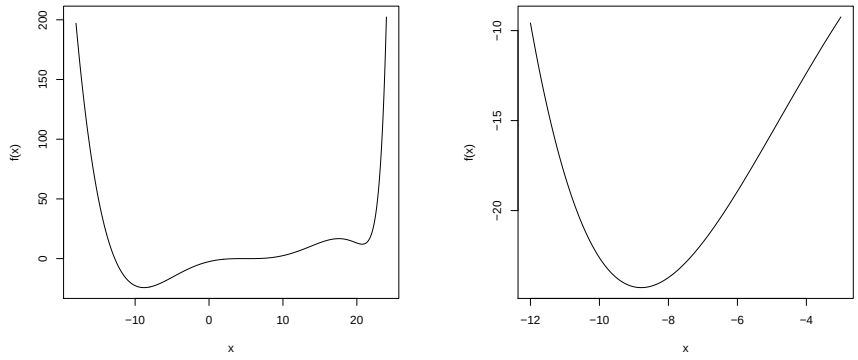

Figure 21: True function over (a) the overall range and (b) $x \in [-12, 8]$

at the global layer. [**Step 2**] Then, at least 6 samples are necessary to estimate a fourth-order polynomial model and a mean squared error.

[**Step 3, 4**] When six samples are collected through the use of minimum energy design (space-filling design), the estimated model with fourth-order polynomial regression is $\hat{y} = -5.48 + 1.12x - 0.0002x^2 - 0.0047x^3 + 0.0003x^4$. Note that the minimum energy design is used for the entire sample collection for consistency. The distribution of the estimated coefficient $\hat{\beta}$ is well known as a multivariate normal distribution, that is,

$$\hat{\beta} \sim MN\left(\mu_{\hat{\beta}}, \Sigma_{\hat{\beta}}\right),$$

where

$$\mu_{\hat{\beta}} = \begin{bmatrix} -5.48 \\ 1.12 \\ -0.0002 \\ -0.0047 \\ 0.0003 \end{bmatrix}$$

$$\Sigma_{\hat{\beta}} = \begin{bmatrix} 0.697 & -0.0056 & -6.59 \cdot 10^{-3} & 9.12 \cdot 10^{-5} & 7.18 \cdot 10^{-6} \\ -5.68 \cdot 10^{-3} & 0.048 & -1.45 \cdot 10^{-3} & -3.73 \cdot 10^{-4} & 1.55 \cdot 10^{-5} \\ -6.59 \cdot 10^{-3} & -0.0015 & 1.58 \cdot 10^{-4} & 1.14 \cdot 10^{-5} & -6.91 \cdot 10^{-7} \\ 9.12 \cdot 10^{-5} & -0.00037 & 1.14 \cdot 10^{-5} & 3.1 \cdot 10^{-6} & -1.31 \cdot 10^{-7} \\ 7.18 \cdot 10^{-6} & 1.5 \cdot 10^{-5} & -6.91 \cdot 10^{-7} & -1.31 \cdot 10^{-7} & 5.95 \cdot 10^{-9} \end{bmatrix}.$$

[Step 5] In this example, the bootstrap number is $m = 100$. Bootstrap resampling enables us to obtain the distribution of $\tilde{X}$ and the estimation of D. **[Step 6]** Based on the bootstrap resampling result, the estimation of D is $\hat{D} = max(\tilde{X}) - min(\tilde{X}) = 22.48$. The histogram and Q-Q plot in Figure 22 show that the distribution of $\tilde{X}$

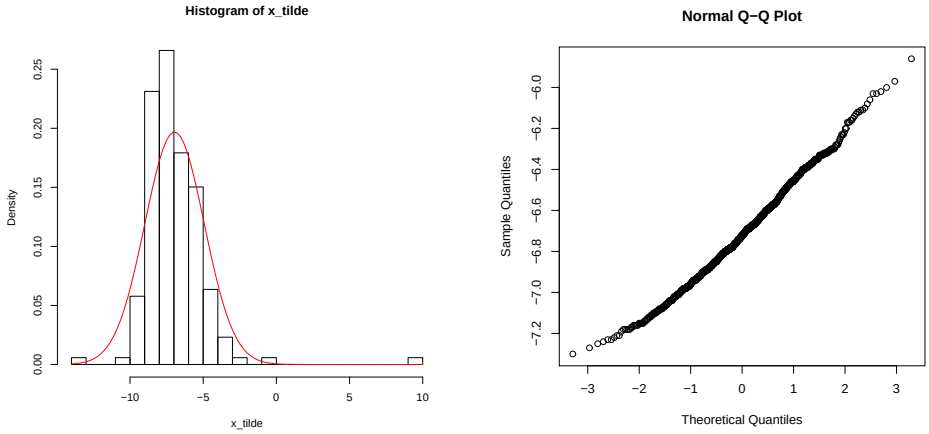

Figure 22: (a) Comparison of the histogram of $\tilde{X}$ and a normal distribution, $N(\mu_{\tilde{X}}, \sigma^2_{\tilde{X}})$; (b) Q-Q plot for the normality check

is approximately a normal distribution. The mean and the standard deviation of $\tilde{X}$ are $\hat{\mu}_{\tilde{X}} = -6.957$ and $\hat{\sigma}_{\tilde{X}} = 2.027$, respectively. Pearson Chi-Square normality test proves that $\tilde{X}$ follows a normal distribution with the value of the Pearson chi-square statistic = 273.55 and p-value = 2.2×10^{-16}.

[Step 7] Because $\hat{D}$ is the range of order statistic, $X_1, \cdots, X_m$, the distribution of the range, $\hat{D}$, is known as Equation (100).

$$f_{\hat{D}}(d) = \int_{-\infty}^{\infty} m(m-1) f_{\tilde{X}}(v+d) f_{\tilde{X}}(v) [F_{\tilde{X}}(v+d) - F_{\tilde{X}}(v)]^{m-2} dv, \qquad (100)$$

where m is the bootstrap number, $f_{\tilde{X}}(x)$ is the probability density function of $\tilde{X}$, and $F_{\tilde{X}}(x)$ is the cumulative density function of $\tilde{X}$. Because the probability and the cumulative density functions of $\tilde{X}$ are derived in Step 6, the mean and the standard deviation of $\hat{D}$ are calculated by Equation (100), i.e., $\hat{\mu}_{\hat{D}} = 10.16531$ and $\hat{\sigma}_{\hat{D}} = 1.226641$, respectively.

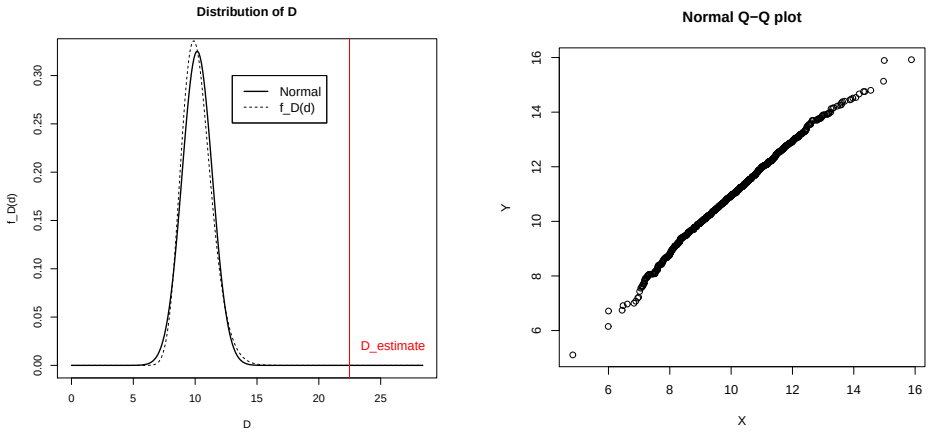

Figure 23: (a) Comparison of $f_D(d)$ and a normal distribution, $N(\mu_D, \sigma_D^2)$; (b) Q-Q plot for the normality check

Figure 23(a) shows the original probability density function of $\hat{D}$ and its approximation. The probability density function of $\hat{D}$ is derived by Equation (100). Normal approximation is used for the distribution of $\hat{D}$ with the mean $\mu_{\hat{D}}$ and the variance $\sigma_{\hat{D}}^2$. That is, the approximated function (solid line in Figure 23(a)) is $\hat{D} \sim N(10.16, 1.227^2)$. The right panel, Figure 23(b), shows the Q-Q plot for the normality check. Additionally, Pearson Chi-Square normality test proves that $\tilde{D}$ follows a normal distribution with the value of a Pearson chi-square statistic = 351.04 and a p-value = 4.6×10^{-12}. The red vertical line in Figure 23(a) is $\hat{D}$, i.e., $x = \hat{D}$. By the definition of $\hat{D}$, $\hat{D}$ is $\hat{D} = 9.15 - (-13.33) = 22.48$ as seen in Figure 22(a). Because the sample size (6) is not enough to generate the stable $\tilde{X}$s, $\hat{D}$ is far from $\mu_{\hat{D}}$.

[Step 8] Equation (96) defines the test statistic. The domain of the global layer

is $x \in [-18, 26]$. Thus, the diameter (width in 1-dimension) of the global layer is $26-(-18) = 44$. The width of the next layer halves by the assumption in Section 4.3, i.e., the diameter of the following layer is 22. A radius of a layer defines the layer size. Thus, the layer size for the following layer, r_1, is 11 in our example. The test statistic by Equation (96) is $T_6 = (\hat{D}-2r)/\hat{\sigma}_{\hat{D}} = (22.48-2\cdot 11)/1.227 = 0.3913$. Recall that $\hat{\sigma}_{\hat{D}}$ is a random variable due to the randomness from the noise variance estimate, $\hat{\sigma}_{\varepsilon}$. We know that the MSE estimates $\hat{\sigma}^2_{\varepsilon}$ and follows a chi-squared distribution. Therefore, it is assumed that $\hat{\sigma}^2_{\tilde{X}}$, follows a chi-squared distribution with the degree of freedom $df = n - p - 1$, where n is the sample size and p is the order of the polynomial. Figure 24 shows the distribution of $\hat{\sigma}^2_{\tilde{X}}$ and compares it to a $\chi^2_{n-p-1} = \chi^2_1$, $n = 6$ and $p = 4$. Kolmogorov-Smirnov test proves that $\hat{\sigma}^2_{\tilde{X}} \sim \chi^2_{n-p-1}$ with p-value= 0.0541 and

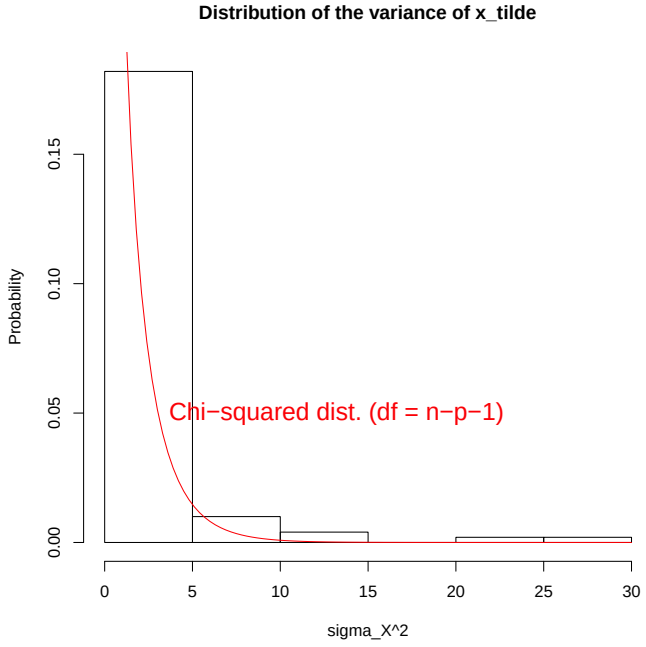

Figure 24: Comparison of χ^2_1 and the distribution of $\hat{\sigma}^2_{\tilde{X}}$

the test-statistic value=0.19. In conclusion, distribution of the test statistic is ratio of a normal distribution over the square root of a chi-squared distribution. Figure 25 shows the distribution of the test statistic with the critical value for $\alpha = 0.1$ and the

test statistic.

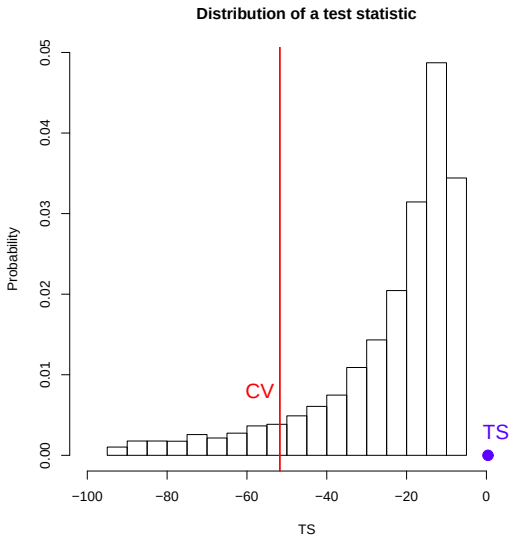

Figure 25: The distribution of a test statistic

[**Step 9**] The hypothesis test is rejected when $T_6 < Critical\ Value$. The critical value can be derived from the distribution of the Test Statistic with a given α. Since the observed value of the test statistic is not in the rejection region, the hypothesis test cannot be rejected when $n = 6$ with $\alpha = 0.1$. That is, the model estimated by 6 samples derives the optimum estimate which will lead into the incorrect direction. Table 5 summarizes the results of the hypothesis test for the different layers and sample sizes. Recall that the hypothesis test is rejected when the test statistic is smaller than the critical value. Both the test statistic and the critical value depend on the sample size. Decision in Table 5 is based on the hypothesis test introduced in Section 4.3. [**Step 10**] From Layer 1, the observed test statistic T_n becomes smaller than the critical value from $n = 8$ based on Table 5. Thus, we collect $n = 8$ samples and estimate a model for Layer 1. The mean of $\tilde{X}$, $\mu_{\tilde{X}}$, derived by the model determines the location of Layer 2, i.e., Layer 2 $= \{x : x \in [\mu_{\tilde{X}} - r_1, \mu_{\tilde{X}} + r_1] = [-19.3914, 2.6086]\}$. Because Layer 1 is $[-18, 26]$, the range of Layer 2 exceeds the global layer. Therefore, Layer 2 will be truncated to $[-18, 2.6086]$. The same

procedure (Step 2 to 9) is performed for the sample size calculation in Layer 2. Figure

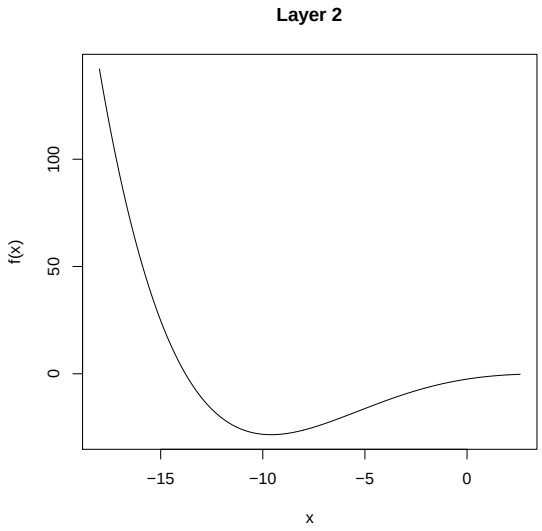

Figure 26: True function over Layer 2 [-18, 2.6]

26 depicts the true function f in Layer 2: [-18, 2.6086]. With a 3rd order polynomial, we can estimate the process optimum, $\tilde{X}$. The distribution of $\tilde{X}$ and Q-Q plot is in Figure 27.

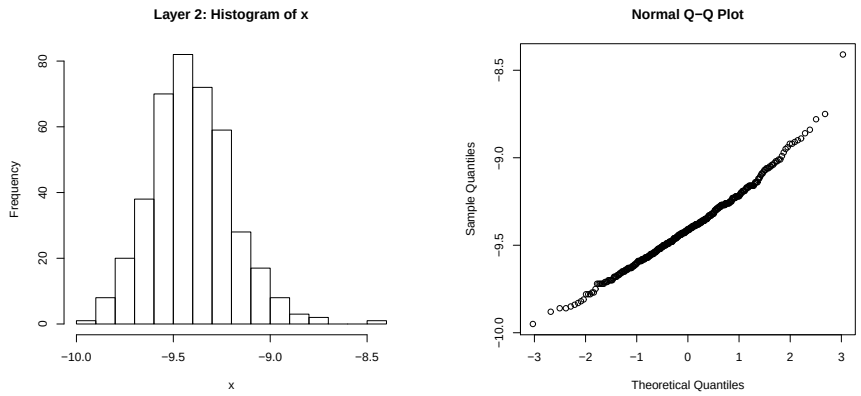

Figure 27: (a) Comparison of the histogram of $\tilde{X}$; (b) Q-Q plot for the normality check

Compared to the model in Layer 1, the model has more confidence, thus the variance of $\tilde{X}$ is smaller in Layer 2. From the bootstrap resampling, the effect size estimate is $\hat{D} = 1.54$, and the mean and the variance of $\tilde{X}$ are $\hat{\mu}_{\hat{D}} = 1.045, \hat{\sigma}_{\hat{D}} = 0.126$, respectively.

In Figure fg:layer2-x, Q-Q plot shows that $\tilde{X}$ follows a normal distribution. The distributions of $\hat{D}$ and $\hat{\sigma}_{\hat{D}}$ are depicted in Figure 28.

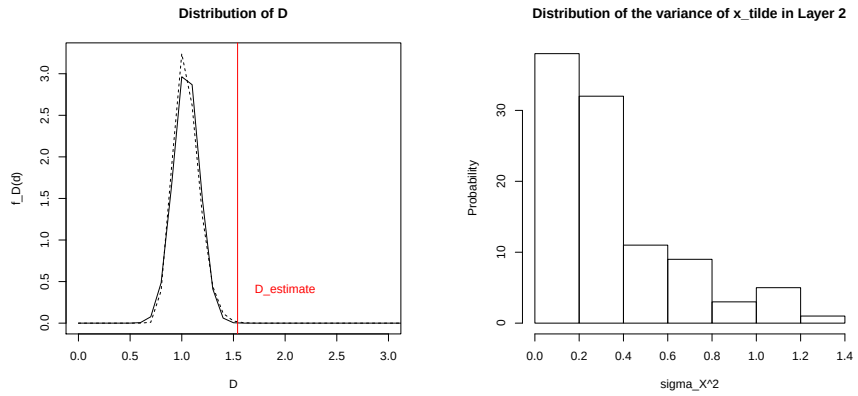

Figure 28: (a) Comparison of $f_D(d)$ (dash line) and a normal distribution, $N(\mu_D, \sigma_D^2)$ (solid line); (b) distribution of $\hat{\sigma}_{\hat{D}}$

As we zoom the region, the model view becomes simpler as explained in Section 4.2. Thus, the underlying model structure for Layer 2 has smaller order than the model for Layer 1. We used a fourth-order polynomial in Layer 1. Thus, a third-order polynomial model is used for Layer 2. The evaluation metric decides the last layer. The evaluation metric decision will be explained in Section 4.7.2. Based on the Table 5, 5 samples are collected from Layer 2.

Table 5: The critical value and the test statistic for different sample sizes

Layer	Sample size	Critival Value	Test Statistic	Decision
Layer 1	6	-51.24156	0.3913	Do not Reject
	7	-39.77246	-33.21269	Do not Reject
	8	-35.95188	-70.57089	Reject
Layer 2	5	-48.43064	-153.38967	Reject
	6	5.770265	4.251640	Reject
	7	1.8722264	0.9471903	Reject

4.7 Methodology properties

4.7.1 Type I and Type II errors

Type I errors (α) and Type II errors (β) describe the particular flaws in a testing process where a true null hypothesis is incorrectly rejected (Type I error) or a false null hypothesis test is not rejected (Type II error). In particular, the power $(1-\beta)$ measures the ability to reject the null hypothesis when it is actually false, i.e., power $= P$ (Reject $H_0 \mid H_0$ is false). That is, Type I and II errors can demonstrate the effectiveness of the proposed methodology.

For the simplicity, let $S = d(\mu, \tilde{X}) - 2r = D - 2r$, and the estimator of S, $\hat{S}$ be $\hat{S} = \hat{D} - 2r$. Then, the null and the alternative hypotheses become $H_0 : S = 0$ versus $H_1 : S < 0$. Since the variances of $\hat{S}$ and $\hat{D}$ are the same, the test statistic is $T.S. = \hat{S}/\hat{\sigma}_{\hat{S}}$. When the test statistic follows a t-distribution, the Type I error for the hypothesis test is as follows:

$$
\begin{aligned}
Type\ I\ error &= P\ (\ Reject\ H_0 \mid H_0\ true) \\
&= P\ (\ T.S. < -t_{\alpha,n-p-1} \mid d(\mu, \tilde{X}) = 2r) \\
&= P\ (\ \frac{\hat{S}}{\hat{\sigma}_{\hat{S}}} < -t_{\alpha,n-p-1} \mid S = 0) \\
&= P\ (\frac{\hat{S} - 0}{\hat{\sigma}_{\hat{S}}} < -t_{\alpha,n-p-1}) \\
&= P\ (Z < -t_{\alpha,n-p-1}) \\
&= \Phi(-t_{\alpha,n-p-1}).
\end{aligned}
$$

If the degree of freedom, $n-p-1$, is large enough, the Type I error equals α, a level of significance. However, in our example, the test statistic does not follow a student t-distribution. Thus, the critical value, $-t_{\alpha,n-p-1}$, will be replaced by the value found in the empirical distribution. In addition, the empirical cumulative density distribution function will substitute $\Phi(\cdot)$ in the above formula.

A similar procedure generates the Type II error as follows:

$$\begin{aligned}
Type\ II\ error &= P\ (\ Do\ not\ reject\ H_0 \mid H_1 true) \\
&= P\ (\ T.S. > -t_{\alpha,n-p-1} \mid d(\mu, \tilde{X}) \leq 2r) \\
&= P\ (\ \frac{\hat{S}}{\hat{\sigma}_{\hat{S}}} > -t_{\alpha,n-p-1} \mid S \leq 0) \\
Type\ II\ error(\tau) &= P\ (\ \frac{\hat{S}}{\hat{\sigma}_{\hat{S}}} > -t_{\alpha,n-p-1} \mid S = \tau) \\
&= P\ (\ \frac{\hat{S} - \tau + \tau}{\hat{\sigma}_{\hat{S}}} > -t_{\alpha,n-p-1} \mid S = \tau) \\
&= P\ (\ \frac{\hat{S} - \tau}{\hat{\sigma}_{\hat{S}}} > -t_{\alpha,n-p-1} - \frac{\tau}{\hat{\sigma}_{\hat{S}}} \mid S = \tau) \\
&\approx 1 - \Phi(-t_{\alpha,n-p-1} - \tau/\hat{\sigma}_{\hat{S}}).
\end{aligned}$$

As the power is $1 - \beta$, the power of the hypothesis test is $\Phi(-t_{\alpha,n-p-1} - \tau/\hat{\sigma}_{\hat{S}})$. According to the formula, the power increases with the values of the parameter τ. For a specific value of τ, we obtain a higher power by increasing the sample size n, because $-t_{\alpha,n-p-1}$ increases as n increases. Because τ is negative, decreasing τ increases the power. Thus, the maximum of the power with a specific sample size n is not estimable, yet we can estimate the infimum value of the power when τ is nearly zero. Table 6 summarizes the power of the hypothesis test in our example.

Table 6: Power for different layers and sample sizes

Sample Size	5	6	7	8
Layer 1	NA	0.131	0.359	0.949
Layer 2	0.970	1.000	1.000	1.000

Recall that eight and five samples are required for the first and second layers, respectively. The corresponding power is 0.949 and 0.970.

4.7.2 Evaluation metric comparison

The evaluation metric for a layer R is introduced in Equation (97). The metric $L(R)$ depends on the layer R and the sample size n. For different layers and different sample sizes, $L(R)$ is calculated and is depicted in Figure 29.

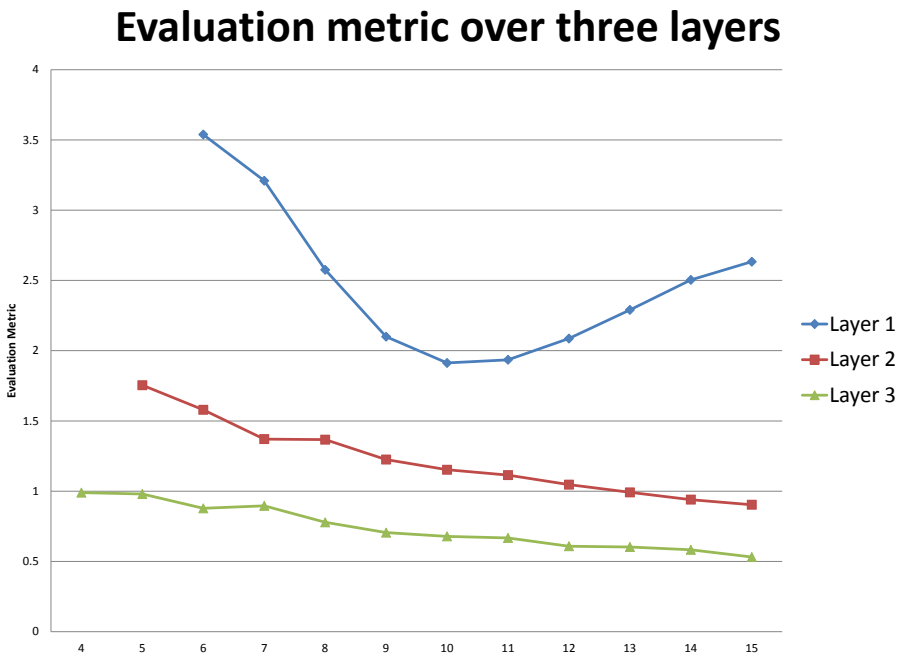

Figure 29: Evaluation metric changes over different sample sizes from three layers

In the global layer, because of the model mismatch, it cannot meet the tolerance requirement even though we increase the sample size. The possible minimum tolerance, $L(R_1)$, is 1.91 with ten samples in Layer 1. When the tolerance requirement is 1 ($\delta = 1$), any models in Layer 1 can never meet the tolerance requirement. Therefore, there is no option to stop at the Layer 1 to satisfy the tolerance requirement objective.

Based on the graph in Figure 29, the metric in Layer 2, $L(R_2)$, is smaller than the requirement when the sample size exceeds $n = 13$. Similarly, the metric in Layer 3 drops below 1 from $n = 4$. Thus, we can stop the multi-layer procedure at the second layer with eight samples from Layer 1 and thirteen samples from Layer 2. Another possible option is that we can lower the multi-layer system until the Layer

3. In that case, eight samples from Layer 1, five samples from Layer 2 (both are based on the model assessment by Table 5), and four samples from Layer 3. The final decision depends on the experimental budget. If the budget can cover 21 experiments (=13+8), then the multi-layer system can stop at the Layer 2. Otherwise, the multi-layer system is performed until the third layer, and 17 (=8+5+4) experiments can estimate the model which can meet the tolerance requirement.

4.7.3 Mean squared error comparison

One of the several objectives in the proposed methodology is to estimate the process as accurate as possible. After estimating the model, we need to evaluate the correctness of the model. MSE is widely used to estimate model error and quantify the model accuracy. Because we are proposing a multi-layer system, we need to compare the model accuracy for each layer. We obtain the MSE by the formula, $MSE = \sum(y_i - \hat{y}_i)^2/(n-p-1)$. However, for the MSE for each layer, we standardize the calculated MSE by dividing the size of the region. That is, the MSE for Layer i is

$$MSE_i = \frac{\sum_{i=1}^{m}(y_k - \hat{y}_k)^2/(m-p-1)}{size(L_i)}, \tag{101}$$

where m is an arbitrary number, and p is the model order in the layer. For this example, we choose $m = 30$. Table 7, which summarizes the results, clearly shows that MSE decreases as we go down the layers, which means that the model becomes more reliable.

4.7.4 With or without multi-layer?

Employing the proposed multi-layer system raises a question, 'what happens when the multi-layer system is not used, so all samples are collected at the global layer?'. Figure 30 shows the results.

Table 7: Mean Squared Error for different layers

	Layer	(8,5,4)
After standardizing	Layer 1	472.2588
	Layer 2	6.245107
	Layer 3	3.147543
Before standardizing	Layer 1	20779.39
	Layer 2	128.7029
	Layer 3	34.62298

Figure 30 depicts three functions in each figure. The solid line is the true function. The black wide-dotted line is the estimated function using the proposed methodology. The sample size for each layer is $(n_1, n_2, n_3) = (8, 5, 4)$. The small-dotted function is the estimated function using 17 samples at the global layer. The total sample usages are the same, while the model estimation significantly improves by using the proposed methodology. Especially, the Layer 3 which is the closest neighborhood of the true optimum and the most important region shows the biggest difference. The model without using the multi-layer system is far from the true model while the model using the system is close to the true model. Therefore, we can conclude that the multi-layer system works effectively for the purpose of the accurate model estimation, especially around the optimum.

4.8 Discussion

This chapter has proposed a sample size calculation based on a multi-layer system. Determining a sample size is an important issue because samples that are too large consume computational time and resources, and samples that are too small can lead to incorrect results. Our proposed method minimized the sample size for the estimation of the process optimum associated to the multi-layer system, and our sample size methodology satisfied the tolerance requirement. The simulation results demonstrated that the proposed methodology successfully estimated the model details and

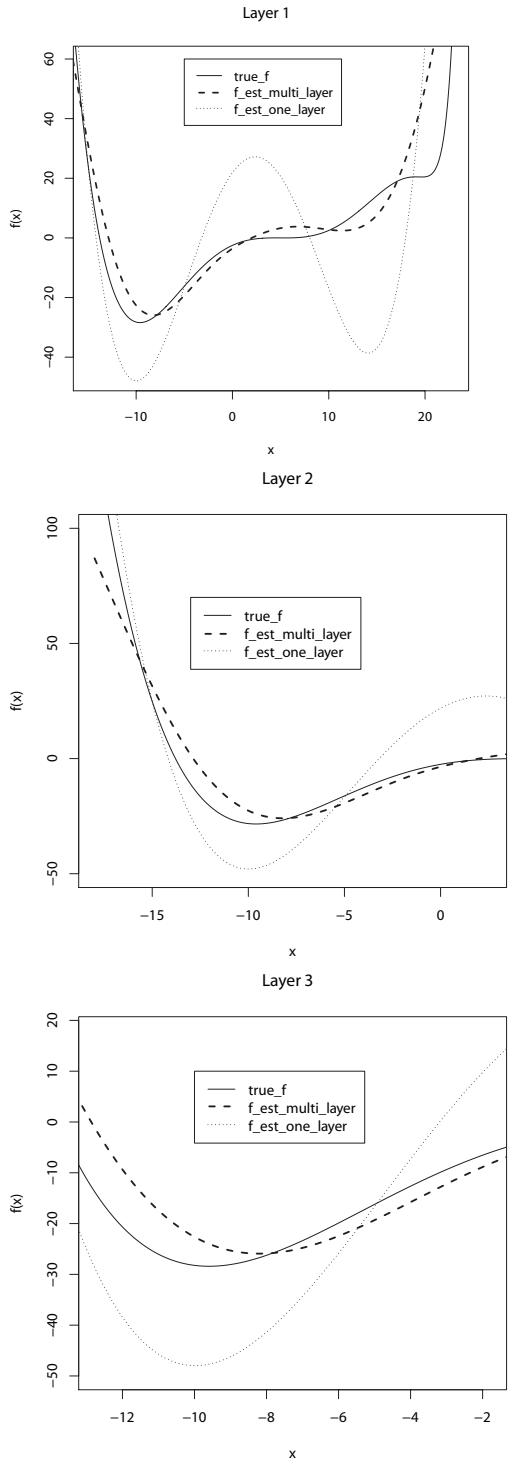

Figure 30: Comparison of three plots in a global layer: the true function f (solid), the estimate function $\hat{f}$ with using the multi-layer (wide-dotted line), and the estimated function $\hat{f}$ without using the multi-layer (small-dotted line) over (a) Layer 1: $x \in [-18, 26]$ (b) Layer 2: $x \in [-18, 2.6086]$ (c) Layer 3: $x \in [-12.7751, -1.7751]$

the optimum satisfying the tolerance requirement.

Several methods determine the sample size. However, the proposed methodology enables to satisfy several objectives at the same time. The first and the most important objective is to estimate an input setting of having an optimum. By estimating the optimum at every layer, we do not miss the optimum and keep including it. In addition, the model estimation is precise with taking the multi-layer system, especially around the optimum region. When the original model has very delicate details, it is easy to miss the optimum even though we are almost close. Therefore, the tight tolerance around the optimum is necessary to get more robust results.

We used a space-filling design to collect samples with a fixed layer size. The space-filling design is useful when model information is unknown. However, as the layers go deeper, model information come from the previous layer as prior knowledge. Therefore we suggest that the optimal design is more suitable than the space-filling design after passing a certain number of layers.

The multi-layer system is an unexplored and very fruitful area of research. We believe that there is a tremendous reservoir of applications where this methodology can be utilized, particularly for sensitive and complex systems.

APPENDIX A

ADDITIONAL NOISE FACTORS

1. **Lack of Knowledge**

   To calculate the solubility constraint which is described in Chapter 3, we use the equation of state [65] for supercritical CO_2 and calculate the density of CO_2, ρ_{CO_2}. The program code has been developed by Weon Shik Han [43] and modified by Andres Hernandez. We highly depend on the code and assume that it is right, but obviously it is possible that some parts are missing.

2. **Particle Size**

   Even though the CNT is ordered from same company, they have different particle size. Some are big and some are fine powder. So mortar and pestle are needed to grind large particle to powder and this may affect the result.

   Recall that α is the concentration of active sites in 19 and 20.

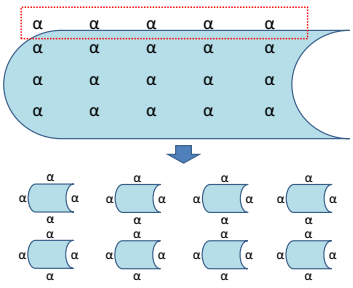

Figure 31: Impact of carbon nanotube particle Size

Figure 31 explains how CNT particle size affects the result. α's are distributed on the CNT. However, when a particle is large enough, only α on the surface is

usable; red square in 31. Making smaller CNT particle increases CNT surface, and it implies increasing of usable α. One more thing we have to consider is, even though 31 shows α's are uniformly distributed on the surface, but in reality, they are not. Thus, number of α differs by choice of CNT surface.

3. **Reactor**

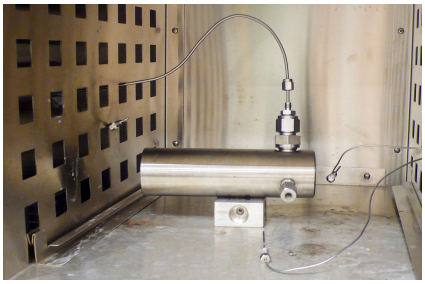

Figure 32: Experimental equipment: vessel in the oven

Figure 32 shows the real experimental image. The cylinder in the middle is the vessel, and all chemicals are reacting in the vessel. During an experiment, the vessel is heated up in the oven, since Erkey uses thermal reduction. As you see in Figure 32, the vessel is putted in horizontal. It is known that vessel position and shape can affect the result, e.g., vertical vessel, or square pillar.

4. **Differential Equation**

   In the program (Matlab) code, *ODE23s* and *ODE45* is used to solve differential equations. If we change the program, not Matlab, or change the solving command - since there are several function for solving differential equation - results can be changed. However, at this moment, we will follows Matlab built-in *ODE* functions, and assume the functions gives a correct answer.

5. Pump fluctuation Hydrogen pressure controller (manually adjusted)

6. Hydrogen pressure measurement ($\pm$ 10 psi)

7. Leaking of gas (affect the equilibrium)

8. Fluid inside reactor is pushed back to pump during the reactors heating up process (raise temperature to T_2)
9. Hydrogen and CO_2 mixture is pushed back to the pump during the mixing
10. Air inside reactor when running experiment (small amount relative to CO_2)
11. Thermal reduction
12. Exposure of CNT to high temperature after reaction
13. Variation among substrates
14. Thermal degradation might lead to weight loss of CNT (at 200C, $< 5\%$)
15. Inconsistent depressurization (done manually-too fast or too slow)

APPENDIX B

ATOMS TO NM CONVERSION

- Density of Platinum : 21.45 g/cm^3

- Platinum Molecular Weight : 195.08 g/mol

- Avo- gadro's number : 6.02 $\times 10^{23}$ $atoms/mol$

$$\frac{21.45g}{1cm^3} \cdot \frac{1mol}{195.08g} \cdot \frac{6.02 \times 10^{23} atoms}{1mol} \cdot (\frac{100cm}{10^9 nm})^3$$
$$= \frac{66.192\ atoms}{1\ nm^3}$$

Therefore, 1 $nm^3 = 66.102\ atoms$, and use formula for volume of a sphere, $\frac{4}{3}\pi r^3$,

$$\frac{4}{3}\pi(\frac{x\ nm}{2})^3 = \frac{y\ atoms}{66.192}$$
$$x\ nm = 2 \times (\frac{y\ atoms}{66.192 \cdot 4/3 \cdot \pi})^{1/3}$$

APPENDIX C

HOW TO CALCULATE THE IMPURITY

The maximum of impurity is 3.22 %.

- Molecular weight of the $Pt(acac)_2$: 393.31 [g/mol]
- Platinum Molecular Weight: 195.08 [g/mol]

If precursor contains no impurity, the amount of platinum should be 49.6 % (= 195.08 / 393.31) theoretically. However, the specifications say 48 % minimum.

$$48\% \quad \leq \quad \frac{195.08}{393.31 \; + \; impurity} \quad \leq \quad 49.6\%$$

$$0 \quad \leq \quad impurity \quad \leq \quad 13.1 \quad \text{[g/mol]}$$

$$0\% \quad \leq \quad \frac{impurity}{393.31 \; + \; impurity} \quad \leq \quad 3.22\%$$

APPENDIX D

HOW TO CALCULATE THE LOAD

- Density of Platinum : 21.45 g/cm^3

- Platinum Molecular Weight : 195.08 g/mol

- Avo- gadro's number : 6.02 $\times 10^{23}$ $atoms/mol$

- Volume : 54 cm^3

$$
\begin{aligned}
& mw_{P_t} \cdot \sum_{i=1}^{n} iC_i \\
= \; & 195.084\ [\frac{g}{mol}] \cdot \sum i\ C_i\ [\frac{mol}{cm^3}] \\
= \; & 195.084\ [\frac{g}{mol}] \cdot \sum i\ C_i\ [\frac{mol}{cm^3}] \cdot 54\ [cm^3] \cdot 1000\ [\frac{mg}{g}] \\
= \; & 195.084 \times 5.4 \times 10^4 \sum iC_i\ [mg]
\end{aligned}
$$

APPENDIX E

DERIVING THE PROBABILITY DENSITY FUNCTION

The distribution of $\boldsymbol{U}$ can be derived by the definition of U and η. The sample space of U is $\{0, 1\}$. The π_i is defined by $\pi_i = P(U_i = 1)$, and $1 - \pi_i = P(U_i = 0$. Thus, $f(U_i = 1|\delta) = \pi_i$, and $f(U_i = 0|\delta) = 1 - \pi_i$. Recall that $\eta_i = log(\pi_i/(1 - \pi_i))$. Then, $exp\{\eta_i\} = \pi_i/(1 - \pi_i)$. If $f(U_i|\delta) = exp\{U_i\eta_i\} \cdot (1 - \pi_i)$, then $f(U = 1|\delta) = exp\{\eta_i\}\cdot(1-\pi_i) = \pi_i/(1-\pi_i)\cdot(1-\pi_i) = \pi_i$ and $f(U = 0|\delta) = exp\{0\}\cdot(1-\pi_i) = 1-\pi_i$. Therefore, $f(U|\delta) = exp\{U_i\eta_i\} \cdot (1 - \pi_i)$. To make $f(U_i|\delta)$ in a function of η_i, the second term, $1-\pi_i$, is transformed to $(1+\exp\{\eta\})^{-1}$. Therefore, $f(U_i|\delta) = exp\{U_i\eta_i\}\cdot (1+\exp\{\eta\})^{-1} = exp\{U_i\eta_i - log[1+exp(\eta_i)]\}$. As mentioned before, U_i is independent for $i = 1, 2, \cdots, m$, so $f(U|\delta)$ is the product of $f(U_i|\delta)$ in Equation (38).

APPENDIX F

DISTRIBUTION OF X TILDE

Proposition: When there is a second order model $y = \beta_0 + \beta_1 x + \beta_2 x^2$, model vertex $\tilde{X} = -\beta_1/2\beta_2$ follows a normal distribution.

Proof: For the simplicity, let $\beta_1 = X$ and $\beta_2 = Y$. Then, $\tilde{X}$ becomes $-X/2Y$. If a random variable W follows a normal distribution, then $-2W$ follows a normal distribution as well. Thus, the normalness of X/Y derives the normalness of $-X/2Y$. Because a normal distribution is symmetric, without loss of generality, the proof is based on assuming $Y > 0$.

We know that the estimated parameters in a linear polynomial model have a normal distribution. Thus, X and Y follows a normal distribution. However, they are correlated. Let the joint density of (X, Y) is $g(x, y)$. Then the c.d.f. of $W = X/Y$, $f(w)$ is

$$\begin{aligned} F_W(w) &= P(W \leq w) \\ &= P(X/Y \leq w) \\ &= P(X \leq wY) \\ &= \int_0^{\infty} \int_{-\infty}^{yw} g(x, y) dx dy. \end{aligned}$$

Thus, the p.d.f. of W is the first derivation of $F_W(w)$ over w, that is, $f(w) =$

$dF_W(w)/dw$.

$$\begin{aligned} f(w) &= \frac{d}{dw}\int_0^\infty \int_{-\infty}^{yw} g(x,y)dxdy \\ &= \int_0^\infty \frac{d}{dw}\int_{-\infty}^{yw} g(x,y)dxdy \\ &= \int_0^\infty y \cdot g(yw,y)dy \\ &= \int_{-\infty}^\infty |y|\ g(wy,y)dy. \end{aligned}$$

When $X \sim N(\theta_1, \sigma_1^2)$ and $Y \sim N(\theta_2, \sigma_2^2)$, the cumulative distribution function $F(W)$ of W is found by direct calculation to be

$$F(W) = L\{\frac{\theta_1 - \theta_2 w}{\sigma_1\sigma_2 a(w)}, -\frac{\theta_2}{\sigma_2}; \frac{\sigma_2 w - \rho\sigma_2}{\sigma_1\sigma_2 a(w)}\} + L\{\frac{\theta_2 w - \theta_1}{\sigma_1\sigma_2 a(w)}, \frac{\theta_2}{\sigma_2}; \frac{\sigma_2 w - \rho\sigma_2}{\sigma_1\sigma_2 a(w)}\},$$

where $L(h, k; \gamma)$ is the standard bivariate normal integral tabulated by the National Bureau of Standards (1959) and $a(w) = (w^2/\sigma_1^2 - 2\rho w/\sigma_1\sigma_2 + 1/\sigma_2^2)^{1/2}$. When $\theta_2/\sigma_2 \to \infty$ [28],

$$F(w) \to \Phi\left(\frac{\theta_2 w - \theta_1}{\sigma_1\sigma_2 a(w)}\right).$$

In the denominator, $a(w)$ is approximately $a(w) \approx a(w_0) + a'(w_0)(w - w_0) + a''(w_0)(w - w_0)^2/2 + \cdots$. When $w_0 = \rho\sigma_1/\sigma_2$, then $a^{(h)}(w_0)$ is zero when $h \geq 1$. Then,

$$F(w) \to \Phi\left(\frac{\theta_2 w - \theta_1}{\sigma_1(1-\rho^2)/\sigma_2}\right).$$

I.e.,

$$W \sim N\left(\frac{\theta_1}{\theta_2}, \left(\frac{\sigma_1(1-\rho^2)/\sigma_2}{\theta_2}\right)^2\right).$$

Condition: $\theta_2/\sigma_2 \to \infty$

REFERENCES

[1] "Benchtop Digital iSeries Thermometer Model." http://www.omega.com/ppt/pptsc.asp?ref=MDSI8_MDSSI8.

[2] "NanoLab,Inc.." http://www.nano-lab.com/cooh-functionalized-nanotubes.html.

[3] "Strem Chenicals, Inc." http://www.strem.com/catalog/v/78-0700/53/.

[4] Alexis, F., Rhee, J., Richie, J., Radovic-Moreno, A., Langer, R., and Farokhzad, O., "New frontiers in nanotechnology for cancer treatment," in *Urologic Oncology: Seminars and Original Investigations*, vol. 26, pp. 74–85, Elsevier, 2008.

[5] Belle, G. and Martin, D., "Sample Size as a Function of Coefficient of Variation and Ratio of Means.," *The American Statistician*, vol. 47, no. 3, 1993.

[6] Berry, S., Carroll, R., and Ruppert, D., "Bayesian smoothing and regression splines for measurement error problems," *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 160–169, 2002.

[7] Binder, H. and Sauerbrei, W., "Adding local components to global functions for continuous covariates in multivariable regression modeing," *Statistics in Medicine*, vol. 29, pp. 808–017, 2008.

[8] Bonami, A. and Estrade, A., "Anisotropic analysis of some gaussian models," *Journal of Fourier Analysis and Applications*, vol. 9, no. 3, pp. 215–236, 2003.

[9] Buckheit, J. and Donoho, D., "Wavelab and reproducible research," *LECTURE NOTES IN STATISTICS-NEW YORK-SPRINGER VERLAG-*, pp. 55–55, 1995.

[10] Campelo, J., Luna, D., Luque, R., Marinas, J., and Romero, A., "Sustainable preparation of supported metal nanoparticles and their applications in catalysis," *ChemSusChem*, vol. 2, no. 1, pp. 18–45, 2009.

[11] Casciato, M., Kim, S., Lu, J., Hess, D., and Grover, M., "Optimization of a carbon dioxide-assisted nanoparticle deposition process using sequential experimental design with adaptive design space," *Industrial & Engineering Chemistry Research*.

[12] Chen, J., Lim, B., Lee, E., and Xia, Y., "Shape-controlled synthesis of platinum nanocrystals for catalytic and electrocatalytic applications," *Nano Today*, vol. 4, no. 1, pp. 81–95, 2009.

[13] Chen, X., "Exact computation of minimum sample size for estimation of binomial parameters," *Journal of Statistical Planning and Inference*, 2011.

[14] Chrastil, J., "Solubility of solids and liquids in supercritical gases," *The Journal of Physical Chemistry*, vol. 86, no. 15, pp. 3016–3021, 1982.

[15] Chrastill, J., "Solubility of solids and liquids in supercritical gases," *Journal of Phys. Chem.*, vol. 86, pp. 3016–3021, 1982.

[16] Dasgupta, T., Ma, C., Joseph, V., Wang, Z., and Wu, C., "Statistical modeling and analysis for robust synthesis of nanostructures," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 594–603, 2008.

[17] Daubechies, I., *Ten lectures on wavelets*. Society for industrial and applied mathematics, 2006.

[18] Davis, C. and Stephens, M., "The covariance matrix of normal order statistics," *Communications in Statistics-Simulation and Computation*, vol. 6, no. 1, pp. 75–81, 1977.

[19] Dewettinck, K., Visscher, A., Deroo, L., and Huyghebaert, A., "Modeling the steady-state thermodynamic operation point of top-spray fluidized bed processing," *Journal of Food Engineering*, vol. 39, no. 2, pp. 131–143, 1999.

[20] Dormand, J. and Prince, P., "A family of embedded runge-kutta formulae," *Journal of computational and applied mathematics*, vol. 6, no. 1, pp. 19–26, 1980.

[21] Ellis, R., "Large deviations for a general class of random vectors," *The Annals of Probability*, vol. 12, no. 1, pp. 1–12, 1984.

[22] Eubank, R., *Nonparametric regression and spline smoothing*, vol. 157. CRC, 1999.

[23] Friedman, J., "Multivariate adaptive regression spines," *The annals of Statistics*, vol. 19, no. 1, pp. 1–141, 1991.

[24] Galdberger, A. and David, R., "Chaos and fractals in human physiology [j]," *Scientific American*, vol. 262, pp. 42–49, 1990.

[25] Gonçalves, P., Riedi, R., and Baraniuk, R., "A simple statistical analysis of wavelet-based multifractal spectrum estimation," in *Signals, Systems & Computers, 1998. Conference Record of the Thirty-Second Asilomar Conference on*, vol. 1, pp. 287–291, IEEE, 1998.

[26] Grunes, J., Zhu, J., and Somorjai, G., "Catalysis and nanoscience," *Chem. Commun.*, no. 18, pp. 2257–2260, 2003.

[27] Hernandez, A. and Grover, M., "Comparison of sampling strategies for gaussian process models, with application to nanoparticle dynamics," *Industrial & Engineering Chemistry Research*, 2011.

[28] Hinkley, D., "On the ratio of two correlated normal random variables," *Biometrika*, vol. 56, no. 3, pp. 635–639, 1969.

[29] Ivanov, P., Amaral, L., Goldberger, A., Havlin, S., Rosenblum, M., Struzik, Z., and Stanley, H., "Multifractality in human heartbeat dynamics," *Arxiv preprint cond-mat/9905329*, 1999.

[30] Javaras, K. and Van Dyk, D., "Multiple imputation for incomplete data with semicontinuous variables," *Journal of the American Statistical Association*, vol. 98, no. 463, pp. 703–715, 2003.

[31] Joseph, V. and Melkote, S., "Statistical adjustments to engineering models," *Journal of Quality Technology*, vol. 41, no. 4, pp. 362–375, 2009.

[32] Jung, Y., Park, Y., Jones, D., Ziegler, T., and Vidakovic, B., "Self-similarity in nmr spectra: an application in assessing the level of cysteine," *Journal of data science: JDS*, vol. 8, no. 1, p. 1, 2010.

[33] Kang, L. and Joseph, V., "Bayesian optimal single arrays for robust parameter design," *Technometrics*, vol. 51, no. 3, pp. 250–261, 2009.

[34] Kennedy, M. and O'Hagan, A., "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001.

[35] Kim, S., "Experimental design methods for nanofabrication processes," 2011.

[36] Lamperti, J., "Semi-stable stochastic processeso)," *Transactions of the American Mathematical Society*, vol. 104, p. 62, 1962.

[37] Lavagnini, I. and Magno, F., "A statistical overview on univariate calibration, inverse regression, and detection limits: Application to gas chromatography/mass spectrometry technique," *Mass spectrometry reviews*, vol. 26, no. 1, pp. 1–18, 2007.

[38] Lehmacher, W. and Wassmer, G., "Adaptive sample size calculations in group sequential trials," *Biometrics*, pp. 1286–1290, 1999.

[39] Lu, J., Jeng, S., Wang, K., and others, "A review of statistical methods for quality improvement and control in nanotechnology," *Journal of Quality Technology*, vol. 41, no. 2, pp. 148–164, 2009.

[40] Lyons, J., Rauh-Pfeiffer, A., Yu, Y., Lu, X., Zurakowski, D., Tompkins, R., Ajami, A., Young, V., and Castillo, L., "Blood glutathione synthesis rates in healthy adults receiving a sulfur amino acid-free diet," *Proceedings of the National Academy of Sciences*, vol. 97, no. 10, p. 5071, 2000.

[41] Mandelbrot, B. and Van Ness, J., "Fractional brownian motions, fractional noises and applications," *SIAM review*, vol. 10, no. 4, pp. 422–437, 1968.

[42] Mao, V. W. and Zhao, L. H., "Free knot polynomial spline confidence intervals," *J.R. Statistics, Soc. B*, vol. 65, no. 4, pp. 901–919, 2003.

[43] McPherson, B., Han, W., and Cole, B., "Two equations of state assembled for basic analysis of multiphase co2 flow and in deep sedimentary basin conditions," *Computers & Geosciences*, vol. 34, no. 5, pp. 427–444, 2008.

[44] Nicolis, O., Ramírez-Cobo, P., and Vidakovic, B., "2d wavelet-based spectra with applications," *Computational Statistics & Data Analysis*, vol. 55, no. 1, pp. 738–751, 2011.

[45] Nychka, D., "Bayesian confidence intervals for smoothing splines," *Journal of the American Statistical Association*, pp. 1134–1143, 1988.

[46] Oakley, J. and O'Hagan, A., "Bayesian inference for the uncertainty distribution of computer model outputs," *Biometrika*, vol. 89, no. 4, pp. 769–784, 2002.

[47] Olsen, M. and Schafer, J., "A two-part random-effects model for semicontinuous longitudinal data," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 730–745, 2001.

[48] Park, Y., Ziegler, T., Gletsu-Miller, N., Liang, Y., Yu, T., Accardi, C., and Jones, D., "Postprandial cysteine/cystine redox potential in human plasma varies with meal content of sulfur amino acids," *The Journal of nutrition*, vol. 140, no. 4, pp. 760–765, 2010.

[49] Raguso, C., Regan, M., and Young, V., "Cysteine kinetics and oxidation at different intakes of methionine and cystine in young adults1," *The American journal of clinical nutrition*, vol. 71, no. 2, pp. 491–499, 2000.

[50] Raudenbush, S., Yang, M., and Yosef, M., "Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace approximation," *Journal of computational and Graphical Statistics*, pp. 141–157, 2000.

[51] Reed, I., Lee, P., and Truong, T., "Spectral representation of fractional brownian motion in n dimensions and its properties," *Information Theory, IEEE Transactions on*, vol. 41, no. 5, pp. 1439–1451, 1995.

[52] Reese, C., Wilson, A., Hamada, M., Martz, H., and Ryan, K., "Integrated analysis of computer and physical experiments," *Technometrics*, vol. 46, no. 2, pp. 153–164, 2004.

[53] Riedi, R., "Multifractal processes," tech. rep., DTIC Document, 1999.

[54] Riedi, R., Crouse, M., Ribeiro, V., and Baraniuk, R., "A multifractal wavelet model with application to network traffic," *Information Theory, IEEE Transactions on*, vol. 45, no. 3, pp. 992–1018, 1999.

[55] Robert B. Gramacy, H. K. H. L., "Adaptive design and analysis of supercomputer experiments," *Technometrics*, vol. 51, no. 2, pp. 130–145.

[56] Sack, J., Schiller, S. B., and Welch, W. J., "Designs for computer experiments," *Technometrics*, vol. 31, no. 1, pp. 41 – 47, 1989.

[57] Sacks, J., Schiller, S., and Welch, W., "Designs for computer experiments," *Technometrics*, pp. 41–47, 1989.

[58] Sacks, J., Welch, W., Mitchell, T., and Wynn, H., "Design and analysis of computer experiments," *Statistical science*, vol. 4, no. 4, pp. 409–423, 1989.

[59] Sarkar, J., John, V., He, J., Brooks, C., Gandhi, D., Nunes, A., Ramanath, G., and Bose, A., "Surfactant-templated synthesis and catalytic properties of patterned nanoporous titania supports loaded with platinum nanoparticles," *Chemistry of Materials*, vol. 20, no. 16, pp. 5301–5306, 2008.

[60] Shampine, L. and Reichelt, M., "The matlab ode suite," *SIAM journal on scientific computing*, vol. 18, no. 1, pp. 1–22, 1997.

[61] Shi, B., "Multiscale statistical analysis of self-similar processes with applications in geophysics and health informatics," 2005.

[62] Shi, B., Moloney, K., Emery, V., Jacko, J., Sainfort, F., and Vidakovic, B., "Multifractal discrimination model of high-frequency pupil-diameter measurements," 2003.

[63] Soltanian, S., Qin, M., Keshavarzi, S., Wang, X., and Dou, S., "Effect of sample size on the magnetic critical current density in nano-sic doped mgb_{2} superconductors," *Physical Review B*, vol. 68, no. 13, p. 134509, 2003.

[64] Somorjai, G. and Park, J., "Colloid science of metal nanoparticle catalysts in 2d and 3d structures. challenges of nucleation, growth, composition, particle shape, size control and their influence on activity and selectivity," *Topics in Catalysis*, vol. 49, no. 3, pp. 126–135, 2008.

[65] Span, R. and Wagner, W., "A new equation of state for carbon dioxide covering the fluid region from the triple-point temperature to 1100 k at pressures up to 800 mpa," *Journal of physical and chemical reference data*, vol. 25, p. 1509, 1996.

[66] Stein, C., "A two-sample test for a linear hypothesis whose power is independent of the variance," *The Annals of Mathematical Statistics*, vol. 16, no. 3, pp. 243–258, 1945.

[67] Tibshirani, R. and Hastie, T., "Local likelihood estimation," *Journal of the American Statistical Association*, pp. 559–567, 1987.

[68] Vidakovic, B., *Statistical modeling by wavelets*. Wiley Online Library, 1999.

[69] Vries, P., "Sampling theory for forest inventory," 1986.

[70] Wahba, G., "Bayesian" confidence intervals" for the cross-validated smoothing spline," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 133–150, 1983.

[71] Wang, H. and West, M., "Bayesian analysis of matrix normal graphical models," *Biometrika*, vol. 96, no. 4, p. 821, 2009.

[72] Wang, Y. and Mei, Y., "Asymptotic optimality theory for decentralized sequential multihypothesis testing problems," *Information Theory, IEEE Transactions on*, vol. 57, no. 10, pp. 7068–7083, 2011.

[73] Wissmann, P. and Grover, M., "A new approach to batch process optimization using experimental design," *AIChE Journal*, vol. 55, no. 2, pp. 342–353, 2009.

[74] Wissmann, P. and Grover, M., "Optimization of a chemical vapor deposition process using sequential experimental design," *Industrial & Engineering Chemistry Research*, vol. 49, no. 12, pp. 5694–5701, 2010.

[75] Wolfe, R. and Carlin, J., "Sample-size calculation for a log-transformed outcome measure," *Controlled clinical trials*, vol. 20, no. 6, pp. 547–554, 1999.

[76] Zhang, Y. and Erkey, C., "Preparation of supported metallic nanoparticles using supercritical fluids: A review," *The Journal of supercritical fluids*, vol. 38, no. 2, pp. 252–267, 2006.

VITA

Hin Kyeol Woo was born in Seoul, South Korea. She received a Bachelor's degree in Mathematics from KAIST. She joined the School of Industrial and Systems Engineering at Georgia Institute of Technology as a master student in 2005 and change the degree for Ph.D. in 2006.