Methods for Irregularly Sampled Continuous Time Processes

A thesis presented for the degree of

Doctor of Philosophy of the University College London

by

Zhen Li

Department of Statistical Science

University College London

1-19 Torrington Place, London WC1E 7HB

MARCH 18, 2014

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Signed:

Copyright

Abstract

This thesis will consider methods associated with irregularly spaced sampling of a real-valued continuous time stationary process. The problem of Monte Carlo simulation as well as parametric estimation under irregularly spaced sampling times will be discussed. For the simulation problem, the focus will be on the spectral simulation method. A novel algorithm has been proposed for the determination of the spectral simulation scheme, which is optimal in the sense of achieving required accuracy with minimal computational costs.

The problem of parametric estimation under irregularly spaced sampling times will also be discussed. We will adapt the framework stochastic sampling times, in which the irregularity of the sampling times is modeled through a renewal point process over the real line. By constructing a second order discrete time stationary process from sampling, a parametric estimation method based on the well-known Whittle log-likelihood function will be proposed. Asymptotic consistency of the resulting estimator will be proved by borrowing existing results from literature of renewal theory.

Moreover the performance issue of this proposed estimation procedure will be investigated further. It will be shown that by calculating the spectral density of the sampled discrete time process through a Discrete Fourier Transform (DFT) approximation, the Whittle log-likelihood function can indeed be evaluated relatively efficiently. This estimation method, however, will induce information loss, which will be shown to be related to the unique properties of the renewal kernel function. Although a accurate analysis of the renewal kernel function is not easy, it is still possible to provide some insights on the determining factors of the information loss through asymptotic calculations.

Contents

Chapter 1

Introduction

Continuous time stochastic processes have for a long time been used to model physical, social and economic phenomena [48]. In this thesis we will be focusing on a particular class of continuous time stationary processes, which are frequently used to model quantities with statistical behavior that are invariant over time [64]. In practice observations of these continuous time phenomenon are usually collected at discrete sampling times, for the purposes of analyzing through digital methods.

In vast majority of existing studies the sampling times are assumed to be equally spaced. Under this simplifying assumption, computational efficient algorithms have been developed [63]. In practical applications however, either being constrained by practical conditions [111, 12, 90] or being deliberately chosen by the experimenters [113, 78, 79], data are also very often being recorded at irregularly spaced sampling times.

A very simple form of irregularity in sampling times arises from the corruption of the originally equally spaced sampling times. For example it is not unusual to encounter in many applications that the observations made at equally spaced sampling times are subject to random drop-outs, thus giving us the missing data problem [92]. Another example arises in signal processing applications, in which the originally equally spaced sampling times are randomly perturbed, and leads to the sampling-jitter problem [130].

On the other hand, it is also common in practice to have sampling times that are intrinsically irregular. This could happen when there are physical constraints, such as in astronomy [73, 111] for example, the observation of a star may be limited by the Earth's rotation and orbiting, as well as equipment faults. Also in the field of

laser Doppler anemometry, the data recording procedure implies intrinsic irregularity in the sampling times [12].

Apart from physical constrains, the experimenters may have incentives to deliberately introduce irregularity into the sampling times, in order to avoid the aliasing problem under equally spaced sampling [79]. *In existing literature [113, 8, 79] this is achieved by modeling the sampling times through a stochastic point process over the real line. It will be showed in Section(6.3) that the variability of the sampling times may help to suppress the high-frequency feature of the continuous time process, hence reducing and even alleviating the aliasing problem. Different frameworks of stochastic sampling times have been proposed in existing studies.* These frameworks are very general and can be used to model a variety of sampling patterns, including the sampling patterns from the missing data problem and the sampling-jitter problem.

Observations with irregularly spaced sampling times are much harder to work with, partly because the established and efficient algorithms developed for equally spaced sampling times are no longer applicable. Being motivated by the above practical considerations, in this thesis we will be discussing problems associated with irregularly spaced sampling of continuous time stationary processes. New algorithms and methods for both the Monte Carlo simulation problem and the parametric estimation problem will be proposed and studied in detail.

1.1 The Simulation Problem

We will first consider the problem of Monte Carlo simulation of finite observations from a continuous time stationary process. Under equally spaced sampling times, finite sample observations can be conveniently and efficiently obtained through applying the circulant embedding technique [140, 35]. Through exploring the particular second order structure of the observations under equally spaced sampling times, the method of circulant embedding can achieve both accuracy and efficiency by utilizing the Fast Fourier Transform (FFT) technique.

When the required sampling times are irregularly spaced, however, the finite sample observations do not in general possess useful second order structures that can be explored. Under this situation, either the matrix factorization method [30] or the spectral method [114] can be used to generate the required finite sample observations. The matrix factorization method is not preferable due to the heavy computational

costs involved in the factorization of the covariance matrix [61]. We will therefore in this thesis focus on the spectral simulation method, which generates realizations through discretizing the spectral representation of the underlying continuous time process.

Being an non-exact simulation method, the spectral method generates covariance functions approximating the target through a quadrature rule [115, 47]. We will provide a detailed review of the major variations of the spectral simulation methods. The focus will be on two particular methods that are based on an equally spaced spectral discretization, and generate covariance functions that approximate the target through Trapezoidal quadrature rules. We will provide some new insights on the statistical accuracy of these two methods, hence arguing that both methods should provide adequate approximation, as long as the discrepancy $\epsilon_{\hat{X}}(\tau)$, between the simulated covariance function and target covariances functions, is controlled at a reasonable level. The exact definition of $\epsilon_{\hat{X}}(\tau)$ will become clear in the discussion of spectral simulation in Section(3.3).

It is clear that $\epsilon_{\hat{X}}(\tau)$ should be controlled through spectral discretization scheme - namely specifying the power cut-off frequency L and the spectral discretization interval Δ_λ. Rough and conservative estimates of L and Δ_λ work, but we believe they tend to provide unnecessary accuracy and hence will incur unnecessary computational costs. In spite of its importance however, the systematic determination of the appropriate spectral discretization scheme has not been discussed in existing literature.

As a major contribution of this thesis, we will provide a solution to the problem of systematical determination of the appropriate spectral discretization scheme for the spectral simulation methods. Instead of using a traditional Taylor series expansion argument, we propose to decompose the discrepancy $\epsilon_{\hat{X}}(\tau)$ into a truncation error component and an aliasing error component, which can be respectively controlled through L and Δ_λ independently.

The determination of L is straightforward and can be obtained through a simple application of the Newton-Raphson recursion, by specifying the desired truncation error component in advance. For the determination of Δ_λ, on the other hand, we will propose a novel algorithm by considering the aliasing error component. Numerical studies will be provided to illustrate the fact that, when the spectral discretization scheme is determined from the proposed algorithms, the spectral simulation method

can potentially be efficient and at the same time provide sufficient accuracy.

1.2 The Estimation Problem

Another problem that we will look at the in this thesis is the problem of parametric estimation from sampled observations. Traditionally when the continuous time stationary process is sampled at equally spaced sampling times, the underlying parametric model can be conveniently estimated through the Whittle log-likelihood function [136, 137, 50], which asymptotically approximates the Gaussian log-likelihood function [127]. This approach is well-known for its computational efficiency, thanks to the use of the FFT algorithm in the evaluation of the likelihood function. However the aliasing effect [95] introduced from equally spaced sampling will give ambiguous interpretations for the parametric models obtained from estimation [20].

On the other hand when the sampling times are not equally spaced, the FFT based technique is no longer generally applicable. Depending on the assumptions imposed on the irregularity of the sampling times, there are different approaches available. One approach is to assume the sampling times being deterministic, and the underlying parametric model is estimated through minimizing the exact Gaussian log-likelihood function. Although being the most generally applicable approach, the computational costs involved in inverting the covariance matrix is not desirable. In order to improve the computational efficiency, various techniques have been employed. For example by assuming the underlying parametric model belonging to the CARMA family, the Gaussian log-likelihood function can be efficiently calculated through Kalman recursion [58, 59]. Alternatively it was also suggested to interpolate the data to equally spaced sampling grid, so that the FFT based method can be readily applied [2]. Although being able to improve the computational efficiency of the estimation procedure, these methods have their drawbacks. The Kalman recursion can only be applied to the CARMA family and hence lacks generality, whereas the interpolation method will create biases that are hard to quantify and control [37].

These disadvantages of the existing methods therefore motivate us to consider the approach under which the sampling times are assumed to be stochastic. It is expected that by regularizing the sampling times through some probabilistic structure, more general and elegant methods can be developed. Historically stochastic sampling times was first considered as an alternative to equally spaced sampling times, in the hope

that the aliasing effect can be alleviated, so that consistent non-parametric spectral estimates can be constructed [113, 8, 79]. The most complete framework of stochastic sampling times is the one proposed by Masry [79], in which the distribution of the sampling times over a time interval is modeled by a stationary point process over the real line. Both non-parametric and parametric spectral estimation problems can be rigorously formulated in this framework [70, 72]. The disadvantage of this framework, however, lies in its practical applicability. In fact, as will be discussed later that, the implementation of Masry's framework requires the sampling point process to be rather simple. This may not be a serious problem under the context of signal processing problems considered by Masry, in which the experimenters usually have full control over the sampling schemes. In other more general data analysis problems however, the data generation mechanism is not under control, and consequently the application of Masry's framework will be rather limited.

This limitation therefore leads us to investigate an earlier framework, originally proposed by Shapiro and Silverman [113], in which the irregularity of the sampling intervals are modeled directly. This framework also includes a variety of sampling patterns, among which the renewal sampling schemes is one of the most important examples. In existing literatures, however, this framework was not preferable for the reason that reliable non-parametric estimates can not be easily constructed [41, 81].

As another major contribution to the existing studies of stochastic sampling times, in this thesis we will analyze the parametric estimation problem under renewal sampling schemes in considerable detail. By constructing a discrete time stationary process, we will show that the parametric estimation problem can be conveniently formulated in this framework. In particular an estimation procedure based on a Whittle log-likelihood function can be implemented for a wide class of renewal sampling schemes.

We will investigate the theoretical properties of this Whittle log-likelihood estimation method, showing that the corresponding estimator is indeed asymptotically consistent under quite general conditions. Moreover by calculating the spectral density $S_Y^{\Delta}(f)$ of the sampled discrete time stationary process through a Discrete Fourier Transform (DFT) approximation, we will show that this proposed estimation method enjoys computational advantage over the more general method of minimizing the exact Gaussian log-likelihood function (assuming the sampling times being deterministic).

In order to understand the finite-sample performance of the proposed estimator, the special features of the renewal sampling kernel function has brought to our attention. We will show that the particular shape of the renewal kernel implies a special aggregation/whitening effect, which will indeed help to eliminate one important source of ambiguity in model identification that is caused by the aliasing effect under the equally spaced sampling times. As a cost, however, we will also show that the renewal sampling schemes can only capture low frequency spectral features of the underlying continuous time process, since the high frequency features will be destroyed by the aggregation/whitening effect. By using a Taylor series based asymptotic argument, we can also determine the main factors determining the resolution of the renewal sampling schemes.

1.3 Organization of the Thesis

This thesis will be organized as follows. Chapter 2 contains a brief review of the theory of stationary processes, focusing on its spectral representations. The purpose will be to provide sufficient background materials, and defining appropriate terms and concepts that will be used in subsequent discussions. Chapter 3 discusses the simulation problem. Traditional technique based on equally spaced sampling times will be reviewed. The focus will be on the spectral simulation methods, and new algorithm will be proposed for its practical implementation. In Chapter 4 we will review the problem of parametric estimation of a continuous time stationary process through sampling. We will review the traditional methods based on equally spaced sampling times, the specialized method of Kalman recursion for CARMA family of models, and also the different frameworks of stochastic sampling times. The problem of parametric estimation under renewal sampling schemes will be discussed in detail in Chapter 5, in which asymptotic consistency of the proposed estimator will be established. Chapter 6 considers the implementation and finite-sample performance of the proposed estimator. Simulation studies will also be provided in this chapter.

Chapter 2

Review of Stationary Processes

2.1 Introduction

In many practical situations one wishes to measure and analyze some time-varying physical quantities, that are intrinsically random, in the sense that repeated measurements (or experiments) lead to different results and one cannot predict later values from those observered earlier. Such time-varying physical quantities are modelled by the concept of stochastic processes, which are families of random variables indexed by an indexing set T. In this thesis we will be fucusing on the most commonly seen scenarios, in which the indexing set T consists of either the set of integers $\mathbb{Z}$, or the set of real numbers $\mathbb{R}$. In the former case the family of random variables will be called a discrete time stochastic process, and will be denoted as $Y = Y_k$. Similarly, when $T = \mathbb{R}$ the corresponding family of random variables is called a continuous time stochastic process, and will be denoted as $X = X(t)$. Moreover in this thesis we will only consider stochastic processes that are real valued. Complex valued or vector valued stochastic processes have also been used in practice to describe more complicated physical phenomena [94], but will not be discussed in this thesis.

Important classes of stochastic processes can be defined by imposing appropriate conditions on their finite dimensional distributions [26]. One set of conditions leads to the classes of stochastic processes that are stationary in the sense that the statistical behaviour, measured by its finite dimensional distributions, is invariant over time. As more or less stringent conditions of invariance are imposed, different types of stationary stochastic processes may be defined. Another class of stochastic process is the class of Gaussian process, and is defined by requiring the finite dimensional

distributions are of the multi-dimensional Gaussian. Apart from the fact that many physical phenomena exhibit Gaussianity, the importance of the class of Gaussian processes in statistical modelling also stems from the simplicity of Gaussian distribution, which usually allows many derived quantities to have their distributions calculated explicitly.

In this Chapter, we will provide a quick review for both continuous time and discrete time real-valued stationary processes. We will be discussing strict and weak stationarity for both continuous and discrete time stochastic processes, their covariance structures, and most importantly, their spectral representations. The purpose will be to provide sufficient background material, and defining appropriate concepts and terms, in order to facilitate subsequent discussions on simulation and sampling of continuous time stationary processes at random instances.

2.2 Continuous Time Stationary Processes

Many random physical quantities can be observed continuously in time, and therefore can be described by continuous time stochastic processes. Moreover, the properties of these physical quantities also exhibit some degree of invariance over time. In statistical terms this invariance is described by the concept of stationarity, which is defined by imposing conditions on the finite dimensional distributions. Depending on how stringent these conditions are, there will be different types of stationarity. In this thesis, however, we will be concerned with two types of stationarity that are widely assumed in practice. The reviewing materials contained in this section is largely based on Cramér and Leadbetter [26] and Koopmans [65].

A continuous time stochastic process $X = X(t)$ is called strictly stationary, if the whole family of its finite dimensional distributions are invariant under a translation of time. As a direct consequence of this strict stationarity, the mean $\mathrm{E}[X(t)]$, when exists, must be constant for all t. Moreover the covariance $\mathrm{Cov}[X(t), X(s)]$, when exists, should only depend on the time difference $\tau = t-s$, thus defining the covariance function $R_X(\tau)$ as

$$R_X(\tau) = \mathrm{Cov}[X(t), X(t+\tau)], \ \ \tau \in \mathbb{R}. \tag{2.1}$$

Without loss of generality, however, in the remaining discussions of this thesis we will always assume the process $X = X(t)$ to have zero mean.

The concept of this strict stationarity is very strong, in the sense that any properties of $X = X(t)$ depending on the finite dimensional distributions should also be independent of the time shift τ. On the other hand, many important properties of $X = X(t)$ can be expressed in terms of its mean $\mathrm{E}[X(t)]$ and covariance $\mathrm{Cov}[X(t), X(s)]$. Therefore the stationarity of mean and covariance will be sufficient to guarantee the stationarity of all these properties. This observation gives the motivation to study the class of continuous time stochastic processes $X = X(t)$ such that $\mathrm{E}[X(t)] = \mu_X$ and $\mathrm{Cov}[X(t), X(s)] = R_X(t-s)$. This class of processes has received various names in history, such as second-order stationary, wide-sense stationary, or weakly stationary stochastic processes. In this thesis we will use the term weakly stationary processes.

Immediately following from the definitions of strict and weak stationarity is the fact that, assuming the existence of first and second order moments, strict stationarity implies weak stationarity but not vice versa. There is however a notable exception when $X = X(t)$ is Gaussian. In this case the strict and weak stationarity becomes equivalent, because the multi-dimensional distributions of a Gaussian process are fully determined by the mean and covariance function [26].

The above relationship between the two types of stationarity implies that any properties of a weakly stationary processes should be automatically shared by the strictly stationary processes. Consequently in subsequent discussions in this thesis, for the purpose of simplicity we will use the term stationary process to mean either strict or weak stationarity. In other words if some results are stated for stationary processes, it means that they will hold for both strictly and weakly stationary processes. When either strict or weak stationarity is required in particular, it will be specified explicitly.

The covariance function $R_X(\tau)$, which describes the covariance structures (or the second order properties) of a real-valued stationary process, is an even and nonnegative definite function that is bounded by its value $R_X(0)$ at the origin. Moreover, if the process is assumed to be mean-square continuous [26], then the covariance function $R_X(\tau)$ can be shown to be continuous everywhere [87]. By Bochner's theorem [100], if the covariance function is integrable then there exists a non-negative function $S_X(\lambda)$, such that

$$R_X(\tau) = \int_{-\infty}^{\infty} e^{i2\pi\lambda\tau} S_X(\lambda) d\lambda, \quad S_X(\lambda) = \int_{-\infty}^{\infty} e^{-i2\pi\lambda\tau} R_X(\tau) d\tau. \tag{2.2}$$

The function $S_X(\lambda)$ is called the spectral density function, and describes the distribution of the power [1] of the process over the frequency domain. It provides an alternative frequency domain description of the second order properties of $X = X(t)$. In particular it shows that the covariance function $R_X(\tau)$ can be decomposed as a superposition of sinusoids $e^{i2\pi\lambda\tau}$, with amplitude at each frequency λ given by $S_X(\lambda)$.

Surprisingly, it can be shown that under the assumption of mean-square continuity, the process $X = X(t)$ also admits a similar real-valued spectral representation

$$X(t) = \int_{-\infty}^{\infty} e^{i2\pi\lambda t} dZ_X(\lambda) \;\; t \in \mathbb{R}, \tag{2.3}$$

where $Z_X(\lambda)$ is a complex-valued stochastic process with orthogonal increments [65]. In this thesis, however, the process $X = X(t)$ will always be assumed to be real-valued, and consequently the above complex-valued spectral representation may not be in its most convenient form. In subsequent discussions, we will therefore use the following real-valued spectral representation, which can be easily derived from the complex-valued representation [65]:

$$X(t) = \int_0^{\infty} \cos(2\pi\lambda t) dU_X(\lambda) + \int_0^{\infty} \sin(2\pi\lambda t) dV_X(\lambda), \;\; t \in \mathbb{R}. \tag{2.4}$$

Here $U_X(\lambda)$ and $V_X(\lambda)$ are uncorrelated real-valued stochastic processes. The increments $dU_X(\lambda)$ and $dV_X(\lambda)$ [2] have the following properties:

1. $\mathrm{E}[dU_X(\lambda)] = \mathrm{E}[dV_X(\lambda)] = 0$, as a consequence of assuming $\mathrm{E}[X(t)] = 0$, for any t;

2. $\mathrm{E}[dU_X(\lambda)dU_X(\mu)] = \mathrm{E}[dV_X(\lambda)dV_X(\mu)] = 2\delta(\lambda - \mu)S_X(\lambda)d\lambda$ [3],

[1] The concept of power measures the activity of a signal, when its energy is not defined. In the case of a continuous time weakly stationary process, power is simply the variance $\sigma^2 = R_X(0)$ of the process. Percival and Walden provides a heuristic argument showing that for a continuous time weakly stationary process, the quantity $dS_{I,X}(\lambda)$ can be regarded as the average contribution (over all realizations) to the power from components with frequencies in a small interval about λ.

[2] Here $dU_X(\lambda)$ and $dV_X(\lambda)$ represents the increment of the real-valued process $U_X(\lambda)$ and $V_X(\lambda)$ over the interval $(\lambda, \lambda + d\lambda]$; in other words, it can be understood as

$$dU_X(\lambda) \approx U_X(\lambda + d\lambda) - U_X(\lambda), \;\text{ and }\; dV_X(\lambda) \approx V_X(\lambda + d\lambda) - V_X(\lambda) \tag{2.5}$$

where $d\lambda$ is an infinitesimal increment in the frequency variable.

[3] Here when $\lambda \neq \mu$, the quantities $dU_X(\lambda)$ and $dU_X(\mu)$ (and similarly $dV_X(\lambda)$ and $dV_X(\mu)$)

where $\delta(\lambda)$ represents the Dirac delta function centering at the origin. Hence the real-valued continuous time stationary process can also be written as a superposition of sinusoids with randomized magnitudes given by $dU_X(\lambda)$ and $dV_X(\lambda)$. This spectral representation is fundamental in the analysis of the continuous time stationary process. In fact in the words of Koopmans [65], the spectral representation is one of the essential reasons for the historical central position held by stationary stochastic processes. Apart from its implications for the analysis of continuous time stationary processes, we will see in Chapter 3 that the spectral representation, given by Equation(2.4), also motivates the spectral simulation method that can be used to generate the realizations of $X = X(t)$.

2.3 Examples of Continuous Time Stationary Processes

Having reviewed the relevant theory of continuous time stationary process and its spectral representation, in this section we will briefly introduce some concrete families of stationary processes that are frequently used in practice to describe random quantities observed from of physical phenomenon. They will also appear in subsequent discussions as concrete examples of continuous time stationary processes.

2.3.1 Continuous Time ARMA Processes

Continuous time auto-regressive moving-average process, abbreviated as CARMA process, provides a very convenient parametric family of stationary processes. It exhibits a very wide range of covariance structures, and has been applied in many distinct applications [96, 131]. A CARMA(p, q) process has an explicit time-domain state-space dynamic, in terms of a pth order linear differential equation [21]. Although the covariance function of a CARMA process can be derived through the state-space formulation, the second order properties can be most conveniently described by its spectral density function, which takes the form of a rational function. In other words $S_X(\lambda)$ can be written as

$$S_X(\lambda) = \frac{|b(i2\pi\lambda)|^2}{|a(i2\pi\lambda)|^2}, \tag{2.6}$$

represent increments over two non-overlapping frequency intervals, i.e. if we write $dU_X(\lambda) \approx U_X(\lambda + d\lambda) - U_X(\lambda)$ and $dU_X(\mu) \approx U_X(\mu + d\mu) - U_X(\mu)$, then implicitly we assume that the interval $(\lambda, \lambda + d\lambda]$ and $(\mu, \mu + d\mu]$ are not overlapping with each other.

where $a(z)$ and $b(z)$ are polynomials of order p and q respectively. Consequently the CARMA family has also been referred to as the family of stationary processes with rational spectral densities. The family of CARMA process will be briefly mentioned in this thesis. In Chapter 3 about spectral simulation methods, a particular example of this rational spectral density has been used to illustrate the proposed algorithm for finding the appropriate spectral discretization scheme. Moreover in Chapter 4, we will briefly review the parametric estimation problem for CARMA process through the Kalman recursion technique.

2.3.2 Matérn Class Processes

Unlike the CARMA process which has an explicit time domain dynamic, the Matérn class of continuous time processes, which is named by Stein [122] after the work of Matérn [86], is only specified by the covariance function $R_X(\tau)$ and the corresponding spectral density $S_X(\lambda)$, which are given respectively by

$$R_X(\tau) = 2\sigma^2 \frac{(\pi\phi\tau)^\nu}{\Gamma(\nu)} K_\nu(2\pi\phi\tau), \tag{2.7}$$

$$S_X(\lambda) = \sigma^2 \frac{\Gamma(\nu + 0.5)\phi^{2\nu}}{2\sqrt{\pi}\Gamma(\nu)} \frac{1}{(\phi^2 + \lambda^2)^{\nu+0.5}}, \tag{2.8}$$

where $K_v(\tau)$ is the modified Bessel function of the second type. Therefore the Matérn class is in fact a parametric family of covariance and spectral models. Each parameters within the Matérn class of models controls different aspects of the properties of the sample function [102]:

- Parameter σ^2: controls the stationary variance of the process.
- Parameter ν: the smoothness parameter and controls the mean-square differentiability (i.e smoothness) of the process; if $\nu > k$ then the corresponding Matérn process will be k times mean-square differentiable;
- Parameter ϕ: the range parameter that controls the decay of the covariance function - higher the value of ϕ, faster the decay of the covariance function.

Note that when the smoothness parameter $\nu \to \infty$ we will obtain the Gaussian covariance function $R_X(\tau) = \sigma^2\phi\sqrt{2\pi}e^{-2\pi^2\phi^2\tau^2}$, which gives infinitely mean-square

differentiable sample path. The Matérn class of models also becomes especially simple when ν is half-integer, i.e. when $\nu = k + \frac{1}{2}$. In this case the covariance function can be shown to be a product of an exponential function and a polynomial of order k. General expression can be found in Abramowitz and Stegun[1]. Being very flexible, the Matérn processes are frequently used as covariance models in areas like spatial interpolation [122], and machine learning [102]. In Chapter 3 we will be discussing the spectral simulation method, and the Matérn process will be used to illustrate the performance of the proposed algorithm of constructing the spectral domain discretization schemes.

2.3.3 Oscillatory Matérn Processes

In Chapter 5 and Chapter 6 we will be studying the parametric estimation problem under the renewal sampling schemes. The Matérn family of models provide ideal examples for the reason that it has flexible control over important aspects of the sample path properties. Moreover, the Matérn process does not possess an explicit time domain dynamic, hence it can be used to show the computational advantages of our proposed estimation method.

On the other hand, in order to demonstrate the unique and interesting properties of the renewal sampling schemes, we will also require a covariance function that has built-in oscillation. Instead of introducing another model, we will simply add an extra oscillation feature into the Matérn family of models, and call the corresponding process the Oscillatory Matérn process. We can do this by shifting the spectral density in the frequency domain. Specifically suppose $\tilde{S}_X(\lambda)$ is the Matérn spectral density function, then for an arbitrary $\lambda_0 > 0$, we define the density function $S_X(\lambda)$ for the Oscillatory Matérn process as

$$\begin{aligned} S_X(\lambda) &= \frac{1}{2}[\tilde{S}_X(\lambda - \lambda_0) + \tilde{S}_X(\lambda + \lambda_0)] \\ &= \sigma^2 \frac{\Gamma(\nu + 0.5)\phi^{2\nu}}{2\sqrt{\pi}\Gamma(\nu)} \left[\frac{1}{(\phi^2 + (\lambda - \lambda_0)^2)^{\nu+0.5}} + \frac{1}{(\phi^2 + (\lambda + \lambda_0)^2)^{\nu+0.5}} \right]. \end{aligned} \tag{2.9}$$

It could then be easily shown that the corresponding covariance function $R_X(\tau)$ of the Oscillatory Matérn process is related to the covariance function $\tilde{R}_X(\tau)$ of the

ordinary Matérn process through

$$\begin{aligned} R_X(\tau) &= \tilde{R}_X(\tau)\cos(2\pi\lambda_0\tau) \\ &= 2\sigma^2\frac{(\pi\phi\tau)^\nu}{\Gamma(\nu)}K_\nu(2\pi\phi\tau)\cos(2\pi\lambda_0\tau). \end{aligned} \tag{2.10}$$

In this way we introduced the extra oscillation term $\cos(2\pi\lambda_0\tau)$, with oscillation frequency λ_0, into the Matérn family of models.

2.4 Discrete Time Stationary Processes

The review in this section of discrete time stationary process largely parallels the review of continuous time process in the last section. A real-valued discrete time stochastic process $Y = Y_k$ is called strictly stationary if the whole family of finite dimensional distributions [26] are invariant under a translation in time t. Therefore any characteristic of the discrete time process that depends solely on the finite dimensional distributions should also be invariant over time. In particular this implies stationarity of the first and second order moments, in the sense that the mean of the process should be constant, and there exists an autocovariance sequence $C_Y(k)$ such that

$$\mathrm{Cov}[Y_k, Y_l] = C_Y(k-l). \tag{2.11}$$

Without loss of generality, we will always assume the discrete time stationary process $Y = Y_k$ have zero mean.

On the other hand, just as in the continuous time case, many important properties of the discrete time process can be expressed in terms of its mean and autocovariance. This therefore leads to the study of the class of discrete time weakly stationary process, which is defined by imposing the stationarity on the first and second order properties of the process. Strict stationarity implies weak stationarity, but not vice versa in general, except for the case when the discrete time stochastic process is Gaussian. Similarly as in the continuous time case, in subsequent discussions we will use the term discrete time stationary process to mean either strict or weak stationarity. When either strict or weak stationarity is required in particular, it will be specified explicitly.

The autocovariance sequence $C_Y(k)$, which describes the covariance structure of

a real-valued discrete time stationary process, is an even and nonnegative definite sequence that is bounded by $C_Y(0)$ [22]. As in the continuous case, the nonnegative definiteness of the autocovariance sequence $C_Y(k)$ implies a similar spectral representation. Assuming the sequence $C_Y(k)$ is absolutely summable, then by Herglotz's theorem [22] there exists a continuous nonnegative function $S_Y^{\Delta_t}(f)$ such that

$$C_Y(k) = \Delta_t \int_{-1/2\Delta_t}^{1/2\Delta_t} e^{i2\pi f k \Delta_t} S_Y(f) df, \quad S_Y^{\Delta_t}(f) = \Delta_t \sum_{k=-\infty}^{\infty} C_Y(k) e^{-i2\pi f k \Delta_t}. \tag{2.12}$$

Here Δ_t is the time interval between adjacent observations Y_k and Y_{k-1} is given by Δ_t, which is usually assumed to be one for simplicity. This spectral density function $S_Y^{\Delta_t}(f)$ is periodic with period $\frac{1}{\Delta_t}$, and describes how the power of the discrete time weakly stationary process distributes over this frequency domain. Moreover, just as in the continuous time case, a discrete time stationary process $Y = Y_k$ itself also admits a similar spectral representation

$$Y_k = \int_{-1/2\Delta_t}^{1/2\Delta_t} e^{i2\pi f k \Delta_t} dZ_Y(f), \tag{2.13}$$

where $Z_Y(f)$ is a complex-valued stochastic process with orthogonal increments. Since the discrete time stationary process considered in this thesis will always be real-valued, the above complex-valued spectral representation can be reduced to the following real-valued form as [22]

$$Y_k = \int_0^{1/2\Delta_t} \cos(2\pi f k \Delta_t) dU_Y(f) + \int_0^{1/2\Delta_t} \sin(2\pi f k \Delta_t) dV_Y(f), \tag{2.14}$$

where $U_Y(f)$ and $V_Y(f)$ are uncorrelated real-valued process with orthogonal increments, defined up to an additive constant over the frequency range $[0, \frac{1}{2\Delta_t}]$, and has similar properties as the orthogonal increment process $U_X(\lambda)$ and $V_X(\lambda)$ for continuous time weakly stationary process:

1. $\mathrm{E}[dU_Y(f)] = \mathrm{E}[dV_Y(f)] = 0$, for any $f \in [0, \frac{1}{2\Delta_t}]$, as a consequence of the assumption that $\mathrm{E}[Y_k] = 0$, for any integer k;

2. $\mathrm{E}[dU_Y(f)dU_Y(\nu)] = \mathrm{E}[dV_Y(f)dV_Y(\nu)] = 2\delta(f-\nu)S_Y^{\Delta_t}(f)df$ for any $f, \nu \in [0, \frac{1}{2\Delta_t}]$.

This discrete time version of spectral representation is also fundamental in the analy-

sis of discrete time weakly stationary processes [22]. In Chapter 3 we will be discussing the circulant embedding simulation technique for generating finite sample of a discrete time stationary process $Y = Y_k$. It will be briefly mentioned that the spectral representation of $Y = Y_k$ provides an alternative interpretation of this simulation technique.

2.5 Summary

This chapter briefly reviewed the frequently encountered results from continuous and discrete time stationary processes, both in strict and in weak sense. The focus has been on the spectral representation of the stationary processes, because of its fundamental role in the analysis of such processes. We will see in the next chapter that the spectral representation also serves as the motivation of spectral simulation algorithms that can be used to generate sample path from Gaussian continuous time stationary processes.

Chapter 3

Simulation Methods for Samples from Continuous Time Stationary Processes

3.1 Introduction

In this chapter we will consider this Monte Carlo simulation problem. We will provide a concise review of methods available under different scenarios, especially the circulant embedding technique under the setting of equally spaced sampling times. We will also focus on the so called spectral simulation method under the setting of unequally spaced sampling times, proposing new algorithm for its practical implementation.

Let us consider a zero-mean real-valued continuous time stationary Gaussian process $\{X = X(t) : t \in \mathbb{R}\}$ defined on the real line, which is characterized by its covariance function $R_X(\tau)$ and spectral density function $S_X(\lambda)$. We will in this chapter further assume that the process under consideration has a short range dependence, in the sense that the covariance function $R_X(\tau)$ decays fast enough so that it is integrable over the real line. This is equivalent to assuming that the spectral density $S_X(\lambda)$ is continuous at $\lambda = 0$. A problem of practical interest is to generate a Monte Carlo sample $\mathbf{X}_n = (X(t_1), \cdots, X(t_n))^T$ of the process with n observations at sampling times $\{t_1, \cdots, t_n\}$. Such a simulation is of great use, for example in geostatistics, in where highly variable quantities such as hydraulic conductivity of a groundwater formation are usually simulated as a continuous time Gaussian stationary processes [64]. Monte

Carlo simulation is also applied to estimate probabilistic characteristics in the output of some linear and nonlinear systems of inputs that can be modeled as a stationary Gaussian processes. For example in engineering problems, many input forces have a random nature and could be modeled by such stationary processes [27, 138]. Also in statistics, simulation studies have been routinely used to assess the finite sample and asymptotic performance of some estimation procedures.

Realizations of a stationary Gaussian process $X = X(t)$ can be obtained in various ways. When this process belongs to the continuous time ARMA (CARMA) family, a convenient time domain dynamic is available in the form of a stochastic differential equation (SDE) [20, 131]. A finite sample of size n can then be obtained by discretizing this SDE at the required sampling times. A potential problem associated with this approach is the need to evaluate matrix exponentials of the form e^{As}, where A is an appropriate matrix. In practical applications, however, the CARMA process will not have a very large order, and consequently the matrix exponentials appearing in the time domain discretization will not have a large size [20]. This in particular implies that, when implemented with modern efficient algorithms [88], the evaluation of these matrix exponentials will not significantly inflate the computational costs of $O(n)$ floating-point operations associated with the time domain discretization [131]. For other more general stationary Gaussian processes, however, such a convenient time domain dynamic is usually not available, and the two most commonly used approaches that are applicable under such situation include the matrix factorization method [30, 57] and the spectral method [114, 142, 115, 47].

The matrix factorization approach is based on the simple fact that the finite sample $\mathbf{X}_n = (X(t_1), \cdots, X(t_n))^T$ forms a zero-mean multivariate Gaussian random vector, with the (j,k)th entry of the corresponding covariance matrix Σ_n given by $R_X(t_j - t_k)$. Being a covariance matrix, it admits the following factorization [112]

$$\Sigma_n = P_n P_n^T, \tag{3.1}$$

where P_n^T is the transpose of the matrix P_n. Then if $\mathbf{Z}_n$ is a vector of n IID standard normal random variables, the linear transformation

$$\mathbf{X}_n = P_n \mathbf{Z}_n \tag{3.2}$$

can be used to obtain the realizations of the finite sample $\mathbf{X}_n$ with exactly the desired covariance structure. However this exact method has its drawbacks: it requires the storage of a potentially large covariance matrix Σ_n, and the decomposition algorithm is generally speaking computationally expensive. Indeed, if a Cholesky decomposition [75] for the covariance matrix Σ_n is used, calculating such a decomposition will require a computational cost of $O(n^3)$ floating-point operations [61], and calculating each particular realization will require a further $O(n^2)$ floating-point operations [35].

When the covariance structure possesses relatively fast decay, or when the sample size n required from simulation is relatively large, it should be expected that most of the entries in the covariance matrix Σ_n are close to zero. Hence a straight-forward simplification is to construct a band-limited matrix Σ_n^B from Σ_n by truncating the covariances in Σ_n to zero at some threshold lag. There are however some potential problems associated with this approach. First of all, we note that due to the irregularity of the sampling times, the truncation lag for each row of Σ_n is not uniform, thus complicating the task of constructing the required band-limited matrix. Moreover, the truncation introduces errors, and in extreme cases will even destroy the positive-definite property required for a covariance matrix. How to quantify these truncation errors, and how to ensure the positive-definiteness of the resulting band-limited matrix, are two important tasks in the implementation of this approach.

The spectral method, on the other hand, uses the spectral density function $S_X(\lambda)$ to provide information of the second order properties of the underlying Gaussian process $X = X(t)$. As pointed out by Percival [93], the spectral simulation method can be convenient under situations when the target process is specified through spectral densities rather than the covariance function. Although from a theoretical point of view, the spectral density $S_X(\lambda)$ can be quite general, in practice almost all spectral models are assumed to be continuous [65]. Hence, in subsequent discussions, unless otherwise stated we will implicitly assume continuity for the spectral density $S_X(\lambda)$ of the target process. This simulation method is motivated from the spectral representation of a zero-mean stationary Gaussian process, which as discussed in Section(2.2), can be written in its real form as

$$X(t) = \int_0^\infty \cos(2\pi\lambda t) dU_X(\lambda) + \int_0^\infty \sin(2\pi\lambda t) dV_X(\lambda), \quad t \in \mathbb{R}, \tag{3.3}$$

where the uncorrelated orthogonal increment processes $U_X(\lambda)$ and $V_X(\lambda)$ are in fact

also Gaussian [117]. The spectral simulation method suggests approximating the target process $X = X(t)$ by a discretized version of Equation(3.3). The spectral discretization is accomplished by first truncating the frequency domain into a finite interval $[0, L]$ and then divide into $m + 1$ grid points $\{\lambda_0, \lambda_1, \cdots, \lambda_m\}$, where $\lambda_0 = 0$ and $\lambda_m = L$. The process $X = X(t)$ can then be approximated by an aggregation of sinusoids of the following form

$$\hat{X}(t) = \sum_{k=0}^{m} [\cos(2\pi\lambda_k t) d\hat{U}_X(k) + \sin(2\pi\lambda_k t) d\hat{V}_X(k)], \tag{3.4}$$

where $d\hat{U}_X(k)$ and $d\hat{V}_X(k)$ are uncorrelated random variables which, roughly speaking, approximate respectively the increments $dU_X(\lambda_k)$ and $dV_X(\lambda_k)$ of the orthogonal processes around the frequency λ_k. As will be discussed in Section(3.3), there are different ways to approximate such increments [115, 47]. Each method will lead to slightly different statistical properties of the simulated realization $\hat{X} = \hat{X}(t)$.

Although the frequency domain discretization grid could be chosen arbitrarily, in practice being motivated from both computational efficiency and simulation error control, it is usually chosen to be equally spaced. Criticism of the spectral simulation method mostly stems from its not being exact: as will be seen in Section(3.3) the covariance function $R_{\hat{X}}(\tau)$ of the simulated process $\hat{X} = \hat{X}(t)$ is only a quadrature approximates to the target covariance function $R_X(\tau)$.

In existing literature, focus has been almost entirely on the scenario in which the sampling times are equally spaced, i.e. when the sampling times $\{t_k : k \in \mathbb{Z}\}$ are generated by

$$t_k = t_{k-1} + \Delta_t, \;\; k \in \mathbb{Z}, \tag{3.5}$$

where Δ_t is the equally spaced sampling interval. Under this scenario, simplifications can be made for both the matrix factorization and the spectral simulation methods, allowing significant reduction of computational costs. The covariance matrix Σ_n, under equally spaced sampling times, has a Toeplitz structure (see, for example Franklin [38]). When such a structure is explored, a Cholesky factorization can be accomplished through the Levinson-Durbin recursion [68, 36, 133], which requires a significantly smaller computational cost of $O(n^2)$ floating-point operations, as compared to the earlier mentioned general case. The most significant reduction of computational costs, however, can be potentially achieved for both matrix factorization and

spectral simulation methods by utilizing the Fast Fourier Transform (FFT) technique (see for example Brigham [17]).

For matrix factorization method, the Toeplitz covariance matrix Σ_n can be embedded in the upper-left corner of an extended circulant covariance matrix Σ_{2M} of size $2M > 2n$, which can then be efficiently factorized by the standard orthonormal FFT matrix [93, 140, 35]. Generation of a realization (of size n) can then be accomplished through a further application of the FFT technique, with a total computational costs of a logarithmic order of $O(2M \log(2M))$ floating-point operations. The key step in applying the circulant embedding technique lies on the construction of the extended circulant embedding matrix Σ_{2M}, which has to be nonnegative-definite and at the same time the size $2M$ is not excessively large. We will present a detailed review of the circulant embedding technique in Section(3.2). On the other hand if we choose an equally spaced spectral discretization scheme, then Equation(3.4) can be written in a particular form to allow the application of the FFT technique [141, 142, 115], so that the finite samples can be generated with a computational costs of $O(m \log(m))$ floating-point operations, where m is the number of grid points used in the spectral simulation method.

The utilization of FFT technique for both circulant embedding based matrix factorization method and the spectral method, under equally spaced sampling in time, suggests that there is a relationship between these two methods. This is indeed the case, as will be discussed in Section(3.2.3), that the circulant embedding simulation formula can also be formulated in a way that resembles the spectral simulation formula given by Equation(3.4), but with the orthogonal increments approximated through covariance function $R_X(\tau)$ rather than through the spectral density $S_X(\lambda)$.

However, the convenient assumption of equally spaced sampling times is not always reasonable. It is well known that equally spaced sampling times will give an aliasing effect, which in turn causes a spectral domain identification problem [95]. As a remedy, many methods have been proposed in signal processing literature [84, 79, 80, 70] to give consistent spectral estimate under randomized sampling times, and the assessment of these methods are usually conducted through Monte-Carlo simulations [83], in which finite samples of a continuous time stationary Gaussian process sampled at randomized sampling times are required to be generated. Under such a scenario, the covariance matrix $\mathbf{\Sigma}_n$ does not in general possess useful structures (such as being Toeplitz) that can be explored to reduce the $O(n^3)$ computational costs in-

volved in the simulation method based on matrix factorization. Moreover, since the sampling times are randomized, each realization will have a different set of sampling times, implying the expensive operation of matrix factorization has to be performed each time a realization is generated. Although being exact and generally applicable, the high computational costs involved in the matrix factorization method are not desirable in this situation. On the other hand, under the scenario of randomized sampling times the FFT technique can no longer be applied to the spectral simulation method. The computational costs of the spectral simulation method, as can be seen from Equation(3.4), then increases linearly with the size of the spectral discretization grid.

We will, however, argue in Section(3.4) that, by choosing an appropriate (equally spaced) spectral discretization scheme, the computational costs of the spectral simulation scheme can be kept at a manageable level. This is accomplished through finding a spectral discretization scheme that provides just the sufficient accuracy, in terms of the discrepancy between the covariance structures of the realizations and that of the target process. Such discrepancy between covariance structures were analyzed in existing literature through the traditional Taylor series expansion approach [115, 52], which will not provide sufficient information for the determination of the appropriate spectral discretization scheme. In this paper, we will look at this discrepancy from a different perspective. As a major contribution in this chapter, we will show that the discrepancy can instead be decomposed into a truncation error component and an aliasing error component, and based on this decomposition we propose novel algorithm to appropriately determine the (equally spaced) spectral discretization scheme. Subsequent numerical studies show that in many situations the spectral simulation method, based on appropriately calculated spectral discretization scheme, will have a definite computational advantage over the matrix factorization method.

The above brief introduction on methods of generating realizations from stationary Gaussian processes is summarized in Table(3.1) for equally spaced sampling times, and Table(3.2) for randomized (unequally) spaced sampling times. Note that the structure of the Table(3.1) is designed in a particular way in order to stress the fact that circulant embedding technique can be interpreted both as a matrix factorization method, and also a spectral simulation method. The remaining of this chapter is organized as follows. In Section(3.2) we briefly review the method of circulant embedding simulation algorithm, in which both the implementation details and spectral

domain interpretations are discussed. In Section(3.3) we review the existing spectral simulation methods, namely the Random Phase Deterministic Amplitude (RPDA) and Random Phase Random Amplitude (RPRA) methods, and discuss their differences. The practical problem of determining the appropriate spectral discretization scheme used for spectral simulation method is discussed in Section(3.4), in which we propose a novel algorithm by which the appropriate discretization scheme can be calculated under mild assumptions. Numerical examples and simulation studies are given in Section(3.5).

Equally Spaced Sampling						
Methods	Algorithm	Applicability	Need.	Exactness	Complexity	Disadvantage
Matrix Factorization	Cholesky Decomposition	General	$R_X(\tau)$	Exact	$O(n^2) + O(n^2)$	Slow for long sample
	Circulant Embedding	General	$R_X(\tau)$	Exact	$O(2M \log(2M))$	Embedding matrix may not be positive definite
Spectral	FFT Spectral Discretization	General	$S_X(\lambda)$	Approx.	$O((m)\log(m))$	No freedom to choose sampling intervals
Model Specific	CARMA	CARMA	Dynamic	Exact	$O(n)$	Not General; Need to evaluate e^{As}

Table 3.1: Summary of Simulation methods available for continuous time stationary Gaussian process, sampled at regularly spaced sampling times. In the table, n is the sample size; $2M$ is the size of the embedding circulant covariance matrix; and m is the number of harmonics used in spectral discretization.

Unequally Spaced Sampling						
Methods	Algorithm	Applicability	Need	Exactness	Complexity	Disadvantage
Matrix Factorization	Cholesky Decomposition	General	$R_X(\tau)$	Exact	$O(n^3) + O(n^2)$	Slow for long sample; Storage of covariance matrix
Spectral	Spectral Discretization	General	$S_X(\lambda)$	Approx.	$O((m)n)$	Not exact
Model Specific	CARMA	CARMA	Dynamic	Exact	$O(n)$	Not General; Need to evaluate e^{As}

Table 3.2: Summary of Simulation methods available for continuous time stationary Gaussian process, sampled at irregularly spaced sampling times. In the table, n is the sample size; and m is the number of harmonics used in the spectral discretization.

3.2 Circulant Embedding

When the sampling times $\{t_k : k \in \mathbb{Z}\}$ are equally spaced, i.e. given by $t_k = t_{k-1} + \Delta_t$, where Δ_t is the time domain sampling interval, then the sampled process $Y = Y_k = X(k\Delta_t)$ is a discrete time stationary Gaussian process. Its autocovariance sequence $C_Y(k)$ is given by $C_Y(k) = R_X(k\Delta_t)$, which is a discretization of the covariance function $R_X(\tau)$ of the underlying continuous time stationary Gaussian process. Without loss of generality, in the discussions of this section we will assume the sampling interval to have unit length, i.e. $\Delta_t = 1$. In response to the drawbacks of both the exact matrix factorization method and the non-exact spectral method, a new technique called circulant embedding has been developed by various authors [32, 140, 35, 94] for generating finite sample realization $\mathbf{Y}_n = (Y_1, \cdots, Y_n)^T$. Potentially, this technique has the ability of achieving both accuracy and efficiency: it can be regarded as an exact matrix factorization method but implemented through the efficient FFT algorithm.

The key idea behind circulant embedding is the observation that the covariance matrix Σ_n of the required finite sample, which has a Toeplitz structure, can be embedded in the upper-left corner of a larger covariance matrix Σ_{2M} of size $2M \geq 2n$. This larger embedding covariance matrix Σ_{2M} is constructed to be circulant, in the sense that the jth row of Σ_{2M} can be obtained by circularly shifting $(j-1)$th row one unit to the right. The reason for constructing a circulant embedding matrix is the fact that it can be decomposed by the standard orthonomal FFT matrix [39]. More precisely if Σ_{2M} is the embedding circulant covariance matrix, then it admits the following eigenvalue decomposition

$$\Sigma_{2M} = F_{2M}^H \Lambda_{2M} F_{2M}, \tag{3.6}$$

where F is the $2M$-by-$2M$ orthonormal FFT matrix with the (j,k)th entry given by

$$[F]_{j,k} = \frac{1}{\sqrt{2M}} e^{-i2\pi(j-1)(k-1)/2M}, \;\; 1 \leq i, j \leq 2M. \tag{3.7}$$

and the subscript $''H''$ represents taking the conjugate-transpose of the matrix F. The matrix Λ_{2M} in the above decomposition is a $2M$-by-$2M$ diagonal matrix whose diagonal entries are simply the standard Discrete Fourier transform (DFT) of the first row $\mathbf{c} = (c_0, c_1, \cdots, c_{2M-1})$ of the matrix Σ_{2M} [39]. Using this eigenvalue decom-

position, an appropriate matrix factorization can then be formulated to generate a finite sample vector $\mathbf{Y}_{2M}$ of size $2M$ using the efficient FFT algorithm, with a computational costs of only $O(2M \log(2M))$ floating-point operations. The covariance structure of this extended finite sample is given by Σ_{2M}, and it is then immediate from the above construction that the first n entries of $\mathbf{Y}_{2M}$ gives the desired finite sample vector $\mathbf{Y}_n$ with the covariance structure Σ_n.

The idea of circulant embedding was believed to be first proposed by Davis and Harte [32] for the purpose of generating finite samples from fractional Gaussian noise processes [76]. The algorithm was, however, only outlined in the appendix of Davis and Hartes' paper, and no further discussion was given therein. Wood and Chan [140] rediscovered exactly the same simulation algorithm from the perspective of circulant embedding, showing that Davis and Harts' formulation is equivalent to using a real-valued factorization based on the eigenvalue decomposition of the embedding matrix Σ_{2M}. Later, another implementation of circulant embedding was proposed, independently, by Dietrich and Newsam [35], in which a complex-valued factorization of the embedding matrix Σ_{2M} was used instead. These two different implementations are outlined in detail in Section(3.2.1). In Section(3.2.2) we review the existing results on embedding strategies, i.e. how the embedding matrix Σ_{2M} can be constructed, and most importantly how to ensure its positive-definiteness. For the purposes of completeness, we also include in Section(3.2.3) a detailed discussion of the spectral domain interpretation of the circulant embedding technique. We will show that both Wood-Chan and Dietrich-Newsam formulations can be regarded as discretizations of the spectral representation of the sampled discrete time stationary processes $Y = Y_k$. We believe this spectral domain interpretation provides intuitive insights into the circulant embedding technique, which is traditionally viewed from a more abstract angle of being a specialized matrix factorization technique.

3.2.1 Implementation of Circulant Embedding

Although based on the same idea of constructing an extended circulant covariance matrix Σ_{2M}, the simulation algorithm for the circulant embedding technique can be formulated in different ways, depending on how Σ_{2M} will be factorized through its eigenvalue decomposition. In this section we will present details of the two implementations for the purpose of future reference.

The earliest Davie and Harte's algorithm, which was later rediscovered by Wood and Chan [140], suggested the following real-valued factorization of Σ_{2M} as

$$\Sigma_{2M} = (F_{2M}^H \Lambda_{2M}^{1/2} F_{2M})(F_{2M}^H \Lambda^{1/2} F_{2M})^T, \tag{3.8}$$

where it could be easily verified that the factor matrix $F_{2M}^H \Lambda_{2M}^{1/2} F_{2M}$ is indeed real-valued. Using this factorization, the extended finite sample $\mathbf{Y}_{2M}$ of size $2M$ can be constructed, from a vector $\mathbf{Z}_{2M}$ of $2M$ IID Gaussian random variables with zero mean and unit variance, through the following transformation:

$$\mathbf{Y}_{2M} = F_{2M}^H \Lambda_{2M}^{1/2} F_{2M} \mathbf{Z}_{2M}. \tag{3.9}$$

In the above Equation the FFT technique has to be applied twice in order to handle matrix multiplications with both F_{2M} and F_{2M}^H. To further save computational costs, Wood and Chan [140] suggested that the complex-valued random vector $\mathbf{W}_{2M} = F_{2M}\mathbf{Z}_{2M}$ in Equation(3.9) can be simulated directly. Let $\{W_0, W_1, \cdots, W_{2M-1}\}$ denotes a set of $2M$ IID Gaussian random variables with zero mean and unit variance, it is then shown by Wood and Chan [140] that the complex-valued random vector $\mathbf{W}_{2M} = F_{2M}\mathbf{Z}_{2M}$ admits the following representation

$$\mathbf{W}_{2M}(j) = \begin{cases} W_0 & \text{for} \quad j = 0, \\ \sqrt{\frac{1}{2}}(W_{2j-1} + iW_{2j}) & \text{for} \quad 1 \leq j < M, \\ W_{2M-1} & \text{for} \quad j = M; \\ \overline{W_{2M-j}} & \text{for} \quad M < j \leq 2M-1, \end{cases} \tag{3.10}$$

(the overbar denotes complex conjugation) and hence can be generated directly. This observation then gives the Wood-Chan simulation formula as

$$\mathbf{Y}_{2M} = F_{2M}^H \Lambda_{2M}^{1/2} \mathbf{W}_{2M}, \tag{3.11}$$

where $\mathbf{W}_{2M}$ is generated directly through Equation(3.10). In this way, the FFT technique is only required to be applied once to calculate the multiplication with the matrix F_{2M}^H, thus further saving computational costs. In subsequent discussions we will be referring to Equation(3.11) as the Wood-Chan implementation. Finally, we mention that this formula, writing in a summation format, is exactly the same as the

original algorithm proposed by Davis and Harte [32, 140].

Later another formulation of circulant embedding algorithm was proposed, independently, by Dietrich and Newsam [35]. Their method factorizes the embedding circulant covariance matrix Σ_{2M} into a complex-valued form as

$$\Sigma_{2M} = (F_{2M}\Lambda_{2M}^{1/2})(F_{2M}\Lambda_{2M}^{1/2})^H, \tag{3.12}$$

Let the complex-valued Gaussian random vector $\mathbf{Z}_{2M}^C$ be defined as

$$\mathbf{Z}_{2M}^C = \mathbf{Z}_{2M}^R + i\mathbf{Z}_{2M}^I, \tag{3.13}$$

where $\mathbf{Z}_{2M}^R$ and $\mathbf{Z}_{2M}^I$ are two independent vectors of $2M$ IID Gaussian random variables with zero mean and unit variance. Dietrich and Newsam [35] suggested generating a complex-valued random vector $\mathbf{Y}_{2M}^C$ of size $2M$ through

$$\mathbf{Y}_{2M}^C = F_{2M}\Lambda_{2M}^{1/2}\mathbf{Z}_{2M}^C, \tag{3.14}$$

in which the multiplication with F_{2M} is performed through the FFT algorithm. It can then be easily shown [35] that both the real-part $\mathrm{Re}(\mathbf{Y}_{2M}^C)$ and the imaginary part $\mathrm{Im}(\mathbf{Y}_{2M}^C)$ form two independent Gaussian random vectors whose covariance matrices are given by Σ_{2M} [35]. Although based on a different factorization of the embedding covariance matrix Σ_{2M}, the Dietrich-Newsam's implementation is essentially equivalent to the Wood-Chan's implementation. The only difference is the fact that, by using a complex-valued factorization, Dietrich-Newsam's implementation generates realizations in multiples of two. However, on a per-realization bases both implementations should be equally efficient, because each realization requires one application of the FFT algorithm acting over a real-valued random vector. Finally we mention that in practice the size $2M$ of embedding matrix Σ_{2M} is usually chosen to be a product of small integers, preferably a power of two, in order to take full advantage of the FFT technique [93, 94]. The construction of the embedding matrix will be discussed in the next section.

3.2.2 Embedding Strategies

The competitiveness of the circulant embedding technique relies on the existence of the embedding circulant covariance matrix Σ_{2M} for which the size $2M$ should not be too much larger than the required sample size n. Dembo, Mallows and Shepp [33] proved that if the original covariance matrix Σ_n is positive definite, then a valid circulant embedding matrix always existed. In practice however a general embedding covariance matrix could be difficult to construct explicitly. For this reason, it is then suggested in literature [35, 94] that, rather than seeking general embedding strategies, we should instead focus on the construction of the so called minimal embedding covariance matrix, which from now on will be denoted by Σ^*_{2M}. Such a matrix is constructed from the $M+1$ terms of the covariance sequences $\{C_Y(0), C_Y(1), \cdots, C_Y(M)\}$ as follows

$$\Sigma^*_{2M} = \begin{bmatrix} C_Y(0) & \cdots & C_Y(M-1) & C_Y(M) & C_Y(M-1) & \cdots & C_Y(1) \\ C_Y(1) & \cdots & C_Y(M-2) & C_Y(M-1) & C_Y(M) & \cdots & C_Y(2) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ C_Y(2) & \cdots & C_Y(M-1) & C_Y(M-2) & C_Y(M-3) & \cdots & C_Y(1) \\ C_Y(1) & \cdots & C_Y(M) & C_Y(M-1) & C_Y(M-2) & \cdots & C_Y(0) \end{bmatrix} \tag{3.15}$$

where we must at least have $M \geq n$. In order for this minimal embedding matrix to be a valid covariance matrix, it is equivalent to requiring its eigenvalues, which is simply the DFT of the first row of the embedding matrix, must at least be nonnegative. As another advantage of using the minimal embedding covariance matrix Σ^*_{2M}, its eigenvalues have an intuitive interpretation, which we will now discuss. Let us denote the eigenvalues of the minimal embedding matrix by $[\Lambda^*_{2M}]_{k,k}$, where Λ^*_{2M} is the corresponding diagonal matrix in its eigenvalue decomposition (see Equation(3.6)). Then it is shown in Percival [93] that these eigenvalues $[\Lambda^*_{2M}]_{k,k}$ can also be expressed in the following simple form

$$[\Lambda^*_{2M}]_{k,k} = C_Y(M)(-1)^k + \int_0^1 D_{2M-1}(\frac{k}{2M} - f) S_Y(f) df, \;\; k = 0, 1, \cdots, 2M-1. \tag{3.16}$$

Here $D_{2M-1}(f)$ is the Dirichlet kernel of order $2M-1$, which is given by

$$D_{2M-1}(f) = \frac{\sin[(2M-1)\pi f]}{\sin(\pi f)}, \tag{3.17}$$

and $S_Y(f)$ is the spectral density of the discrete sampled processes $Y = Y_k$, defined by Equation(2.12) with the assumption that $\Delta_t = 1$. Although $S_Y(f)$ is always non-negative, the Dirichlet kernel $D_{2M-1}(f)$ oscillates between positive and negative values, and also because the term $C_Y(M)(-1)^k$ could be negative, it is then possible that some of the eigenvalues $[\Lambda^*_{2M}]_{k,k}$ may become negative, hence causing the minimal embedding matrix to be negative-definite [93]. On the other hand, it is well known that when acting on sufficiently smooth functions, the Dirichlet kernel $D_{2M-1}(f)$ behaves approximately as a Dirac Delta function for large M, and giving us the following limit (see, for example Bojanic [14] and Zygmund [144]):

$$\lim_{M\to\infty} \int_0^1 D_{2M-1}(f - f')S_Y(f')df = S_Y(f), \text{ uniformly for } f \in [0, 1), \tag{3.18}$$

The above asymptotic results, together with the fact that $C_Y(M)$ tends to zero as M tends to infinity, implies that the eigenvalues $[\Lambda^*_{2M}]_{k,k}$ of the minimal embedding matrix can be regarded as an approximation to the power of the discrete time sampled processes $Y = Y_k$ at frequency $f = \frac{k}{2M}$, i.e. we have

$$\lim_{M\to\infty} [\Lambda^*_{2M}]_{k,k} = S_Y(\frac{k}{2M}), \text{ uniformly in } k. \tag{3.19}$$

This interpretation of the eigenvalues $[\Lambda^*_{2M}]_{k,k}$ hence provides an intuitive framework for the discussion of the positive-definiteness of the minimal embedding matrix Σ^*_{2M}. In particular it has been shown that under the following assumptions over the covariance sequence $\{C_Y(k) : k \in \mathbb{Z}\}$, the minimal embedding matrix Σ^*_{2M} is indeed nonnegative definite:

- The autocovariance sequence $\{C_Y(k) : k \in \mathbb{Z}\}$ is no-negative, decreasing and convex for non-zero lags, then the embedding circulant matrix Σ^*_{2M} is non-negative definite for all M [35]; such a autocovariance sequence can be obtained from discretizing the exponential covariance model [132]

- The autocovariance sequence $\{C_Y(k) : k \in \mathbb{Z}\}$ vanishes for all lags $k \geq k_0$, then for all $M \geq k_0$ the minimal embedding matrix Σ^*_{2M} is nonnegative definite [35]; examples of such a autocovariance sequence arises from discretizing the spherical covariance function [60] and power covariance function (see Warnes [135])

- If the autocovariance sequence $\{C_Y(k) : k \in \mathbb{Z}\}$ is non-positive for $k \neq 0$, then

the embedding circulant matrix Σ^*_{2M} is non-negative definite for all $M > 0$; for example, a fractional differenced Gaussian noise with difference parameter $d \in [-1/2, 0)$ gives rise to such a negative autocovariance sequence, and hence can be conveniently generated by using the circulant embedding technique [25].

All these results on positive-definiteness of Σ^*_{2M} are sufficient rather than necessary, and address particular cases rather than being generally applicable. There exists abundant covariance models that are in frequent use, but their discretization do not give autocovariance sequences satisfying the above listed conditions. Examples include the Gaussian and Matérn covariance models, the latter of which is a frequently used covariance model in Geostatistics [86]. To deal with more general situations in which the above listed conditions are not met, several methods have been proposed in the literature.

The most straightforward method starts from the additional assumption that the spectral density $S_Y(f)$ is strictly positive. Then many authors [93, 140, 35] suggested that a simple procedure which is based on the first principal of increasing M until the matrix Σ^*_{2M} is positive definite. This simple method works because, as shown in Equation(3.19), the eigenvalues $[\Lambda^*_{2M}]_{k,k}$ of the minimal embedding matrix converge uniformly to $S_Y(\frac{k}{2M})$ as M increases. Hence, if $S_Y(f)$ is strictly positive for all $f \in [0, 1]$, it is then expected that the eigenvalues $[\Lambda^*_{2M}]_{k,k}$ should all be strictly positive as long as M is large enough. The positivity of the eigenvalues can be easily checked by taking the FFT of the first row of the minimal embedding matrix [93, 94]. The size $2M$ of the minimal embedding matrix is usually chosen to be a power of two in order to take full advantage of the FFT technique [93, 94].

For most of the practical simulation scenarios, this first principle based simple approach should work without producing a very large size $2M$ of the embedding covariance matrix [35]. There are, however, other situations in which the required size $2M$ of the embedding covariance matrix could be excessively large. This could happen if the target covariance matrix Σ_n of the desired finite sample is ill-conditioned as a result of, for example, the covariance sequence $C_Y(k)$ being obtained through discretizing a smooth covariance function at relatively small intervals [103, 91]. In order to deal with these extreme situations, Wood and Chan [140] proposed an approximate embedding procedure that does not require the minimal embedding matrix Σ_{2M} to be nonnegative definite: simply set negative eigenvalues to zeros, and appropriately rescale the positive eigenvalues (so that the simulated processes has the desired vari-

ance). More recently, Stein [123] proposed an alternative exact method of cut-off embedding (and intrinsic embedding for generating intrinsic stationary processes), which is based on modifying the covariance function $R_X(\tau)$ of the underlying continuous time stationary processes, so that it has a compact support [45]. Stein's idea was studied in detail by Gneiting et al [46] in which they showed that the cut-off embedding could outperform the standard minimal embedding under certain circumstances, but it is not quite generally applicable, because an appropriate intrinsic embedding could be difficult to construct for general covariance functions. Moreover Gneiting et al [46] also pointed out that the smoother the processes, the harder it is to find the appropriate compactly supported covariance function, because a covariance function that is smooth at the origin requires even greater smoothness elsewhere(see also for example Stein [122] and Gneiting [44]).

3.2.3 Spectral Domain Interpretation of Circulant Embedding

The circulant embedding technique has been considered in existing literature as a variation of the time domain simulation method based on efficient matrix factorization. What makes it more interesting, however, is the fact that when the minimal embedding covariance matrix Σ^*_{2M} is used, the circulant embedding technique also admits a spectral domain interpretation. In particular, under the minimal embedding strategy, Percival [93] implicitly showed that the Wood-Chan's real-valued formulation can be written as

$$\mathbf{Y}_{2M}(j) = \sum_{k=0}^{M} \cos(2\pi jk/2M) d\hat{U}_Y(k) + \sum_{k=0}^{M} \sin(2\pi jk/2M) d\hat{V}_Y(k), \tag{3.20}$$

where the random variables $d\hat{U}_Y(k)$ and $d\hat{V}_Y(k)$, appropriately defined for each implementations, are uncorrelated Gaussian random variables, with zero mean and variances given by

$$\mathrm{Var}[d\hat{U}_Y(k)] = \mathrm{Var}[d\hat{V}_Y(k)] = \begin{cases} \frac{1}{2M}[\Lambda^*_{2M}]_{0,0} & \text{for} \quad k = 0 \\ \frac{1}{M}[\Lambda^*_{2M}]_{k,k} & \text{for} \quad 0 < k < M \\ \frac{1}{2M}[\Lambda^*_{2M}]_{M,M} & \text{for} \quad k = M \end{cases} \tag{3.21}$$

It has already been discussed in Section(3.2.2) that the eigenvalues $[\Lambda^*_{2M}]_{k,k}$, under minimal embedding, approximates $S_Y(\frac{k}{2M})$. Therefore, it is easy to recognize the fact

that these random variables $d\hat{U}_Y(k)$ and $d\hat{V}_Y(k)$ can be regarded as approximations (in a probabilistic sense) to the increments of the Gaussian processes $U_Y(f)$ and $V_Y(f)$, which appear in Equation(2.14) of the spectral representation of the discrete time stationary process $Y = Y_k$. On the other hand, using the same trick it can be easily established that the Dietrich-Newsam's complex-valued formulation can also be re-expressed in the form of Equation(3.20). Being motivated by this spectral domain interpretation, Percival [93] proposed an approximate spectral simulation method for discrete time stationary processes, by replacing $[\Lambda^*_{2M}]_{k,k}$ with $S_Y(\frac{k}{2M})$. In next section, we will discuss this spectral simulation method for continuous time stationary processes.

3.3 Spectral Simulation Method

In this section we give a detailed discussion of the spectral simulation method, which utilizes the knowledge of the spectral density function $S_X(\lambda)$, instead of the covariance function $R_X(\tau)$, of the underlying continuous time stationary Gaussian process. Unless explicitly stated, the spectral density function $S_X(\lambda)$ will be assumed to be continuous for $\lambda \in \mathbb{R}$. To start the discussion we first recall that the spectral simulation method, which was briefly introduced in Section(3.1), discretizes the spectral representation of the underlying process $X = X(t)$. The discretization involves first truncating the positive spectral domain $[0, \infty)$ into a compact interval $[0, L]$, and then dividing into a grid $\{\lambda_0, \lambda_1, \cdots, \lambda_m\}$, where $\lambda_0 = 0$ and $\lambda_m = L$. The simulated realization $\hat{X} = \hat{X}(t)$, given by Equation(3.4), takes the form of an aggregation of sinusoids with random amplitudes $d\hat{U}_X(k)$ and $d\hat{V}_X(k)$.

The statistical properties of the simulated realization is produced through these random amplitudes. They can be regarded as random variables modeling the probabilistic properties of the orthogonal increments $dU_X(\lambda_k)$ and $dV_X(\lambda_k)$ of the underlying process over the spectral grid. In other words, we have the following approximations in a probabilistic sense:

$$d\hat{U}_X(k) \approx dU_X(\lambda_k), \;\; d\hat{V}_X(k) \approx dV_X(\lambda_k), \;\; k = 0, 1, \cdots, m. \tag{3.22}$$

For the purpose of creating similar covariance structures, the above probabilistic approximation requires that $d\hat{U}_X(k)$s and $d\hat{V}_X(k)$s are mutually uncorrelated, and also

formally match the second order properties of the increments $dU_X(\lambda_k)$ and $dV_X(\lambda_k)$. This match can be achieved through imposing the following second order conditions:

$$\mathrm{E}[d\hat{U}_X(k)] = \mathrm{E}[d\hat{V}_X(k)] = 0, \;\; k = 0, 1, \cdots, m, \tag{3.23}$$

and

$$\mathrm{Var}[d\hat{U}_X(k)] = \mathrm{Var}[d\hat{V}_X(k)] = S_X(\lambda_k)d_k \approx S_X(\lambda_k)d\lambda_k, \;\; k = 0, 1, \cdots, m, \tag{3.24}$$

where $\{d_1, d_2, \cdots, d_m\}$ are scaling constant compensating for the discretization of the spectral domain, in other words d_k can be regarded as an approximation of the frequency increments $d\lambda_k$. Using these imposed second order properties, an easy calculation shows that the covariance function $R_{\hat{X}}(\tau)$ of the simulated process $\hat{X} = \hat{X}(t)$ can be written as

$$R_{\hat{X}}(\tau) = \sum_{k=0}^{m} \cos(2\pi\lambda_k\tau)S_X(\lambda_k)d_k, \tag{3.25}$$

which is simply in the form of a quadrature approximation to the target covariance function $R_X(\tau)$. Therefore, from a second order point of view, how well the simulated process $\hat{X} = \hat{X}(t)$ resembles the target process $X = X(t)$ can be measured by the properties of the quadrature approximation $R_{\hat{X}}(\tau)$ to the target covariance function $R_X(\tau)$. The properties of this approximation are determined by the power cut-off frequency L, the spectral domain discretization grid $\{\lambda_0, \lambda_1, \cdots, \lambda_m\}$ as well as the scaling constants $\{d_0, d_1, \cdots, d_m\}$.

A common choice most frequently used in practice is an equally spaced spectral domain discretization grid [115, 47], given explicitly by

$$\lambda_k = k\Delta_\lambda, \;\; k = 0, 1, \cdots, m, \tag{3.26}$$

where Δ_λ is the equally spaced spectral discretization interval. Moreover the corresponding scaling constants d_k are chosen in such a way so that the variances of the random variables $d\hat{U}_X(k)$s and $d\hat{V}_X(k)$ satisfy

$$\mathrm{Var}[d\hat{U}_X(k)] = \mathrm{Var}[d\hat{V}_X(k)] = \begin{cases} \Delta_\lambda S_X(k\Delta_\lambda), & \text{for } k = 0, \\ 2\Delta_\lambda S_X(k\Delta_\lambda), & \text{for } k = 1, 2, \cdots, m. \end{cases} \tag{3.27}$$

With such a choice, the covariance function of the simulated process can be easily verified to take the following form

$$R_{\hat{X},Trap}(\tau) = \Delta_\lambda \left[S_X(0) + \sum_{k=1}^{m} 2S_X(k\Delta_\lambda)\cos(2\pi k\Delta_\lambda \tau) \right]. \tag{3.28}$$

The subscript $''Trap''$ indicates the fact that this quadrature approximation, written in complex-exponential form, is in fact a Trapezoidal quadrature approximation to the target covariance function $R_X(\tau)$ through its complex-valued Fourier representation in Equation(2.2), over the two-sided finite frequency domain $[-L, L]$. Because of its popularity in existing literature [114, 116, 115], in subsequent discussions this particular choice of frequency domain discretization will be referred to as the Trapezoidal discretization.

The historical preference towards the Trapezoidal discretization can be justified for two reasons. First of all, under the frequently considered scenario of equally spaced sampling times, many authors [141, 142, 115] suggested that the spectral simulation method has modifications so that the efficient FFT technique can be applied. Detailed discussion of the implementation of the FFT technique can be found in Shinozuka and Deodatis [115]. On the other hand, although this thesis focuses on the case of unequally (randomized) sampling times and hence the computational advantage of the FFT technique is no longer available, we still find the above Trapezoidal discretization useful in terms of analyzing the discrepancies between $R_{\hat{X},Trap}(\tau)$ and $R_X(\tau)$. We will see in Section(3.4) that such an analysis will be extremely useful in determining the appropriate Δ_λ in the Trapezoidal discretization.

Having said that, it is still interesting to ask the question of whether any frequency domain discretization is feasible. We find that the clue to this question is closely related to the time domain aliasing effect introduced from the equally spaced griding in the Trapezoidal discretization. This aliasing effect causes the covariance function $R_{\hat{X},Trap}(\tau)$ to be periodic with period $T_0 = \frac{1}{\Delta_\lambda}$. Consequently the discrepancies between $R_{\hat{X},Trap}(\tau)$ and $R_X(\tau)$ increases as τ moves away from zero.

Now suppose we keep the equally spaced spectral discretization grid, and modify the scaling constants so that $R_{\hat{X}}(\tau)$ approximates $R_X(\tau)$ through a higher-order Newton-Cotes quadrature scheme (e.g. a Simpson's Rule, etc) [124]. Although being possible, we found that such a choice will not provide practical advantage because of the secondary aliasing that will be introduced from higher-order Newton-Cotes

quadrature approximations [3]. To put it simply, higher-order Newton-Cotes quadrature schemes will introduce a secondary periodicity much shorter than that introduced from the Trapezoidal quadrature approximations. This in particular implies that the discrepancies between $R_{\hat{X}}(\tau)$ and $R_X(\tau)$ increases much faster than otherwise when a Trapezoidal approximation is used. A graphical illustration of the secondary aliasing introduced from Simpson's quadrature approximation is given in Figure(3.1).

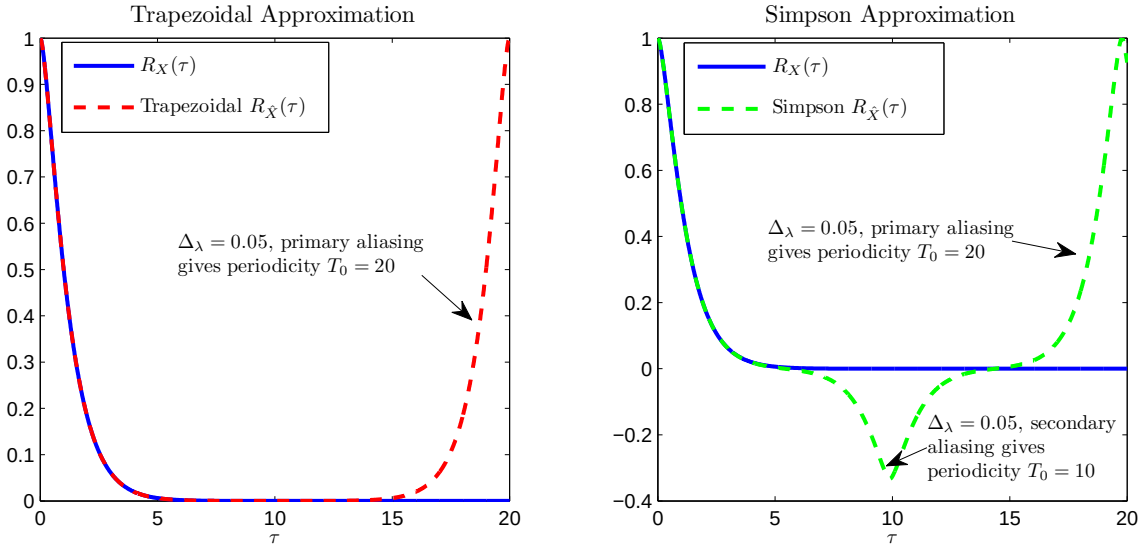

Figure 3.1: Illustration of secondary aliasing introduced from Simpson's quadrature scheme, using a Matèrn covariance function with parameters $\sigma = 1, \phi = 1, \nu = 1$ and a spectral discretization interval $\Delta_\lambda = 0.05$.

Alternatively, we could borrow results from signal processing literature and choose a randomized frequency domain discretization grid, designed to reduce the aliasing effect introduced from equally spaced spectral griding [129, 82, 85]. Apart from the extra computational costs involved in generating the randomized spectral grid, the resulting covariance function $R_{\hat{X}}(\tau)$ will be subject to more significant statistical variations as τ increases, making it very difficult to control the discrepancies between $R_{\hat{X}}(\tau)$ and $R_X(\tau)$.

Therefore in the remaining discussion of this section on spectral simulation, we will only be focusing on simulation method based on equally spaced spectral discretization, such that the covariance function $R_{\hat{X}}(\tau)$ of the simulated process approx-

imates the target covariance function $R_X(\tau)$ through a Trapezoidal quadrature rule of Equation(3.28). Under this particular choice, however, there will still be two different ways, in the existing literature, to specify the uncorrelated random variables $d\hat{U}_X(k)$ and $d\hat{V}_X(k)$, and will lead to slightly different statistical properties for the simulated realization.

3.3.1 Rice's Random Phase Deterministic Amplitude (RPDA) Formula

In engineering literature there is a popular and widely used method, known as the Random Phase Deterministic Amplitude(RPDA) method. It is based on the concept proposed by Rice in his famous paper [105]. Although the idea of the RPDA method has existed for some time, it was Shinozuka and Jan [114, 116] who first applied it for the purpose of simulating stationary Gaussian processes, including the multi-dimensional case. An excellent review of the properties of this RPDA method can be found in Shinuzuka and Deodatis [115].

Under the RPDA specifications, the random variables $d\hat{U}_X(k)$ and $d\hat{V}_X(k)$, which in this case will be denoted as $d\hat{U}_X^{RPDA}(k)$ and $d\hat{V}_X^{RPDA}(k)$ for the purpose of differentiation, are specified in Rice's RPDA formula as

$$d\hat{U}_X^{RPDA}(k) = \begin{cases} \sqrt{2S_X(0)\Delta_\lambda}\cos(\theta_0), & \text{for } k=0, \\ \sqrt{4S_X(k\Delta_\lambda)\Delta_\lambda}\cos(\theta_k), & \text{for } k=1,2,\cdots,m, \end{cases} \tag{3.29}$$

and

$$d\hat{V}_X^{RPDA}(k) = \begin{cases} -\sqrt{2S_X(0)\Delta_\lambda}\sin(\theta_0), & \text{for } k=0, \\ -\sqrt{4S_X(k\Delta_\lambda)\Delta_\lambda}\sin(\theta_k), & \text{for } k=1,2,\cdots,m, \end{cases} \tag{3.30}$$

where $\{\theta_0,\cdots,\theta_m\}$ are IID random variables that are uniformly distributed over $(0,2\pi)$. Putting Equations(3.29) and (3.30) back into Equation(3.4), we can then, with the help of elementary trigonometric identities, write the realization generated by RPDA method, denoted by $\hat{X}_{RPDA} = \hat{X}_{RPDA}(t)$, as follows

$$\hat{X}_{RPDA}(t) = \sqrt{2S_X(0)\Delta_\lambda}\cos(\theta_0) + \sum_{k=1}^{m}\sqrt{4S_X(k\Delta_\lambda)\Delta_\lambda}\cos(2\pi k\Delta_\lambda t + \theta_k). \tag{3.31}$$

Note that $\hat{X}_{RPDA} = \hat{X}_{RPDA}(t)$ consists of a sum of harmonics with random phase and deterministic amplitude, hence comes the name. Looking back at Equation(3.29) and (3.30), a simple calculation shows that $d\hat{U}_X^{RPDA}(k)$ and $d\hat{V}_X^{RPDA}(k)$ are indeed mutually uncorrelated random variables with zero mean and variances given by Equation(3.27). Consequently the covariance function of the simulated realization is given by $R_{\hat{X},Trap}(\tau)$ in Equation(3.28).

Since the stationary process $X = X(t)$ we want to simulate is assumed to be Gaussian, it is known that the orthogonal increments $dU_X(\lambda)$ and $dV_X(\lambda)$ in the spectral representation are also Gaussian [117]. However, one distinctive feature of the Rice's RPDA formula is the fact that the random variables $d\hat{U}_X^{RPDA}(k)$ and $d\hat{V}_X^{RPDA}(k)$ do not attempt to model the Gaussianity of the increments $dU_X(\lambda_k)$ and $dV_X(\lambda_k)$, respectively. It is only through the application of the Central Limit Theorem that the simulated process $\hat{X}_{RPDA} = \hat{X}_{RPDA}(t)$ can achieve Gaussianity in an asymptotic sense [142, 115, 47]. Although with this drawback, the RPDA method is still popular in engineering literature [10, 11]. One of the reasons is that the asymptotic Gaussianity should be relatively easy to achieve for most of the spectral models used in practice [115, 47]. Moreover, with a slight modification the sample function generated from RPDA method enjoys strong ergodicity - a property that is believed to be valuable in engineering applications [115]. Section(3.3.3) will have a more detailed discussion on both the asymptotic Gaussianity and ergodicity issues.

3.3.2 Random Phase Random Amplitude(RPRA) Formula

Another spectral simulation method, called the Random Phase Random Amplitude(RPRA) method, was also studied in considerable details by various authors: for example Grigoriul [47], Sun and Chaika [125] discussed the simulation of continuous time stationary Gaussian processes, whereas Percival [93] discussed the same method for generating finite samples from discrete time stationary Gaussian processes. The random variables $d\hat{U}_X(k)$ and $d\hat{V}_X(k)$, which in this case are denoted by $d\hat{U}_X^{RPRA}(k)$ and $d\hat{V}_X^{RPRA}(k)$, are defined as follows

$$d\hat{U}_X^{RPRA}(k) = \begin{cases} \sqrt{S_X(0)\Delta_\lambda}N_0^U, & \text{for } k = 0, \\ \sqrt{2S_X(k\Delta_\lambda)\Delta_\lambda}N_k^U, & \text{for } k = 1, 2, \cdots, m, \end{cases} \tag{3.32}$$

and

$$d\hat{V}_X^{RPRA}(k) = \begin{cases} \sqrt{S_X(0)\Delta_\lambda}N_0^V, & \text{for } k=0, \\ \sqrt{2S_X(k\Delta_\lambda)\Delta_\lambda}N_k^V, & \text{for } k=1,2,\cdots,m, \end{cases} \tag{3.33}$$

where $\{N_k^U, N_k^V : k = 0, 1, \cdots, m\}$ are IID Gaussian random variables with zero mean and unit variance. Therefore, the realization generated from RPRA formula, denoted by $\hat{X}_{RPRA} = \hat{X}_{RPRA}(t)$, can be written in the following form

$$\hat{X}_{RPRA}(t) = \sqrt{S_X(0)\Delta_\lambda}N_0^U + \sum_{k=1}^{m}\sqrt{2S_X(k\Delta_\lambda)\Delta_\lambda}\left[\cos(2\pi k\Delta_\lambda t)N_k^U + \sin(2\pi k\Delta_\lambda t)N_k^V\right]. \tag{3.34}$$

This simulation formula however is seldom implemented in practice because it can be further simplified for the purpose of saving computational costs [115]. This is achieved through writing the terms in the square brackets of Equation(3.34) as

$$\cos(2\pi k\Delta_\lambda)N_k^U + \sin(2\pi k\Delta_\lambda)N_k^V = D_k\cos(2\pi k\Delta_\lambda t\theta_k). \tag{3.35}$$

Here the random amplitudes

$$D_k = \sqrt{(N_k^U)^2 + (N_k^V)^2} \tag{3.36}$$

are IID Rayleigh random variables, and the random phase variables

$$\theta_k = \tan^{-1}(N_k^U/N_k^V) \tag{3.37}$$

are IID uniform $(0, 2\pi)$ random variables. With these newly defined quantities, the RPRA simulation formula can also be written in a more compact form as

$$\hat{X}_{RPRA}(t) = \sqrt{S_X(0)\Delta_\lambda}N_0^U + \sum_{k=1}^{m}\sqrt{2S_X(k\Delta_\lambda)\Delta_\lambda}D_k\cos(2\pi k\Delta_\lambda t + \theta_k), \tag{3.38}$$

which involves a sum of harmonics with random phases and random amplitudes, hence comes the name. Although Equation(3.34) and (3.38) are equivalent, in practice Equation(3.38) will be more computationally efficient, since it involves a smaller number of sinusoidal functions, which is known to be computationally more expensive to evaluate than simple algebraic operations [124].

The RPRA method aims to provide a better approximation to the spectral rep-

resentation of the target process. In fact apart from matching the second order properties in Equation(3.27), the random variables $d\hat{U}_X^{RPRA}(k)$s and $d\hat{V}_X^{RPRA}(k)$ also explicitly models the Gaussianity of the increments $dU_X(\lambda_k)$ and $dV_X(\lambda_k)$ of the orthogonal processes over the equally spaced spectral grid. Hence the realization $\hat{X}_{RPRA} = \hat{X}_{RPRA}(t)$ generated by the RPRA formula is exactly Gaussian, and has covariance function given by $R_{\hat{X},Trap}(\tau)$. Although having the desired exact Gaussianity, the RPRA formula seems to be less popular in existing literature, perhaps for the reason that the sample functions do not have the strong ergodicity property [115, 47]. Section(3.3.3) will provide a detailed comparison between the RPDA and RPRA simulation methods.

3.3.3 Comparison Between RPDA and RPRA Methods

Efficiency and Gaussianity

To compare these two simulation methods, we first notice that RPDA formula given by Equation(3.31) should be slightly more efficient since it involves generating only $m+1$ uniformly distributed random variables, whereas the RPRA method of Equation(3.38) needs to generate an extra m Rayleigh random variables. However, numerical experiments conducted in Section(3.5.4) shows that this computational efficiency of the RPDA formula is not significant. On the other hand, the RPDA method has a potentially significant drawback: the simulated process $\hat{X}_{RPDA} = \hat{X}_{RPDA}(t)$ is not exactly Gaussian, as compared to the exact Gaussianity of $\hat{X}_{RPRA} = \hat{X}_{RPRA}(t)$ generated by RPRA method. For stationary Gaussian processes with a continuous spectral density $S_X(\lambda)$ this should not be a serious problem. This is because the RPDA formula consists of a sum of independent random variables, and it has been proved that the simulated process $\hat{X}_{RPDA} = \hat{X}_{RPDA}(t)$ is convergent in distribution to a Gaussian process [1] as the number m of grid points tends to infinity (see for example Yang [142], Shinozuka and Deoadatis [142, 115], Grigoriu [47] for more details). Numerical studies have also been performed in the indicated references and suggested that, independent of the particular form of the spectral density $S_X(\lambda)$, the distributional properties of the realization $\hat{X}_{RPDA} = \hat{X}_{RPDA}(t)$ is sufficiently close to the true

[1] The simulated process $\hat{X} = \hat{X}(t)$ converging to a Gaussian process, in the sense that for any finite set of sampling times $\{t_1, \cdots, t_n\}$, the vector $(\hat{X}(t_1), \cdots, \hat{X}(t_n))^T$ converges in distribution to a multivariate Gaussian vector.

Gaussian distribution of the target process $X = X(t)$, as long as the number m of terms is large enough (say $m \geq 1000$).

These remarks consequently indicate that it is not generally justified to reject the use of RPDA method because of its non-Gaussianity. One exception, however, occurs when the target process has a discrete spectral density of the form

$$S_X(\lambda) = \sum_{j=1}^{J} S_j \delta(\lambda - \lambda_j), \tag{3.39}$$

where S_j are positive real numbers and $\delta(\lambda)$ is the Dirac Delta function. In this case, it has been pointed out by Grigoriu [47] that $\hat{X}_{RPDA} = \hat{X}_{RPDA}(t)$ may not provide satisfactory approximation to the target Gaussian process because the RPDA simulation formula contains at most J terms, which may not be large enough to ensure the asymptotic Gaussianity through the Central Limit Theorem [47].

Ergodicity

A stochastic process is said to be ergodic relative to some ensemble properties if these properties can be obtained from taking temporal averages over a single realization [26, 100]. For a zero-mean stationary Gaussian process $G = G(t)$ with covariance function $R_G(\tau)$, the two most important ergodic properties are relative to its ensemble mean and covariance [100, 115]. To be more specific, defining the temporal mean $M_{G,T}$ and covariance $R_{G,T}(\tau)$ as follows:

$$M_{G,T} = \frac{1}{T} \int_0^T G(t) dt, \tag{3.40}$$

$$R_{G,T}(\tau) = \frac{1}{T} \int_0^T G(t+\tau) G(t) dt, \tag{3.41}$$

where the subscripts $''G''$ and $''T''$ in both $M_{G,T}$ and $R_{G,T}(\tau)$ indicate the temporal averages are taken with respect to the process G over an averaging period of T. Then the stationary Gaussian process $G = G(t)$ is said to be ergodic with respect to mean and covariance if

$$\lim_{T\to\infty} M_{G,T} = 0, \text{ and } \lim_{T\to\infty} R_{G,T}(\tau) = R_G(\tau). \tag{3.42}$$

Another distinction between the RPDA and RPRA simulation method that has been discussed in engineering literature [115, 47] concerns the mean and covariance ergodic properties of the realizations. Those existing literature showed a preferences towards the use of RPDA formula, for the reasons that, when the power at zero frequency $S_X(0)$ being set to zero, the realizations generated from the RPDA formula of Equation(3.31) are mean and covariance ergodic in the strong sense (i.e. almost surely). More specifically it is derived in Shinozuka and Deodatis [115] that

$$\lim_{T\to\infty} M_{\hat{X}_{RPDA},T} = \sqrt{2S_X(0)\Delta_\lambda}\cos(\theta_0), \tag{3.43}$$

and

$$\lim_{T\to\infty} R_{\hat{X}_{RPDA},T}(\tau) = \Delta_\lambda[S_X(0)\cos^2(\theta_0) + \sum_{k=1}^{m} 2S_X(k\Delta_\lambda)\cos(2\pi k\Delta_\lambda\tau)]. \tag{3.44}$$

By imposing the condition that $S_X(0) = 0$, it was then argued that we have [115]

$$\lim_{T\to\infty} M_{\hat{X}_{RPDA},T} = 0, \text{ almost surely}, \tag{3.45}$$

and

$$\lim_{T\to\infty} R_{\hat{X}_{RPDA},T}(\tau) = R^0_{\hat{X},Trap}(\tau), \text{ almost surely}. \tag{3.46}$$

Here $R^0_{\hat{X},Trap}(\tau)$ is defined as the covariance function implied from the RPDA method, under the additional assumption that $S_X(0) = 0$, i.e.

$$R^0_{\hat{X},Trap}(\tau) = \Delta_\lambda \sum_{k=1}^{m} 2S_X(k\Delta_\lambda)\cos(2\pi k\Delta_\lambda\tau). \tag{3.47}$$

On the other hand, it has also been shown [115, 47] that the temporal mean $M_{\hat{X}_{RPRA},T}$ of the realization generated from RPRA method satisfies

$$\lim_{T\to\infty} M_{\hat{X}_{RPRA},T} = \sqrt{S_X(0)\Delta_\lambda}N_0^A, \tag{3.48}$$

where N_0^A is a normal random variable with zero mean and unit variance. The corresponding expression for the temporal covariance function $R_{\hat{X},T}(\tau)$ is given by [115, 47]

$$\lim_{T\to\infty} R_{\hat{X}_{RPRA},T}(\tau) = \Delta_\lambda[S_X(0)(N_0^A)^2 + \sum_{k=1}^{m} S_X(k\Delta_\lambda)D_k^2\cos(2\pi k\Delta_\lambda\tau)], \tag{3.49}$$

where D_ks are IID Rayleigh random variables, so that D_k^2s are IID Chi-square random variables with 2 degrees of freedom. Therefore it is immediate from Equation(3.48) that, under the condition $S_X(0) = 0$ the realization generated from RPRA method is also mean ergodic in a strong sense, i.e.

$$\lim_{T\to\infty} M_{\hat{X}_{RPRA},T} = 0, \text{ almost surely.} \tag{3.50}$$

However, because of the appearance of the Chi-square random variables, we will not be able to obtain the covariance ergodicity in strong sense, regardless of the value of $S_X(0)$. In other words

$$\lim_{T\to\infty} R_{\hat{X}_{RPRA},T}(\tau) \neq \begin{cases} R^0_{\hat{X},Trap}(\tau) & \text{for} \quad S_X(0) = 0 \\ R_{\hat{X},Trap}(\tau) & \text{for} \quad S_X(0) \neq 0 \end{cases}. \tag{3.51}$$

Therefore each and average sample function generated from RPRA method is not covariance ergodic in the strong sense [47].

Although the strong ergodicity property of the RPDA realization, obtained from the additional assumption of $S_X(0) = 0$, is theoretically desirable [115], the RPRA simulation method should not be completely rejected based on the absence of this strong ergodicity. In fact, if we use the notation $\mathrm{l.i.m}_{T\to\infty}$ to represent the operation of taking mean-square limit, then it has been shown by Grigoriu [47] that, regardless of the value of $S_X(0)$, the RPRA realization enjoy the following ergodicity properties in a weaker sense

$$\underset{T\to\infty}{\mathrm{l.i.m}}\, M_{\hat{X}_{RPRA},T} = 0, \tag{3.52}$$

and

$$\underset{T\to\infty}{\mathrm{l.i.m}}\, R_{\hat{X}_{RPRA},T}(\tau) = R_{\hat{X},Trap}(\tau). \tag{3.53}$$

Based on this weaker ergodicity, Grigoriu [47] then concluded that the RPRA realization should be regarded as being sufficiently ergodic in both mean and covariance.

The above arguments about the ergodicity of realizations, however, have drawbacks. For one thing, they do not establish a direct link with the ergodicity of the target process, i.e. the covariance ergodicity of both the RPDA and RPRA realizations (i.e. Equation(3.46) and (3.53)) are not with respect to the target covariance function $R_X(\tau)$. Moreover, these arguments are only valid in the asymptotic sense when $T \to \infty$, whereas in practice realizations of only finite length will be required.

In other words, there has been no direct evidence showing that both the RPDA and the RPRA realizations can give good approximation to the ergodicity properties of the target process, under the more practical setting of finite sample length. The discussions presented in the next subsection, however, addresses this issue and therefore can be regarded as a complement to the arguments given by Grigoriu [47].

Statistical Properties of Temporal Averages

In this section, as a complement to the arguments given in Grigoriu [47], we will show that for a fixed but finite averaging period T, the means and variances of the temporal averages $M_{\hat{X}_{RPRA},T}$ and $R_{\hat{X}_{RPRA},T}(\tau)$ will be close to the means and variances of the same temporal averages of the target process, as long as the discrepancy between $R_X(\tau)$ and $R_{\hat{X},Trap}(\tau)$ are controlled at a small level. We will borrow some intermediate results from Grigoriu [47], which are summarized in the following lemma:

Lemma 3.1. *Suppose $G = G(t)$ is a continuous time stationary Gaussian process, with temporal averages $M_{G,T}$ and $R_{G,T}(\tau)$ defined through Equation(3.40) and (3.41), respectively. Then the random variable $M_{G,T}$ has mean and variance given by*

$$E[M_{G,T}] = 0, \tag{3.54}$$

$$Var[M_{G,T}] = \frac{2}{T}\int_0^T (1 - \frac{\tau}{T})R_G(\tau)d\tau. \tag{3.55}$$

On the other hand, the random variable $R_{G,T}(\tau)$ has mean and variance given by

$$E[R_{G,T}(\tau)] = R_G(\tau), \tag{3.56}$$

$$Var[R_{G,T}(\tau)] = \frac{2}{T}\int_0^T (1 - \frac{t}{T})[R_G(t)^2 + R_G(t+\tau)R_G(t-\tau)]dt, \tag{3.57}$$

The proof of these results can be found in Grigoriu [47] and therefore will be omitted. However, it has to be emphasized that these exact expressions depend crucially on the Gaussianity of the process $G = G(t)$. Because the RPRA realization is exactly Gaussian with covariance function given by $R_{\hat{X},Trap}(\tau)$, we can then use the above lemma to derive the following result:

Proposition 3.1. *For any $t > 0$, Define the discrepancy $\epsilon_{\hat{X}}(\tau)$ between $R_X(\tau)$ and*

$R_{\hat{X},Trap}(\tau)$ *over* $\tau \in [0,t]$ *as*

$$\epsilon_{\hat{X}}(\tau) = R_X(\tau) - R_{\hat{X},Trap}(\tau). \tag{3.58}$$

Then regardless of the value of $S_X(0)$*, it can be shown that for the temporal mean* $M_{\hat{X}_{RPRA},T}$ *we have*

$$E[M_{X,T}] = E[M_{\hat{X}_{RPRA},T}] = 0 \tag{3.59}$$

and

$$|\,Var[M_{X,T}] - Var[M_{\hat{X}_{RPRA},T}]| \leq \max_{\tau\in[0,T]} |\epsilon_{\hat{X}}(\tau)|. \tag{3.60}$$

On the other hand, for the temporal covariance $R_{\hat{X}_{RPRA},T}(\tau)$ *we have*

$$E[R_{X,T}(\tau)] - E[R_{\hat{X}_{RPRA},T}(\tau)] \leq \epsilon_{\hat{X}}(\tau), \tag{3.61}$$

and

$$|\,Var[R_{X,T}(\tau)] - Var[R_{\hat{X}_{RPRA},T}(\tau)]| \leq 2 \max_{\tau\in[0,T+\tau]} |\epsilon_{\hat{X}}(\tau)|[R_X(0) + R_{\hat{X},Trap}(0)]. \tag{3.62}$$

The proof of this proposition is a straightforward application of Lemma(3.1) and hence is left in Section(8.1) in the appendix of this thesis. Therefore as an immediate consequence of this result, when the discrepancy $\epsilon_{\hat{X}}(\tau)$ between $R_X(\tau)$ and $R_{\hat{X},Trap}(\tau)$ is controlled at a small level for all τ within some appropriate range (which depends on T and τ), we will then be guaranteed that the statistical behaviour of the temporal averages $M_{\hat{X}_{RPRA},T}$ and $R_{\hat{X}_{RPRA},T}(\tau)$ are close to those of the target process. Hence it provides further and more direct evidence, from a finite sample point of view, to justify the use of the RPRA simulation method, regardless of its lack covariance ergodicity in strong sense.

On the other hand, because of the lack of exact Gaussianity, it will be difficult if not impossible to prove the above results for the RPDA realization. However, as pointed out by Grigoriu [47], Lemma(3.1) should also hold approximately for the RPDA realization, because as discussed in Section(3.3.3) the RPDA realization approaches Gaussianity as $m \to \infty$. Consequently the results in Proposition(3.1) and the subsequent remarks together imply the fact that in practical situations (i.e. finite sample length) it is more important to control the discrepancy between $R_X(\tau)$ and $R_{\hat{X},Trap}(\tau)$ for both the RPDA and the RPRA simulation methods. Setting $S_X(0) = 0$

will not be helpful in this regard, because as will be discussed in Section(3.5.3), this traditional practice will tend to produce larger magnitude for $\epsilon_{\hat{X}}(\tau)$. Therefore based on the arguments given in this subsection, we believe that in the practical situations when only finite sample length is required, the traditional practice of setting $S_X(0) = 0$ should not be preferred.

3.3.4 Summary of RPDA and RPRA Methods

In this section we discussed and compared two spectral simulation methods, namely the RPDA and RPRA methods, and their properties are summarized in Table(3.3). Based on the arguments presented in this section(and in the indicated references), we concluded that in most practical situations both methods will be able to provide an adequate approximation to the target process. Cautions however have to be taken when the target process has a discrete spectral density, in which case the RPRA method is recommended due to its exact Gaussianity. We also argued in this section that, from a more practical perspective of finite sample length, the traditional practice of setting $S_X(0) = 0$ in order to achieve strong mean and covariance ergodicity is not well justified. Instead, we believe that efforts should be focused on determining appropriate spectral discretization scheme so that the discrepancy between $R_X(\tau)$ and $R_{\hat{X},Trap}(\tau)$ is controlled at an appropriate level. This will be discussed in considerable detail in the next section.

3.4 Determination of Frequency Discretization

The successful application of the spectral simulation method depends on an appropriate spectral discretization scheme, i.e. the cut-off frequency L and the discretization interval Δ_λ. They should be determined systematically from the consideration of both the simulation accuracy and simulation efficiency, such that

- the covariance function $R_{\hat{X},Trap}(\tau)$ should provide a good approximation to the target covariance function $R_X(\tau)$;
- the number m of the terms in the simulation formula should not be excessively large.

Summary of RPDA and RPRA Methods			
Properties/Methods		RPDA Method	RPRA Method
Frequency Discretization		Equally Spaced	Equally Spaced
Required Info		$S_X(\lambda)$	$S_X(\lambda)$
Applicability		General with continuous $S_X(\lambda)$	General
Input		$m+1$ uniform$(0, 2\pi)$ r.v.s	m uniform$(0, 2\pi)$ r.v.s m Rayleigh r.v.s
Covariance Fcn		Trapezoidal Approx.	Trapezoidal Approx.
Path Properties	Periodicity	$\frac{1}{\Delta_\lambda}$	$\frac{1}{\Delta_\lambda}$
	Gaussianity	Asymptotically Gaussian	Exactly Gaussian
	Ergodicity	Strong Ergodicity when $S_X(0)$ set to zero	Weak Ergodicity

Table 3.3: Summary of spectral RPDA method and RPRA method

The only way to simultaneously achieve the above listed goals is to choose an appropriate spectral discretization scheme such that $R_{\hat{X},Trap}(\tau)$ provides just sufficient accuracy for a particular simulation task. It therefore involves finding easy and clear descriptions of the discrepancy $\epsilon_{\hat{X}}(\tau)$, defined in Equation(3.58), between $R_X(\tau)$ and $R_{\hat{X},Trap}(\tau)$.

We start the discussion by first decomposing $\epsilon_{\hat{X}}(\tau)$ as follows:

$$\epsilon_{\hat{X}}(\tau) = \epsilon_{Trap}(\tau) + \epsilon_L(\tau), \tag{3.63}$$

where the components $\epsilon_{Trap}(\tau)$ and $\epsilon_L(\tau)$ are respectively defined as

$$\epsilon_{Trap}(\tau) = 2\int_0^L S_X(\lambda)\cos(2\pi\lambda\tau)d\lambda - R_{\hat{X},Trap}(\tau), \tag{3.64}$$

and

$$\epsilon_L(\tau) = 2\int_L^\infty S_X(\lambda)\cos(2\pi\lambda\tau)d\lambda. \tag{3.65}$$

Therefore $\epsilon_{\hat{X}}(\tau)$ is decomposed into the term $\epsilon_{Trap}(\tau)$ which describes the Trapezoidal approximation error over the compact interval $[0, L]$, and the term $\epsilon_L(\tau)$ which

measures the spectral domain truncation error. This decomposition has been used implicitly in existing literature [115, 124]. Moreover, it also has been a standard practice in these literature to assume that the spectral density $S_X(\lambda)$ has a compact support, so that the power cut-off frequency L is given a priori and the truncation error $\epsilon_L(\tau)$ is ignored. Hence the focus has been almost entirely on the Trapezoidal approximation error $\epsilon_{Trap}(\tau)$, which has been described [115, 53, 124] through a traditional Taylor series based argument as

$$\epsilon_{Trap}(\tau) = \frac{L}{12}\Delta_\lambda^2 \frac{\partial^2}{\partial \lambda^2}[S_X(\lambda)\cos(2\pi\lambda\tau)]_{\lambda^*}, \quad \lambda^* \in (0, L), \tag{3.66}$$

where the symbol $[\ ]_{\lambda^*}$ denotes evaluation of the derivative inside the brackets at $\lambda = \lambda^*$. Such a description requires the spectral density $S_X(\lambda)$ having a continuous second order derivative in the interval $(0, L)$, which should be satisfied for most of the spectral models used in practice [65, 115, 95]. One potentially useful piece of information that can be extracted from Equation(3.66) is the fact that when τ is not large, the magnitude of $\epsilon_{Trap}(\tau)$ has an order of $O(\Delta_\lambda^2)$, and hence can be used to decide whether the Δ_λ is sufficiently small [115].

In the more practical situations, however, the spectral density $S_X(\lambda)$ rarely has a compact support, and hence the truncation error has to be taken into account. Moreover, although the Trapezoidal approximation error $\epsilon_{Trap}(\tau)$ contains information about Δ_λ, because of the ambiguity involved in λ^* the global description given by Equation(3.66) is still too vague to be useful for finding appropriate values for Δ_λ. Such difficulties therefore motivates us to look at the discrepancy $\epsilon_{\hat{X}}(\tau)$ from another perspective.

The key point is to recognize the fact that equally spaced sampling in the frequency domain will introduce aliasing effect into the time domain. Let us denote $R_X^{\Delta_\lambda}(\tau)$ as the aliased version of the covariance function $R_X(\tau)$, and is defined as

$$R_X^{\Delta_\lambda}(\tau) = \sum_{k=-\infty}^{\infty} R_X\left(\tau + \frac{k}{\Delta_\lambda}\right). \tag{3.67}$$

Then we can define the aliasing error $\epsilon_{alias}^{\Delta_\lambda}(\tau)$ as follows

$$\epsilon_{alias}^{\Delta_\lambda}(\tau) = R_X(\tau) - R_X^{\Delta_\lambda}(\tau) = -\sum_{k=1}^{\infty}\left[R_X(\tau + \frac{k}{\Delta_\lambda}) + R_X(\tau - \frac{k}{\Delta_\lambda})\right]. \tag{3.68}$$

Moreover, by applying the Poisson summation formula [120], we can easily write $R_X^{\Delta_\lambda}(\tau)$ in terms of $S_X(\lambda)$ as

$$\begin{aligned} R_X^{\Delta_\lambda}(\tau) &= \Delta_\lambda \sum_{k=-\infty}^{\infty} S_X(k\Delta_\lambda) e^{i2\pi k\Delta_\lambda \tau} \\ &= \Delta_\lambda \left\{ S_X(0) + \sum_{k=1}^{\infty} 2S_X(k\Delta_\lambda)\cos(2\pi k\Delta_\lambda \tau) \right\} \\ &= R_{\hat{X},Trap}(\tau) + 2\Delta_\lambda \sum_{k=m+1}^{\infty} S_X(k\Delta_\lambda)\cos(2\pi k\Delta_\lambda \tau). \end{aligned} \tag{3.69}$$

For the purpose of notational convenience, let us denote

$$\epsilon_{L,\Delta_\lambda}(\tau) = 2\Delta_\lambda \sum_{k=m+1}^{\infty} S_X(k\Delta_\lambda)\cos(2\pi k\Delta_\lambda \tau), \tag{3.70}$$

and it can be easily recognized that $\epsilon_{L,\Delta_\lambda}(\tau)$ clearly approximates the truncation error $\epsilon_L(\tau)$ over an equally spaced spectral grid.

Hence using these newly defined quantities we obtain the following alternative decomposition of the the discrepancy $\epsilon_{\hat{X}}(\tau)$

$$\begin{aligned} \epsilon_{\hat{X}}(\tau) &= R_X(\tau) - R_{\hat{X},Trap}(\tau) \\ &= R_X(\tau) - R_X^{\Delta_\lambda}(\tau) + R_X^{\Delta_\lambda}(\tau) - R_{\hat{X},Trap}(\tau) \\ &= \epsilon_{alias}^{\Delta_\lambda}(\tau) + \epsilon_{L,\Delta_\lambda}(\tau). \end{aligned} \tag{3.71}$$

Therefore this alternative decomposition suggests that the discrepancy $\epsilon_{\hat{X}}(\tau)$ can also be analyzed from the perspective of time domain aliasing effect.

Given a practical simulation task, we usually have an idea of how long the realization has to be. Hence $\tau_{\max} = t_n - t_1$, which is the maximum lag we want to realize, is usually known in advance. As argued in Section(3.3.3), under this finite sample length scenario it is important to control the maximum magnitude of the discrepancy $\max_{\tau\in[0,\tau_{\max}]} |\epsilon_{\hat{X}}(\tau)|$, which using the decomposition in Equation(3.71), can be conveniently bounded by

$$\max_{\tau\in[0,\tau_{\max}]} |\epsilon_{\hat{X}}(\tau)| \leq \epsilon_{alias}^{\tau_{max}} + \epsilon_{L,\Delta_\lambda}(0), \tag{3.72}$$

where for convenience we have introduced the following short-hand notation

$$\epsilon_{alias}^{\tau_{max}} = \max_{\tau\in[0,\tau_{\max}]} |\epsilon_{alias}^{\Delta_\lambda}(\tau)|. \tag{3.73}$$

Note in Equation(3.72) we have also used the simple fact that the magnitude of the term $\epsilon_{L,\Delta_\lambda}(\tau)$ reaches its maximum at $\tau = 0$, i.e. we have

$$|\epsilon_{L,\Delta_\lambda}(\tau)| \leq \epsilon_{L,\Delta_\lambda}(0), \text{ for all } \tau. \tag{3.74}$$

This upper bound for $\max_{\tau\in[0,\tau_{\max}]} |\epsilon_{\hat{X}}(\tau)|$, based on the alternative decomposition of $\epsilon_{\hat{X}}(\tau)$, will turn out to be extremely useful in the sense that a spectral discretization scheme can be conveniently and accurately determined by specifying appropriate values for $\epsilon_{alias}^{\tau_{max}}$ and $\epsilon_{L,\Delta_\lambda}(0)$. We will see in Section(3.4.1) that by imposing a mild assumption on the covariance function $R_X(\tau)$, an appropriate Δ_λ can be determined, by a novel algorithm, through pre-specifying a maximum aliasing error.

On the other hand, for the component $\epsilon_{L,\Delta_\lambda}(0)$ we obviously have the following approximation

$$\epsilon_{L,\Delta_\lambda}(0) \approx \epsilon_L(0) = \int_L^\infty S_X(\lambda)d\lambda, \tag{3.75}$$

where $\epsilon_L(0)$ apparently measures the loss of variance as a result of spectral domain truncation. Hence the above approximation suggests that an appropriate L can be determined by pre-specifying a reasonable loss of variance. We will see in Section(3.4.1) that this can be accomplished through an easy application of the Newton-Rahpson recursion [124].

Although this procedure does not give a precise control over $\epsilon_{L,\Delta_\lambda}(0)$, in practice the difference will most likely be negligible. This is because for most of the spectral models used in practice, the spectral density $S_X(\lambda)$ will become monotonic decreasing when λ becomes large enough. Consequently with a reasonably small pre-specified loss of variance, the appropriate L will usually be large enough so that $S_X(\lambda)$ becomes monotonic decreasing for $\lambda \geq L$. Then by an elementary argument [31] it can be derived that

$$0 \leq \epsilon_L(0) - \epsilon_{L,\Delta_\lambda}(0) \leq \Delta_\lambda S_X(L), \tag{3.76}$$

thus showing the difference between $\epsilon_L(0)$ and $\epsilon_{L,\Delta_\lambda}(0)$ will indeed be negligible.

3.4.1 Determining the Power Cut-Off Frequency

As pointed out earlier, the power cut-off frequency L should be determined by pre-specifying the loss of variance $\epsilon_L(0)$, which can be regarded as an approximation to the upper-bound for the component $\epsilon_{L,\Delta_\lambda}(\tau)$. In practice, from the consideration of either computational costs or statistical accuracy, the loss of variance $\epsilon_L(0)$ should not be set at an extremely small level. It is suggested by Shinozuka and Deodatis [115] that $\epsilon_L(0) = 0.01$ or 0.001, relative to a unit variance, should be sufficient for most of the practical simulation scenarios. Apart from this practical guideline of the appropriate size of $\epsilon_L(0)$, there is no further discussions in existing literature about how the power cut-off frequency L can be obtained in a systematical way. Perhaps this is because even with the most general spectral density $S_X(\lambda)$, the determination of the appropriate L can be reduced to a simple application of the Newton-Raphson recursion [124]. However, determining appropriate L according to a pre-specified $\epsilon_L(0)$ is still a very important step in the implementation of the spectral simulation methods. Although a larger-than-necessary rough estimate of L works, there will be extra computational cost involved, which is certainly not desirable. Consequently for the purposes of completeness we therefore still present the details of the specialized implementation of the Newton-Raphson recursion.

Let $\epsilon_L(0)$ be the pre-specified loss of variance, and define the function

$$F(\nu) = \int_{\nu}^{\infty} S_X(\lambda)d\lambda - \frac{\epsilon_L(0)}{2}, \tag{3.77}$$

whose derivative is given by

$$F'(\nu) = -S_X(\nu). \tag{3.78}$$

It is then immediate that the appropriate power cut-off frequency L solves the following equation

$$F(L) = \int_{L}^{\infty} S_X(\lambda)d\lambda - \frac{\epsilon_L(0)}{2} = 0. \tag{3.79}$$

Starting from an arbitrary initial value ν_0, the Newton-Raphson recursion can be written as follows

$$\nu_{k+1} = \nu_k - \frac{F(\nu_k)}{F'(\nu_k)} = \nu_k + \frac{F(\nu_k)}{S_X(\nu_k)} \quad \text{for} \quad k > 0. \tag{3.80}$$

The value $F(\nu_k)$ in the $(k+1)$th iteration can written as

$$\begin{aligned} F(\nu_k) &= \int_{\nu_k}^{\infty} S_X(\lambda)d\lambda - \frac{\epsilon_L(0)}{2} \\ &= \int_0^{\infty} S_X(\lambda)d\lambda - \int_0^{\nu_k} S_X(\lambda)d\lambda - \frac{\epsilon_L(0)}{2} \\ &= \frac{1}{2}R_X(0) - \int_0^{\nu_k} S_X(\lambda)d\lambda - \frac{\epsilon_L(0)}{2}, \end{aligned} \tag{3.81}$$

and therefore can be conveniently evaluated through applying a numerical quadrature. The above procedure is summarized as follows:

Algorithm 3.1.

1. *Determine the appropriate loss of variance $\epsilon_L(0)$;*
2. *Determine the tolerance ϵ as the stopping criteria for the Newton-Raphson recursion;*
3. *Initialize the recursion with an arbitrary ν_0 (for example $\nu_0 = 0$);*
4. *Update ν_{k+1} from ν_k through Equation(3.80);*
5. *Terminate the recursion when $F(\nu_{k+1}) < \epsilon$.*

Typically when the spectral density $S_X(\lambda)$ does not decay very slowly, or when the loss of variance $\epsilon_L(0)$ is not extremely small, this algorithm outlined above converges reasonably fast. Moreover, when we want to generate multiple realizations, this algorithm needs only to be executed once, therefore will not have significant impact on the computational costs of the spectral simulation methods.

3.4.2 Determining the Spectral Discretization Interval

Previously we have briefly discussed the problems associated with determining the appropriate spectral discretization interval Δ_λ. Although the Taylor series based error estimate of Equation(3.66) gives a rate of convergence, it can not be used to determine an appropriate Δ_λ. We then argued that the appropriate Δ_λ can be obtained from the consideration of the aliasing-error, introduced from the equally spaced spectral discretization, and the detailed derivation will be provided in this section.

To start with explaining the calculation of Δ_λ, note that for a given maximum lag τ_{max} required in a particular simulation task, the maximum aliasing error $\epsilon_{alias}^{\tau_{max}}$ can be expressed as

$$\epsilon_{alias}^{\tau_{max}} = \max_{0\le\tau\le\tau_{max}} \left| \sum_{k=1}^{\infty} [R_X(\tau + \frac{k}{\Delta_\lambda}) + R_X(\tau - \frac{k}{\Delta_\lambda})] \right|. \tag{3.82}$$

In many situations, however, the above expression may not be easy to deal with, especially when the covariance function $R_X(\tau)$ contains oscillations, so that it can assume both positive and negative values. In order to handle these complicated situations, we will therefore introduce the following simplifying assumption:

Assumption 3.1. *We assume that the absolute value of the covariance function $|R_X(\tau)|$ is bounded by an envelope function $H_X(\tau)$ for all $\tau \in \mathbb{R}$, where $H_X(\tau)$ is positive, even, and monotonic decreasing for $\tau \ge 0$ (hence monotonic increasing for $\tau \le 0$).*

Using the envelope function $H_X(\tau)$, a more conservative upper-bound for $\epsilon_{alias}^{\tau_{max}}$ can be obtained as

$$\epsilon_{alias}^{\tau_{max}} \le \max_{0\le\tau\le\tau_{max}} \sum_{k=1}^{\infty} [H_X(\tau + \frac{k}{\Delta_\lambda}) + H_X(\tau - \frac{k}{\Delta_\lambda})]. \tag{3.83}$$

Equation(3.83) is still too complicated to be useful in practice, since it involves taking maximum of a complicated function over a finite interval. Therefore further simplification is needed before it can be used in general cases to obtain an appropriate Δ_λ. This can be done by noticing that for $\tau > 0$ we have

$$|\tau - \frac{k}{\Delta_\lambda}| < \tau + \frac{k}{\Delta_\lambda}, \text{ for } k \ge 1. \tag{3.84}$$

This observation, together with the fact that the function $H_X(\tau)$ is positive, even and monotonic decreasing for $\tau > 0$, lead us to the following inequality:

$$\begin{aligned} H_X(\tau + \frac{k}{\Delta_\lambda}) + H_X(\tau - \frac{k}{\Delta_\lambda}) &\le 2H_X(\tau - \frac{k}{\Delta_\lambda}) \\ &= 2H_X(\frac{k}{\Delta_\lambda} - \tau) \\ &\le 2H_X(\frac{k}{\Delta_\lambda} - \tau_{max}), \text{ for } k \ge 1, \end{aligned} \tag{3.85}$$

where we have implicitly assumed that $\Delta_\lambda < \frac{1}{\tau_{max}}$, which is required to avoid the periodicity of $R_{\hat{X},Trap}(\tau)$. Consequently we managed to obtain a more conservative but simpler error bound as follows

$$\epsilon_{alias}^{\tau_{max}} \leq 2\sum_{k=1}^{\infty} H_X(\frac{k}{\Delta_\lambda} - \tau_{max}). \tag{3.86}$$

For some special cases, the envelope function $H_X(\tau)$ admits a simple analytical form, so that the upper-bound in the right-hand-side of Equation(3.86) can be calculated analytically (see Section(3.5.1) for an example). In other more general situations, however, the infinite sum on the right-hand-side of Equation(3.86) has to be approximated by taking a finite sum of, say first M terms. In other words $\epsilon_{alias}^{\tau_{max}}$ can be approximately bounded by

$$\epsilon_{alias}^{\tau_{max}} \leq 2\sum_{k=1}^{M} H_X(\frac{k}{\Delta_\lambda} - \tau_{max}). \tag{3.87}$$

Therefore given a pre-specified maximum aliasing error $\epsilon_{alias}^{\tau_{max}}$, a conservative estimate of the spectral discretization interval Δ_λ can be found by solving the following equation

$$\epsilon_{alias}^{\tau_{max}} = 2\sum_{k=1}^{M} H_X(\frac{k}{\Delta_\lambda} - \tau_{max}), \;\; \Delta_\lambda \in (0, \frac{1}{\tau_{max}}). \tag{3.88}$$

In practice, just as the case of pre-specifying the loss of variance, the value of $\epsilon_{alias}^{\tau_{max}}$ need not be set at a very small value either. For most of the practical simulation scenarios setting $\epsilon_{alias}^{\tau_{max}}$ equal to 0.01 or 0.001 relative to a unit variance should be sufficient, because such an error will most likely be dominated by the variation due to the finite length of the realization. Moreover the number M of terms used in the above equation does't need to be large at all if the covariance function $R_X(\tau)$ and its envelope function $H_X(\tau)$ decays relatively fast.

A rough and conservative estimate of M can be obtained as follows. First note that we must have $\Delta_\lambda < \frac{1}{\tau_{max}}$ so that due to monotonic decreasing of $H_X(\tau)$ we have

$$H_X(\frac{k}{\Delta_\lambda} - \tau_{max}) \leq H_X((k-1)\tau_{max}), \;\; \text{for} \;\; k \geq 1. \tag{3.89}$$

Then we simply choose M such that $H_X((M-1)\tau_{max})$ becomes much smaller as

compared to the pre-specified $\epsilon_{alias}^{\tau max}$. Having determined the appropriate M, then note that the right hand side of Equation(3.88) is monotonic increasing with respect to Δ_λ for $\Delta_\lambda \in (0, \frac{1}{\tau_{max}})$, and hence a standard numerical algorithm (e.g. fzero() in Matlab) can be readily applied to find the appropriate value of Δ_λ.

To conclude the discussion in this section, we summarize the above procedure for finding Δ_λ in the following algorithm:

Algorithm 3.2.

1. *Specify the maximum aliasing error $\epsilon_{alias}^{\tau_{max}}$ (relative to a unit variance) at a reasonable level;*
2. *Find the envelope function $H_X(\tau)$ for the covariance function $R_X(\tau)$ of the continuous time process;*
3. *Determine the maximum lag $\tau_{\max}$ we want to realize;*
4. *Find M such that $H_X((M-1)\tau_{max})$ becomes much smaller than $\epsilon_{alias}^{\tau_{max}}$;*
5. *Solve for $\Delta_\lambda \in (0, \frac{1}{\tau_{max}})$ according to Equation(3.88) through a numerical zero finding routine.*

3.4.3 Further Discussions

So far we have introduced the principles and the algorithms we can use to construct the spectral discretization scheme for spectral simulation method. In this section we will give intuitive discussions, providing more insights on how does our proposed method controls the discrepancy $\epsilon_{\hat{X}}(\tau)$. Moreover, implicit in existing literature [115, 124] is another traditional choice of $\Delta_\lambda = \frac{1}{2\tau_{\max}}$, which is the time-domain equivalent of the Nyquist frequency used in engineering applications [15]. We will give an intuitive explanation on how our proposed method, by removing unnecessary accuracy, can improve efficiency as compared to the traditional choice.

First of all notice that as discussed earlier, at $\tau = 0$ the value of $\epsilon_{L,\Delta_\lambda}(\tau)$ reaches its maximum, which will be numerically almost identical to the pre-specified loss of variance $\epsilon_L(0)$ in most cases. Furthermore, looking back at Equation(3.68) which defines the aliasing error $\epsilon_{alias}^{\Delta_\lambda}(\tau)$, it is immediately recognized that $\epsilon_{alias}^{\Delta_\lambda}(\tau)$ consists of covariances at aliased lags $R_X(\tau - \frac{1}{\Delta_\lambda}), R_X(\tau + \frac{1}{\Delta_\lambda}), \cdots$. Consequently due to the

decay of the covariance function, for τ close to zero, these aliased covariances, and hence the aliasing error $\epsilon_{alias}^{\Delta_\lambda}(\tau)$, will tend to have small magnitudes. In practice when the loss of variance is set at the recommended level of 0.01 or 0.001 (relative to unit variance), it is extremely likely that the aliasing error $\epsilon_{alias}^{\Delta_\lambda}(\tau)$ for τ close to zero will be negligible as compared to the approximate truncation error $\epsilon_{L,\Delta_\lambda}(\tau)$, because most of the covariance function $R_X(\tau)$ used in practice will have a relatively fast decay. This observation therefore suggests that the discrepancy $\epsilon_{\hat{X}}(\tau)$ for τ close to zero will be dominated by the component $\epsilon_{L,\Delta_\lambda}(\tau)$.

On the other hand, when τ moves away from zero, the approximate truncation error $\epsilon_{L,\Delta_\lambda}(\tau)$ will have a magnitude strictly smaller than than its maximum at $\tau = 0$, due to the cancellation effect introduced from the cosine function in Equation(3.70). The aliasing error $\epsilon_{alias}^{\Delta_\lambda}(\tau)$ as τ moves away from zero, however, will tend to increase, mostly because one of the particular aliased covariances $R_X(\tau - \frac{1}{\Delta_\lambda})$ will be increasing and approaching $R_X(0)$, which is the largest magnitude of the covariance function. Hence when the maximum aliasing error $\epsilon_{alias}^{\tau_{max}}$ is specified at the recommended level of 0.01 or 0.001 (relative to unit variance), it is likely that the discrepancy $\epsilon_{\hat{X}}(\tau)$ for τ close to τ_{max} will be dominated by the aliasing error component $\epsilon_{alias}^{\Delta_\lambda}(\tau)$. Figure(3.2) contains a graphical illustration of what has been discussed so far. The dominance of $\epsilon_{L,\Delta_\lambda}(\tau)$ for τ close to zero, and the dominance of $\epsilon_{alias}^{\Delta_\lambda}(\tau)$ for τ close to $\tau_{\max}$, are clearly indicated in the figure. Because L is determined from controlling $\epsilon_{L,\Delta_\lambda}(\tau)$ whereas Δ_λ from controlling $\epsilon_{alias}^{\Delta_\lambda}(\tau)$, we can therefore conclude that our proposed approach for determining the spectral discretization scheme roughly amounts to controlling the discrepancy $\epsilon_{\hat{X}}(\tau)$ at the two ends of the required lag interval $[0, \tau_{max}]$.

Figure(3.2) also contains a graphical illustration of what happens when the traditional choice of $\Delta_\lambda = \frac{1}{2\tau_{max}}$ is used. For a fixed Δ_λ, this traditional choice implicitly supports a maximum lag of only $\tau'_{\max} = \frac{1}{2\Delta_\lambda}$. However, as shown in the figure, at this $\tau'_{\max}$ both $\epsilon_{L,\Delta_\lambda}(\tau)$ and $\epsilon_{alias}^{\Delta_\lambda}(\tau)$ could be unnecessarily small, implying an overall accuracy that is too much more than necessary for statistical purposes. For one thing at $\tau'_{\max}$ the approximate truncation error $\epsilon_{L,\Delta_\lambda}(\tau)$ can be written as

$$\epsilon_{L,\Delta_\lambda}(\tau'_{max}) = \sum_{k=m+1}^{\infty} (-1)^k S_X(k\Delta_\lambda), \tag{3.90}$$

which due to the cancellation effect is much less than its maximum value at $\tau = 0$.

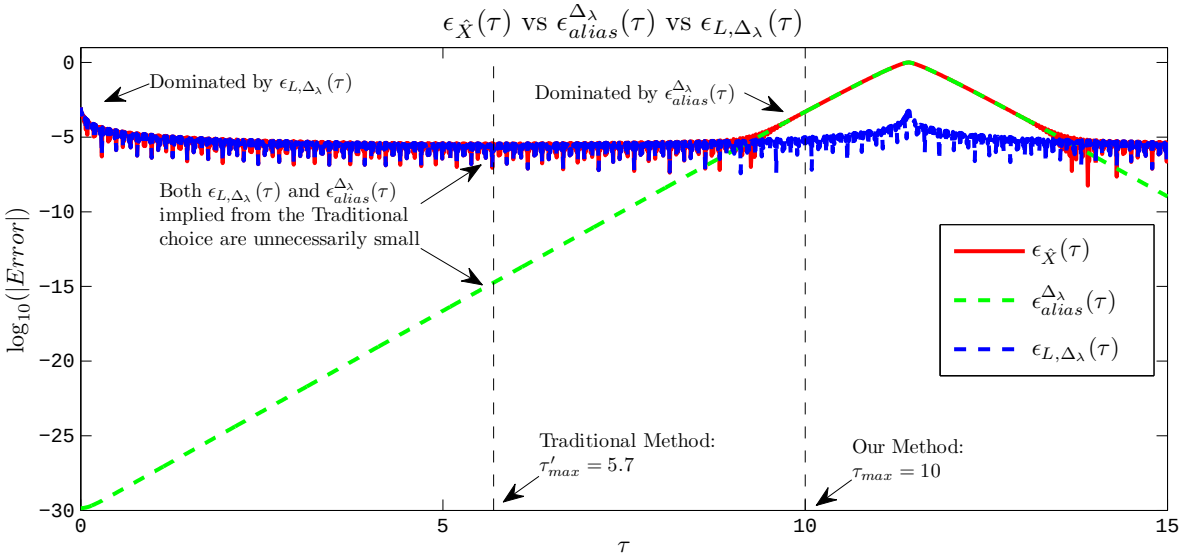

Figure 3.2: Illustration of the components of $\epsilon_{\hat{X}}(\tau)$ using a Matérn process with parameters $\sigma = 1, \phi = 1, \nu = 1$. The spectral discretization interval $\Delta_\lambda = 0.0875$, which using our proposed method supports $\tau_{\max} = 10$ at an error of 0.001 relative to unit variance, significantly larger than $\tau'_{\max} = 5.7$ suggested from the traditional method.

Furthermore, the aliasing error $\epsilon^{\Delta_\lambda}_{alias}(\tau)$ at $\tau = \tau'_{max}$ can be written as

$$\begin{aligned}
\epsilon^{\Delta_\lambda}_{alias}(\tau'_{\max}) &= -\sum_{k=1}^{\infty}\left[R_X\left(\tau'_{max} + \frac{k}{\Delta_\lambda}\right) + R_X\left(\tau'_{max} - \frac{k}{\Delta_\lambda}\right)\right] \\
&= -R_X(\tau'_{max} - \frac{1}{\Delta_\lambda}) - R_X(\tau'_{max} + \frac{1}{\Delta_\lambda}) - \cdots \\
&= -R_X(\tau'_{max}) - R_X(3\tau'_{max}) - \cdots . \qquad (3.91)
\end{aligned}$$

which could also be unnecessarily small if the covariance function $R_X(\tau)$ possesses relatively fast decay - a situation that is very often the case in practical applications. Consequently, both Equation(3.90) and (3.91) indicated that for a fixed Δ_λ the spectral simulation method should be able to produce covariance structure over a longer maximum lag than otherwise implied by the traditional choice of $\tau'_{max} = \frac{1}{2\Delta_\lambda}$. By achieving an accuracy just sufficient for statistical purposes, as indicated in the figure our proposed method therefore extends the maximum lag from τ'_{max} to τ_{max} , and

consequently improves the efficiency of the spectral simulation method by allowing it to work on its full capacity.

3.5 Numerical Examples

In this section we present numerical examples to illustrate points we made earlier. We will be showing that the proposed algorithms for finding the frequency domain discretization scheme will provide desired accuracy, while at the same time reduce the computational costs. The stationary Gaussian processes used in the numerical examples are

1. Narrow Band process (see Section(2.3.1)), in which a semi-analytical solution can be obtained for the frequency interval Δ_λ;

2. Oscillatory Matérn process (see Section(2.3.3)), in which a numerical solution has to be used to find the frequency interval Δ_λ.

For each stationary Gaussian processes under consideration, we will in Section(3.5.1) and (3.5.2) calculate the frequency domain discretization scheme according to our proposed algorithms. In order to illustrate the fact that the frequency domain discretization scheme obtained this way will give sufficient accuracy, we will also calculate and tabulate the following error measures:

- ShortLag-MaxErr: defined as the maximum absolute discrepancy between $R_X(\tau)$ and $R_{\hat{X},Trap}(\tau)$ for $\tau \in [0, \tau_{max}/2)$;

- FarLag-MaxErr: defined as the maximum absolute discrepancy between $R_X(\tau)$ and $R_{\hat{X},Trap}(\tau)$ for $\tau \in [\tau_{max}/2, \tau_{max}]$.

Moreover for comparison purposes, we will also repeat these calculations under the traditional choice of $\Delta_\lambda = \frac{1}{2\tau_{max}}$. By doing this we want to show that although this more conservative Δ_λ works, it tends to provide an accuracy that is more than necessary, at the cost of a significantly larger number m of terms used in the simulation formula.

Then in Section(3.5.3) we will consider the traditional practice of setting $S_X(0) = 0$. Using the Oscillatory Matérn process as an example, we will show that such a practice will tend to introduce more error to the approximating covariance function

$R_{\hat{X},Trap}(\tau)$. Finally in Section(3.5.4) we compare the CPU times between the spectral method and the time domain simulation method based on Cholesky factorization. We will show that the RPDA method is slightly more efficient than the RPRA method, and in many scenarios the spectral simulation method (both RPDA and RPRA) will enjoy significant computational advantage over the matrix factorization based time domain simulation method.

3.5.1 Narrow Band Process

A narrow band process [83] is characterized by a rational spectral density of the form

$$S_X(\lambda) = \sigma^2 \frac{\alpha\lambda^2}{\alpha^2\lambda^2 + \pi^2(\lambda^2 - \lambda_0^2)^2}, \tag{3.92}$$

and is one of the special examples of the CARMA family processes. The corresponding covariance function is given by

$$R_X(\tau) = \sigma^2 e^{-\alpha|\tau|}[\cos(\omega_0|\tau|) - \frac{\alpha}{\omega_0}\sin(\omega_0|\tau|)], \;\; \omega_0 = \sqrt{4\pi^2\lambda_0^2 - \alpha^2}. \tag{3.93}$$

It can be seen from Equation(3.92) that the spectral density of the narrow band process has a single peak, of magnitude $\frac{\sigma^2}{\alpha}$, at frequency $\lambda = \lambda_0$, and hence introduces oscillation in the corresponding covariance function. In this particular example we will set the parameter values to the following

$$\sigma = 1, \;\; \alpha = 1, \;\; \lambda_0 = 1, \tag{3.94}$$

and a graphical illustration of both the covariance function and the spectral density is given in Figure(3.3). Using the modified Newton-Rahpson algorithm outlined in Section(3.4.1) we could easily find the appropriate power cut-off frequency L, according to different pre-specified loss of variance:

1. $\epsilon_L(0) = 0.01$: $L \approx 21$;
2. $\epsilon_L(0) = 0.001$: $L \approx 203$;
3. $\epsilon_L(0) = 0.0001$: $L \approx 2027$.

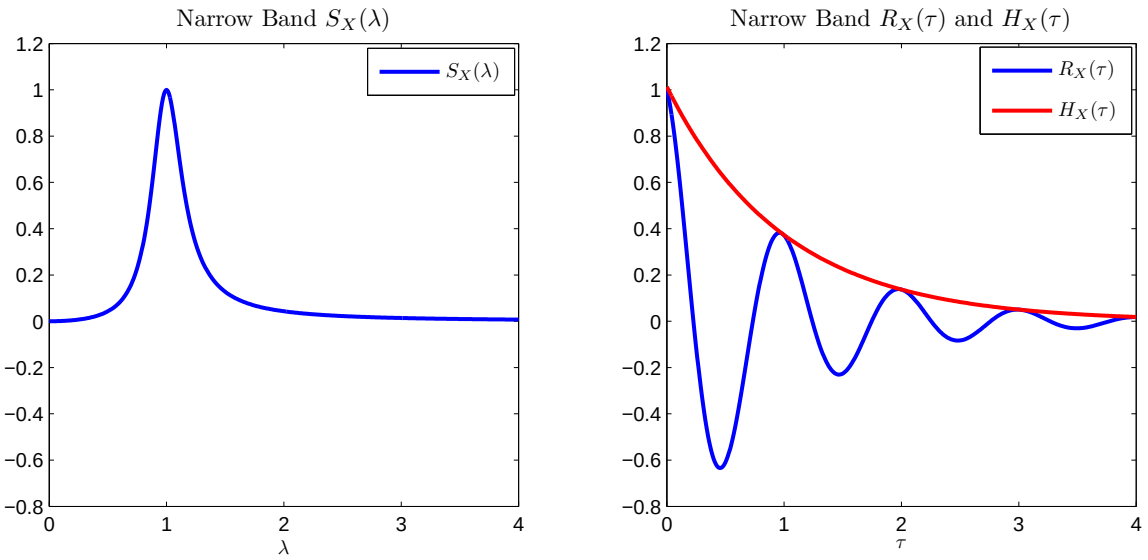

Figure 3.3: Shape of the spectral density $S_X(\lambda)$, covariance function $R_X(\tau)$, and the covariance envelope $H_X(\tau)$, for the Narrow Band process with parameters $\sigma = 1, \alpha = 1, \lambda_0 = 1$.

As the required loss of variance becomes smaller, the power cut-off frequency L increases, as expected. What makes the Narrow Band process special is the fact that its covariance envelope function $H_X(\tau)$ takes a simple form

$$H_X(\tau) = Ce^{-\alpha|\tau|}, \text{ where } C = \sigma^2 \frac{2\pi\lambda_0}{\sqrt{4\pi^2\lambda_0^2 - \alpha^2}}, \tag{3.95}$$

which is also illustrated in Fig(3.3). This simple $H_X(\tau)$ will give a semi-analytical solution for Δ_λ. To be more specific, note that using the above equation for $H_X(\tau)$, the conservative upper-bound for $\epsilon^{max}_{aliasing}$, given by Equation(3.86), can be written as

$$2\sum_{k=1}^{\infty} H_X(\frac{k}{\Delta_\lambda} - \tau_{max}) = 2\sum_{k=1}^{\infty} Ce^{-\alpha(k/\Delta_\lambda - \tau_{max})} = 2C\frac{e^{\alpha\tau_{max}}}{e^{\alpha/\Delta_\lambda} - 1}. \tag{3.96}$$

Equating this simple upper-bound to a pre-specified $\epsilon^{max}_{aliasing}$, we can then conveniently obtain an expression for Δ_λ as follows:

$$2C\frac{e^{\alpha\tau_{max}}}{e^{\alpha/\Delta_\lambda}-1} = \epsilon^{max}_{aliasing} \Rightarrow \Delta_\lambda = \frac{\alpha}{\log(1+2Ce^{\alpha\tau_{max}}/\epsilon^{max}_{aliasing})}. \tag{3.97}$$

Hence in this case appropriate Δ_λ can be easily found, without resorting to any numerical procedures.

Table(3.4) tabulates the details of the frequency domain discretization for this Narrow Band process ($\sigma = 1, \alpha = 1, \lambda_0 = 1$). The loss of variance is set at $\epsilon_L(0) = 0.001$, corresponding to a power cut-off frequency given by $L = 203$; and the maximum aliasing error is also set at $\epsilon^{\tau_{max}}_{aliasing} = 0.001$, for a set of different τ_{max}. As mentioned earlier, the discretization details corresponding to the traditional choice of $\Delta_\lambda = \frac{1}{2\tau_{max}}$ is also included as a comparison.

The first thing to note from the table is the fact that the spectral discretization scheme calculated through our proposed algorithms managed to control the discrepancy $\epsilon_{\hat{X}}(\tau)$ at the desirable level: the ShortLag-MaxErr is almost exactly given by the pre-specified loss of variance $\epsilon_L(0) = 0.001$, and the FarLag-MaxErr is well (but not excessively) below the pre-specified level $\epsilon^{max}_{aliasing} = 0.001$. Although the traditional choice of $\Delta_\lambda = \frac{1}{2\tau_{max}}$ also achieves the desired accuracy, it is overly conservative in the sense that the error measure FarLag-MaxErr is unnecessarily small. Consequently by just meeting the desired accuracy level, the discretization scheme obtained through our proposed method contains a much smaller number m of terms, thus reduces computational costs.

$L = 203$	$\tau_{max} = 10$	$\tau_{max} = 50$	$\tau_{max} = 100$	$\tau_{max} = 500$
Δ_λ	0.0568	0.0174	0.0093	0.002
m	3574	11667	21828	101500
ShortLag-MaxErr	0.001	0.001	0.001	0.001
FarLag-MaxErr	0.00048	0.00048	0.00048	0.00048
$\Delta_\lambda = 1/2\tau_{max}$	0.05	0.01	0.005	0.001
m	4060	20300	40600	203000
ShortLag-MaxErr	0.001	0.001	0.001	0.001
FarLag-MaxErr	0.000042	2.467e-08	1.227e-08	2.459e-09

Table 3.4: Frequency domain discretization scheme for Narrow Band process with parameters $\sigma = 1, \alpha = 1, \lambda_0 = 1$, corresponding to $\epsilon_L(0) = 0.001$ and $\epsilon^{\tau_{max}}_{aliasing} = 0.001$.

3.5.2 Oscillatory Matérn Process

In this particular numerical example, we will use an Oscillatory Matérn process (see Section(2.3.3)) to demonstrate our proposed algorithm. We will set the parameter values to be the following

$$\sigma = 1, \phi = 1, \nu = 1, \lambda_0 = 1, \tag{3.98}$$

and Figure(3.4) gives a graphical illustration of the shapes of $S_X(\lambda)$ and $R_X(\tau)$ for this particular model. Again using the modified Newton-Raphson algorithm we find

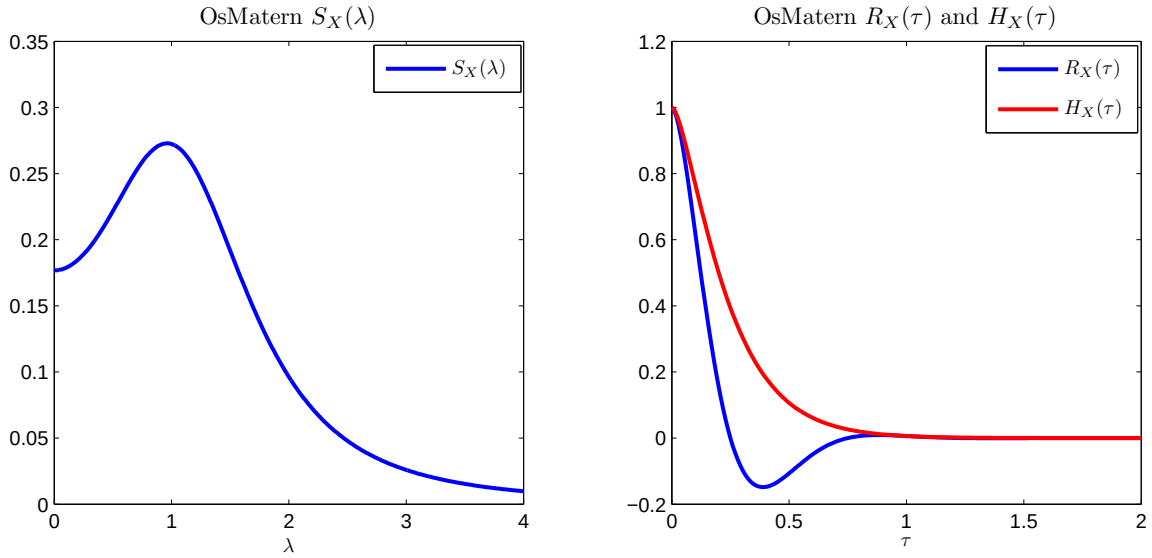

Figure 3.4: Shape of the spectral density $S_X(\lambda)$, covariance function $R_X(\tau)$, and the covariance envelope $H_X(\tau)$, for the Oscillatory Matérn process with parameters $\sigma = 1, \phi = 1, \nu = 1, \lambda_0 = 1$.

the appropriate power cut-off frequency L, corresponding to different pre-specified loss of variance $\epsilon_L(0)$ as

1. $\epsilon_L(0) = 0.01$: $L \approx 8$;
2. $\epsilon_L(0) = 0.001$: $L \approx 23$;
3. $\epsilon_L(0) = 0.0001$: $L \approx 71$.

The power cut-off frequency L at each level of loss of variance $\epsilon_L(0)$ are much smaller than the corresponding truncation error for the Narrow Band process, because for our chosen parameters the spectral density $S_X(\lambda)$ of the Oscillatory Matérn process decays much faster. To find the appropriate frequency domain discretization interval Δ_λ, we need to consider the envelope function $H_X(\tau)$, which according to Equation(2.10) can be easily determined as

$$H_X(\tau) = 2\sigma^2 \frac{(\pi\phi\tau)^\nu}{\Gamma(\nu)} K_\nu(2\pi\phi\tau). \tag{3.99}$$

This envelope function is also illustrated in Figure(3.4). Unlike the Narrow Band process, the envelope function $H_X(\tau)$ for the Oscillatory Matérn process contains the Bessel function, and consequently does not allow a semi-analytical solution for Δ_λ. Therefore we have to use the numerical solution outlined in Section(3.4.2), through solving Equation(3.88).

Table(3.5) contains the details of the frequency domain discretization schemes, with the power cut-off frequency determined as $L \approx 23$, corresponding to a loss of variance set at $\epsilon_L(0) = 0.001$. Just as in the case of Narrow Band processes, the frequency domain discretization interval Δ_λ are calculated, for different valued of τ_{max}, in two ways: first using a maximum aliasing error $\epsilon^{max}_{aliasing} = 0.001$, we calculate Δ_λ through solving Equation(3.88), with $M = 2$; then the traditional choice of $\Delta_\lambda = \frac{1}{2\tau_{max}}$ is also included as a comparison.

$L = 23$	$\tau_{max} = 10$	$\tau_{max} = 50$	$\tau_{max} = 100$	$\tau_{max} = 500$
Δ_λ	0.0875	0.0194	0.0099	0.00199
m	263	1186	2324	11500
ShortLag-MaxErr	0.0009927	0.0009913	0.0009914	0.001
FarLag-MaxErr	0.0004432	0.0004487	0.0004489	0.0004496
$1/2\tau_{max}$	0.05	0.01	0.005	0.001
m	460	2300	4600	23000
ShortLag-MaxErr	0.0009965	0.0009992	0.0009996	0.001
FarLag-MaxErr	2.229e-06	4.4424e-07	2.239e-07	4.4820e-08

Table 3.5: Frequency domain discretization scheme for Oscillatory Matérn process with parameters $\sigma = 1, \phi = 1, \nu = 1, \lambda_0 = 1$, corresponding to $\epsilon_L(0) = 0.001$ and $\epsilon^{\tau_{max}}_{aliasing} = 0.001$.

Again our proposed method for calculating Δ_λ shows its advantage: the frequency domain discretization interval Δ_λ calculated from our proposed method is not overly small, yet it gives an approximating covariance function $R_{\hat{X},Trap}(\tau)$ that is close enough to the target covariance function $R_X(\tau)$.

3.5.3 Ignoring the Power $S_X(0)$ at Zero Frequency

It has been discussed in Section(3.3.3) that, in existing literatures the power $S_X(0)$ at zero-frequency has been ignored in the RPDA simulation formula, in order to achieve the strong ergodicity for the generated realization. However as we have argued there that when the finite sample length is taken into account, it is more important to control the discrepancy $\epsilon_{\hat{X}}(\tau)$ between $R_X(\tau)$ and $R_{\hat{X},Trap}(\tau)$. Assuming $S_X(0) = 0$ will not help to reduce the discrepancy.

To see this, remember from Equation(3.47) that $R^0_{\hat{X},Trap}(\tau)$ is defined as the approximating covariance function with $S_X(0) = 0$. Then we can write

$$\begin{aligned} R_X(\tau) - R^0_{\hat{X},Trap}(\tau) &= R_X(\tau) - 2\Delta_\lambda \sum_{k=1}^{m} 2S_X(k\Delta_\lambda)\cos(2\pi k\Delta_\lambda\tau) \\ &= R_X(\tau) - \int_{-1/2\Delta_\lambda}^{1/2\Delta_\lambda} D^{\Delta_\lambda}_{2m+1}(\tau - t)R^{\Delta_\lambda}_X(t)dt + \Delta_\lambda S_X(0) \\ &= R_X(\tau) - R_{\hat{X},Trap}(\tau) + \Delta_\lambda S_X(0) \\ &= \epsilon_{\hat{X}}(\tau) + \Delta_\lambda S_X(0), \end{aligned} \tag{3.100}$$

where $R^{\Delta_\lambda}_X(\tau)$ is given by Equation(3.67). Consequently it is possible that the discrepancy between $R_X(\tau)$ and $R^0_{\hat{X},Trap}(\tau)$ will be dominated by the term $\Delta_\lambda S_X(0)$. Such a domination by $\Delta_\lambda S_X(0)$ is mostly likely to occur when the calculated Δ_λ is not small enough, and when the target continuous time Gaussian process contains large power at zero frequency. As an example, we consider the Oscillatory Matérn spectral density $S_X(\lambda)$ with the parameter choice

$$\sigma = 1, \phi = 1, \nu = 1, \lambda_0 = 0 \tag{3.101}$$

which is in fact the ordinary Matérn process without oscillation in the covariance function. We set $\lambda_0 = 0$ here since the corresponding spectral density function $S_X(\lambda)$ reaches its maximum at $\lambda = 0$, and therefore the term $\Delta_\lambda S_X(0)$ will have a significant

impact.

Table(3.6) compares the accuracy of the approximating covariance functions, depending on whether $S_X(0)$ being set to zero or not. In each case, the power cut-off frequency L and frequency discretization interval Δ_λ are calculated through our proposed procedures, according to a loss of variance $\epsilon_L(0) = 0.001$, and a maximum aliasing error $\epsilon_{aliasing}^{\tau_{max}} = 0.001$. With such an accuracy level, it can then be immediately observed from the table that the term $\Delta_\lambda S_X(0)$ becomes the dominating source of error. Such a domination, however, becomes less significant as τ_{max} increases. This is because a larger τ_{max} will result in a smaller Δ_λ in order to achieve the same level of accuracy, and hence reducing the magnitude of the term $\Delta_\lambda S_X(0)$. This particular numerical study therefore confirms our earlier suggestion (in Section(3.3.3)) that the traditional practice of setting $S_X(0) = 0$ is not generally helpful in reducing the simulation error, and consequently should be avoided.

$L = 23$	$\tau_{max} = 10$	$\tau_{max} = 50$	$\tau_{max} = 100$	$\tau_{max} = 500$
Δ_λ	0.0875	0.0194	0.0099	0.002
m	263	1186	2324	11500
$ShortLag - MaxErr$	0.0009927	0.0009913	0.0009914	0.001
$FarLag - MaxErr$	0.0004432	0.0004487	0.0004489	0.0004496
$ShortLag - MaxErr^0$	0.0448	0.0107	0.0059	0.002
$FarLag - MaxErr^0$	0.04375	0.0097	0.0049	0.001
$\Delta_\lambda S_X(0)$	0.04375	0.0097	0.00495	0.001

Table 3.6: Frequency domain discretization scheme for Oscillatory Matérn process with parameters $\sigma = 1, \phi = 1, \nu = 1, \lambda_0 = 0$, corresponding to $\epsilon_L(0) = 0.001$ and $\epsilon_{aliasing}^{\tau_{max}} = 0.001$. $ShortLag - MaxErr^0$ and $FarLag - MaxErr^0$ are equivalent to $ShortLag - MaxErr$ and $FarLag - MaxErr$, respectively, except $S_X(0) = 0$ in their definitions.

3.5.4 Computational Costs of Spectral Simulation Methods

The proposed algorithm for finding the appropriate frequency domain discretization scheme is potentially most useful in scenarios when the sampling times are randomized and hence varies across different realizations. Under this situation, for a stationary Gaussian process without convenient time-domain dynamics, either a matrix factorization method, or the spectral method are the only available alternatives. The matrix factorization method has a computational costs of order $O(n^3)$, whereas the

spectral method incurs a computational costs of order $O(mn)$, where n is the sample size, and m is the number of terms used in the spectral simulation formula. Hence the spectral simulation method could potentially be more efficient than the matrix factorization method, when the required number m of the terms is not very large. This could happen when the spectral density $S_X(\lambda)$ decays fast to zero as $\lambda \to \infty$ (implying a small power cut-off frequency L), or when the maximum length of the sample path, measured by $\tau_{\max}$, is not large (thus implying a relatively larger Δ_λ). This is shown in Table(3.7) and (3.8), in which both the RPDA and RPRA simulation methods are compared with the matrix factorization method (Cholesky factorization, to be specific) in terms of CPU times needed to generate a sample size of a particular length.

The stationary Gaussian process under consideration is the Oscillatory Matérn process, with fixed parameters

$$\sigma = 1, \phi = 1, \lambda_0 = 1, \tag{3.102}$$

and variable smoothness parameters

$$\nu = 0.7, \;\; 1, \;\; \text{and} \;\; 2. \tag{3.103}$$

Note that as the smoothness parameter increases, the spectral density $S_X(\lambda)$ will demonstrate faster decay as λ increases. Two different sample sizes are considered: a moderate $n = 1000$ for Table(3.7) and a reasonably large $n = 5000$ for Table(3.8). For each sample size, the sampling times are generated through a Poisson sampling scheme, in which the sampling intervals are IID exponentially distributed. The average sampling intervals are set to the following different values:

$$\Delta = \mathrm{E}[\Delta_k] = 0.1, \;\; 0.2, \;\; 0.5, \;\; \text{and} \;\; 1. \tag{3.104}$$

The frequency domain discretization schemes (i.e. L and Δ_λ) are calculated through our proposed procedures in Section(3.4.1) and (3.4.2). First thing to notice from these two tables are the observation that RPRA method is slightly less computationally efficient than the corresponding RPDA method, because of the extra m Rayleigh random variables that has to be generated. The time discrepancies between RPDA and RPRA methods, however, are not significant and hence can be

ignored in practice. Moreover, under the scenarios in which a relatively large number m of harmonics as compared to the sample size n (i.e. when $L = 94$ and $\Delta = 0.5$ or $\Delta = 1$) are required, the spectral simulation methods (both RPDA and RPRA) may be less efficient than the Cholesky factorization method. Under other scenarios when the number m of harmonics are moderate, the spectral simulation methods could give significant reduction of computational costs. The most significant reduction of computational costs happens, as expected, when the spectral density $S_X(\lambda)$ decays fast (e.g. $\nu = 2$), and when the average sampling interval is small (e.g. $\Delta = 0.1$).

$n = 1000$					
$L = 94$		$\Delta = 0.1$	$\Delta = 0.2$	$\Delta = 0.5$	$\Delta = 1$
$\nu = 0.7$	Δ_λ	0.0097	0.0051	0.002	0.0011
	m	9691	18432	47000	85455
	RPDA(sec)	0.15	0.25	0.5	1.12
	RPRA(sec)	0.2	0.29	0.57	1.16
	CHOL(sec)	0.597	0.568	0.51	0.471
$L = 23$					
$\nu = 1$	Δ_λ	0.0102	0.0052	0.0021	0.001
	m	2255	4424	10953	23000
	RPDA(sec)	0.07	0.14	0.2	0.29
	RPRA(sec)	0.1	0.18	0.25	0.34
	CHOL(sec)	0.543	0.517	0.45	0.42
$L = 5$					
$\nu = 2$	Δ_λ	0.01	0.0051	0.0021	0.00091
	m	500	981	2381	5495
	RPDA(sec)	0.02	0.03	0.07	0.14
	RPRA(sec)	0.06	0.07	0.11	0.18
	CHOL(sec)	0.5541	0.5298	0.4815	0.4476

Table 3.7: Comparison of computation times between spectral simulation method and Cholesky factorization based time domain simulation method, with a relatively small sample size $n = 1000$.

3.5.5 Examples of Realizations

In this section we present examples of realizations generated from spectral simulation methods. Both the RPDA and the PRRA realizations, together with the curve of discrepancy $\epsilon_{\hat{X}}(\tau)$, will be included. The target Gaussian process is chosen to have a Matérn covariance function, with the variance and decay parameters fixed at $\sigma =$

$n = 5000$					
$L = 94$		$\Delta = 0.1$	$\Delta = 0.2$	$\Delta = 0.5$	$\Delta = 1$
$\nu = 0.7$	Δ_λ	0.002	0.000995	0.0004	0.000198
	m	47000	94473	235000	474748
	RPDA(sec)	2.38	5.37	14.08	64.79
	RPRA(sec)	2.52	5.61	15.03	65.64
	CHOL(sec)	15.55	13.63	12.14	11.94
$L = 23$					
$\nu = 1$	Δ_λ	0.002	0.001	0.00039	0.0002
	m	11500	23000	58975	115000
	RPDA(sec)	0.89	1.28	2.99	5.94
	RPRA(sec)	0.95	1.43	3.09	6.73
	CHOL(sec)	14.18	12.11	11.12	10.69
$L = 5$					
$\nu = 2$	Δ_λ	0.002	0.001	0.0004029	0.0002
	m	2500	5000	12411	25013
	RPDA(sec)	0.35	0.66	0.98	1.48
	RPRA(sec)	0.38	0.71	1.07	1.59
	CHOL(sec)	14.61	12.75	11.73	11.36

Table 3.8: Comparison of computation times between spectral simulation method and Cholesky factorization based time-domain simulation method, with a relatively large sample size $n = 5000$.

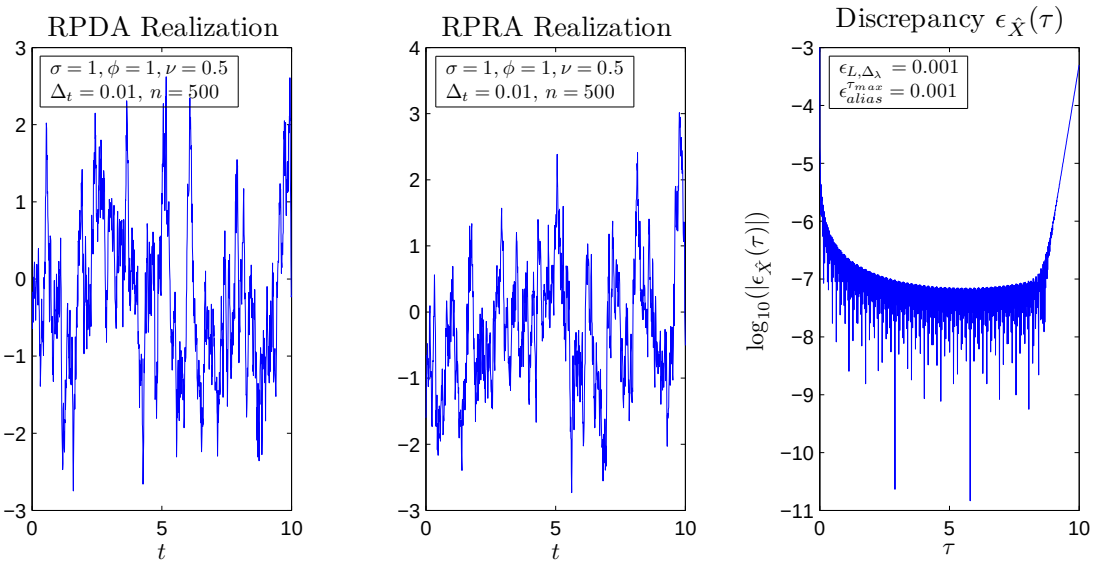

Figure 3.5: Realization of Matérn process, with parameters $\sigma = 1, \phi = 1, \nu = 0.5$.

$1, \phi = 1$, and the smoothness parameter chosen from $\nu = 0.5, 1, 2, 3$. Sample sizes are 500 in all cases, and the irregular sampling times are uniformly distributed over the time interval $[0, 10]$. It can be seen from Figure(3.5)-(3.8) that the realizations becomes smoother as ν increases in value.

3.6 Summary

In this chapter we reviewed methods for generating realizations from continuous time stationary Gaussian processes. When sampling times are equally spaced, the method of circulant embedding is the most promising in existing literature, thanks to its utilization of the efficient FFT technique. Although the method of circulant embedding is a time domain method, it has a spectral domain interpretation in the sense that the simulation formula can be regarded as an approximation to the spectral representation of the discrete time stationary process $Y = Y_k$ obtained from equally spaced sampling. One interesting feature of this spectral domain interpretation is the fact that the power of the process $Y = Y_k$ is approximated from the knowledge of covariance sequences.

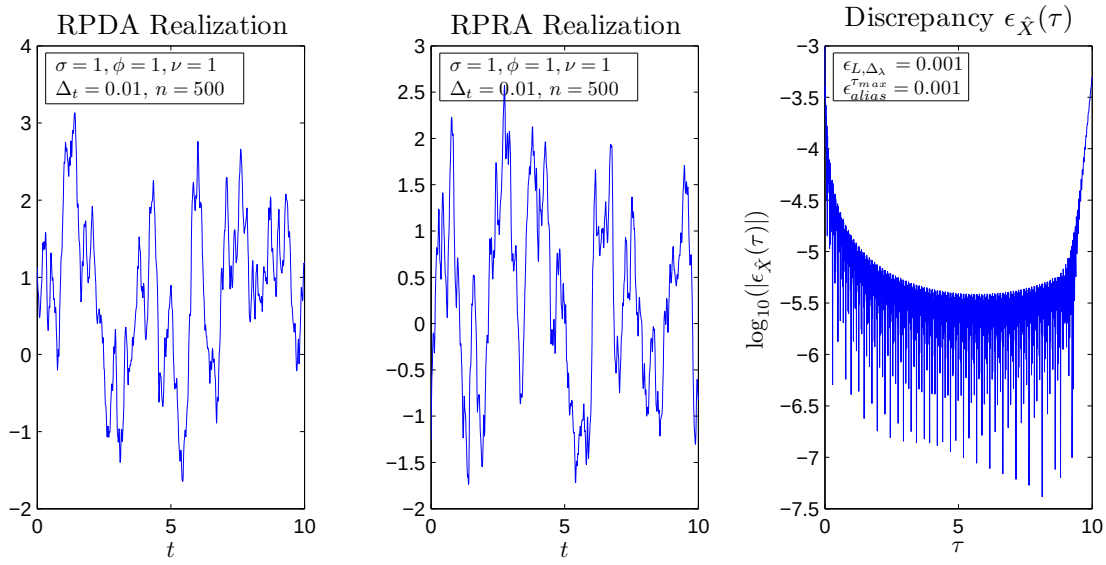

Figure 3.6: Realization of Matérn process, with parameters $\sigma = 1, \phi = 1, \nu = 1$.

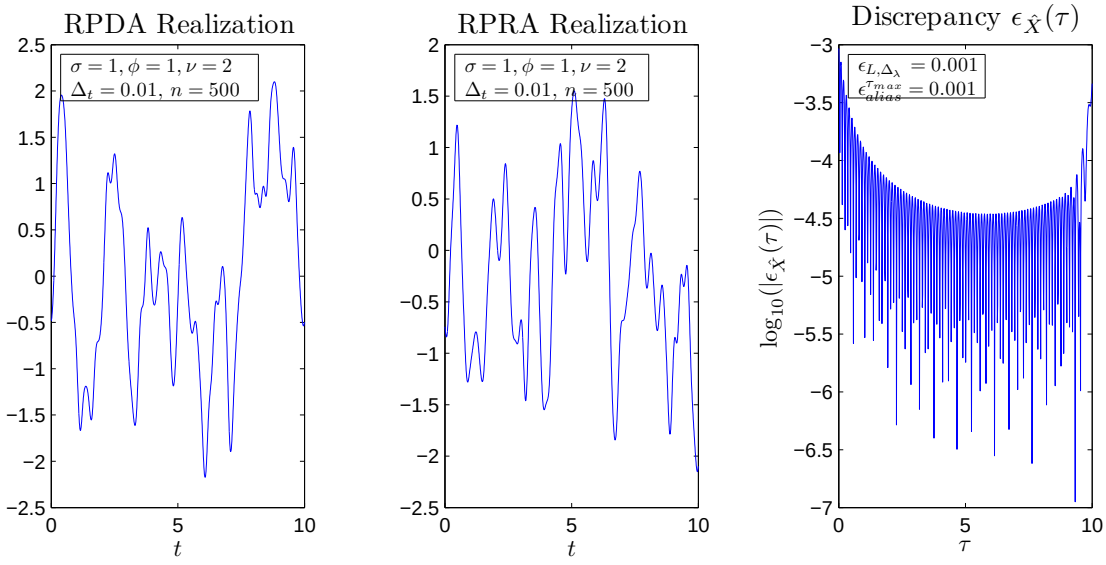

Figure 3.7: Realization of Matérn process, with parameters $\sigma = 1, \phi = 1, \nu = 2$.

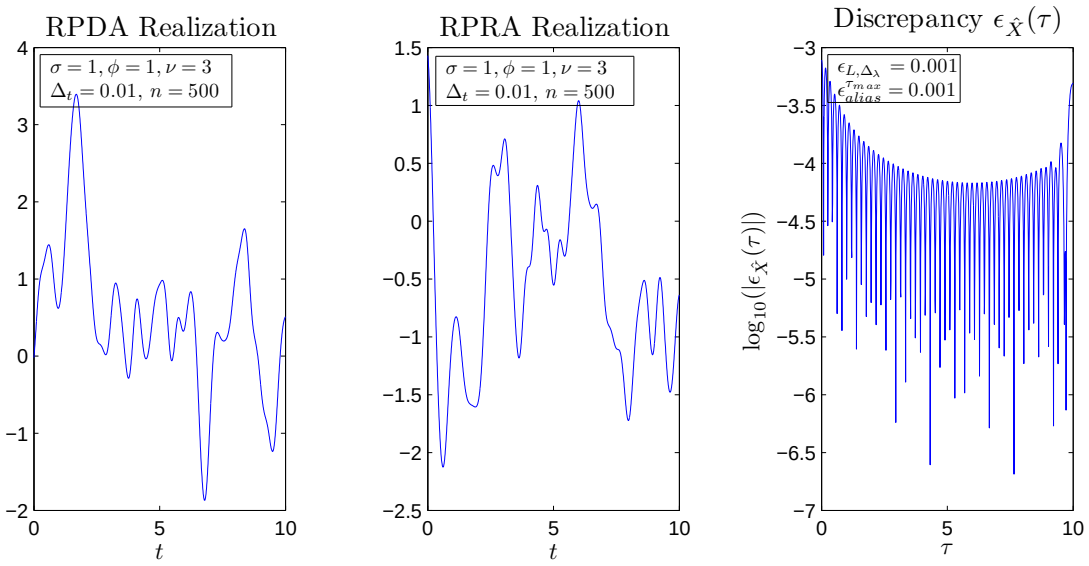

Figure 3.8: Realization of Matérn process, with parameters $\sigma = 1, \phi = 1, \nu = 3$.

On the other hand when the sampling times are unequally spaced, the method of circulant embedding is no longer applicable. Hence we need to use either the exact matrix factorization method or the non-exact spectral method. We focused in this chapter on the spectral simulation method, which generates realizations from a discretization of the spectral representation of the target process. We gave detailed review of the two major variations, namely the RPDA and RPRA methods. After discussions about their properties, we concluded that in situations most often encountered in practice, both methods should be able to provide adequate approximations to the statistical properties of the target process, as long as the discrepancy $\epsilon_{\hat{X}}(\tau)$ between $R_{\hat{X},Trap}(\tau)$ and $R_X(\tau)$ is controlled at a reasonable level.

Then we proceeded to discuss the implementation of the spectral simulation method, focusing on the construction of the spectral domain discretization scheme. We showed that the power cut-off frequency L can be conveniently obtained from an application of the Newton-Raphson recursion. Moreover a novel algorithm was proposed to determine appropriate frequency domain discretization interval Δ_λ, from the consideration of the time domain aliasing error. We then argued that, when the spectral simulation method is implemented with these proposed algorithms, we will

have sufficient control over the discrepancy $\epsilon_{\hat{X}}(\tau)$, while at the same time keep the computational costs of the spectral simulation method at a manageable level.

Chapter 4

Sampling and Estimation of Continuous Time Stationary Processes

4.1 Introduction

Having completed the discussion of simulation problem under irregularly spaced sampling times, we now turn our attention to the second main problem of this thesis, namely the problem of parametric estimation under irregularly spaced sampling times. This chapter will provide a concise overview of the parametric estimation problem under various sampling schemes. We will outline the drawbacks of the existing methods, and propose the idea of a new general framework for the parametric estimation problem under irregularly spaced sampling times.

A common approach to perform parametric estimation is to construct a log-likelihood function in terms of the unknown parameter $\boldsymbol{\theta}$, whose minimum should give the finite sample estimator of the true parameter [20]. When the sampling times are regarded as deterministic, the traditional approach is to construct the classical Gaussian log-likelihood function of the form

$$GLL_n(\boldsymbol{\theta}) = \frac{1}{n}\log(|\Sigma_n(\boldsymbol{\theta})|) + \frac{1}{n}\mathbf{X}_n^T\Sigma_n(\boldsymbol{\theta})^{-1}\mathbf{X}_n, \tag{4.1}$$

where $\mathbf{X}_n = (X(t_1), \cdots, X(t_n))^T$ is the finite sample of size n, with covariance matrix given by $\Sigma_n(\boldsymbol{\theta})$. When the underlying continuous time stationary process is assumed

to be Gaussian, the above likelihood function is exact, and the corresponding estimator is the well-known maximum likelihood estimator [1]. On the other hand, even if the underlying process is not Gaussian, the same procedure is still reasonable and will lead to the so called "maximum Gaussian likelihood estimates" of the parameters of the process [127, 20].

However, because the inversion of the covariance matrix has to be calculated, numerical evaluation of this Gaussian log-likelihood function is in general very expensive [66]. One way to avoid this potentially heavy computational costs is to assume some convenient time domain dynamic for the underlying continuous time stationary process. Yet another way is to regulate the sampling schemes, by assuming the sampling times being equally spaced. Under this assumption of equally spaced sampling times, the above Gaussian log-likelihood function can be approximated, at least for relatively large sample size, by the Whittle log-likelihood function which can be calculated very efficiently [136, 137, 50]. This computational efficiency, however, comes with a significant model identification issue that is caused by the aliasing effect [93, 20].

Being motivated by the problems with the traditional approach based on Gaussian log-likelihood function, we also investigated the methods in which the sampling times are modeled by a stochastic point process over the real line [113, 8, 79]. We first outlined the more recent and also the most complete framework of stochastic sampling times proposed by Masry [79], in which both non-parametric and parametric estimation problems can be rigorously formulated [70, 72]. After explaining how a pseudo log-likelihood function can be constructed, we then pointed out that the practical implementation of this approach can be very difficult. This observation therefore leads us to consider the historically earlier framework proposed by Shapiro and Silverman [113], which will be reviewed in considerable detail. Although it has been argued in existing literature that this earlier framework is not ideal for non-parametric estimation [41], we believe that the problem of parametric estimation can be conveniently formulated within this framework.

This chapter will be organized as follows. Section(4.2) discusses the parametric estimation problem under equally spaced sampling times, in which a concise review

[1]The Gaussian log-likelihood function is in fact the negative value of the logarithm of the Gaussian likelihood function. Hence the minimum of the Gaussian log-likelihood function defined here corresponds to the maximum of the Gaussian likelihood function.

of Whittle's log-likelihood function will be provided. We will also discuss the model identification issue caused by the aliasing effect. Section(4.3) discusses the parametric estimation problem under the assumption that the irregularly spaced sampling times are considered as deterministic. Section(4.4) reviews the frameworks of stochastic sampling times, outlining their advantages and disadvantages.

4.2 Equally Spaced Sampling

Equally spaced sampling, in which the sampling times are given by $t_k = t_{k-1} + \Delta_t$, is perhaps the most commonly assumed sampling schemes in practice [65, 95]. Under such a simplifying assumption, the sampled process $Y = Y_k = X(k\Delta_t)$ is a discrete time stationary process with autocovariance sequence given by

$$C_{Y,\Delta_t}(k;\boldsymbol{\theta}) = R_X(k\Delta_t;\boldsymbol{\theta}), \;\; k \in \mathbb{Z}. \tag{4.2}$$

Here $R_X(\tau;\boldsymbol{\theta})$ is the covariance function of the underlying continuous time stationary process $X = X(t)$, specified through the parameter vector $\boldsymbol{\theta}$. Estimating unknown values for $\boldsymbol{\theta}$ based on a set of finite sample $\mathbf{Y}_n = (Y_1, Y_2, \cdots, Y_n)^T = (X(t_1), X(t_2), \cdots, X(t_n))^T$, obtained from equally spaced sampling, can be efficiently performed through minimizing the Whittle's log-likelihood function [136, 137, 50].

4.2.1 Review on Whittle's Log-Likelihood Function

Whittle's log-likelihood function was originally proposed for a regular discrete time stationary process of the following form

$$Y_k = \sum_{j \geq 0} \alpha_j(\boldsymbol{\eta}) Z_{k-j}, \;\; k \in \mathbb{Z}, \tag{4.3}$$

where $\boldsymbol{\eta}$ is a vector of unknown parameters, and Z_k are IID random innovations with mean zero and variance s^2. Moreover the parameter vector $\boldsymbol{\eta}$ and the innovation variance s^2 are assumed to be independent of each other, and the coefficient $\alpha_0(\boldsymbol{\eta})$ is set to one in order to avoid model identification issue [49]. For such a class of discrete time stationary processes, Whittle [136, 137] proposed an estimator of the

true parameter vector $\boldsymbol{\eta}_0$ a value $\hat{\boldsymbol{\eta}}_n$ which minimizes the quantity

$$W_n^{\Delta_t}(\boldsymbol{\eta}) = \int_0^{1/\Delta_t} \frac{I_n^{\Delta_t}(f)}{K_Y^{\Delta_t}(f;\boldsymbol{\eta})} df. \tag{4.4}$$

Here Δ_t is the appropriate equally spaced sampling interval between observations (which in time series literature is usually assumed to be one for convenience). The function $K_Y^{\Delta_t}(f;\boldsymbol{\eta})$ is defined as

$$K_Y^{\Delta_t}(f;\boldsymbol{\eta}) = |\sum_{j\geq 0} \alpha_j(\boldsymbol{\eta}) e^{i2\pi jf}|, \tag{4.5}$$

which is proportional to the spectral density of $Y = Y_k$ [22]. On the other hand the function $I_n^{\Delta_t}(f)$ is the periodogram of the finite sample $\mathbf{Y}_n = (Y_1, Y_2, \cdots, Y_n)^T$ of size n, and is defined by

$$I_n^{\Delta_t}(f) = \frac{\Delta_t}{n} \left| \sum_{k=1}^{n} Y_k e^{i2\pi fk\Delta_t} \right|^2. \tag{4.6}$$

Whittle's interest in minimizing Equation(4.4) with respect to $\boldsymbol{\eta}$ stems from the following observation. Let $GLL_n(\boldsymbol{\eta}, s^2)$ be the classical Gaussian log-likelihood function corresponding to the finite sample $\mathbf{Y}_n$, written in this case as

$$GLL_n(\boldsymbol{\eta}, s^2) = \frac{1}{n}\log(|\Sigma_n(\boldsymbol{\eta}, s^2)|) + \frac{1}{n}\mathbf{Y}_n^T \Sigma_n(\boldsymbol{\eta}, s^2)^{-1}\mathbf{Y}_n. \tag{4.7}$$

Here $\Sigma_n(\boldsymbol{\eta}, s^2)$ is the covariance matrix of the finite sample. Then it has been shown by Hannan [49] that $\lim_{n\to\infty} \log(|\Sigma_n(\boldsymbol{\eta}, s^2)|) = s^2$, and hence the first term in $GLL_n(\boldsymbol{\eta}, s^2)$ (i.e. the log-determinant term) will be asymptotically independent of $\boldsymbol{\eta}$. Hence the information about the parameter vector $\boldsymbol{\eta}$ is mostly contained in the quadratic term. This term can be shown as being asymptotically approximated by $W_n^{\Delta_t}(\boldsymbol{\eta})$ up to a scaling constant [49, 127]. Whittle's 1962 paper [137] gave the first systematic study of the asymptotic properties of the corresponding estimator, and was subsequently made more rigorous by Walker [134] and Hannan [49]. Brockwell and Davis [22] provide detailed discussions of the Whittle estimator with a focus on the ARMA model. Finally we mention the comprehensive reference of the asymptotic properties of the Whittle estimator by Taniguch and Kakizawa [127].

The particular form of Whittle's original log-likelihood function $W_n^{\Delta_t}(\boldsymbol{\eta})$ depends on the assumption that the innovation variance s^2 and the parameter vector $\boldsymbol{\eta}$ are

independent of each other. This however may not always be the case, and in general it is possible that both s^2 and $\boldsymbol{\eta}$ depend on a common parameter vector $\boldsymbol{\theta}$. Responding to such scenarios, Hosoya [50] proposed to minimize the following functional of the periodogram

$$\int_0^{1/\Delta_t} \left\{ \log(S_Y^{\Delta_t}(f;\boldsymbol{\theta})) + \frac{I_n^{\Delta_t}(f)}{S_Y^{\Delta_t}(f;\boldsymbol{\theta})} \right\} df \tag{4.8}$$

with respect to the parameter vector $\boldsymbol{\theta}$, where $S_Y^{\Delta_t}(f;\boldsymbol{\theta})$ is the spectral density of the discrete time process under consideration. Asymptotic properties of the corresponding estimator were also derived under appropriate regularity conditions.

In practice however, the integral in the above equation is usually replaced with a quadrature approximation over the set of Fourier frequencies $\{\frac{k}{n\Delta_t} : k = 0, 1, \cdots, n-1\}$. Such a discretized approximation over the set of Fourier frequencies will be denoted as $WLL_n^{\Delta_t}(\boldsymbol{\theta})$, and is given explicitly as:

$$WLL_n^{\Delta_t}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{k=0}^{n-1} \log\left(S_Y^{\Delta_t}\left(\frac{k}{n\Delta_t}; \boldsymbol{\theta} \right) \right) + \frac{1}{n}\sum_{k=0}^{n-1} \frac{I_n^{\Delta_t}(\frac{k}{n\Delta_t})}{S_Y^{\Delta_t}(\frac{k}{n\Delta_t};\boldsymbol{\theta})}, \tag{4.9}$$

Just as in the case of $W_n^{\Delta_t}(\boldsymbol{\eta})$, it can also be shown when n is relatively large that, $WLL_n^{\Delta_t}(\boldsymbol{\theta})$ approximates the Gaussian log-likelihood function [127].

The motivation for using Equation(4.9) is from the consideration of computational efficiency: instead of inverting a covariance matrix every time the exact Gaussian log-likelihood function is calculated, the Whittle's log-likelihood function only requires a one-time evaluation of the periodogram $I_n^{\Delta_t}(f)$ over the set of Fourier Frequencies, which can be done efficiently through the FFT technique with a computational costs of $O(n\log(n))$ floating-point-operations. On the other hand, being only an approximation to the exact Gaussian log-likelihood function, Whittle's log-likelihood function requires the sample size n to be relatively large, especially when the sampling procedure resulting a discrete time sampled process $Y = Y_k$ that possesses slowly-decaying autocovariance sequences (see for example Contreras-Cristan,Gutierrez-Pena and Walker [24]).

In such situations, as a result of the Heisenberg uncertainty principle(see for example Gasquet and Witomski [40]), this slowly-decaying autocovariance sequences implies a significant concentration of power over a narrow frequency range. Hence the sample size has to be large enough so that the set of Fourier frequencies can provide

sufficient resolution at the frequency region where the power is concentrating. Moreover, in order to take full advantage of the Whittle's log-likelihood function, the spectral density $S_Y^{\Delta_t}(f;\boldsymbol{\theta})$ has to be relatively easy to evaluate. This, however, may not be always possible, because of the aliasing effect introduced from the equally spaced sampling in time domain will not in general give $S_Y^{\Delta_t}(f)$ a closed form expression[113].

4.2.2 Aliasing Effect and Model Ambiguity

Aliasing effect in signal processing refers to the fact that under equally spaced sampling, the spectral density $S_Y^{\Delta_t}(f)$ of the sampled process $Y = Y_k$ is given by [95]

$$S_Y^{\Delta_t}(f) = \sum_{k=-\infty}^{\infty} S_X(f + \frac{k}{\Delta_t}). \tag{4.10}$$

The above equation describes the aliasing effect that has been discussed intensively in signal processing and engineering literature [65, 100, 95]: the power at frequency $f \in \left(-\frac{1}{2\Delta_t}, \frac{1}{2\Delta_t}\right]$ of the sampled process $Y = Y_k$ consists of the power of the continuous time process at frequency $\lambda = f$, together with the power over a countable set of aliased frequencies $\{\lambda = f + \frac{j}{\Delta_t} : j = \pm 1, \pm 2, \cdots\}$. The frequency $\lambda \in (-\frac{1}{2\Delta_t}, \frac{1}{2\Delta_t}]$ of the continuous time stationary process is then said to be aliased with each of the frequencies $\{\lambda + \frac{j}{\Delta_t}; j = \pm 1, \pm 2, \cdots\}$. The latter frequencies are called the aliases of the frequency $\lambda \in \left(-\frac{1}{2\Delta_t}, \frac{1}{2\Delta_t}\right]$. The highest frequency of the continuous time stationary process that is not an alias of a lower frequency is $\lambda = \frac{1}{2\Delta_t}$, often called the Nyquist frequency or the folding frequency.

This aliasing effect will pose a problem if the interest is to obtain information of the underlying spectral density $S_X(\lambda)$ via digital methods through equally spaced sampling. From the point of view of non-parametric estimation, it is expected that the spectral density $S_Y^{\Delta_t}(f)$ is the best we can recover from observing $Y = Y_k$, and in general there is no one-to-one correspondence between $S_Y^{\Delta_t}(f)$ and $S_X(\lambda)$ [113]. Consequently even very accurate estimation of $S_Y^{\Delta_t}(f)$ may be very poor estimates of $S_X(\lambda)$, because $S_Y^{\Delta_t}(f)$ contains biases introduced from aliased frequencies.

On the other hand, this aliasing effect will also cause potential model identification issues for the parametric estimation problem. This is because the high frequency spectral features of $X = X(t)$ (i.e. spectral features beyond the Nyquist frequency λ_N) will be reproduced as low frequency spectral features of $Y = Y_k$. Consequently

more than one spectral density $S_X(\lambda)$ of $X = X(t)$ may give more or less the same second order properties for $Y = Y_k$, thus creating multiple local minimums over the likelihood surface.

As an illustration of this model identification problem, we consider two spectral densities corresponding to two different Oscillatory Matérn processes (introduced in Section(2.3.3)). These processes are sampled through an equally spaced sampling scheme with $\Delta_t = 0.25$, and the corresponding Nyquist frequency is given by $\lambda = 2$. The first spectral density $S_X^1(\lambda)$ has a spectral peak at $\lambda_0 = 1$, and the second spectral density $S_X^2(\lambda)$ has a spectral peak at $\lambda_0 = 3$. Figure(4.1) plots the spectral density $S_Y^{\Delta_t}(f)$ of the sampled process, together with $S_X^1(\lambda)$ and $S_X^2(\lambda)$ of the underlying continuous time processes. Note that on the left hand side of Figure(4.1), most of the

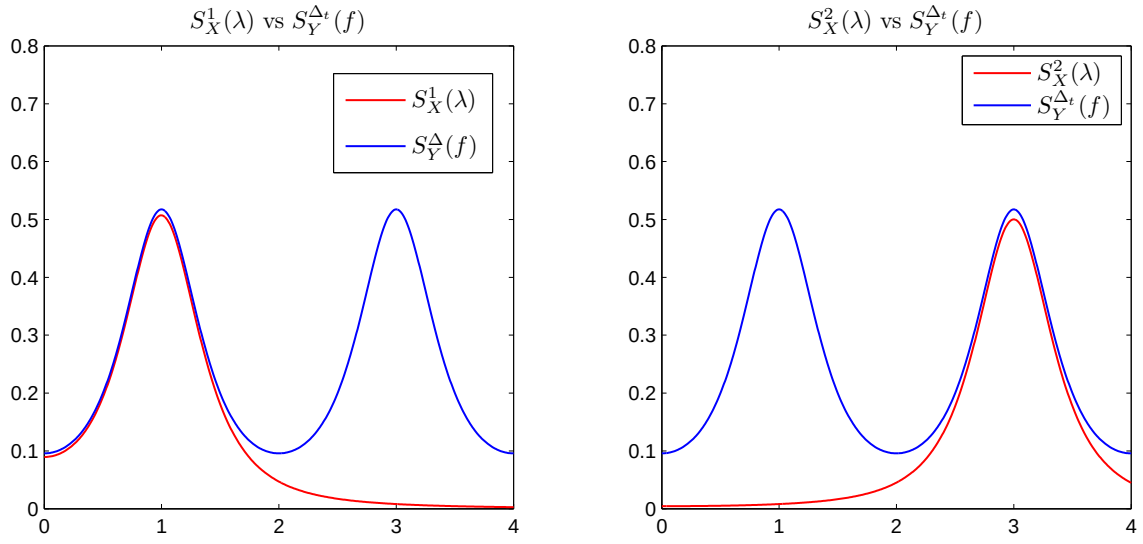

Figure 4.1: Demonstration of aliasing under equally spaced sampling scheme, using Oscillatory Matérn process, sampled with $\Delta_t = 0.25$.

power of the continuous time process is concentrated below the Nyquist frequency. Hence the spectral density $S_Y^{\Delta_t}(f)$ captures the spectral features of $S_X^1(\lambda)$ by showing a similar spectral peak at frequency $f = 1$. On the other hand, as shown by the right hand side of the same figure, an exactly the same peak of $S_Y^{\Delta_t}(f)$ at $f = 1$ may also be caused by a spectral peak of the continuous time process at a higher

aliased frequency $\lambda = 3$ beyond the Nyquist frequency. Consequently the Whittle log-likelihood estimation procedure based on equally spaced sampling will not be able to differentiate between, say $S_X^1(\lambda)$ and $S_X^2(\lambda)$ (and other spectral models with spectral peak at even higher frequencies).

4.3 Deterministic Irregularly Spaced Sampling Times

Having discussed the parametric estimation problem under equally spaced sampling schemes, we now turn our attention to the case of unequally spaced sampling times. Depending on whether the sampling times are regarded as deterministic or random, there may be different methods available to solve the parametric estimation problem.

When the sampling intervals are not regular, we can always treat the sampling times as deterministic, and then perform the parametric estimation through constructing the full Gaussian log-likelihood function. This general approach is however, computationally very expensive, especially when the sample size n is large. This is because the calculation of the Gaussian log-likelihood function involves taking inversion of the n-by-n covariance matrix, and generally requires an $O(n^3)$ floating-point-operations [66].

This computational issue of inverting a large covariance matrix may be avoided if we assume the underlying continuous time process $X = X(t)$ has certain dynamic structure. A common model for such a convenient dynamic structure is given by the family of Continuous time ARMA (CARMA) processes, which is briefly introduced in Section(2.3). By writing it in terms of an appropriate state-space form, the exact Gaussian log-likelihood function can be conveniently calculated recursively through applying the Kalman recursion technique [22].

This approach has been used by Jones [58] for the likelihood fitting of an CAR process with arbitrary sampling patterns, and also by Jones and Ackerson [59] in their analysis of longitudinal data using CARMA processes. Although Kalman recursion offers an analytical way of calculating the exact Gaussian log-likelihood function, there are still some potential computational issues. For one thing this method requires to calculate integrals with respect to matrix exponents of the form $\exp(As)$ for some known matrix A, which could be a numerically challenging object. Jones[58] used the approach of Jordan decomposition of the matrix A to evaluate $\exp(As)$ and the corresponding integrals. More sophisticated algorithms, representing the last decades

development, can be found in Moler and Van Loan [88]. Also when implementing the numerical minimization routine, care has to be taken with respect to the parameter space in order to ensure that stationarity is enforced [89, 97, 5]. An excellent review about the implementations of the Kalman recursion technique to the parametric estimation problem of CARMA processes is provided by Tomasson[131].

The family of CARMA process is very flexible, in the sense that its rational spectral densities can be used to capture a variety of spectral features [96, 131]. However, the CARMA family is by no means the most general, and there are important properties of the samle path that can not be effectively captured. One of the examples is the degree of smoothness of the sample path, which can can be most effectively modeled by using the Matérn class (see Section(2.3.2)). For other class of continuous time stationary process other than the CARMA family, the technique of Kalman recursion is not available due to the lack of convenient time domain dynamic.

Under such situations, a common and heuristic approach in existing literature is to interpolate the data to an equally spaced sampling grid [2], and has been successfully applied to irregularity caused by missing values [109]. While it may be reasonable to deal with the minor irregularity in sampling times caused by missing values, the interpolation procedure will typically change the dynamic of the underlying process, hence creating biases for the estimated parameters [37]. When the sampling times are truly unequally spaced, the biases caused by interpolation procedure can be more significant, and there is little understanding of which particular interpolation method that is the most appropriate on a given data set [37].

4.4 Stochastic Sampling Times

The problems discussed in previous section motivates us to investigate the approach of stochastic sampling times. In existing literature [113, 8, 79], the approach of stochastic sampling times has most often been related to the problem of reducing aliasing effect incurred from equally spaced sampling times. Historically this aliasing effect has been most intensively discussed under the context of non-parametric spectral estimation, in which an estimate of the spectral density is constructed from taking local weighted averages of the periodogram [95]. As being mentioned in Section(4.2), this non-parametric spectral estimate is however not generally consistent, because of the biases introduced from the aliasing effect [37]. Such biases can only be reduced by increasing

the sampling rate (or equivalently reducing the sampling interval) to such a point that only a negligibly small amount of power of the continuous time process falls outside the Nyquist frequency range. By deliberately introducing randomness into the sampling times, researchers hoped to find consistent non-parametric estimates of the spectral density of the underlying continuous time process could be constructed with significantly lower sampling rate.

In the remainder of this section, we will provide a concise review of the stochastic sampling times frameworks that have been discussed in existing literature. The focus, however, will be on parametric estimation problem instead of the non-parametric spectral estimation that have already been intensively discussed.

4.4.1 Masry's Framework of Stochastic Sampling Times

Instead of following the chronicle order, we will first introduce the more recent framework of modeling the sampling times through a stationary point process, which is assumed to be independent of the underlying continuous time process. This framework, which was proposed by Masry [79], was being motivated by Brillinger's work [19] on spectral analysis of stationary interval functions, and also by Beulter and Leneman's work [9, 67] on stationary point process as a model of stochastic sampling times. It can also be regarded as an effort to generalize the earlier work [41, 84, 80] on Poisson sampling times, in which the sampling times are generated through a Poisson point process over the real line. In fact both Roberts [41] and Masry [80] implicitly used this approach as an alternative way to analyze the sampling times generated by a Poisson point process.

One distinctive feature of this framework is the fact that the samples are taken over a fixed time span $(0, t]$, and the number $n(t)$ of samples observed is taken as random. In other words this approach models the random distribution of the irregularly spaced sampling times over the fixed interval $(0, t]$. Let $N(dt)$ be the counting measure [28] introduced by the sampling stationary point process, Masry suggested to analyze the sampled data by considering the following stationary increment process

$$Z(B) = \int_B X(t)N(dt), \tag{4.11}$$

where B is an arbitrary Borel set over the real line. The reason for considering this stationary increment process $Z = Z(\cdot)$ stems from the fact that, under rather

general conditions, it admits a spectral density $S_Z(\lambda)$ that characterize the second order properties. In fact it has been derived by Masry [79] that $S_Z(\lambda)$ can be written as

$$S_Z(\lambda) = \rho^2 S_X(\lambda) + \rho R_X(0) + \int_{-\infty}^{\infty} R_X(\tau) c_N(\tau) e^{-i2\pi\lambda\tau} d\tau \tag{4.12}$$

$$= \rho^2 S_X(\lambda) + \rho R_X(0) + \int_{-\infty}^{\infty} S_X(\lambda - u)\psi_N(u) du. \tag{4.13}$$

The dependence of $S_Z(\lambda)$ on the second order properties of the continuous time stationary process can be seen from the appearance of the covariance function $R_X(\tau)$ or spectral density $S_X(\lambda)$ from the above equations. Moreover $S_Z(\lambda)$ also depends on the second order properties of the sampling stationary point process through the mean intensity rate $\rho = \mathrm{E}[N((0,1])]$ and the covariance density $c_N(\tau)$, which is usually assumed to be integrable. The function $\psi_N(\lambda) = \int_{-\infty}^{\infty} c_N(u) e^{-i2\pi\lambda u} du$ is the Fourier transform of the covariance density $c_N(u)$.

This elegant expression for $S_Z(\lambda)$ allows rigorous discussions of both non-parametric and parametric estimation problems under this stochastic sampling times framework:

Nonparametric Spectral Estimation

Under rather general conditions over the sampling point process, Masry[79] showed that there exists a one-to-one functional correspondence between $S_X(\lambda)$ and $S_Z(\lambda)$. With extra regularity conditions imposing on the covariance function $R_X(\tau)$ and the covariance density $c_N(\tau)$, an expression for $S_X(\lambda)$ in terms of $S_Z(\lambda)$ was also derived in Masry[79] as:

$$S_X(\lambda) = \frac{1}{\rho^2}\left\{ S_Z(\lambda) - \rho R_X(0) - \int_{-\infty}^{\infty} \Gamma_N(\lambda - u)[S_Z(u) - \rho R_X(0)] du \right\}, \tag{4.14}$$

where $\Gamma_N(\lambda)$ is the Fourier transform of the function $\gamma_N(\tau) = \frac{c_N(\tau)}{\rho^2 + c_N(\tau)}$, which is assumed to be integrable. Using this expression, Lii and Masry[72] proposed a non-parametric estimator for $S_X(\lambda)$, based on finite samples, simply by replacing the quantities $S_Z(\lambda)$ and $R_X(0)$ in Equation(4.14) with finite sample estimates. Mean square consistency and other asymptotic properties of this non-parametric estimator were also derived in the indicated reference.

Parametric Estimation

Suppose the underlying continuous time stationary process is parametrized by a parameter vector $\boldsymbol{\theta}$ with true value $\boldsymbol{\theta}_0$. Assuming the sampling point process is known, then it is obvious from Equation(4.14) that $S_Z(\lambda)$ should also depend on $\boldsymbol{\theta}$, and hence will be written as $S_Z(\lambda;\boldsymbol{\theta})$. Using $S_Z(\lambda;\boldsymbol{\theta})$, Lii and Masry[70] showed that the finite parameter model can be estimated by considering the pseudo-likelihood function $K_t(\boldsymbol{\theta})$ of the form

$$K_t(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \frac{\log(S_Z(\lambda;\boldsymbol{\theta}))}{1+\lambda^2} I_t(\lambda) d\lambda. \tag{4.15}$$

Here $I_{t,Z}(\lambda)$ is the periodogram of the sampled increment process, and is defined by

$$I_t(\lambda) = \frac{1}{t}\left|\int_0^t e^{-i2\pi\lambda u} Z(du)\right|^2 = \frac{1}{t}\left|\sum_{k=1}^{n(t)} X(t_k) e^{-i2\pi\lambda t_k}\right|^2. \tag{4.16}$$

Note that in the above equation $n(t)$ is the random number of sampling points within the sampling time interval $(0,t]$. The weight function $\frac{1}{1+\lambda^2}$ is needed because $\log(S_Z(\lambda;\boldsymbol{\theta}))$ is only bounded but not necessarily integrable. The rational behind this pseudo-likelihood function was believed to be first proposed by Ibragimov[54]. The periodogram $I_t(\lambda)$ appeared in $K_t(\boldsymbol{\theta})$ serves as a proxy for the spectral density $S_Z(\lambda;\boldsymbol{\theta}_0)$ of the sampled increment process, evaluated at the true parameter value $\boldsymbol{\theta}_0$. In fact it is proved in Lii and Masry[70] that under suitable conditions, the pseudo-likelihood function $K_t(\boldsymbol{\theta})$ converges in probability to a deterministic function $K(\boldsymbol{\theta})$ as follows

$$K_t(\boldsymbol{\theta}) \xrightarrow{\mathbb{P}} K(\boldsymbol{\theta}) := \int_{-\infty}^{\infty} \frac{\log(S_Z(\lambda;\boldsymbol{\theta}))}{1+\lambda^2} S_Z(\lambda;\boldsymbol{\theta}_0) d\lambda. \tag{4.17}$$

Due to the property of the logarithmic function $\log(x) \leq x-1$, it can be easily shown that the above limit function $K(\boldsymbol{\theta})$ reaches a maximum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Therefore the parametric estimation problem, under this stationary point process sampling framework, can be solved by numerically finding the point $\hat{\boldsymbol{\theta}}_0$ which maximize the pseudo-likelihood function $K_t(\boldsymbol{\theta})$. Both the asymptotic consistency and normality of the corresponding estimator have been rigorously established by Lii and Masry[70].

Further Discussions of Masry's Framework

Masry's theory of sampling through stationary point process represents the most complete framework developed so far, within which both non-parametric and parametric estimation problems can be addressed rigorously. A wide variety of sampling patterns can be described within this framework. For example, Masry[79], Lii and Masry[71] considered an example, in which the sampling times being generated by a delayed renewal process. This delayed renewal point process is almost the same as the ordinary renewal point process, except that the distribution of the interval $\Delta_0 = t_0 - t_{-1}$ (which contains the origin) is modified so that the corresponding point process has stationary counting measure [28]. (the renewal sampling scheme is defined in Section(4.4.2) and further discussed in Chapter 5 and 6 in considerable more detail).

Masry[79] has detailed discussions about the condition under which the delayed renewal process induces one-to-one relationship between $S_X(\lambda)$ and $S_Z(\lambda)$; moreover Lii and Masry[72] contains an explicit example of a delayed renewal sampling scheme, which gives a non-parametric spectral estimate that has a smaller mean-square-error than the popular Poisson sampling process. Guidelines for the selection of specific delayed renewal sampling process in the non-parametric estimation of broadband and narrow-band spectral density functions can be found in Lii and Masry[71]. Moreover Thomson and Robinson [130] showed that the jittered sampling times (which will be clarified in the next section) can also be modified appropriately to fit into the Masry's new framework.

Although being a complete framework, its implementation however, may not be a straightforward task: both the non-parametric and the parametric estimation depend on the covariance density $c_N(u)$ of the sampling point process, which is generally not available in closed form. For example when the sampling is done through the delayed renewal process based on the IID interval with density $p(u)$ it has been shown that [28, 79] the covariance density $c_N(u)$ is given by

$$c_N(u) = \rho[h(u) - \rho], \tag{4.18}$$

where $h(u)$ is the renewal density function of the delayed renewal process (which is the same as the renewal density of an ordinary renewal process; see also Section(5.3)). It is known that this renewal density is generally not available in close form [28], unless the experimenters are able to choose rather simple probability density function $p(u)$

for the IID sampling intervals. The ability to control the sampling scheme is certainly out of question in most of the practical situations when the experimenters only collect data passively.

Furthermore, even the sampling scheme can be modeled by a simple stationary point process, so that the covariance density $c_N(u)$ is known explicitly, the calculation of the pseudo-likelihood function $K_t(\boldsymbol{\theta})$ may be computationally expensive. According to Equation(4.13), the value of $S_Z(\lambda;\boldsymbol{\theta})$ at each λ requires the numerical computation of an integral; after that the function $S_Z(\lambda;\boldsymbol{\theta})$ has to be integrated numerically over whole spectral domain $(-\infty,\infty)$ in order to obtain the pseudo-likelihood function $K_t(\boldsymbol{\theta})$. This heavy computational costs involved in the evaluation of $S_Z(\lambda;\boldsymbol{\theta})$ and $K_t(\boldsymbol{\theta})$ is certainly not desirable in practical situations.

4.4.2 An Old Framework of Stochastic Sampling Times

The practical difficulties in the implementation of Masry's more recent framework leads us to investigate an historically earlier framework, which was first considered by Shapiro and Silverman [113], and later made more rigorous by Beutler[8]. This earlier framework differs from Masry's in that the irregularity of the sampling times are modeled through stochastic sampling intervals. In particular the authors considered stochastic sampling times $\{t_k\}$, independent of $X = X(t)$, such that the sampling intervals $\{\Delta_{k,m} = t_{m+k} - t_m\}$ have distributions that are independent of m. Assuming zero mean for $X = X(t)$, it then follows from Beutler [8] that the discrete time sampled process $Y = Y_k = X(t_k)$ is also a zero mean discrete time stationary process, with autocovariance sequence given by

$$C_Y(k) = \mathrm{E}[X(t_{m+k})X(t_m)], \quad k \in \mathbb{Z}, \tag{4.19}$$

where the expectation is taken over both $X = X(t)$ and $\{t_k\}$. Through taking repeated expectations, it can be easily shown [8] that

$$C_Y(0) = R_X(0), \tag{4.20}$$

and

$$C_Y(k) = \int_0^\infty R_X(\tau)p_k(\tau)d\tau, \quad k \in \mathbb{Z}, \tag{4.21}$$

where $p_k(\tau)$ is the probability density function (independent of m) of the sampling interval $\Delta_{m,k} = t_{m+k} - t_m$. Note that under this framework of stochastic sampling times, when forming the discrete time sampled process, the actual sampling times have been ignored, and only the orders of the observations have been taken into account. This is one of the distinctive characteristic of the traditional framework of stochastic sampling times, which includes the following most commonly seen stochastic sampling patterns:

1. Jittered sampling scheme(Shapiro and Silverman [113], Beutler [8], Thomson and Robinson [130]), in which initially equally spaced sampling times are randomly perturbed, i.e. $\{t_k\}$ are generated through

$$t_k = k\Delta + \epsilon_k, \quad k \in \mathbb{Z}, \tag{4.22}$$

   where $\{\epsilon_k\}$ are the IID random noise imposed on sampling times. The standard deviations of the sampling noise has to be small, relative to Δ, in order to maintain the correct order of the sampling times, i.e. to make sure that we have $t_k > t_{k-1}$ for all k.

2. Renewal sampling scheme(Shapiro and Silverman [113], Beutler [8]), Robinson [107]), in which the consecutive sampling intervals $\Delta_k = t_k - t_{k-1}$ are IID non-negative random variables; in other words the sampling times $\{t_k\}$ are generated through

$$t_k = t_{k-1} + \Delta_k, \quad k \in \mathbb{Z}. \tag{4.23}$$

   The Poisson sampling scheme, in which the IID sampling intervals have a common exponential distribution, is an important example of the renewal sampling schemes.

3. Correlated sampling intervals (Masry [78], Robinson [107]), in which the sampling intervals $\Delta_k = t_k - t_{k-1}$ are generated by taking moving averages of a sequence of IID non-negative random variables $\{\Theta_j\}$:

$$\Delta_k = \sum_{j=-\infty}^{\infty} a_{k-j}\Theta_j, \quad k \in \mathbb{Z}. \tag{4.24}$$

4. Independent skip sampling (Beutler [8], Masry[78]), in which each sample, gen-

erated by applying either equally spaced sampling scheme, or by any of the above stochastic sampling schemes, has a probability q, $0 \leq q < 1$ of being expunged, and deletions occur independently.

The works of Shapiro and Silverman [113], Beutler [8], and later on by Masry [78], have been focused on finding the so called alias-free stochastic sampling schemes, which was first given clear formal definition by Beutler [8] as follows

Definition 4.1. *(Beutler [8]) The stochastic sampling scheme* $\{t_k\}$*, with the sampling interval* $\Delta_{m,k} = t_{m+k} - t_m$ *having distributions independent of* m*, is said to be alias-free relative to a family* $\mathcal{S}$ *of spectral distributions, if no two continuous time process* $X = X(t)$ *with different spectral distributions in* $\mathcal{S}$ *yielding the same autocovariance sequence* $C_Y(k)$ *of the sampled process* $Y = Y_k = X(t_k)$*.*

Therefore, the historical focus of this traditional concept of alias-free sampling schemes aims to establish a theoretical one-to-one correspondence between the second order properties of $X = X(t)$ and the second order properties of $Y = Y_k$. Criteria for a particular sampling scheme being alias-free, under this traditional framework, have been derived by various authors. For example for the special case renewal sampling schemes, Shapiro and Silverman [113] derived a condition in terms of the characteristic function of the common distribution of the IID sampling intervals. It was shown that the Poisson sampling scheme is alias-free with respect to the family of spectral distributions $\mathcal{S}_2$ that is absolutely continuous with spectral densities $S_X(\lambda) \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$. Later Beutler [8] extended this result by deriving alias-free conditions for the more general sampling schemes considered in Definition(4.1), and showed that the Poisson sampling scheme is in fact alias-free relatively to the family of all spectral distributions. Yet more complicated situations were considered by Masry[78], in which correlated sampling intervals of the form given by Equation(4.24) were considered, and the corresponding alias-free condition was also derived. Moreover Masry[78] also showed that the alias-free property of a stochastic sampling scheme is invariant under independent random skips with skip probability $q < 1$.

Non-Parametric Estimation

Although very general conditions for a sampling scheme to be alias-free have been established, the following specialized result is of particular interest, because it suggests

a practical way of reconstruction of the covariance function $R_X(\tau)$, or equivalently the spectral density $S_X(\lambda)$:

Proposition 4.1. *(Beutler [8]) Let $\mathcal{S}_2$ denote the family of absolutely continuous spectral distributions with spectral density $S_X(\lambda) \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$, and let $p_k(t)$ be probability density functions such that $p_k(t) \in L_1(\mathbb{R}^+) \cap L_2(\mathbb{R}^+)$. If the set of density functions $\{p_k(t) : k = 1, 2, \cdots\}$ is closed in $L_2(\mathbb{R}^+)$, then the sampling times will be alias-free in the sense that if two spectral densities differ over a set of positive measure, then they will not produce the same autocovariance sequence for the sampled discrete time process.*

Consequently under the assumptions of Proposition(4.1), the reconstruction of $R_X(\tau)$ or $S_X(\lambda)$ can be carried out, at least theoretically, from the knowledge of the autocovariances $C_Y(k)$ through a Gram-Schmidt orthogonalization procedure over the sequence of density functions $\{p_k(t) : k = 1, 2, \cdots\}$. See Gaster [41] or Masry [80] for more details of this procedure. In practical situations however, we only have a finite number of observations, and it was then suggested that the finite sample estimates $\hat{c}_Y(k)$ of the autocovariance sequence should be used to replace $C_Y(k)$ within the reconstruction algorithm [41, 81].

Such a reconstruction procedure, however, is potentially problematic for two reasons. First of all note that the Gram-Schmidt orthogonalization procedure is generally very difficult to perform, either analytically or numerically. Moreover, although under the special case of Poisson sampling scheme this Gram-Schmidt orthogonalization of densities $\{p_k(t) : k = 1, 2, \cdots\}$ can be carried out explicitly, the resulting estimate for $R_X(\tau)$ and $S_X(\lambda)$, as pointed out by Gaster and Roberts [41], turned out to be numerically very unstable. This issue was later partially answered by the theoretical work of Masry[81], who showed that under Poisson sampling scheme this numerical instability was a consequence of the slow logarithmic mean-square convergence rate for estimating $R_X(\tau)$ under the Poisson sampling scheme. It was also suggested by Lii and Masry [70] that this slow convergence may be caused by the practice that the traditional framework of stochastic sampling times ignores the actual sampling instances, and hence causes information loss.

Parametric Estimation

Therefore, this earlier framework of stochastic sampling times is not ideal for non-parametric spectral estimation. However, the fact that a discrete time stationary process can be constructed within this framework suggests the possibility of performing parametric estimation through the traditional Whittle log-likelihood function (as discussed in Section(4.2)). This possibility will be explored in considerable detail in the next two chapters. We will consider in particular the renewal sampling scheme, which can be described in both stochastic sampling frameworks. We will show in the next chapter that the parametric estimation problem can indeed be conveniently formulated under quite general conditions, and the theoretical validity of the corresponding estimator will be established. Moreover, we will argue that the proposed renewal sampling framework, for the purpose of parametric estimation, enjoys a practical advantage in the sense that the proposed method can be implemented for a wide class of renewal sampling schemes. Moreover, a novel analysis of the spectral property under renewal sampling schemes will be performed, providing more insights over the information loss under the renewal sampling scheme.

4.5 Summary

This chapter briefly reviewed the general methods for parametric estimation of a continuous time stationary process under different sampling schemes. For equally spaced sampling times, the Gaussian log-likelihood function can be approximated through the Whittle log-likelihood function and hence allows efficient estimation. The resulting aliasing effect, however, will cause model identification issue through creating multiple extrema over the likelihood surface. On the other hand, in practical situations the sampling times are frequently not equally spaced. Parametric estimation method under this setting depends on the underlying dynamic of the continuous time process, as well as whether the sampling times are treated as deterministic or stochastic. For the CARMA family of processes, Kalman recursion can be applied to allow easy evaluation of the Gaussian log-likelihood function. For more general stationary processes, however, parametric estimation can be performed through minimizing the full Gaussian log-likelihood function. This approach involves inverting a covariance matrix, and hence can be computationally very expensive when the sample size is

relatively large.

We also reviewed the existing studies about the stochastic sampling times, which historically was developed to alleviate the aliasing effect introduced from equally spaced sampling. The most complete framework for stochastic sampling times is believed to be the one proposed by Masry [79], in which the sampling times are generated from a stationary point process over the real line. Both non-parametric and parametric estimation problem has been rigorously formulated in this framework. This theoretical advantage is however coupled with a practical disadvantage, in the sense that the implementation of this framework is only possible when the experimenter can choose a relatively simple sampling scheme.

We also reviewed the framework developed by Shaprio and Silverman [113], as well as Beutler [8], in which a discrete time stationary process is constructed from the stochastic sampling times. Although this framework is not quite convenient for the purpose of non-parametric estimation of the underlying spectral density, we believe it does provide a useful framework to formulate parametric estimation problem. In particular, we will be showing in the next two chapters that, within the Shaprio, Silverman and Beutler's framework, a Whittle log-likelihood estimation procedure can be constructed under the renewal sampling schemes. Such an estimation procedure will be shown to be versatile in terms of allowing a large family of renewal sampling schemes, and also has much better large sample computational efficiency as compared to the estimation procedure based on the exact Gaussian log-likelihood function.

Chapter 5

Whittle Estimation Under Renewal Sampling

5.1 Introduction

As discussed in the last chapter, the renewal sampling scheme [113, 8] does not provide a sound framework for non-parametric estimation of the spectral density $S_X(\lambda)$ of the underlying continuous time process $X = X(t)$. In this and the next chapters, however, we will argue that the parametric estimation problem can be conveniently formulated within this renewal sampling framework. In particular, because the sampled process $Y = Y_k = X(t_k)$ is also second order stationary, we will show that it is possible to perform parametric estimation through minimizing Whittle's log-likelihood function [49, 22].

Historically Robinson [107] discussed this approach and applied it to parametric estimation problem of Continuous time AR (CAR) process with simple distinct poles in its spectral density, but detailed statistical analysis and finite sample performance were not provided. Moreover the technique used in Robinson's approach depends on the particular structure of the CAR process and therefore can not be extended to more general processes. The material presented in this chapter can be regarded as an extension of Robinson's discussion. By combining existing results from the theory of renewal point processes [118, 119], we will show that the parametric estimation problem can indeed be rigorously formulated under quite general conditions.

This chapter will be organized as follows. In Section(5.2) we will outline the working assumption on the underlying continuous time stationary processes, together with

the particular form of the Whittle log-likelihood function that we will be working with. Then in Section(5.3) we will establish a fundamental convergence result, allowing us to prove the absolute summability of the autocovariance sequence $C_Y(k)$ of the sampled process $Y = Y_k = X(t_k)$. As a consequence of this convergence result, the spectral density $S_Y^{\Delta}(f)$ of the sampled process is well defined as the Discrete Fourier Transform (DFT) of the autocovariance sequence $C_Y(k)$. In Section(5.4) we will derive an alternative frequency domain expressions for $C_Y(k)$ and $S_Y^{\Delta}(f)$, which will provide more insights over the relationship between $X = X(t)$ and $Y = Y_k = X(k\Delta_k)$. We will also see in next chapter that these frequency domain expressions will provide valuable information for both the implementation and understanding of this proposed estimation procedure. Then by using the the convergence result proved earlier, we will in Section(5.5) establish the asymptotic consistency of the proposed Whittle estimator. Finally, although we have not been successful in proving the asymptotic normality of the estimator, this will still be briefly explored in Section(5.6) through simulation studies.

5.2 Whittle Log-Likelihood Function

Consider a continuous time stationary process $X = X(t)$ with zero mean and covariance function $R_X(\tau)$. In the following discussions we will always assume this process has a short range dependence, in the sense that the covariance function is integrable over the real line, i.e. $R_X(\tau) \in L_1(-\infty, \infty)$. This is also equivalent to assuming that the spectral density $S_X(\lambda)$ can be expressed as the Fourier transform of the covariance function, and consequently is everywhere continuous [26].

When such a process is sampled through a renewal point process $\{t_k : k = \mathbb{Z}\}$ over the real line, the sampling instances are generated according to the following recursive relation $t_k = t_{k-1} + \Delta_k$, where $\{\Delta_k : k = \mathbb{Z}\}$ is a sequence of IID non-negative random variables, with common probability density function $p(\tau)$. Furthermore, to facilitate future discussions let us denote the common mean and the common variance of the sampling intervals as Δ and $\mathrm{Var}[\Delta]$, i.e. we have

$$\Delta = \mathrm{E}[\Delta_k], \text{ and } \mathrm{Var}[\Delta] = \mathrm{Var}[\Delta_k], \ k \in \mathbb{Z}. \tag{5.1}$$

Then it has been mentioned in Section(4.4) that the parametric estimation problem

could be addressed through a Whittle log-likelihood function, which is constructed from the sampled discrete time process $Y = Y_k = X(t_k)$ by ignoring the actual sampling instances.

Let us first recall the review of Whittle log-likelihood function in Section(4.2.1). It is then easy to see that based on the nature of the problem, an explicit parameterization in the form of a linear process, given by Equation(4.3), will not generally be possible. Consequently the Whittle log-likelihood $WLL_n^{\Delta_t}(\boldsymbol{\theta})$ given by Equation(4.9) should be considered in our scenario. However, because the sampling times are irregularly spaced, the quantity Δ_t is not defined. In order to make sure both the discrete time sampled process and the underlying continuous time process are on the same time/frequency scale, at least in an average sense, we will therefore take the average sampling interval Δ into account. Hence the Whittle's log-likelihood function considered in this thesis, which will be denoted by $WLL_n^{\Delta}(\boldsymbol{\theta})$, is of the following form

$$WLL_n^{\Delta}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{k=0}^{n-1} \log\left(S_Y^{\Delta}\left(\frac{k}{n\Delta};\boldsymbol{\theta}\right)\right) + \frac{1}{n}\sum_{k=0}^{n-1} \frac{I_n^{\Delta}(\frac{k}{n\Delta})}{S_Y^{\Delta}(\frac{k}{n\Delta};\boldsymbol{\theta})}, \tag{5.2}$$

where the periodogram is given by

$$I_n^{\Delta}(f) = \frac{\Delta}{n}\left|\sum_{k=1}^{n} Y_k e^{i2\pi f k\Delta}\right|^2. \tag{5.3}$$

The corresponding estimator obtained from minimizing this Whittle log-likelihood $WLL_n^{\Delta}(\boldsymbol{\theta})$ will be called the Whittle estimator, and will be denoted as $\hat{\boldsymbol{\theta}}_n$, in other words we have the following definition:

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, WLL_n^{\Delta}(\boldsymbol{\theta}). \tag{5.4}$$

The validity of $\hat{\boldsymbol{\theta}}_n$ as an estimator depends first and foremost on the existence and positivity of the spectral density $S_Y^{\Delta}(f;\boldsymbol{\theta})$. In the next section, we will prove a key convergence result that will allow us to show the absolute convergence of the autocovariance sequence $C_Y(k;\boldsymbol{\theta})$, hence establishing the existence of $S_Y^{\Delta}(f;\boldsymbol{\theta})$. Moreover, we will see that $S_Y^{\Delta}(f;\boldsymbol{\theta})$ defined above will admit a spectral domain expression, which can be used to easily show its positivity under rather general conditions.

5.3 Existence of $S_Y^{\Delta}(f)$

It has been discussed in Section(4.4) that the sampled discrete time process $Y = Y_k$ is also weakly stationary with autocovariance sequence given by

$$C_Y(k) = \int_0^{\infty} R_X(\tau) p_{|k|}(\tau) d\tau, \;\; k \in \mathbb{Z}. \tag{5.5}$$

Under the particular case of renewal sampling, $p_{|k|}(\tau)$ denotes the $|k|$-fold convolution of the density $p(\tau)$ with itself, and $p_0(\tau) = \delta_0(\tau)$, the Dirac delta function centered at zero. In order to apply the Whittle's log-likelihood function, it is important that the spectral density $S_Y^{\Delta}(f)$ of the sampled process $Y = Y_k$ exists and is positive for all f. This section therefore is devoted to a discussion of the existence of the function $S_Y^{\Delta}(f)$.

A sufficient condition is to require the autocovariances $C_Y(k)$ being absolutely summable [22]. In this case the spectral density $S_Y^{\Delta}(f)$ exists as the Discrete Fourier transform (DFT) of the autocovariances [22]:

$$S_Y^{\Delta}(f) = \Delta \sum_{k=-\infty}^{\infty} C_Y(k) e^{-i2\pi f k \Delta}. \tag{5.6}$$

For a CAR process, Robinson [107] showed that $C_Y(k)$ has the form of the autocovariance sequence of a discrete time ARMA process. This in particular implies that $C_Y(k)$ decays exponentially fast and hence is absolutely summable [22]. Masry and Lui [84] on the other hand showed that, under Poisson sampling scheme in which the sampling intervals are IID exponentially distributed, the autocovariance sequence $C_Y(k)$ is also absolutely summable for general covariance functions $R_X(\tau)$. Their technique, however, depends on the particular properties of the density of the exponential distribution, and therefore cannot be generalized.

In this section we will extend these existing discussions to show the absolute summability of the autocovariance sequence $C_Y(k)$ under more general conditions. This is achieved through establishing a more general convergence result, which will also play a key role in establishing the asymptotic consistency of the Whittle log-likelihood estimator $\hat{\boldsymbol{\theta}}_n$.

To begin the discussion, let's first note that we have the following preliminary

calculation:

$$\lim_{M\to\infty}\sum_{k=1}^{M}|C_Y(k)| = \int_0^\infty |R_X(\tau)| \lim_{M\to\infty}\sum_{k=1}^{M} p_k(\tau)d\tau, \tag{5.7}$$

which can be easily justified by applying the Monotone Convergence Theorem [143], since the non-negativity of $p_k(\tau)$ implies the partial sum $\sum_{k=1}^{M} p_k(\tau)$ is non-decreasing as M increases. Therefore it can be easily seen that the absolute summability of the autocovariance sequence $C_Y(k)$ is closely related to the properties of the so called renewal density function $h(\tau)$, which is defined in the renewal process literature [118, 119] as

$$h(\tau) = \lim_{M\to\infty}\sum_{k=1}^{M} p_k(\tau). \tag{5.8}$$

When the sampling scheme is a Poisson process, Masry and Lui[84] showed that $h(\tau) = \frac{1}{\Delta}$. [1] Consequently the absolute summability of $C_Y(k)$ can be trivially established. For other more general renewal sampling schemes, however, the renewal density $h(\tau)$ may not have a closed form expression, and hence the absolute summability of the autocovariance sequence $C_Y(k)$ may not be so obvious. However, because of the assumed integrability of $R_X(\tau)$, it can be immediately recognized from Equation(5.7) that a sufficient condition for the absolute summability of $C_Y(k)$ would be the boundedness of the renewal density $h(\tau)$. This desired boundedness property has been established by Wold [139] in his studies of point process. We shall start by listing the set of assumptions required for his results to hold:

Assumption 5.1.

1. $\Delta = E[\Delta_k] = \int_0^\infty \tau p(\tau)d\tau < \infty$
2. $\lim_{\tau\to\infty} p(\tau) = 0.$

[1] The argument is straightforward and goes as follows: under the assumption of a Poission sampling scheme, the function $p_k(\tau)$ is a gamma density and can be written as

$$p_k = \frac{\rho^k}{(k-1)!}\tau^{k-1}e^{-\rho\tau}, \tag{5.9}$$

where $\rho = \frac{1}{\Delta}$. Then summing this $p_k(\tau)$ over $k = 1, 2, \cdots$ gives

$$\sum_{k=1}^{\infty} p_k(\tau) = \rho e^{-\rho\tau}\sum_{k=1}^{\infty}\frac{\rho^{k-1}}{(k-1)!}\tau^{k-1} = \rho e^{-\rho\tau}e^{\rho\tau} = \rho = \frac{1}{\Delta}. \tag{5.10}$$

3. *$p(\tau)$ is bounded for all τ;*

Using the above listed assumptions, Wold derived the following boundedness property for the renewal density:

Lemma 5.1. *Under Assumption(5.1) the renewal density function $h(\tau)$ has the following properties*

1. *The renewal density function $h(\tau) = \sum_{k=1}^{\infty} p_k(\tau)$ converges uniformly on any finite interval, and hence is continuous.*

2. *The renewal density function $h(\tau)$ is bounded.*

While the first and second of Wold's assumption can be regarded as purely technical and regularizes $p(\tau)$ by removing those pathological densities, the third assumption turns out to be restrictive. For example when we are modeling the IID sampling interval by a Gamma(α, β) distribution, it may happen that the average sampling interval Δ may be smaller than the standard deviation $\sqrt{\text{Var}[\Delta]}$, where Var$[\Delta]$ denotes the variance of the sampling interval. In this case we will have $0 < \alpha < 1$, so that the gamma density function is not bounded at the origin, and consequently is not covered by Lemma(5.1). The next section is therefore devoted to a remedy for this situation, and an alternative boundedness property will be developed to allow a wider class of sampling interval densities.

5.3.1 Boundedness of the Renewal Density

The key to the relaxation of the Wold's condition is the observation that $|R_X(\tau)|$ is bounded and $p_k(\tau)$ is a probability density function for each k. Consequently their product $|R_X(\tau)|p_k(\tau)$ is integrable, and we could allow the first few $p_k(\tau)$s to be unbounded, knowing that their product with $|R_X(\tau)|$ will still give us finite integrals. We will therefore work with a set of less strict assumptions over the sampling interval density $p(\tau)$:

Assumption 5.2.

1. *$\Delta = E[\Delta_t] = \int_0^{\infty} \tau p(\tau) d\tau < \infty$;*

2. $\lim_{\tau\to\infty} p(\tau) = 0$;

3. $p(\tau) \in \mathcal{L}_{1+\delta}(0,\infty)$ *for some* $\delta > 0$;

4. $p(\tau)$ *has infinite support, and hence* $P(t) < 1$ *for all* $t > 0$, *where* $P(t)$ *is the cumulative distribution function of the density* $p(\tau)$.

Part (1) and (2) of Assumption(5.2) carry over from Wold's conditions and removing those pathological densities. The uniform boundedness of $p(\tau)$ in part (3) of Assumption(5.1) is replaced by the more general integrability in part (3) of Assumption(5.2), which we will soon see allows removing the singularity in the density $p(\tau)$ by repeated convolution of the density with itself. Part (1), (2) and (3) of Assumption(5.2) together ensures that the classical renewal theorem applies so that the renewal density $h(\tau)$ is well behaved as $\tau \to \infty$. Part (4) of Assumption(5.2) is purely technical, and is required by the particular proof technique employed here. It excludes sampling interval density with a compact support.

Because of this last technical assumption, Assumption (5.2) can not completely replace the older Assumption(5.1) of Wold. For example, although Assumption(5.2) allows the IID sampling interval Δ_k to have a Gamma(α,β) distribution with $\alpha < 1$, the simple case of Δ_k being Uniform(a,b) distributed is only being covered by the old Assumption(5.1). The one case that is not included in both set of assumptions is when Δ_k has an unbounded density over a compact support, e.g. a Beta(α,β) density with $0 < \alpha < 1$ or $0 < \beta < 1$. The extension to include this last case involves finding a new proof technique and is still under investigation.

Following Assumption(5.2) the probability density $p(\tau)$ possesses the following properties that will be used later in proving the boundedness of the renewal density.

Lemma 5.2.

1. *Under Assumption(5.2)(2) and (3), there exists an integer* j *such that* $p_k(\tau)$ *is continuous and bounded from above, by* M *say, for all* $\tau \geq 0$ *and all* $k \geq j$;

2. *Under Assumption(5.2)(3), we have* $\lim_{\tau\to\infty} p_k(\tau) = 0$ *for all* k.

3. *Under Assumption(5.2)(1-3), the renewal density* $h(\tau)$ *satisfies* $\lim_{\tau\to 0} h(\tau) = \frac{1}{\Delta}$.

The boundedness of $p_k(\tau)$ in Lemma(5.2)(1) follows from Assumption(5.2)(3), and the proof can be found in Lemma(8) of Smith [118] and the ensuing remarks. Continuity follows from the fact that if k is large enough so that $p_{k-1}(\tau)$ is bounded by M, then

$$\begin{aligned}\lim_{\epsilon\to 0}|p_k(\tau+\epsilon)-p_k(\tau)| &\leq \lim_{\epsilon\to 0}\int |p(\tau+\epsilon-u)-p(\tau-u)|p_{k-1}(u)du \\ &\leq \lim_{\epsilon\to 0} M\int |p(\tau+\epsilon-u)-p(\tau-u)|du \\ &= 0 \end{aligned} \tag{5.11}$$

where the last limit comes from the integrability of the density $p(\tau)$, and this shows the continuity of $p_k(\tau)$. See for example Gasquet and Witomski [40] for a complete proof of this statement. The proof of Lemma(5.2)(2) can be found in Smith [119]. Lemma(5.2)(3) is the standard renewal theorem in terms of renewal density, and can also be found in Smith [118]. With these desired properties of $p(\tau)$, we can now state a new boundedness property for the renewal density $h(\tau)$, similar to the one in Lemma(5.1). This is given as below:

Lemma 5.3. *Under Assumption(5.2) we have the following results:*

1. *There exists a positive integer j such that $h_j(\tau)=\sum_{k=j}^{\infty}p_k(\tau)$ converges uniformly on any finite interval, and hence is continuous.*

2. *The function $h_j(\tau)=\sum_{k=j}^{\infty}p_k(\tau)$ is bounded.*

Proof: Part(1): Let j be the positive integer in Lemma(5.2)(1), such that $p_k(\tau)\leq K$ for all $k\geq j$. Also let T be the right end point of an arbitrary but fixed finite interval. Then for any $\tau\leq T$ we have

$$p_{j+1}(\tau)=\int_0^{\tau} p_j(\tau-u)p(u)du \leq KP(\tau)\leq KP(T). \tag{5.12}$$

By a straightforward mathematical induction it follows immediately that

$$p_{j+l}(\tau)\leq KP(T)^l, \tag{5.13}$$

for any positive integer l, and where $P(\tau)$ is the cumulant distribution function of the density $p(\tau)$. By Assumption(5.2)(4) we have $P(T)<1$, and this gives

$\sum_{l=0}^{\infty} P(T)^l < \infty$. Hence the sequence of functions $\{p_{j+l}(\tau) : l = 0, 1, \cdots\}$ is bounded by a absolutely summable sequence of real numbers $\{KP(T)^l : l = 0, 1, \cdots\}$. Consequently by Weierstrass M-test [110] we conclude that $\sum_{k=j}^{\infty} p_k(\tau)$ converges uniformly over any finite interval. Since each $p_k(\tau)$ is continuous for $k \geq j$, the function $h_j(\tau)$ is also continuous by property of uniform convergence [110].

Part(2): By Lemma(5.2)(2) and (3) the function $h_j(\tau)$ tends to $\frac{1}{\Delta}$ as $\tau \to \infty$. Then the continuity of $h_j(\tau)$ immediately implies boundedness. This completes the proof of this Lemma. □.

This hence concludes the discussion of the boundedness properties of the renewal density function $h(\tau)$. As a summary, note that Lemma(5.1) and (5.3) established the desired boundedness property of $h(\tau)$ under different set of assumptions for the density function of the IID sampling intervals. This boundedness property of $h(\tau)$ will be used in the next section to establish the absolute summability of the autocovariance sequence $C_Y(k)$, hence implying the existence of the spectral density function $S_Y^{\Delta}(f)$.

5.3.2 A Convergence Result

The boundedness property in Lemma(5.1) and (5.3) will allow us to establish the desired absolute summability of the autocovariance sequence $C_Y(k)$, and hence the spectral density $S_Y^{\Delta}(f)$ exists as the DFT of the autocovariance sequence. We will, however, state this result in a slightly more general form, which will also be used to prove the mean-square consistency of the sample autocovariance sequence in a later section.

Proposition 5.1. *Suppose $p(\tau)$ is the probability density function of the IID sampling intervals, satisfying either Assumption(5.2) or Assumption(5.1). Let $g(\tau) \in L_1(0, \infty)$ be bounded by a positive constant G. Then the sequence $\{D_k, k = 1, 2, \cdots\}$ defined by*

$$D_k = \int_0^{\infty} g(\tau) p_k(\tau) d\tau, \;\; k = 1, 2, \cdots, \tag{5.14}$$

is absolutely summable.

Proof: To show absolute convergence, let j equal to one if Assumption(5.1) holds, or j equal to the positive integer in Lemma(5.3) when Assumption(5.2) holds. Then

by applying the Monotone Convergence Theorem [143] we can write

$$\begin{aligned}\lim_{M\to\infty}\sum_{k=1}^{M}|D_k| &\le \lim_{M\to\infty}\sum_{k=1}^{M}\int_0^\infty |g(\tau)|p_k(\tau)d\tau \\ &= \int_0^\infty |g(\tau)|\lim_{M\to\infty}\sum_{k=1}^{M}p_k(\tau)d\tau \\ &= \int_0^\infty |g(\tau)|\sum_{k=1}^{j}p_k(\tau)d\tau + \int_0^\infty |g(\tau)|h_j(\tau)d\tau,\end{aligned}$$

Since $|g(\tau)|$ is bounded by G, the we have

$$\begin{aligned}\int_0^\infty |g(\tau)|\sum_{k=1}^{j}p_k(\tau)d\tau &\le G\int_0^\infty \sum_{k=1}^{j}p_k(\tau)d\tau \\ &= G\sum_{k=1}^{j}\int_0^\infty p_k(\tau)d\tau \\ &= jG < \infty, \end{aligned} \tag{5.15}$$

since each $p_k(\tau)$ integrates to one. On the other hand by Lemma(5.3) the function $h_j(\tau)$ is bounded by a constant K_j, say, and hence

$$\int_0^\infty |g(\tau)|h_j(\tau)d\tau \le K_j\int_0^\infty |g(\tau)|d\tau < \infty, \tag{5.16}$$

since $g(\tau) \in L_1(-\infty,\infty)$. This shows the desired absolute convergence and completes the proof. □

Note that the covariance function $R_X(\tau)$ of the underlying continuous time weakly stationary process satisfies all of the assumptions imposed on $g(\tau)$ in the above proposition: $R_X(\tau)$ is integrable because of the assumed short range dependence, and it is obviously bounded by $R_X(0)$. Hence replacing $g(\tau)$ with $R_X(\tau)$ in the above proposition immediately give us the following corollary, which establishes the existence of the spectral density $S_Y^{\Delta}(f)$ of the discrete time process $Y = Y_k$ obtained from a renewal sampling scheme:

Corollary 5.1. *Suppose the IID renewal sampling interval has a probability density function that satisfies either Assumption(5.2) or Assumption(5.1). Because the underlying continuous time weakly stationary process is assumed to have short range dependence, then the sampled discrete time weakly stationary process* $Y = Y_k$ *obtained*

from this renewal sampling scheme will also have short range dependence in the sense that its autocovariance sequence $C_Y(k)$ is absolutely summable. Consequently the spectral density $S_Y^{\Delta}(f)$ exists and is given by Equation(5.6).

5.4 Spectral Domain Expression for $C_Y(k)$ and $S_Y^{\Delta}(f)$

In the previous sections, the convergence of the autocovariance sequence $C_Y(k)$ was proved by using the time domain expression of Equation(5.5), because such a formulation will allow us to use the existing results from renewal theory. In this section we derive the corresponding spectral domain expressions for the autocovariance sequence $C_Y(k)$ and spectral density $S_Y^{\Delta}(f)$, in terms of the characteristic function $\phi_{\Delta}(\lambda)$ of the probability density $p(\tau)$. These expressions, especially the spectral domain expression for $C_Y(k)$, will turn out to be useful for the purpose of numerically evaluating the spectral density $S_Y^{\Delta}(f)$.

We will however use a slightly different definition of the characteristic function, which corresponds to the particular form of the Fourier transform used through out the thesis. Specifically the characteristic function $\phi_{\Delta}(\lambda)$ of the IID random variables Δ_k with common density $p(\tau)$ is defined as

$$\phi_{\Delta}(\lambda) = \int_0^{\infty} e^{i2\pi\lambda\tau} p(\tau) d\tau, \;\; \lambda \in \mathbb{R}. \tag{5.17}$$

Following Shapiro and Silverman [113], we can use the spectral representation of the integrable covariance function $R_X(\tau)$, which is given by Equation(2.2), to rewrite Equation(5.5) for $C_Y(k)$ as follows:

$$C_Y(k) = \int_0^{\infty} R_X(\tau) p_{|k|}(\tau) d\tau = \int_{-\infty}^{\infty} S_X(\lambda) \int_0^{\infty} e^{i2\pi\lambda\tau} p_{|k|}(\tau) d\tau d\lambda, \tag{5.18}$$

where in the last step the operation of changing the order of integrations can be justified by an application of Fubini's Theorem [143]. It is then immediate to recognize the fact that the inner integral is the characteristic function of a sum of k IID random variables with common density $p(\tau)$, and hence can be written as

$$\int_0^{\infty} e^{i2\pi\lambda\tau} p_{|k|}(\tau) d\tau = \phi_{\Delta}(\lambda)^{|k|}, \;\; \lambda \in \mathbb{R}. \tag{5.19}$$

This therefore gives us the alternative frequency domain expression for the autoco-

variance $C_Y(k)$ as

$$C_Y(k) = \int_{-\infty}^{\infty} S_X(\lambda)\phi_{\Delta}(\lambda)^{|k|}d\lambda, \quad k \in \mathbb{Z}, \tag{5.20}$$

which specifies the relationship between $C_Y(k)$ and the spectral density $S_X(\lambda)$ of the underlying continuous time process.

Using the frequency domain expression for $C_Y(k)$, together with the fact that the spectral density $S_X(\lambda)$ is an even function around the origin, Shapiro and Silverman [113] formally derived a more compact expression for $S_Y^{\Delta}(f)$ as

$$S_Y^{\Delta}(f) = \Delta \int_{-\infty}^{\infty} S_X(\lambda)K_{\Delta}(f,\lambda)d\lambda. \tag{5.21}$$

The function $K_{\Delta}(f,\lambda)$ in the above equation, which will be referred in subsequent discussions as the renewal kernel function, is given by

$$K_{\Delta}(f,\lambda) = \mathrm{Re}\left\{\frac{1+e^{-i2\pi f\Delta}\phi_{\Delta}(\lambda)}{1-e^{-i2\pi f\Delta}\phi_{\Delta}(\lambda)}\right\}. \tag{5.22}$$

It is obvious from Equation(5.22) that the kernel function $K_{\Delta}(f,\lambda)$ is well defined except perhaps when $f = 0$, in which case it may become indeterminant at $\lambda = 0$.

Using a contour integration argument it is shown by Martin [77] that, assuming Δ_k has finite moments of all orders, the function $K_{\Delta}(0,\lambda)$ contains a delta function component $\frac{\delta(\lambda)}{\Delta}$. [2] On the other hand when $f \neq 0$ the denominator of $K_{\Delta}(f,\lambda)$ is not zero, we can then use standard technique to rationalize the denominator in

[2]The argument presented in Martin [77] is heuristic: for λ close but not equal to zero, the characteristic function can be approximated by $\phi_{\Delta}(\lambda) \approx 1 + i2\pi\Delta\lambda + \cdots$; hence the kernel function $K_{\Delta}(0,\lambda)$ can be written approximately as

$$K_{\Delta}(0,\lambda) \approx \mathrm{Re}\left\{\frac{-2}{i2\pi\Delta\lambda} + \cdots\right\}. \tag{5.23}$$

where $\cdots$ in the above equation represents powers in λ. Consider the integral of the kernel function $K_{\Delta}(0,\lambda)$ over an interval $[-\epsilon,\epsilon]$ around the origin. By taking a contour Γ_{ϵ} lying on the upper-half plane, we then have

$$\int_{-\epsilon}^{\epsilon} K_{\Delta}(0,\lambda)d\lambda \approx \mathrm{Re}\left\{\int_{\Gamma_{\epsilon}}\left(\frac{-2}{i2\pi\Delta z} + \cdots\right)dz\right\} \approx \mathrm{Re}\left\{-\pi i \operatorname*{Res}_{z=0}\left(\frac{-2}{i2\pi\Delta z}\right)\right\} = \frac{1}{\Delta}. \tag{5.24}$$

Note that here the contour Γ_{ϵ} has to be taken over the upper-half plane to ensure that $|\phi(z)| < 1$ for $z \in \Gamma_{\epsilon}$.

Equation(5.22) as

$$\begin{aligned}\mathrm{Re}\left\{\frac{1+e^{-i2\pi f\Delta}\phi_\Delta(\lambda)}{1-e^{-i2\pi f\Delta}\phi_\Delta(\lambda)}\right\} &= \mathrm{Re}\left\{\frac{[1+e^{-i2\pi f\Delta}\phi_\Delta(\lambda)][1-\overline{e^{-i2\pi f\Delta}\phi_\Delta(\lambda)}]}{|1-e^{-i2\pi f\Delta}\phi_\Delta(\lambda)|^2}\right\} \\ &= \frac{1-|\phi_\Delta(\lambda)|^2+\mathrm{Re}\left\{e^{-i2\pi f\Delta}\phi_\Delta(\lambda)-\overline{e^{-i2\pi f\Delta}\phi_\Delta(\lambda)}\right\}}{|1-e^{-i2\pi f\Delta}\phi_\Delta(\lambda)|^2} \\ &= \frac{1-|\phi_\Delta(\lambda)|^2}{|1-e^{-i2\pi f\Delta}\phi_\Delta(\lambda)|^2}, \text{ for } f\neq 0, \end{aligned} \tag{5.25}$$

Consequently the renewal kernel function $K_\Delta(f,\lambda)$ can be written in terms of a sum of two components as

$$K_\Delta(f,\lambda) = \frac{\delta(\lambda)I_0(f)}{\Delta} + \tilde{K}_\Delta(f,\lambda), \tag{5.26}$$

where $I_0(\cdot)$ is the indicator function of the set $\{0\}$, and $\tilde{K}_\Delta(f,\lambda)$ denotes the right-hand side of Equation(5.25), i.e. we have

$$\tilde{K}_\Delta(f,\lambda) = \frac{1-|\phi_\Delta(\lambda)|^2}{|1-e^{-i2\pi f\Delta}\phi_\Delta(\lambda)|^2}. \tag{5.27}$$

Moreover, the characteristic function admits the following Taylor series expansion

$$\phi_\Delta(\lambda) = 1 + i2\pi\Delta\lambda - 2\pi^2\lambda^2(\mathrm{Var}[\Delta]+\Delta^2) + o(\lambda^2), \tag{5.28}$$

so that we could easily derive

$$\lim_{\lambda\to 0}\tilde{K}_\Delta(0,\lambda) = \lim_{\lambda\to 0}\frac{1-|\phi_\Delta(\lambda)|^2}{|1-\phi_\Delta(\lambda)|^2} = \lim_{\lambda\to 0}\frac{\lambda^2\mathrm{Var}[\Delta]+o(\lambda^2)}{\lambda^2\Delta^2+o(\lambda^2)} = \frac{\mathrm{Var}[\Delta]}{\Delta^2}. \tag{5.29}$$

Hence the component $\tilde{K}_\Delta(f,\lambda)$ can be taken as continuous in λ for all f. In other words, Equation(5.26) decomposes the renewal kernel function $K_\Delta(f,\lambda)$ as the sum of a singular component and a regular (i.e. continuous) component. Using this decomposition, the frequency domain expression for the spectral density $S_Y^\Delta(f)$ of the discrete time sampled process can then be written as

$$S_Y^\Delta(f) = S_X(0)I_0(f) + \Delta\int_{-\infty}^{\infty} S_X(\lambda)\tilde{K}_\Delta(f,\lambda)d\lambda. \tag{5.30}$$

As already mentioned before, the above alternative expression for $S_Y^{\Delta}(f)$ was first formally derived by Shapiro and Silverman [113], but the delta function component when $f = 0$ was not explicitly taken into account. On the other hand, although the delta function component for $K_{\Delta}(0, \lambda)$ was identified by Martin [77], the meaning of the spectral density $S_Y^{\Delta}(f)$ was incorrectly interpreted there.

One particular important aspect, which was not obvious from the DFT form of $S_Y^{\Delta}(f)$ in Equation(5.6), but becomes almost trivially true from the spectral domain expression in Equation(5.30), is the positivity property of the spectral density $S_Y^{\Delta}(f)$. Such a positivity of $S_Y^{\Delta}(f)$ is of great theoretical importance. In particular it implies that

$$\int_0^{1/\Delta} \log(S_Y^{\Delta}(f))df > -\infty, \tag{5.31}$$

and hence ensuring the asymptotic finiteness of the Whittle's log-likelihood function. It also implies that the sampled process is non-deterministic, in the sense that the future cannot be perfectly predicted from the past [22]. Under rather mild conditions over $S_X(\lambda)$ and $\phi_{\Delta}(\lambda)$, we will show in the following lemma that the spectral density $S_Y^{\Delta}(f)$ is indeed bounded away from zero.

Lemma 5.4. *Suppose that the IID sampling intervals have common probability density that satisfies the assumptions of Corollary(5.1), with finite moments of all orders, and are not identically zero; also suppose the spectral density $S_X(\lambda)$ does not vanish for all λ (i.e. $S_X(\lambda)$ is not identically zero), then there exists a positive constant $c > 0$ such that $S_Y^{\Delta}(f) > c$, for all f.*

Proof: With the assumptions on the common density of the IID random variables Δ_k, then $S_Y^{\Delta}(f)$ exists and can be written by Equation(5.30). To get a lower bound for the spectral density $S_Y^{\Delta}(f)$ we can ignore the delta function component and write

$$S_Y^{\Delta}(f) \geq \Delta \int_{-\infty}^{\infty} S_X(\lambda)\tilde{K}_{\Delta}(f, \lambda)d\lambda. \tag{5.32}$$

Because $|1 - e^{-i2\pi f\Delta}\phi_{\Delta}(\lambda)| \leq (1 + |\phi_{\Delta}(\lambda)|)$, the kernel function $\tilde{K}_{\Delta}(f, \lambda)$ can be bounded from below by

$$\tilde{K}_{\Delta}(f, \lambda) = \frac{1 - |\phi_{\Delta}(\lambda)|^2}{|1 - e^{-i2\pi f\Delta}\phi_{\Delta}(\lambda)|^2} \geq \frac{1 - |\phi_{\Delta}(\lambda)|^2}{(1 + |\phi_{\Delta}(\lambda)|)^2} = \frac{1 - |\phi_{\Delta}(\lambda)|}{1 + |\phi_{\Delta}(\lambda)|}. \tag{5.33}$$

Hence the lower bound for $S_Y^{\Delta}(f)$ can also be written as

$$S_Y^{\Delta}(f) \geq \Delta \int_{-\infty}^{\infty} S_X(\lambda) \frac{1 - |\phi_{\Delta}(\lambda)|}{1 + |\phi_{\Delta}(\lambda)|} d\lambda. \tag{5.34}$$

Because the IID random variables Δ_k are not degenerate, the characteristic function $\phi_{\Delta}(\lambda) = 1$ only when $\lambda = 0$ (see Shiryaev [117]). This observation, together with the assumption that $S_X(\lambda)$ being not identically zero, imply that the integrand in the above lower-bound will be strictly positive almost everywhere, hence the lower-bound itself will also be strictly positive. Setting the right-hand side of Equation(5.7) as the positive constant c completes the proof of this lemma. □

5.5 Asymptotic Consistency of Whittle Estimator

Having established the existence and positivity of $S_Y^{\Delta}(f)$ we know that the Whittle's log-likelihood function $WLL_n^{\Delta}(\boldsymbol{\theta})$, given by Equation(5.2), is well defined. In this section therefore we will be focusing on the asymptotic consistency of the Whittle estimator $\hat{\boldsymbol{\theta}}_n$ that is obtained through minimizing $WLL_n^{\Delta}(\boldsymbol{\theta})$.

Asymptotic consistency of the Whittle estimator has a long history. For a linear regular process of the form of Equation(4.3), Whittle [137] argued heuristically that under suitable regularity conditions, the asymptotic consistency of the estimator should be expected by analogy with the classical asymptotic theory of maximum likelihood estimation for data consisting of IID observations. Whittle's interesting and ingenious heuristic argument was first made rigorous by Walker[134] and then by Hannan [49] under a set of less stringent conditions. See also Brockwell and Davis[22] for similar results with a focus on ARMA family of processes. On the other hand in our situation the discrete time sampled process $Y = Y_k$ generated from a renewal sampling scheme does not possess an explicit linear form, and hence those arguments are not directly applicable. The asymptotic consistency results that are mostly relevant to our situation was the one given by Ibragimov[54], in which it was argued that the Whittle's log-likelihood function $WLL_n^{\Delta}(\boldsymbol{\theta})$ were applicable to arbitrary wide sense stationary processes. In this section we will therefore apply Ibragimov's asymptotic consistency theorem to our sampled processes.

5.5.1 Asymptotic Consistency Theorem

Asymptotic consistency of the Whittle estimator will be established by making use of Theorem(1) in Ibragimov[54]:

Theorem 5.1. *Suppose all values of the parameter vector $\boldsymbol{\theta}$ belong to a compact set $\boldsymbol{\Theta} \in \mathbb{R}^p$, where p is the dimension of the parameter space, with the true parameter value given by $\boldsymbol{\theta}_0$, and the following assumptions are satisfied:*

1. *Continuity condition: the functions $\frac{1}{S_Y^{\Delta}(f;\boldsymbol{\theta})}$ and $\log(S_Y^{\Delta}(f;\boldsymbol{\theta}))$ are continuous on $[0, \frac{1}{\Delta}] \times \boldsymbol{\Theta}$.*

2. *Convergence condition: for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ the Whittle log-likelihood function $WLL_n^{\Delta}(\boldsymbol{\theta})$ converges in probability to a deterministic function, which we will denote by $AWLL(\boldsymbol{\theta})$, as $n \to \infty$;*

3. *Identifiability condition: the limit function $AWLL(\boldsymbol{\theta})$ attains a unique minimum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$;*

then the statistic $\hat{\boldsymbol{\theta}}_n$ is a consistent estimate for the true value $\boldsymbol{\theta}_0$ in the sense that $\hat{\boldsymbol{\theta}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_0$.

The idea of the above theorem is simple: since $AWLL(\boldsymbol{\theta})$ has a unique minimum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, and $WLL_n^{\Delta}(\boldsymbol{\theta})$ approximates $AWLL(\boldsymbol{\theta})$ for large n, we would then expect the value $\hat{\boldsymbol{\theta}}_n$, which minimizes $WLL_n(\boldsymbol{\theta})$, to be close to the true value $\boldsymbol{\theta}_0$ which minimizes $AWLL(\boldsymbol{\theta})$. The continuity condition listed above is simply a technical device to facilitate $\hat{\boldsymbol{\theta}}_n$ approaching $\boldsymbol{\theta}_0$. Theorem(5.1) in its original form was proved for the Whittle log-likelihood function of a linear process, given by Equation(4.4). However, as indicated by the author in the paper, the same proof can be used almost without modification to cover the more general case. In the remainder of this section, we will discuss the three conditions in detail, and find appropriate conditions on $X = X(t)$, and the renewal sampling schemes, so that these asymptotic consistency conditions are satisfied.

5.5.2 Continuity Condition

The continuity condition is simply a technical device that serves two purposes: it allows the convergence of $WLL_n^{\Delta}(\boldsymbol{\theta})$ to $AWLL(\boldsymbol{\theta})$; and it allows the convergence of

the Whittle estimator $\hat{\boldsymbol{\theta}}_n$ to the true parameter value $\boldsymbol{\theta}_0$. This continuity condition was usually assumed as a property of the underlying model, which in existing literature [54, 70] was the discrete time stationary process. In our scenario, it no longer makes any sense to simply impose the required continuity condition on the discrete time process, because it is obtained from the underlying continuous time process through sampling. We will see, however, that with some very general assumptions, the continuity condition can indeed be easily established.

First of all we note that the continuity condition over the function $\frac{1}{S_Y^{\Delta}(f;\boldsymbol{\theta})}$ and $\log(S_Y^{\Delta}(f;\boldsymbol{\theta}))$ can be reduced to the continuity of $S_Y^{\Delta}(f;\boldsymbol{\theta})$, provided that it is bounded away from zero for all $(f,\boldsymbol{\theta}) \in [0, \frac{1}{\Delta}] \times \boldsymbol{\Theta}$. This is established in the following lemma with mild assumptions:

Lemma 5.5. *Suppose the assumptions of Lemma(5.4) hold, and further suppose there exists a non-negative integrable function $S(\lambda)$ such that $S_X(\lambda;\boldsymbol{\theta}) \leq S(\lambda)$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then the spectral density $S_Y^{\Delta}(f;\boldsymbol{\theta})$ is bounded away from zero for all $(f,\boldsymbol{\theta}) \in [0, \frac{1}{\Delta}] \times \boldsymbol{\Theta}$.*

Proof: We can apply Lemma(5.4) to conclude that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ we have

$$S_Y^{\Delta}(f;\boldsymbol{\theta}) \geq \Delta \int_{-\infty}^{\infty} S_X(\lambda;\boldsymbol{\theta}) \frac{1-|\phi_{\Delta}(\lambda)|}{1+|\phi_{\Delta}(\lambda)|} d\lambda. \tag{5.35}$$

Denote the right hand side of the above inequality as $LB_{S_Y}(\boldsymbol{\theta})$. Then since $S_X(\lambda;\boldsymbol{\theta})$ is bounded by $S(\lambda)$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we can apply Dominated Convergence Theorem [143] to see that $LB_{S_Y}(\boldsymbol{\theta})$ is indeed continuous for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Being a continuous function over a compact space, it admits its a minimum value at $\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$, and consequently we have

$$S_Y^{\Delta}(f;\boldsymbol{\theta}) \geq LB_{S_Y}(\boldsymbol{\theta}) \geq LB_{S_Y}(\tilde{\boldsymbol{\theta}}). \tag{5.36}$$

Note that by Lemma(5.4) we must have $LB_{S_Y}(\tilde{\boldsymbol{\theta}}) > 0$, and hence this completes the proof of this result. □

Having established the boundedness property of $S_Y^{\Delta}(f;\boldsymbol{\theta})$, we now proceed to show that it is also continuous for $(f,\boldsymbol{\theta}) \in [0, \frac{1}{\Delta}] \times \boldsymbol{\Theta}$, provided that some mild assumption about $R_X(\tau;\boldsymbol{\theta})$ hold.

Lemma 5.6. *Suppose the assumptions of Corollary(5.1) hold; and further suppose the covariance function $R_X(\tau;\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and is bounded in magnitude by*

a positive integrable function $B_X(\tau)$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then the spectral density $S_Y^{\Delta}(f;\boldsymbol{\theta})$ is continuous for all $(f,\boldsymbol{\theta}) \in [0, \frac{1}{\Delta}] \times \boldsymbol{\Theta}$.

Proof: We will prove the continuity of $S_Y^{\Delta}(f;\boldsymbol{\theta})$ from the first principle. First note that because the assumptions of Corollary(5.1) hold, then $S_Y^{\Delta}(f;\boldsymbol{\theta})$ exists for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Let $(f_1,\boldsymbol{\theta}_1)$ and $(f_2,\boldsymbol{\theta}_2)$ be two points in the space $[0, \frac{1}{\Delta}] \times \boldsymbol{\Theta}$. Then we have

$$\begin{aligned}
&|S_Y^{\Delta}(f_1;\boldsymbol{\theta}_1) - S_Y^{\Delta}(f_2;\boldsymbol{\theta}_2)| \\
&= |S_Y^{\Delta}(f_1;\boldsymbol{\theta}_1) - S_Y^{\Delta}(f_1;\boldsymbol{\theta}_2) + S_Y^{\Delta}(f_1;\boldsymbol{\theta}_2) - S_Y^{\Delta}(f_2;\boldsymbol{\theta}_2)| \\
&\leq |S_Y^{\Delta}(f_1;\boldsymbol{\theta}_1) - S_Y^{\Delta}(f_1;\boldsymbol{\theta}_2)| + |S_Y^{\Delta}(f_1;\boldsymbol{\theta}_2) - S_Y^{\Delta}(f_2;\boldsymbol{\theta}_2)|. \qquad (5.37)
\end{aligned}$$

The spectral density $S_Y^{\Delta}(f_1;\boldsymbol{\theta})$, being defined through Equation(5.6), is obviously continuous in f for every $\boldsymbol{\theta}$, and hence for the second term in the above equation we have

$$\lim_{(f_1,\boldsymbol{\theta}_1)\to(f_2,\boldsymbol{\theta}_2)} |S_Y^{\Delta}(f_1;\boldsymbol{\theta}_2) - S_Y^{\Delta}(f_2;\boldsymbol{\theta}_2)| = 0. \qquad (5.38)$$

On the other hand, for the first term $|S_Y^{\Delta}(f_1;\boldsymbol{\theta}_1) - S_Y^{\Delta}(f_1;\boldsymbol{\theta}_2)|$ we have

$$\begin{aligned}
|S_Y^{\Delta}(f_1;\boldsymbol{\theta}_1) - S_Y^{\Delta}(f_1;\boldsymbol{\theta}_2)| &\leq \Delta \sum_{k=-\infty}^{\infty} |C_Y(k;\boldsymbol{\theta}_1) - C_Y(k;\boldsymbol{\theta}_2)| \\
&\leq \Delta \sum_{k=-\infty}^{\infty} \int_0^{\infty} |R_X(\tau;\boldsymbol{\theta}_1) - R_X(\tau;\boldsymbol{\theta}_2)| p_{|k|}(\tau) d\tau \\
&= \Delta \int_0^{\infty} |R_X(\tau;\boldsymbol{\theta}_1) - R_X(\tau;\boldsymbol{\theta}_2)| \sum_{k=-\infty}^{\infty} p_{|k|}(\tau) d\tau \qquad (5.39)
\end{aligned}$$

On the other hand, since $R_X(\tau;\boldsymbol{\theta})$ is assumed to be bounded by $B_X(\tau)$, which is assumed to be integrable, we can apply Proposition(5.1) to conclude that

$$\Delta \int_0^{\infty} |R_X(\tau;\boldsymbol{\theta}_1) - R_X(\tau;\boldsymbol{\theta}_2)| \sum_{k=-\infty}^{\infty} p_{|k|}(\tau) d\tau \leq 2\Delta \int_0^{\infty} B_X(\tau) \sum_{k=-\infty}^{\infty} p_{|k|}(\tau) d\tau < \infty. \qquad (5.40)$$

Therefore we can apply Bounded Convergence Theorem [143] to derive that

$$\begin{aligned}
&\lim_{(f_1,\boldsymbol{\theta}_1)\to(f_2,\boldsymbol{\theta}_2)} |S_Y^{\Delta}(f_1;\boldsymbol{\theta}_1) - S_Y^{\Delta}(f_1;\boldsymbol{\theta}_2)| \\
&\leq \lim_{(f_1,\boldsymbol{\theta}_1)\to(f_2,\boldsymbol{\theta}_2)} \Delta \int_0^{\infty} |R_X(\tau;\boldsymbol{\theta}_1) - R_X(\tau;\boldsymbol{\theta}_2)| \sum_{k=-\infty}^{\infty} p_{|k|}(\tau) d\tau \\
&= \Delta \int_0^{\infty} \lim_{(f_1,\boldsymbol{\theta}_1)\to(f_2,\boldsymbol{\theta}_2)} |R_X(\tau;\boldsymbol{\theta}_1) - R_X(\tau;\boldsymbol{\theta}_2)| \sum_{k=-\infty}^{\infty} p_{|k|}(\tau) d\tau \\
&= 0, \qquad (5.41)
\end{aligned}$$

because of the assumed continuity of $R_X(\tau;\boldsymbol{\theta})$. This therefore completes the proof of the continuity of $S_Y^{\Delta}(f;\boldsymbol{\theta})$. □

Consequently, the continuity conditions of Theorem(5.1) on $\frac{1}{S_Y^{\Delta}(f;\boldsymbol{\theta})}$ and $\log(S_Y^{\Delta}(f;\boldsymbol{\theta}))$ can then be trivially established by applying Lemma(5.5) and (5.6). The following lemma therefore summarizes the continuity results that we established in this section:

Proposition 5.2. *Suppose the assumptions of both Lemma(5.5) and (5.6) are satisfied. Then the continuity condition of Theorem(5.1) will be satisfied, in other words, the functions* $\frac{1}{S_Y^{\Delta}(f;\boldsymbol{\theta})}$ *and* $\log(S_Y^{\Delta}(f;\boldsymbol{\theta}))$ *will be continuous for all* $(f,\boldsymbol{\theta}) \in [0, \frac{1}{\Delta}] \times \boldsymbol{\Theta}$.

5.5.3 Convergence Condition

In this section we will be discussing the convergence condition in Theorem(5.1). Existing literature [54, 49] suggests that this convergence of the Whittle log-likelihood function $WLL_n^{\Delta}(\boldsymbol{\theta})$ depends on the convergence of the sample autocovariance sequence of the discrete time process $Y = Y_k$, which is defined for a zero-mean process as

$$\hat{c}_Y(k) = \begin{cases} \frac{1}{n}\sum_{j=1}^{n-k} Y_{j+k}Y_j & \text{for} \quad 0 \leq k \leq n-k \\ 0 & \text{for} \quad k > n-k \end{cases} . \qquad (5.42)$$

The following technical assumptions are needed to establish the desired convergence of the sample autocovariance sequence:

Assumption 5.3.

1. $|R_X(\tau)| \leq H_X(\tau)$, *where the function* $H_X(\tau)$ *is an even, non-negative, integrable function which is non-increasing over* $[0,\infty)$;

2. *$X = X(t)$ is fourth-order stationary with the fourth-order cumulant function $Q_X(\tau_1, \tau_2, \tau_3)$ satisfying $|Q_X(\tau_1, \tau_2, \tau_3)| \leq H_Q(\tau_1, \tau_2, \tau_3)$, where H_Q is even and nonincreasing on $[0, \infty)$ in each variable such that $\int_0^\infty H_Q(0, \tau, 0)d\tau \leq \infty$.*

Here the fourth-order cumulant function of a weakly stationary process is defined as

$$\begin{aligned} Q_X(\tau_1, \tau_2, \tau_3) = E[X(t)X(t+\tau_1)X(t+\tau_2)X(t+\tau_3)] - R_X(\tau_1)R_X(\tau_3 - \tau_1) \\ - R_X(\tau_2)R_X(\tau_1 - \tau_3) - R_X(\tau_3)R_X(\tau_2 - \tau_1). \end{aligned} \tag{5.43}$$

Note that if the underlying process $X = X(t)$ is Gaussian, then the fourth-order cumulant function $Q_X(\tau_1, \tau_2, \tau_3)$ simply vanishes [80]. Moreover also note that here the bounding function $H_X(\tau)$ and $H_Q(\tau_1, \tau_2, \tau_3)$ also depend on $\boldsymbol{\theta}$, but its appearance has been suppressed for notational convenience. This practice will be carried out throughout this section when discussing the convergence condition, unless otherwise stated. We now would like to show that for any fixed k, the sample autocovariance $\hat{c}_Y(k)$ converges to the true-value $C_Y(k)$ in mean-square, and consequently also in probability [117]. Mean-square consistency is guaranteed if we show separately that both the bias and variance tend to zero as the sample size n tends to infinity. The computation for bias is straightforward and will be stated in the following lemma:

Lemma 5.7. *The sample autocovariance sequence $\hat{c}_Y(k)$ is asymptotically unbiased for any fixed k, as the sample size n tends to infinity.*

Proof: For each fixed $k < n$, we have

$$\text{bias}[\hat{c}_Y(k)] = E[\hat{c}_Y(k)] - C_Y(k) = -\frac{k}{n}C_Y(k) \to 0 \text{ as } n \to \infty, \tag{5.44}$$

and this completes the proof. □

We now consider the variance of the sample autocovariance $\hat{c}_Y(k)$, which is given by

$$\text{Var}[\hat{c}_Y(k)] = E[\hat{c}_Y^2(k)] - E^2[\hat{c}_Y(k)]. \tag{5.45}$$

By taking expectation with respect to $X = X(t)$ and using Equation(5.43) for the fourth-order moments, we could write

$$E[\hat{c}_Y^2(k)] = \sum_{m=1}^{4} A_{n,m}(k), \tag{5.46}$$

where the terms $A_{n,m}(k), m = 1, 2, 3, 4$ are defined as

$$A_{n,1}(k) = \frac{1}{n^2} \sum_{j,l=1}^{n-k} E[R_X(t_{j+k} - t_j) R_X(t_{l+k} - t_l)], \tag{5.47a}$$

$$A_{n,2}(k) = \frac{1}{n^2} \sum_{j,l=1}^{n-k} E[R_X(t_l - t_j) R_X(t_{l+k} - t_{j+k})], \tag{5.47b}$$

$$A_{n,3}(k) = \frac{1}{n^2} \sum_{j,l=1}^{n-k} E[R_X(t_{l+k} - t_j) R_X(t_{k+k} - t_l)], \tag{5.47c}$$

$$A_{n,4}(k) = \frac{1}{n^2} \sum_{j,l=1}^{n-k} E[Q_X(t_{j+k} - t_j, t_l - t_j, t_{l+k} - t_j)], \tag{5.47d}$$

and hence the expression for the variance becomes

$$\mathrm{Var}[\hat{c}_Y(k)] = \sum_{m=1}^{4} A_{n,m}(k) - E^2[\hat{c}_Y(k)]. \tag{5.48}$$

In order to investigate the asymptotic behaviour of the variance $\mathrm{Var}[\hat{c}_Y(k)]$, we have to deal with the expectations with respect to the renewal sampling process $\{t_k\}$. We do this by decomposing the (square) region $\Pi = \{(j, l) : j, l = 1, 2, \cdots, n - k\}$ into five disjoint subregions according to the three (sub)diagonals: (1) the main diagonal $j = l$; (2) the upper sub-diagonal $j = l + k$; and (3) the lower sub-diagonal $j = l - k$. The part of the region Π below and including the lower sub-diagonal constitutes the first subregion Π_1; the low-middle region between the main-diagonal and the lower sub-diagonal being the second subregion Π_2; the main-diagonal is the third subregion Π_3, etc. Specifically these five subregions are defined by

$$\Pi_1 = \{(j, l) : 0 < j < j + k \leq l < l + k\}, \tag{5.49a}$$

$$\Pi_2 = \{(j, l) : 0 < j < l < j + k < l + k\}, \tag{5.49b}$$

$$\Pi_3 = \{(j, l) : 0 < j = l < j + k = l + k \leq n - k\}, \tag{5.49c}$$

$$\Pi_4 = \{(j, l) : 0 < l < l + k \leq j < j + k\}, \tag{5.49d}$$

$$\Pi_5 = \{(j, l) : 0 < l < j < l + k < j + k\}. \tag{5.49e}$$

Then the double summation operator $\sum_{j,l=1}^{n-k} := \sum_{\Pi}$ appearing in Equations(5.47) can be written as $\sum_{\Pi} = \sum_{\Pi_1} + \cdots + \sum_{\Pi_5}$. Consequently we can write for $m = 1, 2, 3$

that

$$A_{n,m}(k) = \sum_{r=1}^{5} A_{n,m}^{r}(k), \tag{5.50}$$

where $A_{n,j}^{r}(k)$ represents the summation over the restricted subregion Π_r of the terms appearing in $A_{n,j}(k)$. In other words we have

$$A_{n,1}^{r}(k) = \frac{1}{n^2} \sum_{j,l \in \Pi_r} \mathrm{E}[R_X(t_{j+k} - t_j) R_X(t_{l+k} - t_l)], \tag{5.51a}$$

$$A_{n,2}^{r}(k) = \frac{1}{n^2} \sum_{j,l \in \Pi_r} \mathrm{E}[R_X(t_{l} - t_j) R_X(t_{l+k} - t_{j+k})], \tag{5.51b}$$

$$A_{n,3}^{r}(k) = \frac{1}{n^2} \sum_{j,l \in \Pi_r} \mathrm{E}[R_X(t_{l+k} - t_j) R_X(t_{j+k} - t_l)], \tag{5.51c}$$

Here the term $A_{n,4}$ is not decomposed according to $\{\Pi_r : r = 1, \ldots, 5\}$, because its asymptotic behaviour can be discussed directly (see Appendix of this chapter for detail). Next we note that the symmetry of the indices j and l in the above equations implies that for $m = 1, 2, 3$ the following relationship hold:

$$A_{n,m}^{1}(k) = A_{n,m}^{4}(k), \tag{5.52a}$$

$$A_{n,m}^{2}(k) = A_{n,m}^{5}(k), \tag{5.52b}$$

and consequently the decomposition of Equation(5.50) can be simplified. For notational clarity we further define

$$U_{n,1}(k) = 2A_{n,1}^{1}(k) - \mathrm{E}^2[\hat{c}_n(k)], \tag{5.53a}$$

$$V_{n,1}(k) = A_{n,1}(k) - 2A_{n,1}^{1}(k) = 2A_{n,1}^{2}(k) + A_{n,1}^{3}(k). \tag{5.53b}$$

Then the variance of the sample autocovariance admits the following decomposition:

$$\mathrm{Var}[\hat{c}_n(k)] = U_{n,1}(k) + V_{n,1}(k) + A_{n,2}(k) + A_{n,3}(k) + A_{n,4}(k). \tag{5.54}$$

The rational for such a complicated decomposition scheme is to make it easier to investigate the asymptotic behaviour of the terms in the expression for variance in each subregion. The plan is to show that the contribution of $\mathrm{E}^2[\hat{c}_Y(k)]$ will be canceled by $2A_{n,1}^{1}(k)$, so that $U_{n,1}(k)$ and all the other terms can be bounded by quantities of order $O(\frac{1}{n})$. The same idea was used in Lui[74], Masry and Lui[84] when dealing with

Poisson sampling scheme, and we generalize it to more general random sampling schemes, by making use of Proposition(5.1) proved in Section(5.3.2). The actual upper bound for each term in Equation(5.54) are presented in the following lemma, with the detailed computation being left in Section(8.2) in the appendix of this thesis:

Lemma 5.8. *Suppose k is any fixed positive integer and without loss of generality assuming that $n \geq 2k$. Then under Assumption(5.3) we have the following expressions:*

$$U_{n,1}(k) = \frac{n - 2k(1+n) + 3k^2}{n^2} C_Y^2(k); \tag{5.55}$$

$$|A_{n,1}^2(k)| \leq \frac{H_X(0)}{n^2} \sum_{j,l=1}^{n-k} E[H_X(t_j - t_l)] \tag{5.56}$$

$$|A_{n,m}^r(k)| \leq \frac{H_X(0)}{n^2} \sum_{j,l=1}^{n-k} E[H_X(t_j - t_l)], \ \ for \ m = 2,3, and \ r = 1,2 \tag{5.57}$$

$$|A_{n,m}^3(k)| \leq \frac{n-k}{n^2} R_X^2(0), \ \ for \ m = 1,2,3, \tag{5.58}$$

$$|A_{n,4}(k)| \leq \frac{1}{n^2} \sum_{j,l=1}^{n-k} E[H_Q(0, t_j - t_l, 0)]. \tag{5.59}$$

The above lemma, together with the expression of the variance of the sampled autocovariance $\hat{c}_n(k)$ in Equation(5.54), immediately gives the following upper bound for the variance of the sample autocovariance sequence as

$$\begin{aligned} \mathrm{Var}[\hat{c}_n(k)] \leq & \frac{n - 2k(1+n) + 3k^2}{n^2} C_Y^2(k) + 3\frac{n-k}{n^2} R_X^2(0) \\ & + 6\frac{H_X(0)}{n^2} \sum_{j,l=1}^{n-k} \mathrm{E}[H_X(t_j - _l)] + \frac{1}{n^2} \sum_{j,l=1}^{n-k} \mathrm{E}[H_Q(0, t_j - t_l, 0)]. \end{aligned} \tag{5.60}$$

It is obvious that for any fixed k, the first two bounding terms in the above expression are of order $O(\frac{1}{n})$. The following lemma, which make crucial use of the convergence result of Proposition(5.1) shows that the remaining terms are also of order $O(\frac{1}{n})$:

Lemma 5.9. *Suppose the assumptions of Proposition(5.1) hold, and further suppose*

$g \in L_1(0,\infty)$ is also bounded, then we have

$$\frac{1}{n^2}\sum_{j,l=1}^{n-k} E[g(t_j - t_l)] = O\left(\frac{1}{n}\right). \tag{5.61}$$

Proof: To show this asymptotic result, we simply rewrite the double summation as

$$\begin{aligned}\frac{1}{n^2}\sum_{j,l=1}^{n-k} \mathrm{E}[g(t_j - t_l)] &= \frac{1}{n^2}\sum_{s=0}^{n-k-1} 2(n-k-s)\mathrm{E}[g(t_s)] \\ &= \frac{2(n-k)}{n^2}\sum_{s=0}^{n-k-1}\mathrm{E}[g(t_s)] - \frac{2}{n^2}\sum_{s=0}^{n-k-1} s\mathrm{E}[g(t_s)]. \end{aligned} \tag{5.62}$$

By Proposition(5.1), the sum $\sum_{s=0}^{n-k-1}\mathrm{E}[g(t_s)]$ converges and hence the first term is of order $O(\frac{1}{n})$. On the other hand by Kronecker lemma [117], the sum $\frac{1}{n}\sum_{s=0}^{n-k-1} s\mathrm{E}[g(t_s)] = o(1)$ and hence the second term is of order $o(\frac{1}{n})$. Combining these two asymptotic results together gives the desired result. □

Lemma 5.10. *Suppose the assumptions of Lemma(5.8) and Lemma(5.9) hold. Then for any fixed integer k we have*

$$Var[\hat{c}_n(k)] = O\left(\frac{1}{n}\right). \tag{5.63}$$

Proof: Since the function $H_X(\tau)$ and $H_Q(0,\tau,0)$ are all bounded and integrable by assumption, we can apply Lemma(5.9) to show that all the terms in Equation(5.60) are of order $O\left(\frac{1}{n}\right)$, and hence this completes the proof. □

Consequently by combining Lemma(5.7) and (5.10) it is immediate to see that for each fixed k the sample autocovariance $\hat{c}_n(k)$ is a consistent estimator of $C_Y(k)$ in the sense that it converges in mean-square to $C_Y(k)$. In particular this implies that $\hat{c}_n(k) \to C_Y(k)$ in probability [104], which we will soon see lead to the convergence of the Whittle log-likelihood function required by Theorem(5.1). The following proposition summarizes this result:

Proposition 5.3. *Suppose the assumptions of Lemma(5.8) and Lemma(5.9) hold. Then the for any fixed k the sample autocovariance sequence $\hat{c}_n(k)$ converges in mean-square, and hence in probability, to the true value $C_Y(k)$.*

Having established the desired convergence of the sample autocovariance sequence, we can now turn to proving the convergence condition in Theorem(5.1) for our Whittle log-likelihood function $WLL_n^{\Delta}(\boldsymbol{\theta})$, which is given by Equation(5.2). The first term of $WLL_n^{\Delta}(\boldsymbol{\theta})$ is simply a Riemann approximation. If the integrand function $\log(S_Y^{\Delta}(f;\boldsymbol{\theta}))$ is continuous, this Riemann approximation converge almost surely. This observation therefore gives the following simple result:

Lemma 5.11. *Suppose the assumptions of Proposition(5.2) hold. Then we have*

$$\frac{1}{n}\sum_{k=0}^{n-1}\log\left(S_Y^{\Delta}\left(\frac{k}{n\Delta};\boldsymbol{\theta}\right)\right)\overset{a.s.}{\to}\Delta\int_0^1\log(S_Y^{\Delta}(f;\boldsymbol{\theta}))df. \tag{5.64}$$

Proof: Since the assumptions of Proposition(5.2) hold, then for any $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ the function $\log(S_Y^{\Delta}(f;\boldsymbol{\theta}))$ is continuous for all $f\in[0,\frac{1}{\Delta}]$. Therefore the desired result is simply a consequence of the convergence of the Riemann approximation for continuous functions [42]. This then completes the proof. □

On the other hand, the convergence of the second term in $WLL_n^{\Delta}(\boldsymbol{\theta})$ can be established by using the convergence of the sample autocovariance sequence that we proved in Proposition(5.3). It is in fact an easy consequence of the following lemma:

Lemma 5.12. *Suppose the assumptions of Proposition(5.2) and Proposition(5.3) hold. Then then we have*

$$\frac{1}{n}\sum_{k=0}^{n-1}\frac{I_n^{\Delta}\left(\frac{k}{n\Delta}\right)}{S_Y^{\Delta}\left(\frac{k}{n\Delta};\boldsymbol{\theta}\right)}\overset{\mathbb{P}}{\to}\Delta\int_0^{1/\Delta}\frac{S_Y^{\Delta}(f;\boldsymbol{\theta}_0)}{S_Y^{\Delta}(f;\boldsymbol{\theta})}df,\ \ \boldsymbol{\theta}\in\boldsymbol{\Theta}. \tag{5.65}$$

Proof: The same result was established by Hannan [49] for linear processes of the form given by Equation(4.3). However the technique used in the proof can be applied in our case without modification. Consequently we will omit the details and only give an outline here, focusing on the technique being used. Note that since the assumptions of Proposition(5.2) are satisfied, the function $\frac{1}{S_Y^{\Delta}(f;\boldsymbol{\theta})}$ will be continuous in f for any $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, and consequently can be uniformly approximated by a Cesaro

sum [121] of the form $Q_m(f;\boldsymbol{\theta}) = \sum_{k=-m}^{m} q(k;\boldsymbol{\theta})e^{i2\pi k f\Delta}$, i.e. we have

$$\frac{1}{n}\sum_{k=0}^{n-1} I_n^{\Delta}\left(\frac{k}{n\Delta}\right) Q_m\left(\frac{k}{n\Delta};\boldsymbol{\theta}\right) \overset{a.s.}{\to} \frac{1}{n}\sum_{k=0}^{n-1} \frac{I_n^{\Delta}\left(\frac{k}{n\Delta}\right)}{S_Y^{\Delta}\left(\frac{k}{n\Delta};\boldsymbol{\theta}\right)}, \text{ uniformly in } m. \tag{5.66}$$

Moreover, because the assumptions of Proposition(5.3) hold, the sample autocovariance sequence $\hat{c}_n(k)$ converges to $C_Y(k)$ in probability. Consequently we have

$$\begin{aligned}
\frac{1}{n}\sum_{k=0}^{n-1} I_n^{\Delta}\left(\frac{k}{n\Delta}\right) Q_m\left(\frac{k}{n\Delta};\boldsymbol{\theta}\right) &\overset{\mathbb{P}}{\to} \sum_{k=-m}^{m} q(k;\boldsymbol{\theta})(1-\frac{|k|}{m})C_Y(k;\boldsymbol{\theta}_0), \text{ as } n\to\infty \\
&= \Delta\int_0^{1/\Delta} \sum_{k=-m}^{m} q(k;\boldsymbol{\theta})(1-\frac{|k|}{m})S_Y(f;\boldsymbol{\theta}_0)e^{i2\pi f\Delta}df \\
&= \Delta\int_0^{1/\Delta} Q_m(f;\boldsymbol{\theta})S(f;\boldsymbol{\theta}_0)df \\
&\overset{a.s}{\to} \Delta\int_0^{1/\Delta} \frac{S(f;\boldsymbol{\theta}_0)}{S(f;\boldsymbol{\theta})}df, \text{ as } m\to\infty,
\end{aligned} \tag{5.67}$$

where the last a.s. convergence is a result of Cesaro sum $Q_m(f;\boldsymbol{\theta})$ approximating $\frac{1}{S_Y^{\Delta}(f;\boldsymbol{\theta})}$ uniformly for all $f \in [0, \frac{1}{\Delta}]$. Combining Equation(5.66) and (5.67) will then give us the desired result. □

Applying the results of Lemma(5.11) and (5.12) together will then allow us to establish the desired convergence condition for Theorem(5.1). This is summarized in the following proposition:

Proposition 5.4. *Suppose the assumptions of Proposition(5.2) and Proposition(5.3) hold. Then the convergence condition for Theorem(5.1) holds with*

$$WLL_n^{\Delta}(\boldsymbol{\theta}) \overset{\mathbb{P}}{\to} AWWL(\boldsymbol{\theta}), \ \ \boldsymbol{\theta}\in\boldsymbol{\Theta}. \tag{5.68}$$

where the limiting deterministic function $AWWL(\boldsymbol{\theta})$ is defined as

$$AWLL(\boldsymbol{\theta}) = \Delta\int_0^{1/\Delta} \log(S_Y^{\Delta}(f;\boldsymbol{\theta}))df + \Delta\int_0^{1/\Delta} \frac{S_Y^{\Delta}(f;\boldsymbol{\theta}_0)}{S_Y^{\Delta}(f;\boldsymbol{\theta})}df. \tag{5.69}$$

5.5.4 Identifiability Condition

Having established convergence of the Whittle log-likelihood function, we now turn to the identifiability implied by the asymptotic deterministic function $AWLL(\boldsymbol{\theta})$. It is in fact easy to show that the true parameter value $\boldsymbol{\theta}_0$ gives the minimum of the function $AWLL(\boldsymbol{\theta})$. Using the inequality $\log(x) \leq x-1$, we can write, for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, that

$$\int_0^{1/\Delta} \log\left[\frac{S_Y^{\Delta}(f;\boldsymbol{\theta}_0)}{S_Y^{\Delta}(f;\boldsymbol{\theta})}\right] df \leq \int_0^{1/\Delta} \left\{\frac{S_Y^{\Delta}(f;\boldsymbol{\theta}_0)}{S_Y^{\Delta}(f;\boldsymbol{\theta})} - 1\right\} df. \tag{5.70}$$

Rearranging the above inequality using the fact that $1 = \frac{S_Y^{\Delta}(f;\boldsymbol{\theta})}{S_Y^{\Delta}(f;\boldsymbol{\theta})}$, we can easily obtain the following inequality:

$$AWLL(\boldsymbol{\theta}_0) \leq AWLL(\boldsymbol{\theta}), \text{ for } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \tag{5.71}$$

and consequently showing that $\boldsymbol{\theta}_0$ gives the minimum of the limit function $AWLL(\boldsymbol{\theta})$. Moreover because $\log(x) = x - 1$ only when $x = 1$, the above inequality becomes equality only when

$$\frac{S_Y^{\Delta}(f;\boldsymbol{\theta}_0)}{S_Y^{\Delta}(f;\boldsymbol{\theta})} = 1 \text{ for almost all } f. \tag{5.72}$$

These discussions therefore give the identifiability condition for Theorem(5.1), which we summarize as the following proposition

Proposition 5.5. *If for any $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ the spectral density $S_Y^{\Delta}(f;\boldsymbol{\theta}_1)$ differs from $S_Y^{\Delta}(f;\boldsymbol{\theta}_2)$ on a set of positive measure, then the limit function $AWLL(\boldsymbol{\theta})$ attains unique minimum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.*

The identifiability assumption is also related to the concept of alias-free sampling that has be discussed in Section(4.4). If the sampling scheme is one of the alias-free sampling scheme, then the one-to-one functional relationship between $S_X(\lambda;\boldsymbol{\theta})$ and $S_Y^{\Delta}(f;\boldsymbol{\theta})$ is guaranteed [78, 8]. As a result the identifiability of the spectral density $S_Y^{\Delta}(f;\boldsymbol{\theta})$ can be established under the assumption that for any $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ the spectral density $S_X(\lambda;\boldsymbol{\theta}_1)$ differs from $S_X(\lambda;\boldsymbol{\theta}_2)$ on a set of positive measure. On the other hand when the sampling is carried out through more general renewal sampling schemes, the relationship between $S_Y^{\Delta}(f;\boldsymbol{\theta})$ and $S_X(\lambda;\boldsymbol{\theta})$ becomes more complicated and no general statement can be made.

5.5.5 Further Discussions

Hence we have completed the discussions of the three conditions required to establish the asymptotic consistency of the Whittle estimator $\hat{\boldsymbol{\theta}}_n$. Proposition(5.2), (5.4) and (5.5) showed, under appropriate technical assumptions, that the continuity, convergence and identifiability condition of Theorem(5.1) will be satisfied. Consequently the Whittle estimator $\hat{\boldsymbol{\theta}}_n$ is indeed a consistent estimator of the true value $\boldsymbol{\theta}_0$.

Note that the consistency result in Theorem(5.1) is in terms of convergence in probability. Such a mode of convergence is practically sufficient to ensure that the Whittle estimator $\hat{\boldsymbol{\theta}}_n$ being a reasonable approximation to the true value $\boldsymbol{\theta}_0$. It is, however, interesting to mention that a stronger mode of convergence can be established under a set of slightly different assumptions, which regularizes the continuous time process $X = X(t)$ by strictly stationarity and mixing condition [16]. Specifically we have the following results, which is proved by Charlog and Rachdi [23]:

Proposition 5.6. *Suppose the continuous time process is strictly stationary and α-mixing, and the renewal sampling IID interval Δ_k satisfies $\mathbb{P}(\Delta_k > 0) = 1$, then the sampled process $Y = Y_k = X(t_k)$ is also strictly stationary and α-mixing.*

The original results in Charlog and Rachdi [23] is more general, considered more mixing types and explicitly gives the corresponding mixing coefficient for the discrete time sampled process $Y = Y_k = X(t_k)$. The strict stationarity, together with α-mixing, implies ergodicity for the discrete time sampled process [29], and consequently for each fixed k the sample autocovariance sequence $\hat{c}_n(k)$ converges to $C_Y(k)$ almost surely [29]. This almost sure convergence result of $\hat{c}_n(k)$, as suggested by Hannan[49] and Ibragimov[54], will then lead to almost sure convergence of $WLL_n^{\Delta}(\boldsymbol{\theta})$ to its limit $AWWL(\boldsymbol{\theta})$. As a consequence when other conditions of Theorem(5.1) are satisfied, this stronger convergence of $WLL_n^{\Delta}(\boldsymbol{\theta})$ will in turn imply that $\hat{\boldsymbol{\theta}}_n$ converges to $\boldsymbol{\theta}_0$ almost surely.

5.6 Asymptotic Normality of Whittle Estimator

The asymptotic normality of the Whittle estimator $\hat{\boldsymbol{\theta}}_n$ seems to be a very difficult problem, because it depends on higher order probabilitistic structure of the discrete time sampled process $Y = Y_k = X(t_k)$. These properties, however, are not yet fully known. We will however still provide a quick and heuristic discussion, starting

from revewing the relevant asymptotic normality theory for Whittle estimator. This review will largely based on the material from Taniguchi and Kakizawa [127], and Rosenblatt [108].

Suppose $WLL_n^{\Delta}(\hat{\boldsymbol{\theta}}_n)$ is sufficiently smooth, it can then be expanded through a Taylor series around $\boldsymbol{\theta}_0$ to obtain

$$0 = \frac{\partial}{\partial \boldsymbol{\theta}} WLL_n^{\Delta}(\hat{\boldsymbol{\theta}}_n) = \frac{\partial}{\partial \boldsymbol{\theta}} WLL_n^{\Delta}(\boldsymbol{\theta}_0) + \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} WLL_n^{\Delta}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \cdots . \quad (5.73)$$

Assuming the higher order terms are negligible, it can then be expected from the above equation that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \approx -[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} WLL_n^{\Delta}(\boldsymbol{\theta}_0)]^{-1} \sqrt{n} \frac{\partial}{\partial \boldsymbol{\theta}} WLL_n^{\Delta}(\boldsymbol{\theta}_0). \quad (5.74)$$

Applying the same technique as in the proof of Lemma(5.12), it is straightforward to show that the second order derivative term $[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} WLL_n^{\Delta}(\boldsymbol{\theta}_0)]$ converges, in probability, to a deterministic matrix $\Sigma_Y(\boldsymbol{\theta}_0)$:

$$\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} WLL_n^{\Delta}(\boldsymbol{\theta}_0) \xrightarrow{\mathbb{P}} \Sigma_Y(\boldsymbol{\theta}_0), \quad (5.75)$$

where $\Sigma_Y(\boldsymbol{\theta}_0)$ is defined as

$$\Sigma_Y(\boldsymbol{\theta}_0) = \Delta \int_0^{1/\Delta} \frac{\frac{\partial}{\partial \boldsymbol{\theta}} S_Y^{\Delta}(f; \boldsymbol{\theta}_0) \frac{\partial}{\partial \boldsymbol{\theta}^T} S_Y^{\Delta}(f; \boldsymbol{\theta}_0)}{S_Y^{\Delta}(f; \boldsymbol{\theta}_0)^2} df. \quad (5.76)$$

This deterministic matrix $\Sigma_Y(\boldsymbol{\theta}_0)$ is known in the literature as the Fisher information matrix [127]. Being a second derivative of the Whittle log-likelihood, it can be regarded as a measure of the curvature [62]. A high curvature generally means that the likelihood function will have a well-defined sharp minimum at the true parameter value $\boldsymbol{\theta}_0$, thus making the estimation more accurate for a given sample size.

On the other hand, it is a standard result in the literature [108, 127] that, under appropriate regularity conditions over the dependence structure of the discrete time sampled process, we should have the following asymptotic normality for the

renormalized first derivative term

$$\begin{aligned}\sqrt{n}\frac{\partial}{\partial\boldsymbol{\theta}}WLL_n^{\Delta}(\boldsymbol{\theta}_0) &= \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\frac{S_Y^{\Delta}(\frac{k}{n\Delta};\boldsymbol{\theta}_0)-I_n^{\Delta}(\frac{k}{n\Delta})}{S_Y^{\Delta}(\frac{k}{n\Delta};\boldsymbol{\theta}_0)^2}\frac{\partial}{\partial\boldsymbol{\theta}}S_Y^{\Delta}\left(\frac{k}{n\Delta};\boldsymbol{\theta}_0\right)\\ &\approx \sqrt{n}\Delta\int_0^{1/\Delta}\frac{S_Y^{\Delta}(f;\boldsymbol{\theta}_0)-I_n^{\Delta}(f)}{S_Y^{\Delta}(f;\boldsymbol{\theta}_0)^2}\frac{\partial}{\partial\boldsymbol{\theta}}S_Y^{\Delta}(f;\boldsymbol{\theta}_0)df\\ &\Rightarrow \mathcal{N}(\mathbf{0},2\Sigma_Y(\boldsymbol{\theta}_0)+\Omega_Y(\boldsymbol{\theta}_0)).\end{aligned} \tag{5.77}$$

Here the covariance matrix component $\Sigma_Y(\boldsymbol{\theta}_0)$ is given by Equation(5.76), and the other component $\Omega_Y(\boldsymbol{\theta}_0)$ is defined as [108, 127]

$$\Omega_Y(\boldsymbol{\theta}_0)=\Delta^2\int_0^{1/\Delta}\int_0^{1/\Delta}\frac{\frac{\partial}{\partial\boldsymbol{\theta}}S_Y^{\Delta}(f;\boldsymbol{\theta}_0)\frac{\partial}{\partial\boldsymbol{\theta}^T}S_Y^{\Delta}(\mu;\boldsymbol{\theta}_0)}{S_Y^{\Delta}(f;\boldsymbol{\theta}_0)^2S_Y^{\Delta}(\mu;\boldsymbol{\theta}_0)^2}Q_Y(f,-\mu,\mu)df\,d\mu, \tag{5.78}$$

where the function $Q_Y(f,\mu,\nu)$ is the fourth-order cumulant spectral density of the discrete time sampled process $Y=Y_k$. Consequently it is obvious from Equation(5.74) that if the asymptotic result in Equation(5.77) holds, then the Whittle estimator $\hat{\boldsymbol{\theta}}_n$ will also be asymptotically Gaussian, i.e. we will have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\Rightarrow\mathcal{N}(\mathbf{0},2\Sigma_Y(\boldsymbol{\theta}_0)^{-1}+\Sigma_Y(\boldsymbol{\theta}_0)^{-1}\Omega_Y(\boldsymbol{\theta}_0)\Sigma_Y(\boldsymbol{\theta}_0)^{-1}). \tag{5.79}$$

Therefore the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ depends crucially on the validity of the asymptotic normality given in Equation(5.77).

Historically there are different approaches showing this asymptotic normality, imposing different technical conditions that amount to asymptotic independence of distant observations. Brillinger[18] proved the desired asymptotic normality for general discrete time strict stationary processes, by assuming a strong cumulant condition of all orders. A second approach is due to Hosoya and Taniguchi[51], in which the authors proved Equation(5.77) for linear processes of the form given by Equation(4.3), with complicated cumulant conditions being imposed on the innovation noises (see also [127]). Yet another approach was used by Rosenblatt[108], and the asymptotic normality of Equation(5.77) was proved by assuming absolute summability of the fourth-order cumulants, and then replacing the higher order cumulant conditions by a set of martingale difference conditions. All these conditions are technical and difficult to verify in our case: although by Wald decomposition [22] our discrete time sampled process $Y=Y_k$ can be written as a linear process, the higher-order probabil-

ity structure of the innovation noises are not known. Because either the higher order cumulant conditions, or the Rosenblatt's martingale difference condition, depend on the higher order probabilistic structure of the sampled process $Y = Y_k$, they are in general are very difficult to verify.

Although a rigorous theoretical investigation of the asymptotic normality was not considered in this thesis, we will still present the result of a simulation study, showing that under practical scenarios in which the underlying process is Gaussian, then the term $\sqrt{n}\frac{\partial}{\partial \boldsymbol{\theta}} WLL_n^{\Delta}(\boldsymbol{\theta}_0)$ will have a distribution that resembles Gaussianity. We have chosen an Oscillatory Matérn Gaussian process (see Section(2.3.3)) with parameters $\sigma = 1, \phi = 1, \nu = 1, \lambda_0 = 1$ as the underlying process $X = X(t)$, sampled through a renewal process with Gamma sampling intervals with fixed $\Delta = 0.25$. Histogram of the components of $\sqrt{n}\frac{\partial}{\partial \boldsymbol{\theta}} WLL_n^{\Delta}(\boldsymbol{\theta}_0)$, based on 1000 independent realizations of $X = X(t)$ each with $n = 1000$ time samples, are plotted in Figure(5.1) and (5.1), corresponding to different variances of the IID sampling intervals. We also included the red curve as the fitted normal distribution.

Judging from the histograms and the fitted normal distribution curve, it seems reasonable to expect the first derivative term $\sqrt{n}\frac{\partial}{\partial \boldsymbol{\theta}} WLL_n(\boldsymbol{\theta}_0)$ to be asymptotically normally distributed, under the assumption that the underlying continuous time process is Gaussian. We have also conducted a number of other simulations with (1)different parameters combinations for the Oscillatory Matérn process; (2)different renewal sampling schemes - Uniform, Gamma, and Inverse Gamma distributions with different mean $\Delta = \mathrm{E}(\Delta_k)$ and variance $\mathrm{Var}(\Delta_k)$. The histograms of the components in $\sqrt{n}\frac{\partial}{\partial \boldsymbol{\theta}} WLL_n(\boldsymbol{\theta}_0)$ did not show serious deviations from the fitted normal distributions, suggesting that overall the asymptotic normality of $\sqrt{n}\frac{\partial}{\partial \boldsymbol{\theta}} WLL_n(\boldsymbol{\theta}_0)$, and consequently the Whittle estimator $\hat{\boldsymbol{\theta}}_n$, can be empirically justified.

In existing literature the asymptotic normality of an estimator is primarily used to quantify finite sample uncertainty through constructing confidence intervals. Looking back at Equation(5.79), we can see that constructing confidence interval involves the evaluation of both the matrix $\Sigma_Y(\hat{\boldsymbol{\theta}}_n)$ and $\Omega_Y(\hat{\boldsymbol{\theta}}_n)$. It will be discussed in next chapter that the evaluation of $S_Y^{\Delta}(f; \hat{\boldsymbol{\theta}}_n)$ can be conveniently done through DFT approximations, and hence the matrix $\Sigma_Y(\hat{\boldsymbol{\theta}}_n)$ can be readily evaluated numerically.

On the other hand, since the discrete time stationary process $Y = Y_k$ is obtained from sampling through a renewal process, it will not in general being Gaussian, regardless of the Gaussianity of the underlying continuous time process $X = X(t)$. Conse-

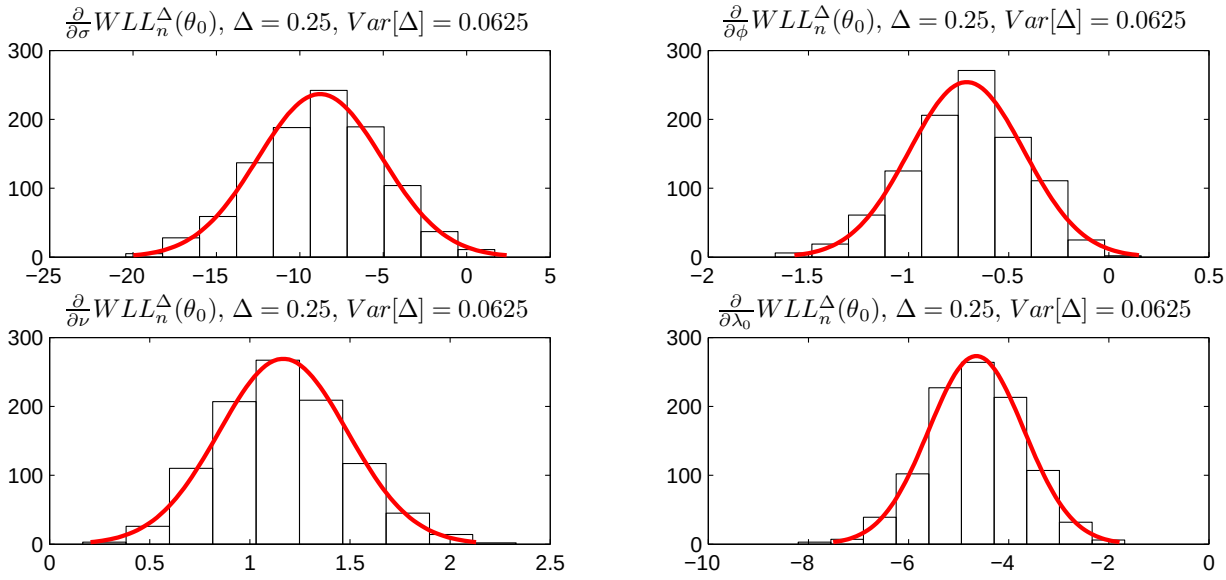

Figure 5.1: Histogram for elements of $\sqrt{n}\frac{\partial}{\partial\boldsymbol{\theta}}WLL_n^{\Delta}(\boldsymbol{\theta}_0)$, from 1000 independent realizations of Oscillatory Matérn process with parameters $\sigma = 1, \phi = 1, \nu = 1, \lambda_0 = 1$. Sample size is $n = 1000$. Renewal sampling intervals have IID Gamma$(1, 4)$ distribution.

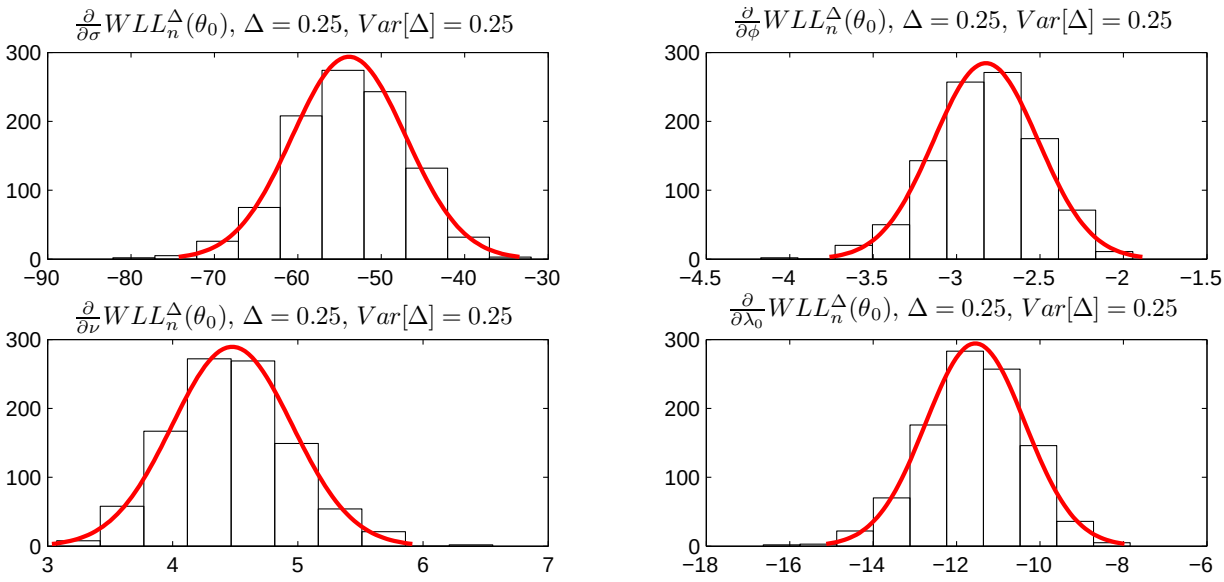

Figure 5.2: Histogram for elements of $\sqrt{n}\frac{\partial}{\partial\boldsymbol{\theta}}WLL_n^{\Delta}(\boldsymbol{\theta}_0)$, from 1000 independent realizations of Oscillatory Matérn process with parameters $\sigma = 1, \phi = 1, \nu = 1, \lambda_0 = 1$. Sample size is $n = 1000$. Renewal sampling intervals have IID Gamma$(0.25, 1)$ distribution.

quently the evaluation of the matrix $\Omega_Y(\hat{\boldsymbol{\theta}}_n)$ will involve the estimation of the a double integral with respect to the fourth order cumulant density $Q_Y(f, \mu, \nu)$. A consistent estimator of such a double integral of $Q_Y(f, \mu, \nu)$ can be found in Taniguchi [126], and in Section(6.4.4) we will use this estimator to construct confidence intervals for our proposed Whittle estimator $\hat{\boldsymbol{\theta}}_n$.

5.7 Summary

This chapter discussed the problem of parametric estimation of a continuous time stationary processes under renewal sampling schemes. By constructing the Whittle log-likelihood function based on the sampled discrete time stationary process, we discovered that the parametric estimation problem can be conveniently formulated under rather general conditions. We established a key convergence result, and showed that the spectral density $S_Y^{\Delta}(f)$ exists and is positive for all f, so that the Whittle log-likelihood function is well defind (i.e. finite). The asymptotic consistency of the Whittle's estimator, obtained from minimizing the Whittle log-likelihood function, were discussed within the framework provided by Ibragimov[54]. We showed that under appropriate assumptions the conditions of Ibragimov's asymptotic consistency theorem will be satisfied. Finally, although we were not successful in proving rigorously the asymptotic normality of the Whittle's estimator, a set of simulation studies were conducted showing that asymptotically normality shoule be achieved in most of the practical situations.

Chapter 6

Performence of the Whittle Log-Likelihood

6.1 Introduction

Having discussed the asymptotic consistency, the focus in this chapter will be turned to the performance issue of the proposed Whittle log-likelihood estimator $\hat{\boldsymbol{\theta}}_n$. We will provide algorithms on the implementation, and insights on the finite sample performence, of the proposed estimation method under renewal sampling schemes.

We will first consider the computational efficiency of the proposed estimation method. In Section(6.2) it will be discussed that when the covariance function $R_X(\tau)$ has relatively fast decay, then the autocovariances $C_Y(k)$ will also admit a relatively fast decay. Consequently the spectral density $S_Y^{\Delta}(f)$ can be efficiently calculated through a Discrete Fourier Transform (DFT) approximations using only a limited number of autocovariances. This therefore allows us to calculate the proposed Whittle log-likelihood function $WLL_n^{\Delta}(\boldsymbol{\theta})$ much more efficiently when compared to the calculation of the traditional Gaussian log-likelihood function.

Next we will give a detailed discussion on the finite sample performance of the proposed estimator $\hat{\boldsymbol{\theta}}_n$. The major contribution of this chapter will be in Section(6.3), in which a detailed analysis of the unique properties of the renewal kernel function $K_{\Delta}(f, \lambda)$ will be given there. These unique properties have not been discussed in existing literature on renewal sampling schemes [113, 8, 106]. In particular we will show that the kernel function has a peak effect and an aggregation effect, so that the power $S_Y^{\Delta}(f)$ of the sampled process can be decomposed into two distinct components

accordingly. Such a decomposition allows us to see intuitively that, as compared to the power relocation implied by equally spaced aliasing effect, the renewal sampling scheme introduces a more complicated power-mixing. Such a power-mixing can be used to conveniently explain the loss of spectral information under renewal sampling schemes as compared to the scenario under equally spaced sampling schemes. Moreover through an asymptotic argument over the renewal kernel function, we will see how the loss of spectral information is affected by the mean and variances of the renewal sampling intervals. Finally in Section(6.4) we will present a set of simulation studies to verify the conclusions obtained in the discussions of previous sections.

6.2 Computing the Whittle Log-Likelihood

This section discusses issues related to numerical computation of the Whittle log-likelihood function $WLL_n^{\Delta}(\boldsymbol{\theta})$, which was defined in the last chapter by Equation(5.2). For each evaluation of $WLL_n^{\Delta}(\boldsymbol{\theta})$, we need to calculate the values of the periodogram $I_n^{\Delta}(f)$ and the spectral density $S_Y^{\Delta}(f;\boldsymbol{\theta})$ at the set of Fourier frequencies $\{\frac{k}{n\Delta} : k = 0, 1, \cdots, n-1\}$. The computational costs associated with the periodogram should be negligible, since it needs to be calculated only once through the efficient FFT technique. Consequently the efficiency of the proposed estimation method is primarily determined by the computational efficiency of evaluating the spectral density $S_Y^{\Delta}(f;\boldsymbol{\theta})$ of the discrete time sampled process.

Unfortunately $S_Y^{\Delta}(f;\boldsymbol{\theta})$ does not usually admit a closed form expression, and hence numerical approximations will be required. In this section, we will argue that when the covariance function $R_X(\tau;\boldsymbol{\theta})$ decays exponentially fast, the autocovariance sequence $C_Y^{\Delta}(k;\boldsymbol{\theta})$ of the sampled discrete time process should also admit a similar exponential decay. Consequently $S_Y^{\Delta}(f;\boldsymbol{\theta})$ can be approximated efficiently through a Discrete Fourier Transform (DFT) approximation with only a few terms of the autocovaruances. We will show through numerical experiments that this finite DFT approximation of $S_Y^{\Delta}(f;\boldsymbol{\theta})$ will give our proposed Whittle log-likelihood estimation method a distinctive computational advantage, as compared with the traditional method based on minimizing the Gaussian log-likelihood function. Finally before going into detailed discussions, we need to mention that for the purpose of notational clarity, we will sometimes suppress the appearance of $\boldsymbol{\theta}$ and simply write $S_Y^{\Delta}(f)$, instead of $S_Y^{\Delta}(f;\boldsymbol{\theta})$, as the spectral density of the discrete time sampled process.

6.2.1 Computing the Spectral Density $S_Y^{\Delta}(f)$

Assuming the renewal sampling intervals have finite moments of all orders, then $S_Y^{\Delta}(f)$ admits the spectral representation given by Equation(5.30). Consequently one straightforward approach to evaluate the $S_Y^{\Delta}(f)$ is to numerically integrate this spectral domain expression. The primary disadvantage of this approach was in terms of computational costs, in particular when the sample size is large. This is because the values of the function $S_Y^{\Delta}(f)$ are required over the set of Fourier frequencies $\{\frac{k}{n\Delta} : k = 0, 1, \cdots, n-1\}$, and each evaluation through Equation(5.30) involves a numerical quadrature approximation. Consequently the total number of numerical integrations that are required to compute the values of $S_Y^{\Delta}(f)$ in this way increases linearly with the sample size n, which is certainly not a desirable situation.

On the other hand, instead of using the spectral domain expression of Equation(5.30), we note that $S_Y^{\Delta}(f)$ can also be calculated, at least in principle, through the fact that it is the DFT of the autocovariance sequence $C_Y(k)$. This observation therefore suggests that we can first calculate $C_Y(k)$ for lags $k = 0, 1, \cdots, M$, where M is an integer such that $|C_Y(k)|$s are negligible for $k > M$. Then an approximation of $S_Y^{\Delta}(f)$ over the set of Fourier frequencies can be obtained from the following expression:

$$S_Y^{\Delta}(\frac{k}{n\Delta}) \approx \Delta \sum_{j=-M}^{M} C_Y(j) e^{-i2\pi kj/n} = \Delta C_Y(0) + 2\Delta \sum_{j=1}^{M} C_Y(j) \cos(2\pi kj/n). \quad (6.1)$$

Note that as soon as the autocovariances are calculated, the values of $S_Y^{\Delta}(\frac{k}{n\Delta})$ can be obtained, from Equation(6.1), with trivial computational costs through the FFT technique. Consequently using this approach the computational efficiency of evaluating $S_Y^{\Delta}(\frac{k}{n\Delta})$ is primarily determined from evaluating $M+1$ autocovariances $C_Y(k)$, and hence will be largely independent of the sample size n. If the required number M of autocovariances is reasonable, the DFT approximation approach is likely to enjoy computational advantage when the sample size is relatively large.

The autocovariance sequence $C_Y(k)$ for $0 \leq k \leq M$ could be most conveniently evaluated, through numerical quadratures, from the corresponding frequency domain expression given by Equation(5.20). Using the fact that $S_X(\lambda)$ is even around the origin, this frequency domain expression can also be written in real form as

$$C_Y(k) = 2 \int_0^{\infty} S_X(\lambda) \mathrm{Re}\left[\phi_{\Delta}(\lambda)^{|k|}\right] d\lambda. \quad (6.2)$$

The major reason of using the spectral domain expression of $C_Y(k)$ is from the consideration of practical implementation. In most cases, the density function $p_k(\tau)$ will not be available in closed form; and even in the case of Gamma sampling, in which the $p_k(\tau)$ takes the form of a Gamma(α, β) density function, its evaluation may be problematic for relatively large k, due to the appearance of the Gamma function $\Gamma(k\alpha)$. [1] Comparing with the corresponding time domain expression for $C_Y(k)$, however, the spectral domain expression depends on the sampling scheme only through the characteristic function $\phi_\Delta(\lambda)$, which usually takes a much simpler form than the density function $p_{|k|}(\tau)$ that enters into the corresponding time domain expression. Once the characteristic function is known, the evaluation of $C_Y(k)$ for $k = 0, 1, \cdots, M$ can be efficiently implemented in Matlab using the vectorized quadrature function quadv().

There are, however, potential problems associated with evaluating $S_Y^\Delta(f)$ through this DFT approximation. For one thing, when the probability density $p(\tau)$ of the IID sampling interval contains discontinuities (e.g. IID sampling interval has uniform or beta distribution), the real part of the characteristic function $\phi_\Delta(f)$ becomes oscillatory, and the oscillation of $\mathrm{Re}[\phi_\Delta(f)^{|k|}]$ in the integrand becomes faster as k increases. Figure(6.1) shows an example of this behaviour of using the characteristic function of a Uniform$(0, 0.5)$ distribution. This potential oscillatory behaviour of the integrand for large k is known to make traditional quadrature algorithm not reliable (see for example Iserles[55, 56]).

Another potential problem concerns the number of maximum lags M that the autocovariance sequence is required to approximate the function $S_Y^\Delta(f)$. Large M implies a higher computational costs, making the proposed Whittle log-likelihood estimation method less computationally appealing. In practice, however, most of the spectral density models are sufficiently smooth, implying that the corresponding covariance functions $R_X(\tau)$ will decay exponentially fast. Naturally in this case it is expected that the autocovariances $C_Y(k)$ should also have relatively fast decay, so that DFT approximation of $S_Y^\Delta(f)$ does not require a large number of terms. This also implies that the potential oscillatory behaviour of $\mathrm{Re}[\phi_\Delta(f)^{|k|}]$ will not impose

[1] When the IID sampling intervals have Gamma(α, β) distribution, then it is immediate that $p_k(\tau)$ takes the form

$$p_k(\tau) = \frac{\beta^\alpha}{\Gamma(k\alpha)} \tau^{k\alpha - 1} e^{-\beta\tau}. \tag{6.3}$$

For k relatively large, the appearance of $\Gamma(k\alpha)$ in the denominator may give us unreliable results in numerical evaluations.

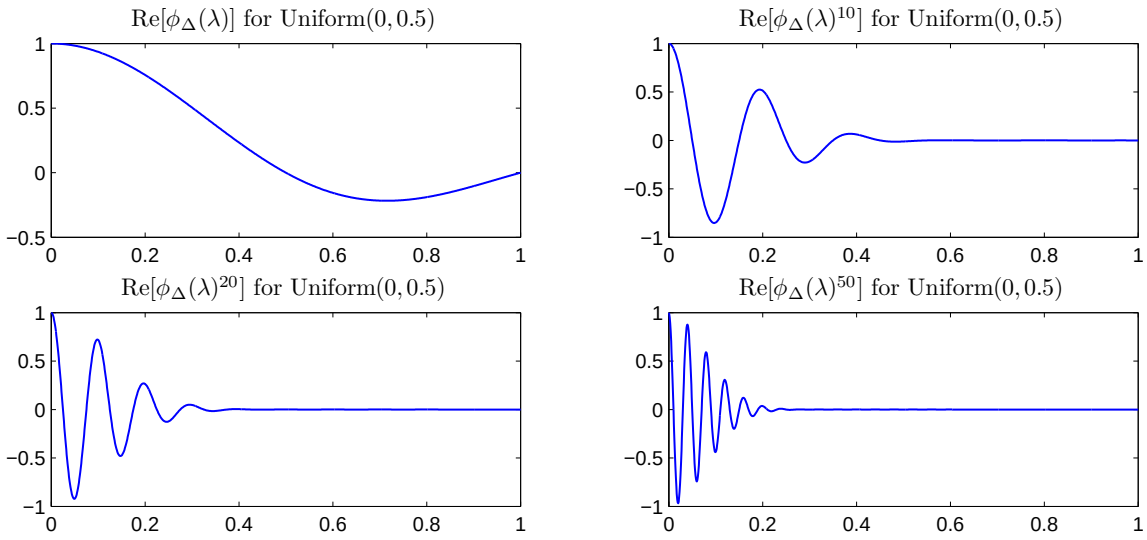

Figure 6.1: Oscillatory behaviour of $\mathrm{Re}[\phi_\Delta(f)^{|k|}]$ as k increases, using Uniform$(0, 0.5)$ distribution as an example.

serious problems in practice. Next section will provide a quantitative description of the decay of $C_Y(k)$ for a large class of $R_X(\tau)$. For this class of covariance functions, we will also propose an algorithm that can be used to determine the number M of autocovariances required to calculate $S_Y^\Delta(f)$ through the DFT approximation.

6.2.2 Decay of the Autocovariance $C_Y(k)$

As mentioned in the previous section, the efficient evaluation of $S_Y^\Delta(f)$ and hence the Whittle log-likelihood function $WLL_n^\Delta(\boldsymbol{\theta})$ depends, to a large extent, on the number M of the autocovariance $C_Y(k)$ that have to be evaluated through numerical integrations. This in turn depends on the decay of the autocovariance sequence, which should be directly related to the decay of the covariance function $R_X(\tau)$ of the underlying continuous time process. This can be most intuitively seen from the time-domain expression for $C_Y(k)$, which is given by

$$C_Y(k) = \int_0^\infty R_X(\tau) p_{|k|}(\tau) d\tau. \tag{6.4}$$

Note that in this time-domain expression, the density function $p_{|k|}(\tau)$ corresponds to a distribution with mean $k\Delta$ and variance $k\text{Var}[\Delta]$. Hence, the value of $C_Y(k)$ is just a weighted average of the covariance function $R_X(\tau)$ over a neighborhood centered at $\tau = k\Delta$. This observation therefore suggests that the decay of $R_X(\tau)$ is going to be one of the important factors that determines the decay of the autocovariance $C_Y(k)$.

Simplifying Approximations

However, the above discussion only gives an intuitive understanding. In order to determine the appropriate M that should be used to evaluate $S_Y^{\Delta}(f)$, we will require a quantitative measure of the decay of $C_Y(k)$. We will see in this section that this can be done by assuming that the covariance function $R_X(\tau)$ admits the following damped oscillation as the asymptotic form as τ being relatively large:

$$R_X(\tau) \propto \tau^{\nu} e^{-\phi\tau} \cos(2\pi\lambda_0\tau + \psi), \tag{6.5}$$

where $\phi > 0, \lambda_0 > 0$, but ν and ψ can be both positive or negative. It has to be pointed out that this asymptotic form includes almost all the decay patterns for integrable short-range dependence covariance models that are frequently used in practice.

For the purpose of obtaining relatively simple expressions, we will further approximate the the density function $p_{|k|}(\tau)$, for k relatively large, by a Gamma density function $g_{|k|}(\tau)$ with mean $k\Delta$ and variance $k\text{Var}[\Delta]$. The reasons for this approximation are two folded: first of all it will be shown shortly that the asymptotic form of $R_X(\tau)$ given by Equation(6.5) can be integrated exactly with $g_{|k|}(\tau)$; on the other hand for relatively large k, the gamma density $g_{|k|}(\tau)$ should approximate $p_{|k|}(\tau)$, because both densities should be close in shape to a normal density with mean $k\Delta$ and variance $k\text{Var}[\Delta]$. The gamma density $g_{|k|}(\tau)$ that we use to approximate $p_{|k|}(\tau)$ takes the following form

$$g_{|k|}(\tau) = \frac{\beta^{\alpha_{|k|}}}{\Gamma(\alpha_{|k|})} \tau^{\alpha_{|k|}-1} e^{-\beta\tau}, \tag{6.6}$$

where $\Gamma(\cdot)$ represent the Gamma function. The location parameter $\alpha_{|k|}$ and the scale parameter β are chosen so that the mean and variance of the density $g_{|k|}(\tau)$ matches

the mean and variance of the density $p_{|k|}(\tau)$:

$$\alpha_{|k|} = |k|\frac{\Delta^2}{\text{Var}[\Delta]}, \text{ and } \beta = \frac{\Delta}{\text{Var}[\Delta]}. \tag{6.7}$$

Note that when the IID sampling intervals have gamma distribution, such an approximation will become exact. We have to, however, emphasize the fact that such an approximation is not meant to be numerically accurate in terms of approximating the values of $C_Y(k)$. It only aims to provide an insightful understanding for factors determining the decay of the autocovariance sequence $C_Y(k)$, so that a rough quantitative measure of the decay of $C_Y(k)$ can be estimated.

Factors Affecting the Decay of $C_Y(k)$

Using the simplifying assumptions discussed previously, we can derive an informative expression for the autocovariance $C_Y(k)$. This is summarized in the following result:

Proposition 6.1. *Suppose $R_X(\tau)$ follows the asymptotic expression given by Equation(6.5). Then using the Gamma density $g_{|k|}(\tau)$ to approximate $p_{|k|}(\tau)$ for relatively large k, we can derive the following asymptotic expression for $C_Y(k)$ as*

$$C_Y(k) \propto \left[\frac{\beta(\alpha_k+\nu-1)}{(\beta+\phi)\sqrt{\beta^2+4\pi^2\lambda_0^2}}\right]^{\nu} \times \left(\frac{\beta}{\beta+\phi}\right)^{\alpha_k}\left(\frac{\beta}{\sqrt{\beta^2+4\pi^2\lambda_0^2}}\right)^{\alpha_k}\cos\left[\arctan\left(\frac{2\pi\lambda_0}{\beta}\right)(\alpha_k+\nu)+\psi\right], \tag{6.8}$$

where α_k and β are given by Equation(6.7).

The derivation of Equation(6.8) amounts to integrating the asymptotic expression of $R_X(\tau)$ with respect to the Gamma density $p_{|k|}(\tau)$, and the details can be found in Section(8.3.1) in the Appendix of this thesis. It is then obvious that this asymptotic expression $C_Y(k)$ can be decomposed into the following four parts:

1. Polynomial part: $\left[\frac{\beta(\alpha_k+\nu-1)}{(\beta+\phi)\sqrt{\beta^2+4\pi^2\lambda_0^2}}\right]^{\nu}$, which corresponds to the τ^{ν} part in Equation(6.5);
2. Exponential damping factor: $(\frac{\beta}{\beta+\phi})^{\alpha_k}$, which corresponds to $e^{-\phi\tau}$ part of Equation(6.5);

3. Oscillation factor: $\cos\left[\arctan\left(\frac{2\pi f_0}{\beta}\right)(\alpha_k+\nu)+\psi\right]$, which corresponds to the $\cos(2\pi\lambda_0\tau+\psi)$ part of Equation(6.5);
4. Extra exponential damping factor: $\left(\frac{\beta}{\sqrt{\beta^2+4\pi^2\lambda_0^2}}\right)^{\alpha_k}$, which arises from the oscillatory in $R_X(\tau)$.

The first three part in the above list resemble the asymptotic decay pattern of $R_X(\tau)$, and the decay of $R_X(\tau)$ is an important factor determining the decay of $C_Y(k)$: when $R_X(\tau)$ decays relatively fast, i.e. ϕ being relatively large, it is obvious that the exponential damping factor $(\frac{\beta}{\beta+\phi})^{\alpha_k}$ in $C_Y(k)$ is more significant, implying a faster decay for $C_Y(k)$. What is more interesting is the fact that $C_Y(k)$ also contains an extra exponential damping factor $\left(\frac{\beta}{\sqrt{\beta^2+4\pi^2\lambda_0^2}}\right)^{\alpha_k}$ that is introduced from the oscillation in $R_X(\tau)$. Hence when the $R_X(\tau)$ contains oscillation, the magnitude of $C_Y(k)$ will have a faster decay because of this extra damping factor. This is, however, expected from the time domain expression of $C_Y(k)$ given by Equation(6.4): being roughly a weighted average of $R_X(\tau)$ around $\tau = k\Delta$, oscillation in $R_X(\tau)$ would imply some degrees of cancellations from positive and negative covariances, hence making $C_Y(k)$ smaller in magnitude. The extra exponential damping factor simply gives a quantitative description of this intuitive explanation. We will also see in Section(6.3.2) that this extra damping factor introduced from oscillation in $R_X(\tau)$ is a particular example of the whitening effect introduced from the renewal sampling scheme.

Other factors influencing the decay of $C_Y(k)$ include the average sampling interval Δ and the variance $\mathrm{Var}[\Delta]$ of the sampling intervals. Assuming $k > 0$, and recall that the expressions for α_k and β from Equation(6.7), it is then immediate that $\alpha_{|k|}$ has the following two expressions in terms of β:

$$\alpha_k = k\Delta\beta, \text{ or } \alpha_k = k\mathrm{Var}[\Delta]\beta^2. \tag{6.9}$$

Using the first expression for α_k, we can express the exponential damping factor as

$$\text{Exponential damping factor: } \left(\frac{\beta}{\beta+\phi}\right)^{k\Delta\beta}. \tag{6.10}$$

On the other hand, using the second expression for α_k we can write the extra damping

factor as

$$\text{Extra damping factor: } \left(\frac{\beta^2}{\beta^2 + 4\pi^2\lambda_0^2}\right)^{k\text{Var}[\Delta]\beta^2} . \tag{6.11}$$

From the above expression for the exponential damping factor, it can be recognized that it will be mostly affected by the average sampling interval Δ. This is completely expected, since a larger Δ means that on average the samples are further apart, and consequently $C_Y(k)$ will decay faster. On the other hand, it can also be observed from the expression for the extra damping factor that it is primarily affected by the variance Var[Δ] of the sampling intervals. Higher value for Var[Δ] implies that the density function $g_k(\tau)$ is more flat in shape, and hence gives faster decay of $C_Y(k)$ by allowing more cancellation effect against the oscillation of $R_X(\tau)$.

6.2.3 Approximating $S_Y^{\Delta}(f)$

Apart from providing intuitive insights on the decay of $C_Y(k)$, the asymptotic expression in Equation(6.8) also suggests a practical way to determine the cut-off lag M, i.e. the number of autocovariances to be included in the DFT approximation of $S_Y^{\Delta}(f)$. Because each autocovariance $C_Y(k)$ is calculated through numerical quadrature with a prespecified precision of, say ϵ, it does not make sense to include autocovariances with magnitude smaller than ϵ. Hence a natural criteria for selecting the number of terms for the DFT approximation is to find a reasonable estimate for M such that $|C_Y(M)| \leq \epsilon$.

Also we note that in order to derive appropriate expressions for finding M, it is legitimate to focus on the envelope of the autocovariance sequence, by ignoring the oscillation component $\cos\left[\arctan\left(\frac{2\pi\lambda_0}{\beta}\right)(\alpha_m + \nu) + \psi\right]$ in the asymptotic expression for $C_Y(k)$. Consequently we would like to find M such that

$$\epsilon = \left[\frac{\beta(M\Delta\beta + \nu - 1)}{(\beta+\phi)\sqrt{\beta^2 + 4\pi^2\lambda_0^2}}\right]^{\nu} \left(\frac{\beta}{\beta+\phi}\right)^{M\Delta\beta} \left(\frac{\beta}{\sqrt{\beta^2 + 4\pi^2\lambda_0^2}}\right)^{M\Delta\beta} , \tag{6.12}$$

where we have used the expression $\alpha_M = M\Delta\beta$. Upon taking logarithm on both

sides, the above equation then becomes

$$\log(\epsilon) = \nu \log\left[\frac{\beta(M\Delta\beta + \nu - 1)}{(\beta+\phi)\sqrt{\beta^2 + 4\pi^2\lambda_0^2}}\right] + M\Delta\beta\left\{\log\left(\frac{\beta}{\beta+\phi}\right) + \log\left(\frac{\beta}{\sqrt{\beta^2+4\pi^2\lambda_0^2}}\right)\right\}, \tag{6.13}$$

which, at least in principal can be easily solved for M by a non-linear solver. It is, however, not necessary, since as mentioned earlier the required M does not need to be extremely accurate - it suffices to come up with a rough estimate of M that is not excessively large. In order to achieve this purpose, we propose the following two-step estimation procedure:

Algorithm 6.1.

1. *Solve Equation(6.13) without the first term to obtain M' as*

$$M' = \frac{\log(\epsilon)}{\Delta\beta\left\{\log\left(\frac{\beta}{\beta+\phi}\right) + \log\left(\frac{\beta}{\sqrt{\beta^2+4\pi^2\lambda_0^2}}\right)\right\}}. \tag{6.14}$$

   Although this M' may not be a very good estimate for M, it could be used to approximate the first term in Equation(6.13), because the logarithm term varies slowly with respect to M.

2. *By using M' to estimate the first term in Equation(6.13), we can then obtain the following estimate for the cut-off lag M:*

$$M = \frac{\log(\epsilon) - \nu \log\left[\frac{\beta(M'\Delta\beta+\nu-1)}{(\beta+\phi)\sqrt{\beta^2+4\pi^2\lambda_0^2}}\right]}{\Delta\beta\left\{\log\left(\frac{\beta}{\beta+\phi}\right) + \log\left(\frac{\beta}{\sqrt{\beta^2+4\pi^2\lambda_0^2}}\right)\right\}}. \tag{6.15}$$

   which should give a slightly more accurate estimate of the cut-off lag M beyond which the magnitude of $C_Y(k)$ becomes negligible.

6.2.4 Numerical Example: Decay and Accuracy

In this section, we consider some concrete examples to illustrate the DFT approximation of $S_Y^{\Delta}(f)$. We will be considering the class of Oscillatory Matérn process, which was introduced in Section(2.3.3). Using the following asymptotic expression for the Bessel function (see Abramowitz and Stegun [1])

$$K_\nu(x) \propto \sqrt{\frac{\pi}{2x}} e^{-x}, \tag{6.16}$$

we can then easily shown that for relatively large τ, the covariance function for the Oscillatory Matérn process admits the following asymptotic form:

$$R_X(\tau) \propto \tau^{\nu-0.5} e^{-2\pi\phi|\tau|} \cos(2\pi\lambda_0\tau). \tag{6.17}$$

Therefore it can be immediately recognized that asymptotically the Oscillatory Matérn covariance function belongs to the damped oscillation class given by Equation(6.5), and consequently the asymptotic results of the last section should apply. In order to illustrate the effects of the oscillation in $R_X(\tau)$ on the decay of $C_Y(k)$, we will consider the following three specific examples of Oscillatory Matérn processes with

1. No oscillation: $\sigma = 1$, $\phi = 0.5$, $\nu = 1$ and $\lambda_0 = 0$. Note that this is simply the ordinary Matérn covariance function frequently used in practice.

2. Slow oscillation: $\sigma = 1$, $\phi = 0.5$, $\nu = 1$ and $\lambda_0 = 0.63$.

3. Relatively high oscillation: $\sigma = 1$, $\phi = 0.5$, $\nu = 1$ and $\lambda_0 = 1.63$.

Moreover, in order to show that the asymptotic results of the previous section is applicable for general renewal sampling schemes, we considered two renewal sampling schemes with Gamma$(3, 12)$ and Uniform$(0, 0.5)$ distributions as the IID sampling intervals. Note that these two distributions have the same means and variances. Figure(6.2) plots the magnitude of both $C_Y(k)$ and $R_X(\tau)$ on a logarithmic scale with a base 10, so that the order of magnitude can be shown more clearly. Also included in the figure are vertical dashed lines, which indicate the cut-off leg M calculated through the two-step algorithm proposed in the previous section, corresponding to an error with a magnitude of $\epsilon = 10^{-9}$.

It is evident from the plot that the autocovariance sequences in all three cases exhibit an exponential decay pattern (i.e. linear in the logarithmic scale) that is

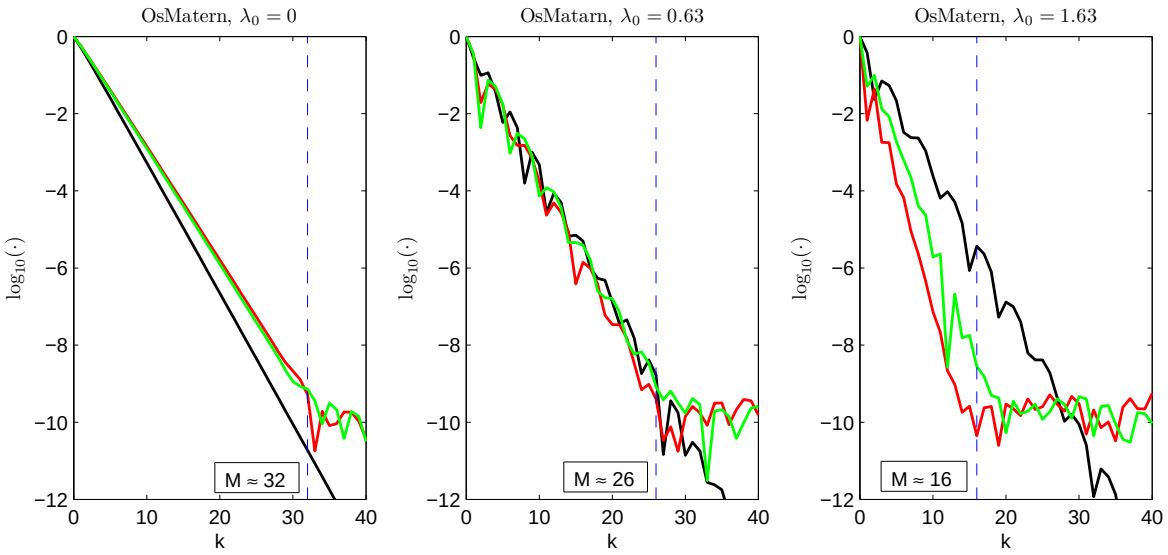

Figure 6.2: Decay of $|C_Y(k)|$ (red curve for Uniform sampling, green curve for Gamma sampling) and $|R_X(\tau)|$ (black curve), on a logarithmic scale with base 10. The vertical blue dashed line indicate the calculated cut-off lag M according to Algorithm(6.1).

consistent with the asymptotic result obtained in last section, regardless of the renewal sampling schemes being used. The effect of the oscillation in the covariance function is also clearly demonstrated: when the oscillation is slow (i.e. $\lambda_0 = 0.63$) the extra damping factor is close to one (0.95) and consequently does not have a significant effect; on the other hand when the oscillation is relatively fast (i.e. $\lambda_0 = 1.63$), it introduces a more significant damping factor (0.76), thus making the decay of $C_Y(k)$ considerably faster than the decay of the corresponding covariance function $R_X(\tau)$.

It can also be observed from Figure(6.2) that the two-step estimation procedure based on our proposed algorithm is quite satisfactory in determining the cut-off lag M, beyond which the magnitude of the auto-covariance sequence reaches the prescribed tolerance level $\epsilon = 10^{-9}$, and hence are negligible. Note that the estimated M are generally not large, and hence it is expected that the spectral density $S_Y^{\Delta}(f)$ approximated from DFT approximation should be relatively efficient, and at the same time will also provide sufficient accuracy.

For the purpose of directly illustrating the accuracy of approximating $S_Y^{\Delta}(f)$, we consider the special case of the Oscillatory Matérn process corresponding to $\nu = 0.5$,

so that the covariance function admits the following simple form:

$$R_X(\tau) = \sigma^2 e^{-2\pi\tau}\cos(2\pi\lambda_0\tau). \tag{6.18}$$

For such a special case, the autocovariance sequence $C_Y(k)$ can be explicitly derived as

$$\begin{aligned} C_Y(k) &= \int_0^\infty \sigma^2 e^{-2\pi\phi\tau}\cos(2\pi\lambda_0\tau)p_{|k|}(\tau)d\tau \\ &= \int_0^\infty \sigma^2 e^{-2\pi\phi\tau}\mathrm{Re}[e^{i2\pi\lambda_0\tau}]p_{|k|}(\tau)d\tau \\ &= \mathrm{Re}\left[\int_0^\infty \sigma^2 e^{i2\pi\left(\lambda_0-\frac{\phi}{i}\right)\tau}p_{|k|}(\tau)d\tau\right] \\ &= \sigma^2\mathrm{Re}\left[\phi_\Delta\left(\lambda_0-\frac{\phi}{i}\right)^{|k|}\right]. \end{aligned} \tag{6.19}$$

Using the above expression, the spectral density $S_Y^\Delta(f)$ of the discrete time sampled process can then be calculated exactly as

$$S_Y^\Delta(f) = \sigma^2\mathrm{Re}\left[\frac{1-r^2}{1-2r\cos(2\pi f)+r^2}\right], \text{ where } r = \phi_\Delta\left(\lambda_0-\frac{\phi}{i}\right). \tag{6.20}$$

This exact expression for $S_Y^\Delta(f)$ can be used to illustrate the accuracy of the DFT approximation for $S_Y^\Delta(f)$. Figure(6.3) and (6.4) show the error of approximation under renewal sampling schemes with IID sampling intervals by a Gamma$(3, 12)$ distribution and an Uniform$(0, 0.5)$ distribution, respectively. The Oscillatory Matérn processes under consideration is chosen to be the following

1. No oscillation: $\sigma = 1$, $\phi = 0.5$, $\nu = 0.5$, and $\lambda_0 = 0$. Note that this simply gives the stationary Gaussian Markov process that is frequently used in practice.

2. Slow oscillation: $\sigma = 1$, $\phi = 0.5$, $\nu = 0.5$ and $\lambda_0 = 0.63$.

3. Relatively high oscillation: $\sigma = 1$, $\phi = 0.5$, $\nu = 0.5$ and $\lambda_0 = 1.63$.

The number M of terms in all the cases are determined from Algorithm(6.1), with the tolerance level being set to $\epsilon = 10^{-9}$. It is then evident from these plots that the DFT approximation of $S_Y^\Delta(f)$ can give us sufficient accuracy.

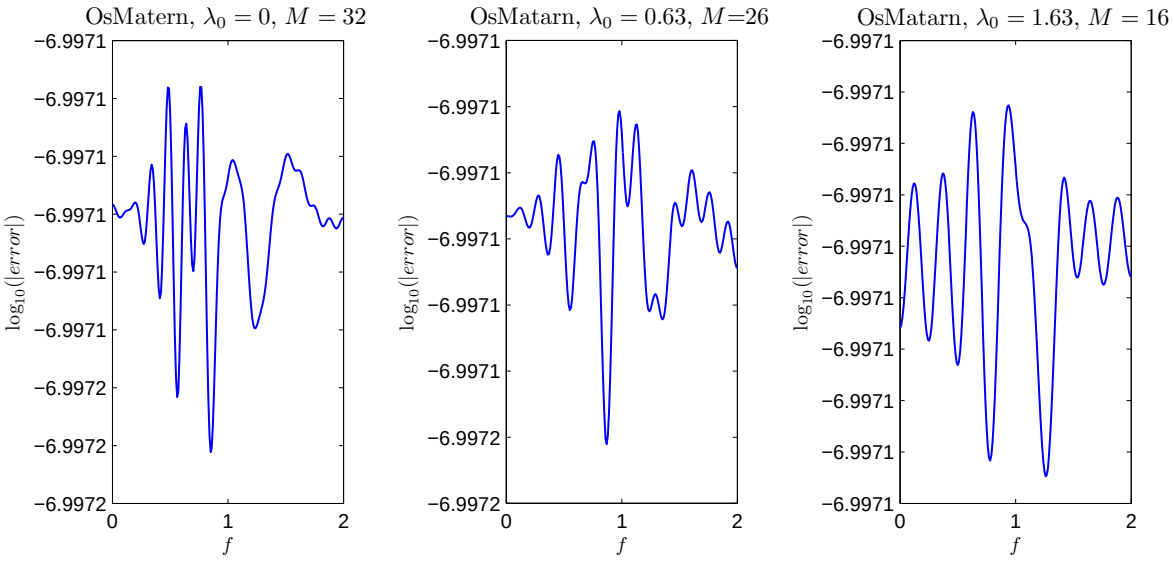

Figure 6.3: Error of approximating $S_Y^{\Delta}(f)$ through the DFT, for Oscillatory Matérn processes sampled through a Gamma$(3, 12)$ renewal sampling scheme.

6.2.5 Numerical Example: Computational Efficiency

Examples in the last section showed that the spectral density $S_Y^{\Delta}(f)$ can be calculated, with good accuracy, through DFT approximation. Note that in all those examples the required number M of autocovariances is not large, consequently implying that $S_Y^{\Delta}(f)$, and consequently the Whittle log-likelihood function $WLL_n^{\Delta}(\boldsymbol{\theta})$ can be calculated with relatively low computational costs. In this section, we will present more examples, showing that this computational advantage indeed exists under quite general situations. We will again use the Oscillatory Mat'ern process as the underlying continuous time process. We will fix the parameters at $\sigma = 1$, $\nu = 1$, $\lambda_0 = 0.63$, and let the covariance decay parameter ϕ to vary. The renewal sampling scheme is chosen to have different mean sampling intervals. Table(6.1) contains the number M of terms, calculated through our proposed two-step algorithm. It can be immediately observed from this table that under all the situations considered, the number M of terms are not large. Moreover it can also be observed that when the mean sampling interval is relatively large, or when the covariance function decays relatively fast, we then required a smaller number M of terms for calculating $S_Y^{\Delta}(f)$.

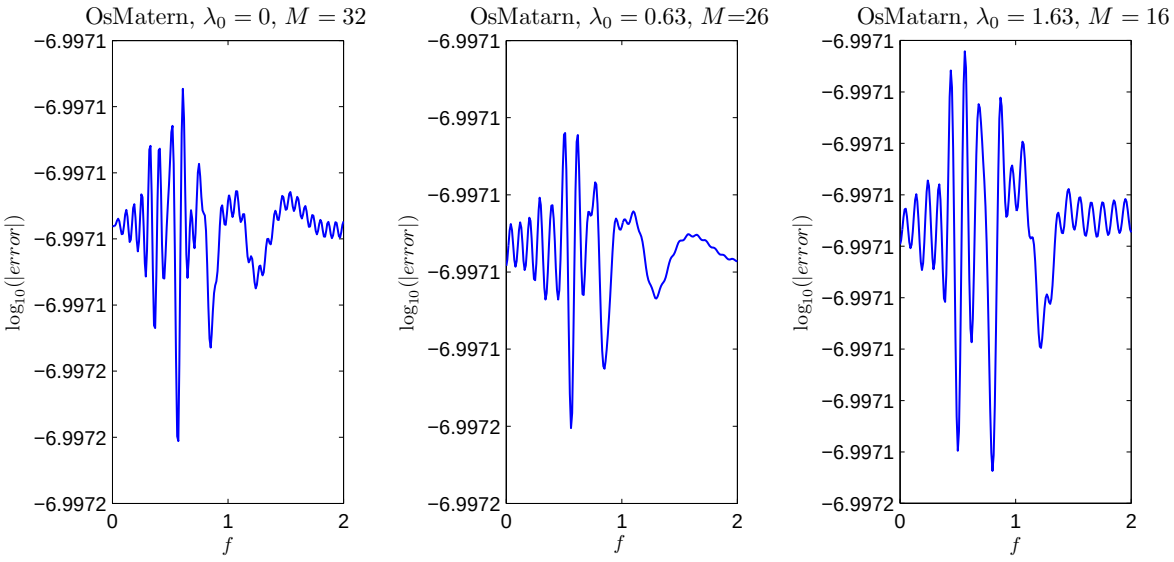

Figure 6.4: Error of approximating $S_Y^{\Delta}(f)$ through DFT, for Oscillatory Matérn processes sampled through a Uniform$(0, 0.5)$ renewal sampling scheme.

The number M of terms tabulated in Table(6.1) along will not be sufficient for showing the computational advantage of the proposed Whittle log-likelihood estimation procedure. We will therefore compare the computational times required to perform one evaluation of our proposed Whittle log-likelihood function, with the computational times needed to calculate the traditional Gaussian log-likelihood function once. The ratios of these two computational times are contained in Table(6.2). Note that each entry of this table corresponds to the corresponding entry in Table(6.1), and the sample size is chosen to be a ressonable $n = 1000$. Even with such a modest sample size, the computational advantage of our proposed Whittle log-likelihood estimation procedure can be clearly seen from Table(6.1): it requires at most about 10% of the computational times that is needed to calculate the log-likelihood function. This computational advantage is more significant when ϕ is large or when Δ is large.

What makes the the proposed Whittle log-likelihood estimation procedure even more appealing is the fact that the number M of autocovariances required to approximate $S_Y^{\Delta}(f)$ does not change with sample size n. This in particular means that by computing the spectral density $S_Y^{\Delta}(f)$ through DFT approximations, the com-

putational costs will increase only linear as the sample size n increases, which is in sharp contrast with the almost $O(n^3)$ increase in computational costs associated with evaluation of the Gaussian log-likelihood function. This computational advantage, however, comes at a cost, in the sense that there will be some loss of information associated with our proposed Whittle log-likelihood estimation procedure. In fact, under the situations when the autocovariance sequence $C_Y(k)$ possesses relatively fast decay, the discrete time sampled process behaves more like a white noise sequence, so that the information about the covariance structure of the underlying continuous time process will be lost to a large extent. This loss-of-information problem of our proposed Whittle log-likelihood estimation procedure will be discussed in detail in next section.

$\sigma = 1, \nu = 1, \lambda_0 = 0.63$	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$	$\phi = 1$
$\Delta = 0.05$	210	126	92	74
$\Delta = 0.10$	105	63	46	37
$\Delta = 0.25$	42	36	25	18
$\Delta = 0.50$	28	19	13	10
$\Delta = 1.00$	15	10	7	6

Table 6.1: Estimated number M of terms used to approximate $S_Y^{\Delta}(f)$, for Oscillatory Matérn processes (with $\lambda_0 = 0.63$) sampled through Uniform renewal sampling schemes with different mean sampling intervals. Different decay parameters ϕ of the covariance function are considered.

$\sigma = 1, \nu = 1, \lambda_0 = 0.63$	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$	$\phi = 1$
$\Delta = 0.05$	0.1315	0.1027	0.1012	0.0981
$\Delta = 0.10$	0.1103	0.0898	0.071	0.0621
$\Delta = 0.25$	0.0936	0.0592	0.0656	0.0634
$\Delta = 0.50$	0.0612	0.0641	0.0755	0.0742
$\Delta = 1.00$	0.0698	0.0758	0.0831	0.0814

Table 6.2: Ratio of computational times of Whittle log-likelihood to exact Gaussian log-likelihood, with a modest sample size $n = 1000$, for Oscillatory Matérn process (with $\lambda_0 = 0.63$) sampled through Uniform renewal sampling schemes with different mean sampling intervals. Different decay parameters ϕ of the covariance function are considered. The spectral density $S_Y^{\Delta}(f)$ used in calculating Whittle log-likelihood is calculated through DFT approximation, with number M of terms calculated in Table(6.1)

6.3 Renewal Sampling Power Mixing Effect

Having discussed the computational aspects of the proposed Whittle log-likelihood estimation method, we now turn our attention to issues related to its finite sample performance. Through studying the spectral properties of the renewal sampling scheme in detail, we will show that the renewal sampling scheme introduces a complicated power-mixing that aggregates high frequency spectral features of $X = X(t)$. This will result in significant information loss on high frequency spectral features of $X = X(t)$. In more extreme cases when $X = X(t)$ contains only high-frequency spectral features, this power-mixing effect will cause the sampled process $Y = Y_k$ to behave like a white noise sequence. Using these observations we can explain and predict the finite sample performance of the proposed Whittle log-likelihood estimation method.

In the remaining discussions we will frequently refer to high/low-frequency features of $X = X(t)$. A frequency close to zero is regarded as low. On the other hand the concept of high-frequencies should be defined with proper reference to the sampling schemes. Under equally spaced sampling with sampling interval Δ_t, the spectral density $S_Y^{\Delta_t}(f)$ is defined (periodically) over the Nyquist frequency range $\left[-\frac{1}{2\Delta_t}, \frac{1}{2\Delta_t}\right]$. Consequently under the resolution provided by this equally spaced sampling scheme, any spectral features of $X = X(t)$ located close to the two ends or beyond this Nyquist frequency range can be regarded as high-frequency features.

Under the renewal sampling schemes, although the sampling intervals are not equal, it still makes intuitive sense to discuss the resolution of the renewal sampling scheme on an average sense. This has been implicitly done in previous discussions by introducing the average sampling interval Δ into the definition of the spectral density $S_Y^{\Delta}(f)$ of the sampled process (see Section(5.2)). As a result $S_Y^{\Delta}(f)$ is periodic with period $\frac{1}{2\Delta}$, thus defining the frequency range $\left[-\frac{1}{2\Delta}, \frac{1}{2\Delta}\right]$. This frequency range will be referred to in the remaining discussions as the equivalent Nyquist frequency range, and defines the averages resolution provided by the renewal sampling schemes. Any spectral features of $X = X(t)$ located close the two ends or beyond this equivalent Nyquist frequency range will be regarded as high-frequency features under the renewal sampling schemes.

6.3.1 Shape of the Renewal Kernel

The spectral properties of the renewal sampling scheme can be most conveniently discussed through the spectral domain expression for the spectral density $S_Y^{\Delta}(f)$ of the sampled process, which takes the following form:

$$S_Y^{\Delta}(f) = \Delta \int_{-\infty}^{\infty} S_X(\lambda) K_{\Delta}(f, \lambda) d\lambda. \tag{6.21}$$

Here the kernel function $K_{\Delta}(f, \lambda)$ is given by

$$K_{\Delta}(f, \lambda) = \mathrm{Re}\left\{ \frac{1 + e^{-i2\pi f\Delta}\phi_{\Delta}(\lambda)}{1 - e^{-i2\pi f\Delta}\phi_{\Delta}(\lambda)} \right\}. \tag{6.22}$$

Therefore the power of the sampled process $Y = Y_k$ at frequency f can be regarded as a weighted average of the power of $X = X(t)$ over all frequencies $\lambda \in \mathbb{R}$, weighted according to the renewal kernel $K_{\Delta}(f, \lambda)$. Unlike the kernel functions encountered in spectral analysis under equally spaced sampling, the shape of the renewal kernel $K_{\Delta}(f, \lambda)$ changes as the discrete time frequency variable f changes, implying a different power composition of $Y = Y_k$ for each different frequency f. Figure(6.5) shows a concrete example of the shapes of $K_{\Delta}(f, \lambda)$ as f changes, for three different sampling interval distributions, namely the Uniform, Inverse Gamma, and Gamma, with the same mean $\Delta = \frac{1}{4}$ and variance $\mathrm{Var}[\Delta] = \frac{1}{48}$ for the IID sampling intervals.

It can be observed from this figure that for each fixed f, the shape of $K_{\Delta}(f, \lambda)$ is generally not symmetric in λ around the origin. In fact this asymmetry is the most significant for $K_{\Delta}(f, \lambda)$ when f is relatively small, and then gradually becomes more symmetric as f increases, until $f = \frac{1}{2\Delta}$ when the renewal kernel becomes perfectly symmetric around the origin. When f moves to the negative range, the same kernel feature used to be in the positive λ region will then reappear in the negative λ region.

This observation, together with the symmetry of the spectral density $S_X(\lambda)$ around $\lambda = 0$, is consistent with the fact that $S_Y^{\Delta}(f)$ is symmetric around $f = 0$. Thus we shall only concentrate on the shape of $K_{\Delta}(f, \lambda)$ for $f \in \left[0, \frac{1}{2\Delta}\right]$. Also note that the renewal kernel function for uniform sampling scheme exhibits oscillatory behaviour. This is because the discontinuity on the probability density of the uniform distribution introduces oscillation into the characteristic function $\phi_{\Delta}(\lambda)$ of the IID sampling intervals (see Figure(6.1)). Apart from this difference, the renewal kernel functions for different sampling schemes share some common features: when $f > 0$, the renewal

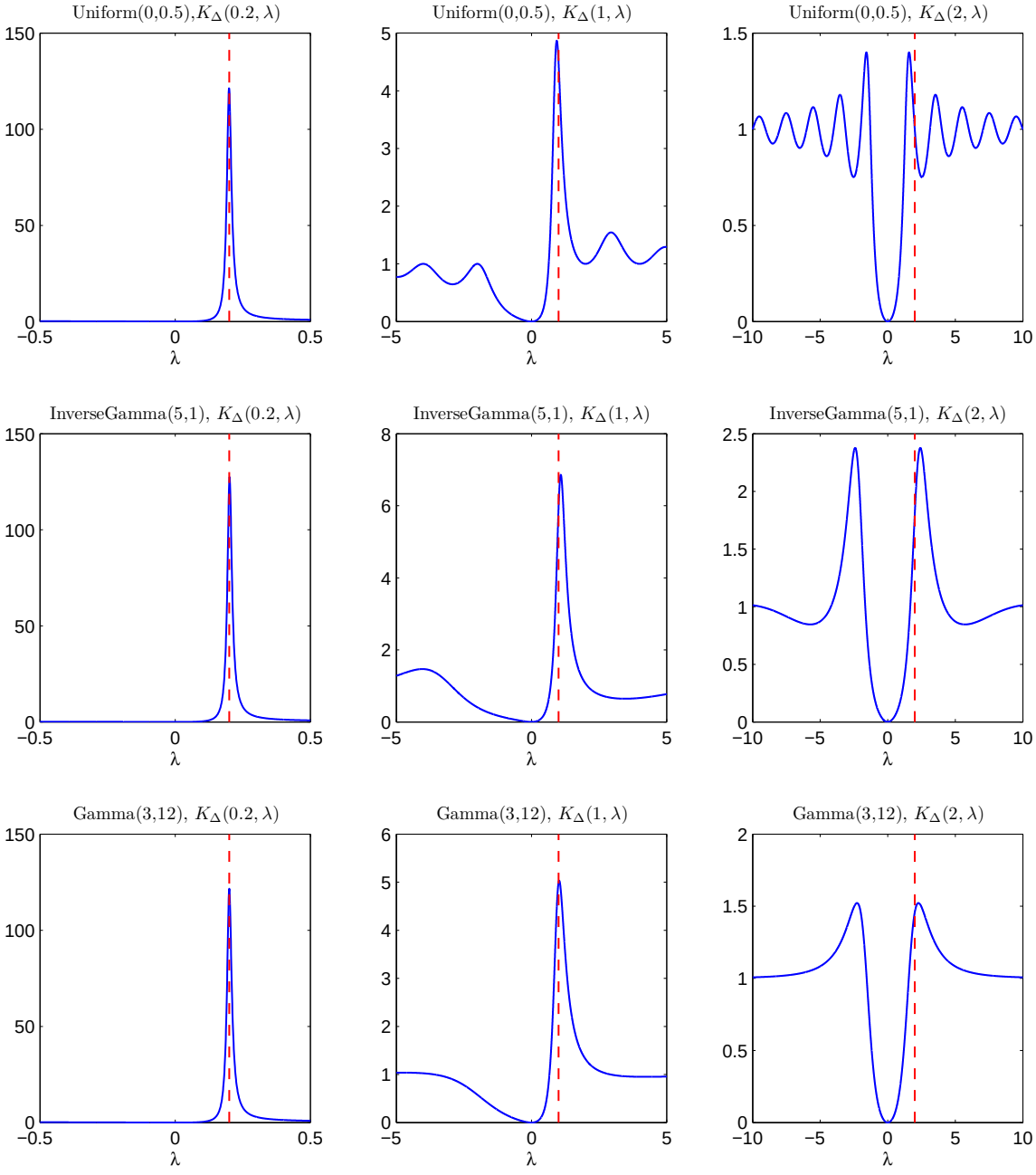

Figure 6.5: Shape of the weighting kernel $K_\Delta(f,\lambda)$, $f = 0.2, 1, 2$, for Uniform, Gamma, and Inverse Gamma sampling schemes. The red vertical line shows the position of the frequency f.

kernel $K_\Delta(f,\lambda)$ has small magnitude for λ close to zero; $K_\Delta(f,\lambda)$ demonstrates a sharp peak when f is not large; and $K_\Delta(f,\lambda)$ does not vanish as λ tends to infinity.

Renewal Kernel Peak Behaviour

The most significant feature of the renewal kernel function $K_\Delta(f,\lambda)$ is a sharp peak when the discrete time frequency variable f is not large. Observing Figure(6.5) more closely, we can see that the peak is more significant when f is close to zero, and gradually losses resolution as f increases. Due to the complexity of the kernel function, we can not offer a precise description of these observations. We will, however, try to provide some asymptotic and empirical results that provide insights on the change of the kernel peak as f increases. For this purposes, we will be focusing on the location λ_p of the kernel peak, as well as the kernel half width d. While the meaning of λ_p is self-evident, the half width d of the kernel peak is defined numerically by solving the following equation

$$K_\Delta(f,\lambda_p - d) = 1. \tag{6.23}$$

The reason for this particular choice of the kernel peak half-width d is the observation that, for a fixed $f > 0$ the kernel function $K_\Delta(f,\lambda)$ will be monotonic increasing in λ when λ is not large. Hence when $0 \leq \lambda < \lambda_p - d$ the kernel function $K_\Delta(f,\lambda)$ has a magnitude smaller than unity, suppressing the power of $X = X(t)$ in this low frequency range. This is completely different from the power modification behaviour implied by the peak of the kernel function.

Using a Taylor series based argument, we can derive the following asymptotic results for the location and half-width of the renewal kernel:

Proposition 6.2. *When $f > 0$ is not far away from zero, then we have the following approximation*

$$\lambda_p \approx f, \tag{6.24}$$

for the peak location λ_p. In this case the kernel half-width d can be approximated by

$$d \approx \frac{\lambda_p}{\frac{\Delta}{\sqrt{Var[\Delta]}} + 1} \approx \frac{f}{\frac{\Delta}{\sqrt{Var[\Delta]}} + 1}. \tag{6.25}$$

The derivation of this result can be found in Section(8.3.2) and Section(8.3.3) in the appendix of the thesis. Therefore, when f is not large, the kernel peak λ_p is close to f. In Figure(6.5) we used a red dashed vertical line to mark the frequency $\lambda = f$, and we can observe that this frequency indeed provides a good estimate for the peak location when f is not large (i.e. when $f = 0.2,\ 1$). Moreover, it is clear from the above result that the kernel-half width d increases linearly with f. How fast the kernel half-width increases is controlled by the ratio of the mean Δ and the standard deviation $\sqrt{\text{Var}[\Delta]}$ of the renewal sampling intervals. This ratio will be referred in subsequent discussions as the sampling certainty ratio, and will be denoted as r. In other words we have

$$r = \frac{\Delta}{\sqrt{\text{Var}[\Delta]}}. \tag{6.26}$$

On the other hand, as discussed in Section(5.4) the kernel function $K_\Delta(f, \lambda)$ contains a delta function component $\delta(\lambda)$ when $f = 0$. It should then be expected that, when f is close but not equal to zero, the sharp peak in the renewal kernel function should have behaviour close to a delta function. As the discrete time frequency variable f increases, however, it can be observed that the kernel peak gradually losses resolution, and hence no longer bear resemblance to the delta function. This resemblance of the kernel peak to a delta function, however, is difficult to investigate analytically. Therefore we perform numerical studies to provide empirical evidence.

We can measure this behaviour of the kernel peak by calculating a local integral $I_d(f)$ of the renewal kernel function around the exact kernel peak

$$I_d(f) = \Delta \int_{\lambda_p - d}^{\lambda_p + d} K_\Delta(f, \lambda) d\lambda, \quad \text{for } f > 0. \tag{6.27}$$

For a fixed f the exact kernel peak λ_p can be found through a numerical optimization routine over $K_\Delta(f, \lambda)$, starting from λ round the approximate peak location $\lambda = f$. It should be expected that when f is close to zero, this local integral should be close to unity.

Table(6.3) shows the values of these local integral around the exact peak location λ_p, for renewal sampling schemes with the IID sampling intervals having Uniform, Gamma, and Inverse Gamma distributions, with mean sampling interval $\Delta = 1/4$ and variance $\text{Var}[\Delta] = 1/48$. From the tabulated values, it is immediate that the when f is not large, the kernel peak located at λ_p does resemble the behaviour of

a delta function: the half width d is small, and the local integrals $I_d(f)$ is close to one. As f increases, we can see that the kernel peak will gradually demonstrate a considerably larger spread, as evidenced by the increasing magnitude of the half-width d. Consequently the kernel peak will no longer behave like a delta function, and is evidenced in the Table(6.3) by the values of the local integral being considerable different from one for f being relatively large.

	Uniform(0,0.5)		Gamma(3,12)		InvGamma(5,1)	
	d	$I_d(f)$	d	$I_d(f)$	d	$I_d(f)$
$f = 0.01$	0.00366	0.996	0.00366	0.99604	0.00366	0.99609
$f = 0.05$	0.01829	0.9793	0.0183	0.97997	0.01831	0.98126
$f = 0.10$	0.03653	0.9568	0.0366	0.95941	0.03663	0.96419
$f = 0.30$	0.10789	0.8542	0.10968	0.87506	0.11055	0.9073
$f = 0.50$	0.17419	0.7432	0.1824	0.79349	0.18564	0.86304
$f = 1.00$	0.29783	0.4921	0.36092	0.63063	0.37627	0.78757
$f = 1.50$	0.33648	0.3097	0.53397	0.53514	0.56485	0.74393
$f = 2.00$	0.2872	0.1839	0.71183	0.50044	0.74508	0.72081

Table 6.3: Half width d and local integral $I_d(f)$ of the renewal kernel for Uniform, Gamma and Inverse Gamma renewal sampling scheme.

Off-Peak Kernel Behaviour

Having discussed the behaviour of the kernel peak, we now turn our attention to the behaviour of $K_\Delta(f, \lambda)$ for λ outside the kernel peak. It can be see from Figure(6.5) that for λ close to zero, the magnitude of the kernel peak will be smaller than unity, and hence implying the power of $X = X(t)$ in this frequency range will be suppressed.

On the other hand unlike traditional sampling kernels, as λ tends to $\pm\infty$ the magnitude of $K_\Delta(f, \lambda)$ does not vanish quickly. In fact it can be easily shown that

$$\lim_{\lambda \to \pm\infty} K_\Delta(f, \lambda) = 1, \text{ for any } f. \tag{6.28}$$

This feature of the renewal kernel function means that the high frequency power of $X = X(t)$, together with the suppressed low frequency power, will be aggregated and added to the power at frequency f of $Y = Y_k$. There will be significant consequences as a result of this power aggregation effect that implied by the flat tail of the renewal kernel function, and will be discussed in the next section.

6.3.2 Implications for Parametric Estimation

It has been discussed in the last section that the special properties of the renewal kernel can be broadly described by its peak behaviour and its off-peak behaviour. In this section we will show how these special properties of the renewal kernel will affect the ability of $S_Y^{\Delta}(f)$ to capture the spectral feature of the underlying continuous time process $X = X(t)$, together with its implication for finite sample performance of the proposed Whittle log-likelihood estimation method.

Decomposing $S_Y(f)$

We can decompose $S_Y(f)$ into the following two components:

1. A local average component $E_d(f)$ implied by the kernel peak, and is defined by

$$E_d(f) = \Delta \int_{\lambda_p - d}^{\lambda_p + d} S_X(\lambda) K_{\Delta}(f, \lambda) d\lambda, \tag{6.29}$$

   where λ_p is the location of the kernel peak, and d is the half width determined by Equation(6.23). Note that this definition bears resemblance to the definition of the local integral $I_d(f)$ given by Equation(6.27).

2. An aggregation component $A(f)$ of high/low frequency power from $S_X(\lambda)$, implied by the off-peak behaviour of the renewal kernel, and will be measured by

$$A(f) = S_Y^{\Delta}(f) - E_d(f). \tag{6.30}$$

   This is simply an aggregation of $S_X(\lambda)$, weighted by $K_{\Delta}(f, \lambda)$, over the frequency range of $(-\infty, \lambda_p - d) \cup (\lambda_p + d, \infty)$.

When f is relatively small, the sharp peak of the kernel function $K_{\Delta}(f, \lambda)$ at around $\lambda = \lambda_p \approx f$ resembles a delta function. Hence the local average component $E_d(f)$ will have a significant contribution with $E_d(f) \approx S_X(\lambda_p) \approx S_X(f)$, implying that $E_d(f)$, and consequently $S_Y^{\Delta}(f)$, will contain sufficient information about the local shape of $S_X(\lambda)$ for λ around f. When f gradually increases (i.e. moving towards $f = \frac{1}{2\Delta}$), however, the kernel peak gradually becomes more diffusive and losses its resolution. Consequently $E_d(f)$ is no longer able to capture the local shape information of $S_X(\lambda)$ for λ around f. Moreover, it is pointed out in Section(8.3.4) in

the appendix that, under some circumstances the renewal kernel could disappear for relatively large f. Consequently the component $E_d(f)$ may not even exist.

On the other hand, the component $A(f)$ contains both the low/high frequency power of $X = X(t)$, off the frequency range of the kernel peak. The low-frequency power will be suppressed due to the magnitude of the kernel being smaller than unity, and the high-frequency feature will be aggregated due to the flat behaviour of the kernel function. As a consequence the component $A(f)$ should not contain much spectral shape information of the underlying continuous time process over the off-peak frequency range.

Renewal Sampling Aggregation and Whitening Effect

It is interesting to compare the above decomposition under renewal sampling with the aliasing-effect introduced from equally spaced sampling schemes with sampling interval Δ_t such that $\Delta_t = \Delta = \mathrm{E}[\Delta_t]$. According to Equation(4.10) in Section(4.2), the spectral density $S_Y^{\Delta_t}(f)$ of the discrete time sampled process under equally spaced sampling can be written as

$$S_Y^{\Delta_t}(f) = S_X(f) + \text{Aliasing Component}. \tag{6.31}$$

Note that the component $S_X(f)$ of $S_Y^{\Delta_t}(f)$ in the above equation is equivalent to the component $E_d(f)$ of $S_Y^{\Delta}(f)$ under renewal sampling scheme. However the resolution of the component $S_X(f)$ of $S_Y^{\Delta_t}(f)$ will not deteriorate when f increases, as in the case of renewal sampling schemes. Hence it should naturally be expected that the finite sample obtained from equally spaced sampling scheme should contain more information about underlying continuous time process $X = X(t)$. This advantage, however, comes from the fact that the aliasing component will contaminate $S_Y^{\Delta_t}(f)$. In fact, as discussed in Section(4.2), it is not possible to identify whether the captured low-frequency feature is genuine or not, since it could be caused by a contamination of the high-frequency spectral features outside the Nyquist frequency range $\left[-\frac{1}{2\Delta_t}, \frac{1}{2\Delta_t}\right]$.

On the other hand, the same problem will not happen under renewal sampling scheme. When $X = X(t)$ contains significant spectral features outside the equivalent Nyquist frequency range $\left[-\frac{1}{2\Delta}, \frac{1}{2\Delta}\right]$, as discussed earlier these spectral features will be mostly contained in the aggregation component $A(f)$. Because the renewal kernel $K_\Delta(f, \lambda)$ will be relatively flat for λ outside the equivalent Nyquist frequency range,

the contribution to $A(f)$ from these high frequency spectral features will be more or less the same for all f. Therefore these high frequency spectral information is lost due to the aggregation effect of the renewal kernel function.

In particular if most of the spectral features of $X = X(t)$ is contained outside the equivalent Nyquist frequency range, the lack of variation of $A(f)$ as f varies implies that $S_Y^{\Delta}(f)$ will be rather flat, and therefore the sampled process $Y = Y_k$ tend to behave like a white noise sequence. This phenomenon will be termed as the renewal sampling whitening effect. It has been discovered from the discussion of the decay of $C_Y(k)$ in Section(6.2.2), in which we showed that the oscillation in $R_X(\tau)$ introduces an extra damping factor into $C_Y(k)$, thus causing faster decay in its magnitude. Because the high frequency spectral features of $S_X(\lambda)$ is usually related to fast oscillation of $R_X(\tau)$, the discussions presented in this section can be regarded as an alternative (and more general) spectral domain explanation of this special property of the renewal sampling scheme. A slightly more rigorous derivation of the renewal sampling whitening effect can be found in Section(8.3.5) in the appendix of this thesis.

Although the renewal sampling whitening effect destroys the high-frequency spectral information through aggregation, we want to stress that it also gives the renewal sampling scheme an edge by alleviating the source of model ambiguity under the equally spaced sampling scheme. In other words, if we detect some low-frequency features of $X = X(t)$ under renewal sampling schemes, we will be assured that these low-frequency spectral features will most likely be genuine. Hence the renewal sampling schemes should help to reduce the model identification issue in the parametric estimation problem.

Graphical Illustration I - Shape of $S_Y^{\Delta}(f)$

As a demonstration of the the above discussion, let us consider the the family of Oscillatory Matérn process (see Section(2.3.3)) with fixed parameters $\sigma = 1, \phi = 0.5, \nu = 1$, and a set of different values for the peak location parameter $\lambda_0 = 0.5$, 1, 1.5, 2, 2.5, 3. Note that as λ_0 increases, the spectral feature (in this case a spectral peak) of $S_X(\lambda)$ will move to higher frequencies. Suppose this family of processes is sampled through a renewal sampling scheme with Gamma$(3, 12)$ distribution. Figure(6.6) shows the decomposition of $S_Y^{\Delta}(f)$ in terms of its components $E_d(f)$ and $A(f)$ as λ_0 changes.

When the spectral peak of the Oscillatory Matérn process is located at relatively

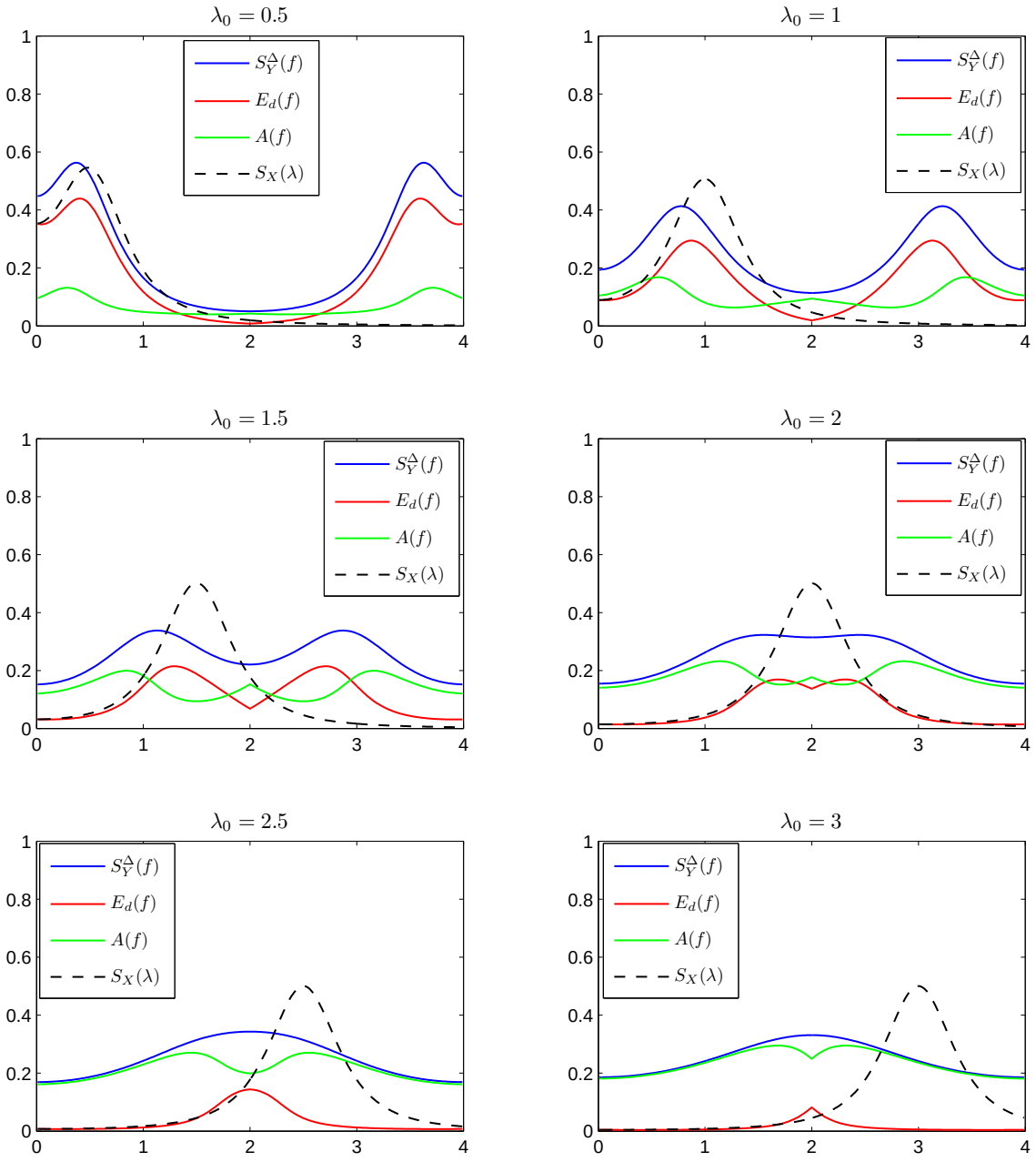

Figure 6.6: Composition of $S_Y^{\Delta}(f)$ for Oscillatory Matérn process, with spectral peak at different locations, and sampled by a renewal process with Gamma$(3, 12)$ distribution.

low frequencies (e.g. $\lambda_0 = 0.5, 1$), they are successfully captured by the component $E_d(f)$ (and consequently by $S_Y^{\Delta}(f)$) which tracks the shape of $S_X(\lambda)$ closely. On the other hand, as the spectral peak of the Oscillatory Matérn process moves to higher frequencies (e.g. $\lambda_0 = 1.5, 2$) where the kernel peak no longer has resolution, then the component $E_d(f)$ (and hence $S_Y^{\Delta}(f)$) is no longer able to capture the shape of $S_X(f)$ closely, hence indicating a loss of information in this case. When the spectral peak of the Oscillatory Matérn process moves to even higher frequencies (e.g. $\lambda_0 = 2.5, 3$) where the kernel function is almost flat, most of the power of the underlying process will be aggregated to give the predominant component $A(f)$, which will then contribute to $S_Y^{\Delta}(f)$ at every frequency by more or less the same amount. Consequently in this last case the spectral density $S_Y^{\Delta}(f)$ will tend to be rather flat across all frequencies, suggesting that the high-frequency spectral information has been destroyed, and the discrete time sampled process will behave like a white noise sequence.

Graphical Illustration II - Fisher Information

This loss of information under renewal sampling can also be more explicitly seen by considering the Fisher information [62], which for any scalar parameter θ is given by

$$FI(\theta) = \Delta \int_0^{1/\Delta} \frac{\frac{\partial}{\partial \theta} S_Y^{\Delta}(f;\theta)^2}{S_Y^{\Delta}(f;\theta)^2} df. \tag{6.32}$$

Note that this Fisher information defined above for a scalar parameter θ is simply the diagonal entry of the Fisher information matrix defined in Equation(5.76). The reason for considering $FI(\theta)$ is simple: assuming θ_0 is the true parameter value, then $FI(\theta_0)$ measures the curvature of the function $AWLL(\theta)$ at the true parameter value $\theta = \theta_0$ (see for example Kay [62]). A large curvature implies that the function $AWLL(\theta)$ has a relatively sharp minimum at $\theta = \theta_0$, and consequently could be captured by the Whittle estimator with relatively low variability. A low curvature on the other hand implies that the minimum at $\theta = \theta_0$ is not very distinctive, hence making it more difficult for the estimation procedure to capture the true parameter value. Using the family of Oscillatory Matérn process considered previously in Figure(6.6), we plot the Fisher information for the parameters ϕ, ν, and λ_0 against varying λ_0 in Figure(6.7). Moreover as a comparison we also included in the figure the the Fisher information under equally spaced sampling scheme.

First of all it can be observed that as λ_0 increases, the Fisher information for ϕ, ν, λ_0 decreases. This is consistent with the fact that as the spectral features of the underlying process moves to higher frequencies, the renewal sampling schemes suffers more severe information loss and the sampled process behaves more like a white noise sequence. Also can be observed is the fact that the Fisher information for ϕ, ν, λ_0 under equally spaced sampling scheme is generally higher than under renewal sampling scheme, clearly indicating the previously discussed point that the renewal sampling schemes generally provide less resolution than the equally spaced sampling schemes.

Finally note that we did not consider the Fisher information for the parameter σ. This is because $FI(\sigma)$ behaves rather differently. In fact a simple calculation gives that

$$FI(\sigma) = \Delta \int_0^{1/\Delta} \frac{\frac{\partial}{\partial \sigma} S_Y^{\Delta}(f;\sigma)^2}{S_Y^{\Delta}(f;\sigma)^2} df = \frac{4}{\sigma^2}, \tag{6.33}$$

hence showing that $FI(\sigma)$ is neither dependent on other parameters, nor on the sampling schemes. Therefore we should expect an almost constant accuracy for estimating σ through our proposed Whittle log-likelihood estimation procedure.

6.3.3 Changing Mean and Variance of Sampling Intervals

The discussion in previous section assumes a fixed renewal sampling scheme, and showed that when the spectral features of $X = X(t)$ move to higher frequencies, significant loss of information will occur due to the whitening effect of the renewal sampling scheme. The same issue can also be discussed in terms of fixing the underlying process $X = X(t)$ and letting the renewal sampling schemes vary. In particular note that the average sampling interval Δ of the renewal sampling scheme determines the equivalent Nyquist frequency range of the discrete time sampled process $Y = Y_k$. Consequently a smaller average sampling interval Δ will allow $Y = Y_k$ to cover a wider frequency range, and hence has the potential ability to capture the spectral feature of $X = X(t)$ at relatively higher frequencies.

Once an average sampling interval Δ is determined and fixed, whether or not the spectral features of $X = X(t)$ can be effectively captured within the frequency range of $Y = Y_k$ is determined by the resolution of the renewal kernel $K_{\Delta}(f, \lambda)$, which as shown in Section(6.3.1) is measured by the half-width d of the kernel peak. Ac-

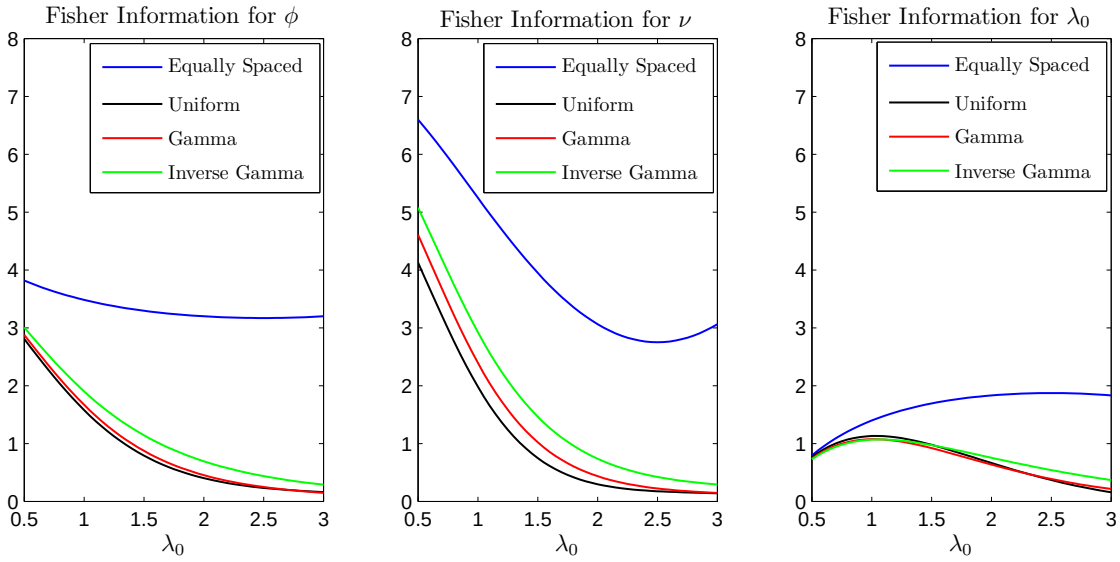

Figure 6.7: Composition of $S_Y^{\Delta}(f)$ for Oscillatory Matérn process, with spectral peak at different locations, and sampled by a renewal process with Gamma$(3, 12)$ distribution.

cording to Proposition(6.2) in Section(6.3.1), the half-width d can be asymptotically approximated by

$$d \approx \frac{\lambda_p}{\frac{\Delta}{\sqrt{\text{Var}[\Delta]}} + 1} \approx \frac{f}{\frac{\Delta}{\sqrt{\text{Var}[\Delta]}} + 1}. \tag{6.34}$$

Consequently, the variance $\text{Var}[\Delta]$ of the IID sampling interval affects the resolution of the renewal kernel through what we call the sampling certainty ratio $r = \frac{\Delta}{\sqrt{\text{Var}[\Delta]}}$. When the sampling scheme is random, the certainty ratio r is positive. This will then cause information loss because in this case (i.e. when $r > 0$) the half-width d of the kernel peak will increase linearly as f increases. Reducing $\text{Var}[\Delta]$ will have the effect of letting the half-width d decrease at a slower pace, hence reducing the extent of loss of information caused by renewal sampling scheme.

The discussion in this section is illustrated, through the plot of Fisher information, using an Oscillatory Matérn process with parameters given by $\sigma = 1, \phi = 0.5, \nu = 1, \lambda_0 = 1$. Such a process is sampled through different renewal sampling schemes with IID sampling intervals having Uniform, Gamma, and Inverse Gamma distributions.

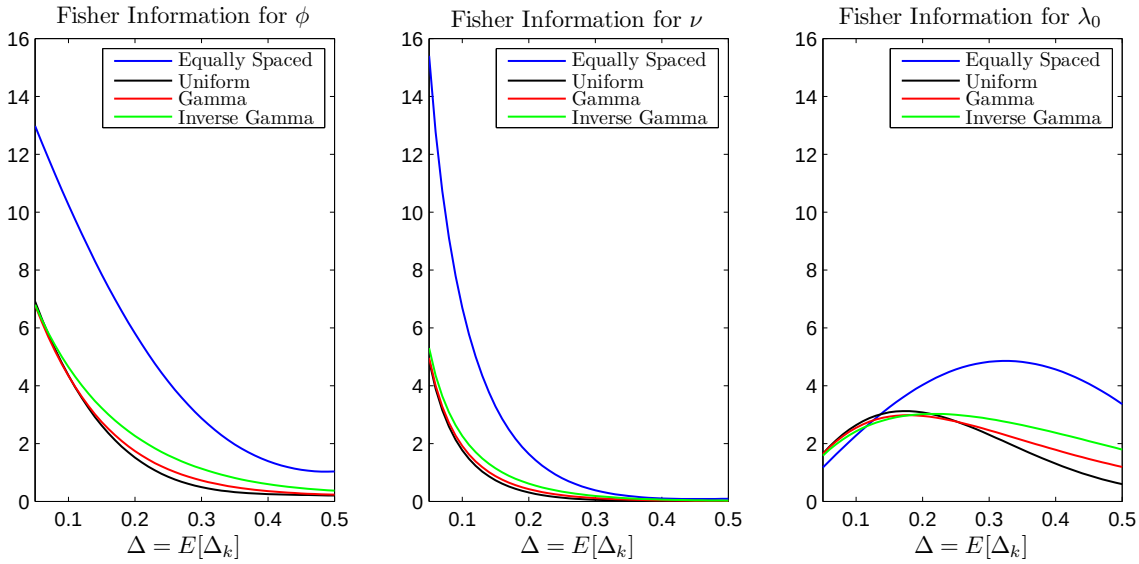

Figure 6.8: Fisher information for ϕ, ν, λ_0 as average sampling interval Δ changes. The sampling certainty ratio is fixed at $r \approx 1.73$.

For comparison purposes we also included the equally spaced sampling scheme with the sampling interval $\Delta_t = \Delta = \mathrm{E}[\Delta_k]$.

Figure(6.8) shows how the Fisher information for parameters ϕ, ν, λ_0 changes as average sampling interval Δ (and hence Δ_t) changes, while keeping the sampling certainty ratio fixed at $r \approx 1.73$ (which is the implicit ratio for all Uniform distributions). Similarly Figure(6.9) shows the Fisher information curves as the sampling certainty ratio r changes, while keeping the average sampling interval at $\Delta = 0.1$. Note that the Uniform sampling interval distribution is not considered in Figure(6.9) for the reason that the its sampling certainty ratio is not variable.

The patterns of the Fisher information demonstrated in both figures are broadly consistent with the discussions given in this section. As Δ decreases, the Fisher information will gradually increase, indicating that it will be easier for the sampled process to capture the spectral information of the underlying continuous time process. On the other hand as the sampling certainty ratio r increases (i.e. the sampling intervals are becoming less random), the Fisher information under renewal sampling gradually increases and approaches the Fisher information under equally spaced sampling

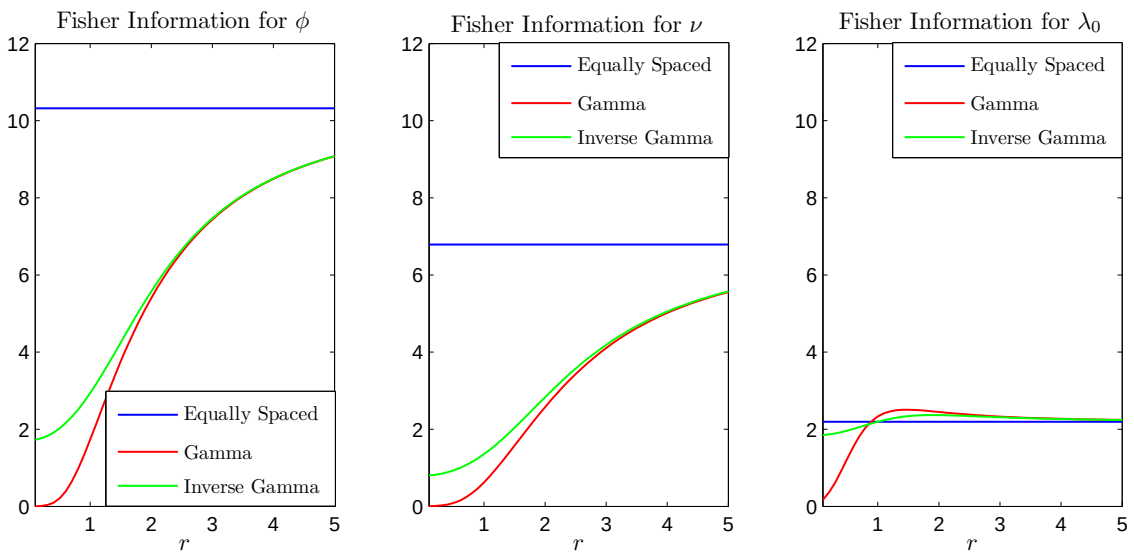

Figure 6.9: Fisher information for ϕ, ν, λ_0 as sampling certainty ratio r changes. The average sampling interval is fixed at $\Delta = 0.1$.

schemes. It is also interesting to note that under some circumstances the renewal sampling scheme gives higher Fisher information for the spectral location parameter λ_0, hence is helpful in its estimation. However, this higher Fisher information will not in general translate into a smaller mean-square error for estimating λ_0 through the proposed Whittle log-likelihood function, because the non-Gaussianity of the sampled process will also contribute to the finite sample variability.

This therefore completes the discussions of the renewal sampling power-mixing effect and its implications for the proposed Whittle log-likelihood parametric estimation method. The next section will be devoted to a set of simulation studies, which can be regarded as companion to the discussions provided in this section, showing that the finite sample performance of the proposed Whittle estimator follows the same pattens as mentioned in this section.

6.4 Simulation Studies

In this section we present a set of simulation studies, showing the finite sample performance of the Whittle log-likelihood estimator $\hat{\boldsymbol{\theta}}_n$. We will be considering the prob-

lem of estimating parameters $\sigma, \phi, \nu, \lambda_0$ from an Oscillatory Matérn process, under different scenarios corresponding to the discussions made in previous sections. The finite sample performance of the Whittle log-likelihood estimator $\hat{\boldsymbol{\theta}}_n$ will be assessed by calculating biases and root mean square errors (rmse) across 500 independent realizations, each contains $n = 5000$ observations. Such a sample size would be problematic under traditional approach of parameter estimation through exact Gaussian log-likelihood function, due to the excessive computational costs involved in inverting the large covariance matrix. Using our proposed estimation method, on the other hand, the computational costs becomes manageable: in fact with randomized starting values, the non-linear minimizer in most cases converges in roughly 20 seconds. The renewal sampling schemes considered in the simulation studies will have IID sampling intervals with Uniform, Gamma and Inverse Gamma distributions. We implement our proposed estimation procedure in Matlab, using the build-in unconstrained optimizer fminunc().

6.4.1 Varying the Spectral Location Parameter

In this subsection we consider the scenario similar to the one in Figure(6.7). We will show the finite sample performance of our proposed Whittle log-likelihood estimation method for the class of Oscillatory Matérn processes with fixed parameter values $\sigma = 1$, $\phi = 0.5$, $\nu = 1$, and variable spectral location parameter $\lambda_0 = 0.5$, 1, 3. The renewal sampling schemes under consideration have Uniform, Gamma and Inverse Gamma distributions for the IID sampling intervals. The mean sampling interval will be $\Delta = 0.1$, and the sampling certainty ratio will be fixed at $r \approx 1.73$, which is implicitly implied by the Uniform sampling intervals. Table(6.4) reports the performance in terms of bias, standard deviation (std) and root-mean-square-error (rmse).

We immediately see from the table that the performance of our proposed Whittle log-likelihood estimator $\hat{\boldsymbol{\theta}}_n$ shows a pattern that is consistent with the discussions given in Section(6.3.2): when the spectral feature of the underlying process is located at low frequencies (i.e. when λ_0 is relatively small) the Whittle estimator $\hat{\boldsymbol{\theta}}_n$ delivers good finite sample performance; on the other hand as the spectral feature moves to higher frequencies (i.e. as λ_0 increases), the accuracy of $\hat{\boldsymbol{\theta}}_n$ deteriorates quickly, especially for ϕ and ν. The almost constant performance for estimating σ can be explained by Equation(6.33) that the Fisher information for σ depends only on σ

OsMatérn, $n = 5000$ $\Delta = 0.1,\ r = 1.73$		$\lambda_0 = 0.5$			$\lambda_0 = 1$			$\lambda_0 = 3$		
		bias	std	rmse	bias	std	rmse	bias	std	rmse
Uniform	σ	0.003	0.005	0.006	-0.001	0.007	0.007	0.001	0.006	0.006
	ϕ	0.005	0.029	0.029	0.013	0.040	0.042	0.002	0.043	0.043
	ν	0.004	0.043	0.043	0.015	0.064	0.066	-0.003	0.156	0.156
	λ_0	-0.004	0.011	0.012	-0.002	0.031	0.031	0.006	0.068	0.068
Gamma	σ	0.002	0.032	0.032	0.002	0.019	0.018	0.018	0.019	0.019
	ϕ	0.002	0.029	0.029	-0.005	0.035	0.045	0.045	0.065	0.066
	ν	0.003	0.045	0.045	-0.007	0.054	0.063	0.064	0.170	0.173
	λ_0	0.006	0.022	0.023	-0.002	0.052	0.042	0.042	0.071	0.071
InvGamma	λ	-0.002	0.018	0.018	0.003	0.020	0.020	0.002	0.018	0.018
	ϕ	0.003	0.048	0.048	-0.001	0.041	0.041	-0.006	0.069	0.069
	ν	0.005	0.037	0.037	0.013	0.054	0.056	0.015	0.132	0.133
	λ_0	0.003	0.034	0.034	0.000	0.048	0.048	0.008	0.053	0.054

Table 6.4: Comparison of the performance of the Whittle log-likelihood estimator, under Uniform, Gamma and Inverse Gamma sampling schemes, for the family of Oscillatory Matérn process indexed by the spectral location parameter λ_0. The sampling schemes have fixed $\Delta = 0.1$ and $r \approx 1.73$.

itself.

Moreover, it seems that the Inverse Gamma renewal sampling schemes deliver slightly better finite sample performance. This perhaps can be explained by a closer look at Table(6.3), in which it can be observed that under Inverse Gamma renewal sampling schemes, the peak of the renewal kernel losses its resolution slightly more slowly as compared with the Uniform and Gamma renewal sampling schemes. This better property of the peak of the Inverse Gamma renewal kernel then gives slightly higher Fisher information for almost all the parameters, as evidenced from Figure(6.8)-(6.9).

6.4.2 Varying Average Sampling Interval and Sampling Certainty Ratio

In this set of simulation studies, we consider the finite sample performance under the scenarios of different combinations of average sampling interval Δ and sampling certainty ratio $r = \frac{\Delta}{\sqrt{\text{Var}[\Delta]}}$. We will be considering the Oscillatory Matérn process with true parameter values $\sigma = 1$, $\phi = 0.5$, $\nu = 1$, and $\lambda_0 = 1$. Such a process is sampled with renewal sampling schemes with IID sampling intervals having Gamma

and Inverse Gamma distributions. Here we the Uniform sampling interval is not considered for the reason that it has a fixed sampling certainty ratio $r = \sqrt{3} \approx 1.73$ and hence does not fit into the simulation scenario. The finite sample performance based on 500 independent realizations of 5000 observations are reported in Table(6.5)-(6.7).

The performance figure reported in these tables are consistent with the discussions made in Section(6.3.3). In particular, it can be seen by comparing these tables that as the mean sampling interval Δ increases, the finite sample performance of the Whittle log-likelihood estimator $\hat{\boldsymbol{\theta}}_n$ generally deteriorates, especially for the parameters ϕ and ν. On the other hand when the average sampling interval Δ is fixed, the performance of $\hat{\boldsymbol{\theta}}_n$ improves as the sampling certainty ratio r increases. Again as observed and explained in the Section(6.4.1), the finite sample performance seems to be slightly better under Inverse Gamma renewal sampling schemes.

OsMatérn, $n = 5000$		$r = 2$			$r = 1.5$			$r = 1$		
$\Delta = 0.05$		bias	std	rmse	bias	std	rmse	bias	std	rmse
Gamma	σ	-0.001	0.024	0.024	0.001	0.025	0.025	0.001	0.025	0.025
	ϕ	0.009	0.039	0.039	-0.006	0.042	0.042	-0.022	0.045	0.049
	ν	0.008	0.032	0.033	0.004	0.046	0.046	-0.036	0.049	0.061
	λ_0	-0.004	0.017	0.017	0.002	0.016	0.017	0.000	0.023	0.023
InvGamma	σ	0.001	0.006	0.007	0.003	0.008	0.009	0.002	0.011	0.011
	ϕ	0.007	0.024	0.025	0.011	0.029	0.031	0.003	0.037	0.038
	ν	0.007	0.033	0.033	0.010	0.040	0.042	0.002	0.058	0.059
	λ_0	-0.006	0.017	0.017	-0.005	0.019	0.021	-0.007	0.021	0.022

Table 6.5: Comparison of the performance of the Whittle log-likelihood estimator, under Gamma and Inverse Gamma sampling schemes. The mean sampling interval is fixed at $\Delta = 0.05$, while the sampling times non-randomness ratio $r = \frac{\Delta}{\sqrt{\text{Var}[\Delta]}}$ varies.

6.4.3 Misspecifying Renewal Sampling Schemes

In practice, the implementation of our proposed Whittle log-likelihood estimation method depends on knowing the renewal sampling scheme that is involved in the data generation process. In other words we have to specify the family of IID sampling interval distributions, and then choose the appropriate parameters. While the parameters of the IID sampling distributions can be estimated from the observed sampling instances, the family of IID sampling interval distribution usually has to be

OsMatérn, $n = 5000$		$r = 2$			$r = 1.5$			$r = 1$		
$\Delta = 0.1$		bias	std	rmse	bias	std	rmse	bias	std	rmse
Gamma	σ	-0.005	0.019	0.020	0.001	0.018	0.018	0.001	0.024	0.024
	ϕ	-0.009	0.036	0.037	-0.005	0.037	0.038	0.004	0.080	0.081
	ν	-0.016	0.044	0.047	-0.004	0.057	0.058	-0.007	0.138	0.139
	λ_0	0.001	0.021	0.021	0.001	0.028	0.028	-0.007	0.034	0.035
InvGamma	σ	0.002	0.005	0.005	0.003	0.005	0.006	0.003	0.008	0.008
	ϕ	0.002	0.030	0.030	0.014	0.033	0.036	0.008	0.050	0.050
	ν	0.006	0.043	0.044	0.020	0.056	0.059	0.024	0.087	0.091
	λ_0	-0.003	0.007	0.007	-0.002	0.022	0.023	-0.001	0.028	0.028

Table 6.6: Comparison of the performance of the Whittle log-likelihood estimator, under Gamma and Inverse Gamma sampling schemes. The mean sampling interval is fixed at $\Delta = 0.1$, while the sampling times non-randomness ratio $r = \frac{\Delta}{\sqrt{\text{Var}[\Delta]}}$ varies.

OsMatérn, $n = 5000$		$r = 2$			$r = 1.5$			$r = 1$		
$\Delta = 0.2$		bias	std	rmse	bias	std	rmse	bias	std	rmse
Gamma	σ	-0.001	0.014	0.014	-0.001	0.015	0.015	-0.002	0.016	0.017
	ϕ	0.001	0.048	0.048	-0.025	0.067	0.071	-0.019	0.079	0.081
	ν	0.005	0.102	0.102	-0.030	0.143	0.148	-0.033	0.175	0.178
	λ_0	0.000	0.023	0.023	-0.003	0.019	0.020	-0.008	0.031	0.032
InvGamma	σ	0.001	0.006	0.006	0.003	0.006	0.007	0.001	0.007	0.007
	ϕ	-0.002	0.042	0.042	-0.007	0.041	0.041	0.011	0.059	0.061
	ν	-0.004	0.061	0.063	-0.008	0.080	0.081	0.031	0.126	0.130
	λ_0	-0.004	0.016	0.016	-0.001	0.021	0.021	-0.003	0.025	0.025

Table 6.7: Comparison of the performance of the Whittle log-likelihood estimator, under Gamma and Inverse Gamma sampling schemes. The mean sampling interval is fixed at $\Delta = 0.2$, while the sampling times non-randomness ratio $r = \frac{\Delta}{\sqrt{\text{Var}[\Delta]}}$ varies.

decided subjectively, and therefore may subject to mistakes. In this section we will, through simulation studies, illustrate the consequences of misspecifying the underlying renewal sampling schemes.

Specifically we will be considering combinations of renewal sampling schemes with sampling intervals having Uniform, Gamma and Inverse Gamma distributions. For each considered sampling scheme responsible for generating finite samples, we will input other (misspecified) sampling schemes into our proposed Whittle log-likelihood estimation procedure. The parameters of the misspecified input sampling schemes are estimated by matching the mean and variance of the observed sampling intervals. The continuous time Gaussian process under consideration is still the Oscillatory Matérn process with true parameters given by $\sigma = 1, \phi = 0.5, \nu = 1, \lambda_0 = 1$.

Table(6.8), (6.9) and (6.10) contain the results of misspecifying the renewal sampling schemes, with varying average sampling interval Δ while keeping the sampling certainty ratio fixed at $r = \sqrt{3} \approx 1.73$. On the other hand Table(6.11) and (6.12) consider the scenario with fixed average sampling interval and varying sampling certainty ratios. These tables are structured into 9-by-9 blocks, each representing one of the possible combinations of the true/input sampling schemes. In particular the diagonal blocks representing cases of no misspecification of sampling schemes.

Observing these tables closely, we have observed the following patterns as a consequence of misspecifying sampling schemes:

- It seems that the smoothness parameter ν and the covariance decay parameter ϕ will be the most significantly affected by misspecifying the sampling schemes, in terms of potentially producing a large bias when the wrong sampling schemes are used in the estimation procedure;

- It seems that under the circumstances when the renewal sampling schemes can deliver relatively good performance (i.e. when Δ is small or when r is large), the Whittle log-likelihood estimator $\hat{\boldsymbol{\theta}}_n$ is relatively robust with respect to misspecification of the sampling schemes;

- When Δ increases or when r decreases, the Whittle log-likelihood estimator $\hat{\boldsymbol{\theta}}_n$ becomes less reliable with respect to misspecification of sampling schemes, and the non-linear optimization routine may even fail to converge in some circumstances.

When the average sampling interval Δ is small, the sampled process $Y = Y_k$ will have a wider frequency range. Consequently more spectral features of the underlying continuous time process will be concentrated at low frequency range of $Y = Y_k$. Moreover note that according to Table(6.3) and the approximating argument in Section(8.3.2) and (8.3.3) of the appendix the kernel peak (which provides resolution) will behave like a delta function in the low frequency range of $Y = Y_k$, regardless of the IID sampling interval distributions. Consequently it should be expected that as Δ becomes smaller, different renewal sampling schemes will tend to provide more or less the same resolution, and hence implying robustness in terms of misspecification of sampling schemes. This explains why the biases introduced from misspecification of sampling schemes tend to be smaller when Δ decreases.

On the other hand, when the sampling certainty ratio r increases, the sampling times will tend to be less uncertain. Hence different renewal sampling schemes will have similar behaviour in the sense that they are all close to the equally spaced sampling scheme. This therefore explains the observation that the biases introduced from misspecification of sampling schemes tend to be smaller when r increases. In other situations when Δ is not relatively small and r is not relatively large, however, correctly specifying the input renewal sampling schemes may become vital in producing consistent estimators using our proposed Whittle log-likelihood estimation method.

6.4.4 Confidence Intervals

Having investigated the finite sample performance in terms of biases and rmse, we now turn to the problem of constructing confidence intervals for the proposed Whittle log-likelihood estimator $\hat{\boldsymbol{\theta}}_n$. In Section(5.6) we have explored the asymptotic normality of the estimator $\hat{\boldsymbol{\theta}}_n$. Although rigorous proof can not be provided, we conducted simulation studies to show that the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ can be reasonably assumed when the underlying continuous time process is Gaussian. Consequently it is reasonable to expect from Equation(5.79) that when the sample size is relatively large the estimated parameter $\hat{\boldsymbol{\theta}}_n$ has the following asymptotic normal distribution:

$$\hat{\boldsymbol{\theta}}_n \overset{d}{\approx} \mathcal{N}(\boldsymbol{\theta}_0, \hat{\Sigma}_n(\hat{\boldsymbol{\theta}}_n)). \tag{6.35}$$

OsMatérn, $n = 5000$			Input Sampling Distributions								
			Uniform			Gamma			InvGamma		
$\Delta = 0.05, r = 1.73$			bias	std	rmse	bias	std	rmse	bias	std	rmse
Sampling Distribution	Uniform	σ	0.004	0.009	0.009	0.003	0.007	0.008	0.005	0.009	0.010
		ϕ	0.007	0.027	0.027	-0.014	0.028	0.032	-0.035	0.020	0.040
		ν	0.007	0.034	0.034	-0.033	0.034	0.047	-0.073	0.022	0.076
		λ_0	-0.006	0.013	0.013	-0.005	0.012	0.013	-0.005	0.010	0.012
	Gamma	σ	0.000	0.006	0.006	-0.003	0.028	0.028	0.003	0.006	0.007
		ϕ	0.021	0.024	0.032	0.003	0.038	0.038	-0.007	0.024	0.025
		ν	0.039	0.035	0.052	0.001	0.040	0.040	-0.029	0.027	0.039
		λ_0	-0.005	0.011	0.012	0.004	0.023	0.023	0.006	0.008	0.010
	InvGamma	σ	-0.001	0.007	0.007	0.001	0.007	0.007	0.004	0.028	0.028
		ϕ	0.040	0.026	0.048	0.019	0.027	0.033	-0.001	0.038	0.038
		ν	0.088	0.034	0.095	0.043	0.039	0.058	0.006	0.047	0.048
		λ_0	-0.003	0.010	0.011	-0.003	0.011	0.011	0.000	0.024	0.024

Table 6.8: Performence of Whittle log-likelihood estimators, under misspecifying of sampling schemes. The sampling schemes under consideration all have mean sampling interval $\Delta = \mathrm{E}[\Delta_k] = 0.05$ and sampling uncertainty ratio $r = \frac{\Delta}{\sqrt{\mathrm{Var}[\Delta]}} \approx 1.73$.

OsMatérn, $n = 5000$			Input Sampling Distributions								
			Uniform			Gamma			InvGamma		
$\Delta = 0.1, r = 1.73$			bias	std	rmse	bias	std	rmse	bias	std	rmse
True Sampling Distributions	Uniform	σ	0.001	0.007	0.007	0.002	0.005	0.006	0.002	0.007	0.007
		ϕ	0.004	0.028	0.029	0.007	0.034	0.035	-0.029	0.018	0.035
		ν	0.008	0.051	0.051	-0.046	0.051	0.069	-0.121	0.032	0.125
		λ_0	-0.002	0.011	0.011	0.002	0.009	0.009	0.007	0.011	0.013
	Gamma	σ	0.001	0.006	0.006	0.002	0.018	0.018	0.000	0.007	0.0073
		ϕ	0.024	0.041	0.048	-0.005	0.045	0.045	-0.006	0.029	0.03
		ν	0.095	0.082	0.126	-0.007	0.064	0.064	-0.056	0.045	0.071
		λ_0	-0.001	0.01	0.011	0.002	0.018	0.018	-0.006	0.009	0.011
	InvGamma	σ	0.001	0.005	0.005	0.001	0.005	0.006	-0.001	0.021	0.021
		ϕ	-0.005	0.035	0.04	0.025	0.03	0.04	-0.003	0.041	0.041
		ν	0.175	0.055	0.183	0.089	0.056	0.105	0.006	0.063	0.063
		λ_0	0.003	0.009	0.009	0.000	0.009	0.01	0.006	0.016	0.017

Table 6.9: Performence of Whittle log-likelihood estimators, under misspecifying of sampling schemes. The sampling schemes under consideration all have mean sampling interval $\Delta = \mathrm{E}[\Delta_k] = 0.1$ and sampling uncertainty ratio $r = \frac{\Delta}{\sqrt{\mathrm{Var}[\Delta]}} \approx 1.73$.

OsMatérn, $n = 5000$			Input Sampling Distributions								
			Uniform			Gamma			InvGamma		
$\Delta = 0.2$,$r = 1.73$			bias	std	rmse	bias	std	rmse	bias	std	rmse
True Sampling Distribution	Uniform	σ	0.001	0.015	0.016	-0.001	0.008	0.008	-0.001	0.007	0.007
		ϕ	-0.014	0.070	0.071	-0.009	0.055	0.055	-0.012	0.039	0.041
		ν	-0.022	0.170	0.172	-0.133	0.101	0.167	-0.205	0.062	0.215
		λ_0	-0.004	0.018	0.018	0.001	0.015	0.015	-0.018	0.012	0.021
	Gamma	σ	0.000	0.007	0.007	-0.001	0.014	0.014	0.004	0.015	0.015
		ϕ	0.011	0.053	0.054	0.022	0.070	0.074	-0.008	0.065	0.065
		ν	0.189	0.170	0.254	0.043	0.150	0.157	-0.090	0.113	0.144
		λ_0	-0.004	0.012	0.012	0.000	0.018	0.018	-0.012	0.017	0.021
	Inv Gamma	σ	0.001	0.006	0.006	0.001	0.014	0.014	0.000	0.012	0.012
		ϕ	0.012	0.059	0.061	0.010	0.054	0.055	-0.008	0.059	0.059
		ν	0.678	0.273	0.731	0.164	0.139	0.215	-0.011	0.112	0.112
		λ_0	0.004	0.012	0.013	0.010	0.012	0.015	0.002	0.017	0.017

Table 6.10: Performence of Whittle log-likelihood estimators, under misspecifying of sampling schemes. The sampling schemes under consideration all have mean sampling interval $\Delta = \mathrm{E}[\Delta_k] = 0.2$ and sampling uncertainty ratio $r = \frac{\Delta}{\sqrt{\mathrm{Var}[\Delta]}} \approx 1.73$.

OsMatérn, n=5000			Input Dist: InvGamma								
			$r = 1$			$r = 1.5$			$r = 2$		
			bias	std	rmse	bias	std	rmse	bias	std	rmse
Sampling Dist.	Gamma $\mathrm{E}(\Delta) = 0.05$	σ	0.003	0.010	0.011	-0.002	0.023	0.023	0.001	0.026	0.026
		ϕ	-0.065	0.033	0.072	-0.028	0.034	0.045	-0.003	0.042	0.043
		ν	-0.196	0.035	0.199	-0.072	0.034	0.079	-0.017	0.036	0.039
		λ_0	-0.019	0.017	0.026	-0.005	0.023	0.023	0.002	0.024	0.024
	Gamma $\mathrm{E}(\Delta) = 0.1$	σ	-0.002	0.017	0.018	0.006	0.020	0.021	0.001	0.021	0.021
		ϕ	-0.065	0.044	0.078	-0.019	0.048	0.052	-0.011	0.035	0.036
		ν	-0.269	0.042	0.272	-0.099	0.049	0.110	-0.036	0.043	0.056
		λ_0	-0.046	0.022	0.051	-0.011	0.023	0.025	0.001	0.014	0.014
	Gamma $\mathrm{E}(\Delta) = 0.2$	σ	-0.001	0.017	0.017	0.004	0.016	0.016	-0.001	0.013	0.013
		ϕ	-0.046	0.058	0.074	-0.012	0.062	0.063	-0.005	0.049	0.049
		ν	-0.317	0.069	0.324	-0.152	0.102	0.183	-0.073	0.098	0.123
		λ_0	-0.107	0.021	0.109	-0.025	0.017	0.031	-0.007	0.017	0.019

Table 6.11: Performence of Whittle log-likelihood estimators, under misspecifying of sampling schemes. Gamma scheme being the underlying sampling scheme, and the Inverse Gamma scheme being the sampling scheme used in Whittle log-likelihood estimation. The performance statistics are tabulated for different sampling non-randomness ratio.

OsMatérn, n=5000			Input Dist: Gamma								
			$r=1$			$r=1.5$			$r=2$		
			bias	std	rmse	bias	std	rmse	bias	std	rmse
Sampling Dist.	InvGamma ED=0.05	σ	-0.007	0.011	0.014	0.001	0.007	0.007	0.001	0.001	0.001
		ϕ	0.148	0.073	0.165	0.032	0.029	0.044	0.023	0.020	0.031
		ν	0.427	0.203	0.473	0.079	0.052	0.094	0.040	0.030	0.051
		λ_0	0.001	0.019	0.020	-0.003	0.012	0.012	-0.003	0.011	0.012
	InvGamma ED=0.1	σ	Not Converging			0.000	0.007	0.007	0.000	0.005	0.005
		ϕ				0.052	0.034	0.063	0.016	0.028	0.032
		ν				0.194	0.092	0.215	0.048	0.050	0.069
		λ_0				0.006	0.012	0.013	-0.001	0.007	0.007
	InvGamma ED=0.2	σ	Not Converging			0.004	0.014	0.015	0.002	0.015	0.015
		ϕ				0.061	0.094	0.111	-0.002	0.057	0.057
		ν				0.460	0.349	0.577	0.075	0.121	0.142
		λ_0				0.021	0.019	0.029	0.004	0.012	0.012

Table 6.12: Performence of Whittle log-likelihood estimators, under misspecifying of sampling schemes. Inverse Gamma scheme being the underlying sampling scheme, and the Gamma scheme being the sampling scheme used in Whittle log-likelihood estimation. The performance statistics are tabulated for different sampling non-randomness ratio.

Here the covariance matrix $\hat{\Sigma}_n(\hat{\boldsymbol{\theta}}_n)$ is given by

$$\hat{\Sigma}_n(\hat{\boldsymbol{\theta}}_n) = \frac{2}{n}\Sigma_Y(\hat{\boldsymbol{\theta}}_n)^{-1} + \frac{1}{n}\Sigma_Y(\hat{\boldsymbol{\theta}}_n)^{-1}\Omega_Y(\hat{\boldsymbol{\theta}}_n)\Sigma_Y(\hat{\boldsymbol{\theta}}_n)^{-1}, \tag{6.36}$$

and the matrices $\Sigma_Y(\hat{\boldsymbol{\theta}}_n)$ and $\Omega_Y(\hat{\boldsymbol{\theta}}_n)$ are given respectively by Equation(5.76) and (5.78). Note that in the above expressions we have used the estimated parameter $\hat{\boldsymbol{\theta}}_n$ as a proxy to the true but unknown parameter $\boldsymbol{\theta}_0$. Suppose the standard deviation of an estimated scalar parameter $\hat{\theta}_n$ is denoted by $\hat{\sigma}(\hat{\theta}_n)$, which is given by the square root of the appropriate diagonal entries of $\hat{\Sigma}_n(\hat{\boldsymbol{\theta}}_n)$. Then using Equation(6.35), the confidence interval for $\hat{\sigma}(\hat{\theta}_n)$ with a significance level of $100(1-\alpha)\%$ can be constructed as

$$CI_\alpha(\hat{\theta}_n) = (\hat{\boldsymbol{\theta}}_n - z_{1-\alpha/2}\hat{\sigma}(\hat{\theta}_n), \hat{\boldsymbol{\theta}}_n + z_{1-\alpha/2}\hat{\sigma}(\hat{\theta}_n)), \tag{6.37}$$

where $z_{1-\alpha/2}$ is the appropriate quantile of a standard normal distribution. In order to calculated this confidence interval, the matrices $\Sigma_Y(\hat{\boldsymbol{\theta}}_n)$ and $\Omega_Y(\hat{\boldsymbol{\theta}}_n)$ have to be calculated.

Note that according to Equation(5.76), the entries of the matrix $\Sigma_Y(\hat{\boldsymbol{\theta}}_n)$ involves

an integral with respect to $\frac{\partial}{\partial \boldsymbol{\theta}} S_Y^{\Delta}(f; \hat{\boldsymbol{\theta}}_n)$. This can be evaluated by a DFT approximation, similar to how $S_Y^{\Delta}(f)$ has been calculated, over the sequence $\frac{\partial}{\partial \boldsymbol{\theta}} C_Y(k; \hat{\boldsymbol{\theta}}_n)$, which in turn can be evaluated through the following expression:

$$\frac{\partial}{\partial \boldsymbol{\theta}} C_Y(k; \hat{\boldsymbol{\theta}}_n) = \int_{-\infty}^{\infty} \frac{\partial}{\partial \boldsymbol{\theta}} S_X(\lambda, \hat{\boldsymbol{\theta}}_n) \phi_{\Delta}(\lambda)^{|k|} d\lambda. \tag{6.38}$$

On the other hand, according to Equation(5.78), the entries of the matrix $\Omega_Y(\hat{\boldsymbol{\theta}}_n)$ take the form of the following double integrals

$$\Delta^2 \int_0^{1/\Delta} \int_0^{1/\Delta} g(f, \nu) Q_Y(f, -\mu, \mu) df d\mu. \tag{6.39}$$

Because the function $Q_Y(f, \mu, \nu)$, which represents the fourth order cumulant density of $Y = Y_k$, is generally unknown, the above double integral must be estimated from the finite sample observations. This can be done by using the mean-square consistent estimator developed by Taniguchi [126]. Therefore the matrix $\Omega_Y(\hat{\boldsymbol{\theta}}_n)$ can also be readily calculated, and the confidence interval for any estimated scale parameter $\hat{\theta}_n$ can then be constructed.

We will evaluate the performance of this asymptotic confidence interval through a simulation study, in which we again consider an Oscillatory Matérn process with true parameters $\sigma = 1$, $\phi = 0.5$, $\nu = 1$, and $\lambda_0 = 1$. With m independent replications, the finite sample performance of the confidence interval can be evaluated through the so called empirical coverage probability (ECP), defined as

$$\text{ECP} = \frac{1}{m} \sum_{j=1}^{m} I(\hat{\theta}_{n,j} \in CI_{\alpha}(\hat{\boldsymbol{\theta}}_{n,j})), \tag{6.40}$$

where $I(\cdot)$ is the indicator function, and $\hat{\theta}_{n,j}$ denotes an entry in the estimated parameter vector $\hat{\boldsymbol{\theta}}_{n,j}$ for the jth independent replication. Table(6.13) reports the ECP for $\hat{\boldsymbol{\theta}}_n$ using $m = 500$ independent replications, with a 95% significance level.

It can be observed from the table that the confidence intervals given by Equation(6.37) provided appropriate coverage probabilities that do not deviate too much from the nominal level in most cases. The reported ECP is closer to nominal level and performed slightly better when average sampling interval Δ is small, and when the sampling certainty ratio r is relatively large. However when Δ increases and r decreases, there seems to be a tendency of under coverage by the calculated confidence intervals.

This may be attributable to the fact that under these situations the finite sample performance of the estimator $\hat{\boldsymbol{\theta}}_n$ is relatively poor (as discussed in Section(6.3.2)), and consequently can not be adequately described by the asymptotic results underlying the calculation of the confidence intervals. We also performed a number of further simulation studies with different settings, including different parameters for both the underlying Oscillatory Matérn process and the sampling interval distributions, and observed similar patterns for the empirical coverage probabilities.

OsMatérn, $n = 5000$		Gamma			Inverse Gamma		
		$r = 2$	$r = 1.5$	$r = 1$	$r = 2$	$r = 1.5$	$r = 1$
$\Delta = 0.05$	σ	0.947	0.939	0.936	0.957	0.945	0.931
	ϕ	0.942	0.945	0.929	0.938	0.929	0.916
	ν	0.956	0.933	0.933	0.956	0.923	0.929
	λ_0	0.948	0.931	0.911	0.978	0.931	0.921
$\Delta = 0.1$	σ	0.937	0.929	0.966	0.937	0.949	0.946
	ϕ	0.932	0.945	0.919	0.932	0.945	0.929
	ν	0.936	0.943	0.923	0.916	0.923	0.903
	λ_0	0.962	0.934	0.919	0.933	0.952	0.916
$\Delta = 0.2$	σ	0.947	0.939	0.942	0.943	0.921	0.975
	ϕ	0.932	0.925	0.912	0.926	0.935	0.929
	ν	0.916	0.923	0.891	0.906	0.913	0.886
	λ_0	0.918	0.925	0.914	0.916	0.911	0.901

Table 6.13: Empirical coverage probability for proposed Whittle log-likelihood estimator $\hat{\boldsymbol{\theta}}_n$, with renewal sampling intervals having Gamma and Inverse Gamma distributions. Different combinations of Δ and r are considered. Coverage are based on 500 replications. Confidence level is 95%.

6.4.5 Long Memory Process

We have demonstrated the performance of our proposed estimation method under the assumption that the underlying continuous time process has a short-range dependence covariance structure. In practice, however, observations from many application fields demonstrate the property of long-range dependence [13, 7]. It is therefore nature to ask whether our proposed method could be extended to capture the long range dependence structure in the underlying process. Difficulties however arises from both theoretical and practical considerations. First of all, it is known that the autocovariance function $R_X(\tau)$ is not integrable under the assumption of long-range

dependence [128]. Consequently it will be more difficult (if not impossible) to determine the summability (and other properties) of the autocovariance sequence $C_Y(k)$ of the sampled process. As a result our proposed method cannot be readily applied in this case.

It is, however, expected that the long memory properties of the underlying continuous time process will be preserved under renewal sampling scheme. Although this statement can not be verified rigorously so far, we have conducted a simulation study and found some evidence to support this conjecture. To be more specific, we have considered the case of a fractional Gaussian noise, with a spectral density given by

$$S_X(\lambda) = 4\sigma^2\Gamma(2H+1)\sin(\pi H)\sin^2(\pi\lambda)|\lambda|^{-2H-1}. \tag{6.41}$$

The parameter H controls the degree of long-range dependence, and is usually called the Hurst parameter. In our simulation study, we choose $H = 0.6$.

Then using the same argument as in Section(4.4.2), it is obvious that the sampled process $Y = Y_k$ is a zero-mean stationary sequence. In order to demonstrate statistically the conjecture that $Y = Y_k$ also possesses long-range dependence, we assumed asymptotically at f close to zero, that the spectral density $S_Y(f)$ of $Y = Y_k$ is taking the following classic form:

$$S_Y(f) \sim |1 - e^{-i2\pi f}|^{-2d} S^*(f). \tag{6.42}$$

Here $d \in (0, 0.5)$ is the fractional difference parameter, and the function $S^*(f)$ is an even, positive, continuous function defined on $(-\frac{1}{2}, \frac{1}{2})$. A value of d belonging to $(0, 0.5)$ indicates long-range dependence by producing an singularity at the origin of $S_Y(f)$.

Using the method proposed by Geweke and Proter-Hudak [43], and an estimate $\hat{d}$ can be obtained by regressing $\log(I_n^{\Delta}(\frac{k}{n\Delta}))$ on $|1 - e^{-i\frac{2\pi k}{n}}|$ for $k = 1, 2, \cdots, m$, where $m = o(n)$. A convenient choice of m suggested in the literature is $m = [\sqrt{n}]$, the integer most close to $\sqrt{n}$. Figure(6.10) and (6.10) shows the histogram of estimated values for the fractional difference parameter d over 500 independent repetitions.

It can be observed that when the underlying process possesses long-range dependence, the fractional difference parameter $\hat{d}$ estimated from the sampled process will most likely be positive as well. In fact the confidence intervals for $\hat{d}$ exclude zero for

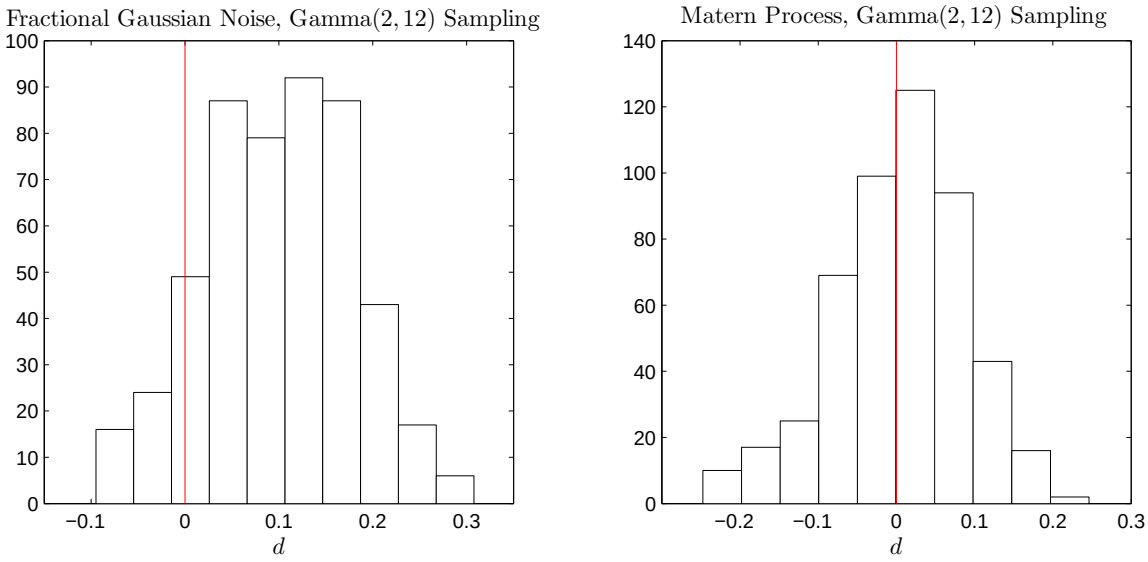

Figure 6.10: Histogram for d with 500 repetitions. Fractional Gaussian noise (H=0.6) sampled through Gamma(2,12) scheme.

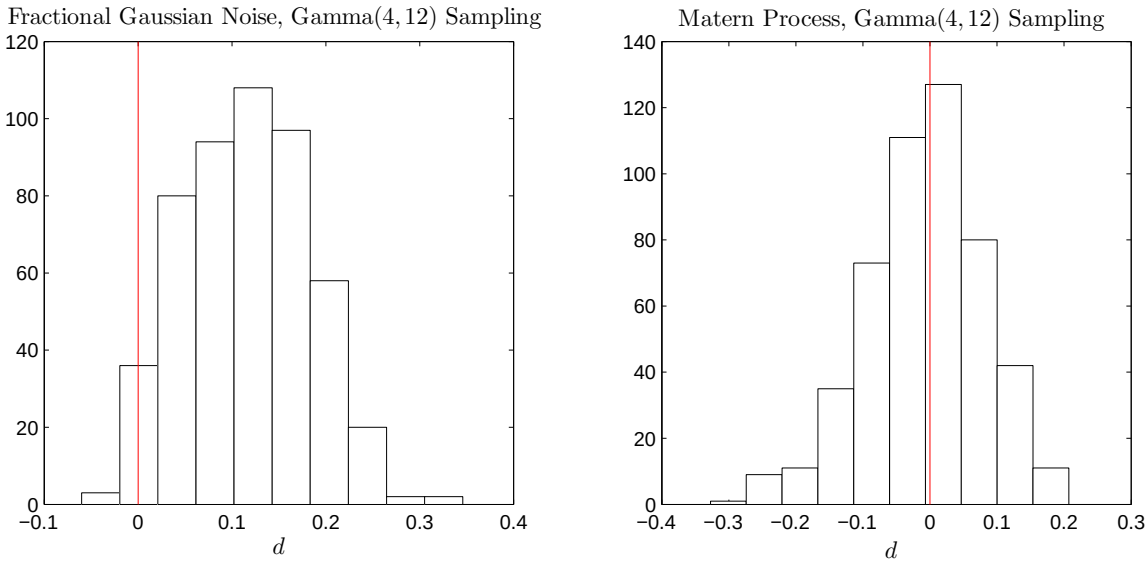

Figure 6.11: Histogram for d with 500 repetitions. Fractional Gaussian noise (H=0.6) sampled through Gamma(4,12) scheme.

most of the 500 repetitions. As a comparison, the parameter $\hat{d}$ is also estimated with a short-memory Matérn process, and observed that the histograms in this case are centered at zero. Moreover, the confidence intervals for $\hat{d}$ in this case include zero in most of the 500 repetitions. Consequently this simple regression test performed in the above simulation study suggested that it is possible to detect whether a long-range dependence property is in place.

A natural question then follows: is there any relationship between the estimated fractional difference parameter $\hat{d}$ and the Hurst parameter H of the underlying fractional Gaussian noise. Recall our analysis in Section(5.4) that the renewal kernel $K_\Delta(0,\lambda)$ contains a delta-function component at $\lambda = 0$, hence the discrete time sampled process should be able to capture low-frequency behavior of the underlying continuous time process. This result in particular suggests that the asymptotic behavior of $S_X(\lambda)$ and $S_Y(f)$ should be similar when both λ and f are close to the origin. Hence it can be expected that the following relationship between H and d should roughly hold (see also Beran [7]):

$$d \approx H - \frac{1}{2}. \tag{6.43}$$

Having a closer look at the histograms, we can see that when the underlying process has a Hurst parameter $H = 0.6$, the histogram of the estimated fractional difference parameters $\hat{d}$ are centered around 0.1. This observation therefore is consistent with the theoretical relationship given in Equation(6.43), suggesting a possibility of estimating the underlying Hurst parameter under the renewal sampling scheme. This is clearly a very interesting direction for future research.

6.5 Summary

In this chapter we discussed the performance issue of the proposed Whittle log-likelihood estimation method. We first showed that by calculating the spectral density $S_Y^\Delta(f)$ through a DFT approximation, the Whittle log-likelihood function for relatively large sample size can be calculated much more efficiently as compared with calculating the traditional Gaussian log-likelihood function. During the discussion of computational efficiency we also discovered the so called renewal sampling whitening effect, in the sense that the autocovariance sequence $C_Y(k)$ of the sampled process

admits faster decay as the covariance function $R_X(\tau)$ contains faster oscillation. This is because oscillatory behaviour of $R_X(\tau)$ will introduce an extra damping factor into $C_Y(k)$.

Then we provided a novel discussion of the special property of the renewal kernel function $K_\Delta(f, \lambda)$. By decomposing the renewal kernel function into a peak effect and an off-peak effect, we showed intuitively that

- the renewal sampling is only able to capture the low frequency spectral feature of the continuous time process $X = X(t)$, because the kernel peak effect has resolution only when f is relatively small; moreover the kernel peak provides better resolution when the IID sampling intervals are less random;

- as f increases the kernel peak gradually losses resolution and consequently the renewal sampling will incur a loss of information as compared to the equally spaced sampling scheme;

- the off-peak effect also provides an alternative explanation to the renewal sampling whitening effect, because as the oscillation of $R_X(\tau)$ increases, more power of $X = X(t)$ will be moved to higher frequencies and hence will be aggregated to give a relatively flat $S_Y^\Delta(f)$.

Consequently this analysis of the renewal kernel suggests the finite sample performance of the proposed Whittle log-likelihood estimation procedure would be better if (1) the average sampling interval Δ is relatively small so that the discrete time sampled process $Y = Y_k$ covers a wider frequency range; (2) the sampling certainty ratio r is relatively large so that the kernel peak losses resolution at slower pace as f increases. These intuitive discussions have subsequently been verified through simulation studies, in which we considered estimating parameters of an Oscillatory Matérn process through Uniform, Gamma and Inverse Gamma renewal sampling schemes.

We also investigated the robustness of the proposed estimation method through simulation studies, and found that correctly specifying the sampling scheme may be vital when the average sampling interval is relatively large or when the sampling certainty ratio is relatively small. Then we considered the problem of constructing asymptotic confidence intervals for the proposed estimator, and showed through simulation studies that the constructed confidence interval generally provided satisfactory

converge probability. Finally we presented a simulation study confirming the possibility of detecting and estimating the Hurst parameter of a underlying continuous process with long-range dependence.

Chapter 7

Conclusion and Future Work

7.1 Conclusions

This thesis has looked at the Monte Carlo simulation and parametric estimation problem associated with unequally spaced sampling over a continuous time stationary process. The process under consideration is assumed to belong to a class, such that a convenient time domain dynamic is not available. Under this assumption, unequally spaced sampling implies that the traditional and computational efficient methods based on equally spaced sampling times, for both simulation and parametric estimation, are no longer applicable.

For the Monte Carlo simulation problem, we considered the spectral simulation method, which generates realizations through discretizing the spectral representation of the underlying continuous time process. The discretization introduces errors, and the covariance functions generated by the spectral method approximates the target covariance function through a quadrature scheme. We reviewed two major variations of the spectral simulation methods, namely the RPDA and RPRA methods. These methods are all based on equally spaced spectral discretization, and the covariance function of the simulated realizations approximate the target through Trapezoidal quadrature rule. We concluded that in practical situations both methods are able to provide adequate approximations to the target process, as long as the discrepancy $\epsilon_{Trap}(\tau)$ between the simulated and the target covariance function is controlled at a reasonable level.

Then we proceeded to the implementation of the spectral simulation method, focusing on the construction of the appropriate spectral discretization scheme. We

showed that $\epsilon_{Trap}(\tau)$ can be decomposed into a truncation error component and an aliasing error component. These components can be controlled through specifying the power cut-off frequency L and the spectral discretization interval Δ_λ respectively. The power cut-off frequency L can be determined through a simple application of the Newton-Raphson recursion, by specifying the desired level of truncation error component. On the other hand, as a major contribution, we proposed a novel algorithm to determine Δ_λ, by considering the aliasing error component. Subsequent numerical and simulation studies showed that when the spectral simulation method is implemented with these proposed algorithms, we would have sufficient control over the discrepancy $\epsilon_{Trap}(\tau)$. As a result it will allow us to keep the computational cost of the spectral simulation method at a manageable level.

For the parametric estimation problem, we considered the approach of modeling the irregularity of the sampling times through a stochastic point process over the real line. We reviewed the existing studies about the stochastic sampling times. These studies were historically motivated from the attempts to alleviate the aliasing effect introduced from equally spaced sampling.

We first reviewed the framework proposed by Masry [79], in which a stationary point process is used to model the random distribution of the sampling times over a fixed time interval. This framework is the most theoretically complete, in the sense that both non-parametric and parametric estimation problem can be rigorously formulated. Such a theoretical advantage is however coupled with a practical disadvantage. It has been argued that the implementation of this framework is only possible when the experimenter can choose a relatively simple sampling scheme.

Being motivated by the practical disadvantages of Masry's framework, we also reviewed an earlier framework proposed by Shapiro and Silverman [113], and in partucular the framework of renewal sampling scheme. This framework is characterized by modeling the irregularity of the sampling intervals through IID random variables. Although this framework is not convenient for the purpose of non-parametric estimation of the underlying spectral density, we showed that the problem of parametric estimation can be conveniently formulated within this framework. In particular an estimation procedure based on a Whittle log-likelihood function can be implemented for a wide class of renewal sampling schemes.

We investigated the theoretical properties of this Whittle log-likelihood estimation method, and proved the asymptotic consistency of the corresponding Whittle esti-

mator. We also discussed the practical implementation of this method. By approximating the spectral density $S_Y^{\Delta}(f)$ of the sampled discrete time stationary process through a DFT approximation, we showed that the Whittle log-likelihood function can be calculated rather efficiently for a large class of renewal sampling schemes. We then investigated the finite sample performance of the proposed Whittle estimator, and discovered the renewal sampling aggregation/whitening effect. This aggregation/whitening effect causes information losses, and implies that the renewal sampling scheme can only capture low-frequency spectral features of the underlying continuous time process. However it also eliminates the souses of ambiguity in model identification that is caused by the aliasing effect under the equally spaced sampling times. By using a Taylor series based asymptotic argument, we found that the resolution of renewal sampling schemes were primarily determined by the average sampling interval Δ and the sampling certainty ratio $r = \frac{\Delta}{\sqrt{\text{Var}[\Delta]}}$. These discussions and discoveries have been verified through simulation studies.

7.2 Future Work

The work presented in this thesis provided some new algorithms and methods to deal with observations obtained from unequally spaced sampling times. There are still much work to be done, both in extending and perfecting the existing framework, and also in discovering new methods and frameworks to deal with more general situations.

7.2.1 Simulation Methods

The discussions on simulation methods in this thesis deals with the simplest case of a real-valued continuous time stationary process. Naturally we can consider extensions to include the stationary vector process and non-stationary processes. Using the corresponding spectral representation of a stationary vector process [26], the spectral simulation method can be easily constructed. In fact Deodatis [34] showed that under equally spaced sampling, the spectral simulation of stationary vector process can be accomplished efficiently through FFT technique. Under unequally spaced sampling patterns, however, the efficiency of the FFT technique is no longer available. Consequently it becomes more important to construct an appropriate spectral discretization scheme, so that the balance between the simulation error and the computational costs

can be effectively controlled. So far it is not very clear how this can be done in detail, and therefore constitutes an interesting direction of future research.

Although stationarity is a theoretical desirable property, in many practice situations this is not always a reasonable assumption [4]. The spectral representation for a non-stationary process has been defined in several ways in the literature [6, 101], with different areas of applications. Among these existing definitions, the most widely used is probably Priestley's evolution spectral representation [98, 99]. Through discretization of this evolution spectral representation, a spectral simulation formula can be easily derived [69]. Again the difficulty of this approach reduces to how the simulation error can be controlled through constructing the appropriate spectral discretization scheme, and has not been discussed in existing literature. Consequently more works has to be done in this direction.

7.2.2 Parametric Estimation

The problem of parametric estimation under stochastic sampling times is a vast area. In this thesis we have confined our attention to the simplified scenario of sampling a real-valued continuous time stationary process through renewal sampling scheme. Although the asymptotic consistency of the Whittle estimator has been established, we did not manage to verify its asymptotic normality. The rigorous derivation of the asymptotic normality of the Whittle estimator relies on the higher order statistical properties of the sampled process, which so far is not fully known. More work has to be done in this direction to obtain more understanding of the probabilistic structure of the sampled process.

Other direction of future research include the extension to underlying process with long-range dependence, which we have briefly considered in Section(6.4.5). On the other hand, a more direct generalization of the proposed Whittle log-likelihood estimation method to vector stationary process should be straight-forward. Suppose a vector stationary process $\boldsymbol{X} = \boldsymbol{X}(t)$ *with autocovariance matrix* $\Sigma_{\boldsymbol{X}}(\tau)$ *is sampled through a renewal sampling scheme. Then a straight-forward calculation (similar to the calculation used to obtain Equation(4.21)) shows that the sampled process* $\boldsymbol{Y} = \boldsymbol{Y}_k$ *is also vector stationary with autocovariance matrix* $\Sigma_{\boldsymbol{Y}}(k)$ *given by*

$$\Sigma_{\boldsymbol{Y}}(k) = \int_0^\infty \Sigma_{\boldsymbol{X}}(\tau) p_k(\tau) d\tau = \int_{-\infty}^\infty F_{\boldsymbol{X}}(\lambda) \phi_\Delta^{|k|}(\lambda) d\lambda, \tag{7.1}$$

where $F_{\boldsymbol{X}}(\lambda)$ is the spectral density matrix of the continuous time vector stationary process $\boldsymbol{X} = \boldsymbol{X}(t)$. Then Proposition(5.1) can be readily applied to arrive at the conclusion that the autocovariance matrix $\Sigma_{\boldsymbol{Y}}(k)$ is absolutely summable. Consequently the spectral density matrix $F_{\boldsymbol{Y}}(\lambda)$ of the sampled process $\boldsymbol{Y} = \boldsymbol{Y}_k$ exists as the Fourier transform of the autocovariance matrix $\Sigma_{\boldsymbol{Y}}(k)$. With the above results established, there should be no difficulty in extending the proposed parametric estimation procedure to this multi-dimensional scenario.

Moreover, using the spectral domain expression for $\Sigma_{\boldsymbol{Y}}(k)$ given in Equation(7.1), the spectral density matrix $F_{\boldsymbol{Y}}(f)$ of the sampled process can be written in terms of the renewal kernel $K_{\Delta}(f, \lambda)$ as

$$F_{\boldsymbol{Y}}(f) = \int_{-\infty}^{\infty} F_{\boldsymbol{X}}(\lambda) K_{\Delta}(f, \lambda) d\lambda. \tag{7.2}$$

Consequently, the discussions in Section(6.3.2) about the renewal sampling aggregation and whitening effects can be readily carried out to each component of the spectral density matrix $F_{\boldsymbol{Y}}(\lambda)$.

On the other hand as an alternative extension of the proposed renewal sampling Whittle log-likelihood estimation method, we have conducted some preliminary work on a particular class of non-stationary process, constructed from integrating a real-valued continuous time stationary process $X = X(t)$ as follows:

$$Y(t) = \int_0^t X(s) ds, \;\; t > 0. \tag{7.3}$$

When sampled under renewal sampling scheme, we can construct the finite-difference process $Z = Z_k$ as follows:

$$Z_k = Y(t_k) - Y(t_{k-1}), \;\; k = 1, 2, \cdots, \tag{7.4}$$

and it can be easily shown that Z_k is also a second-order stationary process. The spectral density $S_Z^{\Delta}(f)$ of $Z = Z_k$ is well defined through the DFT of the autocovariance sequence, and consequently the Whittle log-likelihood function can be readily calculated. However, because of the finite-difference procedure introduces a more complex dependence of $Z = Z_k$ over the renewal sampling scheme, we can not establish rigorously the convergence of the sampled autocovariance. This therefore does not allow us to claim the asymptotic consistency of the corresponding estimator. Trying to

theoretically justify the Whittle log-likelihood estimation method for this particular case of non-stationarity is therefore an interesting direction of future research.

The renewal sampling scheme is just a particular example of stochastic sampling times that fits into the general framework of Shapiro, Silverman and Beutler [113, 8]. There are other stochastic sampling patterns that are included in this framework. Consequently another interesting direction of future research is to extend the proposed Whittle log-likelihood estimation procedure to include these sampling patterns. For example, according to the calculation in Thomson and Robinson [130], under the scenario of jittered sampling, the spectral density of the sampled discrete time stationary process, and hence the Whittle log-likelihood function, can be very efficiently calculated. Moreover because of the simple dependence of second order properties of the discrete time sampled process over the sampling jitter, we believe the theoretical verification of the validity of the Whittle log-likelihood estimation procedure should not be difficult. Furthermore, Shapiro, Silverman and Beutler's framework also includes the sampling times generated from a filtered renewal process, which can be used to model the dependent sampling intervals. Extension of the proposed estimation method over such scenario of dependent sampling intervals will be more interesting from a practical point of view.

Finally we mention the fact that the Shapiro, Silverman and Beutler's framework is far from perfect. Because the actual sampling times are ignored, as shown in this thesis through the example of renewal sampling times, the resulting information loss in some circumstance may be significant. Although Masry's framework (which models the sampling times through a stationary point process) takes into account the actual sampling times, as discussed in the thesis it is not clear how the corresponding parametric estimation procedure can be implemented efficiently. This will involve a deeper understanding of the statistical inference of the sampling point process, and has to be investigated in future work. The ultimate aim of the future research is to explore the possibility of developing a generalized framework, so that the problem of parametric estimation under stochastic sampling times can be addressed with both accuracy and efficiency.

Chapter 8

Appendix

This chapter contains detailed derivations of some of the results appeared in the previous discussions.

8.1 Derivation for Chapter 3

In this section, we give detailed derivation of results contained in Proposition(3.1) of Section(3.3.3). We first consider the temporal mean $M_{\hat{X}_{RPRA},T}$. Since both $X = X(t)$ and $\hat{X}_{RPRA} = \hat{X}_{RPRA}(t)$ are zero-mean Gaussian processes, by an application of Lemma(3.1) we have

$$\mathrm{E}[M_{X,T}] = \mathrm{E}[M_{\hat{X}_{RPRA},T}] = 0, \tag{8.1}$$

and

$$\begin{aligned} |\mathrm{Var}[M_{X,T}] - \mathrm{Var}[M_{\hat{X}_{RPRA},T}]| &= |\frac{2}{T}\int_0^T (1-\frac{\tau}{T})[R_X(\tau) - R_{\hat{X},Trap}(\tau)]d\tau| \\ &\leq \frac{2}{T}\int_0^T (1-\frac{\tau}{T})|R_X(\tau) - R_{\hat{X},Trap}(\tau)|d\tau \\ &\leq \frac{2}{T}\int_0^T (1-\frac{\tau}{T}) \max_{\tau\in[0,T]} |\epsilon_{\hat{X}}(\tau)|d\tau \\ &= \max_{\tau\in[0,T]} |\epsilon_{\hat{X}}(\tau)|. \end{aligned} \tag{8.2}$$

This therefore proves the desired results for the temporal means, and we now turn our attention to the temporal covariances. Without loss of generality, assume that

$\tau \geq 0$. Then note that by Equation(3.56) in Lemma(3.1), we can immediately write

$$\mathrm{E}[R_{X,T}(\tau)] - \mathrm{E}[R_{\hat{X}_{RPRA},T}(\tau)] = R_X(\tau) - R_{\hat{X}_{RPRA}}(\tau). \tag{8.3}$$

To discuss the difference of variances between $R_{X,T}(\tau)$ and $R_{\hat{X}_{RPRA},T}(\tau)$, note that by Equation(3.57) of Lemma(3.1) we have

$$|\mathrm{Var}[R_{X,T}(\tau)] - \mathrm{Var}[R_{\hat{X}_{RPRA},T}(\tau)]| \leq \frac{2}{T}\int_0^T (1 - \frac{\tau}{T})|P_\tau(t)|dt, \tag{8.4}$$

where to simplify discussion we defined the function $P_\tau(t)$ as

$$P_\tau(t) = R_X(t)^2 + R_X(t+\tau)R_X(t-\tau) - R_{\hat{X},Trap}(t)^2 - R_{\hat{X},Trap}(t+\tau)R_{\hat{X},Trap}(t-\tau). \tag{8.5}$$

By adding and subtracting the term $R_{\hat{X},Trap}(t+\tau)R_{\hat{X},Trap}(t-\tau)$, we now derive an upper bound for $P_\tau(t)$ with $t \in [0, T]$ as follows:

$$\begin{aligned}
|P_\tau(t)| &\leq |R_X(t)^2 - R_{\hat{X},Trap}(t)^2| + |R_X(t+\tau)R_X(t-\tau) - R_{\hat{X},Trap}(t+\tau)R_{\hat{X},Trap}(t-\tau)| \\
&\leq |R_X(t)^2 - R_{\hat{X},Trap}(t)^2| + |R_X(t+\tau)R_X(t-\tau) - R_{\hat{X},Trap}(t+\tau)R_X(t-\tau)| \\
&\quad + |R_{\hat{X},Trap}(t+\tau)R_X(t-\tau) - R_{\hat{X},Trap}(t+\tau)R_{\hat{X},Trap}(t-\tau)| \\
&\leq |R_X(t) - R_{\hat{X},Trap}(t)||R_X(t) + R_{\hat{X},Trap}(t)| \\
&\quad + |R_X(t-\tau)||R_X(t+\tau) - R_{\hat{X},Trap}(t+\tau)| \\
&\quad + |R_{\hat{X},Trap}(t+\tau)||R_X(t-\tau) - R_{\hat{X},Trap}(t-\tau)| \\
&\leq \max_{s\in[0,T]} |\epsilon_{\hat{X}}(s)|[R_X(0) + R_{\hat{X},Trap}(0)] + \max_{s\in[\tau,T+\tau]} |\epsilon_{\hat{X}}(s)|R_X(0) \\
&\quad + \max_{s\in[-\tau,T-\tau]} |\epsilon_{\hat{X}}(s)|R_{\hat{X},Trap}(0),
\end{aligned} \tag{8.6}$$

where in the last inequality we have used the fact that the magnitude of both $R_X(\tau)$ and $R_{\hat{X},Trap}(\tau)$ reaches maximum when $\tau = 0$. Now, using the simple fact that $\epsilon_{\hat{X}}(s)$ is symmetric around $s = 0$, we should have the following relations

$$\max_{s\in[0,T]} |\epsilon_{\hat{X}}(s)| \leq \max_{s\in[0,T+\tau]} |\epsilon_{\hat{X}}(s)|, \tag{8.7}$$

$$\max_{s\in[\tau,T+\tau]} |\epsilon_{\hat{X}}(s)| \leq \max_{s\in[0,T+\tau]} |\epsilon_{\hat{X}}(s)|, \tag{8.8}$$

and

$$\max_{s\in[-\tau,T-\tau]} |\epsilon_{\hat{X}}(s)| \leq \max_{s\in[0,T+\tau]} |\epsilon_{\hat{X}}(s)|. \tag{8.9}$$

Using these inequalities in Equation(8.6) then gives

$$|P_\tau(t)| \leq \max_{s\in[0,T+\tau]} |\epsilon_{\hat{X}}(s)|[R_X(0) + R_{\hat{X},Trap}(0)], \quad t \in [0,T]. \tag{8.10}$$

Putting this upper-bound into Equation(8.4) then gives the desired result. This complete the proof of Proposition(3.1).

8.2 Derivation for Chapter 5

In this section of the appendix, we will provide detailed proof of Lemma(5.8) in Section(5.5.3). This result is crucial in providing upper-bound for the variance of the sample autocovariance sequence $\mathrm{Var}[\hat{c}_n(k)]$, which is given by Equation(5.60).

We first focus on the term $U_{n,1}(k) = 2A^1_{n,1}(k) - \mathrm{E}^2[\hat{c}_n(k)]$. Note first that for $k \in \Pi_1$, the random variables $t_{j+k} - t_j$ and $t_{l+k} - t_l$ are independent of each other, hence we should be able to factorize the expectations in $A^1_{n,1}(m)$ and write

$$\begin{aligned}
2A^1_{n,1}(k) &= \frac{2}{n^2} \sum_{j,l\in\Pi_1} \mathrm{E}[R_X(t_{j+k} - t_j)R_X(t_{l+k} - t_l)] \\
&= \frac{2}{n^2} \sum_{j,l\in\Pi_1} \mathrm{E}[R_X(t_{j+k} - t_j)]\mathrm{E}[R_X(t_{l+k} - t_l)] \\
&= \frac{2}{n^2} \sum_{j,l\in\Pi_1} C^2_Y(k) \\
&= \frac{(1+n-2k)(n-2k)}{n^2} C^2_Y(k),
\end{aligned} \tag{8.11}$$

since there are altogether $\frac{(1+n-2k)(n-2k)}{2}$ terms in the summation $\sum_{j,l\in\Pi_1}$. On the other hand the term $\mathrm{E}^2[\hat{c}_n(k)]$ can be easily written as

$$\mathrm{E}^2[\hat{c}_n(k)] = (\frac{n-k}{n})^2 C^2_Y(k). \tag{8.12}$$

Hence we can immediate write $U_{n,1}(k)$ as follows:

$$U_{n,1}(k) = 2A^1_{n,1}(k) - \mathrm{E}^2[\hat{c}_n(k)] = \frac{n - 2k(1+n) + 3k^2}{n^2} C^2_Y(k), \tag{8.13}$$

which is the required expression for $U_{n,1}(k)$. Next we work on the bound for of $A^2_{n,1}(m)$, which is given by

$$A^2_{n,1}(k) = \frac{1}{n^2} \sum_{j,l \in \Pi_2} \mathrm{E}[R_X(t_{j+k} - t_j) R_X(t_{l+k} - t_l)]. \tag{8.14}$$

In general the random variables $t_{j+k} - t_j$ and $t_{l+k} - t_l$, for $j, l \in \Pi_2 = \{(j,l) : 0 < j < l < j+k < l+k\}$, are not independent of each other, because they represent sampling intervals that are overlapping. Therefore the expectations in the expressions for $A^2_{n,1}(k)$ can not be directly factorized. However, a trick can be applied by utilizing the assumption that $|R_X(\tau)|$ is bounded by an even, non-negative, integrable function function $H_X(\tau)$, which is non-increasing for $\tau > 0$. To be more specific, we first use the triangle-inequality to write

$$\begin{aligned} |A^2_{n,1}(k)| &\leq \frac{1}{n^2} \sum_{j,l \in \Pi_2} \mathrm{E}[|R_X(t_{j+k} - t_j)||R_X(t_{l+k} - t_l)|] \\ &\leq \frac{1}{n^2} \sum_{j,l \in \Pi_2} \mathrm{E}[H_X(t_{j+k} - t_j) H_X(t_{l+k} - t_l)] \end{aligned} \tag{8.15}$$

Being even and non-increasing for $\tau > 0$, the function $H_X(\tau)$ allows us to (1)replace the random variable $t_{j+k} - t_j$ with $t_l - t_j$; (2)replace the random variable $t_{l+k} - t_l$ with 0, so that the inequality will remain hold. In other words we can write

$$\begin{aligned} |A^2_{n,1}(k)| &\leq \frac{1}{n^2} \sum_{j,l \in \Pi_2} \mathrm{E}[H_X(t_{j+k} - t_j) H_X(t_{l+k} - t_l)] \\ &\leq \frac{1}{n^2} \sum_{j,l \in \Pi_2} \mathrm{E}[H_X(t_l - t_j)] \mathrm{E}[H_X(0)] \\ &= \frac{H_X(0)}{n^2} \sum_{j,l=1}^{n-k} \mathrm{E}[H_X(t_j - t_l)] \end{aligned}$$

which is the desired upper-bound for $|A^2_{n,1}(k)|$. Using very similar techniques, we can easily derive the corresponding expressions for the upper-bounds of $|A^r_{n,m}(k)|, m = 2, 3,\ r = 1, 2$. Next we turn to the upper-bound for $|A^3_{n,m}(k)|, m = 1, 2, 3$, which is

relatively easy because of the simple structure of the subregion Π_3.

$$\begin{aligned}|A^3_{n,m}(k)| &\leq \frac{1}{n^2}\sum_{j,l\in\Pi_3}\mathrm{E}[|R_X(t_{l+k}-t_j)||R_X(t_{j+k}-t_l)|] \\ &= \frac{1}{n^2}\sum_{j,l\in\Pi_3}\mathrm{E}[|R_X(t_{j+k}-t_j)|^2] \\ &\leq \frac{n-k}{n^2}R_X^2(0), \end{aligned} \tag{8.16}$$

where the last inequality holds because $|R_X(\tau)| \leq |R_X(0)|$ for any τ. The last term $|A_{n,4}(k)|$ involves the fourth-order cumulant function $Q_X(\tau_1,\tau_2,\tau_3)$, which is assumed to be bounded by $H_Q(\tau_1,\tau_2,\tau_3)$. Because $H_Q(\tau_1,\tau_2,\tau_3)$ is even and non-increasing on $[0,\infty)$ in each variable such that $\int_0^\infty H(0,\tau,0)d\tau < \infty$, We can then write

$$\begin{aligned}|A_{n,4}(k)| &\leq \frac{1}{n^2}\sum_{j,l=1}^{n-k}\mathrm{E}[|H_Q(t_{j+n}-t_j,t_l-t_j,t_{l+n}-t_j)|] \\ &\leq \frac{1}{n^2}\sum_{j,l=1}^{n-k}\mathrm{E}[H_Q(t_{k+n}-t_j,t_l-t_j,t_{l+n}-t_j)] \\ &\leq \frac{1}{n^2}\sum_{j,l=1}^{n-k}\mathrm{E}[H_Q(0,t_l-t_j,0)] \\ &= \frac{1}{n^2}\sum_{j,l=1}^{n-k}\mathrm{E}[H_Q(0,t_j-t_l,0)] \end{aligned} \tag{8.17}$$

which is the desired bound for $A_{n,4}(k)$. This then completes the proof of Lemma(5.8).

8.3 Derivation for Chapter 6

8.3.1 Asympotic Decay of the Autocovariance $C_Y(k)$

This section will provide the detailed derivation of the asymptotic decay of $C_Y(k)$, which is contained in Proposition(6.1) of Section(6.2.2). Without loss of generality we can assume $k > 0$. Then we can integrate the asymptotic form of $R_X(\tau)$ given by Equation(6.5) with respect to the gamma density to obtain the following asymptotic

expression for $C_Y(k)$ as:

$$\begin{aligned} C_Y(k) &\propto \int_0^\infty \tau^\nu e^{-\phi\tau} \cos(2\pi\lambda_0\tau + \psi) g_{|k|}(\tau) d\tau \\ &\propto \frac{\beta^{\alpha_k}}{\Gamma(\alpha_k)} \int_0^\infty \cos(2\pi\lambda_0\tau + \psi) \tau^{\alpha_k+\nu-1} e^{-(\beta+\phi)\tau} d\tau \\ &= \left(\frac{\beta}{\beta+\phi}\right)^{\alpha_k} \frac{\Gamma(\alpha_k+\nu)}{\Gamma(\alpha_k)} \left(\frac{1}{\beta+\phi}\right)^{\nu} \times \\ &\quad \int_0^\infty \frac{(\beta+\phi)^{\alpha_k+\nu}}{\Gamma(\alpha_k+\nu)} \cos(2\pi\lambda_0\tau + \psi) \tau^{\alpha_k+\nu-1} e^{-(\beta+\phi)\tau} d\tau. \end{aligned} \tag{8.18}$$

We first consider the integral in the above equation. Assuming k is large enough so that $\alpha_k+\nu-1 > 0$, then the integral in the above equation can be easily related to the characteristic function $\phi_k(\lambda)$ corresponding to the gamma density $g_k(\tau)$ as follows:

$$\begin{aligned} &\int_0^\infty \frac{(\beta+\phi)^{\alpha_k+\nu}}{\Gamma(\alpha_k+\nu)} \cos(2\pi\lambda_0\tau + \psi) \tau^{\alpha_k+\nu-1} e^{-(\beta+\phi)\tau} d\tau \\ &= \int_0^\infty \frac{(\beta+\phi)^{\alpha_k+\nu}}{\Gamma(\alpha_k+\nu)} \cos(2\pi\lambda_0\tau) \cos(\psi) \tau^{\alpha_k+\nu-1} e^{-(\beta+\phi)\tau} d\tau \\ &\quad - \int_0^\infty \frac{(\beta+\phi)^{\alpha_k+\nu}}{\Gamma(\alpha_k+\nu)} \sin(2\pi\lambda_0\tau) \sin(\psi) \tau^{\alpha_k+\nu-1} e^{-(\beta+\phi)\tau} d\tau \\ &= \cos(\psi) \mathrm{Re}\left(\phi_k(\lambda_0)\right) - \sin(\psi) \mathrm{Im}\left(\phi_k(\lambda_0)\right). \end{aligned} \tag{8.19}$$

The characteristic function $\phi_k(\lambda_0)$ of the gamma density $g_k(\tau)$ can be easily evaluated as

$$\begin{aligned} \phi_k(\lambda_0) &= \left(\frac{\beta}{\beta - 2\pi i\lambda_0}\right)^{\alpha_k+\nu} \\ &= \exp\left((\alpha_k+\nu)\mathrm{Log}\left(\frac{\beta}{\beta - i2\pi\lambda_0}\right)\right) \\ &= \left(\frac{\beta}{\sqrt{\beta^2 + 4\pi^2\lambda_0^2}}\right)^{\alpha_k+\nu} \exp\left[i(\alpha_k+\nu)\arctan\left(\frac{2\pi\lambda_0}{\beta}\right)\right] \end{aligned}$$

where $\arctan\left(\frac{2\pi\lambda_0}{\beta}\right)$ is the principal value of the argument of the complex number $z = \frac{\beta}{\beta - i2\pi\lambda_0}$. This then gives the expression for both $\mathrm{Re}\{\phi_k(\lambda_0)\}$ and $\mathrm{Im}\{\phi_k(\lambda_0)\}$ as

$$\mathrm{Re}\left(\phi_k(\lambda_0)\right) = \left(\frac{\beta}{\sqrt{\beta^2 + 4\pi^2\lambda_0^2}}\right)^{\alpha_k+\nu} \cos\left[(\alpha_k+\nu)\arctan\left(\frac{2\pi\lambda_0}{\beta}\right)\right], \tag{8.20}$$

$$\mathrm{Im}\left(\phi_m(\lambda_0)\right) = \left(\frac{\beta}{\sqrt{\beta^2 + 4\pi^2\lambda_0^2}}\right)^{\alpha_k+\nu} \sin\left[(\alpha_k+\nu)\arctan\left(\frac{2\pi\lambda_0}{\beta}\right)\right]. \tag{8.21}$$

Now putting Equation(8.20) and (8.21) back into the expression given in Equation(8.19) will allow us to derived the asymptotic expression for the integral part in Equation(8.18):

$$\begin{aligned}
&\int_0^\infty \frac{(\beta+\phi)^{\alpha_k+\nu}}{\Gamma(\alpha_k+\nu)} \cos(2\pi\lambda_0\tau + \psi)\tau^{\alpha_k+\nu-1}e^{-(\beta+\phi)\tau}d\tau \\
&= \left(\frac{\beta}{\sqrt{\beta^2 + 4\pi^2\lambda_0^2}}\right)^{\alpha_k+\nu} \cos\left[\arctan\left(\frac{2\pi\lambda_0}{\beta}\right)(\alpha_k+\nu) + \psi\right].
\end{aligned} \tag{8.22}$$

Having simplified integral part of Equation(8.18), we now turn our attention to the multiplicative constant term. To simplify this term, we can use the Stirling's formula $\Gamma(z) \approx \sqrt{2\pi z}(\frac{z}{e})^z$ to approximate the ratio of the gamma functions $\Gamma(\alpha_k+\nu)/\Gamma(\alpha_k)$ as

$$\begin{aligned}
\frac{\Gamma(\alpha_k+\nu)}{\Gamma(\alpha_k)} &= \sqrt{\frac{\alpha_k+\nu-1}{\alpha_k-1}}(\frac{\alpha_k+\nu-1}{\alpha_k-1})^{\alpha_k-1}(\frac{\alpha_k+\nu-1}{e})^\nu \\
&= \sqrt{\frac{\alpha_k+\nu-1}{\alpha_k-1}}(1+\frac{\nu}{\alpha_k-1})^{\alpha_k-1}(\frac{\alpha_k+\nu-1}{e})^\nu \\
&\approx e^\nu \left(\frac{\alpha_k+\nu-1}{e}\right)^\nu \\
&= (\alpha_k+\nu-1)^\nu,
\end{aligned} \tag{8.23}$$

where for the purpose of simplification we have used the classical result (see for example Gaughan [42]) that $\lim_{x\to}(1+\frac{\nu}{x})^x = e^\nu$. Substituting Equation(8.22) and Equation(8.23) back into Equation(8.18) then gives the desired asymptotic expression for $C_Y(k)$.

8.3.2 Approximating Location of the Kernel Peak

This section will be devoted to a discussion of approximating the location of the kernel peak of a renewal kernel function $K_\Delta(f, \lambda)$, under the asymptotic assumption that the discrete time frequency variable f is not far away from zero. From the plots in Figure(6.5) in Section(6.3.1), it seems that, when f is small, the location λ_p of the kernel peak is at a frequency λ close to f. This observation suggests a closer look at the behaviour of the kernel function $K_\Delta(f, \lambda)$ when f and λ are both small, through approximating the characteristic function $\phi_\Delta(\lambda)$ for small λ. Instead of Taylor expanding the characteristic function directly, we find it more convenient to expand the logarithm of the characteristic function (see Shiraev[117]) as

$$\log(\phi_\Delta(\lambda)) = i2\pi\Delta\lambda - 2\pi^2 \mathrm{Var}[\Delta]\lambda^2 + o(|\lambda|^2). \tag{8.24}$$

Taking exponential on both sides then gives

$$\phi_\Delta(\lambda) \approx e^{-2\pi^2 \mathrm{Var}[\Delta]\lambda^2} e^{i2\pi\Delta\lambda}, \tag{8.25}$$

so that the renewal kernel $K_\Delta(f, \lambda)$ for small $f > 0$ and small $\lambda > 0$ can be written approximately as

$$\begin{aligned} K_\Delta(f, \lambda) &\approx \mathrm{Re}\left\{ \frac{1 + e^{-\pi^2 \mathrm{Var}[\Delta]\lambda^2} e^{i2\pi\Delta(\lambda - f)}}{1 - e^{-\pi^2 \mathrm{Var}[\Delta]\lambda^2} e^{i2\pi\Delta(\lambda - f)}} \right\} \\ &= \frac{1 - e^{-4\pi^2 \mathrm{Var}[\Delta]\lambda^2}}{|1 - e^{-2\pi^2 \mathrm{Var}[\Delta]\lambda^2} e^{i2\pi\Delta(\lambda - f)}|^2} \\ &= \frac{1 - e^{-4\pi^2 \mathrm{Var}[\Delta]\lambda^2}}{1 - 2e^{-2\pi^2 \mathrm{Var}[\Delta]\lambda^2} \cos(2\pi\Delta(\lambda - f)) + e^{-4\pi^2 \mathrm{Var}[\Delta]\lambda^2}} \\ &\approx \frac{\mathrm{Var}[\Delta]\lambda^2}{\pi^2 \mathrm{Var}^2[\Delta]\lambda^4 + \Delta^2(\lambda - f)^2}. \end{aligned} \tag{8.26}$$

Note that in the last step we have further used the Taylor series expansion $e^x \approx 1 + x$ and $\cos(x) \approx 1 - \frac{x^2}{2}$ for x close to zero. To find the approximate location λ_p of the peak, we simply note that at λ_p the derivative of Equation(8.26) with respect to λ must be (close to) zero. Therefore, after tediously differencing Equation(8.26) with

respect to λ and set the numerator to zero, we then obtain the following equation

$$f^2\Delta^2 - f\Delta^2\lambda_p - \pi^2\mathrm{Var}^2[\Delta]\lambda_p^4 \approx 0. \tag{8.27}$$

It would be difficult to find the expression for λ_p by directly solving the above equation for λ_p. However it can be noted that this equation cal also be regarded as a quadratic equation in terms of the variable f, which can be readily solved as

$$\begin{aligned} f &\approx \frac{\Delta^2\lambda_p \pm \sqrt{\Delta^4\lambda_p^2 + 4\Delta^2\pi^2\mathrm{Var}^2[\Delta]\lambda_p^4}}{2\Delta^2} \\ &= \frac{\lambda_p \pm \lambda_p\sqrt{1 + 4\pi^2\frac{\mathrm{Var}^2[\Delta]}{\Delta^2}\lambda_p^2}}{2} \end{aligned} \tag{8.28}$$

Because f has to be positive, we must take the positive root. Moreover it can be observed from Equation(8.28) that when f is assumed to be small, so is λ_p. Consequently we can use the the Taylor series expansion $\sqrt{1+x} \approx 1 + \frac{x}{2}$ to further simplify the above expression of f in terms of λ_p as

$$f \approx \lambda_p + \pi^2\frac{\mathrm{Var}^2[\Delta]}{\Delta}\lambda_p^3. \tag{8.29}$$

Hence we immediately have $f \geq \lambda_p$. This implies that when f is small so is λ, and consequently the term $\pi^2\frac{\mathrm{Var}^2[\Delta]}{\Delta}\lambda_p^3$ will be negligible as compared to λ_p. Although the approximation (8.29) is not a direct expression for the location λ_p of the kernel peak, it provides sufficient information to show that when f is small, the peak location λ_p is indeed close to f.

8.3.3 Approximating Half Width of the Kernel Peak

In section(6.3) we have used the quantity d, defined through

$$K_\Delta(f, \lambda_p - d) = 1, \tag{8.30}$$

as a measure of the half width of the kernel peak. Assuming that f is small, it is possible to obtain an elegant approximating formula for the half-width d of the kernel function $K_\Delta(f, \lambda)$, by making use of the approximating equation(8.26). In fact setting

Equation(8.26) to one, we then obtain the following equation

$$\text{Var}[\Delta](\lambda_p - d)^2 = \pi^2 \text{Var}^2[\Delta](\lambda_p - d)^4 + \Delta^2 d^2. \tag{8.31}$$

Since $(\lambda_p - d)$ is assumed to be small, the term containing $(\lambda_p - d)^4$ should be negligible and hence the above equation can be replaced by the following approximation

$$\text{Var}[\Delta](\lambda_p - d)^2 \approx \Delta^2 d^2. \tag{8.32}$$

This then immediately gives the following elegant approximating expression for the half width d of the kernel peak as

$$d \approx \frac{\sqrt{\text{Var}[\Delta]}}{\Delta + \sqrt{\text{Var}[\Delta]}} \lambda_p = \frac{\lambda_p}{\frac{\Delta}{\sqrt{\text{Var}[\Delta]}} + 1}. \tag{8.33}$$

Using the fact that was derived from previous section that $\lambda_p \approx f$ when f is close to zero, the half width can then be further approximated by

$$d \approx \frac{f}{\frac{\Delta}{\sqrt{\text{Var}[\Delta]}} + 1}, \tag{8.34}$$

Table(8.1) compares the exact half-width with the approximation of (8.33), for three different renewal sampling intervals with the same average sampling interval $\Delta = 0.25$ and variance $\text{Var}[\Delta] = 1/48$. It is as expected that the approximating formula(8.33) does a good job in estimating the true half-width of the renewal kernel peak when f is small. When f is relatively large, the approximating formula(8.33) is not meant to be accurate. It will, however, still provide intuitive understanding that the sampling certainty ratio $\frac{\Delta}{\sqrt{\text{Var}[\Delta]}}$, which measures the randomness of the IID sampling intervals, is an important determining factor of the resolution of the renewal kernel peak. Higher the ratio means less uncertainty of the sampling intervals, and the resulting narrower peak half-width gives higher resolution.

8.3.4 Existence of Renewal Kernel Peak

The approximation argument in the last two subsections is a valid description of the shape of the kernel $K_\Delta(f, \lambda)$ for λ close to zero, under the assumption that f is also

	Uniform		Gamma		Inverse Gamma	
	true d	approx. d	true d	approx. d	true d	approx. d
$f = 0.01$	0.00366	0.0037	0.00366	0.0037	0.00366	0.0037
$f = 0.05$	0.01829	0.0183	0.0183	0.0183	0.01831	0.0183
$f = 0.1$	0.03653	0.0366	0.0366	0.0366	0.03663	0.0367
$f = 0.3$	0.10789	0.1091	0.10968	0.1100	0.11055	0.1113
$f = 0.5$	0.17419	0.1799	0.1824	0.1841	0.18564	0.1886
$f = 1$	0.29783	0.3420	0.36092	0.3749	0.37627	0.3965
$f = 1.5$	0.33648	0.4747	0.53397	0.5829	0.56485	0.6271
$f = 2$	0.2872	0.5779	0.71183	0.8313	0.74508	0.8822

Table 8.1: Comparing True and Approximate Half-Width d

close to zero. When f increases, these approximations are no longer valid, to the extent that the kernel peak may even disappear. Numerical studies suggest that the disappearance of the kernel peak happens when the sampling intervals are having a large variance relative to the mean sampling interval. Figure(8.1) shows the shape of the kernel function $K_\Delta(f, \lambda)$ for various different f, under Gamma$(1, 4)$ sampling scheme. Such a sampling scheme has the same average sampling interval $\Delta = 0.25$ as the Gamma$(3, 12)$ renewal sampling scheme considered in previous examples, but the sampling interval certainty ratio is significantly smaller at $r = 1$ as compared to the corresponding ratio of the Gamma$(3, 12)$ renewal sampling scheme. The plot shows that with this small certainty ratio, the peak of the renewal kernel disappear for some f between $f = 1.5$ and $f = 2$.

Hence the half-width d of the peak of the kernel function $K_\Delta(f, \lambda)$ may not be defined for f relatively large, indicating that the local average component $E_d(f)$ of $S_Y^\Delta(f)$ of the sampled process may not exist.

8.3.5 Whitening Through Renewal Sampling

The renewal sampling scheme will have a whitening effect, in the sense that the sampled process will behave like a white noise sequence if the spectral features of the underlying continuous time process are located at relatively high frequencies. To give a more rigorous arguments about this whitening effect, consider a family of spectral densities $S_{\lambda_0}(\lambda)$ defined through

$$S_{\lambda_0}(\lambda) = \frac{1}{2}\left[S_X(\lambda - \lambda_0) + S_X(\lambda + \lambda_0)\right], \tag{8.35}$$

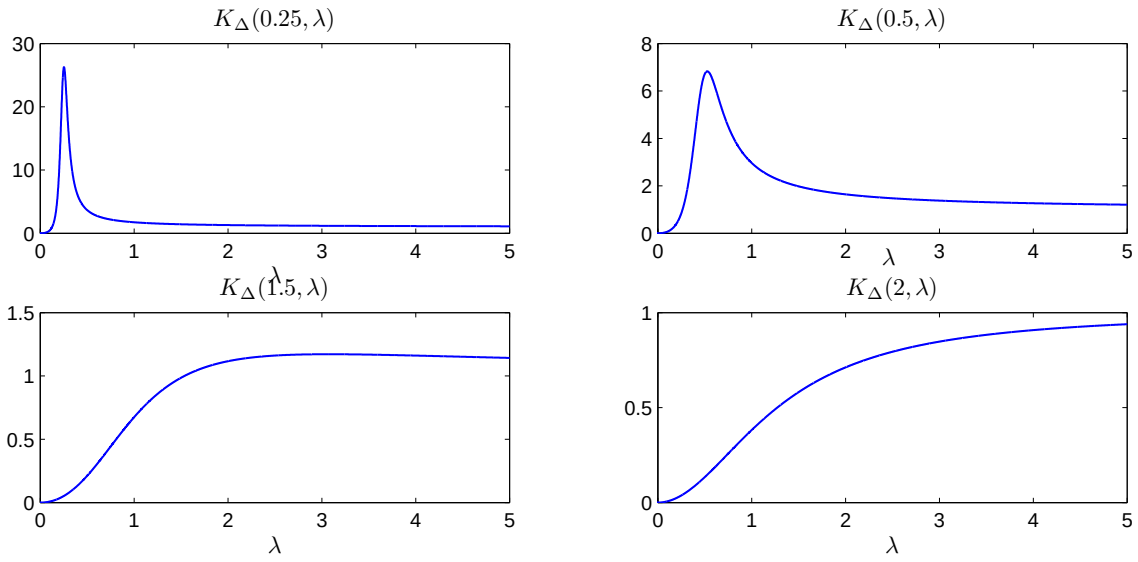

Figure 8.1: Shape of the renewal kernel under a Gamma$(1,4)$ renewal sampling scheme, in which the mean sampling interval $\Delta = 0.25$, and the sampling certainty ratio $r = 1$.

where $S_X(\lambda)$ is any valid spectral density function (i.e. being positive and integrable). consequently $S_{\lambda_0}(\lambda)$ is constructed by shifting the spectral features of $S_X(\lambda)$ by a frequency displacement λ_0. Note that the same procedure was in Section(2.3.3) to introduce the Oscillatory Matérn process.

Then the spectral density $S_Y^\Delta(f;\lambda_0)$ of the sampled process corresponding to this family of continuous time spectral densities is immediately given by

$$\begin{aligned} S_Y^\Delta(f;\lambda_0) &= \Delta \int_{-\infty}^{\infty} S_{\lambda_0}(\lambda) K_\Delta(f,\lambda) d\lambda \\ &= S_{\lambda_0}(0) I_0(f) + \Delta \int_{-\infty}^{\infty} S_{\lambda_0}(\lambda) \tilde{K}_\Delta(f,\lambda) d\lambda, \end{aligned} \tag{8.36}$$

where the definition of $K_\Delta(f,\lambda)$ and $\tilde{K}_\Delta(f,\lambda)$ are given by Equation(5.22) and (5.27) respectively. Being a spectral density function, $S_X(\lambda)$ must be symmetric about

$\lambda = 0$, and hence we must have

$$\begin{aligned}\lim_{\lambda_0 \to \infty} S_{\lambda_0}(0) &= \lim_{\lambda_0 \to \infty} \frac{1}{2} \left[S_X(-\lambda_0) + S_X(\lambda_0) \right] \\ &= \lim_{\lambda_0 \to \infty} S_X(\lambda_0) \\ &= 0, \end{aligned} \tag{8.37}$$

On the other hand, to show the limiting behaviour of the integral in Equation(8.36) as λ_0 tends to infinity, we first note that for any f, the kernel function $\tilde{K}_\Delta(f, \lambda)$ is continuous in λ and are globally bounded (although the upper bound is not uniform in f). Hence we may apply the Dominated Convergence Theorem [143] to derive that for a fixed f we have

$$\begin{aligned}&\lim_{\lambda_0 \to \infty} \int_{-\infty}^{\infty} S_{\lambda_0}(\lambda) \tilde{K}_\Delta(f, \lambda) d\lambda \\ &= \lim_{\lambda_0 \to \infty} \frac{1}{2} \int_{-\infty}^{\infty} S_X(\lambda - \lambda_0) \tilde{K}_\Delta(f, \lambda) d\lambda + \lim_{\lambda_0 \to \infty} \frac{1}{2} \int_{-\infty}^{\infty} S_X(\lambda + \lambda_0) \tilde{K}_\Delta(f, \lambda) d\lambda \\ &= \lim_{\lambda_0 \to \infty} \frac{1}{2} \int_{-\infty}^{\infty} S_X(\lambda) \tilde{K}_\Delta(f, \lambda + \lambda_0) d\lambda + \lim_{\lambda_0 \to \infty} \frac{1}{2} \int_{-\infty}^{\infty} S_X(\lambda) \tilde{K}_\Delta(f, \lambda - \lambda_0) d\lambda \\ &= \frac{1}{2} \int_{-\infty}^{\infty} S_X(\lambda) \lim_{\lambda_0 \to \infty} \tilde{K}_\Delta(f, \lambda + \lambda_0) d\lambda + \frac{1}{2} \int_{-\infty}^{\infty} S_X(\lambda) \lim_{\lambda_0 \to \infty} \tilde{K}_\Delta(f, \lambda - \lambda_0) d\lambda \\ &= \int_{-\infty}^{\infty} S_X(\lambda) d\lambda. \end{aligned} \tag{8.38}$$

Combining the limiting result of Equation(8.37) and the above equation together, we immediately have

$$\lim_{\lambda_0 \to \infty} S_Y(f; \lambda_0) = \Delta \int_{-\infty}^{\infty} S_X(\lambda) d\lambda \tag{8.39}$$

for any fixed f. In other words, as the frequency features of the continuous time process moves to higher frequencies, the spectral density of the sampled process tends to become flat, and hence resembles a white noise sequence.

Bibliography

[1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions, Graphs, and Mathematical Tables*. Dover, New York, 1965.

[2] H.M. Adorf. Interpolation of irregularly sampled data series-a survey. In R.A. Shaw, H.E. Payne, and J.J.E. Hayes, editors, *Astronomical Data Analysis Software and Systems IV*, pages 460–463. ASP Conference Series 77, 1995.

[3] W.H. Bakun and A. Eisenberg. Fourier integrals and quadrature-introduced aliasing. *Bulletin of Seismological Society of America*, 60(4):1291–1296, 1970.

[4] M. Barbato and J.P. Conte. Spectral characteristics of non-stationary random processes: Theory and applications to linear structural models. *Probabilistic Engineering Mechanics*, 23(4):416–426, 2008.

[5] J. Belcher, J. Hampton, and W. Tunnicliffe. Parameriztion of cont autoregressive models for irregularly sampled time series data. *Journal of the Royal Statistical Association, series B*, 56(1):141–155, 1994.

[6] J. Bendat and A.G. Piersol. *Random Data: Analysis and Measurement Procedures*. Wiley, New York, 1986.

[7] Jan Beran. *Statistics for long-memory processes*, volume 61. CRC Press, 1994.

[8] F.J. Beutler. Alias free randomly times sampling of stochastic processes. *IEEE Trans. Inform. Theory*, IT-16:147–152, 1970.

[9] F.J. Beutler and O.A. Leneman. Random sampling of random processes: Stationary point processes. *Information and Control*, 9, 1966.

[10] K.Y.R. Billah and M. Shinozuka. Numerical method for colored-noise generation and its application to a bistable system. *Physical Review A*, 42(12):7492–7495, 1990.

[11] K.Y.R. Billah and M. Shinozuka. Fluctuation of dynamic pressure and white-noise assumption in flow-induced vibration problems. *Journal of Sound and Vibration*, 147:179–183, 1991.

[12] P. Biron, A. Roy, and J.L. Best. A scheme for resampling, filtering, and subsampling unevenly spaced laser doppler anemometer data. *Mathematical geology*, 27(6):731–748, 1995.

[13] Peter Bloomfield. Trends in global temperature. *Climatic change*, 21(1):1–16, 1992.

[14] B. Bojanic. An estimate of the rate of convergence for fourier series of functions of bounded variation. *Publications De L'institut Mathematique*, 26(40):57–60, 1979.

[15] R. Bracewell. *The Fourier Transform and its Applications*. McGraw-Hill, third edition edition, 1999.

[16] R.C. Bradley. Basic properties of strong mixing conditions. Technical report, DTIC Document, 1985.

[17] E.O. Brigham. *The Fast Fourier Transform and its Applications*. Prentice Hall, 1988.

[18] D.R. Brillinger. *Time Series Data Analysis and Theory*. SIAM, 1965.

[19] D.R. Brillinger. The spectral analysis of stationary interval functions. In *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob., Vol. 1*. University of California Press, 1972.

[20] P.J Brockwell. Continuous-time arma processes. In D.N.Shanbhag and C.R.Rao, editors, *Handbook ofStatistics 19 - Stochastic Processes: Theory and Methods*, pages 249–276. Elsevier, 2001.

[21] P.J. Brockwell and J. Hannig. Carma(p,q) generalized random processes. *Journal of Statistical Planning and Inference*, 140(12):3613 – 3618, 2010.

[22] P.J. Brockwell and Davis R.A. *Time Series: Theory and Methods*. Springer, second edition edition, 1991.

[23] M. Charlot and M. Rachdi. On the statistical properties of a stationary process sampled by a stationary point process. *Statistics & Probability Letters*, 78:456–462, 2008.

[24] A. Contreras-Cristan, E. Gutierrez-Pena, and S.G. Walker. a note on whittle's likelihood. *Communications in Statistics - Simulation and Computation*, 35(4):857–875, 2006.

[25] P.F. Craigmile. Simulating a class of stationary gaussian processes using the davis-harte algorithm, with application to long memory processes. *Journal of Time Series Analysis*, 24(5):505–511, 2003.

[26] H. Cramer and M.R. Leadbetter. *Stationary and Related Stochastic Processes - Sample Function Properties and Their Applications*. John Wiley Sons, Inc, New York, 1967.

[27] S.H. Crandall and W.D. Mark. *Random Vibration in Mechanical Systems*. Academic Press, 1963.

[28] D.J. Daley and D. Vere-Jones. A summary of the theory of point processes. In P.A.W. Lewis, editor, *Stochastic Point Processes*. New York: Wiley, 1972.

[29] J. Davidson. *Stochastic Limit Theory - An Introduction for Econometricians*. Oxford University Press, 1994.

[30] M.W. Davis. Production of conditional simulation via the lu-triangular decomoposotion of the covariance matrix. *Mathematical Geoscience*, 19(2):91–98, 1987.

[31] P.F. Davis and P. Rainowitz. *Methods of Numerical Integration*. Academic Press, Inc, second edition edition, 1984.

[32] R.B. Davis and D.S. Harte. Tests for hurst effect. *Biometrika*, 74(1):95–101, 1987.

[33] A. Dembo, C.L. Mallows, and L.A. Shepp. Embedding nonnegative definite toeplitz matrices in nonnegative definite circulant matrices, with application to covariance estimation. *IEEE Transactions on Information Theory*, 35(6):1206–1212, 1989.

[34] G. Deodatis. Simulation of ergodic multivariate stochastic processes. *Journal of engineering mechanics*, 122(8):778–787, 1996.

[35] C.R. Dietrich and G.N. Newsam. Fast and exact simulation of stationary gaussian processes through circulant embedding of the covariance matrix. *SIAM: Journal on Scientific Computing*, 18(4):1088–1107, July 1997.

[36] J. Durbin. The fitting of time-series models. *Revue de l'Institut International de Statistique*, pages 233–244, 1960.

[37] E. Erdogan, S. Ma, A. Beygelzimer, and I. Rish. Statistical models for unequally spced time series. In *Proceedings of the Fifth SIAM International Conference on Data Mining*, 2004.

[38] J.N. Franklin. Numerical simulation of stationary and nonstationary gaussian random processes. *SIAM Review*, 7(1):68–80, 1965.

[39] M.W. Frazier. *An introduction to wavelets through linear algebra*. Springer Verlag, New York, 1999.

[40] C. Gasquet and P. Witomski. *Fourier Analysis and Applications - Filtering, Numerical Computation, Wavelets*. Springer, 1999.

[41] M. Gaster and J.B. Roberts. Spectral analysis of randomly sampled signals. *J.Inst.Maths Applics*, 15:195–216, 1975.

[42] E. Gaughan. *Introduction to Analysis*. Brooks/Cole Publishing Company, fifth edition edition, 1997.

[43] John Geweke and Susan Porter-Hudak. The estimation and application of long memory time series models. *Journal of time series analysis*, 4(4):221–238, 1983.

[44] T. Gneiting. On the derivatives of radial positive definite functions. *Journal ofMathematical Analysis and Applications*, 236:86–93, 1999.

[45] T. Gneiting. Compactly supported correlation functions. *Journal of Multivariate Analysis*, 83(2):493–508, January 2002.

[46] T. Gneiting, H. Sevcikova, D. Percival, M. Schlather, and Y. Jiang. Fast and exact simulation of large gaussianlattice systems in r^2: Exploring the limits. *Journal of Computational and Graphical Statistics*, 15(3):483–501, 2006.

[47] M. Grigoriu. On the spectral representation method in simulation. *Probabilistic Engineering Mechanics*, 8:75–90, 1993.

[48] M. Grigoriu. *Stochastic Calculus - Applications in Science and Engineering*. Birkhauser, Boston, 2002.

[49] E.J. Hannan. The asymptotic theory of linear time-series models. *Journal of Applied Probability*, 10(1):130–145, 1995.

[50] Y. Hosoya. *Estimation problems on Stationary Time-Series Models*. PhD thesis, Yale University, 1974.

[51] Y. Hosoya and M. Taniguchi. A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics*, 10(1):132–153, 1982.

[52] B. Hu and W. Schiehlen. On the simulation of stochastic processes by spectral representation. *Probabilistic engineering mechanics*, 12(2):105–113, 1997.

[53] B. Hu and W. Schiehlen. On the simulation of stochastic processes by spectral representation. *Probabilistic Engineering Mechanics*, 12(2):105–113, 1997.

[54] I.A. Ibragimov. On maximum likelihood estimation of parameters of the spectral density of stationary time series. *Theory of Probability and its Applications*, 12:115–119, 1967.

[55] A. Iserles. On the numerical quadrature of highly oscillating integrals i: Fourier transforms. *IMA Journal of Numerical Analysis*, 24:365–391, 2004.

[56] A. Iserles. On the numerical quadrature of highly oscillating integrals ii: Irregular oscillators. *IMA Journal of Numerical Analysis*, 25:25–44, 2005.

[57] G.E. Johnson. Constructions of particular random processes. *Proceedings of the IEEE*, 82:270–285, 1994.

[58] R.H. Jones. Fitting a continuous time autoregression to discrete data. In D.F.Findley, editor, *Applied Time Series Analysis II*, pages 651–682. Academic Press, 1981.

[59] R.H. Jones and L.M. Ackerson. Serial correlationin unequally spaced longitudinal data. *Biometrika*, 77:721–732, 1990.

[60] A.G. Journel and C.T. Huijbregts. *Mining Geostatistics*. Academic Press, 1978.

[61] T. Kailath. Theorem of I. Schur and its impact on modern signal processing. In I. Gohberg, editor, *Schur Methods in Operator Theory and Signal Processing I*, pages 9–30. Birkhauser-Verlag, Basel, 1986.

[62] S.M. Kay. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.

[63] S.M. Kay. *Modern Spectral Estimation - Theory and Application*. Prentice-Hall, 1999.

[64] P.R. King and P.J Smith. Generation of correlated properties in heterogeneous porous media. *Mathematical Geosciences*, 20(7):863–877, 1988.

[65] L.H. Koopmans. *The Spectral Analysis of Time Series*. Academic Press, 1974.

[66] K. Lange. *Numerical Analysis for Statisticians*. Springer, 2010.

[67] O.A. Leneman. Random sampling of random processes: Impulse processes. *Information and Control*, 9, 1966.

[68] N. Levinson. The wiener rms (root mean square) error criterion in filter design and prediction. *Journal of Mathematical Physics*, 25(4):261–278, 1947.

[69] J. Liang, S.R. Chaudhuri, and M. Shinozuka. Simulation of nonstationary stochastic processes by spectral representation. *Journal of Engineering Mechanics*, 133(6):616–627, 2007.

[70] K.S.. Lii and E. Masry. Model fitting for continuous-time stationary processes from discrete-time data. *Journal of Multivariate Analysis*, 41, 1992.

[71] K.S.. Lii and E. Masry. On the selection of random sampling schemes for the spectral estimation of continuous time processes. *Journal of Time Series Analysis*, 16(3), 1994.

[72] K.S. Lii and E. Masry. Spectral estimation of continuous-time stationary processes from random sampling. *Stochastic Processes and Their Applications*, 52, 1994.

[73] N.R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and space science*, 39(2):447–462, 1976.

[74] M.C. Lui. *Spectral Estimation of Continuous Parameter Processes from Randomly Timed Observations*. PhD thesis, University of Carlifornia, 1974.

[75] F. Ma, M. S. Wei, and W. H. Mills. Correlation structuring and the statistical analysis of steady-state groundwater flow. *SIAM journal on scientific and statistical computing*, 8(5):848–867, 1987.

[76] B.B. Mandelbrot and J.R. Wallis. Computer experiments with fractional gaussian noises, part 1: Averages and variances. *Water Resources Research*, 5(1):228–241, 1969.

[77] R.J. Martin. *Irregularly Sampled Signals:Theories and Techniques for Analysis*. PhD thesis, UCL, 1998.

[78] E. Masry. Random sampling and reconstruction of spectra. *Information and Control*, 19, 1971.

[79] E. Masry. Alias-free sampling: An alternative conceptualization and its applications. *IEEE Transactions on Information Theory*, IT-24(3), 1978.

[80] E. Masry. Poisson sampling and spectral estimation of continuous time processes. *IEEE Transactions on Information Theory*, IT-24(2), 1978.

[81] E. Masry. Discrete-time spectral estimation of continuous-time processes - the orthogonal series method. *Ann. Statist*, 8, 1980.

[82] E. Masry. Random sampling of deterministic signals: Statistical analysis of fourier transform estimates. *Signal Processing, IEEE Transactions on*, 54(5):1750–1761, 2006.

[83] E. Masry, D. Klamer, and C. Mirabile. Spectral estimation of continuous-time proceses: Performence comparision between periodic and poisson sampling schemes. *IEEE Transactions on Automatic Control*, 23(4):679–685, 1978.

[84] E. Masry and C. Lui. Discrete-time spectral estimation of continuous time process - a new consistent estimate. *IEEE Transactions on Information Theory*, IT-22(3), 1975.

[85] E. Masry and A. Vadrevu. Random sampling estimates of fourier transforms: antithetical stratified monte carlo. *Signal Processing, IEEE Transactions on*, 57(1):194–204, 2009.

[86] B. Matern. *Spatial Variation*. Springer-Verlag, Berlin, second edition edition, 1960.

[87] K.S. Miller. *Complex Stochastic Processes: An Introduction to Theory and Application*. Addison-Wesley Pub. Co., 1974.

[88] C. Moler and C. Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, 45(1):1–46, 2003.

[89] J.F. Monahan. A note on enforcing stationarity in autoregressive-moving average models. *Biometrika*, 71(2):403–404, 1984.

[90] M.I. Moore, P.J. Thomson, and T.G.L. Shirtcliffe. Spectral analysis of ocean profiles from unequally spaced data. *Journal of Geophysical Research: Oceans (1978–2012)*, 93(C1):655–664, 1988.

[91] G.N. Newsam. On the distribution of the eigenvalues of discretizations of a compact operator. In *Proc. Center for Mathematical Analysis, Vol. 17: Special Program on Inverse Problems*. Australian National Univeristy, Canberra, Australia, 1988.

[92] E. Parzen. *Proceedings of Time Series Analysis of Irregularly Observed Data, Lecture Notes in Statistics 25*. New York:Springer Verlag, 1983.

[93] D.B. Percival. Simulating gaussian random processes with specified spectra. *Computing Science and Statistics*, 24:534–538, 1992.

[94] D.B. Percival. Exact simulation of complex-valued gaussian stationary processes via circuant embedding. *Signal Processing*, 86(7):1470–1476, 2006.

[95] D.B. Percival and A.T. Walden. *Spectral Analysis for Phisical Applications - Multitaper and Conventional Univariate Techniques*. Cambridge University Press, 1993.

[96] M.S. Phadke and S.M. Wu. Modeling of continuous stochastic processes from discrete observations with application to sunspots data. *Journal of the American Statistical Association*, 69(346):325–329, 1974.

[97] D.T. Pham and A. Le Breton. Levinson-durbin-type algorithms for continuous-time autoregressive models and applications. *Mathematics of Control, Signals and Systems*, 4(1):69–79, 1991.

[98] M.B. Priestley. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 204–237, 1965.

[99] M.B. Priestley. Power spectral analysis of non-stationary random processes. *Journal of Sound and Vibration*, 6(1):86–97, 1967.

[100] M.B. Priestley. *Spectral Analysis and Time Series*. Academic Press, 1982.

[101] M.B. Priestley. *Non-Linear and Non-Stationary Time Series Analysis*. Academic Press, 1988.

[102] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[103] J.B. Read. On the sharpness of weyl's estimates for eigtnvalues of smooth kernels. *SIAM Journal on Mathematical Analysis*, 19(3):627–631, 1988.

[104] S.I. Resnick. *A Probability Path*. Birkhauser, 1999.

[105] S.O. Rice. Mathematical analysis of random noises. In N. Was, editor, *Selected Papers on Noise and Stochastic Processes*. New York:Dover, 1954.

[106] A. Rivoira and G.A. Fleury. A consistent nonparametric spectral estimator for randomly sampled signals. *IEEE Transactions on Signal Processing*, 52(9):2383–2395, 2004.

[107] P.M. Robinson. Continuous model fitting from discrete data. In D.R.Brillinger and G.C.Tiao, editors, *Institute of Mathematical Statistics Report on Directions in Time Series*, pages 263–278. Institute of Mathematical Statistics, 1978.

[108] M. Rosenblatt. *Stationary Sequences and Random Fields*. Birkhauser, 1985.

[109] D.B. Rubin. *Statistical Analysis With Missing Data*. Wiley, New York, 2002.

[110] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, third edition edition, 1976.

[111] J.D. Scargle. Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835–853, 1982.

[112] G.A.F. Seber. *A Matrix Handbook for Statisticians*. Wiley, 2008.

[113] H.S. Shapiro and Silverman R.A. Alias-free sampling of random noise. *J.Soc.Indust.Appl.Math*, 8(2):225–248, 1960.

[114] M. Shinozuka. Monte carlo solution of structural dynamics. *Computers and Structures*, 2(5-6):855–874, 1972.

[115] M. Shinozuka and G. Deodatis. Simulation of stochastic processes by spectral representation. *Applied Mechanics Reviews*, 44(4):191–203, 1991.

[116] M. Shinozuka and C.M. Jan. Digital simulation of random processes and its applications. *Journal of Sound and Viberation*, 25(1):111–128, 1972.

[117] A.N. Shiryaev. *Probability*. Springer, 1996.

[118] L.W. Smith. Asymptotic renewal theorems. *Proceedings of the Royal Society of Edinburgh. Section A. Mathematical and Physical Sciences*, 64:9–48, 1954.

[119] L.W. Smith. Extensions of a renewal theorem. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51:629–638, 1955.

[120] E. Stein. *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton University Press, 1971.

[121] E.M. Stein and Shakarchi R. *Fourier analysis: an Introduction*. Princeton University Press, 2003.

[122] M.L. Stein. *Interpolationof Spatial Data - Some Theory of Kriging*. Springer, 1999.

[123] M.L. Stein. Fast and exact simulation of fractional brownian surfaces. *Journal of Computational and Graphical Statistics*, 11(3):527–531, 2002.

[124] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Springer, third edition edition, 2002.

[125] T.C. Sun and M. Chaika. On simulation of a gaussian stationary process. *Journal of Time Series Analysis*, 18(1):79–93, 1995.

[126] M. Taniguchi. On estimation of the integrals of the fourth order spectral density. *Biometrika*, 69:117–122, 1982.

[127] M. Taniguchi and Y. Kakizawa. *Asymptotic Theory of Statistical Inference for Time Series*. Springer-Verlag New York, 2000.

[128] Murad S Taqqu. Fractional brownian motion and long-range dependence. *Theory and applications of long-range dependence*, pages 5–38, 2003.

[129] A. Tarczynski and N. Allay. Spectral analysis of randomly sampled signals: Suppression of aliasing and sampler jitter. *Signal Processing, IEEE Transactions on*, 52(12):3324–3334, 2004.

[130] P.J. Thomson and P.M. Robinson. Estimation of second-order properties from jittered time series. *Annals of the Institute of Statistical Mathematics*, 48(1):29–48, 1996.

[131] H. Tómasson. Some computational aspects of gaussian modelling. Economics Series 274, Institute for Advanced Studies, September 2011.

[132] A.F.B Tompson, R. Ababou, and L.W. Gelhar. Implementation of the three-dimansional turning bands field generator. *Water Resources Research*, 25:2227–2243, 1989.

[133] W.F. Trench. An algorithm for the inversion of finite toeplitz matrices. *Journal of the Society for Industrial & Applied Mathematics*, 12(3):515–522, 1964.

[134] A.M. Walker. Asymptotic peroperties of least-square estimates of parameters of the spectrum of a stationary nondeterministic time seires. *Journal of the Australian Mathematical Society*, 4(3):363–384, 1964.

[135] J.J. Warnes. A sensitivity analysis for universal kriging. *Mathematical Geology*, 18:653–676, 1986.

[136] P. Whittle. Estimation and information in stationary time series. *Ark. Math*, 2(5):423–434, 1953.

[137] P. Whittle. Gaussian estimation in stationary time series. *Bull. Inst. Internat. Statist.*, 39:105–129, 1962.

[138] P.H. Wirshing and A.M. Shehata. Fatigue under wide band random stress using the rainflow method. *Journal of the Structureal Division*, 106(7):1593–1607, 1980.

[139] D.E. Wold. On smooth estimation of renewal density.

[140] A. Wood and G. Chan. Simulation of stationary gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics*, 3(4):409–432, 1994.

[141] J.N. Yang. Simulation of random envelope processes. *Journal of Sound and Viberation*, 21(1):73–85, 1972.

[142] J.N. Yang. On the normality and accuracy of simulated random processes. *Journal of Sound and Vibration*, 26(3):417–428, 1973.

[143] J. Yeh. *Real Analysis - Theory of Measure and Integration*. World Scientific, second edition edition, 2006.

[144] A. Zygmund. *Trigonometric Series*. Cambridge University Press, third edition edition, 2002.

List of Figures

List of Tables