Winter 2008

Improved estimation of Fourier coefficients for ill-posed inverse problems

Mikhail V. Shubov
University of New Hampshire, Durham

Improved Estimation of Fourier Coefficients for Ill-Posed Inverse Problems

BY

Mikhail V. Shubov

Baccalaureate Degrees (BS in Mathematics and Physics), Texas Tech University (1998)
Master's Degree (MS in Physics), Texas Tech University (2000)

DISSERTATION

Submitted to the University of New Hampshire
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

in

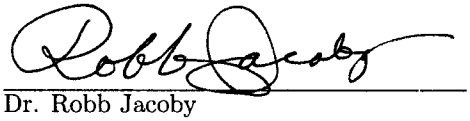
Mathematics

December 2008

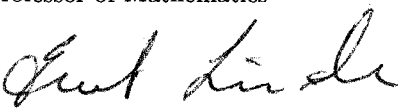
This dissertation has been examined and approved.

Thesis Director Dr. Eric L Grinberg
Professor of Mathematics

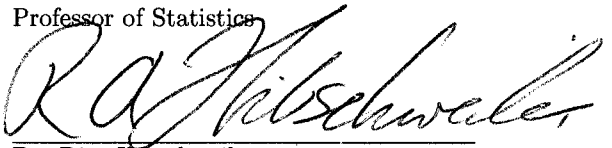
Thesis Co-Director Dr. Fritz Ruymgaart
Horn Professor of Mathematics



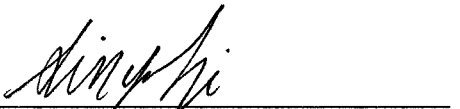
Dr. Robb Jacoby
Professor of Mathematics



Dr. Ernst Linder
Professor of Statistics



Dr. Rita Hibschweiler
Professor of Mathematics



Dr. Linyuan Li
Associate Professor of Statistics

11/19/08
Date

Acknowledgements

First and foremost I would like to thank my advisors Dr. Eric Grinberg and Dr. Fritz Ruymgaart for being wonderful advisors. Dr. Grinberg has given me countless hours of his time and has taught me many skills. Dr. Ruymgaart worked with me not only when I was in Texas Tech, but also when I was a UNH student. I would also like to thank Dr. Robb Jacoby, a great professor who made his courses most interesting. Dr. Jacoby also helped me with my Minor and Major presentations and my Defence. I would also like to thank Dr. Rita Hibschweiler for her help with my dissertation. I would also like thank Dr. Earnst Linder for his help with exploring employment opportunities in Statistics. I also thank Dr. Linyuan Li for his help with my defence. I would also like to thank my parents for their support during my work on PhD.

ABSTRACT

Improved Estimation of Fourier Coefficients for Ill-Posed Inverse Problems

by

Mikhail V. Shubov
University of New Hampshire, December, 2008

In this dissertation we present and solve an ill-posed inverse problem which involves reproducing a function $f(x)$ or its Fourier coefficients from the observed values of the function. The observations of the $f(x)$ are made at n equidistant points on the unit interval with p observations being made at each point. The observations are effected by a random error with a known distribution.

First of all we present a very simple estimator for the Fourier coefficients of $f(x)$. Then we present an iteration algorithm for improving the estimator for the Fourier coefficients. We show that the improved estimator we use is a simplified and improved version of the Maximum Likelihood Estimator.

Second, we introduce the mean squared error(MSE) for the estimators, which is the main measure of estimator performance. We show that a singly iterated estimator has a smaller MSE then a non-iterated estimator and a multiply iterated estimator has a smaller MSE then a singly iterated estimator. We also prove that the errors in estimating the Fourier Coefficients by the singly and multiply improved methods are normally distributed.

Third, we prove a theorem showing that as the sample size goes to infinity, the MSE of our estimator asymptotically approaches the theoretical minimum. That shows that our results are theoretically the "best possible" results.

Fourth, we perform simulations which numerically approximate MSE for a given set of f, error distributions, as well as the number of observation points. We approximate the MSE for the non-iterated error coefficient approximation as well as the singly iterated and multiply iterated ones. We show that indeed the MSE decreases with each iteration. We also plot an error histogram in each case showing that the errors are normally distributed.

Finally, we look at some ways in which our problem can be expanded. Possible expansions include working on the problem in multiple dimensions, taking measurements of f at random points, or both of the above.

Foreward:
Ill-Posed Inverse Problems

We begin this section by giving a general definition of ill-posed inverse problems. Then we give several examples of practical applications of these problems. These applications demonstrate the great importance of ill-posed inverse problems in many aspects of life. Although the focus of this dissertation is restricted to a mathematical study of ill-posed inverse problems, we find it important to demonstrate the significance of ill-posed inverse problems first. Thus we demonstrate the significance of progress in the field of ill-posed inverse problems.

The subject of my work is the statistical approach to ill-posed inverse problems. Quoting [1], "Inverse problems are concerned with determining causes for a desired or an observed effect or calibrating the parameters of a mathematical model to reproduce observations. Ill-posed problems do not fulfill Hadamard's postulates of well-posedness." According to [2] p.85, Hadamard's postulates of well-posedness:

1. The problem has a solution in the strict sense.

2. The solution is unique.

3. The solution depends continuously on the data.

For instance, if a little change in data can produce a large change in solution, then the problem is ill-posed.

Natterer [2] p.85 gives a purely abstract example of an ill-posed problem. Let H, K be Hilbert spaces and let A be a bounded linear operator from H into K Given $g \in K$ find $f \in H$ such that $Af = g$, when A^{-1} does not exist. Several methods have been developed for obtaining f which minimize $\|Af - g\|_K$.

Ill-posed inverse problems have also been discussed by O'Sullivan [3]. These problems involve recovering a whole function given a finite number of noisy measurements on functionals [3, p. 502]. As we will see in section 2, the problem we will consider in this thesis is of the type described above. In these cases, including the case we will present in this work, developing simple algorithms for inversion can be straightforward, while developing an optimal estimator is an open statistical problem. According to O'Sullivan, the traditional least squares or maximum likelihood solutions for an ill-posed reverse problem may not be unique and may be subject to large changes caused by small changes in data.

O'Sullivan also gives some examples of how this seemingly abstract mathematical problem has many practical applications. He gives an example of finding the distribution of tumor sizes with limited data at

hand. In his model, the tumors are spheres randomly distributed through the tissue. We can only observe the radii observed in slices. Given our observation we have to determine the three-dimensional distribution of the tumors in the tissue.

Neutrino tomography is another possible future application. If better neutrino detectors are developed, the internal structure of the Earth could be determined by the Earth's absorption of cosmic neutrinos or neutrinos generated by particle accelerators. Then determining the Earth's density as a function of position would involve solving a three-dimensional ill-posed inverse problem. X-ray tomography involves exactly the same mathematical problem as discussed in my example.

Another ill-posed inverse problem which is still unsolved involves determining the repulsive forces between two atoms as a function of the distance between their nuclei based on scattering angle data. I have encountered this problem during my research on proton scattering in summer of 1996.

Abda [4] describes an ill-posed inverse problem in which the location of cracks in a tested material is deduced by measuring the material's thermal resistivity when heat is applied to different points in the material. In this problem it is necessary to invert the heat equation to get the material's structure from the temperature at the object's boundaries. Unlike my problem, however, Abda' s work is highly theoretical and has no computation. Alessandrini [5] described a similar problem in which electric resistivity over various paths in the material is used in place of surface temperature. Both of the problems described above have important applications in working with precious stones.

Jonas [6] describes an ill-posed inverse problem in phase contrast tomography. This tomography is used to X-ray weakly absorbing materials which change the phase of the X-rays passing through them. The object studied is irradiated with a weak beam of monochromatic X-rays, and the phase distribution of the beam leaving the object is measured. From that information, the structure of the object has to be obtained. Phase contrast tomography is an important field since it has the advantage of using smaller intensity of X-rays than conventional tomography. That decreases the damage to the object studied.

Kurylev [7] discusses a very interesting topic magnetoencephalography – determining the neural current inside the brain from measurements of the magnetic field outside the head. Magnetoencephalography has enormous potential in medicine and perhaps in law enforcement. Currently there is a device at the Veterans Affairs Hospital in Albuquerque, New Mexico, that uses 306 superconducting sensors to track every impulse in a subject's brain [8]. Such devices may be used in the future to monitor the wearer's thoughts. According

to [9], sharks have electric sensors which allow them to find their pray by the electrical impulses within its brain and muscles. All of these scientific advances, however, would be impossible without developing a solution for the inverse problem involved: obtaining source currents from the changing magnetic fields by inverting Maxwell's equations [10].

Kirchgraber [11] gives another example of an ill-posed reverse problem. While I will not go into his extreme simplification, I would like to present the problem. This problem involves estimating the Earth's density distribution from the motion of Earth's artificial satellites. This would be very useful in searching for oil and ore deposits. It will be even more useful in finding out the location of over 14 trillion tons of methane hydrate underneath the ocean floor – the world's vastest and most elusive fossil fuel source [12].

By a very detailed observation of the motion of close artificial satellites, we could find the acceleration due to Earth's gravity at any point above the Earth's atmosphere as a function of the satellite's position in space relative to Earth:

$$g(r) = \frac{d^2r}{dt^2}$$

In order to get meaningful results for Earth's density distribution, we need great accuracy in determining $g(r)$. A $100m$ thick oil, ore or methane hydrate deposit, which changes the ground density by $500kg/m^2$ will change g in the area right above it by 4 parts per million. At the altitude of the satellites, the effect will be much less. Hence, our own measure of $g(r)$ should be accurate to 30 parts per billion or about $3 \times 10^{-7} m/s^2$.

In order to get accurate results for the satellite's position at a given time, and hence its acceleration as a function of position, we will need to install super accurate clocks and pulsating lasers on both the satellite and observation points. Clocks which measure time up to picoseconds and are accurate to one part in 10^{-15} are soon to be available [13]. Lasers which can emit 10 femtosecond pulses are available already. With such technological capabilities, we can measure the satellite's velocity with accuracy of 10^{-15} which would mean $3 \times 10^{-7} m/s$ [14]. From its velocity, we can measure the satellite's acceleration with an accuracy of up to $3 \times 10^{-7} m/s2$. Hence the Earth density mapping project can be technically accomplished.

We know the relation between the Earth's density ρ and the gravitational acceleration g at any point r:

$$g(\,r) = \int\int\int G\rho(x)\,\frac{x-r}{\|x-r\|^3}\,d^3x. \tag{0.0.1}$$

Hence, by solving the inverse problem, we calculate the Earth's density ρ as a function of coordinates with the center of our coordinate system being the Earth's center. Due to the fact that our problem does not have a solution in a strict sense, our problem is ill-posed.

Since numerical estimations can not easily deal with infinitely many points, we will have to approximate the integral (0.0.1) by a finite sum. We will estimate the integral at M points:

$$g(r_j) = \sum_{i=1}^{N} G m_i(x_i) \frac{x_i - r_i}{\|x_i - r_i\|^3}, \quad j = 1...M. \tag{0.0.2}$$

In (0.0.1) the mass distribution of the Earth is approximated by N mass points each containing mass m_i at the point r_i. Since the volume of the Earth is 2.6×10^{11} cubic miles, in order to get a resolution of 10 miles, we will have to consider hundreds of millions of points. Likewise M, or the number of points at which g is measured, should be in hundreds of millions of points, preferably considerably greater than N as we should have more equations than unknowns. If we want a resolution closer to a mile, we will have to look at tens of billions of points.

In essence, (0.0.2) is a system of several hundred million to several tens of billions of linear equations with a somewhat smaller number of unknowns $\{m_i\}$. Our goal is to find the best or least expected error solution. Since the problem is ill-posed, the least square error solution may not be the best, and finding the best solution will require lots of analytic work. Moreover, given the great expense of supercomputers, there is a need to develop the best algorithms using as few calculations as possible.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Our main object of interest is the problem of reconstruction of a one-dimensional function $f : [0, 1] \to \mathbb{R}$ from observations affected by a random error of known distribution. The problem we study is very similar to the one-dimensional version of the image restoration problem [15].

Even though the function $f : [0, 1] \to \mathbb{R}$ we wish to reconstruct can be described by an infinite Fourier series, we will have to estimate it by a finite series:

$$f(x) = \sum_{k=1}^{\infty} v_k \phi_k(x) \approx \sum_{k=1}^{m} v_k \phi_k(x),$$

where $\{\phi_i\}_{i=1\ldots\infty}$ is an orthonormal basis in $L^2[0, 1]$ and $\{v_k\}_{k=1}^{\infty}$ is the set of the Fourier coefficients of f(x). In regular Fourier analysis, the Fourier coefficients can be determined precisely [16], but in our case precise determination is out of the question. Hence, our goal is to find the best possible estimator for the coefficients $\{v_k\}_{k=1}^{m}$.

The estimators we wish to find are denoted $\{\hat{v}_k\}_{k=1}^{m}$. By the best estimator we mean the estimator which minimizes the mean squared error (MSE) of the coefficients. The error is given by

$$\text{MSE} = \text{Exp}\left((v_k - \hat{v}_k)^2\right).$$

Within this work we obtain several important results. First, we present a simple estimator for the Fourier coefficients $\{v_k\}_{k=1}^{m}$. Then we present an iterative algorithm for improving the estimator. We determine the MSE for our estimators. We also show that the algorithm we use does lower the MSE for the estimator of Fourier coefficients with each iteration. As we continue to iterate our results the MSE converges but not to zero. Then we present and prove a theorem showing that there is a limit on how well Fourier coefficients can be estimated and how small an estimator's MSE can be theoretically. Thus

we prove that for the sample size going to infinity, our estimator is in fact the best. Finally, we perform simulations to verify our results.

In Section 2 we present a simple algorithm for obtaining a set of Fourier coefficient estimators from our data set. Then we derive an approximate likelihood function $L\left(\{\hat{v}_k\}_{k=1}^m\right)$ for our estimator set $\{\hat{v}_k\}_{k=1}^m$. The exact likelihood function can not be obtained. Having a simple estimator along with an approximate likelihood function, we can find an approximate maximum likelihood estimator (AMLE) of the Fourier coefficients of f(x). AMLE is an estimator which maximizes an approximate likelihood function. In general, AMLE estimation is used very rarely. While we were not able to find any papers using AMLE in ill-posed inverse problems, we found a work on AMLE estimation being used in linear regression [17]. Given the unprecise nature of the estimation we will devote Section 3 to showing that it still gives an asymptotically best result.

There are two ways of finding the AMLE. One way involves maximizing an approximate likelihood function itself. That is difficult as it involves working with the function $L\left(\{\hat{v}_k\}_{k=1}^m\right)$ which may contain a multiple product. So in Section 2 we show that AMLE can be obtained by maximizing an approximate log likelihood function

$$LL\left(\{\hat{v}_k\}_{k=1}^m\right) = \log\left(L\left(\{\hat{v}_k\}_{k=1}^m\right)\right).$$

Hence we derive an iterative algorithm for obtaining $\{\hat{v}_k\}_{k=1}^m$ which maximize the approximate log likelihood function, and thus we obtain the AMLE. Our initial iteration contains a very complicated formula, but in the end of Section 2 we derive a much simpler iteration and show that it arrives at AMLE as well as its complicated counterpart.

In Section 3 we calculate the MSE for the estimator $\{\hat{v}_k\}_{k=1}^m$ improved by a single iteration or multiple iterations. We show that each iteration decreases the mean squared error, and that these errors converge to a limit. We also show that when the sample size is large the limit obeys the following relation for any k:

$$\lim_{n\to\infty} npI(\psi)E\left(\left(\hat{v}_k - v_k\right)^2\right) = 1. \tag{1.0.1}$$

That result is very important, since in our maximum likelihood estimator we have used an approximate likelihood function, which could have produced a result with unreasonable error. Finally we show that the estimator error $v_k - \hat{v}_k$ for the iterated estimators has a normal distribution.

In Section 4, we determine the minimum possible MSE of our estimator based on the error distribution

function and our sample size. Then we show that the value we have in (1.0.1) is indeed the minimum. From that we conclude that the method we have developed does produce the best estimator at least for large sample sizes.

In Section 5 we perform simulations which show that our analytic results do hold for some representative choices of parameters. In these simulations we show that MSE is indeed reduced by each iteration. We also show that the estimator errors are normally distributed for large sample sizes.

The simulations we perform cover several sample sizes, several functions $f(x)$, several distributions of the error term, and two sets of orthonormal basis of $L^2[0,1]$. Hence, our simulations are broad and relatively comprehensive.

Chapter 2

The Main Estimation Problem

2.1 Introduction

We begin this section by describing the problem we have worked on throughout the thesis. We will describe the data set we observe and the coefficients we wish to obtain from that data set. Then we will introduce a simple estimator for the coefficients we wish to obtain and derive the mean squared error of that estimator. After that we will define the maximum likelihood estimator for the desired coefficients. Finally we will derive an iteration step from simple estimator to the maximum likelihood estimator. Dr Ruymgaart [18] has studied the general version of our problem.

2.2 The functional to be estimated

2.2.1 The description of observed data

We have the following data set $\{Y_{ij}\}_{j=1\ldots p}^{i=1\ldots n}$ given by

$$Y_{ij} = (Kg)\left(\frac{i}{n}\right) + \epsilon_{ij} \qquad i = 1, ..., n \qquad j = 1, ..., p. \tag{2.2.1}$$

In (2.2.1), ϵ_{ij} is an error term which has density ψ. The error distribution function ψ is symmetric and has the finite Fisher information:

$$0 < I(\psi) = \int_{-\infty}^{\infty} \frac{\psi'^2(\epsilon)}{\psi(\epsilon)} \; d\epsilon < \infty.$$

Moreover, the error function obeys the following assumptions:

$$\left[\frac{\psi'(\epsilon)}{\psi(\epsilon)}\right]' \text{ is uniformly continuous on } \mathbb{R}.$$

$$\int_{-\infty}^{\infty} \left|\frac{\psi'(\epsilon)}{\psi(\epsilon)}\right|^{2+\alpha} \psi(\epsilon) \; d\epsilon < \infty \text{ for some } \alpha > 0.$$

The function $g(x)$ mentioned in (2.2.1) is a continuous and hence bounded function on [0,1].

The type of operator $K : L^2[0,1] \to L^2[0,1]$ we have in mind in (2.2.1) is Hermitian. The eigenvectors of K form an orthonormal basis $\{\phi_i\}_{i=1...\infty}$, while the eigenvalues $\{\rho_i\}_{i=1...\infty}$ are such that $\sup\{\rho_i\}_{i=1...\infty} < \infty$. Two examples of such operators are given below.

Example 1. K is an operator transforming any function in $L^2[0,1]$

$$g(x) = \sum_{k=1}^{\infty} g_k \frac{1}{\sqrt{2}} \sin(\pi k x)$$

into the function

$$f(x) = \sum_{v=1}^{\infty} \exp\left(-A(k)\right) g_k \frac{1}{\sqrt{2}} \sin(\pi k x). \tag{2.2.2}$$

In this example, the eigenfunctions of the operator K form a sinusoidal basis $\left\{\frac{1}{\sqrt{2}} \sin(\pi k x)\right\}_{k=1}^{\infty}$. The eigenvalues are $\{\exp\left(-A(k)\right).\}_{k=1}^{\infty}$. Even though this example seems artificial, it actually describes the absorption of sound waves by the air. In that case $A(k)$ in (2.2.2) is the frequency-related absorption coefficient multiplied by the distance to the sound source.

The absorption coefficient $A(k)$ decreases very rapidly with sound frequency [19]. The reason why a distant thunder sounds like a low rumble is that the high frequency sounds are absorbed by the air along the way [20]. Very low frequency waves can travel very long distances before being absorbed – the pressure wave from a Soviet 50 megaton thermonuclear bomb test circled the Earth three times [21].

Example 2. The operator K is given below. First, let us define the extension of the function $g(x) : [0,1] \to \mathbb{R}$ into a function $g_{\text{ext}}(x) : \mathbb{R} \to \mathbb{R}$:

$$g_{\text{ext}}(x) = \begin{cases} -g(-x) & x \in [-1,0) \\ g\left(x - 1 - 2\lfloor \frac{x-1}{2} \rfloor\right) & |x| > 1 \end{cases}. \tag{2.2.3}$$

The operator is given by

$$(Kg)(x) = \lim_{s \to +0} \int_{-\infty}^{x} e^{-ys} \left[\lim_{t \to +0} \int_{-\infty}^{y} e^{-zt} g_{\text{ext}}(z) dz \right] dy. \tag{2.2.4}$$

This operator transforms a function in $L^2[0,1]$

$$g(x) = \sum_{k=1}^{\infty} g_k \frac{1}{\sqrt{2}} \sin(\pi k x)$$

into

$$
\begin{aligned}
(Kg)(x) &= \lim_{s\to+0}\int_{-\infty}^{x} e^{-ys}\left[\lim_{t\to+0}\int_{-\infty}^{y} e^{-zt}\sum_{k=1}^{\infty} g_k\frac{1}{\sqrt{2}}\sin(\pi kz)dz\right]dy \\
&= \sum_{k=1}^{\infty} g_k\frac{1}{\sqrt{2}}\lim_{s\to+0}\int_{-\infty}^{x} e^{-ys}\left[\lim_{t\to+0}\int_{-\infty}^{y} e^{-zt}\sin(\pi kz)dz\right]dy \\
&= \sum_{k=1}^{\infty} g_k\frac{1}{\sqrt{2}}\lim_{s\to+0}\int_{-\infty}^{x} e^{-ys}\left[-\frac{1}{\pi k}\cos(\pi ky)dz\right]dy \\
&= -\sum_{k=1}^{\infty} g_k\frac{1}{\sqrt{2}}\frac{1}{\pi^2k^2}\sin(\pi ky).
\end{aligned}
$$

In this example, the eigenfunctions of the operator K form a sinusoidal basis $\left\{\frac{1}{\sqrt{2}}\sin(\pi kx)\right\}_{k=1}^{\infty}$. The eigenvalues are $\left\{\frac{1}{\pi^2k^2}\right\}_{k=1}^{\infty}$.

2.2.2 The Fourier coefficients

Let us recall that $\{\phi_i\}_{i=1...\infty}$ is an orthonormal basis on $L^2[0,1]$. Then the function $g(x)$ is determined by the series of its Fourier coefficients:

$$
v_{ft} = \int_0^1 \phi_t(x)g(x)\,dx \qquad t = 1, 2, ..., \infty. \tag{2.2.5}
$$

Our goal is to estimate v_{ft} for $t = 1, 2, ..., m$ using the observed data $\{Y_{ij}\}_{j=1...p}^{i=1...n}$.

2.2.3 The transformation of data

First, define the function $f = Kg$. Our data set given by (2.2.1) becomes

$$
f\left(\frac{i}{n}\right) = (Kg)\left(\frac{i}{n}\right) + \epsilon_{ij} \qquad i = 1, ..., n \qquad j = 1, ..., p. \tag{2.2.6}
$$

Our goal is to estimate the coefficients of g(x):

$$
\begin{aligned}
v_{ft} &= < g, \phi(t) > = \frac{1}{\rho_t} < g, \rho_t\phi(t) > = \frac{1}{\rho_t} < g, K\phi(t) > \\
&= \frac{1}{\rho_t} < Kg, \phi(t) > = \frac{1}{\rho_t} < f, \phi(t) > = \frac{v_t}{\rho_t}.
\end{aligned} \tag{2.2.7}
$$

Thus we can estimate the coefficients v_{ft} by estimating the coefficients v_t.

2.3 The simple estimator

2.3.1 The definition of the simple estimator

A simple estimator for the t-th Fourier coefficient of $f(x)$ is given by

$$
\hat{v}_t^{(1)} = \frac{1}{n}\sum_{i=1}^{n} \text{median}\left(\{Y_{tj}\}_{j=1...p}\right)\phi_t\left(\frac{i}{n}\right). \tag{2.3.1}
$$

Having the estimator for Fourier coefficients of $f(x)$, we can estimate the function itself:

$$f_{[m]}^{(1)}(x) = \sum_{k=1}^{m} \hat{v}_k^{(1)} \phi_k(x). \tag{2.3.2}$$

2.3.2 Error of the simple estimator

Claim 1: *If a simple estimator for the t-th Fourier coefficient of $f(x)$ is given by (2.3.1), then the error of the simple estimator is limited by*

$$v_t - \hat{v}_t = O\left(\frac{1}{\sqrt{n}}\right).$$

Proof. Combining (2.2.1), (2.2.5), and (2.3.1) we get

$$\begin{aligned}
v_t - \hat{v}_t &= \int_0^1 \phi_t(x) f(x)\, dx - \frac{1}{n}\sum_{i=1}^{n} \text{median}\left(\{Y_{tj}\}_{j=1\ldots p}\right) \phi_t\left(\frac{i}{n}\right) \\
&= \int_0^1 \phi_t(x) f(x)\, dx - \frac{1}{n}\sum_{i=1}^{n} \text{median}\left(\left\{f\left(\frac{i}{n}\right) + \epsilon_{ij}\right\}_{j=1\ldots p}\right) \phi_t\left(\frac{i}{n}\right) \\
&= \int_0^1 \phi_t(x) f(x)\, dx - \frac{1}{n}\sum_{i=1}^{n} f\left(\frac{i}{n}\right) \phi_t\left(\frac{i}{n}\right) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n} \text{median}\left(\{\epsilon_{ij}\}_{j=1\ldots p}\right) \phi_t\left(\frac{i}{n}\right) \\
&= O\left(\frac{1}{n}\right) + O\left(\frac{1}{\sqrt{n}}\right) = O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

□

2.4 The improved estimator: Maximum Likelihood Estimator

2.4.1 Likelihood and Log-likelihood functions

The likelihood function $L(v_1, ..., v_m)$ is the probability density for the occurrence of the random data $\{Y_{ij}\}_{i=1..n}^{j=1..p}$ we observe given that the probability density for the observed data is known and dependent on the set $(v_1, ..., v_m)$.

The Maximum Likelihood Estimator(MLE) [22] of $(v_1, ..., v_m)$ is the set $(\hat{v}_1, ..., \hat{v}_m)$ such that

$$L(\hat{v}_1, ..., \hat{v}_m) = \max_{\text{All possible } (v_1,...,v_m)} \left(L(v_1, ..., v_m)\right).$$

The method of maximum likelihood was developed by R. A. Fisher between 1912 and 1922 [23].

Sometimes it is more convenient to work with the log likelihood function

$$LL(v_1, ..., v_m) = \ln\left(L(v_1, ..., v_m)\right).$$

Given that the logarithm is a strictly increasing function on $(0, \infty)$, we can find coefficients maximizing the likelihood estimator by finding the coefficients which maximize the log likelihood function.

2.4.2 Log-Likelihood Function for our problem

Let us recall that we observe a data set $\{Y_{ij} = f\left(\frac{i}{n}\right) + \epsilon_{ij}\}_{j=1...p}^{i=1...n}$ mentioned in subsection 2.1.1. Let us also recall that the function $f(x)$ is approximated by

$$f(x) \approx \sum_{k=1}^{m} \hat{v}_k \phi_k(x), \tag{2.4.3}$$

where $\{\phi_k\}_{k=1}^{\infty}$ is an orthonormal basis and $(v_k)_{k=1}^{\infty}$ are the Fourier coefficients we wish to estimate.

By substituting (2.2.1) into (2.4.3) we can obtain an expression for the random error terms based on the unknown Fourier coefficients:

$$\epsilon_{ij} \approx Y_{ij} - \sum_{k=1}^{m} \hat{v}_k \phi_k\left(\frac{i}{n}\right) \tag{2.4.4}$$

The significance of $\approx$ in (2.4.3) and (2.4.4) will be discussed in the next subsection.

From our knowledge of the probability density of the error terms $\epsilon_j^{(i)}$ and from (2.4.4), we can obtain the cumulative likelihood function for the unknown Fourier coefficients:

$$L(\hat{v}_1, ..., \hat{v}_m) \approx \prod_{i=1}^{n}\prod_{j=1}^{p} \psi\left(Y_{ij} - \sum_{k=1}^{m} \hat{v}_k \phi_k\left(\frac{i}{n}\right)\right).$$

The log likelihood estimator will be given by

$$LLE(\hat{v}_1, ..., \hat{v}_m) \approx \sum_{i=1}^{n}\sum_{j=1}^{p} \ln\psi\left(Y_{ij} - \sum_{k=1}^{m} \hat{v}_k \phi_k\left(\frac{i}{n}\right)\right).$$

The log–likelihood function is expressed as a sum, while the likelihood is expressed as a product. Thus, the log likelihood function is more convenient to work with then the likelihood function.

2.4.3 Approximate Maximum Likelihood estimator

The method of approximate maximum likelihood estimator was developed earlier [24], but we have developed its application to the problem described in this chapter. As we have mentioned in the previous section, we do not have the exact likelihood function, thus we will have to work with an approximate one.

That is due to the fact that (2.4.4) presents only an approximation of the error term rather then the term itself. We can rewrite (2.4.4) in the following way:

$$\hat{\epsilon}_{ij} = Y_{ij} - \sum_{k=1}^{m} \hat{v}_k \phi_k \left(\frac{i}{n}\right),$$

where $\hat{\epsilon}_{ij} \approx \epsilon_{ij}$ is the approximation of the error term. We are interested in finding an approximate expected squared error of the approximation. That is done in the derivation below:

$$\begin{aligned}
&E\left((\hat{\epsilon}_{ij} - \epsilon_{ij})^2\right) \\
=&E\left(\left(\left(Y_{ij} - \sum_{k=1}^{m} \hat{v}_k \phi_k \left(\frac{i}{n}\right)\right) - \left(Y_{ij} - f\left(\frac{i}{n}\right)\right)\right)^2\right) \\
=&E\left(\left(f\left(\frac{i}{n}\right) - \sum_{k=1}^{m} \hat{v}_k \phi_k \left(\frac{i}{n}\right)\right)^2\right) \\
=&E\left(\left(\sum_{k=1}^{\infty} v_k \phi_k \left(\frac{i}{n}\right) - \sum_{k=1}^{m} \hat{v}_k \phi_k \left(\frac{i}{n}\right)\right)^2\right) \\
=&E\left(\left(\sum_{k=m}^{\infty} v_k \phi_k \left(\frac{i}{n}\right) + \sum_{k=1}^{m} (v_k - \hat{v}_k) \phi_k \left(\frac{i}{n}\right)\right)^2\right)
\end{aligned} \tag{2.4.5}$$

Recalling that $\{\phi\}_{k=1}^{\infty}$ is an orthonormal basis of [0,1], we can simplify (2.4.5):

$$\begin{aligned}
&E\left((\hat{\epsilon}_{ij} - \epsilon_{ij})^2\right) \\
=&E\left(\left(\sum_{k=1}^{m} (v_k - \hat{v}_k) \phi_k \left(\frac{i}{n}\right) + \sum_{k=m}^{\infty} v_k \phi_k \left(\frac{i}{n}\right)\right)^2\right) \\
=&\sum_{k=1}^{m} (v_k - \hat{v}_k)^2 + \sum_{k=m}^{\infty} v_k^2 + O\left(\frac{1}{n}\right).
\end{aligned} \tag{2.4.6}$$

Later it will be shown that $(v_k - \hat{v}_k)^2 = O\left(\frac{1}{n}\right)$. Hence, (2.4.6) can be simplified

$$E\left((\hat{\epsilon}_{ij} - \epsilon_{ij})^2\right) = \sum_{k=m}^{\infty} v_k^2 + O\left(\frac{m}{n}\right). \tag{2.4.7}$$

Now we have to estimate $\sum_{k=m}^{\infty} v_k^2$ and find m minimizing (2.4.7). We know that the function f(x) is continuous. Nevertheless, unless we also know that there is a finite number of subintervals of [0,1] on which it is increasing or decreasing, nothing can be said about the rate of convergence of $\sum_{k=m}^{\infty} v_k^2$. In almost all practical cases, however, it will be increasing/decreasing only on a finite number of subintervals of [0,1], so its Fourier coefficients obey the following rule [25] p. 441,476:

$$v_k = O\left(\frac{1}{k}\right). \tag{2.4.8}$$

If (2.4.8) holds, then

$$\sum_{k=m}^{\infty} v_k^2 = O\left(\frac{1}{k}\right),$$

hence by setting $m = \sqrt{n}$, we get

$$E\left(\left(\hat{\epsilon}_{ij} - \epsilon_{ij}\right)^2\right) = \sum_{k=m}^{\infty} v_k^2 + O\left(\frac{m}{n}\right)$$
$$= O\left(\frac{1}{\sqrt{n}}\right) + O\left(\frac{\sqrt{n}}{n}\right) = O\left(\frac{1}{\sqrt{n}}\right).$$

Thus in most cases the root mean square difference between ϵ_{ij} and $\hat{\epsilon}_{ij}$ is at most $O\left(\frac{1}{n^{1/4}}\right)$. This rate of convergence is not very fast, and for some functions the convergence will be slower. Thus instead of the maximum likelihood estimator we will have to use an approximate maximum likelihood estimator. In the subsequent chapter we will prove that we still obtain the best possible estimator for the Fourier coefficients.

2.4.4 Newton's Method in Optimization

Our goal is to maximize the function $L(v_1, ..., v_m)$, where $v_k \in \mathbb{R}$ for $k = 1...m$.

Even though the precise formula for finding the set $(\hat{v}_1, ..., \hat{v}_m)$ which maximizes the likelihood function does not exist, we can obtain the MLE via Newton's method in optimization.

We will use Newton's Method in optimization iteration for finding an extremum, which is different from the iteration for finding a zero.

The method has been described by Avriel [26] and Nocedal [27]. The steps of the Newton's method in optimization will be described below.

2.4.5 The steps of the Newton's Method in Optimization

1. Find or guess a set $\mathbf{v}_1 = \left(v_1^{(1)}, ..., v_m^{(1)}\right)$ which is close to the desired result.

2. Apply the following iteration to the above set:

   (2.4.9) $$\mathbf{v}_2 = \mathbf{v}_1 - [HL(\mathbf{v}_1)]^{-1} \nabla L(\mathbf{v}_1),$$

   where the gradient is

   $$\nabla L(\mathbf{v}_1) = \left(\frac{\partial}{\partial v_k} L(\mathbf{v}_1)\right)_{k=1}^{m},$$

   and the Hessian is a matrix given by

   $$HL(\mathbf{v}_1) = \begin{pmatrix} \frac{\partial^2 L(\mathbf{v}_1)}{\partial v_1 \partial v_1} & \cdots & \frac{\partial^2 L(\mathbf{v}_1)}{\partial v_m \partial v_1} \\ \vdots & . & \vdots \\ \frac{\partial^2 L(\mathbf{v}_1)}{\partial v_m \partial v_1} & \cdots & \frac{\partial^2 L(\mathbf{v}_1)}{\partial v_m \partial v_m} \end{pmatrix}.$$

The formula can be used only in the cases where Hessian $HL(\mathbf{v}_1)$ is invertible.

3. Repeat iteration (2.4.9) to obtain the sets $\mathbf{v}_t$ until they converge:

   $\lim_{t\to\infty} \mathbf{v}_t = \mathbf{v}$.

Calculating the Hessian and its inverse for large m may be computationally very difficult. Hence a simplified version of Newton's method in optimization was developed. That method is called the method of **gradient descent [28].** It will be discussed in the next subsection.

2.4.6 The Gradient Descent Method

According to the gradient descent method, performing iterations we take the next "step" in the direction of the function's greatest ascent. This direction is given by $\nabla L(\mathbf{v}_1)$. The size of the step may vary. The steps are the same as in the Newton's method for optimization except that the iteration is different and is given by

$$\mathbf{v_2} = \mathbf{v}_1 - \text{Stepsize}(\psi, \mathbf{v}_1)\nabla L(\mathbf{v}_1),$$

where ψ is the error distribution function and the step size is selected in such a way as to ensure convergence at a reasonable rate. In the diagram below we illustrate iterations via the gradient descent method leading from an initial guess to a maximum.

Determining the best step size will be discussed on the next page.

2.4.7 Finding the best step size for the Gradient Descent Method

Recall that we wish to minimize the function $L(\mathbf{v}) : \mathbb{R}^m \to \mathbb{R}$ via iterations of the form

$$\mathbf{v}_{t+1} = \mathbf{v}_t - \text{Stepsize}(\psi, \mathbf{v}_t)\nabla L(\mathbf{v}_t), \tag{2.4.10}$$

where $\mathbf{v}_{t+1}$, $\mathbf{v}_{t+1}$ and $\nabla L(\mathbf{v}_t)$ are vectors while Stepsize is a scalar. We want to find a step size which maximizes $L(\mathbf{v}_{t+1})$.

The function below is similar to (2.4.10) with Stepsize$(\psi, \mathbf{v}_t)$ replaced by x for simplicity:

$$g(x) = L(\mathbf{v}_t - x\nabla L(\mathbf{v}_t)). \tag{2.4.11}$$

Given that (2.4.11) is a function of one variable, the step used to maximize it is given by Newton's method for optimization:

$$x_{\text{new}} = x - \frac{g'(x)}{g''(x)}.$$

Since the step size x is 0 at the onset, we have

$$\text{Stepsize}(\psi, \mathbf{v}_t) = \left.\frac{g'(x)}{g''(x)}\right|_{x=0}, \tag{2.4.12}$$

where the function $g(x)$ is given by $g(x) = L(\mathbf{v}_t - x\nabla L(\mathbf{v}_t))$.

In order to compute the step size, we need to derive $g'(x)$ and $g''(x)$, both of which are scalars, at $x = 0$. The computation for $g'(x)$ is below:

$$\begin{aligned}
g'(x)|_{x=0} &= \left.\frac{d}{dx} L(\mathbf{v}_t - x\nabla L(\mathbf{v}_t))\right|_{x=0} = \left.\frac{d}{dx} L\left(\mathbf{v}_t - x\nabla L(v_1, ..., v_m)\right)\right|_{x=0} \\
&= \left.\frac{d}{dx} L\left(\left[v_i - x\frac{d}{dv_i} L(v_1, ..., v_m)\right]_{i=1}^{m}\right)\right|_{x=0} \\
&= \left.\sum_{i=1}^{m} \frac{d}{dv_i} L(v_1, ..., v_m) \frac{d}{dv_i} L\left(\left[v_j - x\frac{d}{dv_j} L(v_1, ..., v_m)\right]_{j=1}^{m}\right)\right|_{x=0} \\
&= \sum_{i=1}^{m} \frac{d}{dv_i} L(v_1, ..., v_m) \frac{d}{dv_i} L(v_1, ..., v_m) = \sum_{i=1}^{m} \left[\frac{d}{dv_i} L(v_1, ..., v_m)\right]^2 \\
&= \|\nabla L(\mathbf{v}_t)\|^2.
\end{aligned} \tag{2.4.13}$$

Using (2.4.13), we will derive the expression for $g''(x)$ on the next page. Let us recall the expression for $g'(x)$:

$$g'(x) = \sum_{i=1}^{m} \frac{d}{dv_i} L(v_1, ..., v_m) \frac{d}{dv_i} L\left(\left[v_j - x\frac{d}{dv_j} L(v_1, ..., v_m)\right]_{j=1}^{m}\right).$$

The derivation for $g''(x)$ is below:

$$\begin{aligned}
g''(x)|_{x=0} &= \left.\frac{d}{dx} \sum_{i=1}^{m} \frac{d}{dv_i} L(v_1, ..., v_m) \frac{d}{dv_i} L\left(\left[v_j - x\frac{d}{dv_j} L(v_1, ..., v_m)\right]_{j=1}^{m}\right)\right|_{x=0} \\
&= \left.\sum_{i=1}^{m} \frac{d}{dv_i} L(v_1, ..., v_m) \frac{d}{dx}\left[\frac{d}{dv_i} L\left(\left[v_j - x\frac{d}{dv_j} L(v_1, ..., v_m)\right]_{j=1}^{m}\right)\right]\right|_{x=0} \\
&= \sum_{i=1}^{m} \frac{d}{dv_i} L(v_1, ..., v_m) \\
&\times \left.\sum_{j=1}^{m} \frac{d}{dv_j} L(v_1, ..., v_m) \frac{d^2}{dv_i dv_j} L\left(\left[v_p - x\frac{d}{dv_p} L(v_1, ..., v_m)\right]_{p=1}^{m}\right)\right|_{x=0} \\
&= \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{d}{dv_i} L(v_1, ..., v_m) \frac{d}{dv_j} L(v_1, ..., v_m) \frac{d^2}{dv_i dv_j} L(v_1, ..., v_m) \\
&= \nabla L(\mathbf{v}_s)[HL(\mathbf{v}_s)](\nabla L(\mathbf{v}_s))^{\mathrm{T}}.
\end{aligned} \tag{2.4.14}$$

Substituting (2.4.13) and (2.4.14) into (2.4.12) we get the expression for the step size:

$$\text{Stepsize}(\psi, \mathbf{v}_t) = \left.\frac{g'(x)}{g''(x)}\right|_{x=0} = \frac{\|\nabla L(\mathbf{v}_t)\|^2}{\nabla L(\mathbf{v}_t)[HL(\mathbf{v}_t)](\nabla L(\mathbf{v}_t))^{\mathrm{T}}}.$$

The iteration step becomes

$$\mathbf{v}_{t+1} = \mathbf{v}_t - \frac{\|\nabla L(\mathbf{v}_t)\|^2}{\nabla L(\mathbf{v}_t)[HL(\mathbf{v}_t)](\nabla L(\mathbf{v}_t))^{\mathrm{T}}}\nabla L(\mathbf{v}_t). \tag{2.4.15}$$

When (2.4.15) diverges, we may use a relaxed method where the step size could be multiplied by a constant $\gamma < 1$ to ensure convergence:

$$\mathbf{v}_{t+1} = \mathbf{v}_t - \gamma\frac{\|\nabla L(\mathbf{v}_t)\|^2}{\nabla L(\mathbf{v}_t)[HL(\mathbf{v}_t)](\nabla L(\mathbf{v}_t))^{\mathrm{T}}}\nabla L(\mathbf{v}_t).$$

2.4.8 Gradient Descent Method step for our problem

Let $f(x)$ be a continuous and continuously differentiable function, $f : [0,1] \to \mathbb{R}$. We will represent the Fourier coefficients of $f(x)$ relative to an orthonormal basis $\{\phi_i\}_{i=1\ldots\infty}$ by the array $(v_1, ..., v_m)$. We will use the gradient descent method to find the maximum likelihood estimator of $(v_1, ..., v_m)$

Theorem 1. *Let $\{Y_{ij}\}_{j=1\ldots p}^{i=1\ldots n}$ be a data set for the function $f(x)$:*

$$Y_{ij} = f\left(\frac{i}{n}\right) + \epsilon_{ij}.$$

Here ϵ_{ij} is an error term with density ψ which is symmetric and has finite Fisher information. Then the best gradient iteration step for each element of $(v_1, ..., v_m)$ is:

$$\begin{aligned} v_k^{(s)} &\approx v_k^{(s-1)} \\ &\quad - \frac{1}{npI(\psi)}\sum_{i=1}^{n}\phi_k\left(\frac{i}{n}\right) \\ &\quad \times \sum_{j=1}^{p}\lambda\left(Y_{ij} - \sum_{k=1}^{m} v_k^{(s-1)}\phi_k\left(\frac{i}{n}\right)\right) \\ &\quad k = 1, \ldots, m, \end{aligned} \tag{2.4.16}$$

where the function $\lambda(x)$ is defined as $\lambda(x) = \frac{\psi'(x)}{\psi(x)}$.

Proof. Recall that if we want to maximize the log–likelihood estimator LLE, then the gradient descent method step is given by:

$$\mathbf{v}_{t+1} = \mathbf{v}_t - \mathrm{Stepsize}(\psi, \mathbf{v}_t)\nabla LLE(\mathbf{v}_t), \tag{2.4.17}$$

where the step size is given by

$$\mathrm{Stepsize}(\psi, \mathbf{v}_t) = \frac{\|\nabla LLE(\mathbf{v}_t)\|^2}{\nabla LLE(\mathbf{v}_t)[H(LLE(\mathbf{v}_t))](\nabla LLE(\mathbf{v}_t))^{\mathrm{T}}}. \tag{2.4.18}$$

Recall the log–likelihood function for the problem we are working on:

$$LLE(v_1,...,v_m) \approx \sum_{i=1}^{n}\sum_{j=1}^{p} \ln\psi\left(Y_{ij} - \sum_{k=1}^{m} v_k\phi_k\left(\frac{i}{n}\right)\right). \tag{2.4.19}$$

By substituting (2.4.19) into (2.4.17) we get

$$\mathbf{v}_{t+1} \approx \mathbf{v}_t - \text{Stepsize}\,(\psi,\mathbf{v})\sum_{i=1}^{n}\sum_{j=1}^{p}\lambda\left(Y_{ij} - \sum_{k=1}^{m} v_k\phi_k\left(\frac{i}{n}\right)\right)\left(\phi_k\left(\frac{i}{n}\right)\right)_{k=1}^{m}.$$

We can express the iteration (2.4.13) by stating that the iteration undergone by every component of the vector $\mathbf{v}_t$ is:

$$(v_k)_{new} \approx v_k - \text{Stepsize}\,(\psi,\mathbf{v})\sum_{i=1}^{n}\phi_k\left(\frac{i}{n}\right)\sum_{j=1}^{p}\lambda\left(Y_{ij} - \sum_{k=1}^{m} v_k\phi_k\left(\frac{i}{n}\right)\right) \tag{2.4.20}$$

for $k = 1,\ldots,n$.

Using (2.4.18) and (2.4.19) to determine or approximate the step size will enable us to complete the iteration (2.4.20). Before we begin the calculation, let us introduce some notation to simplify the expressions. Let $\lambda'(x)$ be the function

$$\lambda'(x) = \frac{d}{dx}\left(\frac{\psi'(x)}{\psi(x)}\right) = \frac{(\psi'(x))^2 - \psi(x)\psi''(x)}{\psi^2(x)}. \tag{2.4.21}$$

Now, let us introduce some simplified notations:

$$\psi_{ij} = \psi\left(Y_{ij} - \sum_{k=1}^{m} v_k\phi_k\left(\frac{i}{n}\right)\right),$$

$$\lambda_{ij} = \lambda\left(Y_{ij} - \sum_{k=1}^{m} v_k\phi_k\left(\frac{i}{n}\right)\right), \tag{2.4.22}$$

and

$$\lambda'_{ij} = \lambda'\left(Y_{ij} - \sum_{k=1}^{m} v_k\phi_k\left(\frac{i}{n}\right)\right).$$

We now derive some preliminary results which will be used for calculating the step size. First, let us calculate any single derivative of the log likelihood estimator:

$$\begin{aligned}
&\frac{d}{dv_a}LLE(v_1,...,v_m)\\
\approx&\frac{d}{dv_a}\sum_{i=1}^{n}\sum_{j=1}^{p}\ln\psi\left(Y_{ij} - \sum_{k=1}^{m} v_k\phi_k\left(\frac{i}{n}\right)\right)\\
=&\sum_{i=1}^{n}\phi_a\left(\frac{i}{n}\right)\sum_{j=1}^{p}\lambda\left(Y_{ij} - \sum_{k=1}^{m} v_k\phi_k\left(\frac{i}{n}\right)\right)\\
=&\sum_{i=1}^{n}\phi_a\left(\frac{i}{n}\right)\sum_{j=1}^{p}\lambda_{ij}.
\end{aligned}$$

Now, let us calculate any double derivative of the log likelihood estimator:

$$
\begin{aligned}
&\frac{d}{dv_a}\frac{d}{dv_b}LLE(v_1,...,v_m)\\
\approx&\frac{d}{dv_a}\sum_{i=1}^{n}\phi_b\left(\frac{i}{n}\right)\sum_{j=1}^{p}\lambda\left(Y_{ij}-\sum_{k=1}^{m}v_k\phi_k\left(\frac{i}{n}\right)\right)\\
=&\sum_{i=1}^{n}\phi_a\left(\frac{i}{n}\right)\phi_b\left(\frac{i}{n}\right)\sum_{j=1}^{p}\lambda'\left(Y_{ij}-\sum_{k=1}^{m}v_k\phi_k\left(\frac{i}{n}\right)\right)\\
=&\sum_{i=1}^{n}\phi_a\left(\frac{i}{n}\right)\phi_b\left(\frac{i}{n}\right)\sum_{j=1}^{p}\lambda'_{ij}.
\end{aligned}
\tag{2.4.23}
$$

The above equation is the same whether $a = b$ or $a \neq b$. Now we can calculate the numerator term for the expression for step size (2.4.18):

$$
\begin{aligned}
\|\nabla L(\mathbf{v}_t)\|^2 =&\sum_{a=1}^{m}\left(\frac{d}{dv_a}L(v_1,...,v_m)\right)^2\\
=&\sum_{a=1}^{m}\left(\sum_{i=1}^{n}\phi_a\left(\frac{i}{n}\right)\sum_{j=1}^{p}\lambda_{ij}\right)^2
\end{aligned}
$$

Now, let us calculate the denominator term for the expression for step size given by (2.4.18):

$$
\begin{aligned}
&\nabla L(\mathbf{v}_s)[HL(\mathbf{v}_s)](\nabla L(\mathbf{v}_s))^{\mathrm{T}}\\
=&\sum_{a=1}^{m}\sum_{b=1}^{m}\frac{d}{dv_a}L(v_1,...,v_m)\frac{d}{dv_b}L(v_1,...,v_m)\frac{d^2}{dv_a dv_b}L(v_1,...,v_m)\\
=&\sum_{a=1}^{m}\sum_{b=1}^{m}\begin{bmatrix}\left(\sum_{i_a=1}^{n}\phi_a\left(\frac{i_a}{n}\right)\sum_{j_a=1}^{p}\lambda_{i_a j_a}\right)\times\\ \left(\sum_{i_b=1}^{n}\phi_b\left(\frac{i_b}{n}\right)\sum_{j_b=1}^{p}\lambda_{i_b j_b}\right)\times\\ \left(\sum_{i=1}^{n}\phi_a\left(\frac{i}{n}\right)\phi_b\left(\frac{i}{n}\right)\sum_{j=1}^{p}\lambda'_{ij}\right)\end{bmatrix}.
\end{aligned}
\tag{2.4.24}
$$

By substituting (2.4.23) and (2.4.24) into (2.4.18) we get the expression for the step size:

$$
\begin{aligned}
&\mathrm{Stepsize}(\psi,\mathbf{v}_t)=\frac{\mathrm{NUM}}{\mathrm{DEN}}\\
=&\frac{\sum_{a=1}^{m}\left(\sum_{i=1}^{n}\phi_a\left(\frac{i}{n}\right)\sum_{j=1}^{p}\lambda_{ij}\right)^2}{\sum_{a=1}^{m}\sum_{b=1}^{m}\begin{bmatrix}\left(\sum_{i_a=1}^{n}\phi_a\left(\frac{i_a}{n}\right)\sum_{j_a=1}^{p}\lambda_{i_a j_a}\right)\times\\ \left(\sum_{i_b=1}^{n}\phi_b\left(\frac{i_b}{n}\right)\sum_{j_b=1}^{p}\lambda_{i_b j_b}\right)\times\\ \left(\sum_{i=1}^{n}\phi_a\left(\frac{i}{n}\right)\phi_b\left(\frac{i}{n}\right)\sum_{j=1}^{p}\lambda'_{ij}\right)\end{bmatrix}}.
\end{aligned}
\tag{2.4.25}
$$

The above equation defines NUM and DEN as numerator and denominator of the expression for the step size.

The calculation of the step size can be very complicated. Thus, we are interested in obtaining a much simpler approximation to (2.4.25). Also, we should notice the fact that even if the approximation of the step size for all steps errs by 50% or more, the gradient method is still likely to converge.

We would like to estimate (2.4.25) by estimating its expectation over the random variables used to obtain (2.4.25). In order to do so, we have to express (2.4.25) in terms of these variables. By substituting (2.2.1) and (2.3.2) into (2.4.22) we get

$$\lambda_{ij} = \lambda\left(Y_{ij} - \sum_{k=1}^{m} v_k \phi_k\left(\frac{i}{n}\right)\right) = \lambda\left(Y_{ij} - f_{[m]}\left(\frac{i}{n}\right)\right) \approx \lambda(\epsilon_{ij}),$$

where, as we have mentioned in the beginning of the document, ϵ_{ij} is an iid random variable with distribution ψ. Recall that the expectation of any function of the random variable ϵ_{ij} is

$$E(g(\epsilon_{ij})) = \int_{-\infty}^{\infty} g(\epsilon_{ij})\psi(\epsilon_{ij})\, d\epsilon_{ij}.$$

Also recall that since the variables ϵ_{ij} are independent,

$$E(g_1(\epsilon_{i_1 j_1})g_2(\epsilon_{i_2 j_2})) = E(g_1(\epsilon_{i_1 j_1}))E(g_2(\epsilon_{i_2 j_2})),$$

unless both $i_1 = j_1$ and $i_2 = j_2$.

Now let us calculate some expectations necessary for finding the expectation of (2.4.25). First, let us calculate the expectation of λ_{ij}:

$$\begin{aligned} E(\lambda_{ij}) = E(\lambda(\epsilon_{ij})) &= \int_{-\infty}^{\infty} \lambda(\epsilon_{ij})\psi(\epsilon_{ij})\, d\epsilon_{ij} \\ &= \int_{-\infty}^{\infty} -\frac{\psi'(\epsilon_{ij})}{\psi(\epsilon_{ij})}\psi(\epsilon_{ij})\, d\epsilon_{ij} = -\int_{-\infty}^{\infty} \psi(\epsilon_{ij})\, d\epsilon_{ij} = 0. \end{aligned} \tag{2.4.26}$$

Now let us calculate the expectation of λ_{ij}^2:

$$\begin{aligned} E(\lambda_{ij}^2) = E(\lambda^2(\epsilon_{ij})) &= \int_{-\infty}^{\infty} \lambda^2(\epsilon_{ij})\psi(\epsilon_{ij})\, d\epsilon_{ij} \\ &= \int_{-\infty}^{\infty} \left(-\frac{\psi'(\epsilon_{ij})}{\psi(\epsilon_{ij})}\right)^2 \psi(\epsilon_{ij})\, d\epsilon_{ij} = I(\psi). \end{aligned} \tag{2.4.27}$$

At this point, we can use (2.4.21) to calculate the expectation of λ'_{ij}:

$$\begin{aligned} E(\lambda'_{ij}) = E(\lambda'(\epsilon_{ij})) &= \int_{-\infty}^{\infty} \left(\frac{(\psi'(x))^2 - \psi(x)\psi''(x)}{\psi^2(x)}\right)\psi(\epsilon_{ij})\, d\epsilon_{ij} \\ &= \int_{-\infty}^{\infty} \left(\frac{\psi'(\epsilon_{ij})}{\psi(\epsilon_{ij})}\right)^2 \psi(\epsilon_{ij})\, d\epsilon_{ij} - \int_{-\infty}^{\infty} \psi''(\epsilon_{ij})\, d\epsilon_{ij} \\ &= I(\psi) - 0 = I(\psi). \end{aligned} \tag{2.4.28}$$

Now let us denote the expectation of $\lambda_{ij}^2\lambda'_{ij}$ by $J(\psi) + I^2(\psi)$:

$$\begin{aligned} E(\lambda_{ij}^2\lambda'_{ij}) &= E(\lambda^2(\epsilon_{ij})\lambda'(\epsilon_{ij})) \\ &= \int_{-\infty}^{\infty} \left(\frac{\psi'(\epsilon_{ij})}{\psi(\epsilon_{ij})}\right)^2 \left(\frac{(\psi'(x))^2 - \psi(x)\psi''(x)}{\psi^2(x)}\right)\psi(\epsilon_{ij})\, d\epsilon_{ij} \\ &= J(\psi) + I^2(\psi). \end{aligned} \tag{2.4.29}$$

In effect, $J(\psi)$ is defined by (2.4.29). We do not know whether $J(\psi)$ is positive or negative, but we do know that it is defined by the function ψ and that it is independent of p and n.

At this point, we can use (2.4.26) and (2.4.27) to calculate the expectation of the numerator of (2.4.25) which expresses the step size:

$$
\begin{aligned}
\text{NUM} \approx & E\left(\sum_{a=1}^{m}\left(\sum_{i=1}^{n}\phi_a\left(\frac{i}{n}\right)\sum_{j=1}^{p}\lambda_{ij}\right)^2\right) \\
= & \sum_{a=1}^{m} E\left(\left(\sum_{i=1}^{n}\sum_{j=1}^{p}\phi_a\left(\frac{i}{n}\right)\lambda_{ij}\right)^2\right) \\
= & \sum_{a=1}^{m} E\left(\sum_{i,i_1=1}^{n}\sum_{j,j_1=1}^{p}\phi_a^2\left(\frac{i}{n}\right)\lambda_{ij}\lambda_{i_1 j_1}\right) \\
= & \sum_{a=1}^{m}\left(\sum_{i_1=1}^{n}\sum_{i,i_1=1}^{n}\sum_{j,j_1=1}^{p}\phi_a^2\left(\frac{i}{n}\right)E\left(\lambda_{ij}\lambda_{i_1 j_1}\right)\right) \\
= & \sum_{a=1}^{m}\left(\sum_{i,i_1=1}^{n}\sum_{j,j_1=1}^{p}\phi_a^2\left(\frac{i}{n}\right)I(\psi)\delta_{ii_1}\delta_{jj_1}\right) \\
= & \sum_{a=1}^{m}\left(\sum_{i=1}^{n}\sum_{j=1}^{p}\phi_a^2\left(\frac{i}{n}\right)I(\psi)\right) = pnI(\psi)\sum_{a=1}^{m}\sum_{i=1}^{n}\frac{1}{n}\phi_a^2\left(\frac{i}{n}\right).
\end{aligned}
\tag{2.4.30}
$$

Now we can use equations (2.4.26) – (2.4.29) to calculate the expectation of the denominator of (2.4.25):

$$
\begin{aligned}
\text{DEN} \approx & E\left(\sum_{a,b=1}^{m}\left[\begin{array}{c}\left(\sum_{i_a=1}^{n}\phi_a\left(\frac{i_a}{n}\right)\sum_{j_a=1}^{p}\lambda_{i_a j_a}\right)\times \\ \left(\sum_{i_b=1}^{n}\phi_b\left(\frac{i_b}{n}\right)\sum_{j_b=1}^{p}\lambda_{i_b j_b}\right)\times \\ \left(\sum_{i=1}^{n}\phi_a\left(\frac{i}{n}\right)\phi_b\left(\frac{i}{n}\right)\sum_{j=1}^{p}\lambda'_{ij}\right)\end{array}\right]\right) \\
= & \sum_{a,b=1}^{m}\sum_{i,i_a,i_b=1}^{n}\sum_{j,j_a,j_b=1}^{p}\phi_a\left(\frac{i}{n}\right)\phi_a\left(\frac{i_a}{n}\right)\phi_b\left(\frac{i}{n}\right)\phi_b\left(\frac{i_b}{n}\right)\times E\left(\lambda_{i_a j_a}\lambda_{i_b j_b}\lambda'_{ij}\right).
\end{aligned}
$$

By combining (2.4.27), (2.4.28), and (2.4.29) we obtain

$$
E\left(\lambda_{i_a j_a}\lambda_{i_b j_b}\lambda'_{ij}\right) = I^2(\psi)\delta_{i_a i_b}\delta_{j_a j_b} + J(\psi)\delta_{i_a i_b}\delta_{j_a j_b}\delta_{i i_a}\delta_{j j_a}.
$$

We can continue the calculation of DEN:

$$
\begin{aligned}
\text{DEN} \approx & \sum_{a,b=1}^{m} \sum_{i,i_a,i_b=1}^{n} \sum_{j,j_a,j_b=1}^{p} \phi_a\left(\frac{i}{n}\right)\phi_a\left(\frac{i_a}{n}\right)\phi_b\left(\frac{i}{n}\right)\phi_b\left(\frac{i_b}{n}\right) \\
& \times (I^2(\psi)\delta_{i_a i_b}\delta_{j_a j_b} + J(\psi)\delta_{i_a i_b}\delta_{j_a j_b}\delta_{i i_a}\delta_{j j_a}) \\
& = \sum_{a,b=1}^{m}\left[\sum_{i,i_a=1}^{n}\sum_{j,j_a=1}^{p} \phi_a\left(\frac{i_a}{n}\right)\phi_b\left(\frac{i_a}{n}\right)\phi_a\left(\frac{i}{n}\right)\phi_b\left(\frac{i}{n}\right) I^2(\psi)\right] \\
& + \sum_{a,b=1}^{m}\left[\sum_{i=1}^{n}\sum_{j=1}^{p}\phi_a^2\left(\frac{i}{n}\right)\phi_b^2\left(\frac{i}{n}\right) J(\psi)\right] \\
& = n^2p^2I^2(\psi)\sum_{a,b=1}^{m}\left[\sum_{i_a=1}^{n}\frac{1}{n}\phi_a\left(\frac{i_a}{n}\right)\phi_b\left(\frac{i_a}{n}\right)\right]\left[\sum_{i=1}^{n}\frac{1}{n}\phi_a\left(\frac{i}{n}\right)\phi_b\left(\frac{i}{n}\right)\right] \\
& + npJ(\psi)\sum_{a,b=1}^{m}\left[\sum_{i=1}^{n}\frac{1}{n}\phi_a^2\left(\frac{i}{n}\right)\phi_b^2\left(\frac{i}{n}\right)\right].
\end{aligned}
\tag{2.4.31}
$$

Equations (2.4.30) and (2.4.31) bring some simplification to the numerator and denominator of the expression for the step size (2.4.25). Nevertheless, given that usually n is large and we do not need great precision, we can approximate the sums by integrals. Before doing so, recall that $\{\phi_v\}_{v=1}^{\infty}$ is an orthonormal basis in $L^2([0,1])$.

Now, we can bring further simplification to the numerator of the expression for the step size:

$$
\begin{aligned}
\text{NUM} &\approx pmI(\psi)\sum_{a=1}^{m}\sum_{i=1}^{n}\frac{1}{n}\phi_a^2\left(\frac{i}{n}\right) \approx \\
&\approx pmI(\psi)\sum_{a=1}^{m}\int_0^1 \phi_a^2(x)\,dx = pmI(\psi)\sum_{a=1}^{m}1 = mpnI(\psi).
\end{aligned}
\tag{2.4.32}
$$

We can also bring further simplification to the denominator of the expression for the step size

$$
\begin{aligned}
\text{DEN} &\approx n^2p^2I^2(\psi)\sum_{a,b=1}^{m}\left[\sum_{i_a=1}^{n}\frac{1}{n}\phi_a\left(\frac{i_a}{n}\right)\phi_b\left(\frac{i_a}{n}\right)\right]\left[\sum_{i=1}^{n}\frac{1}{n}\phi_a\left(\frac{i}{n}\right)\phi_b\left(\frac{i}{n}\right)\right] \\
&+npJ(\psi)\sum_{a,b=1}^{m}\left[\sum_{i=1}^{n}\frac{1}{n}\phi_a^2\left(\frac{i}{n}\right)\phi_b^2\left(\frac{i}{n}\right)\right] \\
&\approx n^2p^2I^2(\psi)\sum_{a,b=1}^{m}\int_0^1\phi_a(x)\phi_b(x)\,dx\int_0^1\phi_a(x)\phi_b(x)\,dx \\
&+npJ(\psi)\sum_{a,b=1}^{m}\int_0^1\phi_a^2(x)\phi_b^2(x)\,dx \\
&= n^2p^2I^2(\psi)\sum_{a=1}^{m}1 + npJ(\psi)\sum_{a,b=1}^{m}\int_0^1\phi_a^2(x)\phi_b^2(x)\,dx \\
&= mn^2p^2I^2(\psi) + npJ(\psi)\sum_{a,b=1}^{m}\int_0^1\phi_a^2(x)\phi_b^2(x)\,dx.
\end{aligned}
\tag{2.4.33}
$$

We can further simplify (2.4.33) by leaving out the term with np to the first power. We get

$$\text{DEN} \approx mn^2p^2I^2(\psi). \tag{2.4.34}$$

By combining (2.4.32) and (2.4.34) we get an approximation for the step size

$$\text{Stepsize} \approx \frac{\text{NUM}}{\text{DEN}} \approx \frac{mnpI(\psi)}{mn^2p^2I^2(\psi)} = \frac{1}{npI(\psi)}. \tag{2.4.35}$$

By combining (2.4.20) and (2.4.35) we get the following iteration for each scalar in the vector $\mathbf{v}_s$:

$$v_k^{(s)} \approx v_k^{(s-1)} - \frac{1}{npI(\psi)} \sum_{i=1}^{n} \phi_k\left(\frac{i}{n}\right) \sum_{j=1}^{p} \lambda\left(Y_{ij} - \sum_{k=1}^{m} v_k^{(s-1)} \phi_k\left(\frac{i}{n}\right)\right), \quad k = 1...m.$$

□

We have arrived at (2.4.16) by using many approximations and simplifications, thus at this point we can not be sure that (2.4.16) is indeed the best iteration for $\mathbf{v}_s$. That will be shown by analytic and computational results in the rest of the thesis.

2.4.9 Improvement via a single iteration

We have a simple estimator for the Fourier coefficients given by (2.3.1). We also have the iteration (2.4.16). By applying this iteration until our coefficients converge, we will obtain the maximum likelihood estimator. However, in Subsection 3.2.2 it will be shown in this thesis that even applying (2.3.1) once will bring us significant improvements. In that section we will also see that after just one iteration the result approaches the best possible as $n \to \infty$.

Chapter 3

The Improved Method

3.1 Introduction

In this section we will define the error and the expected error for the estimators of the Fourier coefficients we are trying to find. Then we will calculate the expected error for the the improved estimator defined in the previous section. Recalling that n is the number of points at which we take observations, we will show that for large n the error for the improved estimator has normal distribution.

3.2 Estimation of the expected squared errors

3.2.1 Expected squared errors

The error for any given estimate is defined as $E_t^{(s)} = \hat{v}_t^{(s)} - v_t$.

The expected square error is given by

$$\left(Ee_t^{(s)}\right)^2 = E\left(\left(\hat{v}_t^{(s)} - v_t\right)^2\right).$$

3.2.2 Expected squared error for the improved estimator

Theorem 2. *Let $\{Y_{ij}\}_{j=1\ldots p}^{i=1\ldots n}$ be a data set for the function $f(x)$:*

$$Y_{ij} = f\left(\frac{i}{n}\right) + \epsilon_{ij}.$$

Here ϵ_{ij} is an error term with density ψ which is symmetric and has finite Fisher information. Let $v_k^{(1)}$ be a simple estimator of Fourier coefficients of $f(x)$ given by

$$\hat{v}_k^{(1)} = \frac{1}{n}\sum_{i=1}^{n} median\left(\{Y_{kj}\}_{j=1\ldots p}\right)\phi_k\left(\frac{i}{n}\right).$$

Let $v_k^{(s)}$ be an iterated estimator of Fourier coefficients of $f(x)$ given by

$$v_k^{(s)} = v_k^{(s-1)} - \frac{1}{npI(\psi)} \sum_{i=1}^{n} \phi_k \left(\frac{i}{n}\right) \sum_{j=1}^{p} \lambda \left(Y_{ij} - \sum_{k=1}^{m} v_k^{(s-1)} \phi_k \left(\frac{i}{n}\right) \right), \qquad k = 1...m.$$

Then $v_k^{(s)}$ satisfies the following relation:

$$\lim_{n\to\infty} \left[npI(\psi)\ E\left((\hat{v}_t^{(s)} - v_t)^2 \right) \right] = 1.$$

Proof. First, let us define the error function

$$\varepsilon(x) = f(x) - f_{[m]}(x).$$

Next, let us introduce another relationship:

$$Y_{ij} - f_{[m]}(i/n) = f(i/n) + \epsilon_{ij} - (f(i/n) - \varepsilon(i/n)) = \epsilon_{ij} + \varepsilon(i/n). \tag{3.2.1}$$

Next, let us introduce some relationships between sums and corresponding integrals for any continuous differentiable function $S(x)$:

$$\begin{gathered} \frac{1}{n} \sum_{i=1}^{n} S(i/n) = \int_0^1 S(x)\ dx + M(S)\ O\left(\frac{1}{n}\right), \\ \text{hence} \\ \frac{1}{n} \sum_{i=n}^{n} \phi_k(i/n)\phi_l(i/n) = \delta_{kl} + M(\phi_k\phi_l)\ O\left(\frac{1}{n}\right), \end{gathered} \tag{3.2.2}$$

where

$$M(s) = \frac{1}{2}|s(0) - s(1)| + \max(s')\ O\left(\frac{1}{n}\right). \tag{3.2.3}$$

Using (3.2.2), we can derive

$$\begin{aligned} &\frac{1}{n} \sum_{i=n}^{n} \varepsilon(i/n)\phi_t(i/n) \\ =&\int_0^1 \phi_t(x)\varepsilon(x)\ dx + M(\varepsilon\phi_t)\ O\left(\frac{1}{n}\right) \\ =&\int_0^1 \phi_t(x)f(x)\ dx - \int_0^1 \phi_t(x)f_m(x)\ dx + M(\varepsilon\phi_t)\ O\left(\frac{1}{n}\right) \\ =&v_t - \hat{v}_t + M(\varepsilon\phi_t)\ O\left(\frac{1}{n}\right). \end{aligned} \tag{3.2.4}$$

By substituting (3.2.4) and (3.2.1) into (3.2.3), we get

$$
\begin{aligned}
\hat{v}_t^{(s)} - v_t =& -\frac{1}{n}\sum_{i=n}^{n}\varepsilon(i/n)\phi_t(i/n) \\
&+\frac{1}{npI(\psi)}\sum_{i=1}^{p}\sum_{j=1}^{n}\lambda(Y_{ij}-f_{[m]}(i/n))\phi_t(i/n) + M(\varepsilon\phi_t)\ O\left(\frac{1}{n}\right) \\
=&\frac{1}{npI(\psi)}\sum_{i=1}^{p}\sum_{j=1}^{n}\phi_t(i/n)\left[\lambda(Y_{ij}-f_{[m]}(i/n)) - I(\psi)\varepsilon(j/n)\right] \\
&+M(\varepsilon\phi_t)\ O\left(\frac{1}{n}\right).
\end{aligned}
\tag{3.2.5}
$$

Before working further on expression (3.2.5), we would like to introduce some relationships regarding the quantities ϵ_{ij} which have iid distribution:

$$
E\left(\frac{1}{c}\sum_i\sum_j S_{ij}(\epsilon_{ij})\right) = \frac{1}{c}\sum_i\sum_j E\left(S_{ij}(\epsilon)\right). \tag{3.2.6}
$$

$$
\mathrm{Var}\left(\frac{1}{c}\sum_i\sum_j S_{ij}(\epsilon_{ij})\right) = \frac{1}{c^2}\sum_i\sum_j \mathrm{Var}\left(S_{ij}(\epsilon)\right). \tag{3.2.7}
$$

By substituting (3.2.7) and (3.2.6) into (3.2.5), we can calculate the expectation and the variance of $\hat{v}_t^{(s)} - v_t$. The expectation of $\hat{v}_t^{(s)} - v_t$ is given by

$$
\begin{aligned}
&E\left(\hat{v}_t^{(s)} - v_t\right) \\
=&E\left\{\frac{1}{npI(\psi)}\sum_{j=1}^{n}\sum_{i=1}^{p}\phi_t(i/n)\left[\lambda(\epsilon+\varepsilon(j/n)) - I(\psi)\varepsilon(j/n)\right]\right\} \\
&+E(M(\varepsilon\phi_t))\ O\left(\frac{1}{n}\right) \\
=&\frac{1}{nI(\psi)}\sum_{j=1}^{n}\phi_t(i/n)\left[E(\lambda(\epsilon+\varepsilon(j/n))) - I(\psi)\varepsilon(j/n)\right] \\
&+E(M(\varepsilon\phi_t))\ O\left(\frac{1}{n}\right) \\
=&\frac{1}{I(\psi)}\int_0^1 \phi_t(x)\left[E(\lambda(\epsilon+\varepsilon(x))) - I(\psi)\varepsilon(x)\right]\ dx \\
&+\left\{M\left[E(\lambda(\epsilon+\varepsilon(j/n)))\phi_t - I(\psi)\varepsilon(x)\right] + E(M(\varepsilon\phi_t))\right\}\ O\left(\frac{1}{n}\right).
\end{aligned}
\tag{3.2.8}
$$

Now, we can evaluate the variance as follows:

$$
\begin{aligned}
&\mathrm{Var}\left(\hat{v}_t^{(s)} - v_t\right) \\
=&\mathrm{Var}\left\{\frac{1}{npI(\psi)}\sum_{j=1}^{n}\sum_{i=1}^{p}\phi_t(j/n)\left[\lambda(\epsilon+\varepsilon(j/n)) - I(\psi)\varepsilon(j/n)\right]\right\} \\
&+\mathrm{Var}(M(\varepsilon\phi_t))\ O\left(\frac{1}{n^2}\right) \\
=&\frac{1}{n^2p^2I^2(\psi)}\sum_{j=1}^{n}\sum_{i=1}^{p}\phi_t^2(j/n)\mathrm{Var}\left[\lambda(\epsilon+\varepsilon(j/n))\right] \\
&+\mathrm{Var}(M(\varepsilon\phi_t))\ O\left(\frac{1}{n^2}\right) \\
=&\frac{1}{npI^2(\psi)}\int_0^1 \phi_t^2(x)\mathrm{Var}\left[\lambda(\epsilon+\varepsilon(x))\right]\ dx \\
&+\left[M\left(\phi_t^2(x)\mathrm{Var}(\lambda(\epsilon+\varepsilon(x)))\right) + \ \mathrm{Var}(M(\varepsilon\phi_t))\right] O\left(\frac{1}{n^2}\right).
\end{aligned}
\tag{3.2.9}
$$

Expressions (3.2.8) and (3.2.9) can be significantly simplified once we derive the mean and variance of $\lambda(\epsilon+\varepsilon(x))$.

Now, we will derive the mean of $\lambda(\epsilon+\varepsilon(x))$:

$$
\begin{aligned}
&E(\lambda(\epsilon+\varepsilon(x))) \\
=&\int_{\infty}^{\infty}\lambda(\epsilon+\varepsilon(x))\psi(\epsilon)\ d\epsilon \\
=&\int_{\infty}^{\infty}\left[-\frac{\psi'(\epsilon)}{\phi(\epsilon)} - \frac{d}{d\epsilon}\left(\frac{\psi'(\epsilon)}{\phi(\epsilon)}\right)\varepsilon(x) + \sum_{p=2}^{\infty}\frac{1}{p!}\lambda^p(\epsilon)\varepsilon^p(x)\right]\psi(\epsilon)\ d\epsilon \\
=&I(\psi)\varepsilon(x) + \sum_{p=2}^{\infty}\frac{1}{p!}\varepsilon^p(x)\int_{\infty}^{\infty}\lambda^p(\epsilon)\psi(\epsilon)\ d\epsilon
\end{aligned}
\tag{3.2.10}
$$

Given that $f_{[m]}(x)$ is estimated from data observed at n points, it can be shown that for large n

$$
\varepsilon(x) = f(x) - f_{[m]}(x) = O\left(\frac{1}{\sqrt{n}}\right). \tag{3.2.11}
$$

Now, by substituting (3.2.10) and (3.2.11) into the main component of (3.2.8)

$$
\begin{aligned}
&\frac{1}{I(\psi)}\int_0^1 \phi_t(x)\left[E(\lambda(\epsilon+\varepsilon(x))) - I(\psi)\varepsilon(x)\right]\ dx \\
=&\frac{1}{I(\psi)}\sum_{p=2}^{\infty}\left[\int_0^1\phi_t(x)\varepsilon^p(x)\ dx\right]\left[\frac{1}{p!}\int_{-\infty}^{\infty}\frac{d^p}{d\epsilon^p}(\lambda(\epsilon))\psi(\epsilon)\ \ d\epsilon\right] = O\left(\frac{1}{n}\right).
\end{aligned}
$$

The error term in the expression (3.2.8) can be obtained by using (3.2.11)

$$
\begin{aligned}
&\left\{M\left[E(\lambda(\epsilon+\varepsilon(j/n)))\phi_t - I(\psi)\varepsilon(x)\right] + E(M(\varepsilon\phi_t))\right\}\ O\left(\frac{1}{n}\right) \\
=&\ O\left(\frac{1}{\sqrt{n}}\right)O\left(\frac{1}{n}\right) = O\left(\frac{1}{n^{1.5}}\right).
\end{aligned}
\tag{3.2.12}
$$

By substituting (3.2.11) and (3.2.12) into (3.2.8) we obtain

$$E\left(\hat{\hat{v}}_t - v_t\right) = O\left(\frac{1}{n^{1.5}}\right) + O\left(\frac{1}{n}\right) = O\left(\frac{1}{n}\right).$$

By substituting (3.2.11) into (3.2.10), we get

$$E(\lambda(\epsilon + \varepsilon(x))) = O\left(\frac{1}{n}\right). \tag{3.2.13}$$

Now, we can calculate the variance of $\lambda(\epsilon+\varepsilon(x))$. The only fact we need is the expectation of $\lambda^2(\epsilon+\varepsilon(x))$:

$$\begin{aligned} E(\lambda^2(\epsilon + \varepsilon(x))) &= E\left(\lambda^2(\epsilon) + \frac{d}{d\epsilon}\left(\lambda^2(\epsilon)\right)\varepsilon(x)\right) \\ +E\left(\frac{d^2}{d\epsilon^2}(\lambda^2(\epsilon))\varepsilon^2(x) + \frac{d^3}{d\epsilon^3}\left(\lambda^2(\epsilon)\right)\varepsilon^3(x)\right) &+ E\left(\sum_{k=4}^{\infty}\frac{d^k}{d\epsilon^k}\left(\lambda^2(\epsilon)\right)\varepsilon^p(x)\right) \end{aligned}$$

Given that $\frac{d}{d\epsilon}\left(\lambda^2(\epsilon)\right)$ and $\frac{d^3}{d\epsilon^3}\left(\lambda^2(\epsilon)\right)$ are odd functions with zero expectation and that

$$\frac{d^k}{d\epsilon^k}\left(\lambda^2(\epsilon)\right)\varepsilon^p(x) = O\left(\frac{1}{n^{k/2}}\right)$$

we derive the following:

$$\begin{aligned} &E(\lambda^2(\epsilon + \varepsilon(x))) \\ =&E\left(\lambda^2(\epsilon)\right) + E\left(\frac{d^2}{d\epsilon^2}(\lambda^2(\epsilon))\varepsilon^2(x)\right) + E\left(\sum_{k=4}^{\infty} O\left(\frac{1}{n^{k/2}}\right)\right) \\ =&\int_{-\infty}^{\infty}\left(\frac{\psi'(\epsilon)}{\psi(\epsilon)}\right)^2\psi(\epsilon)\,d\epsilon + \int_{-\infty}^{\infty}\left(\frac{\psi'(\epsilon)}{\psi(\epsilon)}\right)^4\psi(\epsilon)\,d\epsilon\;E(\varepsilon^2(x)) \\ &+ O\left(\frac{1}{n^2}\right)I(\psi) + I(\psi)K(\psi)E(\varepsilon^2(x)) + O\left(\frac{1}{n^2}\right), \end{aligned} \tag{3.2.14}$$

where

$$K(\psi)I(\psi) = \int_{-\infty}^{\infty}\left(\frac{\psi'(\epsilon)}{\psi(\epsilon)}\right)^4\psi(\epsilon)\,d\epsilon.$$

By substituting (3.2.14) and (3.2.13) into the expression for variance we get

$$\begin{aligned} &\mathrm{Var}(\lambda(\epsilon + \varepsilon(x))) \\ =&E(\lambda^2(\epsilon + \varepsilon(x))) - \left(E(\lambda^2(\epsilon + \varepsilon(x)))\right)^2 \\ =&I(\psi) + I(\psi)K(\psi)E(\varepsilon^2(x)) + O\left(\frac{1}{n^2}\right). \end{aligned} \tag{3.2.15}$$

By substituting (3.2.15) into (3.2.9), we can find the variance of our estimator

$$E\left[\left(\hat{v}_t^{(s)} - v_t\right)^2\right] = \mathrm{Var}\left(\hat{v}_t^{(s)} - v_t\right) = \frac{1}{npI^2(\psi)} \int_0^1 \phi_t^2(x)\mathrm{Var}\left[\lambda(\epsilon + \varepsilon(x))\right]\, dx +$$
$$\left[M\left(\phi_t^2(x)\mathrm{Var}(\lambda(\epsilon + \varepsilon(x)))\right) + \mathrm{Var}(M(\varepsilon\phi_t))\right] O\left(\frac{1}{n^2}\right)$$
$$= \frac{1}{npI^2(\psi)} \int_0^1 \phi_t^2(x)\left[I(\psi) + I(\psi)K(\psi)E(\varepsilon^2(x)) + O\left(\frac{1}{n^2}\right)\right]\, dx + O\left(\frac{1}{n^2}\right)$$
$$= \frac{1}{npI(\psi)}\left(1 + K(\psi)\int_0^1 \phi_t^2(x)E(\varepsilon^2(x))\, dx\right) + O\left(\frac{1}{n^2}\right).$$

Finally, let us summarize our results in the following form:

$$\text{(3.2.16)} \qquad \begin{cases} E\left[\left(\hat{v}_t^{(s)} - v_t\right)^2\right] \\ = \frac{1}{npI(\psi)}\left[1 + K(\psi)\int_0^1 \phi_t^2(x)E(\varepsilon^2(x))\, dx\right] + O\left(\frac{1}{n^2}\right) \\ \\ E(\lambda(\epsilon + \varepsilon(x))) = O\left(\frac{1}{n}\right) \end{cases},$$

where

$$K(\psi) = \frac{1}{I(\psi)} \int_{-\infty}^{\infty} \left(\frac{\psi'(\epsilon)}{\psi(\epsilon)}\right)^4 \psi(\epsilon)\, d\epsilon \;>\; 0.$$

Given that

$$\int_0^1 \phi_t^2(x)E(\varepsilon^2(x))\, dx = O\left(\frac{1}{n}\right),$$

we can conclude from (3.2.16) that

$$\lim_{n\to\infty}\left\{npI(\psi)E\left[\left(\hat{v}_t^{(s)} - v_t\right)^2\right]\right\} = 1.$$

□

3.3 Analyzing the result

As we recall from the last section, the expected error after iteration T is given by:

$$\text{(3.3.1)} \qquad \begin{aligned} \left(Ee_t^{(s)}\right)^2 &= E\left[\left(v_t^{(T)} - v_t\right)^2\right] \\ &= \frac{1}{npI(\psi)}\left[1 + K(\psi)\int_0^1 \phi_t^2(x)E(\varepsilon_{(T-1)}^2(x))\, dx\right] + O\left(\frac{1}{n^3}\right), \end{aligned}$$

where $K(\psi)$ is given by

$$K(\psi) = \frac{1}{I(\psi)} \int_{-\infty}^{\infty} \left(\frac{\psi'(\epsilon)}{\psi(\epsilon)}\right)^4 \psi(\epsilon)\, d\epsilon \;>\; 0,$$

while $I(\psi)$ is given by

$$I(\psi) = \int_{-\infty}^{\infty} \frac{(\psi'(x))^2}{\psi(x)}\, dx,$$

and $\varepsilon_{(T-1)}(x)$ is given by

$$\varepsilon_{(T-1)}(x) = f(x) - f_m^{(T)}(x) = f(x) - \sum_{t=1}^{m} v_t^{(T)} \psi_t.$$

In this subsection we will work with the following hypothesis regarding error:

$$\int_0^1 \phi_t^2(x) E(\varepsilon_{(T-1)}^2(x))\, dx = (C_0 + C_1 E E_{(T-1)})^2, \tag{3.3.2}$$

where

$$C_0 = O\left(\frac{1}{n}\right) \quad \text{and} \quad C_1 = O\left(\frac{1}{n}\right).$$

Then, by substituting (3.3.2) into (3.3.1), we get

$$npI(\psi) E E_T^2 = 1 + K(\psi)(C_0 + C_1 E E_{(T-1)})^2 + O\left(\frac{1}{n^3}\right).$$

We have found that, for $s \geq 2$, the following relation holds for the expected square error:

$$\lim_{n\to\infty} \left[n \left(E e_t^{(s)} \right)^2 \right] = \frac{1}{pI(\psi)}. \tag{3.3.3}$$

Considering (3.3.3), we would like to introduce the scaled expected error given by:

$$ee_t = \sqrt{npI(\psi)}\; E e_t^{(s)}$$

Moreover, we have found the relation between the expected errors after different numbers of iterations:

$$\begin{aligned} ee_T^2 &= 1 + (C_2 + C_3(ee_{T-1} - 1))^2 + O\left(\frac{1}{n^3}\right) \\ &\approx 1 + (C_2 + C_3(ee_{T-1} - 1))^2. \end{aligned} \tag{3.3.4}$$

In the above equation, the coefficients C_2 and C_3 are such that

$$\begin{cases} C_2 = O\left(\frac{1}{\sqrt{n}}\right) \\ C_3 = O\left(\frac{1}{\sqrt{n}}\right) \end{cases}.$$

If we can estimate the coefficients $\{ee_v\}$,we can find the coefficients C_2 and C_3. First, let us introduce the limit

$$ee_\infty = \lim_{t\to\infty} ee_t.$$

Even though we do not have a mathematical proof of the existence of the limit above, in all the simulations we have done the above limit does exist. So we will assume that the limit exists. From the recurrence relation (3.3.4), ee_∞ is given by

$$ee_\infty^2 \approx 1 + (C_2 + C_3(ee_\infty - 1))^2. \tag{3.3.5}$$

Let us also introduce the rate of convergence:

$$R = \lim_{T\to\infty} \frac{ee_\infty - ee_{T+1}}{ee_\infty - ee_T}. \tag{3.3.6}$$

The rate of convergence has to be less then 1, and the smaller it is, the faster the sequence converges. Once again in all the simulations we have done, the above limit exists. Nevertheless, we do not have a mathematical proof of the existence of the limit above. It is quite possible that for a few cases not covered by our simulations the above limit does not exist. So we will assume that the limit exists. We can rewrite (3.3.6) as

$$R = \lim_{T\to\infty} \frac{ee_\infty - ee_{T+1}}{ee_\infty - ee_T} = \lim_{T\to\infty} \frac{ee_{T+1} - ee_T}{ee_T - ee_{T-1}}.$$

At this point let us rewrite (3.3.4) in two different ways

$$\begin{cases} ee_{T+1}^2 \approx 1 + (C_2 + C_3(ee_T - 1))^2 \\ ee_T^2 \approx 1 + (C_2 + C_3(ee_{T-1} - 1))^2 \end{cases}. \tag{3.3.7}$$

By subtracting the two equations in (3.3.7) we obtain:

$$(ee_{T+1} - ee_T)(ee_{T+1} + ee_T) \approx C_3(ee_{T+1} - ee_T)[2C_2 + C_3(ee_{T+1} + ee_T - 2)].$$

by taking the limit of the above expression as $T \to \infty$ we obtain:

$$ee_\infty R \approx C_3(C_2 + C_3(ee_\infty - 1)). \tag{3.3.8}$$

By solving (3.3.5) and (3.3.8) for C_2 and C_3 we get

$$\begin{cases} C_2 \approx \frac{ee_\infty^2 - 1 - (ee_\infty - 1)ee_\infty R}{\sqrt{ee_\infty^2 - 1}} \\ C_3 \approx \frac{ee_\infty R}{\sqrt{ee_\infty^2 - 1}} \end{cases}.$$

In most simulations, determining ee_∞ fairly accurately is easy, but determining R is very hard and requires a huge sample size.

Another way to estimate the coefficients C_2 and C_3 involves rewriting (3.3.4):

$$\sqrt{ee_T^2 - 1} \approx C_2 + C_3\sqrt{ee_{T-1} - 1}$$

and solving the above equation for several values of T.

3.4 Normal distribution of error

Theorem 3. *Let $\{Y_{ij}\}_{j=1...p}^{i=1...n}$ be a data set for the function $f(x)$:*

$$Y_{ij} = f\left(\frac{i}{n}\right) + \epsilon_{ij}.$$

Here ϵ_{ij} is an error term with density ψ which is symmetric and has finite Fisher information. Let $v_k^{(1)}$ be a simple estimator of Fourier coefficients of $f(x)$ given by

$$\hat{v}_k^{(1)} = \frac{1}{n}\sum_{i=1}^{n} median\left(\{Y_{kj}\}_{j=1...p}\right)\phi_k\left(\frac{i}{n}\right).$$

Let $v_k^{(s)}$ be an iterated estimator of Fourier coefficients of $f(x)$ given by

$$v_k^{(s)} = v_k^{(s-1)} - \frac{1}{npI(\psi)}\sum_{i=1}^{n}\phi_k\left(\frac{i}{n}\right)\sum_{j=1}^{p}\lambda\left(Y_{ij} - \sum_{k=1}^{m} v_k^{(s-1)}\phi_k\left(\frac{i}{n}\right)\right), \qquad k = 1...m.$$

Then $v_k^{(s)}$ satisfies the following relation:

Distribution $\left[\sqrt{npI(\psi)}\ (\hat{v}_t^{(s)} - v_t)\right]$ *goes to* Normal$(0,1)$ *as n goes to ∞.*

Proof. The fact that the limiting variance and expected value of

$\sqrt{npI(\psi)}\ (\hat{v}_t^{(s)} - v_t)$ is 1 is proved in Theorem 1.

Now we have to prove that the errors are indeed normally distributed. By combining (3.2.5), (3.2.11) and (3.2.1) we obtain the following result:

$$\begin{aligned}
&\hat{v}_t^{(s)} - v_t \\
&= \frac{1}{npI(\psi)}\sum_{i=1}^{p}\sum_{j=1}^{n}\phi_t(j/n)\left[\lambda(Y_{ij} - f_{[m]}(j/n)) - I(\psi)\varepsilon(j/n)\right] + O\left(\frac{1}{n^{1.5}}\right) \\
&= \frac{1}{npI(\psi)}\sum_{i=1}^{p}\sum_{j=1}^{n}\phi_t(j/n)\left[\lambda(\epsilon_{ij} + \varepsilon(j/n)) - I(\psi)\varepsilon(j/n)\right] + O\left(\frac{1}{n^{1.5}}\right) \qquad (3.4.1) \\
&= \frac{1}{npI(\psi)}\sum_{i=1}^{p}\sum_{j=1}^{n}\phi_t(j/n)\left[\lambda(\epsilon_{ij}) + \lambda'(\epsilon_{ij})\varepsilon(j/n) + O(\varepsilon^2(j/n)) - I(\psi)\varepsilon(j/n)\right] + O\left(\frac{1}{n^{1.5}}\right) \\
&= \frac{1}{npI(\psi)}\sum_{i=1}^{p}\sum_{j=1}^{n}\phi_t(j/n)\left[\lambda(\epsilon_{ij}) + (\lambda'(\epsilon_{ij}) - I(\psi))\varepsilon(j/n)\right] + O\left(\frac{1}{n}\right).
\end{aligned}$$

Now, let us define the function

$$F(\epsilon, x) = \phi_t(x)\left[\lambda(\epsilon) + (\lambda'(\epsilon) - I(\psi))\varepsilon(x)\right]. \qquad (3.4.2)$$

By substituting (3.4.1) into (3.4.2) we get

$$\sqrt{npI(\psi)}\ (\hat{v}_t^{(s)} - v_t) = \frac{1}{\sqrt{npI(\psi)}}\sum_{i=1}^{p}\sum_{j=1}^{n}F(\epsilon_{ij}, j/n). \qquad (3.4.3)$$

The variables $\{\epsilon_{ij}\}_{i=1\ldots p}^{j=1\ldots n}$ are independent identically distributed random variables. By The Central Limit Theorem,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} F(\epsilon_{ij}, j/n)$$

has a normal limiting distribution. Hence, (3.4.3) also has a normal limiting distribution.

□

Chapter 4

Convolution Theorem

4.1 Forward

The general result showing asymptotic normality and asymptotically minimal variance for the indirect regression has not been obtained by anyone before us. Nevertheless, it would have been possible to derive that result as well as the Convolution Theorem we present in this chapter from the abstract results in literature[29]. Such results were produced by McNeney [30], van der Vaart [31], and Beran [32]. McNeney [30] studies asymptotic efficiency for estimators in semi-parametric models is the study of lower bounds on asymptotic variances via convolution theorems. van der Vaart [31] presents the theory of weak convergence and empirical processes with applications to many applications in statistics. Beran [32] studies asymptotically optimal estimators in abstract metric spaces. Hajek [33] studies limiting distributions of regular estimators, but we determine the exact limiting distribution for the best estimator for the problem we are studying.

In this chapter we work with the concept of "limits of experiments". For instance, in this chapter we prove that the estimator we have developed is the best estimator in limiting case of the sample size going to infinity. That concept was developed by LeCam [34]. We also study infinite dimensional parameters developed by Millar [35] and LeCam [36].

4.2 Preliminary definitions

4.2.1 The Class of Error Probability Densities Ψ

Recall the data $\{Y_{ij}\}_{j=1...p}^{i=1...n}$ given by

$$Y_{ij} = f\left(\frac{i}{n}\right) + \epsilon_{ij} \qquad i = 1, ..., n \qquad j = 1, ..., p.$$

In the above expression, ϵ_{ij} is an error term which has density ψ described in Subsection 2.1.1. The error function obeys the following assumptions:

$$0 < I(\psi) = \int_{-\infty}^{\infty} \frac{\psi'^2(\epsilon)}{\psi(\epsilon)} \, d\epsilon < \infty. \tag{4.2.1}$$

$$\left[\frac{\psi'(\epsilon)}{\psi(\epsilon)}\right]' \text{ is uniformly continuous on } \mathbb{R}. \tag{4.2.2}$$

$$\int_{-\infty}^{\infty} \left|\frac{\psi'(\epsilon)}{\psi(\epsilon)}\right|^{2+\alpha} \psi(\epsilon) \, d\epsilon < \infty \text{ for some } \alpha > 0. \tag{4.2.3}$$

The function $f(x)$ mentioned in (2.2.1) is a finite and continuous function on [0,1].

Define Ψ as the set of all probability densities satisfying (4.2.1)-(4.2.3).

4.2.2 The δ_1 Function

Let us consider the notion of a δ_1 function $\delta_1 : \mathbb{R} \to \mathbb{R}$, which has the following properties:

$$\delta_1, \delta_1', \delta_1'' \text{ are continuous and bounded on } \mathbb{R} \tag{4.2.4}$$

$$\int_{-\infty}^{\infty} \delta_1(x)\psi(x) \, dx = \int_{-\infty}^{\infty} \delta_1'(x)\psi(x) \, dx = 0, \tag{4.2.5}$$

$$\lim_{x \to \pm\infty} \psi(x)\delta_1(x) = 0. \tag{4.2.6}$$

The norm of δ is 1:

$$||\delta_1||_\psi^2 = \int_{-\infty}^{\infty} \delta_1^2(x)\psi(x) \, dx = 1. \tag{4.2.7}$$

Notice that the δ_1 function we use here is different from the conventional δ functional defined by the relations below:

$$\begin{cases} \text{supp}(\delta) = \{0\} \\ \int_{-\infty}^{\infty} \delta(x)f(x) \, dx = f(0) \text{ for all continuous functions } f(x) \end{cases}$$

The functions δ_1 and δ are not only different but also unrelated having only a part of their name in common.

An example of such a function $\delta_1 : \mathbb{R} \to \mathbb{R}$ is

$$\delta_1(x) = 1 - \sqrt{2}\exp\left(-\frac{x^2}{2}\right),$$

and then (4.2.4) - (4.2.7) are verified whenever ψ is a N(0,1) distribution.

In this chapter we will work with the general δ_1 function. Even though it may have been preferable to work with a particular function, as we have seen in (4.2.4) - (4.2.7), the choice of δ_1 function depends on $\psi(x)$. Hence a single function for all choices of $\psi(x)$ may not exist.

The conditions (4.2.4) - (4.2.6) imply that

$$\psi_c(x) = \psi(x)(1 + c\delta_1(x)) \in \Psi, \tag{4.2.8}$$

for c sufficiently small.

4.2.3 The coefficient we wish to estimate

As mentioned before, we wish to estimate Fourier coefficients of $f(x)$. In this section we are going to generalize that problem. We will consider an estimator of

$$T(f, \psi) =< f, \phi >= \int_0^1 f(x)\phi(x)\, dx, \tag{4.2.9}$$

where ϕ obeys the conditions below. First, ϕ has a norm of 1:

$$||\phi|| = \sqrt{\int_0^1 \phi^2(x)\, dx} = 1.$$

Second, $\phi \in C^{(1)}([0,1])$. Note that ϕ is bounded on $[0,1]$.

4.2.4 The Log-Likelihood Ratio

For $\theta \geq 0, \delta_1$ as in (4.2.8), and ϕ as in (4.2.9) let us define

$$f_{p,n,\theta} = f + \frac{\theta}{\sqrt{pn}}\phi, \tag{4.2.10}$$

and

$$\psi_{p,n,\theta} = \psi + \frac{\theta}{\sqrt{pn}}\psi\delta_1.$$

For sufficiently large n, $f_{p,n,\theta} \in L^2([0,1])$ and $\psi_{p,n,\theta} \in \Psi$.

Let us call the hypothesis that $\theta = 0$ as the "null hypothesis" and the hypothesis that $\theta \neq 0$ as the "local alternative". Now we can write the likelihood functions in these two cases.

Under the null hypothesis $\theta = 0$, the likelihood function is:

$$LE_{p,n,0}\left(\{Y_{ij}\}_{j=1\ldots p}^{i=1\ldots n}\right) =$$

$$LE_{f_{p,n,0},\psi_{p,n,0}}\left(\{Y_{ij}\}_{j=1\ldots p}^{i=1\ldots n}\right) = \prod_{i=1}^{n}\prod_{j=1}^{p}\psi\left(Y_{ij} - f\left(\frac{i}{n}\right)\right).$$

Under the local alternative $\theta \neq 0$, the likelihood function is:

$$LE_{p,n,\theta}(Y) = LE_{p,n,\theta}\left(\{Y_{ij}\}_{j=1\ldots p}^{i=1\ldots n}\right) =$$
$$LE_{f_{p,n,\theta},\psi_{p,n,0}}\left(\{Y_{ij}\}_{j=1\ldots p}^{i=1\ldots n}\right) = \prod_{i=1}^{n}\prod_{j=1}^{p}\psi_{p,n,\theta}\left(Y_{ij} - f_{p,n,\theta}\left(\frac{i}{n}\right)\right).$$

The log-likelihood ratio between the two hypotheses is defined below:

$$\begin{aligned}\Lambda_{p,n,\theta} &= \sum_{i=1}^{n}\sum_{j=1}^{p}\log\frac{\psi_{p,n,\theta}(Y_{ij} - f_{p,n,\theta}\left(\frac{i}{n}\right))}{\psi(Y_{ij} - f\left(\frac{i}{n}\right))} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{p}\log\frac{\psi\left(Y_{ij} - f_{p,n,\theta}\left(\frac{i}{n}\right)\right)}{\psi\left(Y_{ij} - f\left(\frac{i}{n}\right)\right)} \\ &+ \sum_{i=1}^{n}\sum_{j=1}^{p}\log\left(1 + \frac{\theta}{\sqrt{np}}\delta_1\left(Y_{ij} - f_{p,n,\theta}\left(\frac{i}{n}\right)\right)\right) \\ &= \Lambda_{p,n,\theta}^{(1)} + \Lambda_{p,n,\theta}^{(2)}.\end{aligned} \tag{4.2.11}$$

Note that (4.2.11) also defines $\Lambda_{p,n,\theta}^{(1)}$ and $\Lambda_{p,n,\theta}^{(2)}$.

4.2.5 Modes of Convergence

Denote convergence in distribution or probability under the null hypothesis ($\theta = 0$) by

$$\rightarrow_{d(0)} \text{ and } \rightarrow_{p(0)} \text{ respectively.}$$

Denote convergence in distribution or probability under local alternatives

($\theta > 0$) by

$$\rightarrow_{d(\theta)} \text{ and } \rightarrow_{p(\theta)} \text{ respectively.}$$

respectively. Throughout the section, $C \in (0, \infty)$ will be used as a generic constant.

4.3 Preliminary relations

Lemma 1. *Let $\{\xi_i\}_{i=1}^{\infty}$ be a set of i.i.d. random variables with finite mean μ and variance $0 < \sigma^2 < \infty$. For any nonzero function $q \in C^{(1)}([0,1])$, we have*

$$\frac{1}{\sqrt{pn}}\sum_{i=1}^{n}\sum_{j=1}^{p} q\left(\frac{i}{n}\right)(\xi_{ij} - \mu) \rightarrow_d N(0, \sigma^2||q||^2),$$

Proof. The condition $q \in C^{(1)}([0,1])$ implies that $q^2 \in C^{(1)}([0,1])$. Hence,

$$\frac{1}{n}\sum_{i=1}^{n} q^2\left(\frac{i}{n}\right) = ||q||^2 + O\left(\frac{1}{n}\right). \tag{4.3.1}$$

From (4.3.1), we conclude that

$$\lim_{n\to\infty}\left[\frac{\max_{i=1,\ldots,n} q^2\left(\frac{i}{n}\right)}{\sum_{i=1}^n q^2\left(\frac{i}{n}\right)}\right]=0.$$

The above equation is called the Noether condition. Given that the Noether condition [37] is satisfied, from the Central Limit Theorem in [38] p. 153 we get

$$\frac{\sum_{i=1}^n\sum_{i=1}^p q\left(\frac{i}{n}\right)(\xi_{ij}-\mu)}{\sqrt{p\,\sigma^2\sum_{i=1}^n q^2\left(\frac{i}{n}\right)}}\to_d N(0,1),\ \text{as } n\to\infty. \tag{4.3.2}$$

Because the left hand side in (4.3.2) can be written as

$$(pn\sigma^2||q||^2)^{-1/2}\times\left\{\sum_{i=1}^n\sum_{j=1}^p q\left(\frac{i}{n}\right)(\xi_{ij}-\mu)\right\}\times||q||\times\left\{\frac{1}{n}\sum_{i=1}^n q^2\left(\frac{i}{n}\right)\right\}^{-1/2}$$

and the product of the last two factors tends to 1 by (4.3.1), the result follows. □

Lemma 2. *The $\Lambda^{(1)}_{p,n,\theta}$ from (4.2.11) can be expressed as*

$$\Lambda^{(1)}_{p,n,\theta}=-\frac{\theta}{\sqrt{pn}}\sum_{i=1}^n\sum_{j=1}^p\phi\left(\frac{i}{n}\right)\lambda(\epsilon_{ij})-\frac{1}{2}\theta^2||\phi||^2I(\psi)+o_{p(0)}(1). \tag{4.3.3}$$

Proof. Recalling (4.2.10), we can expand $\Lambda^{(1)}_{p,n,\theta}$ from (4.2.11) in a Taylor expansion:

$$\begin{aligned}\Lambda^{(1)}_{p,n,\theta}&=\sum_{i=1}^n\sum_{j=1}^p\left\{\log\psi\left(\epsilon_{ij}-\frac{\theta}{\sqrt{pn}}\phi\left(\frac{i}{n}\right)\right)-\log\psi(\epsilon_{ij})\right\}\\&=-\frac{\theta}{\sqrt{pn}}\sum_{i=1}^n\sum_{j=1}^p\phi\left(\frac{i}{n}\right)\lambda(\epsilon_{ij})+\frac{\theta^2}{2pn}\sum_{i=1}^n\sum_{j=1}^p\phi^2\left(\frac{i}{n}\right)\lambda'(\epsilon_{ij})\\&\quad+\frac{\theta^2}{2pn}\sum_{i=1}^n\sum_{j=1}^p\phi^2\left(\frac{i}{n}\right)\left\{\lambda'(\epsilon'_{ij})-\lambda'(\varepsilon_{ij})\right\},\end{aligned} \tag{4.3.4}$$

where ε'_{ij} is between ϵ_{ij} and $\epsilon_{ij}-(\theta/\sqrt{pn})\phi(i/n)$. The last term in (4.3.4) is the residue term.

The uniform continuity of λ' implies that

$$\max_{1\le i\le n}|\lambda'(\epsilon'_{ij})-\lambda'(\epsilon_{ij})|=o(1),\ \text{as } n\to\infty. \tag{4.3.5}$$

From smoothness of ϕ we obtain

$$\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^m\phi^2\left(\frac{i}{n}\right)=m||\phi||^2+O\left(\frac{1}{n}\right). \tag{4.3.6}$$

Using (4.3.5) and (4.3.6) we obtain the convergence of the residue term in (4.3.4) to zero:

$$\begin{aligned}&\left|\frac{\theta^2}{2pn}\sum_{i=1}^n\sum_{j=1}^p\sum_{k=1}^m\phi^2\left(\frac{i}{n}\right)\left\{\lambda'(\epsilon'_{ij})-\lambda'(\epsilon_{ij})\right\}\right|\\&\le\max_{1\le i\le n}|\lambda'(\epsilon'_{ij})-\lambda'(\epsilon_{ij})|\left|\frac{\theta^2}{2pn}\sum_{i=1}^n\sum_{j=1}^p\sum_{k=1}^m\phi^2\left(\frac{i}{n}\right)\right|\to 0,\ \text{as}\ \ n\to\infty.\end{aligned} \tag{4.3.7}$$

Now, we are going to show the following about the second term of the right hand side of (4.3.4):

$$\frac{\theta^2}{2pn}\sum_{i=1}^{n}\sum_{j=1}^{p}\phi^2\left(\frac{i}{n}\right)\lambda'(\epsilon_{ij}) \rightarrow_{p(0)} -\frac{1}{2}\theta^2||\phi||^2 I(\psi). \tag{4.3.8}$$

First we derive the expected value of $\lambda'(\epsilon_{ij})$:

$$\begin{aligned} E(\lambda'(\epsilon_{ij})) &= \int_{-\infty}^{\infty}\left(\frac{\psi'(\epsilon)}{\psi(\epsilon)}\right)'\psi(\epsilon)\,d\epsilon \\ &= \int_{-\infty}^{\infty}\frac{\psi''(\epsilon)\psi(\epsilon)-(\psi'(\epsilon))^2}{\psi^2(\epsilon)}\psi(\epsilon)\,d\epsilon \\ &= \int_{-\infty}^{\infty}\psi''(\epsilon)\,d\epsilon - \int_{-\infty}^{\infty}\frac{(\psi'(\epsilon))^2}{\psi(\epsilon)}\,d\epsilon = -I(\psi). \end{aligned}$$

Now we can apply Lemma 1 with

$$\begin{cases} \xi_{ij} = \lambda'(\epsilon_{ij}) \\ \mu = -I(\psi) \\ \sigma^2 = \mathrm{Var}(\lambda'(\epsilon_{ij})) \\ q = \phi^2 \end{cases},$$

to obtain

$$(pn)^{-1/2}\times\sum_{i=1}^{n}\sum_{j=1}^{p}\phi^2\left(\frac{i}{n}\right)(\lambda'(\epsilon_{ij})+I(\psi)) \rightarrow_d N(0,\sigma^2||q||^2).$$

The above implies that

$$(pn)^{-1}\times\sum_{i=1}^{n}\sum_{j=1}^{p}\phi^2\left(\frac{i}{n}\right)(\lambda'(\epsilon_{ij})+I(\psi)) \rightarrow_{p(0)} 0. \tag{4.3.9}$$

We can rewrite (4.3.9) as

$$\frac{\theta^2}{2pn}\sum_{i=1}^{n}\sum_{j=1}^{p}\phi^2\left(\frac{i}{n}\right)\lambda'(\epsilon_{ij}) \rightarrow_{p(0)} -\frac{1}{2}\theta^2 I(\psi)\sum_{i=1}^{n}\phi^2\left(\frac{i}{n}\right). \tag{4.3.10}$$

Combining (4.3.10) with (4.3.6) for $q=\phi$ we obtain the claim in (4.3.8). Combining (4.3.4), (4.3.7), and (4.3.8) we obtain

$$\Lambda^{(1)}_{p,n,\theta} = -\frac{\theta}{\sqrt{pn}}\sum_{i=1}^{n}\sum_{j=1}^{p}\phi\left(\frac{i}{n}\right)\lambda(\epsilon_{ij}) - \frac{1}{2}\theta^2||\phi||^2 I(\psi) + o_{p(0)}(1).$$

□

Lemma 3. *The $\Lambda^{(2)}_{p,n,\theta}$ from (4.2.11) can be expressed as*

$$\Lambda^{(2)}_{p,n,\theta} = \frac{\theta}{\sqrt{pn}}\sum_{i=1}^{n}\sum_{j=1}^{p}\delta_1(\epsilon_{ij}) - \frac{1}{2}\theta^2||\delta_1||^2_{\psi} + o_{p(0)}(1). \tag{4.3.11}$$

Proof. First, recall the Taylor expansion for the log function:

$$\log(1+x) = x - \frac{1}{2}x^2 + \frac{1}{6}x^3 R(x),\ |R(x)| \le 16 \text{ for } |x| \le \frac{1}{2}. \tag{4.3.12}$$

Using (4.3.12) on $\Lambda^{(2)}_{p,n,\theta}$ from (4.2.11) we obtain

$$\begin{aligned}\Lambda^{(2)}_{p,n,\theta} &= \frac{\theta}{\sqrt{pn}}\sum_{i=1}^{n}\sum_{j=1}^{p}\delta_1\left(\epsilon_{ij}-\frac{\theta}{\sqrt{pn}}\phi\left(\frac{i}{n}\right)\right)\\ &\quad -\frac{1}{2}\frac{\theta^2}{pn}\sum_{i=1}^{n}\sum_{j=1}^{p}\delta_1^2\left(\epsilon_{ij}-\frac{\theta}{\sqrt{pn}}\phi\left(\frac{i}{n}\right)\right)+O\left(\frac{1}{\sqrt{pn}}\right).\end{aligned} \tag{4.3.13}$$

From the Taylor expansion of (4.3.13) we obtain

$$\begin{aligned}\Lambda^{(2)}_{n,\theta} &= \frac{\theta}{\sqrt{pn}}\sum_{i=1}^{n}\sum_{j=1}^{p}\delta_1(\epsilon_{ij})-\frac{\theta^2}{pn}\sum_{i=1}^{n}\sum_{j=1}^{p}\phi\left(\frac{i}{n}\right)\delta_1'(\epsilon_{ij})\\ &\quad +\frac{1}{2}\frac{\theta^3}{(pn)^{1.5}}\sum_{i=1}^{n}\sum_{j=1}^{p}\phi^2\left(\frac{i}{n}\right)\delta_1''(\epsilon_{ij}'')\\ &\quad -\frac{1}{2}\frac{\theta^2}{pn}\sum_{i=1}^{n}\sum_{j=1}^{p}\delta_1^2(\epsilon_{ij})+\frac{1}{2}\frac{\theta^3}{(pn)^{1.5}}\sum_{i=1}^{n}\sum_{j=1}^{p}(\delta_1^2)'(\epsilon_{ij}''')+O\left(\frac{1}{\sqrt{pn}}\right).\end{aligned} \tag{4.3.14}$$

Given that δ_1' is bounded we conclude that

$$\frac{1}{2}\frac{\theta^3}{(pn)^{1.5}}\sum_{i=1}^{n}\sum_{j=1}^{p}\phi^2\left(\frac{i}{n}\right)\delta_1''(\epsilon_{ij}'')+\frac{1}{2}\frac{\theta^3}{(pn)^{1.5}}\sum_{i=1}^{n}\sum_{j=1}^{p}(\delta_1^2)'(\epsilon_{ij}''')=O\left(\frac{1}{\sqrt{n}}\right).$$

Hence we can simplify (4.3.14) to obtain:

$$\begin{aligned}\Lambda^{(2)}_{n,\theta} &= \frac{\theta}{\sqrt{pn}}\sum_{i=1}^{n}\sum_{j=1}^{p}\delta_1(\epsilon_{ij})-\frac{\theta^2}{pn}\sum_{i=1}^{n}\sum_{j=1}^{p}\phi\left(\frac{i}{n}\right)\delta_1'(\epsilon_{ij})\\ &\quad -\frac{1}{2}\frac{\theta^2}{pn}\sum_{i=1}^{n}\sum_{j=1}^{p}\delta_1^2(\epsilon_{ij})+O\left(\frac{1}{\sqrt{pn}}\right).\end{aligned} \tag{4.3.15}$$

We apply Lemma 1 with

$$\begin{cases}\xi_{ij}=\delta_1'(\epsilon_{ij})\\ \mu=\int_{-\infty}^{\infty}\delta_1'(\epsilon)\psi(\epsilon)\,d\epsilon=0\\ \sigma^2=\mathrm{Var}(\delta_1'(\epsilon))=E(\delta_1'(\epsilon)^2)\\ q=\phi\end{cases},$$

we obtain the following:

$$\frac{1}{\sqrt{pn}}\sum_{i=1}^{n}\sum_{j=1}^{p}\phi\left(\frac{i}{n}\right)\delta_1'(\epsilon_{ij})\rightarrow_{d(0)}N(0,||\phi||^2p^2E(\{\delta_1'(\epsilon_{ij})\}^2)). \tag{4.3.16}$$

From (4.3.16) we see that

$$\frac{\theta^2}{pn}\sum_{i=1}^{n}\sum_{j=1}^{p}\phi\left(\frac{i}{n}\right)\delta_1'(\epsilon_{ij})\rightarrow_{p(0)}0. \tag{4.3.17}$$

Now, we simplify the third term of (4.3.15) via the the weak law of large numbers:

$$-\frac{1}{2}\frac{\theta^2}{pn}\sum_{i=1}^{n}\sum_{j=1}^{p}\delta_1^2(\epsilon_{ij})\rightarrow_{p(0)}-\frac{1}{2}\theta^2||\delta_1||_\psi^2. \tag{4.3.18}$$

Combining (4.3.15), (4.3.17), and (4.3.18) we obtain:

$$\Lambda^{(2)}_{p,n,\theta} = \frac{\theta}{\sqrt{pn}} \sum_{i=1}^{n} \sum_{j=1}^{p} \delta_1(\epsilon_{ij}) - \frac{1}{2}\theta^2 ||\delta_1||^2_\psi + o_{p(0)}(1).$$

□

Theorem 4. *Under the null hypothesis* ($\theta = 0$) *we have*

$$\Lambda_{p,n,\theta} \rightarrow_{d(0)} N\left(-\frac{1}{2}\,\theta^2\Delta^2, \theta^2\Delta^2\right), \text{ as } n \rightarrow \infty, \tag{4.3.19}$$

where $\Delta^2 = ||\phi||^2 I(\psi) + ||\delta_1||^2_\psi$.

Proof. Combining (4.2.11), Lemma 2, and Lemma 3 we obtain:

$$\begin{aligned}
\Lambda_{p,n,\theta} &= \Lambda^{(1)}_{p,n,\theta} + \Lambda^{(2)}_{p,n,\theta} \\
&= -\frac{\theta}{\sqrt{pn}} \sum_{i=1}^{n} \sum_{j=1}^{p} \phi\left(\frac{i}{n}\right) \lambda(\epsilon_{ij}) - \frac{1}{2}\theta^2 ||\phi||^2 I(\psi) + o_{p(0)}(1) \\
&+ \frac{\theta}{\sqrt{pn}} \sum_{i=1}^{n} \sum_{j=1}^{p} \delta_1(\epsilon_{ij}) - \frac{1}{2}\theta^2 ||\delta_1||^2_\psi + o_{p(0)}(1) \\
&= \frac{\theta}{\sqrt{pn}} \sum_{i=1}^{n} \sum_{j=1}^{p} \left\{ \delta_1(\epsilon_{ij}) - \phi\left(\frac{i}{n}\right) \lambda(\epsilon_{ij}) \right\} \\
&- \frac{1}{2}\theta^2 \left[||\phi||^2 I(\psi) + ||\delta_1||^2_\psi \right] + o_{p(0)}(1) \\
&= \frac{\theta}{\sqrt{pn}} \sum_{i=1}^{n} \sum_{j=1}^{p} \left\{ \delta_1(\epsilon_{ij}) - \phi\left(\frac{i}{n}\right) \lambda(\epsilon_{ij}) \right\} \\
&- \frac{1}{2}\theta^2 \Delta^2 + o_{p(0)}(1).
\end{aligned} \tag{4.3.20}$$

Recalling (4.2.4) we derive the following

$$\begin{aligned}
&E\left(\delta_1(\epsilon_{ij}) - \phi\left(\frac{i}{n}\right) \lambda(\epsilon_{ij})\right) \\
&= \int_{-\infty}^{\infty} \delta_1(\epsilon)\psi(\epsilon)\, d\epsilon - \phi\left(\frac{i}{n}\right) \int_{-\infty}^{\infty} \frac{\psi'(\epsilon)}{\psi(\epsilon)} \psi(\epsilon)\, d\epsilon \\
&= 0 - \phi\left(\frac{i}{n}\right) \int_{-\infty}^{\infty} \psi'(\epsilon)\, d\epsilon = 0.
\end{aligned} \tag{4.3.21}$$

From (4.2.5) and (4.2.6) we derive

$$E(\delta_1(\varepsilon)\lambda(\varepsilon)) = \int_{-\infty}^{\infty} \delta_1(x)\psi'(x)\, dx = \delta_1(x)\psi(x)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \delta_1'(x)\psi(x)\, dx = 0.$$

Hence,

$$
\begin{aligned}
&E\left(\left\{\delta_1(\epsilon_{ij})-\phi\left(\frac{i}{n}\right)\lambda(\epsilon_{ij})\right\}^2\right)\\
&=E\left(\delta_1^2(\epsilon_{ij})\right)+\phi^2\left(\frac{i}{n}\right)E\left(\{\lambda(\epsilon_{ij})\}^2\right)-2\phi\left(\frac{i}{n}\right)E(\delta_1(\varepsilon)\lambda(\varepsilon))\\
&=E\left(\delta_1^2(\epsilon_{ij})\right)+\phi^2\left(\frac{i}{n}\right)E\left(\{\lambda(\epsilon_{ij})\}^2\right).
\end{aligned}
\tag{4.3.22}
$$

Given that δ_1 is bounded (4.2.4) and that $\lambda(\epsilon)$ has a moment of order $2+\alpha$ (4.2.3), we conclude that for some constant $C<\infty$:

$$
E\left(|\delta_1(\epsilon_{ij})-\phi\left(\frac{i}{n}\right)\lambda(\epsilon_{ij})|^{2+\alpha}\right)<C
$$

for all choices of i and j. Hence,

$$
\begin{aligned}
&\lim_{n\to\infty}\frac{\sum_{i=1}^n\sum_{j=1}^p E\left(|\delta_1(\epsilon_{ij})-\phi\left(\frac{i}{n}\right)\lambda(\epsilon_{ij})|^{2+\alpha}\right)}{\left[\sum_{i=1}^n\sum_{j=1}^p E\left(|\delta_1(\epsilon_{ij})-\phi\left(\frac{i}{n}\right)\lambda(\epsilon_{ij})|^{2}\right)\right]^{1+\alpha/2}}\\
&\le\lim_{n\to\infty}\frac{Cnp}{\left[\sum_{i=1}^n\sum_{j=1}^p E\left(|\delta_1(\epsilon_{ij})-\phi\left(\frac{i}{n}\right)\lambda(\epsilon_{ij})|^{2}\right)\right]^{1+\alpha/2}}\\
&\le\lim_{n\to\infty}\frac{Cnp}{\{np\,E(\delta_1(\epsilon_{ij})^2)\}^{1+\alpha/2}}=\frac{C}{E(\delta_1(\epsilon_{ij})^2)^{1+\alpha/2}}\lim_{n\to\infty}(np)^{-\alpha/2}=0.
\end{aligned}
\tag{4.3.23}
$$

From (4.3.20), (4.3.21) and (4.3.23) we see that the conditions of Lyapunov Central Limit Theorem are fulfilled. Application of the theorem yields

$$
\frac{\Lambda_{p,n,\theta}+\frac{1}{2}\theta^2\Delta^2}{\left[\frac{\theta^2}{pn}\sum_{i=1}^n\sum_{j=1}^p\mathrm{Var}\left\{\delta_1(\epsilon_{ij})-\phi\left(\frac{i}{n}\right)\lambda(\epsilon_{ij})\right\}\right]^{1/2}}\to_{d(0)}N(0,1).
\tag{4.3.24}
$$

From (4.3.21) and (4.3.22) we derive

$$
\begin{aligned}
&\frac{\theta^2}{pn}\sum_{i=1}^n\sum_{j=1}^p\mathrm{Var}\left(\delta_1(\epsilon_{ij})-\phi\left(\frac{i}{n}\right)\lambda(\epsilon_{ij})\right)\\
&=\frac{\theta^2}{pn}\sum_{i=1}^n\sum_{j=1}^p E\left(\left\{\delta_1(\epsilon_{ij})-\phi\left(\frac{i}{n}\right)\lambda(\epsilon_{ij})\right\}^2\right)\\
&=\theta^2E\left(\delta_1^2(\epsilon)\right)+\theta^2\frac{1}{pn}\sum_{i=1}^n\sum_{j=1}^p\phi^2\left(\frac{i}{n}\right)E\left(\{\lambda'(\varepsilon)\}^2\right)\\
&=\theta^2(||\phi||^2I(\psi)+||\delta_1||_\psi^2)+O\left(\frac{1}{n}\right)=\theta^2\Delta^2+O\left(\frac{1}{n}\right).
\end{aligned}
\tag{4.3.25}
$$

Combining (4.3.24) and (4.3.25) we obtain

$$
\frac{\Lambda_{p,n,\theta}+\frac{1}{2}\theta^2\Delta^2}{\theta\Delta}\to_{d(0)}N(0,1)\ \text{ as } n\to\infty.
$$

The theorem follows. □

Remark 1.

According to the general theory as developed by [39], under the local alternatives $\theta \neq 0$, the asymptotic distribution of the log-likelihood ratio is given by

$$\Lambda_{p,n,\theta} \rightarrow_{d(\theta)} N\left(\frac{1}{2}\theta^2\Delta^2, \theta^2\Delta^2\right), \text{ as } n \rightarrow \infty.$$

Recall that under the null hypothesis $\theta = 0$,

$$\Lambda_{p,n,\theta} \rightarrow_{d(\theta)} N\left(-\frac{1}{2}\theta^2\Delta^2, \theta^2\Delta^2\right), \text{ as } n \rightarrow \infty.$$

4.4 Convolution Theorem

4.4.1 Regular Sequence of Estimators

Most concepts useful for this theorem were defined in subsection 4.1 (Preliminary Definitions). Nevertheless, the "regular sequence of estimators" will be defined here as it is an important concept in the Convolution Theorem.

A **regular sequence of estimators** is a sequence of real-valued estimators $\{\hat{T}_n\}$ of the functional $T =< f, \phi >$ such that

$$\sqrt{pn}\{\hat{T}_n - T(f_{p,n,\theta}, \psi_{p,n,0})\} \rightarrow_{d(\theta)} \mathcal{L}_{f,\psi} \text{ as } n \rightarrow \infty, \tag{4.4.1}$$

for each $\theta \geq 0$, where $\mathcal{L}_{f,\psi}$ is a probability distribution on $\mathbb{R}$ that does not depend on $\theta \geq 0$ but depends on the choice of the function f and the error distribution ψ. Note that if we substitute $\theta = 0$ into (4.4.1) we will get

$$\sqrt{pn}\{\hat{T}_n - T(f, \psi)\} \rightarrow_{d(0)} \mathcal{L}_{f,\psi} \text{ as } n \rightarrow \infty.$$

4.4.2 The statement of the Convolution Theorem

Theorem 5. *The weak limit $\mathcal{L}_{f,\psi}$ of any regular sequence of estimators satisfies*

$$\mathcal{L}_{f,\psi} = \mathcal{N}\left(0, \frac{1}{\Delta^2}\right) * \mathcal{D}_{f,\psi},$$

for some probability distribution $D_{f,\psi}$ on $\mathbb{R}$.

4.4.3 Proof of the Convolution Theorem

This proof is patterned on the proof of Theorem 6 in Beran [32]. Recalling the expression for $f_{p,n,\theta}$ from (4.2.10) we derive:

$$\begin{aligned}\sqrt{pn}\{T(f_{p,n,\theta},\psi_{p,n,0})-T(f,\psi)\}=\sqrt{pn}\,[<f_{p,n,\theta},\phi>-<f,\phi>]=\\ \sqrt{pn}\,[<f+\frac{\theta}{\sqrt{pn}}\,\phi,\phi>-<f,\phi>]=\sqrt{pn}<\frac{\theta}{\sqrt{pn}}\phi,\phi>=\theta.\end{aligned} \tag{4.4.2}$$

From (4.4.2) we conclude that

$$\sqrt{pn}\{\hat{T}_n-T(f_{p,n,\theta},\psi_{p,n,\theta})\}=\sqrt{pn}\{\hat{T}_n-T(f,\psi)\}-\theta. \tag{4.4.3}$$

Now, we write the characteristic function of the left hand side of (4.4.3) as:

$$\begin{aligned}&E_\theta\left(\exp\left(it\sqrt{pn}\{\hat{T}_n-T(f_{p,n,\theta},\psi_{p,n,0})\}\right)\right)\\ &=e^{-it\theta}\int\exp\left(it\sqrt{pn}\{\hat{T}_n-T(f,\psi)\}\right)dP_{p,n,\theta}\\ &=e^{-it\theta}\int\exp\left(it\sqrt{pn}\{\hat{T}_n-T(f,\psi)\}+\Lambda_{p,n,\theta}\right)dP_{p,n,0},\end{aligned} \tag{4.4.4}$$

where $\Lambda_{p,n,\theta}$ is as defined in (4.2.11).

Throughout the proof $P_{p,n,0}$ and $P_{p,n,\theta}$ are the probability distributions on $(\mathbb{R}^p,\mathbb{R}^n,\mathcal{B}^n)$ with densities $p_{p,n,0}$ and $p_{p,n,\theta}$ respectively. The densities $p_{p,n,0}$ and $p_{p,n,\theta}$ represent the combined densities of all the error terms $\{\varepsilon_{ij}\}_{i=1\ldots p}^{j=1..n}$ under the null hypothesis and local alternatives respectively.

At this point, let us introduce several notations.

1. The expected value of δ_1: $E(\delta_1^2(\epsilon))=||\delta_1||_\psi^2$.
2. The Fisher Information: $I(\psi)=I$.
3. The scaled estimate error: $\sqrt{pn}\{\hat{T}_n-T(f,\psi)\}=S_n$.
4. The Normal distribution from (4.3.19): $N\left(-\frac{1}{2}\,\theta^2\Delta^2,\theta^2\Delta^2\right)=\mathcal{N}$.

At this point we express several relations in new terms.According to (4.4.1) we have

$$S_{n\to d(0)}\mathcal{L}_{f,\psi}\text{ as }n\to\infty. \tag{4.4.5}$$

According to (4.3.19),

$$\Lambda_{p,n,\theta}\to_{d(0)}\mathcal{N},\text{ as }n\to\infty. \tag{4.4.6}$$

From (4.4.5) and (4.4.6) we can show that for every $\varepsilon > 0$ there is a radius R_ε such that if we select a random pair (x, y) from distribution $(S_n, \Lambda_{p,n,\theta})$ then

$$P(x^2 + y^2 < R_\varepsilon^2) > 1 - \varepsilon.$$

Hence the pairs $(S_n, \Lambda_{p,n,\theta})$ induce a tight collection of probability distributions in the plane.

Thus, by Prokhorov's theorem [40], there exists a subsequence that has a weak limit. Call the limit (S, Λ_θ). Hence, there is a subsequence $(n_k)_{k\geq 1}$ such that

$$(S_{n_k}, \Lambda_{p,n_k,\theta}) \rightarrow_{d(0)} (S, \Lambda_\theta), \tag{4.4.7}$$

From (4.4.5), (4.4.6) and (4.4.7) we conclude that

$$\begin{cases} S =_d \mathcal{L}_{f,\psi} \\ \Lambda_\theta =_d \mathcal{N} \end{cases}. \tag{4.4.8}$$

For convenience, let us define Z as

$$Z = \frac{\Lambda_\theta + \frac{1}{2}\theta^2\Delta^2}{\theta\Delta}. \tag{4.4.9}$$

Combining (4.4.8) with the definition of the distribution $\mathcal{N}$ we find the distribution for Z:

$$Z =_d \mathcal{N}(0, 1).$$

From the above equation along with (4.4.9) we get

$$\Lambda_\theta = -\frac{1}{2}\theta^2\Delta^2 + \theta\Delta Z.$$

At this point, let us recall the Skorokhod's representation theorem [41]:

Theorem 6. (Skorokhod's Representation Theorem) *Let $(S_n, \Lambda_n)_{n=1}^\infty$ be a sequence of probability measures on a topological space $\mathbb{R}^{n\times p}$; suppose that (S_n, Λ_n) converges weakly to some probability measure (S, Λ) on $\mathbb{R}^{n\times p}$ as $n \rightarrow \infty$. Suppose also that the support of μ is separable. Then there exist random variables $(\tilde{S}, \tilde{\Lambda}), (\tilde{S}_n, \tilde{\Lambda}_n)$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that*

- $(\tilde{S}_n, \tilde{\Lambda}_n)_*(\mathbb{P}) = (S_n, \Lambda_n)$ *(i.e. $(\tilde{S}_n, \tilde{\Lambda}_n)$ is the distribution of (S_n, Λ_n));*
- $(\tilde{S}, \tilde{\Lambda})_*(\mathbb{P}) = (S, \Lambda)$ *(i.e. $(\tilde{S}, \tilde{\Lambda})$ is the distribution of (S, Λ)); and*
- $(\tilde{S}_n, \tilde{\Lambda}_n)(\omega) \rightarrow (\tilde{S}, \tilde{\Lambda})(\omega)$ *as $n \rightarrow \infty$ for every $\omega \in \Omega$.*

Note that in the original theorem, general probability spaces other then $\mathbb{R}^{n\times p}$ may be used.

According to Skorokhod's representation theorem there exist pairs of random variables $(\tilde{S}_{n_k}, \tilde{\Lambda}_{p,n_k,\theta}), (\tilde{S}, \tilde{\Lambda}_\theta)$, all defined on one and the same probability space, such that $(\tilde{S}_{n_k}, \tilde{\Lambda}_{p,n_k,\theta}) =_d (S_{n_k}, \Lambda_{p,n_k,\theta}), (\tilde{S}, \tilde{\Lambda}_\theta) =_d (\tilde{S}, \tilde{\Lambda}_\theta)$, and

$$(\tilde{S}_{n_k}, \tilde{\Lambda}_{n_k,\theta}) \to_{a.s.} (\tilde{S}, \tilde{\Lambda}_\theta). \tag{4.4.10}$$

Let us denote expectation on this new probability space by $\tilde{E}$.

Now observe that

$$\begin{aligned}
&\tilde{E}\left(\left|\exp\left[it\tilde{S}+\tilde{\Lambda}_\theta\right]\right|\right) = E\left(\exp\left[-\frac{1}{2}\theta^2\Delta^2+\theta\Delta Z\right]\right)\\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\theta^2\Delta^2+\theta\Delta z-\frac{1}{2}z^2\right)dz\\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}(z-\theta\Delta)^2\right)dz = 1.
\end{aligned} \tag{4.4.11}$$

From (4.4.10) and (4.4.11) it follows that

$$\tilde{E}\left(\left|\exp\left(it\tilde{S}_{n_k}+\tilde{\Lambda}_{p,n_k,\theta}\right)\right|\right) = E\left(\exp\left(\Lambda_{p,n_k,\theta}\right)\right) = 1,$$

Therefore, by Vitali's theorem [42] we obtain

$$\begin{aligned}
&E\left(\exp\left[itS_{n_k}+\Lambda_{n_k,\theta}\right]\right) = \tilde{E}\left(\exp\left[it\tilde{S}_{n_k}+\tilde{\Lambda}_{n_k,\theta}\right]\right) \to\\
&\tilde{E}\left(\exp\left[it\tilde{S}+\tilde{\Lambda}_\theta\right]\right) = E\left(\exp\left[itS+\theta\Delta Z-\frac{1}{2}\theta^2\Delta^2\right]\right).
\end{aligned} \tag{4.4.12}$$

Due to regularity of estimators, we see that

$$\begin{aligned}
&E_\theta\left(\exp\left(it\sqrt{pn}\{\hat{T}_n - T(f_{p,n,\theta},\psi_{p,n,0})\}\right)\right)\\
&\to E(\exp(it\sqrt{pn}\{\hat{T}_n - T(f,\psi)\})) = E(\exp(itS)).
\end{aligned} \tag{4.4.13}$$

Note that the limit in (4.4.13) does not depend on θ. From (4.4.12) we see that

$$\begin{aligned}
&e^{-it\theta}\int\exp\left(it\sqrt{pn}\{\hat{T}_n - T(f,\psi)\}+\Lambda_{p,n,\theta}\right)dP_{p,n,0}\\
&= e^{-it\theta}E\left(\exp\left[itS_{n_k}+\Lambda_{n_k,\theta}\right]\right)\\
&\to \exp(-it\theta-\frac{1}{2}\theta^2\Delta^2)\times E(\exp(itS+\theta\Delta Z)).
\end{aligned} \tag{4.4.14}$$

Substituting (4.4.13) and (4.4.14) into (4.4.4) we obtain

$$E\left(\exp\left[itS\right]\right) = \exp\left[-it\theta-\frac{1}{2}\theta^2\Delta^2\right]\cdot E\left(\exp\left[itS+\theta\Delta Z\right]\right), \tag{4.4.15}$$

for all $\theta > 0$ and each $t \in \mathbb{R}$.

Define the constant function $\gamma(\zeta) = E(\exp(itS))$. Also define the function $\mathcal{X}(\zeta) = \exp(-it\zeta - \frac{1}{2}\zeta^2\Delta^2) \times E(\exp(itS+\zeta\Delta Z))$, for $\zeta \in \mathbb{C}$. The function γ is analytic on $\mathbb{C}$, since it is a constant function. The function $\mathcal{X}$ is also analytic as can be seen by expanding $\exp(\zeta\Delta Z)$ in a power series and exploiting the values of the absolute moments of the standard normal distributions. By (4.4.15) we can see that γ and $\mathcal{X}$ are equal on $(0,\infty)$. Therefore, by the Identity Theorem [43] we conclude that $\gamma(\zeta) = \mathcal{X}(\zeta)$ on $\mathbb{C}$. So we rewrite (4.4.15) as

$$E(\exp(itS)) = \exp\left(-it\zeta - \frac{1}{2}\zeta^2\Delta^2\right) \cdot E\left(\exp\left(itS + \zeta\Delta Z\right)\right), \tag{4.4.16}$$

for all $\zeta \in \mathbb{C}$. By substituting $\zeta = -it\eta$, into (4.4.16) we obtain

$$E(\exp(itS)) = \exp\left(-t^2\eta + \frac{1}{2}t^2\eta^2\Delta^2\right) \cdot E\left(\exp\left(it(S - \eta\Delta Z)\right)\right), \tag{4.4.17}$$

for all $\eta \in \mathbb{R}$. By further substituting $\eta = 1/\Delta^2$ into (4.4.17) we obtain

$$E(\exp(itS)) = \exp\left(-\frac{t^2}{2\Delta^2}\right) \cdot E\left(\exp\left(it\left(S - \frac{Z}{\Delta}\right)\right)\right),$$

for all $t \in \mathbb{R}$. Note the following:

1. $E(\exp(itS))$ is the characteristic function of $\mathcal{L}_{f,\psi}$.
2. $\exp\left(-\frac{t^2}{2\Delta^2}\right)$ is the characteristic function of $\mathcal{N}(0, 1/\Delta^2)$.
3. $E\left(\exp\left(it\left(S - \frac{Z}{\Delta}\right)\right)\right)$ is the characteristic function of another, not further specified, distribution $\mathcal{D}_{f,\psi}$.

This proves the Convolution Theorem.

4.5 Implications of the Convolution Theorem

At this point we present a brief summary of some of our results from sections 2 and 3. We have a data set $\{Y_{ij}\}_{j=1\ldots p}^{i=1\ldots n}$ for the function $f(x)$:

$$Y_{ij} = f\left(\frac{i}{n}\right) + \epsilon_{ij}.$$

Here ϵ_{ij} is an error term with density ψ which is symmetric and has finite Fisher information. Our goal is estimating the Fourier coefficients

$$v_t = \int_0^1 \phi_t(x) f(x)\, dx \qquad t = 1, 2, \ldots, \infty,$$

where $\{\phi_i\}_{i=1...\infty}$ is orthonormal basis on $L^2[0,1]$. We estimate $(v_1,...,v_m)$ by finding an initial estimate $(\hat{v}_1^{(1)},...,\hat{v}_m^{(1)})$ and then recursively applying an iteration:

$$\hat{v}_k^{(s)} = \hat{v}_k^{(s-1)} - \frac{1}{npI(\psi)} \sum_{i=1}^{n} \phi_k\left(\frac{i}{n}\right) \times \sum_{j=1}^{p} \lambda\left(Y_{ij} - \sum_{k=1}^{m} \hat{v}_k^{(s-1)} \phi_k\left(\frac{i}{n}\right)\right) \qquad k = 1,\ldots,m,$$

where the function $\lambda(x)$ is defined as $\lambda(x) = \frac{\psi'(x)}{\psi(x)}$. In subsection 3.1 we have shown the following about the quality of our estimator:

$$\lim_{n\to\infty} \left[npI(\psi)\ E\left((\hat{v}_t^{(s)} - v_t)^2\right)\right] = 1. \tag{4.5.18}$$

According to the Convolution Theorem, for any estimator $\hat{v}_t$ of v_t, the following holds:

$$\lim_{n\to\infty} \left[npI(\psi)\ E\left((\hat{v}_t - v_t)^2\right)\right] \geq 1,$$

which shows that at the limit $n \to \infty$, our estimator for the Fourier Coefficients is in fact the best.

Moreover, the Convolution Theorem implies that as $n \to \infty$ the random variable

$$\text{Err} = \sqrt{npI(\psi)}\ (\hat{v}_t - v_t)$$

can be expressed as a sum

$$\text{Err} = z + \text{Err}_1, \tag{4.5.19}$$

where z has $N(0,1)$ distribution and Err_1 is a random variable with an unspecified distribution. From (4.5.19) we obtain the squared mean of Err_1:

$$E((\text{Err}_1)^2) = E(\text{Err}^2) - E(z^2) = E(\text{Err}^2) - 1. \tag{4.5.20}$$

From (4.5.18) we see that for the improved estimator used throughout this work, $E(\text{Err}^2) = 1$. Hence, from (4.5.20), we conclude that $E((\text{Err}_1)^2) = 0$ and so

$$\text{Err}_1 = 0. \tag{4.5.21}$$

From (4.5.19) and (4.5.21) we conclude that $\sqrt{npI(\psi)}\ (\hat{v}_t^{(s)} - v_t)$ converges in distribution to $N(0,1)$ as $n \to \infty$. Even though the above was separately proved in section 3.3, here we have derived it from the Convolution Theorem.

4.6 A possible extension of the Convolution Theorem.

Extended Convolution Conjecture. *Let $\{Y_{ij}\}_{j=1\ldots p}^{i=1\ldots n}$ be a data set for the function $f(x) \in L^2[0,1]$:*

$$Y_{ij} = f\left(\frac{i}{n}\right) + \epsilon_j^{(i)}.$$

Here $\epsilon_j^{(i)}$ is an error term with density ψ which is symmetric and has finite Fisher information. Let a continuous function $\phi : [0,1] \to \mathbb{R}$ have a norm $||\phi|| = 1$. Let $\theta =< f\phi >$, and let $\hat{\theta}$ be an estimator for θ obtained via the data set $\{Y_{ij}\}_{j=1\ldots p}^{i=1\ldots n}$. Then the error of the estimator obeys the following:

$$npI(\psi)E((\hat{\theta} - \theta)^2) \geq 1. \tag{4.6.22}$$

The Convolution Theorem proved above implies a weaker version of (4.6.22):

$$\lim_{n\to\infty} npI(\psi)E((\hat{\theta} - \theta)^2) \geq 1.$$

We do not have a proof of Extended Convolution Conjecture. We only have an idea of how the conjecture may be proved if it is indeed true. First we have to prove that the average information regarding $\hat{\theta}$ contained in one observation is given by $I(\psi)$. From that and the additivity of the Fisher Information we will conclude that the total information is $pnI(\psi)$. From that we will obtain the following:

$$E((\hat{\theta} - \theta)^2) \geq \frac{1}{npI(\psi)},$$

and the conjecture will be proved.

Chapter 5

Simulations

5.1 Purpose of simulations

Within our work we claim and theoretically show that we can improve an estimator of Fourier coefficients for the problem described in subsection 1.5. Moreover, we claim that the improved estimators have an expected error which obeys the following relation:

$$\lim_{n\to\infty} npI(\psi)E\left((\hat{v}_t^* - v_t)^2\right) = 1,$$

where v_t is the t-th Fourier coefficient of $f(x)$ and $\hat{v}_t^*$ is the iterated estimator of v_t. The main purpose of our simulation is to test our claims numerically. Another item we will show by our simulation is the normal distribution of the error given by

$$\text{Err} = \hat{v}_t^* - v_t.$$

We will show the normality of error in some cases for the following estimators:

1. Simple estimator.

2. Singly iterated estimator.

3. Multiply iterated estimator.

There are many distribution functions $f(x)$ and error functions $\psi(x)$. We will be able to test our hypothesis only in a few cases. In our simulations we will use three distributions. First of all we will use the normal distribution, since most errors in nature are normally distributed.

Then we will use a heavy-tailed distribution. A heavy-tailed distribution $\psi(\epsilon)$ is a distribution for which the expression

$$\int_x^\infty \psi(\epsilon)\, d\epsilon$$

converges to 0 as $x \to \infty$ very slowly. Samples from heavy-tailed distributions are likely to contain some very large values. These distributions present a difficulty for analysis. We must use them to show that our method produces useful results even for the difficult error distributions. Our choice of the heavy tailed distribution will be given by

$$\psi(\epsilon) = \frac{C}{(\epsilon^2+1)^{\frac{b+1}{2}}},$$

where C is the normalization constant. The closer b is to 0, the heavier is the tail.

Finally, we use an almost uniform distribution. The uniform distribution itself does not occur in nature. It has an infinite Fisher Information. Nevertheless, error distributions close to uniform are common. We will use the distribution given by

$$\psi(\epsilon) = \frac{b}{2+2b} \begin{cases} \exp[b(\epsilon+1)] & \epsilon < -1 \\ 1 & -1 < \epsilon < 1 \\ \exp[-b(\epsilon-1)] & \epsilon > 1 \end{cases},$$

where b is the steepness parameter. The larger b is, the closer is our distribution is to uniform in L^2 norm.

5.2 Description of the programs used for simulations

In our simulations we use the following six programs: HEAVYTAIL_SIN.m, HEAVYTAIL_LEG.m, NORMAL_SIN.m, NORMAL_LEG.m,

ALMOSTUNIFORM_SIN.m, and ALMOSTUNIFORM_LEG.m. First, these programs generate the random error values $\{\epsilon_{ij}\}_{i=1..n}^{j=1..p}$ from the distribution ψ given by the parameters we choose inside the input portion of the program. The programs are similar in every respect except for the random error distribution and basis used. HEAVYTAIL_SIN.m and HEAVYTAIL_LEG.m use a heavy-tailed distribution given by

$$\psi(\epsilon) = \frac{C}{(1+\epsilon^2)^{\frac{1+b}{2}}},$$

where C is the normalization coefficient and b is the user-defined tail heaviness parameter. ALMOSTUNIFORM_SIN.m and ALMOSTUNIFORM_LEG.m use the "almost uniform" distribution given by

$$\psi(\epsilon) = \frac{b}{2+2b} \begin{cases} \exp[b(\epsilon+1)] & \epsilon < -1 \\ 1 & -1 < \epsilon < 1 \\ \exp[-b(\epsilon-1)] & \epsilon > 1 \end{cases},$$

where b is the user-defined steepness parameter. NORMAL_SIN.m and NORMAL_LEG.m use N(0,1) distribution.

Then, given either the Fourier coefficients for $f(x)$ or $f(x)$ itself, the programs generate $\{Y_{ij}\}_{i=1..n}^{j=1..p}$:

$$Y_{ij} = f\left(\frac{i}{n}\right) + \epsilon_{ij}$$

With that information the programs determine the Fourier coefficient estimators $\left\{\hat{f}_v\right\}_{v=1..m}$ after one to four iterations. The Fourier coefficients are defined as

$$v_t = \int_0^1 f(x)\phi_t(x)\, dx,$$

where $\{\phi_t\}_{t=1}^{\infty}$ is an orthonormal basis of $L^2[0,1]$. The programs ending in _SIN use sinusoidal basis, while the programs ending in _LEG use the normalized Legendre polynomial basis.

The first estimator is given by

$$\hat{v}_t = \frac{1}{n}\sum_{i=1}^{n}\phi_t(x_i)\ \text{median}\left(\{Y_{ij}\}_{i=1..n}^{j=1..p}\right).$$

The recursive relation between estimators is given by

$$\hat{v}_t^{(s+1)} = v_t{}^{(s)} + \frac{1}{npI(\psi)}\sum_{i=1}^{p}\sum_{j=1}^{n}\lambda(Y_{ij} - f_{[m]}(x_i))\phi_t(x_i)$$

where,

$$\lambda(x) = -\frac{\psi'(x)}{\psi(x)}$$

and the Fisher Information is

$$I = \int_{-\infty}^{\infty}\frac{(\psi'(x))^2}{\psi(x)}\, dx.$$

The programs calculate it's overall error for each estimator:

$$E_n = \sqrt{\sum_{t=1}^{m}(\hat{v}_t - v_t)^2}$$

The output errors are divided by the minimum expected error:

$$E_{\text{Exp}} = \sqrt{\frac{m}{pnI}},$$

where I is the Fisher information. The programs perform a given number of runs of such simulations on the same function and calculate the mean square average error of the estimators over N runs. They display the errors averaged over a given number of runs in the tabular format.

5.3 Choice of a basis

In our simulation we will need to use an orthonormal basis for the space of $L^2[0,1]$. Even though there are uncountably many such bases, we will use only a few of them:

1. Sinusoidal basis.

   The basis is given by $\{\psi_\nu\} = \{\sin(\pi\nu x)\}$ for $\nu = 1, ..., \infty$.

2. Normalized Legendre Polynomials.

   The Legendre Polynomials are given by the formula:

$$P_\nu^*(x) = \frac{d^\nu}{dy^\nu}\left[\left(y^2-1\right)^n\right]\Big|_{y=2x-1}.$$

   The Normalized Legendre Polynomials are given by the formula:

$$\psi_\nu(x) = \frac{P_\nu^*(x)}{\int_0^1 P_\nu^{*2}(y)dy}$$

   for $\nu = 0, ..., \infty$.[44]

These bases were chosen since they are the ones most commonly used in mathematics. The sinusoidal wave basis is commonly used for determining Fourier coefficients in diverse mathematical problems. This is how complex Fourier coefficients are determined in [25],[45], and [46].

The Legendre basis approximates a function by a polynomial. The polynomial approximations of functions are also commonly used [47] and [48].

There are uncountably many other orthonormal bases of $L^2[0,1]$ which we will not use in our simulations. Any non-orthonormal basis can be transformed into an orthonormal one via Gram – Schmidt process [49].

5.4 Examples of simulations

Example I.

In this example we will estimate the Fourier coefficients of the function

$$f(x) = \frac{1}{1 + (x - 1/2)^2}.$$

We will use the basis $\{\sqrt{2}\sin(\pi v x)\}_{v=1}^{\infty}$ to find the Fourier coefficients of $f(x)$. In our simulation p=21 and n=50. We use the normal error distribution $N(0,1)$. We estimate m=10 terms of our series. We do simulations over 1,000 runs.

The Expected error is $E_{\text{Exp}} = .0976$. The ratio of mean squared error and expected error after v iterations is denoted Eratio_v. It is given in the table below:

Eratio_0	Eratio_1	Eratio_2	Eratio_3	Eratio_4
1.1984	0.9916	0.9727	0.9714	0.9713

Table 5.1: Error ratios for estimating $f(x) = \left(1 + (x - 1/2)^2\right)^{-1/2}$ from data effected by normal error with respect to sinusoidal basis.

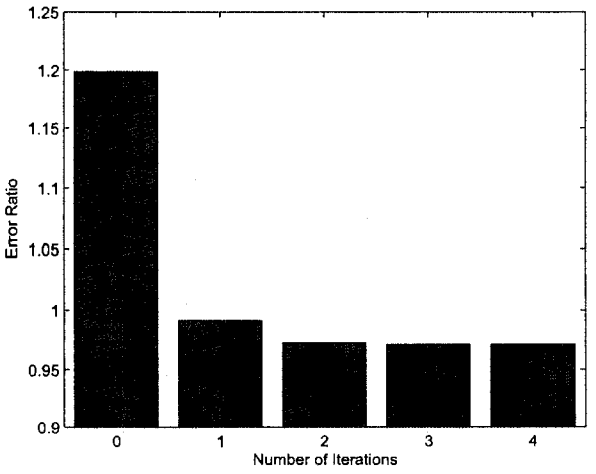


Figure 5.1: Error ratios for estimating $f(x) = \left(1 + (x - 1/2)^2\right)^{-1/2}$ from data effected by normal error with respect to sinusoidal basis.

Normal error distribution produces a reasonably good initial estimate – only 20% above the minimum. The very first iteration brings the error to the expected value.

Below we present the histograms of error distribution showing that these distributions are approximately normal:

1. Uniterated simulation error histogram:

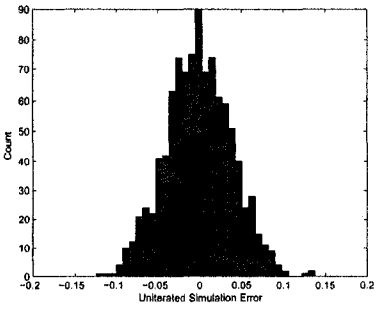


Figure 5.2: Uniterated simulation error histogram estimating $f(x) = \left(1 + (x - 1/2)^2\right)^{-1/2}$ from data effected by normal error with respect to sinusoidal basis.

2. Singly iterated simulation error histogram:

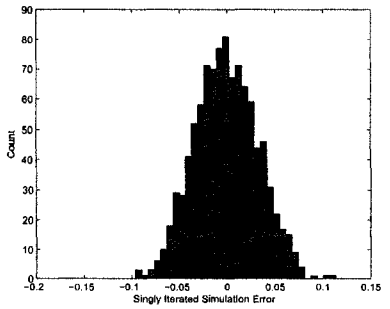


Figure 5.3: Singly iterated simulation error histogram estimating $f(x) = \left(1 + (x - 1/2)^2\right)^{-1/2}$ from data effected by normal error with respect to sinusoidal basis.

3. Multiply iterated simulation error histogram:

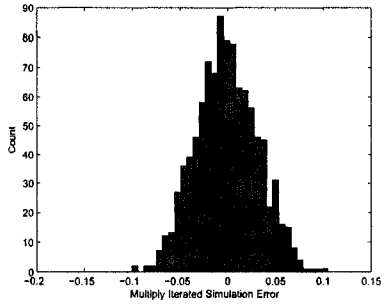


Figure 5.4: Multiply iterated simulation error histogram estimating $f(x) = \left(1 + (x - 1/2)^2\right)^{-1/2}$ from data effected by normal error with respect to sinusoidal basis.

Example II.

In this example we will estimate the Fourier coefficients of the function

$$g(x) = \frac{\sin(x)}{x} \quad \text{for } x \in [0, \pi].$$

The function g is similar to

$$f(x) = \frac{\sin(\pi x)}{\pi x} \quad \text{for } x \in [0, 1].$$

We will use the basis $\{\sqrt{2}\sin(\pi v x)\}_{v=1}^{\infty}$ to find the Fourier coefficients of $f(x)$. In our simulation p=11 and n=20. We use the normal error distribution $N(0, 1)$. We estimate m=10 terms of our series. We do simulations over 2,000 runs.

The Expected error is $E_{\text{Exp}} = .2132$ The ratio of mean squared error and expected error after v iterations is denoted Eratio_v. It is given in the table below:

Eratio_0	Eratio_1	Eratio_2	Eratio_3	Eratio_4
1.1994	0.9987	0.9803	0.9790	0.9789

Table 5.2: Error ratios for estimating $f(x) = \sin(x)/x$ from data effected by normal error with respect to sinusoidal basis.

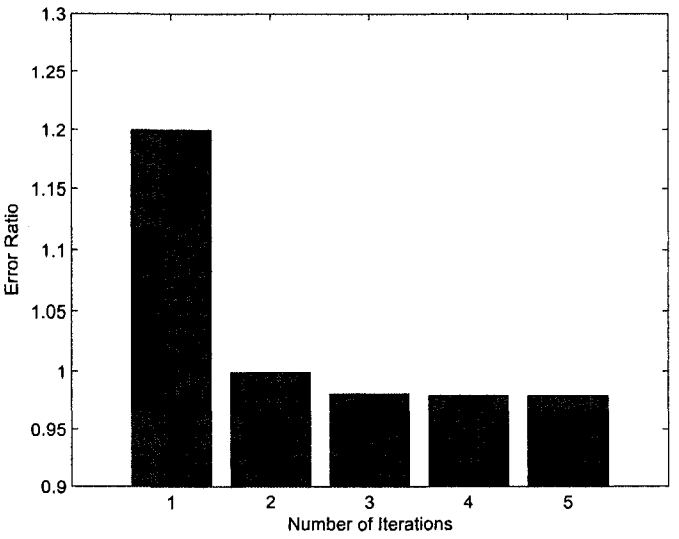


Figure 5.5: Error ratios for estimating $f(x) = \sin(x)/x$ from data effected by normal error with respect to sinusoidal basis.

As in the previous example, a reasonably good initial estimate results -- only 20% above the minimum. The very first iteration brings the error to the expected value.

Example III.

In this example we will estimate the Fourier coefficients of the function

$$f(x) = x^x.$$

We will use the basis $\{\sqrt{2}\sin(\pi v x)\}_{v=1}^{\infty}$ to find the Fourier coefficients of $f(x)$. In our simulation p=21 and n=40. We use different error distributions in parts 1) and 2). We estimate m=10 terms of our series. We do simulations over 1,000 runs.

1. The error function in this example is a heavy tailed function given by

$$\psi(\epsilon) = \frac{C}{(\epsilon^2+1)^{9/14}},$$

where C is the normalization constant.

The Expected error is $E_{\text{Exp}} = .3263$ The ratio of mean squared error and expected error after v iterations is denoted Eratio_v. It is given in the table below:

Eratio_0	Eratio_1	Eratio_2	Eratio_3	Eratio_4
2.5446	1.7705	1.4186	1.2506	1.1641

Table 5.3: Error ratios for estimating $f(x) = x^x$ from data effected by error with distribution $C * (x^2 + 1)^{-9/14}$ with respect to sinusoidal basis.

The initial estimate has a relatively large error, 2.5 times the theoretical minimum. Four iterations bring the error down to only 1.16 times the theoretical minimum.

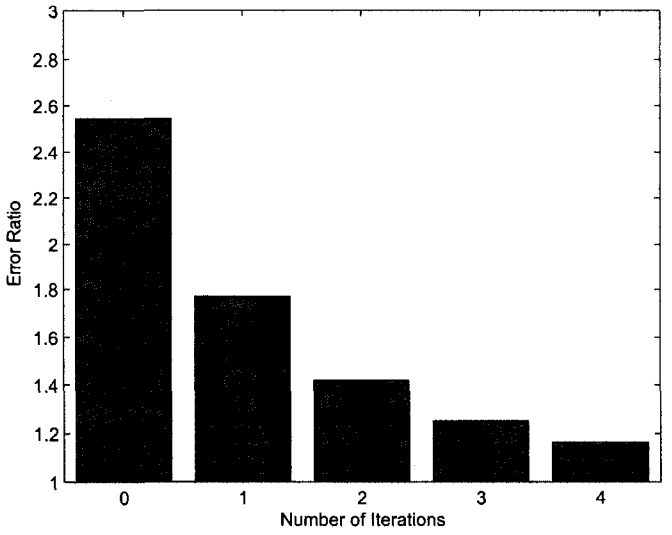


Figure 5.6: Error ratios for estimating $f(x) = x^x$ from data effected by error with distribution $C * (x^2 + 1)^{-9/14}$ with respect to sinusoidal basis.

Below, we present the histograms of error distribution showing that these distributions are approximately normal:

(a) Uniterated simulation error histogram:

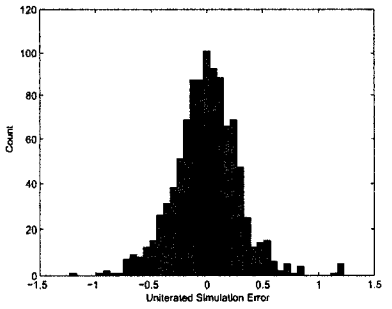


Figure 5.7: Uniterated simulation error histogram for estimating $f(x) = x^x$ from data effected by error with distribution $C * (x^2 + 1)^{-9/14}$ with respect to sinusoidal basis.

(b) Singly iterated simulation error histogram:

(c) Multiply iterated simulation error histogram:

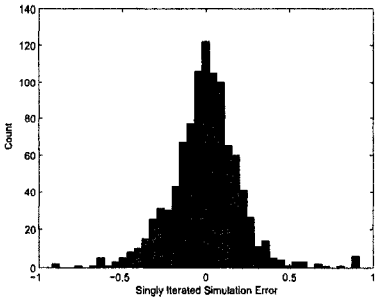


Figure 5.8: Singly iterated simulation error histogram for estimating $f(x) = x^x$ from data effected by error with distribution $C * (x^2 + 1)^{-9/14}$ with respect to sinusoidal basis.

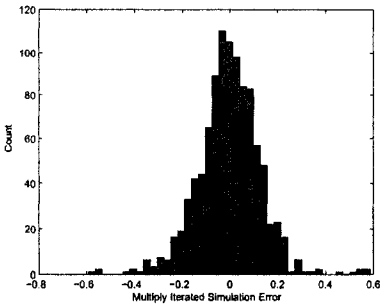


Figure 5.9: Multiply iterated simulation error histogram for estimating $f(x) = x^x$ from data effected by error with distribution $C * (x^2 + 1)^{-9/14}$ with respect to sinusoidal basis.

2. The error function in this example is a heavy tailed function given by

$$\psi(\epsilon) = \frac{b}{2+2b} \begin{cases} \exp[b(\epsilon+1)] & \epsilon < -1 \\ 1 & -1 < \epsilon < 1 \\ \exp[-b(\epsilon-1)] & \epsilon > 1 \end{cases},$$

where b is the steepness constant. In our case b=10.

The Expected error is $E_{\text{Exp}} = .0362$. The ratio of mean squared error and expected error after v iterations is denoted Eratio_v. It is given in the table below:

Eratio_0	Eratio_1	Eratio_2	Eratio_3	Eratio_4
3.0945	1.9911	1.5583	1.4272	1.3886

Table 5.4: Error ratios for estimating $f(x) = x^x$ from data effected by error with exponentially tailed distribution with respect to sinusoidal basis.

The initial estimate has a relatively large error 3.1 times the theoretical minimum. Four iterations bring the error down to 1.4 times the theoretical minimum – still a large error, but obviously an improvement.

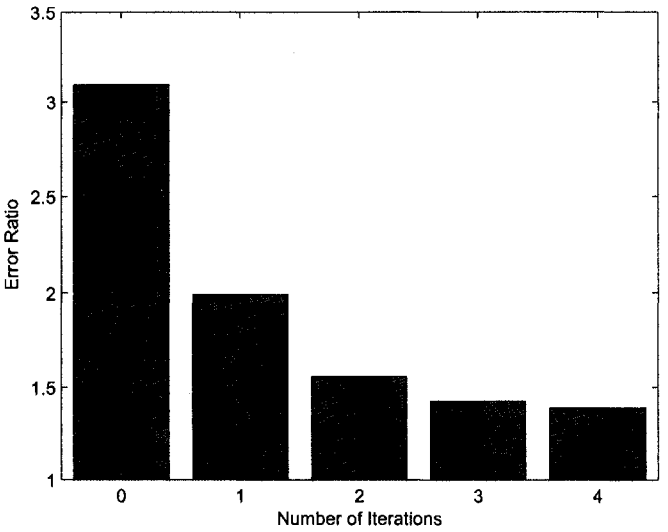


Figure 5.10: Error ratios for estimating $f(x) = x^x$ from data effected by error with exponentially tailed distribution with respect to sinusoidal basis.

Below, we present the histograms of error distribution showing that these distributions are approximately normal:

(a) Uniterated simulation error histogram:

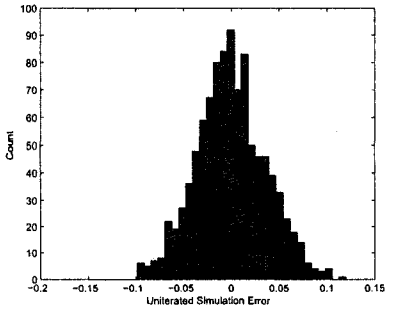


Figure 5.11: Uniterated simulation error histogram for estimating $f(x) = x^x$ from data effected by error with exponentially tailed distribution with respect to sinusoidal basis.

(b) Singly iterated simulation error histogram:

(c) Multiply iterated simulation error histogram:

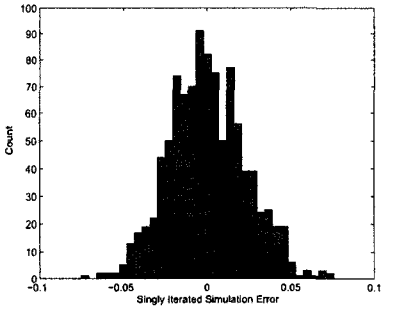


Figure 5.12: Singly iterated simulation error histogram for estimating $f(x) = x^x$ from data effected by error with exponentially tailed distribution with respect to sinusoidal basis.

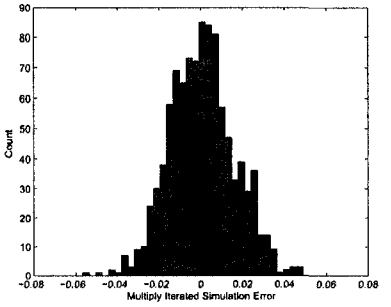


Figure 5.13: Multiply iterated simulation error histogram for estimating $f(x) = x^x$ from data effected by error with exponentially tailed distribution with respect to sinusoidal basis.

Example IV.

In this example we will estimate the Fourier coefficients of the function

$$f(x) = \sin(50x^2).$$

Unlike other relatively smooth functions we have used, this is a rapidly oscillating function. We will use the basis $\{\sqrt{2}\sin(\pi v x)\}_{v=1}^{\infty}$ to find the Fourier coefficients of $f(x)$. In our simulation p=21 and n=40. We use different error distributions in parts 1) and 2). We estimate m=10 terms of our series. We do simulations over 1,000 runs.

1. The error function in this example is a heavy tailed function given by

$$\psi(\epsilon) = \frac{C}{(\epsilon^2+1)^{9/14}},$$

where C is the normalization constant.

The Expected error is $E_{\text{Exp}} = .3263$. The ratio of mean squared error and expected error after v iterations is denoted Eratio_v. It is given in the table below:

Eratio_0	Eratio_1	Eratio_2	Eratio_3	Eratio_4
2.5448	1.8688	1.5620	1.4153	1.3395

Table 5.5: Error ratios for estimating $f(x) = \sin(50 * x^2)$ from data effected by error with distribution $C * (x^2 + 1)^{-9/14}$ with respect to sinusoidal basis.

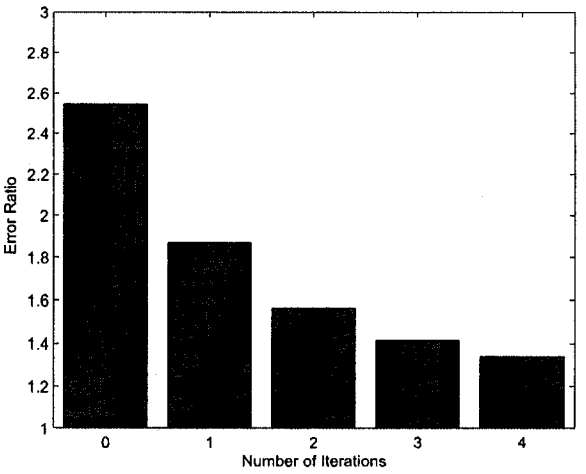


Figure 5.14: Error ratios for estimating $f(x) = \sin(50 * x^2)$ from data effected by error with distribution $C * (x^2 + 1)^{-9/14}$ with respect to sinusoidal basis.

The initial estimate has a relatively large error, 2.5 times the theoretical minimum. Four iterations bring the error down to 1.33 times the theoretical minimum – a considerable improvement.

2. The error function in this example is a heavy tailed function given by

$$\psi(\epsilon) = \frac{b}{2+2b} \begin{cases} \exp[b(\epsilon+1)] & \epsilon < -1 \\ 1 & -1 < \epsilon < 1 \\ \exp[-b(\epsilon-1)] & \epsilon > 1 \end{cases},$$

where b is the steepness constant. In our case b=10.

The Expected error is $E_{\text{Exp}} = .0362$. The ratio of mean squared error and expected error after v iterations is denoted Eratio_v. It is given in the table below:

Eratio_0	Eratio_1	Eratio_2	Eratio_3	Eratio_4
3.0922	2.6318	2.5007	2.4788	2.4843

Table 5.6: Error ratios for estimating $f(x) = \sin(50 * x^2)$ from data effected by error with exponentially-tailed distribution with respect to sinusoidal basis.

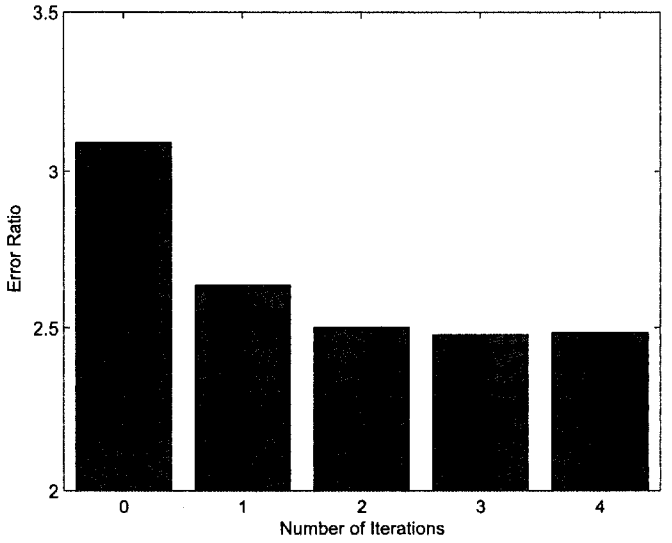


Figure 5.15: Error ratios for estimating $f(x) = \sin(50 * x^2)$ from data effected by error with exponentially-tailed distribution with respect to sinusoidal basis.

The initial estimate has a relatively large error, 3.1 times the theoretical minimum. Four iterations bring the error down to 2.5 times the theoretical minimum. The improvement in this case is insignificant.

Example V.

In this example we will estimate the Fourier coefficients of the function

$$f(x) = |\sin(20x^2)|.$$

Unlike other smooth and differentiable functions we have used, this function is not differentiable at several points. We will use the basis $\{\sqrt{2}\sin(\pi v x)\}_{v=1}^{\infty}$ to find the Fourier coefficients of $f(x)$. In our simulation p=21 and n=40. We use different error distributions in parts 1) and 2). We estimate m=10 terms of our series. We do simulations over 1,000 runs.

1. The error function in this example is a heavy tailed function given by

$$\psi(\epsilon) = \frac{C}{(\epsilon^2+1)^{9/14}},$$

where C is the normalization constant.

The Expected error is $E_{\text{Exp}} = .3263$. The ratio of mean squared error and expected error after v iterations is denoted Eratio_v. It is given in the table below:

Eratio_0	Eratio_1	Eratio_2	Eratio_3	Eratio_4
2.5450	1.7831	1.4378	1.2715	1.1849

Table 5.7: Error ratios for estimating $f(x) = |\sin(20 * x^2)|$ from data effected by error with distribution $C * (x^2+1)^{-9/14}$ with respect to sinusoidal basis.

The initial estimate has a relatively large error, 2.5 times the theoretical minimum. Four iterations bring the error down to 1.2 times the theoretical minimum – a considerable improvement.

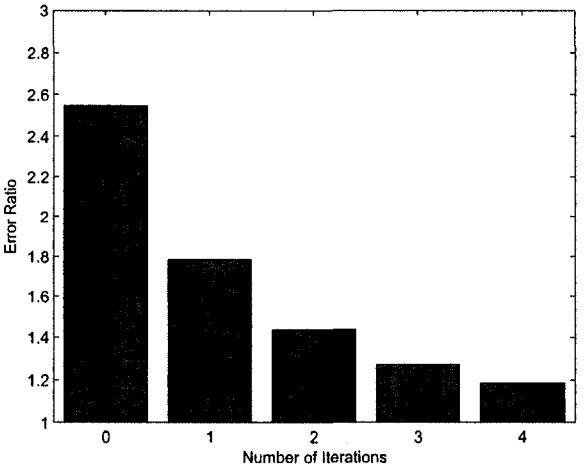


Figure 5.16: Error ratios for estimating $f(x) = |\sin(20 * x^2)|$ from data effected by error with distribution $C * (x^2 + 1)^{-9/14}$ with respect to sinusoidal basis.

Below, present the histograms of error distribution showing that these distributions are approximately normal:

(a) Uniterated simulation error histogram:

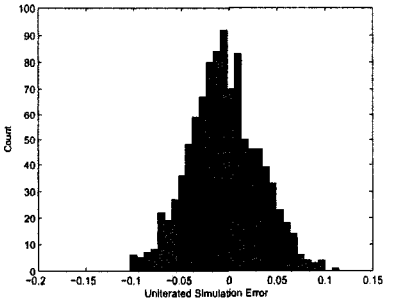


Figure 5.17: Uniterated simulation error histogram for estimating $f(x) = |\sin(20*x^2)|$ from data effected by error with distribution $C*(x^2+1)^{-9/14}$ with respect to sinusoidal basis.

(b) Singly iterated simulation error histogram:

(c) Multiply iterated simulation error histogram:

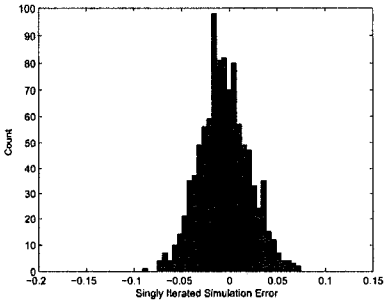


Figure 5.18: Singly iterated simulation error histogram for estimating $f(x) = |\sin(20 * x^2)|$ from data effected by error with distribution $C * (x^2 + 1)^{-9/14}$ with respect to sinusoidal basis.

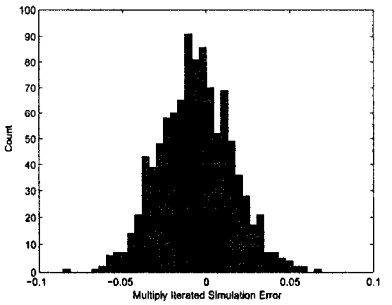


Figure 5.19: Multiply iterated simulation error histogram for estimating $f(x) = |\sin(20 * x^2)|$ from data effected by error with distribution $C * (x^2 + 1)^{-9/14}$ with respect to sinusoidal basis.

2. The error function in this example is a heavy tailed function given by

$$\psi(\epsilon) = \frac{b}{2+2b} \begin{cases} \exp[b(\epsilon+1)] & \epsilon < -1 \\ 1 & -1 < \epsilon < 1 \\ \exp[-b(\epsilon-1)] & \epsilon > 1 \end{cases},$$

where b is the steepness constant. In our case b=10.

The Expected error is $E_{\text{Exp}} = .0362$. The ratio of mean squared error and expected error after v iterations is denoted Eratio_v. It is given in the table below:

Eratio_0	Eratio_1	Eratio_2	Eratio_3	Eratio_4
3.0995	2.2100	1.9059	1.8193	1.8007

Table 5.8: Error ratio for estimating $f(x) = |\sin(20 * x^2)|$ from data effected by error with exponential tail distribution with respect to sinusoidal basis.

The initial estimate has a relatively large error, 3.1 times the theoretical minimum. Four iterations bring the error down to 1.8 times the theoretical minimum. That is a small improvement.

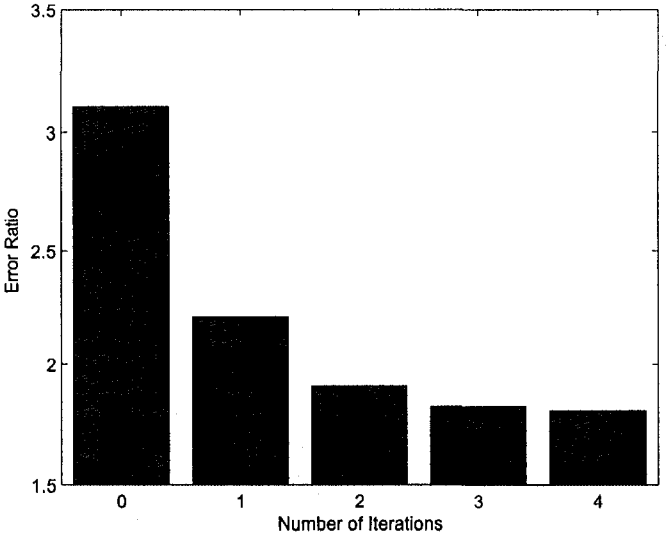


Figure 5.20: Error ratio for estimating $f(x) = |\sin(20 * x^2)|$ from data effected by error with exponential tail distribution with respect to sinusoidal basis.

Example VI.

In this example we will estimate the Fourier coefficients of the function

$$f(x) = \sin(\pi x).$$

Unlike other simulations in which we have used the sinusoidal basis, in this example we use the Legendre polynomial basis. We will use the basis

$$\left\{ \frac{P_v(2x-1)}{\int_0^1 P_v^2(2x-1)\, dx} \right\}_{v=1}^{\infty},$$

where P_v is the v-th Legendre polynomial to find the Fourier coefficients of $f(x)$.

All the functions of the Legendre basis except the first one are oscillatory functions. Nevertheless, unlike the sinusoidal function, the amplitude of the oscillation of the basis functions increases toward the edges. This phenomenon causes the maximum likelihood estimator of the Fourier coefficients of $f(x)$ to be further from the true value. In order for the maximum likelihood estimator $\left\{\hat{f}_v\right\}_{v=1..m}$ to be close to the actual value $\{f_v\}_{v=1..m}$ we should choose a large value of n. When we chose n=20, we got an unreasonably large error.

In our simulation with Legendre basis, p=21 and n=200. We use different error distributions in parts 1) and 2). We estimate m=10 terms of our series. We do simulations over 500 runs.

1. The error function in this example is a heavy tailed function given by

$$\psi(\epsilon) = \frac{C}{(\epsilon^2 + 1)^{9/14}},$$

where C is the normalization constant.

The Expected error is $E_{\text{Exp}} = .1459$. The ratio of mean squared error and expected error after v iterations is denoted Eratio_v. It is given in the table below:

Eratio_0	Eratio_1	Eratio_2	Eratio_3	Eratio_4
2.8096	1.3468	1.1553	1.1080	1.0947

Table 5.9: Error ratios for estimating $f(x) = \sin(\pi x)$ from data effected by error with distribution $C * (x^2 + 1)^{-9/14}$ with respect to Legendre basis.

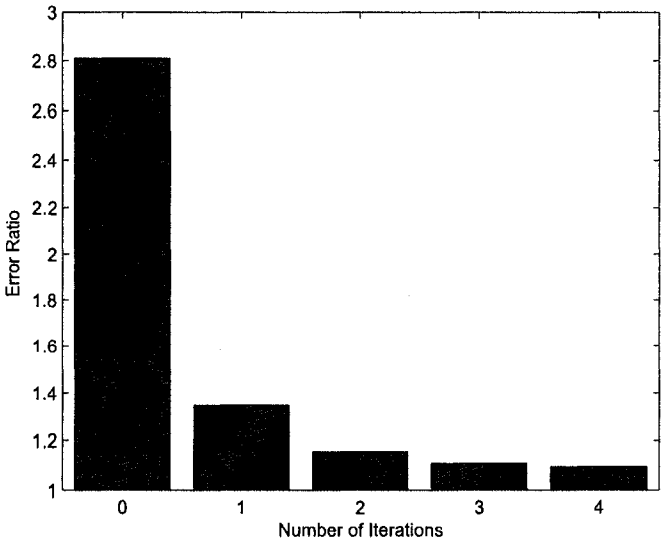


Figure 5.21: Error ratios for estimating $f(x) = \sin(\pi x)$ from data effected by error with distribution $C * (x^2 + 1)^{-9/14}$ with respect to Legendre basis.

The initial estimate has a relatively large error, 2.8 times the theoretical minimum. The first iteration brings the mean squared error to 1.3. Four iterations bring the error down to 1.1 times the theoretical minimum.

2. The error function in this example is $N(0, 1)$.

The Expected error is $E_{\text{Exp}} = .0362$. The ratio of mean squared error and expected error after v iterations is denoted Eratio_v. It is given in the table below:

Eratio_0	Eratio_1	Eratio_2	Eratio_3	Eratio_4
1.1892	1.0130	1.0253	1.0276	1.0285

Table 5.10: Error ratios for estimating $f(x) = \sin(\pi x)$ from data effected by normally distributed error with respect to Legendre basis.

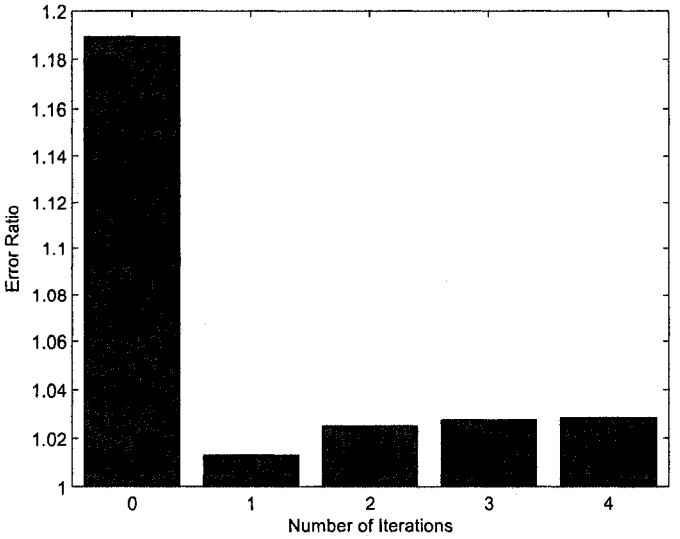


Figure 5.22: Error ratios for estimating $f(x) = \sin(\pi x)$ from data effected by normally distributed error with respect to Legendre basis.

We have reasonably good initial estimate – only 20% above the minimum. The very first iteration brings the error to the expected value.

Below, we present the histograms of error distribution showing that these distributions are approximately normal:

(a) Uniterated simulation error histogram:

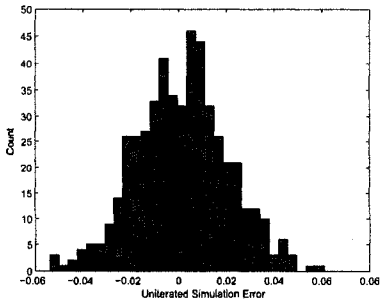


Figure 5.23: Uniterated simulation error histogram for estimating $f(x) = \sin(\pi x)$ from data effected by normally distributed error with respect to Legendre basis.

(b) Singly iterated simulation error histogram:

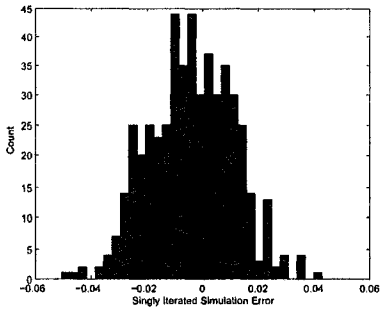


Figure 5.24: Singly iterated simulation error histogram for estimating $f(x) = \sin(\pi x)$ from data effected by normally distributed error with respect to Legendre basis.

(c) Multiply iterated simulation error histogram:

5.5 Results derived from Simulations

From the simulations we notice that for different bases and different functions $f(x)$ we see the same pattern. First of all, the iterations do reduce the expected error. In some cases the error reduction can be quite significant – sometimes the error is reduced 2.5 times. That result fits very well with our theoretical prediction that the iteration (2.4.16) will improve our estimator.

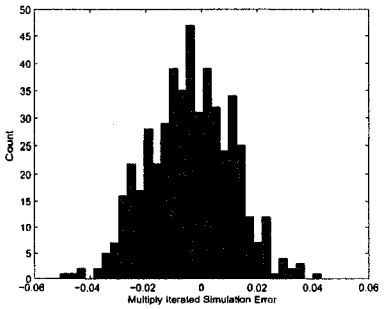


Figure 5.25: Multiply iterated simulation error histogram for estimating $f(x) = \sin(\pi x)$ from data effected by normally distributed error with respect to Legendre basis.

Second, the mean squared error given by

$$E = \sqrt{\text{Exp}\left[\sum_{t=1}^{m} (\hat{v}_t - v_t)^2\right]}$$

can almost never be reduced below the "expected error" defined by

$$E_{\text{Exp}} = \sqrt{\frac{m}{pnI}}.$$

It has been shown theoretically in **Subsection 2.3** that $E \to E_{\text{Exp}}$ as $n \to \infty$. Now we have seen the same result via the simulations.

Finally, we have seen that the errors $\hat{v}_t - v_t$ have distributions close to Normal distribution, especially in case n is large. That agrees with our theoretical prediction from **Subsection 2.5** that as $n \to \infty$, the distribution of $\hat{v}_t - v_t$ approaches the Normal distribution.

Chapter 6

Conclusion

6.1 Summary of our work

We have considered the problem of reconstruction of a one-dimensional function $f : [0,1] \rightarrow \mathbb{R}$ from observations affected by a random error of known distribution. We have derived an algorithm for estimating the Fourier coefficients of $f(x)$. The estimators are denoted by $\{\hat{v}_k\}_{k=1}^m$. That algorithm includes an initial rough estimation of the coefficients and an iteration formula which is applied to $\{\hat{v}_k\}_{k=1}^m$ recursively in order to produce a better result.

Then we have calculated the expected error for the coefficients obtained by our algorithm. We have shown that with every iteration of $\{\hat{v}_k\}_{k=1}^m$, the expected error decreases. We have also shown that at least for $n \rightarrow \infty$ the expected error obtained by our estimation is the theoretical minimum for an error obtained by any estimation.

Finally, we have verified our theoretical results via simulations. Given that we have used different sample sizes, different functions $f(x)$, different error functions and different bases, our simulations are relatively comprehensive.

6.2 A generalization of the main problem and a possible solution

6.2.1 The generalized problem

The problem we have worked on in this dissertation can be generalized in several ways. First we can introduce random observations. Our observations on $[0,1]$ in the original were taken on n points equidistant from each other, so we can take observations on any n points on $[0,1]$. Second, we can expand our problem

from being one dimensional to being D dimensional. Thus, the n observation points will be located in the subset $[0,1]^D \subset \mathbb{R}^D$, and the function we wish to estimate will be a continuous function $f : [0,1]^D \to \mathbb{R}$. Third, in our original problem we have considered a similar number of observations at each point, but in the generalized problem the number of observations may vary. Now, we can state the generalized version of the problem we have worked on in this dissertation.

We have n observation points $\{\mathbf{x}_i\}_{i=1}^n \in [0,1]^D$. At each observation point $\mathbf{x}_i$ there are p_i observations given by:

$$Y_{ij} = f(\mathbf{x}_i) + \epsilon_j^{(i)} \qquad j = 1, ..., p_i. \tag{6.2.1}$$

In (6.2.1), $\epsilon_j^{(i)}$ is an error term which has density ψ. The error distribution function ψ is symmetric and has the finite Fisher information. Moreover, the error function obeys the following assumptions:

$$\left[\frac{\psi'(\epsilon)}{\psi(\epsilon)}\right]' \text{ is uniformly continuous on } \mathbb{R}.$$

$$\int_{-\infty}^{\infty} \left|\frac{\psi'(\epsilon)}{\psi(\epsilon)}\right|^{2+\alpha} \psi(\epsilon)\, d\epsilon < \infty \text{ for some } \alpha > 0.$$

The function $f(x)$ mentioned in (6.2.1) is a finite and continuous function on $[0,1]^D$.

Let $\{\phi_i\}_{i=1...\infty}$ be an orthonormal basis on $L^2\left([0,1]^D\right)$. Then the function $f(x)$ would be determined by the series of its Fourier coefficients:

$$v_t = \int \cdots \int_{[0,1]^D} \phi_t(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \qquad t = 1, 2, ..., \infty.$$

Our goal is to estimate v_t for $t = 1, 2, ..., m$ using the observed data $\{Y_{ij}\}$.

6.2.2 A possible solution

By analogy with the one-dimensional case we have considered, we can propose formulas for simple and improved estimators. A simple estimator for the k-th Fourier coefficient of $f(\mathbf{x})$ is given by

$$\hat{v}_k^{(1)} = \frac{1}{\sum_{i=1}^n p_i} \sum_{i=1}^n p_i * \text{median}\left(\{Y_{ij}\}_{j=1...p_i}\right) \phi_k(\mathbf{x}_i), \quad k = 1, 2, ..., m. \tag{6.2.2}$$

The estimator should be improved by repetitive application of the following iteration:

$$\begin{aligned} v_k^{(s)} &\approx v_k^{(s-1)} - \frac{1}{I(\psi) \sum_{i=1}^n p_i} \sum_{i=1}^n \phi_k(\mathbf{x}_i) \\ &\sum_{j=1}^{p_i} \lambda\left(Y_{ij} - \sum_{k=1}^m v_k^{(s-1)} \phi_k(\mathbf{x}_i)\right), \quad k = 1, 2, ..., m, \end{aligned} \tag{6.2.3}$$

where the function $\lambda(\epsilon)$ is defined as $\lambda(\epsilon) = \frac{\psi'(\epsilon)}{\psi(\epsilon)}$. We must keep in mind, however, that showing that (6.2.2) and (6.2.3) are the best expressions for obtaining estimators for Fourier coefficients of $f(\mathbf{x})$ remains an open problem.

Intuitively, we can expect that under certain conditions the following relation will hold:

$$\lim_{n\to\infty}\left[\sum_{i=1}^{n} p_i\right] I(\psi)E\left((\hat{v}_k - v_k)^2\right) = 1. \tag{6.2.4}$$

Nevertheless, determining under which conditions (6.2.4) holds and proving or disproving that it does hold under any conditions remains an open problem.

Finally, we will have to determine whether (6.2.4) is indeed a characteristic of the best result or whether there can be an estimator for which

$$\lim_{n\to\infty}\left[\sum_{i=1}^{n} p_i\right] I(\psi)E\left((\hat{v}_k - v_k)^2\right) < 1.$$

APPENDICES

The program HEAVYTAIL_SIN.m

```
% ANYSIMNORM.m
% THIS PROGRAM GENERATES A RANDOM ERROR FUNCTION
% WITH DISTRIBUTION GIVEN BY phi.
% THEN THE PROGRAM ANALYSES AVERAGE ERROR FOR
% f_v^hat, f_v^2hat, f_v^3hat, f_v^4hat
% BUT THIS PROGRAM DISPLAYS RESULTS DIVIDED BY
% THE MINIMUM EXPECTED VARIANCE

tic
% WITHIN THIS PART OF THE PROGRAM WE WILL GENERATE
% RANDOM NUMBERS WITH DISTRIBUTION phi
rand('state',0);

    %%          INPUT DATA BELOW            %%
    %%          INPUT DATA BELOW            %%
    %%      FOR RANDOM VARIABLE GENERATOR     %%
    b=2/7;                              %  *INPUT* b
    D=.01;                              %  *INPUT* D FOR THE INTEGRAL
    phi=inline('(y.^2+1).^(-1/2-b/2)','y','b');%  *INPUT* FUNCTION f(y,t)
    f=inline('I*exp(-t)/(y.^2+1).^(-1/2-b/2)','t','y','b','I');
                                            %  *INPUT* dy/dt
    N=2000;                       %  *INPUT* NUMBER OF STEPS FOR ADAMS
    Nrand=10000;                  %  *INPUT* NUMBER OF RANDOM NUMBERS

    %%      FOR THE MAIN SIMULATION         %%
    LAMBDA=inline('(1+b)*y./(1+y.^2)','y','b');
              %  *INPUT* LAMBDA=-phi'(x)/phi(x)

%%      END OF INPUT        %%

y0=0;                                       %  y VARIABLE AT t=ln(2)
a=log(2);                                   %  INITIAL t
b1=log(2)+12;                               %  t=INFINITY

%%  INTEGRATING phi;  OBTAINING FISHER INFORMATION      %%
L=10+1/b;                                   %  LIMITS OF INTEGRATING phi
t1=-L:D:L;                                  %  INTEGRAND
Iphi=sum(phi(t1.*exp(t1.^2/2),b).*(t1.^2+1).*exp(t1.^2/2))*D
                                            %  INTEGRAL OF phi
IF=1/Iphi*sum(LAMBDA(t1.*exp(t1.^2/2),b).^2.*phi(t1.*exp(t1.^2/2),b)
.*(t1.^2+1).*exp(t1.^2/2))*D                 % FISHER INFORMATION
%%  INTEGRATING phi;  OBTAINING FISHER INFORMATION      %%

t=a:((b1-a)/N):b1;                          %  TIME
h=(b1-a)/N;                                 %  STEPSIZE
S=size(y0);                                 %  SIZE OF Y VECTOR
S=S(2);
W=zeros(N+1,S);                             % APPROXIMATION OF Y
%% t(i)=a+(i-1)*(b1-a)/N i=1...N+1
%% W(i)~=y(t(i))

%%  RUNGA-KUTTA
W(1)=y0;                            % W(1) (WO BY OLD COUNT)

for i=1:3                           % W(2)-W(4) (W(1)-W(3) BY O.C.)
    K1=h*f(t(i),W(i),b,Iphi);
    K2=h*f(t(i)+h/2,W(i)+K1/2,b,Iphi);
    K3=h*f(t(i)+h/2,W(i)+K2/2,b,Iphi);
    K4=h*f(t(i)+h,W(i)+K3,b,Iphi);
    W(i+1)=W(i)+(K1+2*K2+2*K3+K4)/6;
```

```
end

%%  PREDICTOR-CORRECTOR

for i=4:N                          % W(4)-W(N+1) (W(3)-W(N) BY O.C.)
    WP=W(i)+h/24*(55*f(t(i),W(i),b,Iphi)-59*f(t(i-1),W(i-1),b,Iphi)
            +37*f(t(i-2),W(i-2),b,Iphi)-9*f(t(i-3),W(i-3),b,Iphi));
    W(i+1)=W(i)+h/24*(9*f(t(i+1),WP,b,Iphi)+19*f(t(i),W(i),b,Iphi)
             -5*f(t(i-1),W(i-1),b,Iphi)+1*f(t(i-2),W(i-2),b,Iphi));
end
toc


% SIMULATION PROGRAM
tic
RATIOS=zeros(1:5,1:4);                  % ERROR RATIOS
ERRORS=RATIOS;                          % UNCERTAINTY IN RATIOS
RUNS=1000;                              % INPUT THE NUMBER OF RUNS

Fvt5s=[1*RUNS]*0;                       % Fvt-Fv
Fvtt5s=[1*RUNS]*0;                      % Fvtt-Fv

disp('    RUN       ExpErr    Error1    Error2    Error3
                         Error4    Error5    Time(min)')
allp=[21];
alln=[40];
for pI=1:sum(allp.^0)
for nI=1:sum(alln.^0)

Error1=(1:RUNS)*0;Error2=(1:RUNS)*0;Error3=(1:RUNS)*0;Error4=
                               (1:RUNS)*0;Error5=(1:RUNS)*0;

for RUN=1:RUNS
p=allp(pI);
n=alln(nI);
TNS=300;        % NUMBER OF TERMS IN THE SERIES DEFINING f(x)
TN=10;          % NUMBER OF TERMS CONSIDERED

M=zeros(n+1,1);
X=zeros(n+1,p);
x=(0:n)/n;                              %  VARIABLE
Y=x*0;                                  %  APPROXIMATION
F=Y;                                    %  FUNCTION

S=sign(rand(n+1,p)-1/2);  % SIGN OF RANDOM NUMBERS
WHY=(rand(n+1,p)+1)/2;    % RANDOM CUMULUTIVE FUNCTION
T=-log(1-WHY);            % RANDOM T
E=spline(t,W,T).*S;       % THE SEQUENCE OF RANDOM NUMBERS FROM phi
X=zeros(n+1,p);           % RANDOM VARIABLE
Expvar=sqrt(TN)/sqrt(p*n*IF); % MINIMUM EXPECTED VARIANCE


D=.0005;
xfunct=0:D:1+1e-8;       % TO CALCULATE FOURIER COEFFICIENTS
funct=abs(sin(20*xfunct.^2));
for i=1:TN
    Fv(i)=sum(funct*2^.5.*sin(xfunct*pi*i))*D;
                              %% COEFFECIENTS ARE DEFINED HERE %%
    F=abs(sin(20*(x+1e-8).^2));              % SERIES SUM
end

%%  CONTROL FULNCTION
%Fv=(1:TN)*0;
```

```
%for i=1:TNS
%   Fv(i)=2^2.5*pi^(-3)/i^3;   %% COEFFECIENTS ARE DEFINED HERE %%
%   Fi=Fv(i).*2^.5*sin(x*pi*i);
%   F=F+Fi;          % SERIES SUM
%end

for I=0:n
    X(I+1,:)=(1:p)*0+F(I+1)+E(I+1,:);              % Xji-s
    M(I+1,:)=median(X(I+1,:));                     % Mi-s
end

FvT=Fv*0;

for I=1:TN
    FvT(I)=sum(M'*2^.5.*sin(I*pi*(0:n)/n))/n;      % Fv TILDA
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%      Fhat[m] APPROXIMATION                                              %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for i=1:TN
    Yi=FvT(i).*2^.5*sin(x*pi*i);
    Y=Y+Yi;     % SERIES SUM
end

%%  IF -- FISHER INFORMATION FOR ERROR DISTRIBUTION %%
%%  e -- ERROR DISTRIBUTION                         %%
%%  LAMBDA=-phi'/phi -- FUNCTION OF ADJUSTED X[i,j] %%

FvT1=FvT;
FvT2=FvT*0;FvT3=FvT*0;FvT4=FvT*0;FvT5=FvT*0;


for ROUND=1:4
    X1=X;
    for i=1:n
        X1(i,:)=X1(i,:)-Y(i);% VALUES FOR X IN APPROXIMATION OF FvTT
    end

    LX=zeros(n+1,p);
    LIKE=LX;
    for i=1:(n+1)
        for j=1:p
            LX(i,j)=LAMBDA(X1(i,j),b);
        end
    end

    Yerror=(sum((Fv-FvT).^2))^.5;

    for v=1:TN
        FvT1(v)=FvT1(v)+sum(sum(LX').*sin(v*pi*(0:n)/n))/(n*p*IF);
    end

    Y=Y*0;
    for i=1:TN
        Yi=FvT1(i).*2^.5*sin(x*pi*i);
        Y=Y+Yi;
    end
    FvT2=(ROUND==1)*FvT1+FvT2;
    FvT3=(ROUND==2)*FvT1+FvT3;
    FvT4=(ROUND==3)*FvT1+FvT4;
    FvT5=(ROUND==4)*FvT1+FvT5;
```

```
end     % ROUND

% ERROR FOR v=2 AFTER 0,2, AND 5 ITERATIONS %%%%%%%%%
Fvt2s(RUN)=FvT(2)-Fv(2);                                % Fvt-Fv
Fvtt2s(RUN)=FvT2(2)-Fv(2);                              % Fvtt-Fv
Fvttt2s(RUN)=FvT5(2)-Fv(2);                             % Fvttt-Fv
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Y1error=(sum((Fv-FvT2).^2))^.5;
Y2error=(sum((Fv-FvT3).^2))^.5;
Y3error=(sum((Fv-FvT4).^2))^.5;
Y4error=(sum((Fv-FvT5).^2))^.5;

Error1(RUN)=Yerror;
Error2(RUN)=Y1error;
Error3(RUN)=Y2error;
Error4(RUN)=Y3error;
Error5(RUN)=Y4error;

disp([RUN/10 Expvar sum(Error1)/RUN/Expvar sum(Error2)/RUN/Expvar
sum(Error3)/RUN/Expvar sum(Error4)/RUN/Expvar sum(Error5)/RUN/Expvar toc/60])

%disp([0  0 (sum(Error1)/RUN/Expvar-1)^.5 (sum(Error2)/RUN/Expvar-1)^.5
(sum(Error3)/RUN/Expvar-1)^.5 (sum(Error4)/RUN/Expvar-1)^.5
(sum(Error5)/RUN/Expvar-1)^.5])

disp([])
end                                                     % RUN
end                                                     % nI
end                                                     % pI
toc
```

A line-by line description of functioning HEAVYTAIL_SIN.m

1. Lines 1-11: Brief description of the program.
2. Line 12: Set random state, so the random numbers will be different each time.
3. Lines 17-22: Input the distribution function ψ – not necessarily normalized. Also input the number of iterations we will use for the Adams method below.
4. Lines 24-27: Input the function $\lambda(x) = \psi'(x)/\psi(x)$.
5. Lines 31-40: Normalize ψ and find the Fisher Information $I(\psi)$.
6. Lines 42-66: Use Runga-Kutta Predictor-Corrector Method to find the cumulative distribution function
$$\Psi(x) = \int_{-\infty}^{x} \psi(s)ds.$$
We find the value $\Psi(x)$ for $x = exp(t)$, where $t = 0, D, 2D, ..., 12$. D is a small number we choose.
7. Line 74: Simulation begins.
8. Lines 75-81: Set the number of runs for which we will calculate the errors. Also define slots for errors and the values of the Fourier coefficient estimates $\hat{v}_s$.
9. Lines 84-87: Set the values of p and n for which we will work. If we want our program to run for more then one set of (p, n), we start a loop which goes over all values of p and n.
10. Line 91: Set a loop to run the simulation a given number of times.
11. Lines 92-93: Get p and n from lines 86-87.
12. Lines 94-95: Get the number of terms in the series defining our function which has to be estimated. Then get m.

13. Lines 97-101: Set spaces for the variable and approximation functions.
14. Lines 104-107: Generate the set of random variables errors $\{\epsilon_{ij}\}_{i=1..n}^{j=1..p}$.
15. Line 109: Calculate the expected error of the estimator for any Fourier coefficient.
16. Lines 113-118: The Fourier coefficients of the "unknown" function $f(x)$ are defined and $f(x)$ is calculated.
17. Lines 120-123: Generates the set of "observations" $\{Y_{ij}\}_{i=1..n}^{j=1..p}$ by adding the error term to $f(i/n)$, where $i = 1, ..., n$.
18. Lines 127-129: Generates the simple estimator for a Fourier coefficient we want to analyze.
19. Lines 134-137: Obtains the approximation $f_{[m]}(x)$ to $f(x)$.
20. Lines 147-177: Performs four rounds of iteration of $\hat{v}_s$ as will be discussed below.
21. Lines 149-151: Estimates the error terms $\epsilon_{ij} \approx X_{ij} - f_{[m]}(i/n)$.
22. Lines 155-159: Estimates $\lambda(X_{ij} - f_{[m]}(i/n))$
23. Lines 163-165: Performs the iteration.
24. Lines 168-171: Calculates the new $f_{[m]}(x)$.
25. Lines 179-188: Calculates the error of the estimates after one to four iterations.
26. Lines 190-196: Produces a vector consisting of errors obtained in different trials. By taking the mean of square of this vector we obtain a mean squared error for the estimate $\hat{v}_s$ after one to four iterations. Then displays the result.

Bibliography

[1] http://www.cmap.polytechnique.fr/ rama/inverse.html As of Sept 29 04 Maintained by: Centre de Mathematiques AppliquesEcolPolytechni F-91128 Palaiseau, France.

[2] Natterer, F. The Mathematics of Computerised Tomography. John Wiley & Sons. New York (1986).

[3] O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion). Statistical Science, 1, 1986, 502-527.

[4] Abda, Ben ; Bui, H.D. Planar crack identification for the transient heat equation. Journal of Inverse and Ill-Posed Problems. Volume: 11 Number: 1 Page: 27 - 31 (March 2003)

[5] Alessandrinit, G; Beretta, E. Stability in crack determination from electrostaticmeasurements at the boundary-a numerical investigation. Inverse Problems I1 (1995) L17-I24

[6] Jonas, P; Louis, A.K. Phase contrast tomography using holographic measurements. Inverse Problems 20 p. 75-102. (20)

[7] Fokasy, A.S.; Gel-fandz, I.M.; Kurylev, Y. Inversion method for magnetoencephalography. Inverse Problems 12 (1996) L9

[8] Wilson, Jim. Watching Thoughts. Popular Mechanics Vol. 181 Issue 9, p28 (September 2004)

[9] http://ladywildlife.com/animal/tigershark.html

[10] Ammari, Habib. An Inverse Source Problem For Maxwell's Equations Magnetoencephalography. Siam Journal of Applied Mathematics. Vol. 62, No. 4, pp. 1369-1382 (2004)

[11] Kirchgraber, Urs. Popularization: The Case of Ill-Posed Inverse Problems. Department of Mathematics, ETH Zurich CH-8092 Zur Switzerland http://www.icme-organisers.dk/taC/ICME-Artikel2.pdf

[12] Kvenvolden, KA. Gas Hydrates, Relevance to World Margin Stability and Climatic Change. Pp. 9-30. (Geological Society of London, 1998).

[13] Cole, Sally. All-optical atomic clock may be world's most precise. Laser Focus World. Vol. 37 Issue 9, p28 (September 2001).

[14] Anderson, Stephen G. Femtosecond pulses attract attention. Laser Focus World, Vol. 36 Issue 8, p5. (August 2000)

[15] Natterer,F; Wubbeling, F. Mathematical Methods in Image Reconstruction. SIAM, Philadelphia, 2001.

[16] Krner, T. W. Fourier Analysis. Cambridge, England: Cambridge University Press, 1988.

[17] Magdalinos, Michael A. Approximate maximum likelihood estimation in linear regression. *Annals of the Institute of Statistical Mathematics* Vol. 45, no. 1, 89–104.

[18] Klaassen, Ch.A.J., Lee, E.-J. & Ruymgaart, F.H. (2004). Asymptotically efficient estimation of linear functionals in inverse regression models. Journal of Nonparametric Statistics 17, 819-831.

[19] Weast, Robert C. CRC Handbook of Chemistry and Physics. CRC Press, INC. Bocca Raton, Florida (1982-83), pp. E46-49.

[20] Piercy, J.E.; Embleton, T.; Sutherland,L. Review of noise propagation in the atmosphere. *Journal of the Acoustical Society of America.* 61(6), June 1977, pp 1403-1418.

[21] Farkas, E. "Transit of Pressure Waves through New Zealand from the Soviet 50 Megaton Bomb Explosion." *Nature.* 4617 (24 February 1962): 765-766.

[22] Lehmann, E. L.; Casella, G. (1998). Theory of Point Estimation. Springer, 2nd ed.

[23] Aldrich, John . "R.A. Fisher and the making of maximum likelihood 1912-1922." *Statistical Science.* Vol 12, Num 3 (1997), 162-176.

[24] Bishwal, Jaya. Parameter Estimation in Stochastic Differential Equations. Springer (2007).

[25] Taibleson, M. Shorter Notes: Fourier Coefficients of Functions of Bounded Variation. Proceedings of the American Mathematical Society, Vol. 18, No. 4. (Aug., 1967), p. 766.

[26] Avriel, Mordecai. (2003). Nonlinear Programming: Analysis and Methods. Dover Publishing

[27] Nocedal, Jorge & Wright, Stephen J. (1999). Numerical Optimization. Springer-Verlag.

[28] Snyman, Jan A. (2005). Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms. Springer Publishing.

[29] A. Khoujmane, F. Ruymgaart, M. Shubov "The Convolution Theorem for Estimating Linear Functionals in Indirect Nonparametric Regression Models." Journal of Statistical Planning and Inference, 137 (2007) 811 820

[30] McNeney, B.; Wellner, J.A. (2000). Application of convolution theorems in semiparametric models with non-i.i.d. data. J. Statist. Pl. Inf. 91, 441-480.

[31] van der Vaart, A.W.; Wellner, J.A. (1996). Weak Convergence and Empirical Processes wtih Applications to Statistics. Springer, New York.

[32] Beran, R. (1977). Robust location estimates. Ann. Statist. 5, 431-444.

[33] Hajek, J. (1970). A characterization of limiting distributions of regular estimates. Zeitschrift Wahrscheinlichkeitsth. Verw. Geb. 14, 323-330.

[34] LeCam, L. (1972). Limits of experiments. Proceedings of Sixth Berkeley Symposium on Mathematical Statististical Probability 1, 245-261. University of California Press, Berkeley.

[35] Millar, P.W. (1985). Non-parametric applications of an infinite-dimensional convolution theorem. Z. Wahrsch. Verw. Gebiete 68, 545-556.

[36] LeCam, L. (1994). An infinite dimensional convolution theorem. In Statistical Decision Theory and Related Topics 5 (S.S. Gupta and J.O. Berger,eds.) 401-411. Springer, New York.

[37] Atiyah, M. F., and I. G. MacDonald, Introduction to Commutative Algebra, Perseus Books, 1969.

[38] Hájek, J. & Sidák, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

[39] LeCam, L (1960). Locally asymptotically normal families of distributions. *University of California Publications in Statistics* **3**, 37-98.

[40] Billingsley, Patrick (1999). Convergence of Probability Measures. New York, NY: John Wiley & Sons, Inc.

[41] Billingsley, P. Convergence of Probability Measures. John Wiley & Sons, Inc., New York.(1999)

[42] Herrlich, Horst. Axiom of Choice. Springer, 2006

[43] Ablowitz, Mark J.; Fokas A. S. (1997). Complex variables: Introduction and applications. Cambridge, UK: Cambridge University Press, 123.

[44] Abramowitz, M.; Stegun, I. eds. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. New York: Dover.(1965)

[45] Szego, G. Concerning the Fourier Coefficients of a Nonnegative Function. The American Mathematical Monthly, Vol. 67, No. 4. (Apr., 1960), pp. 365-366.

[46] Juncosa, M.; Young,D. A Uniform Approximation to Fourier Coefficients. Proceedings of the American Mathematical Society, Vol. 4, No. 3. (Jun., 1953), pp. 373-374.

[47] Feinerman, R.P. and Newman, D.J., "Polynomial Approximation", Williams and Wilkins, Baltimore 1974.

[48] Villadsen, J., Michelsen, M. (1978), Solution of differential equation models by polynomial approximation, Prentice Hall, New Jersey.

[49] Lloyd N. Trefethen and David Bau, III, Numerical Linear Algebra, Society for Industrial and Applied Mathematics, 1997.