Bayesian nonparametric modelling of covariance functions, with application to time series and spatial statistics

by

Gudmund Horn Hermansen

THESIS

presented for the degree of

MASTER OF SCIENCE

under the supervision of

Professor Nils Lid Hjort

Statistics Division
Department of Mathematics, University of Oslo

June 2008

Avdeling for statistikk
Matematisk institutt, Universitetet i Oslo

Preface

First of all I would like to thank my supervisor and mentor, professor Nils Lid Hjort, who has allowed me to work independently and has always given me the necessary amount of pushes in the right direction so that I have been able to find solutions to most of the different problems involved in this work. Without his help and support this thesis would not have been what it has finally become.

I would also like to thank my fellow students, my friends and my family, who have been of great help solving all kinds of problems with the writing of my master thesis. Most of all I would like to thank my beloved girlfriend Siv, who has been of great support through all these long and demanding days of work.

To work on this thesis has been an exciting and challenging task, I feel that I have learned a lot and developed much during this time. My interest in and respect for mathematical and applied statistics has increased and this makes me sincerely wish to work further in this field in the future.

Oslo, June 2008,
Gudmund Horn Hermansen

Contents

CHAPTER 1

Introduction and summary

1. Thesis overview

In this thesis we will attempt to apply the nonparametric Bayesian modeling strategy for the estimation of the dependency structure for stationary time series and for second order stationary random fields.

It is quite common in spatial models to use a parametric approach when modeling the dependency structure. This usually consists of choosing a class of parametric functions, based on some a priori information about the phenomenon under study, and the estimating of the unknown parameters in the model from a set of observations. In this thesis we will only consider spatial models that are second order stationary and that also have isotropic covariance function. It is not uncommon to make rather strong model assumptions in spatial models to compensate for the incomplete observations. The second order stationary supposition is in many situations plausible and a quite usual assumption, see **?** or **?**. In such spatial models the class of exponential covariance functions are by many regarded as the most useful choice of covariance functions in $\mathbb{R}^1$, cf. **?**. For second order stationary spatial models on $\mathbb{R}^d$, where $d > 1$, with isotropic covariance functions, the class of functions known as Matérn is considered the natural choice of parametric covariance functions, see **?**, **?** or **?**. The Bayesian strategy for such parametric models will involve placing a prior distribution on the set of unknown parameters. For simple models this set will typically consist of three parameters, say (μ, σ, ρ), which represent expectation, variance and correlation.

Various journal articles have dealt with nonparametric modeling of trend functions in spatial models, also from the Bayesian perspective, see **?**. These approaches are actually semiparametric, since they use a nonparametric model for the trend and combine this with a typically parametric model for the covariance function. Rather fewer attempts have been made to model both the trend and the covariance function nonparametrically, and I am not aware of a single journal article that deals with such an approach from the Bayesian perspective.

It is perhaps more common to use the nonparametric strategy when estimating the dependency in time series models, since this can easily be done from the nonparametric estimation of the power spectrum. We will restrict ourselves to only consider the class of stationary Gaussian time series. For such models **?** has shown that a large class of parametric covariance functions with finite number of parameters falls into the class known as "locally asymptotic normal" considered by Le Cam (1960 - 1970). This essentially means that the estimated parameters in such parametric covariance models satisfy similar asymptotic properties as the maximum likelihood estimators.

The reason we will use the nonparametric Bayesian modeling scheme is that we would like to have both the advantage of the flexible nonparametric models and also the benefit of a Bayesian

approach. By using nonparametric Bayesian models we are avoiding the problem of making critical dependency assumptions by the possibility of choosing a wrong model.

The main goal of this thesis is not to change the world, but rather to suggest an alternative solution to how we can model the dependency structure in some time series and spatial models. The idea we present here will become easy to use and the a priori information may be included in the model in a natural way. We will have some extra focus on the class of exponential covariance functions since they are often used and possess some especially nice properties regarding the construction of suitable prior distribution. We will also derive some of the large-sample properties for the estimators in the stationary Gaussian time series models and show that these satisfy exactly the properties we would hope for.

As mentioned, the idea is to use the nonparametric Bayesian approach to model the covariance function in some time series and in spatial models. It is not obvious how this should proceed or how one may attack such a problem since we want to be able to protect ourselves from making covariance functions that are not valid, where valid means positive definite.

The solution to the positive definiteness problem is to use the connection between the spectral density and the covariance function for random fields. The methods will turn out to be partly similar to the nonparametric Bayesian distribution function estimation methods presented in the articles of **?** and **?**. From **?** we know that the covariance function for a spatial model on $\mathbb{R}^d$ and the spectral density are connected with each other through the Fourier transformation pair

$$C(\boldsymbol{h}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(i\boldsymbol{u}^T\boldsymbol{h}) f(\boldsymbol{u})\, d\boldsymbol{u}, \tag{1.1}$$

and

$$f(\boldsymbol{u}) = (2\pi)^{-d} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(-i\boldsymbol{u}^T\boldsymbol{h}) C(\boldsymbol{h})\, d\boldsymbol{h},$$

for vectors $\boldsymbol{u}, \boldsymbol{h} \in \mathbb{R}^d$, see also **?** and **?**. It is clear from equation (1.1) that if we want to place a prior distribution on the set of covariance functions it is equivalent to place a corresponding prior distribution on the set of spectral densities. The main result we will use in this thesis, that also will ensure that the strategy above is acceptable, is known as Bochner's theorem. Bochner's theorem states that a function $C(\boldsymbol{h})$ is positive definite if and only if it has a spectral representation given by

$$C(\boldsymbol{h}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(i\boldsymbol{u}^T\boldsymbol{h}) dF(\boldsymbol{u}), \tag{1.2}$$

where $\boldsymbol{u}, \boldsymbol{h} \in \mathbb{R}^d$ and $dF(\boldsymbol{u})$ is a bounded symmetric measure. In the case of real valued random fields with isotropic covariance function, representation (1.2) from Bochner's theorem simplifies into functions which satisfy the spectral representation

$$C(\boldsymbol{h}) = 2^{d/2-1}\Gamma(d/2) \int_0^{\infty} (uh)^{-(d/2-1)} J_{d/2-1}(uh)\, dF(u), \quad \text{where } h = ||\boldsymbol{h}|| \tag{1.3}$$

and $J_\nu(t)$ is the Bessel function of the first kind and F is a nondecreasing function that is bounded on $[0, \infty)$ and satisfy $F(0) = 0$, see **?**. This means that in order to place a prior distribution on the space of isotropic covariance functions it is sufficient to place a prior distribution on the set of spectral measurers F that are nondecreasing, bounded on $[0, \infty)$ and satisfy $F(0) = 0$, i.e. viewing F as a positive increment process bounded on $[0, \infty)$. Let F be a Lévy process (in our context, this is a process with independent and nonnegative increments) that satisfies

$\Pr\{F(\infty) < \infty\} = 1$, then F is a positive and independent increment process that is bounded with probability 1 on $[0, \infty)$ and can therefore work as our prior distribution on the set of spectral measures. Also from Bochner's theorem we know that every random covariance function that is constructed in this manner will become valid.

To be more specific we will throughout most of the thesis and in the examples assume that F is a Gamma process, meaning that the F is an independent increment process where

$$dF(u) \sim \text{Ga}(\alpha(u), \beta(u))$$

and $\text{Ga}(\alpha, \beta)$ is the Gamma distribution with shape parameter α and rate parameter β, see **?**, **?** or **?** for definition and examples of the use of the Gamma processes. The main reason for the use of the Gamma process is the uncomplicated expressions for the expectation and variance. This will make it straightforward and easy to determine how we should specify the parameters in the process in such a way that it will reflect our a priori beliefs. Let F be a Gamma process with parameters $\alpha(u) = b(u)f_0(u)\,du > 0$ and $\beta(u) = b(u) > 0$, where $f_0(u)$ is the spectral density function that corresponds to our favorite and a priori guess for the isotropic covariance function $C_0(h)$. Then from the properties of the Gamma distribution we have that the expected covariance function is determine by

$$E[C(\boldsymbol{h})] = 2^{d/2-1}\Gamma(d/2)\int_0^\infty (uh)^{-(d/2-1)} J_{d/2-1}(uh)\, E[dF(u)] = C_0(\boldsymbol{h}), \quad \text{where } h = ||\boldsymbol{h}||$$

and $h \in [0, \infty)$, since we may interchange mean value operation and integral here, by the Fubini theorem. Note that in the simple stationary time series model the necessary and sufficient require (1.3) simplifies to finding a positive nondecreasing function $F(u)$ defined on $[0, \pi]$, where $F(0) = 0$ and $F(\pi) < \infty$, such that

$$C(h) = 2\int_0^\pi \cos(uh)\, dF(u),$$

for $h = 0, \pm 1, \pm 2, \ldots$, see **?**.

The thesis is divided into two main parts, in Chapter 2 and Chapter 3 we will consider stationary Gaussian time series with unknown covariance function. In Chapter 4 we will try to extend the ideas from Chapter 2 to the spatial models. To be specific we will consider second order stationary Gaussian random fields over a continuous domain $D \subset \mathbb{R}^d$ with unknown and isotropic covariance functions.

In Section 1 of Chapter 2 we will introduce the main concepts and theory we will need regarding stationary time series, most of this are from the books **?** and **?**. In Section 1.1 we will establish the connection between the power spectrum and the covariance function. Section 1.2 will discuss how we can use the periodogram as a nonparametric estimator for the unknown power spectrum and therefore also the covariance function. Further in Section 1.3 we will establish some properties of the spectral measure, the integrated power spectrum, and extend some of the results from **?** in such a way that it will fit our nonparametric Bayesian framework.

In Section 2.1 we will introduce the main concepts of nonparametric Bayesian modeling and give a short introduction to distribution function estimation. In section 2.2 and 2.3 we will show how we can define a prior distribution on the set of valid covariance functions by viewing the spectral measure as a random process, especially a Gamma process, and show how we can make posterior inference through the use of Markov chain Mote Carlo simulations.

We will in Chapter 3 use the "principal part" given in **?** as an approximation to the full multivariate Gaussian log-likelihood and show how this can be used to obtain asymptotic inference about the spectral measure and the covariance function. We will also show that we obtain similar asymptotic result if we use the discrete version of the "principal part" approximation, this is often referred to as the Whittle approximation, to honor P. Whittle who first suggested this approximation in the early fifties.

Chapter 4 will be a natural extension of the ideas of Chapter 2 and will also be a bit shorter since some of the general ideas and concepts are already discussed in detail in Chapter 2. The main reason why this Chapter 4 is shorter is that we do not have the nice approximation for the multivariate Gaussian log-likelihood as we did in Chapter 3. In particular, various large-sample results reached in Chapter 3 become much more complicated in the framework of Chapter 4, therefore some of the topics dealt with of Chapter 3 will not be pursued in Chapter 4.

Section 1 will introduce the basic theory regarding spatial data analysis and random fields with some extra attention on spatial prediction and Kriging. In Section 2 we will again show how we can use the Gamma processes to construct prior distributions for the isotropic covariance functions through the spectral measure. We will also show that two special classes of covariance functions possess some especially nice properties and we will complete the section with an extensive example as an illustration of the main ideas.

Appendix A is a short introduction to the Metropolis-Hastings algorithm, which is the MCMC sampler we will use to make approximative inference about the posterior spectral measure. There are some calculations and general theory in Appendix B and in Appendix C we give a short introduction and a list of all the functions and routines written in R. Even though it is quite common for master thesis, the R code will not be included in the appendix. The main reason is that this would add another hundred pages to the paper. Those who are interested in a copy of source files that contains all the routines, functions and examples used in the thesis may send me an email at `gudmunhh@student.matnat.uio.no`.

2. Conclusions

As already mentioned in the previous section the intention of this thesis is to relax some of the harsh conditions that are commonly assumed for stationary time series and in spatial data analysis. The methods we have applied are outside the normal modeling framework. We have nevertheless succeeded in the sense that we have been able to define a reasonable model, established easy to use and meaningful prior distributions and we have also managed to show how we can obtain posterior inference by the use of simulations. These three steps, model, prior and posterior inference through simulations, can be thought of as the minimal demand for Bayesian analysis. In many situations these stages are sufficient and there exist several articles that deals with Bayesian statistics that do not continue the discussions after theses goals are achieved.

The natural extension and the fourth step will be to derive the asymptotic or large-sample properties for the posterior parameters. In several models this becomes so complicated that it is impractical to obtain and the researcher has to be satisfied with the numerical approximations from the simulations. In the thesis we are able for the stationary time series model, by using

some clever approximations, to obtain the large-sample properties associated with the posterior distribution of the unknown covariance function.

The thesis is written within the classical framework of a master thesis. This is perhaps not the most efficient medium in which to communicate new results, since they become parts of a long and detailed story. Chapter 3, in particular has various new results that might be published separately. I also hope that I will be able to write (together with my supervisor) an article this autumn, where we present the main results from the thesis.

See also the concluding remarks at the end of Chapter 3 and 4 for more details and a longer discussion.

CHAPTER 2

Discrete time

This chapter is divided into two parts. The first section is a general introduction to the central properties of stationary time series with some extra attention on the stationary Gaussian time series and the connection between the power spectrum and the covariance function. In the second part we will introduce the basic concepts regarding nonparametric Bayesian estimation. We will derive meaningful prior distributions for both the covariance and correlation function by placing equivalent prior distribution on the spectral measure (integrated power spectrum) and also show how we can obtain posterior inference based on simulations.

In Section 1 will introduce the basic ideas and definitions regarding stationary time series, we will also very briefly show how we can make predictions about future outcomes based on a observed sequence. In Section 1.1 we will introduce the power spectrum, discuss some of its properties and show how this function is related to the covariance and correlation function. Section 1.2 and 1.3 will be used to show how we can estimate the power spectrum and spectral measure from the periodogram function based on a observed series. We will also introduce some of the large-sample properties, which we will extend in a way that it will fit our nonparametric Bayesian framework. Most of the theory presented in this section and some of the notation are based on the books **?** and **?**.

1. Stationary time series

A time series $Y(t)$ is defined as a family or sequence of random variables, $\{Y(t)\}$, indexed by t, where t is often referred to as time and belongs to the index set $T = \{0, \pm 1, \pm 2, \ldots\}$. We will write the expectation at the time point t as $E[Y(t)] = \mu(t)$ and denote the variance at the same time point by $\mathrm{Var}(Y(t)) = \sigma^2(t)$. A time series $Y(t)$ is defined to be *strictly stationary* or *strongly stationary* if the distribution of the time series is invariant under translation, i.e.

$$\Pr\{Y(t_1) < y_1, \ldots, Y(t_k) < y_k\} = \Pr\{Y(t_1 + h) < y_1, \ldots, Y(t_k + h) < y_k\}$$

for any choice of time points $t_1, \ldots, t_k$, where $k \in \mathbb{N}$, and $h = 0, \pm 1, \pm 2 \ldots$. A time series $Y(t)$ is said to be *second order stationary* if the expectation and covariance satisfy

$$E[Y(t)] = \mu(t) = \mu \text{ and } \mathrm{Cov}(Y(t + h),\, Y(t)) = C(h)$$

for $t, h = 0, \pm 1, \pm 2 \ldots$ and where the covariance is defined in the usual way, i.e. $\mathrm{Cov}(Y(t + h),\, Y(t)) = E[(Y(t + h) - \mu)(Y(t) - \mu]$. Normally we will work under the assumption that the time series is stationary and has zero mean, this assumption is quite common, see **?**. Suppose $Y(0), \ldots, Y(n - 1)$ is a sample form a stationary time series $Y(t)$ with expectation μ, then $X(t) = Y(t) - \mu$ will become a zero mean time series with the same dependency structure as $Y(t)$, note that if μ is unknown it is not unusual to replace μ with the sample mean $\bar{Y}(t)$.

Therefore, given a sample from a stationary time series with expectation μ we are always able to construct a zero mean series $X(t)$ and work with this instead. In the case of a second order stationary time series we will refer to the function $C(h)$ as the *covariance function*, also if $Y(t)$ is second order stationary we have that $C(0) = \sigma^2(t) = \sigma^2$ and the *correlation function* will be defined as the function $R(h) = C(h)/C(0)$.

Let $Y(0), \ldots, Y(n-1)$ be a sample of size n from $Y(t)$, a quite common estimate for the covariance is

$$\hat{C}(h) = \frac{1}{n-|h|} \sum_{u=1}^{n-|h|} (Y(u) - \bar{Y})(Y(u+|h|) - \bar{Y}) \tag{1.1}$$

for $h = 0, \pm 1, \ldots, \pm(n-1)$. The estimator (1.1) is in some literature known as the *unbiased covariance estimator*, where the *biased covariance estimator* is estimator (1.1) where the $(n-|h|)$ term in the fraction is replaced with n. The biased estimators is often preferred since it has less bias for large $|h|$ and since in most situation it also has the least total mean squared error, see Chapter 5 in **?**. A perhaps more common method in covariance estimation is to assume that $C(h) = C(h|\boldsymbol{\theta})$, i.e. that the covariance function belong to a class of parametric functions, and estimate the unknown parameter $\boldsymbol{\theta}$ from the observed data, see **?** for a complete discussion of parametric covariance estimation and the asymptotic behavior of the estimated parameters. We will refer to the covariance function as *valid* if it satisfies the positive definiteness condition, i.e.

$$\sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j C(t_i - t_j) \geq 0 \tag{1.2}$$

for any set of locations $(t_1, \ldots, t_k)$, real numbers $(a_1, \ldots, a_k)$ and $k \in \mathbb{N}$. It is obvious that any covariance function has to satisfy this property since

$$\operatorname{Var}\left(\sum_{i=1}^{k} a_i Y(t_i) \right) = \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j \operatorname{Cov}(Y(t_i),\ Y(t_j)) = \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j C(t_i - t_j),$$

which is defined to be nonnegative.

The next results sum up some of the main properties for the covariance function of a second order stationary time series.

LEMMA 1.1. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a second order stationary time series with covariance function given by $C(h)$. Then for $h = 0, \pm 1, \ldots$*

i) $C(0) \geq 0$.
ii) $C(h) = C(-h)$.
iii) $C(0) \geq |C(h)|$.
iv) $C(h) = \operatorname{Cov}(Y(t),\ Y(t+h)) = \operatorname{Cov}(Y(0),\ Y(h))$.
v) *If $C_i(h)$ is a valid covariance function for $i = 1, \ldots, k$, then $\sum_{i=1}^{k} a_i C_i(h)$ is a valid covariance function if $a_i \geq 0$ and $a_i \in \mathbb{R}$ for all $i = 1, \ldots, k$.*
vi) *If $C_i(h)$ is a valid covariance function for $i = 1, \ldots, k$, then $\prod_{i=1}^{k} C_i(h)$ is a valid covariance function.*

PROOF. i) - v): Most of these properties are straightforward to verify and proofs for all of them can be found in statistical textbooks treating time series.

vi): It is sufficient to show that this is true for two arbitrary choices of valid covariance functions since the general result will follow directly from this property. To prove this property we will use a general result stated below, from Corollary 1.10 we have that $C(h)$ is a valid covariance function if and only if it can be expressed as

$$C(h) = \int_{-\pi}^{\pi} \exp(iuh)\, dF(u)$$

where $F(u)$ is a function with properties similar to a distribution function, see Corollary 1.10 for details. If both $C_1(h)$ and $C_2(h)$ are valid covariance functions, then

$$\begin{aligned} C_1(h)C_2(h) &= \int_{-\pi}^{\pi} \exp(iuh)\, dF_1(u) \int_{-\pi}^{\pi} \exp(iuh)\, dF_2(u) \\ &= \frac{1}{F_1(\pi)F_2(\pi)} \int_{-\pi}^{\pi} \exp(iuh)\, dH_1(u) \int_{-\pi}^{\pi} \exp(iuh)\, dH_2(u) \\ &= \frac{1}{F_1(\pi)F_2(\pi)} E[\exp(ihX_1)]E[\exp(ihX_2)] \\ &= \frac{1}{F_1(\pi)F_2(\pi)} E[\exp(ih(X_1 + X_2))] = \int_{-\pi}^{\pi} \exp(iuh)\, dF_3(u) \end{aligned}$$

where X_1 and X_2 are random variables with cumulative distribution functions given by $H_1(u)$ and $H_2(u)$. The function $F_3(u)$ is defined as $F_3(u) = H_3(u)/[F_1(\pi)F_2(\pi)]$, where $H_3(u)$ is the convolution of $H_1(u)$ and $H_2(u)$. Since $C_1(h)C_2(h)$ has the representation required from Corollary 1.10 we know that the product of two valid covariance functions becomes a valid covariance function, which completes the proof. □

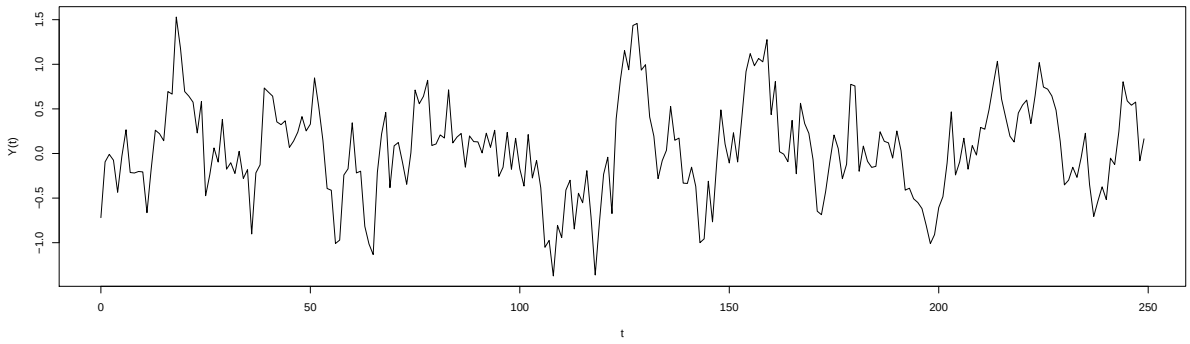

Figure 1.1: Stationary Gaussian time series with $\mu = 0$ and covariance function $C(h) = \rho^{|h|}$, where $\sigma = 0.55$, $\rho = 0.77$ and $n = 250$.

EXAMPLE 1.2. *(Gaussian time series)*
We will say that the random process $Y(t)$ is *Gaussian time series* if the cumulative distribution function

$$\Pr\{Y(t_1) < y_1, \ldots, Y(t_k) < y_k\}$$

is equal to the distribution of k-variate Gaussian random variable for any $k \in \mathbb{N}$, note that this implies that every $Y(t_i)$ is an univariate Gaussian random variable for every $i = 1, \ldots, k$.

We will now give two examples of simulated stationary Gaussian time series. The series in Figure 1.1 have covariance with longer range and lower variance than the series in Figure 1.3. Both are displayed with their respective estimated covariances, Figure 1.4 and Figure 1.2, based on both the biased and the unbiased nonparametric estimators given by equation 1.1.

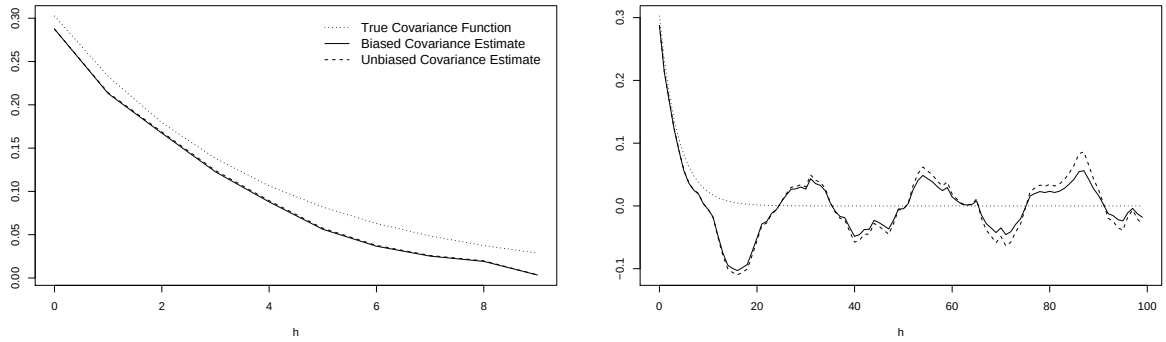

Figure 1.2: Estimated covariance based on the estimator (1.1) for the stationary time series in Figure 1.3, plotted on the interval $[0, 8]$ (left panel) and $[0, 100]$ (right panel).

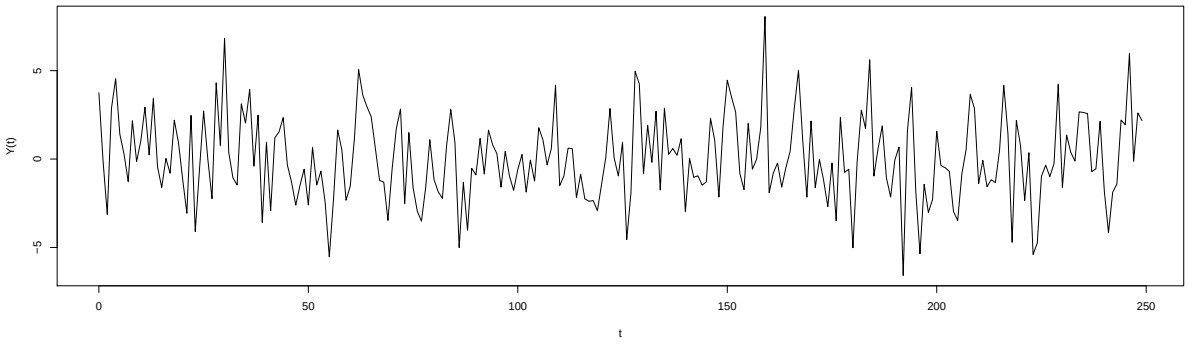

Figure 1.3: Stationary Gaussian time series with $\mu = 0$ and covariance function $C(h) = \rho^{|h|}$, where $\sigma = 2.22$ and $\rho = 0.11$.

The last concept that will be given in this section is the definition of a version of cumulants. The reason for this is that **?** among others uses results based on the cumulants throughout the theory, in many concepts, and as a tool to prove several results related to time series.

Suppose $\boldsymbol{Y}(t)$, where $t = 0, \pm 1, \ldots$, is an r vector valued time series with components $Y_a(t)$ that satisfy $E[|Y_a|^r] < \infty$ for all $a = 1, \ldots, r$, then the *joint cumulant function of order* k is defined as

$$\begin{aligned} c_{a_1,\ldots,a_k}(t_1, \ldots, t_k) &= c_{Y_{a_1},\ldots,Y_{a_k}}(t_1, \ldots, t_k) \\ &= \operatorname{cum}(Y_{a_1}(t_1), \ldots, Y_{a_k}(t_k)) \\ &= \sum_P (-1)^{|P|-1}(|P|-1)! \prod_{p \in P} E\Big[\prod_{j \in p} Y_{a_j}(t_j)\Big] \end{aligned} \tag{1.3}$$

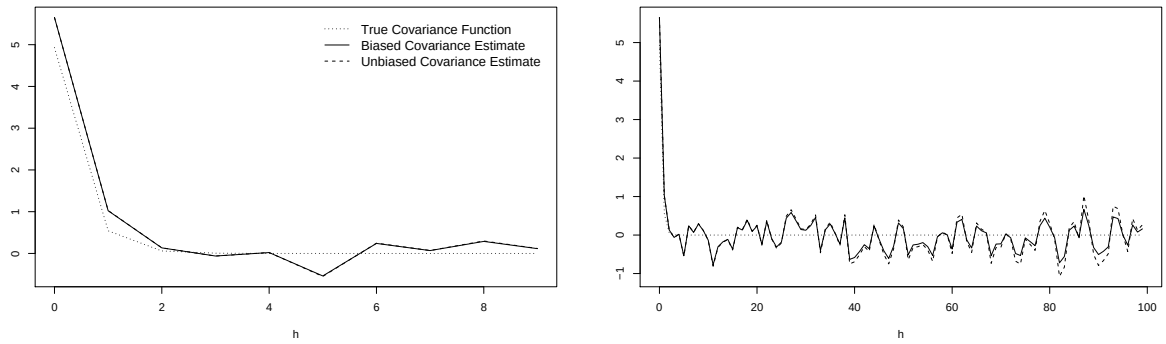

Figure 1.4: Estimated covariance based on the estimator (1.1) for the stationary time series in Figure 1.3, plotted on the interval $[0, 8]$ (left panel) and $[0, 100]$ (right panel).

where P runs through the the list of all partitions of $(a_1, \ldots, a_k)$ and p runs through the list of all blocks of partitions of P, the notation cum stands for cumulant, or joint cumulant. We will now work out some of the properties of cumulant functions through some examples, for a more complete introduction to theory and other related properties see **?**.

EXAMPLE 1.3. *(Properties of the cumulant functions)*
The expectation, variance and covariance of a random variable can be expressed through the first orders of their joint cumulant functions. Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a time seres, then from the definition of the cumulant (1.3) and for any choice of integers, l, l' and l'' we have that

$$
\begin{aligned}
c_Y(t_l) &= \sum_{\{\{l\}\}} (-1)^{1-1}(1-1)! \prod_{p \in \{\{l\}\}} E[\prod_{j \in p} Y(t_j)] = E[Y(t_j)], \\
c_{YY}(t_l, t_{l'}) &= \sum_{P} (-1)^{|P|-1}(|P|-1)! \prod_{p \in P} E\left[\prod_{j \in p} Y_{a_j}(t_j)\right], \quad \text{where } P = \{\{l, l'\}, \{\{l\}, \{l'\}\}\} \\
&= (-1)^{1-1}(1-1)!E[Y(t_l)Y(t_{l'})] + (-1)^{2-1}(2-1)!E[Y(t_l)]E[Y(t_{l'})] \\
&= E[Y(t_l)Y(t_{l'})] - E[Y(t_l)]E[Y(t_{l'})] \\
&= \mathrm{Cov}(Y(t_l),\ Y(t_{l'})) \\
c_{YYY}(t_l, t_{l'}, t_{l''}) &= \sum_{P} (-1)^{|P|-1}(|P|-1)! \prod_{p \in P} E\left[\prod_{j \in p} Y_{a_j}(t_j)\right], \\
&\qquad \text{where } P = \{\text{the set of all all partitions of the set } \{l, l', l''\}\} \\
&= E[Y(t_l)Y(t_{l'})Y(t_{l''})] - E[Y(t_l)]E[Y(t_{l'})Y(t_{l''})] - E[Y(t_{l'})]E[Y(t_l)Y(t_{l''})] \\
&\quad - E[Y(t_{l''})]E[Y(t_l)Y(t_{l'})] + 2E[Y(t_l)]E[Y(t_{l'})]E[Y(t_{l''})].
\end{aligned}
$$

REMARK 1.4. There exists another and perhaps more common definition of the joint cumulant than equation (1.3). From **?** we have that the joint cumulant, $\mathrm{cum}(Y(t_1), \ldots, Y(t_k))$, can be expressed as the coefficients in the Taylor expansion of $\log(E[\exp(i \sum_{j=1}^{k} Y(t_j)\theta_j)])$. To be precise, the different orders of a joint cumulant can be obtained from the coefficients in the Taylor expansion of the logarithm of the characteristic function. In the the one dimensional case this

become

$$\log(E[\exp(iY(t)\theta)]) = \sum_{j=1}^{\infty} c_{Y_1 Y_2 \cdots Y_j}(t_1, \ldots, t_j)\frac{1}{j!}(i\theta)^j = i\mu_Y\theta - \sigma_Y^2\frac{1}{2}\theta^2 + \ldots,$$

where $Y_j = Y$ and $t_j = t$ for all $j = 1, 2, \ldots$. In the case of a Gaussian time series we have from the characteristic function that

$$\log(E[\exp(i\sum_{j=1}^{k} Y(t_j)\theta_j)]) = i\boldsymbol{\mu}^T\boldsymbol{\theta} - \frac{1}{2}\boldsymbol{\theta}^T\Sigma\boldsymbol{\theta},$$

which implies that derivatives of order greater than 2 vanish in the Taylor expansion and the joint cumulant of order greater than 2 does not exist for a Gaussian time series.

1.0.1. *Prediction.* In this subsection we will very briefly explain the prediction setup within the framework of Gaussian time series. The prediction of future outcomes in time series is a large and important topic, but since we will always assume that we are within the class Gaussian time series, the ideas and methods become fairly simple.

Let $Y(t)$, where $t = 0, \pm1, \pm2$, be a stationary Gaussian time series with known expectation and dependency structure given by the covariance function $C(h)$, where $h = 0, \pm1, \pm2 \ldots$. Suppose we have observed n steps of the process $\boldsymbol{Y} = (Y(t), (Y(0), \ldots, Y(n-1))$, and that we wish to predict the outcome at the next, or a future, location m, typically m is in the set $m = n, n+1, \ldots$. We will denote a predictor for the unobserved value $Y(m)$ by $p(\boldsymbol{Y}, Y(m))$ and let $L(Y(m),\ p(\boldsymbol{Y}, Y(m)))$, to be precise we will assume squared-error loss, i.e

$$L(Y(m),\ p(\boldsymbol{Y}, Y(m))) = (Y(m) - p(\boldsymbol{Y}, Y(m)))^2.$$

It is well known from classic decision theory that the optimal solution, the one that minimizes the expected loss or Bayes Risk, is the conditional mean, i.e.

$$p_0(\boldsymbol{Y},\ Y(m)) = E[Y(m)|\boldsymbol{Y}], \tag{1.4}$$

where $p_0(\boldsymbol{Y}, Y(m))$ denotes the optimal predictor, see for example **?**. We will illustrate the concept with an example.

Example 1.5. *(Prediction in stationary Gaussian time series)*
Let $Y(t)$ be a Gaussian time series with expectation $E[Y(t)] = \mu(t)$, where $t = 0, \pm1, \pm2, \ldots$, and with covariance function $C(h)$, for $h = 0, \pm1, \pm2 \ldots$. Suppose a sample $\boldsymbol{Y} = (Y(t), (Y(0), \ldots, Y(n-1))$ is observed, let $E[\boldsymbol{Y}] = \boldsymbol{\mu}$ and denote the covariance matrix of the vector $\boldsymbol{Y}$ by $\Sigma_{\boldsymbol{Y}}$. The goal is to predict the outcome at a future location m from the observed $\boldsymbol{Y}$, where m typically is in the set $m = n, n+1, \ldots$. We will assume squared-error loss and from (1.4) we have that the optimal predictor is the conditional mean. In order to determine the optimal predictor we will need the joint distribution of the observations and the unknown outcome. Under the assumption of a Gaussian time series the joint distribution is given by

$$\begin{bmatrix} Y(m) \\ \boldsymbol{Y} \end{bmatrix} \sim N_{n+1}\left(\begin{bmatrix} \mu(m) \\ \boldsymbol{\mu} \end{bmatrix}, \Sigma_{m\boldsymbol{Y}} = \begin{bmatrix} \sigma_m^2 & \sigma_{m\boldsymbol{Y}}^T \\ \sigma_{m\boldsymbol{Y}} & \Sigma_{\boldsymbol{Y}} \end{bmatrix}\right),$$

where $\mu(m) = E[Y(m)]$, $\sigma_m^2 = \mathrm{Var}(Y(m))$ and $\sigma_{m\boldsymbol{Y}} = \mathrm{Cov}(Y(m), \boldsymbol{Y})$. From Remark 1.6 below it is now straightforward to verify that the conditional expectation and variance are determine

by the equations

$$E[Y(m)|\boldsymbol{Y}] = \mu(m) + \boldsymbol{\sigma}_{m\boldsymbol{Y}}^T \Sigma_{\boldsymbol{Y}}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}) \text{ and}$$
$$\text{Var}(Y(m)|\boldsymbol{Y}) = \sigma_m^2 - \boldsymbol{\sigma}_{0\boldsymbol{Y}}^T \Sigma_{\boldsymbol{Y}}^{-1} \boldsymbol{\sigma}_{m\boldsymbol{Y}}.$$

Under squared-error loss the optimal prediction for the outcome at m is now given by

$$p_0(\boldsymbol{Y}, Y(m)) = \mu(m) + \boldsymbol{\sigma}_{m\boldsymbol{Y}}^T \Sigma_{\boldsymbol{Y}}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}), \tag{1.5}$$

note that the optimal predictor (1.5) simplifies to $p_0(\boldsymbol{Y}, Y(m)) = \boldsymbol{\sigma}_{m\boldsymbol{Y}}^T \Sigma_{\boldsymbol{Y}}^{-1} \boldsymbol{Y}$ for time series with expectation zero. We will return to this example later in the thesis when the complete model is presented.

The following remark can be found in several textbooks in statistics and will be stated without proof.

REMARK 1.6. Let W be a $(n \times 1)$ Gaussian random vector, that can be partition into $W = (U, V)^T$ with U of dimension $(u \times 1)$ and V of dimension $(v = (n - u) \times 1)$. If

$$E[W] = E\begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} \mu_u \\ \mu_v \end{bmatrix} \text{ and } \text{Var}(W) = \begin{bmatrix} \Sigma_u & \Sigma_{uv} \\ \Sigma_{uv}^T & \Sigma_v \end{bmatrix},$$

then $V|U$ follows a Gaussian distribution with mean and variance given by

$$E[U|V] = \mu_u + \Sigma_{uv}\Sigma_v^{-1}(V - \mu_v) \text{ and } \text{Var}(U|V) = \Sigma_u - \Sigma_{uv}\Sigma_v^{-1}\Sigma_{uv}^T. \tag{1.6}$$

The next assumption is the Assumption 2.6.1 in **?** and is fundamental in many of his results and is usually the condition required in order to prove several theorems stated later.

ASSUMPTION 1.7. The times series $\boldsymbol{Y}(t)$ is a strictly stationary r vector-valued series with components $Y_j(t)$, where $j = 1, \ldots r$, all of whose moments exist, and satisfying

$$\sum_{u_1,\ldots,u_{k-1}=-\infty}^{\infty} |c_{a_1,\ldots,a_k}(u_1, \ldots, u_{k-1})| < \infty$$

for $a_1, \ldots, a_k = 1, \ldots, r$ and $k = 2, 3, \ldots$.

In this thesis we will not be too much concerned with general r vector-valued series, we will mostly focus on the one-dimensional case where the series is also Gaussian, then, according to **?**, the assumption above is simplified and we may then rewrite Assumption 1.7 as follows.

ASSUMPTION 1.8. The time series $Y(t)$ is a strictly stationary Gaussian series where all moments exist and satisfy

$$\sum_{h=-\infty}^{\infty} |C(h)| < \infty.$$

1.1. The power spectrum. We have now established most of the basic concepts regarding stationary time series and we will now discuss one of the main topics needed in this thesis. Throughout this section we will assume that $Y(t)$, where $t = 0, \pm 1, \ldots$, is a real-valued time series with mean $E[Y(t)] = \mu_Y$ and covariance function $\mathrm{Cov}(Y(t+h), Y(t)) = C(h)$, defined for $h = 0, \pm 1, \ldots$. If the covariance function satisfies the restriction

$$\sum_{h=-\infty}^{\infty} |C(h)| < \infty,$$

then the *power spectrum* of the series $Y(t)$ at frequency u is defined to be the Fourier transformation

$$f(u) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \exp(-iuh)C(h) \text{ for } -\infty < u < \infty. \tag{1.7}$$

The power spectrum is non-negative, even and of period 2π with respect to u, which means that we may take the interval $[0, \pi]$ as the fundamental domain for $f(u)$ since the whole function is determined by its behavior on this interval, see **?** or **?**. We can also invert expression (1.7) and write the covariance function as a function of the power spectrum

$$C(h) = \int_{-\pi}^{\pi} \exp(iuh) f(u)\, du \text{ for } h = 0, \pm 1, \ldots. \tag{1.8}$$

As we shall see later, equation (1.8) will become of great importance and in combination with Corollary 1.10 below we will have a very useful and general tool to create valid covariance functions. Note that the variance of $Y(t)$ can be expressed as

$$\mathrm{Var}(Y(t)) = \sigma_Y^2 = \int_{-\pi}^{\pi} f(u)\, du.$$

We will define the *integrated spectrum* or *spectral measure* in the natural way

$$F(u) = \int_{-\pi}^{u} f(v)\, dv \text{ for } -\pi < u < \pi, \tag{1.9}$$

because of the periodicity and the evenness of $f(u)$ we could also take the interval $[0, \pi]$ as the fundamental domain for $F(u)$. It is quite common to work within the complex numbers when studying the power spectrum and its related properties, the reason for this is that it is often easier to wok within this framework and the results become more elegant. In the case of real valued processes we can rewrite expression (1.7) and (1.8)

$$f(u) = \frac{C(0)}{2\pi} + \frac{1}{\pi} \sum_{h=1}^{\infty} \cos(uh)C(h) \text{ for } -\pi < u < \pi \tag{1.10}$$

and

$$C(h) = 2 \int_{0}^{\pi} \cos(uh) f(u)\, du \text{ for } h = 0, \pm 1, \ldots. \tag{1.11}$$

see **?** for details. We will define *normalized power spectrum* as $f(u)/\sigma_Y^2 = f(u)/F(\pi) = h(u)$, for $u \in (-\infty, \infty)$, with fundamental domain $u \in [0, \pi]$, then there exists a similar relation between the normalized power spectrum and the correlation function as between the power spectrum and the covariance function. From **?** we have that

$$h(u) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \exp(-iuh)R(h) \text{ for } -\pi < u < \pi.$$

and

$$R(h) = \int_{-\pi}^{\pi} \exp(iuh)h(u)\,du \text{ for } h = 0, \pm 1, \ldots.$$

The *normalized integrated spectrum* is defined in the same way as the integrated power spectrum (1.9), and for real valued processes there exist an equivalent simplification as (1.10) and (1.11) for the normalized power spectrum and correlation function. For a complete introduction to spectral analysis see **?**.

From what we have seen so far it is clear that if the expectation and the dependency structure or spectral measure is given we know everything we need to know about a certain time series. This means that we in some sense are free to choose how we will define or work with time series, we can choose to work within the time domain (covariance or correlation functions) or in the frequency domain (power spectrum and spectral measures). The choice of domain will probably depend on the situation and the study, and in some situations it might become necessary to work back and fourth between the two domains.

The following two results reveal another strong argument for why it is desirable, and sometimes necessary, to work in the frequency domain, both results can be found in **?** with an argument and are in a sense the discrete version of the Wiener-Khintchine Theorem given in Section 1 of Chapter 4.

THEOREM 1.9. *(Wold's Theorem) A necessary and sufficient condition for the sequence $R(h)$, where $h = 0, \pm 1, \ldots$, to be the correlation function for a discrete time process $Y(t)$, where $t = 0, \pm 1, \ldots$, is that there exists a function $H(u)$, having the properties of a distribution function on the interval $(-\pi, \pi)$, (i.e. $H(-\pi) = 0$, $H(\pi) = 1$, and $H(u)$ is non-decreasing), such that*

$$R(h) = \int_{-\pi}^{\pi} \exp(iuh)\,dH(u), \quad \text{for } h = 0, \pm 1, \ldots. \tag{1.12}$$

COROLLARY 1.10. *A necessary and sufficient condition for the sequence $C(h)$, where $h = 0, \pm 1, \ldots$, to be the covariance function for a discrete time process $Y(t)$, where $t = 0, \pm 1, \ldots$, is that there exists a function $F(u)$, having the similar properties of a distribution function on the interval $(-\pi, \pi)$, (i.e. $F(-\pi) = 0$, $F(\pi) < \infty$, and $F(u)$ is non-decreasing), such that*

$$C(h) = \int_{-\pi}^{\pi} \exp(iuh)\,dF(u), \quad \text{for } h = 0, \pm 1, \ldots. \tag{1.13}$$

In particular for real valued time series equation (1.13) simplifies to

$$C(h) = 2\int_{0}^{\pi} \cos(uh)\,dF(u), \quad \text{for } h = 0, \pm 1, \ldots. \tag{1.14}$$

Note that Corollary 1.10 follows directly from Theorem 1.9.

The two results provide us with some quite flexible and general tools to construct or check new covariance functions. Suppose we are given a sample from a time series with unknown dependency structure and that the parametric function $C(h|\boldsymbol{\theta})$ is suggested as the covariance function, then if it is clear that $C(h|\boldsymbol{\theta})$ does not possess the representation given in Corollary 1.10 we should not use it to explain the dependency.

Corollary 1.10 and Theorem 1.9 are perhaps more important as tools to construct new and valid covariance or correlation functions. The advantage of the two results are the weak conditions for

the functions $F(u)$ and $H(u)$ that are fairly easy to check compared to the positive definiteness property. Note that we do not even require that $F(u)$ is smooth, $F(u)$ could for example be a step function, which suggests the use of nonparametric methods in estimation of the covariance function.

In the next section we will begin to discuss some ideas regarding the estimation of the power spectrum and we will also derive some of its large-sample properties. In this thesis will mostly be concerned with the covariance function and the power spectrum, but most of the ideas are straightforward to use and generalize to the case of correlation functions. Before we end this section we will show how we can obtain the power spectrum for some given covariance functions.

EXAMPLE 1.11. Suppose $Y(t)$, where $t = 0, \pm 1, \ldots$, is a time series with covariance function $C(h) = \sigma^2 \rho^{|h|}$, and we would like to find $f(u)$. If we assume that $Y(t)$ is a real valued process, then

$$f(u) = \frac{\sigma^2}{2\pi} \sum_{h=-\infty}^{\infty} \exp(-iuh)\rho^{|h|} = \frac{\sigma^2}{2\pi}\left[1 + 2\sum_{h=1}^{\infty} \cos(uh)\rho^h\right].$$

This motivates us to find the limit as the real part of the limit of the of the equivalent infinite complex valued sum, which is much easier to solve, let $c = \log(\rho)$ then

$$\begin{aligned} 1 + \sum_{h=1}^{\infty} \cos(uh)\rho^h &= \mathrm{Re} \sum_{h=0}^{\infty} \exp(ihu)\exp(cu) = \mathrm{Re} \sum_{h=0}^{\infty} \exp(ih + c)^u = \mathrm{Re}\, \frac{1}{1 - \exp(c + iu)} \\ &= \mathrm{Re}\, \frac{1 - \exp(c - iu)}{1 - \exp(c + iu) - \exp(c - iu) + \exp(2c)} = \frac{1 - \rho\cos(u)}{1 - 2\rho\cos(u) + \rho^2}. \end{aligned}$$

An explicit expression for the power spectrum is now given by

$$f(u) = \frac{\sigma^2}{2\pi}\left[1 + 2\frac{1 - \rho\cos(u)}{1 - 2\rho\cos(u) + \rho^2} - 2\right] = \frac{\sigma^2(1 - \rho^2)}{2\pi(1 - 2\rho\cos(u) + \rho^2)}$$

If we choose $\rho = e^{-\alpha}$ we will get the common exponential covariance function $C(h) = \sigma^2 \exp(-\alpha|h|)$ for $h = 0, \pm 1, \ldots$. Below we will give two concrete examples of the power spectrum and the spectral measure and their respective covariance functions, see Figure 1.5 and 1.6

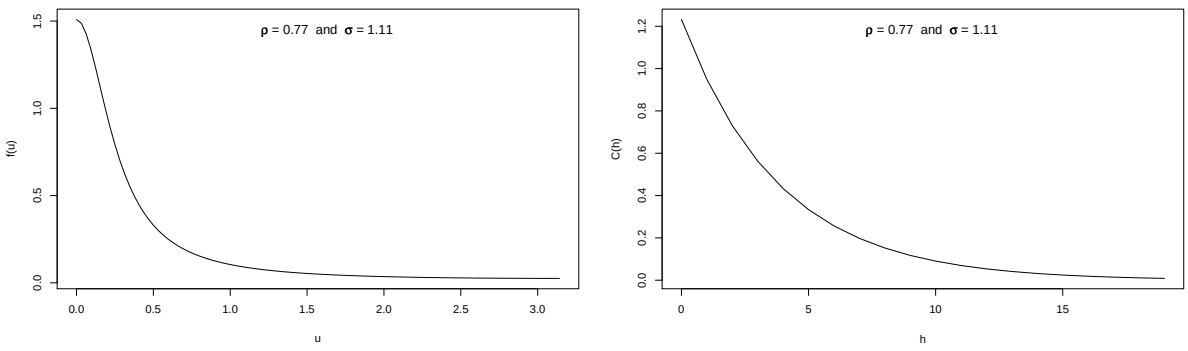

Figure 1.5: Plot of power spectrum (left panel) and the corresponding covariance function (right panel) with dependency with long range and medium variation.

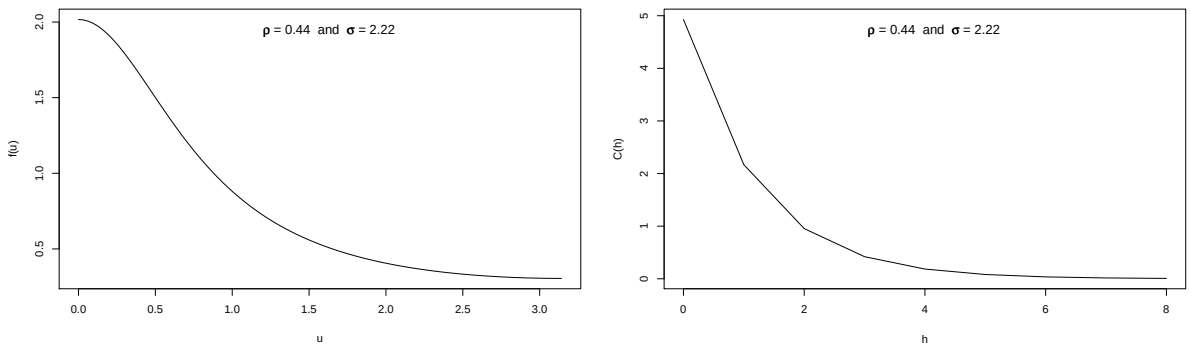

Figure 1.6: Plot of power spectrum (left panel) and the corresponding covariance function (right panel) with dependency with short range and large variation.

EXAMPLE 1.12. From, for example **?**, we have that for $c = -\log(\rho)$ the following integral has an exact solution

$$\begin{aligned}\lim_{M\to\infty}\int_{-M}^{M}\exp(iuh)\frac{1}{c\pi(1+(u/c)^2)}\,du \\ = \lim_{M\to\infty} 2\int_0^M \cos(uh)\frac{1}{c\pi(1+(u/c)^2)}\,du = \exp(-ch) = \rho^h.\end{aligned} \tag{1.15}$$

For a reasonably large constant M it follows form the above that

$$2\int_0^{M\pi}\cos(uh)\frac{1}{c\pi(1+(u/c)^2)}\,du = 2\int_0^{\pi}\cos(Muh)\frac{M}{c\pi(1+(Mu/c)^2)}\,du \approx \rho^h.$$

Let $f(u) = \sigma^2 M/(c\pi(1+(Mu/c)^2)$, then from Corollary 1.10 we know that the covariance function, $C(h)$, given by

$$C(h) = 2\int_0^{\pi}\cos(uh')f(u)\,du = 2\int_0^{\pi}\cos(uh')\frac{\sigma^2 M}{c\pi(1+(Mu/c)^2)}\,du$$

where $h' = Mh$, is a valid covariance function and that $C(h) \approx \rho^h$.

EXAMPLE 1.13. In Section 2 in Chapter 4 we will introduce a concept known as the *aliasing effect*. The aliasing effect is a term used for the problem that arise when we sample a continuous time process at equidistant time points, see Section 2 of Chapter 4 or **?** for a explanation. We will now use the 'solution' to the aliasing effect problem to construct valid covariance functions for time series models from covariance functions defied for general spatial models.

Let $C(h)$, where $h \in (-\infty, \infty)$ be a covariance function for a continuous time process over $\mathbb{R}^1$. Suppose we know that this specific covariance function has power spectrum (spectral density) given by $f(u)$. Unfortunately we cannot use the given power spectrum $f(u)$ directly to construct covariance functions for stationary time series by equation (1.8) such that the discrete time covariance function $C_1(h)$ satisfy $C_1(h) = C(h)$, for $h = 0, \pm 1, \pm 2, \ldots$. Let $f_1(u)$ be a function obtained from the power spectrum $f(u)$ in the following way

$$f_1(u) = \sum_{j=-\infty}^{\infty} f(u + 2\pi j), \tag{1.16}$$

where $u \in [-\pi, \pi]$. From **?** we have that the function $f_1(u)$ is exactly the power spectrum that through equation (1.8) will result in a covariance function $C_1(h)$ which satisfies $C_1(h) = C(h)$, for $h = 0, \pm 1, \pm 2, \ldots$. It is clear that equation (1.16) provides us with a general method to construct covariance functions for stationary time series with the same properties as covariance functions defied for continuous time processes.

As an concrete example we will show a third method to how we can obtain the power spectrum for the covariance function $C(h) = \rho^{|h|}$. Let $c = -\log(\rho)$, then from equations (1.16) and (1.15) we have that

$$f_1(u) = \sum_{j=-\infty}^{\infty} \frac{1}{c\pi(1 + ([u + 2\pi j]/c)^2)} \quad \text{and} \quad C(h) = 2\int_0^{\pi} \cos(uh) f_1(u)\, du = \rho^{|h|}$$

where $u \in [-\pi, \pi]$.

EXAMPLE 1.14. As a final example we will study a slightly different situation. Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a real valued time series with covariance function given by

$$C(h) = \begin{cases} \sigma^2 \rho^{|h|}, & \text{if } h \text{ is even} \\ 0, & \text{if } h \text{ is odd} \end{cases} \tag{1.17}$$

then the spectral function, $f(u)$, is given as

$$f(u) = \frac{\sigma^2}{2\pi} \sum_{h=-\infty}^{\infty} \exp(-iuh)\rho^{|h|} = \frac{\sigma^2}{2\pi}\left[1 + 2\sum_{h=0}^{\infty} \cos(u2h)\rho^{2h}\right] = \frac{\sigma^2}{2\pi}\left[1 + 2\sum_{h=0}^{\infty} \cos(2uh)(\rho^2)^h\right].$$

From Example 1.11 it follows directly that we could write the spectral measure, $f(u)$, as

$$f(u) = \frac{\sigma^2}{\pi}\left[\frac{1 - \rho^2\cos(2u)}{1 - 2\rho^2\cos(2u) + \rho^4} - \frac{1}{2}\right] = \frac{\sigma^2(1 - \rho^4)}{2\pi(1 - 2\rho^2\cos(2u) + \rho^4)}, \tag{1.18}$$

for $u \in (-\infty, \infty)$.

1.2. The periodogram. Given a sample $Y(0), \ldots, Y(n-1)$ of size n from the time series $Y(t)$, in this section we will start to study how we can estimate the power spectrum $f(u)$ from a given sample. Perhaps the most obvious and natural idea is to use the sample to estimate covariance function $\hat{C}(h)$ from the equation (1.1) and then use relation (1.7) to estimate power spectrum. Following this idea an estimate for the power spectrum is given by

$$\hat{f}(u) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \exp(-iuh)\hat{C}(h), \tag{1.19}$$

even though the approach described here is natural, we will not follow this here, see **?** for a complete discussion of this approach.

Let $Y(t)$, where $t = 0, \pm 1, \pm 2, \ldots$, be a stationary time series with $E[Y(t)] = \mu$ and true power spectrum $f_0(u)$, where $u \in [0, \pi]$. Let $Y(0), \ldots, Y(n-1)$ be a sample of size n from the time series $Y(t)$, then the *periodogram* or *second-order periodogram* will be defined as

$$I_n(u) = \frac{1}{2\pi n}\left|\sum_{j=0}^{n-1} \exp(-iuj)Y(j)\right|^2. \tag{1.20}$$

The periodogram was first introduced by **?** and was then used as a tool to find hidden periodicities. The periodogram is now well known and widely used as an estimate for the power

spectrum $f_0(u)$, but because of its nice properties we will in practice often rather use a weighted or smoothed version of the periodogram, see below.

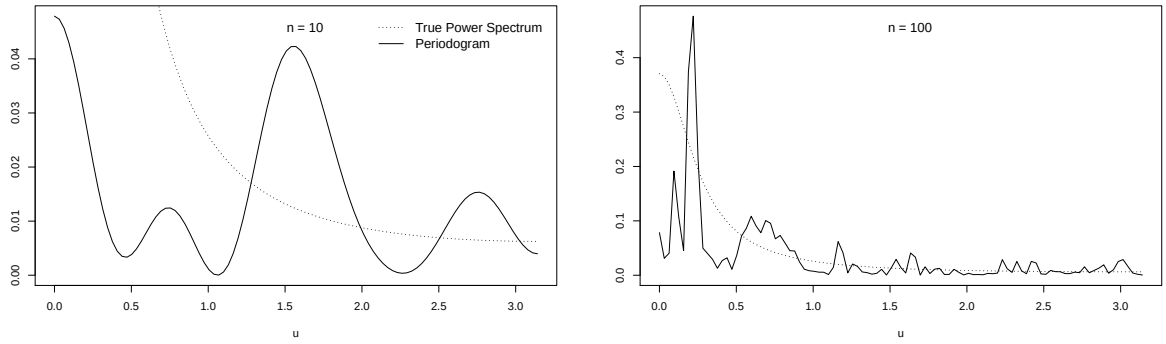

Figure 1.7: The periodogram function based on the n first observations from the data of Figure 1.1

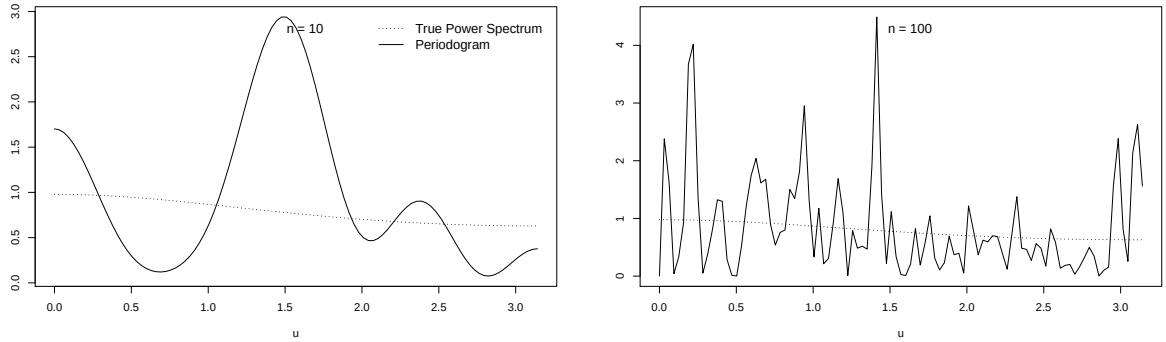

Figure 1.8: The periodogram function based on the n first observations from the data of Figure 1.3

The following results show some of the main statistical properties of the periodogram, its expectation, variance and the correlation between periodogram at different frequencies. The results will reveal why the periodogram can be used as an estimator for the power spectrum, but also why it is not the optimal choice of estimator. The results below are stated without proof and can be all be found in **?**.

THEOREM 1.15. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a time series with mean $E[Y(t)] = \mu_Y$ and covariance function given by $\mathrm{Cov}(Y(t+h), Y(t)) = C(h)$, where $u, h = 0, \pm 1, \ldots$. Suppose that the covariance function satisfies*

$$\sum_h |h||C(h)| < \infty \tag{1.21}$$

then the periodogram is an asymptotically unbiased estimate for $f(u)$, where $u \not\equiv 0 \,(\mathrm{mod}\ 2\pi)$, and we have

$$E[I_n(u)] = f(u) + \frac{1}{2\pi n}\left[\frac{\sin(nu/2)}{\sin(u/2)}\right]^2 \mu_Y^2 + O(n^{-1}).$$

The need of the inequality (1.21) is not necessary for the periodogram to be an unbiased estimator, it is sufficient to assume that $\sum_h |C(h)| < \infty$, see **?**.

THEOREM 1.16. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a real valued time series satisfying Assumption (1.7) and let $I_n(u)$ be the periodogram given by 1.20, then for u, v where $u \leq v$ and $u, v \not\equiv 0 \pmod{2\pi}$*

$$\begin{aligned}&\mathrm{Cov}(I_n(u),\ I_n(v))\\ &\quad= \left(\left[\frac{\sin(n(u+v)/2)}{n\sin((u+v)/2)}\right]^2 + \left[\frac{\sin(n(u-v)/2)}{n\sin((u-v)/2)}\right]^2\right) f(u)^2 + O(n^{-1})\end{aligned}$$

COROLLARY 1.17. *Assume the conditions in Theorem 1.15 and Theorem 1.16 are satisfied. Let r, s be integers such that r, s, $r \pm s \not\equiv 0 \pmod{n}$ and define $u_r = 2\pi r/n$ and $v_s = 2\pi s/n$. Then*

$$\begin{aligned}E[I_n(u_r)] &= f(u) + O(n^{-1}),\\ \mathrm{Var}(I_n(u_r)) &= f(u)^2 + O(n^{-1}) \text{ and}\\ \mathrm{Cov}(I_n(u_r),\ I_n(v_s)) &= O(n^{-1}).\end{aligned}$$

From the results above it is clear that the periodogram will work as an estimator for the power spectrum, but it is not optimal in the sense that it is not a consistent estimator. The variance of $I_n(u)$ will tend to a finite limit, i.e $\lim_{n\to\infty} \mathrm{Var}(I_n(u)) = f_0(u)^2$, for all $u \in [0, \pi]$. It will therefore become impossible to reduce the uncertainty in the estimate below this limit by increasing the number of observations. The low dependency between periodogram functions at different frequencies explain why the periodogram has such irregular behavior, see Figure 1.7 and 1.8, even though the true $f_0(u)$ will be expected to be fairly smooth. In the next theorem asymptotic distribution of the periodogram will be established.

THEOREM 1.18. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a real valued time series satisfying Assumption (1.7) and with true power spectrum $f_0(u)$. Let $s_j(n)$ be an integer with $u_j(n) = 2\pi s_j(n)/n \to u_j$ as $n \to \infty$ for $j = 1, \ldots, J$. Suppose $2u_j(n)$, $u_j(n) \pm u_k(n) \not\equiv 0 \pmod{2\pi}$ for $1 \leq j \leq k \leq J$ and $n = 1, 2, \ldots$. Let the periodogram, $I_n(u)$ be as defined in 1.20, then for $j = 1, \ldots, J$ $I_n(u_j(n))$ is asymptotical $f_0(u) \times \mathrm{Exp}(1)$ and asymptotically independent of each other.*

1.2.1. *The smoothed periodogram and other consistent estimators.* In the previous section we saw that the variance of the raw periodogram did not approach zero, this is a desirable property of an estimator, and it suggests that the periodogram itself might not be the perfect choice as an estimator for the power spectrum. The common solution to this problem is to choose a smoothed or weighed version of the periodogram as an estimator. Let $\omega(u)$ be a suitable weight function, see **?** for several choices of $\omega(u)$, then the following expression will be referred to as a weighted/smoothed estimator for the power spectrum

$$\hat{f}_\omega(u) = \sum_{u_i \in U_u} \omega(u_i) I_n(u_i),$$

where U_u is a set containing values in an neighborhood around u. In **?** and **?** it is shown that for several classes of "natural" weight functions the weight periodogram is a consistent estimator for the power spectrum. In the first paragraph of this section we mentioned an estimator for the power spectrum based on the estimated covariance function (1.19). **?** shows that this method will create an unbiased estimate for the power spectrum, but again it is not consistent, and the

solution is to use a wighted version. Let ω be an appropriate choice of weight function, see **?** for examples, then the estimator given by

$$\hat{f}_\omega(u) = \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \omega(h) \exp(-iuh) \hat{C}_n(h),$$

is an unbiased estimator for the power spectrum. For a more complete introduction to power spectrum functions based on the estimated covariance see **?**.

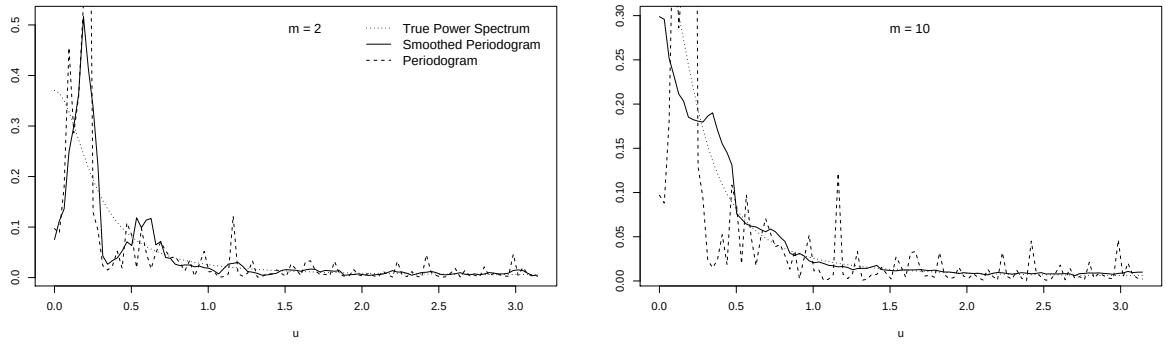

Figure 1.9: The Periodogram together with two different degrees of smoothed periodogram functions based on the function (1.22). The estimates are based on the whole set of observations displayed in Figure 1.1

In this thesis we will not be too much concerned with the different types of weight estimators, but there is one type we will study in a little more detail. The following estimator for the power spectrum will be known as the *smoothed periodogram*. Assume that there exist integers $r(n)$ such that $2\pi r(n)/n$ are close to u, let m be some integer, where $0 < m < n$, then the smoothed periodogram is defined as

$$\hat{f}_m(u) = \frac{1}{2m+1} \sum_{j=-m}^{m} I_n\left(\frac{2\pi[r(n)+j]}{n}\right) \text{ for } 0 < u_i < \pi. \tag{1.22}$$

The next result is from **?** and establishes some of the properties of smoothed periodogram.

THEOREM 1.19. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a real valued series satisfying Assumption 1.7 with true power spectrum $f_0(u)$. Let $\hat{f}_m(u)$ be given by (1.22), then for $0 < u_1, \ldots, u_J < \pi$ let $r_j(n)$ be integers such that $u_j(n) = 2\pi r_j(n)/n \to u_j$ for $j = 1, \ldots, J$, then*

$$E[\hat{f}_m(u_j(n))] = f_0(u_j) \text{ and } \mathrm{Var}(\hat{f}_m(u_j(n))) = f_0(u_j)^2/m + O(n^{-1}).$$

Also $\hat{f}_m(u_j(n))$ and $\hat{f}_m(u_k(n))$, for $j, k = 1, \ldots, J$ and $j \neq k$, are asymptotically independent and

$$\hat{f}_m(u_j(n)) \xrightarrow{L} f_0(u_j) \times \chi^2_{4m+2}/(4m+2).$$

We will end this section with an example of the estimated covariance function based on the periodogram and the smoothed periodogram, Figure 1.10. Note that estimate to the right, the covariance estimated from the raw and unsmoothed periodogram, is the same as the biased

covariance estimate in Figure 1.2. This is a common method to use when estimating the non-parametric covariance function from data since it is several times faster, see also the *fast Fourier transformation* in **?** or **?**.

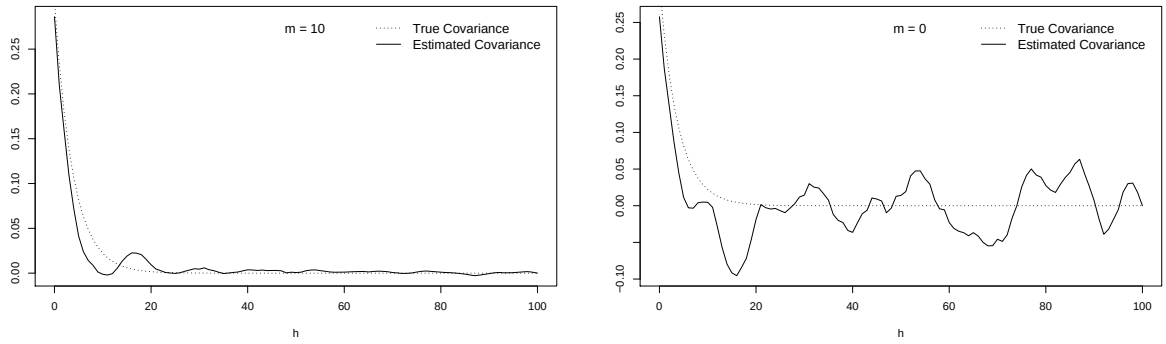

Figure 1.10: Estimated covariance based on the periodogram (left panel) and smoothed periodogram (right panel) from the data displayed in Figure 1.1

1.3. Spectral measure estimation. We will now study an estimator for the spectral measure, $F(u)$, where $u \in [-\pi, \pi]$. Because of the symmetric and periodic properties of $f(u)$ we will rewrite and redefine the spectral measure to the interval $[0, \pi]$. The *spectral measure* of a real valued time series $Y(t)$, where $t = 0, \pm 1, \ldots$, with power spectrum $f(u)$ is given by

$$F(u) = \int_0^u f(v)\, dv, \quad \text{for } u \in [0, \infty]. \tag{1.23}$$

From the properties of the periodogram, $I_n(u)$, it seems reasonable to use this as an estimator for the spectral measure. From the previous section we know that the periodogram is not a consistent estimator, but Theorem 1.19 suggests that as an estimator for the spectral measure the periodogram will work fine. We will denote the *Estimated Spectral Measure* by $\hat{F}(u)$, where

$$\hat{F}(u) = \frac{2\pi}{n} \sum_{0 < \frac{2\pi j}{n} < u} I_n\left(\frac{2\pi j}{n}\right), \quad \text{for } u \in [0, \infty], \tag{1.24}$$

see **?**. The reason we define the sum over the values $2\pi j/n$ is that this provides the estimator with some nice and easy to use properties, see Corollary 1.17 and Chapter 4 in **?**. Before we state the next theorem regrading the asymptotic properties of the estimator (1.24), we need to introduce some new concepts.

Suppose $\boldsymbol{Y}(t)$, where $t = 0, \pm 1, \ldots$, is an r vector valued time series with components $Y_a(t)$ that satisfy $E[|Y_a|^r] < \infty$ for all $a = 1, \ldots, r$, then we define the *k-th order cumulant spectrum* $f_{a_1,\ldots,a_k}(u_1, \ldots, u_{k-1}) \equiv f_{Y_{a_1},\ldots,Y_{a_k}}(u_1, \ldots, u_{k-1})$ by

$$\begin{aligned} &f_{Y_{a_1},\ldots,Y_{a_k}}(u_1, \ldots, u_{k-1}) \\ &\qquad = (2\pi)^{-k+1} \sum_{h_1,\ldots,h_{k-1}=-\infty}^{\infty} c_{a_1,\ldots,a_k}(h_1, \ldots, h_{k-1}) \exp\left(-i \sum_{j=1}^{k-1} h_j u_j \right) \end{aligned} \tag{1.25}$$

for $-\infty < h_j < \infty$, $a_1, \ldots, a_k = 1, \ldots, r$, $k = 2, 3, \ldots$. The definition (1.25) can be extended to the case $k = 1$, by setting $f_a = c_a = E[X_a(t)]$, for $a = 1, \ldots, r$.

THEOREM 1.20. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a real valued series with true power spectrum $f_0(u)$ that satisfying Assumption 1.7 and let $\hat{F}(u)$ be as defined in (1.24). Then for $0 < u_1, \ldots, u_J < \pi$ define $(v_1, \ldots, v_J)$ where $v_j = \sqrt{n}(\hat{F}(u_j) - F_0(u_j))$, and*

$$F_0(u_j) = \int_0^{u_j} f_0(v)\, dv.$$

Then the vector $(v_1, \ldots, v_J)$ is asymptotical multivariate Gaussian with expectation zero and covariance matrix Σ with elements given by

$$\Sigma_{k,l} = 2\pi \int_0^{\min(u_k, u_l)} f_0(v)^2\, dv + 2\pi \int_0^{u_k} \int_0^{u_l} f_{YYYY}(v, w, -v, 0)\, dv\, dw.$$

In the next corollary we state what happens in the case of a Gaussian time series, note that the result follows almost directly from Theorem 1.20.

COROLLARY 1.21. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a real valued series with true power spectrum $f_0(u)$ that satisfying Assumption 1.8. For $0 < u_1, \ldots, u_J < \pi$ let $r_j(n)$ be an integer such that $r_j(n)$ satisfy $2\pi r_j(n)/n \to u_j$ as $n \to \infty$.*

i) *Then, for all choice of j, k, where $j \neq k$ we have that $I_n(u_j)$ is distributed as $f(u_j) \times \text{Exp}(1)$ and $I_n(u_j)$ and $I_n(u_k)$ are asymptotically independent as $n \to \infty$.*

ii) *Let $\hat{F}(u)$ be given by (1.24) and for $0 < u_1, \ldots, u_J < \pi$ define $(v_1, \ldots, v_J)$, where $v_j = \sqrt{n}(\hat{F}(u_j) - F_0(u_j))$ and $F_0(u_j)$ is the true spectral measure. Then as n approaches infinity the vector $(v_1, \ldots, v_J)$ converges in distribution to a multivariate Gaussian with expectation zero and covariance matrix Σ with elements given by*

$$\Sigma_{k,l} = 2\pi \int_0^{\min(u_k, u_l)} f_0(v)^2\, dv.$$

PROOF. i) For a sample of size n of $Y(t)$, where $t = 0, \ldots, n-1$, the *Finite Fourier Transformation* $d_n(u)$ is define as

$$d_n(u) = \sum_{t=0}^{n-1} \exp(-iut) Y(t) \ \ -\infty < u < \infty. \tag{1.26}$$

From **?** we have that for $(u_1, \ldots, u_k) \in (-\infty, \infty)$ the joint cumulant function of $(d_n(u_1), \ldots, d_n(u_k))$ is given by

$$\text{cum}(d_n(u_1), \ldots d_n(u_k)) = (2\pi)^{k-1} \Delta_n\left(\sum_{j=1}^{k} u_j\right) f_{Y \ldots Y}(u_1, \ldots, u_{k-1}) + O(1) \tag{1.27}$$

where $f_{Y \ldots Y}$ is given by (1.25) and $\Delta_n(u) = \sum_{t=0}^{n-1} \exp(-iut)$. Note that we have that $|\Delta(u)| = |\sin(nu/2)/\sin(u/2)|$, $\Delta_n(u) = T$, for $u \equiv 0 \pmod{2\pi}$, $|\Delta_n(u)| \leq 1/|\sin(u/2)|$ for all $u \in (-\infty, \infty)$ and $\Delta_n(2\pi r/n) = 0$ for an integer $r \not\equiv 0 \pmod{n}$. From the result (1.27) we find that the expectation and covariance of the finite Fourier transformation of a strict stationary time series $Y(t)$ are given by

$$E[d_n(u)] = \Delta_n(u) E[Y(t)]$$

and

$$\text{Cov}(d_n(u),\ d_n(v)) = 2\pi \Delta_n(u - v) f(u) + O(1).$$

Since $u_j = 2\pi r_j(n)/n$ and $0 < u_j < \pi$, for $j = 1, \ldots, J$, it follows that $E[d_n(u_j)] = 0$ and the variance and covariance satisfy $n^{-1}\operatorname{Var}(u_j) = n^{-1}\operatorname{Cov}(d_n(u_j),\ d_n(u_j)) = 2\pi f(u_j)$ and $n^{-1}\operatorname{Cov}(d_n(u_k),\ d_n(u_l)) = O(n^{-1})$, where $k \neq l$. Since $d_n(u_j)$ is a sum of Gaussian variables with complex weights, the finite Fourier transformation, $d_n(u_j)$, follows a complex Gaussian distributed with expectation 0 and variance $2\pi n f(u_j)$. From the properties of the complex Gaussian distribution we know that if X follows a complex Gaussian distribution, $X \sim N_c(\mu, \sigma^2)$, then $\operatorname{Re} X$ and $\operatorname{Im} X$ are independent and $\operatorname{Re} X \sim N(\operatorname{Re}\mu, \sigma^2/2)$ and $\operatorname{Im} X \sim N(\operatorname{Im}\mu, \sigma^2/2)$. Further properties of complex numbers gives that the periodogram (1.20) could be written as

$$\begin{aligned} I_n(u_j) = \frac{1}{2\pi n}\big|d_n(u_j)\big|^2 &= \frac{1}{2\pi n}\big|\operatorname{Re} d_n(u_j)\big|^2 + \big|\operatorname{Im} d_n(u_j)\big|^2 \\ &= f(u_j)\big|\operatorname{Re} d_n(u_j)/\sqrt{2\pi n f(u_j)}\big|^2 + \big|\operatorname{Im} d_n(u_j)/\sqrt{2\pi n f(u_j)}\big|^2 \end{aligned}$$

which implies that $I_n(u_j)$ is distributed as $f(u_j) \times \operatorname{Exp}(1)$ for $j = 1, \ldots, J$. Also $I_n(u_k)$ and $I_n(u_l)$, for $k, l = 1, \ldots, J$ where $k \neq l$, are asymptotically independent, which follows from the asymptotically independence of the finite Fourier transformation.

ii) The result follows as a direct consequence of Theorem 1.20 and Remark 1.4. □

All the results above follows the approach of **?**, where everything is defined according to the values $u_j(n) = 2\pi r_j(n)/n$ where $u_j(n) \to u_j$ as $n \to \infty$. Later in this thesis we will need equivalent results for general vales u and the next two results consider this slightly different idea.

THEOREM 1.22. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a real valued series satisfying Assumption 1.8 and with true spectral measure $F_0(u)$, then for any choice of u where $u \in (0, \pi)$*

$$\tilde{F}(u) = \int_0^u I_n(w)\, dw \xrightarrow{a.s} \int_0^u f_0(w)\, dw = F_0(u). \tag{1.28}$$

PROOF. An argument is given in **?** page 418. □

The following result is the natural extension of Corollary 1.21 to the case of general frequencies $u \in [0, \pi]$ and the new estimator $\tilde{F}(u)$ for the power spectrum.

LEMMA 1.23. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a real valued series with true power spectrum $f_0(u)$ satisfying Assumption 1.8 and let $\tilde{F}$ be as defined in Theorem 1.22. Let $u_1, \ldots, u_m$ be an arbitrary sequence of frequencies where $-\pi < u_i < \pi$ for $i = 1, \ldots, m$ and positive integer m. Let $v_j = \sqrt{n}(\tilde{F}(u_j) - F_0(u_j))$, for all $j = 1, \ldots, m$, then $(v_1, \ldots, v_m)$ is asymptotically multivariate Gaussian distributed with exception zero and covariance matrix Σ with elements*

$$\Sigma_{k,l} = 2\pi \int_0^{\min(u_k, u_l)} f_0(w)^2\, dw.$$

Also if $u, v \in (0, \pi)$ and $u \neq v$ then $\tilde{F}(u)$ and $\tilde{F}(v)$ are asymptoticaly independent.

PROOF. (Sketch) From Theorem 1.22 it is clear that all $(v_1, \ldots, v_m)$ will have expectation zero. If we write

$$\tilde{F}(u_j) = \int_0^{u_j} I_n(w)\, dw = \lim_{m \to \infty} (2\pi/m) \sum_{2\pi j/m < u_j} I_n(2\pi j/m)$$

it is easy to see that it follows from Corollary 1.21 that $(v_1, \ldots, v_m)$ will become asymptotically multivariate Gaussian distributed. All that remains to show is what happens with the covariance matrix of $(v_1, \ldots, v_m)$. From Theorem 1.16 we have an explicit expression for the covariance between to periodogram functions at different frequencies u_k and u_l, where $u_k, u_l \in [0, \pi]$, then

$$\mathrm{Cov}(I_n(u_k),\ I_n(u_l)) = \begin{cases} \mathrm{Cov}(I_n(u_k),\ I_n(u_l)) \leq 2f_0(u_k)^2, & \text{if } 0 < |u_k - u_l| < \epsilon_n \\ f_0(u_k)^2, & \text{if } u_k = u_l \\ 0, & \text{else} \end{cases}.$$

where $\epsilon_n \to 0$ as $n \to \infty$. Let $w_i = 2\pi i/m$ then it follows from **?** and Theorem 1.16 that

$$\begin{aligned} &\mathrm{Cov}(\tilde{F}(u_k), \tilde{F}(u_l)) \\ &\quad = \lim_{m\to\infty} (2\pi/m)^2 \sum_{w_i<u_k} \sum_{w_j<u_l} \mathrm{Cov}(I_n(w_i), I_n(w_j)) \\ &\quad = \lim_{m\to\infty} (2\pi/m)^2 \Bigg(\sum_{w_i<\min(u_k,u_l)} f_0(w_i)^2 \\ &\qquad + n^{-2} \sum_{i\neq j} \sum_{j\neq i} \mathrm{Cov}(I_n(w_i),\ I_n(w_j)) + 2\pi n f_{YYYY}(w_i, w_j, -w_i, 0) \Bigg) + O(n^{-2}) \\ &\quad = \lim_{m\to\infty} (2\pi/m)^2 \Bigg(\sum_{w_i<\min(u_k,u_l)} f_0(w_i)^2 \\ &\qquad + n^{-2} \sum_{i=1}^{m} \sum_{j\in U_i} \mathrm{Cov}(I_n(w_i),\ I_n(w_j)) + 2\pi n f_{YYYY}(w_i, w_j, -w_i, 0) \Bigg) + O(n^{-2}). \end{aligned}$$

where $U_i = \{j \,|\, j = 1, \ldots, m \text{ and } |w_i - w_j| < \epsilon_n\}$. Assume that n approaches infinity so fast that $n/m \to 1$ and $U_i = \emptyset$, then since we have limited ourselves to the study of Gaussian time seres, we have that

$$\lim_{n\to\infty} n\,\mathrm{Cov}(\tilde{F}(u_k),\ \tilde{F}(u_l)) = \lim_{m,n\to\infty} n\,(2\pi/m)^2 \sum_{w_i<\min(u_k,u_l)} f_0(w_i)^2 = 2\pi \int_0^{\min(u_k,u_l)} f_0(w)^2\,dw,$$

which completes the proof. □

The last result in this section will be a generalization of Lemma 1.23 to covariance functions, which still has the main focus in this thesis.

COROLLARY 1.24. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a real valued series hat satisfying Assumption 1.8 and has true covariance function $C_0(h)$ based on the power spectrum $f_0(u)$. Define the estimator $\tilde{C}(h)$ for the true covariance function by*

$$\tilde{C}(h) = \int_{-\pi}^{\pi} \exp(-iuh) I_n(u)\,du = 2\int_0^{\pi} \cos(uh) I_n(u)\,du. \tag{1.29}$$

Let $h_1, \ldots, h_m$, where $h_j = 0, \pm 1, \pm 2, \ldots$ and $j = 1, \ldots, m$, be an arbitrary set of lag distance and define $v_j = \sqrt{n}[\tilde{C}(h_j) - C_0(h_j)]$, for $j = 1, \ldots, m$. Then $(v_1, \ldots, v_m)$ is distributed according to a multivariate Gaussian distribution with expectation zero and covariance matrix Σ with elements given by

$$\Sigma_{k,l} = 4\pi \int_0^{\pi} [\cos(u(h_k - h_l)) + \cos(u(h_k + h_l))] f_0(u)^2\,du.$$

PROOF. (Sketch) The asymptotic normality follows from Lemma 1.23 and it is clear that the expectation of $\tilde{C}(h)$, for any $i = 1, \ldots, m$ is the true covariance function

$$E[\tilde{C}(h_j)] = \int_{-\pi}^{\pi} \exp(-iuh_j) E[I_n(u)]\, du = 2\int_0^{\pi} \cos(uh_j) f_0(u)\, du = C_0(h_j).$$

The remaining thing to show is the covariance between two arbitrary lag distances h_k and h_l. We will use the same approach as in Lemma 1.23, define

$$\tilde{C}(h_j) = \int_{-\pi}^{\pi} \exp(-iuh_j) I_n(u)\, du = \lim_{m\to\infty} \frac{4\pi}{m} \sum_{i=1}^{m} \cos(u_i h_j) I_n(u_i), \quad \text{where } u_i = 2\pi i/m$$

then

$$\begin{aligned}
&\operatorname{Cov}(\tilde{C}(h_k),\ \tilde{C}(h_l)) \\
&\quad = \lim_{m\to\infty} \operatorname{Cov}\Big(\frac{4\pi}{m} \sum_{i=1}^{m} \cos(u_i h_k) I_n(u_i),\ \frac{4\pi}{m} \sum_{i=1}^{m} \cos(u_i h_l) I_n(u_i)\Big) \\
&\quad = \lim_{m\to\infty} \frac{16\pi^2}{m^2} \sum_{i=1}^{m}\sum_{j=1}^{m} \cos(u_i h_k)\cos(u_j h_l) \operatorname{Cov}(I_n(u_i),\ I_n(u_j)) \\
&\quad = \lim_{m\to\infty} \frac{8\pi^2}{m^2} \sum_{i=1}^{m}\sum_{j=1}^{m} \big[\cos(u_i h_k - u_j h_l) + \cos(u_i h_k + u_j h_l)\big] \operatorname{Cov}(I_n(u_i),\ I_n(u_j)).
\end{aligned}$$

Under the same assumption as in Lemma 1.23 we know that if n approaches infinity fast enough, i.e. such that $n/m \to 1$ and the sets U_i defined in Lemma 1.23 are the empty sets for all $i = 1, \ldots, m$, it follows that

$$\begin{aligned}
&n \operatorname{Cov}(\tilde{C}(h_k),\ \tilde{C}(h_l)) \\
&\quad = n \lim_{m\to\infty} \frac{8\pi^2}{m^2} \sum_{i=1}^{m}\sum_{j=1}^{m} \big[\cos(u_i h_k - u_j h_l) + \cos(u_i h_k + u_j h_l)\big] \operatorname{Cov}(I_n(u_i),\ I_n(u_j)) \\
&\quad = 4\pi \int_0^{\pi} \big[\cos(u(h_k - h_l)) + \cos(u(h_k + h_l))\big] f_0(u)^2\, du.
\end{aligned}$$

□

As a final remark we will state another natural choice of estimator for covariance function.

REMARK 1.25. Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a real valued time series that satisfies the conditions of Theorem 1.22. Then

$$\begin{aligned}
\tilde{C}(h) = \int_{-\pi}^{\pi} \exp(-iuh) d\tilde{F}(u) &= 2\int_0^{\pi} \cos(uh)\big[\tilde{F}(u + du) - \tilde{F}(u)\big] \\
&\xrightarrow{a.s} 2\int_0^{\pi} \cos(uh)\big[F_0(u + du) - F_0(u)\big] = 2\int_0^{\pi} \cos(uh) dF_0(u) = C_0(h),
\end{aligned}$$

for $h = 0, \pm 1, \pm 2, \ldots$.

2. Bayesian methods

We have already mentioned that there are two things that determine a stationary time series, this is the expectation and the dependency structure. The easiest, and perhaps most common, way to model these, given a sample of observations, is to estimate the expectation by the sample mean and assume a parametric class of covariance, correlation or semivariogram functions for the dependency and estimate the unknown parameters from the data. It is of course possible to use Bayesian methods where the unknown parameters under study are assumed to be random variables from a suitable a priori distribution that reflects our a priori belief, see **?** and **?** for a complete introduction and discussion of the Bayesian approach to statistical inference. By restricting ourselves to a specific class of parametric models we may make critical assumptions about the phenomena under study that may lead to incorrect conclusions and results. To avoid the perhaps crucial parametric model assumption we can use the more sensitive and robust class of nonparametric models, see **?** and **?**.

In this section we will introduce and discuss how to use nonparametric Bayesian modeling to make statistical inference about the dependency structure in stationary time series. In Section 2.1 there will be a short introduction to nonparametric and semiparametric Bayesian methods with some focus on distribution function estimation, because of its historical value and the similarity to our nonparametric approach to the dependency estimation in stationary time seres. In Section 2.2 we will introduce some a priori distribution for the class of covariance and the correlation functions and discuss some of its properties. Section 2.3 will mostly be concerned with how we can obtain posterior inference after a sample from a stationary time series is observed through the use of Markov chain Monte Carlo simulation, also known as MCMC simulation, see Appendix A.

2.1. Nonparametric and semiparametric Bayesian methods. Nonparametric Bayesian models usually refer to Bayesian constructions with large numbers on unknown parameters, often more than the number of observations or even infinitely many. It is therefore in some situations natural to think of nonparametric Bayesian models as probability models on a suitable function space. In nonparametric Bayesian analysis we do not make any model assumptions and we will therefore avoid possible misclassifications as a result of an incorrect parametric model. We are also able to include valuable a priori information which may lead, if our prior belief is not far off, to more accurate models and precise inference. The nonparametric Bayesian approach has become quite popular and is used in several different types of analysis, and with fast modern computers it is now usually straightforward and fairly easy to obtain posterior inference through simulation. For a brief introduction to some of the fields where nonparametric Bayesian modeling is used see **?**.

The term semiparametric model is used when referring to models that consist of parametric and nonparametric parts. A semiparametric Bayesian setup for a stationary time seres could perhaps be the case where a nonparametric Bayesian model was used to estimation the correlation function while the mean and variance where assumed to be unknown random parameters from a prior density.

As an illustration we will very short introduce and explain the nonparametric approach to distribution function estimation. Let $Y_1, \ldots, Y_n$ be a random sample of independent and identically distributed variables from an unknown distribution function F, the Bayesian approach requires a probability model for the unknown distribution function F. In a nonparametric Bayesian framework this involves placing a probability measure on the collection of distribution functions, i.e viewing F as a stochastic process, see **?**. Such probability measures are also often referred to as *random probability measures*. **?** introduced the Dirichlet process as a random probability measure and a solution to the nonparametric Bayesian distribution function estimation. The random distribution function F is said to be generated by a *Dirichlet process* if for any partition $B_1, \ldots, B_k$ of the sample space, the joint random vector $(F(B_1), \ldots, F(B_k))$ follows a Dirichlet distribution with parameters $(aF_0(B_1), \ldots, aF_0(B_k))$, where F_0 is our a priori guess for F and a is a weight representing our a priori precision. The interpretation of F_0 and a becomes clear from the expressions for the expectation and variance for a given subset B_i, from the properties of the Dirichlet distribution we have that

$$E[F(B_i)] = F_0(B_i) \quad \text{and} \quad \text{Var}(F(B_i)) = F_0(B_i)(1 - F_0(B_i)/(a+1).$$

? also suggest two properties that would be desirable for a random probability measures to satisfy: i) the support of the prior distribution should be large, and ii) the posterior distribution obtained given a sample of observations from the true probability distribution should be manageable analytically. Today, 35 years later, property ii) could be replaced by a requirement that a efficient simulation routine must be available.

See **?** and **?** for a more complete treatment of distribution function estimation and **?** for a extension of property i) by the use of mixed models. Other interesting articles concerned with the topic of nonparametric Bayesian modeling are **?**, **?** and **?**.

2.2. Prior distributions. In the following sections we will try to apply the nonparametric Bayesian ideas as an approach for estimating the dependency structure in stationary Gaussian time series. To be more precise, our main goal will be to model the covariance function from the spectral measure by using ideas similar to the nonparametric Bayesian approach to density estimation. The reason why we will model the covariance through the spectral measure will soon become clear. We have already mentioned that in order to use the nonparametric Bayesian design we need to place a probability measure on the function space consisting of all valid covariance functions, this is done by specifying a random process as the prior distribution for the covariance function, i.e. we will view the covariance function as a random process. Before we start the real discussion on how to construct suitable nonparametric prior distributions, we will show that a fairly simple and reasonable construction might lead to improper and useless results.

EXAMPLE 2.1. Given a sample form a stationary Gaussian time series and suppose we are only interested in the unknown covariance structure. Let the covariance functions that are within the class of exponential functions be our prior guess. Then the set of covariance functions of main interest is given by

$$\{C(h) = \sigma^2 \exp(-\phi h) \mid \phi > 0 \text{ and } h = 1, 2, \ldots\}.$$

Note that this is a special case of the covariance function given in Example 1.11 and therefore we know that all the functions in the set are valid. We now need to place a probability measure on

the space of valid covariance functions and try to somehow concentrate it around the exponential, which means we have to define a suitable stochastic process as the prior distribution for $C(h)$. Assume that σ^2 is known and define the random $C(h)$ by

$$C(h) = \sigma^2 \exp(-G(h)), \quad \text{for } h = 0, \pm 1, \ldots$$

where $G(h)$ might for example be a Gamma Process. To be more specific we will concentrate the process around $G_0(h) = \phi h$, where ϕ is determined from our a priori information. It is clear that the random function $C(h)$ is close to our prior belief $C_0(h) = \sigma^2 \exp(-\phi h)$ and the method is quite easy to use. Unfortunately the random covariance functions described here will not necessarily satisfy the positive definite condition, in other words, the random function $C(h)$ will not always become a valid covariance function, i.e. $\Pr\{C(h) \text{ is a valid covariance function}\} < 1$. The easiest way to test if $C(h)$ is valid is to use another method than the definition (1.2) to check the positive definiteness property.

If a covariance function $C(h)$ is valid, then for any $k \in \mathcal{N}$ and vector of locations $(0, 1, \ldots, k)$ the covariance matrix Σ obtained from $C(h)$, i.e. element i, j of Σ is given by $\Sigma_{i,j} = C(|i - j|)$, will only have positive eigenvalues.

Let $C(h)$ be a random covariance function from the construction above, it is then easy to check with a statistical software package, that for $k = 3$ it is possible for the corresponding covariance matrix to have negative eigenvalues with a positive probability. This implies that the method described above is not a correct routine to use in order to obtain random covariance functions.

Example 2.1 illustrates the fact that we have to be a little careful when we construct prior distributions for nonparametric covariance functions since we always have to make sure that we are within the class of valid functions. Motivated from Corollary 1.10 it makes sense to limit ourselves to the set of covariance functions that are constructed from the spectral measures $F(u)$ defined by (1.9). By doing this we will have some simple and quite week conditions our process need to satisfy in order to verify that our prior distribution will generate valid covariance function. Since we know how to find the spectral measure, given a specific covariance function, it will become easy to see what we have to concentrate our a priori spectral measure around. If the posterior process is updated in the frequency domain we can also always be sure that our posterior covariance function will become a valid function. Note that this limitation is not necessarily a problem, or a real restriction (see Remark **??**), first of all it will become quite easy to include our a priori information regarding the covariance function by equation 1.7 and as we shall see it is not to difficult to obtain posterior inference by using MCMC simulations.

Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a stationary Gaussian time series with expectation zero and true covariance function $C_0(h)$, where $h = 0, \pm 1, \ldots$, determined by the spectral measure $F_0(u)$, where $u \in [0, \pi]$. In order to use the nonparametric Bayesian framework we need to place a probability measure on the space of covariance functions. As we have argued earlier we will do this by placing a probability measure on the collection of spectral measures in order to always be sure that we are only working with the valid covariance functions. In other words we have to view F as a stochastic process and it is clear from equation (1.8) that $C(h)$ constructed in this way also will become a valid random covariance function.

Let $\bar{F}(u)$ be a function on the interval $[-\pi, \pi]$ defined by the function $F(u)$, on $u \in [0, \pi]$, in the following way

$$\bar{F}(u) = \begin{cases} 0 & u = -\pi \\ F(\pi) - F(u) & -\pi < u < 0 \\ F(\pi) + F(u) & 0 \leq u \leq \pi \end{cases} . \tag{2.1}$$

Let $\mathcal{F}$ be the set of all functions $F(u)$ such that $\bar{F}(u)$ defined according to (2.1) and satisfy the conditions of Corollary 1.10, i.e. $\bar{F}(u)$ is non-decreasing, $\bar{F}(-\pi) = 0$ and $\bar{F}(\pi) < \infty$. Then $\mathcal{F}$ is the set of spectral measures with fundamental domain $[0, \pi]$ that will lead to valid covariance function and $\mathcal{F}$ is the set of functions which we will need to place a probability distribution. Before we continue, note that we can in almost the same way construct the set $\mathcal{H}$ of normalized spectral measures with fundamental domain $[0, \pi]$ that will lead to valid correlation functions, see Theorem 1.9. We will motivate the general idea by a simple example of how we can construction random spectral measures as random step functions with a finite number of jumps.

EXAMPLE 2.2. Let $0 = u_0 < u_1 < \ldots < u_{M-1} < u_M = \pi$ define a partition of the interval $[0, \pi]$ and let $(\Delta F(u_1), \ldots, \Delta F(u_m))$, where $\Delta F(u_i) = F(u_i) - F(u_{i-1})$, be a vector of independent random variables, where $\Delta F(u_i)$ has prior density function given by $\pi_i(v_i)$ and let the $\pi(\boldsymbol{v}) = \prod_{i=1}^{m} \pi_i(v_i)$ define the joint density function for the whole vector. Let $F(u)$ be the random function defined as

$$F(u) = \sum_{u_i < u} \Delta F(u_i), \quad \text{for } 0 < u < \pi.$$

Note that $F(u)$ defines a random step function on the interval $[0, \pi]$. Assume that the distributions of $\Delta F(u_i)$, for all $i = 1, \ldots, m$, is such that $F(u)$ satisfies the conditions of Corollary 1.10, $F(-\pi) = 0$, $F(\pi) < \infty$ and $F(u)$ have to be non-decreasing, i.e. $\Delta F(u_i) \geq 0$ and $\Delta F(u_i)$ must be bounded in probability for all $i = 1, \ldots, m$. If the function $F(u)$ is extended in the natural way to the interval $[-\pi, \pi]$, by $F(-\pi) = 0$ and $F(-u) \equiv F(u)$ for $-\pi < u < 0$, then $F(u)$ becomes a random spectral measure, it is clearly in the set $\mathcal{F}$. The random covariance function $C(h)$ constructed by

$$C(h) = \int_{-\pi}^{\pi} \exp(iuh)\, dF(u) = 2 \sum_{i=1}^{m} \cos(u_i h) \Delta F(u_i)$$

will from Corollary 1.10 always become a valid covariance function. The density $\pi(\boldsymbol{v})$ will be specified according to our prior beliefs, i.e.

$$E[F(u)] = E[\sum_{u_i < u} \Delta F(u_i)] = \sum_{u_i < u} E[\Delta F(u_i)] = \sum_{u_i < u} \Delta F_0(u_i) = F_0(u)$$

and

$$\mathrm{Var}(F(u)) = \mathrm{Var}(\sum_{u_i < u} \Delta F(u_i)) = \sum_{u_i < u} \mathrm{Var}(\Delta F(u_i)) = \sigma^2(u).$$

where $F_0(u)$ is the spectral measure that corresponds to $C_0(u)$, i.e

$$E[C(h)] = 2 \sum_{i=1}^{m} \cos(u_i h) E[\Delta F(u_i)] = C_0(h).$$

In order to place a probability measure on the collection of functions contained in $\mathcal{F}$, we will have to specifically a random process that will generate random functions F, defined on the interval $[0, \pi]$ that when extended in the natural way to $[-\pi, \pi]$ will satisfy the conditions of Corollary

1.10. There is no reason not to believe that there should not exist several types of processes that will be suitable to generate such random spectral measures.

We can define a quite general class of possible prior distributions by letting F be a finite Lévy process, i.e. a nonnegative independent increment process, that satisfies $\Pr\{F(\pi) < \infty\} = 1$. In order to be specific we will assume that the random spectral measures F is a Gamma process defined on the interval $[0, \pi]$.

A random process $G(u)$ with increments given by $dG(u) = G(u + du) - G(u)$ will be referred to as a *Gamma process*, if $G(u)$ is an independent increment process where the increments are distributed according to a Gamma distribution, i.e.

$$dG(u) \sim \text{Ga}(\alpha(u)\, du,\ \beta(u)), \quad \text{for } u > 0 \tag{2.2}$$

where $\text{Ga}(\alpha(u)\, du,\ \beta(u))$ is the Gamma distribution with shape parameter $\alpha(u)\, du > 0$ and rate parameter $\beta(u) > 0$, for $u > 0$. The Gamma process is especially a Lévy process and is therefore a nonnegative independent increment process. The process will also be finite and from simple expressions for the expectation and variance it will become quite easy to translate and include our a priori information about the dependency into the prior distribution. See for example **?**, **?** or **?** for a complete definition and different uses of the Gamma processes.

Let now F be a Gamma process with increments given as $dF(u) = F(u+du) - F(u)$, and where the increment distribution is specified by

$$dF(u) \sim \text{Ga}(\beta(u)\ dF_0(u), \beta(u)),$$

where $dF_0(u) = F_0(u+du) - F_0(u)$ and $F_0(u) \in \mathcal{F}$ represent our prior belief. The rate parameter $\beta(u)$ will determine the precision or strength in our a priori guess, see below. The random function defined by

$$F(u) = \int_0^u dF(u), \quad \text{for } u \in [0, \pi] \tag{2.3}$$

will be referred to as a *random spectral measure* and a *random covariance function* will be the random function

$$C(h) = 2 \int_0^\pi \cos(uh)\, dF(u), \quad \text{for } h = 0, \pm 1, \pm 2, \ldots. \tag{2.4}$$

From the properties of the Gamma Distribution and the independence of the increments it is now straightforward to determine how the a priori information shall be included in the model. For every $h = 0, 1, 2, \ldots$ we have that

$$\begin{aligned} E[C(h)] &= E[2 \int_0^\pi \cos(uh)\, dF(u)] \\ &= 2 \int_0^\pi \cos(uh)\, E[dF(u)] = 2 \int_0^\pi \cos(uh)\, dF_0(u) = C_0(h), \end{aligned}$$

and the precision will be determined from the variance

$$\begin{aligned} \text{Var}(C(h)) &= 4 \int_0^\pi \cos(uh)^2\ \text{Var}(dF(u)) \\ &= 2 \int_0^\pi (1 + \cos(2uh))\, dF_0(u)/\beta(u) \\ &= 2 \int_0^\pi \beta(u)^{-1}\, dF_0(u) + 2 \int_0^\pi \cos(u2h)\beta(u)^{-1}\, dF_0(u) \end{aligned} \tag{2.5}$$

There is one problem with the construction presented here, the expression for the variance, that represents our strength in the a priori guess, consists of two parts where the first term does not depend on h and will therefore become a constant. If we assume for a moment that the rate parameter is a constant, i.e. $\beta(u) = \beta$, then the last term of expression (2.5) will become a scaled version of $C_0(h)$ that approaches zero twice as fast, this means that

$$\mathrm{Var}(C(h)) \to 2\int_0^\pi \beta^{-1}\, dF_0(u) = 2F(\pi)/\beta, \quad \text{as } h \to \infty.$$

In other words, suppose we are particularly certain that the dependence will become arbitrary close to zero as the separation between two different time points becomes large. We will then probably wish to include this information in our model, but this will not be possible in the given framework. The model we have described so far is in a sense "not good enough", since it is impossible to capture a very common belief about the dependency structure. Note that there is at least one obvious reason why it might be desirable to let $b(u)$ be an increasing function and not a constant. The high end of the spectrum corresponds to short and rapid oscillations in the covariance, then by forcing the posterior process to be close to the prior process for high frequencies we might smooth out some of the noise from the observation. There is one important topic we have not discussed, this is the support of our construction, unfortunately it was not enough time to discuss this properly in this thesis, see the concluding remark at the end of the next chapter for a remark on this subject. We will illustrate the ideas we have presented so far with two examples that also have a solution to the problem with the constant variance.

Example 2.3. Assume that our prior guess for the covariance function is the class of functions given by

$$\mathrm{Cov}(Y(l),\, Y(k)) = C_0(|l-k|) = C_0(h) = \sigma^2\rho^h, \quad \text{for } h = 0, 1, 2, \ldots.$$

We will now follow the ideas from the discussion above and define a prior distribution on the spectral measure by a Gamma process and we will specify the distribution of the increments to reflect our prior belief. Let $dF(u) = F(u + du) - F(u)$ and $dF(u) \sim \mathrm{Ga}(\alpha(u)\, du, \beta(u))$, where $\alpha(u) = b(u)f_\pi(u)$ and $\beta(u) = b(u)$, where $f_\pi(u)$ is the power spectrum from Example 1.11 (the index π indicates that this is our a priori guess) and $b(u)$ is a positive real valued function. Then the random covariance function is given by

$$C(h) = 2\int_0^\pi \cos(uh)\, dF(u), \quad \text{for } h = 0, 1, 2, \ldots$$

which will always become a valid covariance function. For any h, where $h = 0, 1, 2, \ldots$, the expectation of $C(h)$ at h is given by the expression

$$E[C(h)] = 2\int_0^\pi \cos(uh)\, E[dF(u)] = 2\int_0^\pi \cos(uh) f_\pi(u)\, du = \sigma^2\rho^h$$

and the variance by

$$\begin{aligned}&\mathrm{Var}(C(h))\\ &\quad = 4\int_0^\pi \cos(uh)^2\, \mathrm{Var}(dF(u)) = 2\sigma^2\int_0^\pi b(u)^{-1} f_\pi(u)\, du + 2\sigma^2\int_0^\pi \cos(u2h) b(u)^{-1} f_\pi(u)\, du.\end{aligned}$$

We have already mentioned the problem related to the variance expression, that it reaches a finite limit for large values of h, in particular this situation the variance problem has an easy

solution. First we will assume that $b(u) = b$ to simplify and make the solution more explicit, then we can rewrite the expression for the variance

$$\mathrm{Var}(C(h)) = \frac{\sigma^2}{b}\big[2F_\pi(\pi) + \rho^{2h}\big].$$

From Lemma 1.1 we know that the product of two valid covariance functions is also always a valid covariance function, therefore if $dF(u) \sim \mathrm{Ga}(\alpha'(u)\,du, \beta(u))$, where $\alpha'(u) = bf_\pi(u)$ and $\beta(u) = b$, where $f_\pi(u)$ is now a new prior guess that corresponding power spectrum for the covariance function $C(h) = \sigma^2\rho^{ah}$, which is a straightforward and easy generalization of Example 1.11. If we now combine the new random covariance function with a suitable choice of deterministic function we obtain the two following results:

$$E[\rho^{h(1-a)}C(h)] = \rho^{h(1-a)}2\int_0^\pi \cos(uh)f_\pi(u)\,du = \sigma^2\rho^{h(1-a)}\rho^{ha} = \sigma^2\rho^h$$

and

$$\begin{aligned}\mathrm{Var}(\rho^{h(1-a)}C(h)) &= \rho^{2h(1-a)}2\sigma^2\int_0^\pi b(u)^{-1}f_\pi(u)\,du + 2\sigma^2\int_0^\pi \cos(u2h)b(u)^{-1}f_\pi(u)\,du \\ &= \frac{\sigma^2}{b}\big[2\rho^{2h(1-a)}F_\pi(\pi) + \rho^{2h}\big].\end{aligned}$$

This finale result gives us a quite large and flexible class of prior distributions for the covariance functions that are of the form $C(h) = \sigma^2\rho^h$. By adjusting the four hyper parameters (a, b, ρ, σ) we are able to construct processes that are able to capture several different types of a priori beliefs. The two parameters σ and ρ represent our prior guess for the unknown covariance function, b controls the largest possible amount of variation, or the precision in our guess, and a determines how fast this variation should converges to zero for increasing separation. There will be some numerical illustrations based on simulations at the end of the next section.

In the following example we will show have we can use the Dirichlet process as a prior distribution on the set of correlation functions.

EXAMPLE 2.4. Let $\mathcal{H}$ be the set of functions $H(u)$ defined on the interval $[0, \pi]$ such that $\bar{H}(u)$, defined by $H(u)$ in the equivalent way as (2.1), on the interval $[-\pi, \pi]$ satisfies the conditions of Theorem 1.9. It is clear that every function $H(u)$ in $\mathcal{H}$ will lead to valid a correlation function $R(h)$ by the expression

$$R(h) = \int_{-\pi}^{\pi} \exp(iuh)\,d\bar{H}(u) = 2\int_0^\pi \cos(uh)\,dH(u), \quad \text{for } h = 0, 1, \ldots.$$

From section 2.1 it is clear that the Dirichlet process will be perfect as a prior distribution since the normalized spectral measure has the same properties as a cumulative distribution function on the interval $[-\pi, \pi]$. We will assume that H is a Dirichlet process and refer to it as a *random normalized spectral measure* and this will be our prior distribution on the set of normalized spectral measures. The covariance function constructed from the random spectral measure

$$R(h) = 2\int_0^\pi \cos(uh)\,dH(u), \quad \text{for } h = 0, 1, \ldots,$$

will be called a *random correlation function* and from Theorem 1.9 we known that $R(h)$ constructed in this way always will become valid.

The Dirichlet process is specified by two parameters, $H_\pi(u)$ and a, that represents the a priori belief and the precision, from the properties of the Dirichlet distribution it is easy to see how we should specify theses parameters.

$$E[R(h)] = 2\int_0^\pi \cos(uh)\, E[dH(u)] = 2\int_0^\pi \cos(uh)\, dH_\pi(u) = R_\pi(h), \quad \text{for } h = 0, 1, \ldots.$$

where $R_\pi(h)$ is our prior guess for the correlation function. Since the increments of a Dirichlet process is not independent the variance of the estimate will not have such a simple structure as in Example 2.3. From Appendix B we know how to find the variance of a finite number of Dirichlet distributed variables, by taking the limit we find that

$$\begin{aligned}
\mathrm{Var}(R(h)) &= 4\int_0^\pi \int_0^\pi \cos(uh)\, \cos(vh)\, \mathrm{Cov}(dH(u),\, dH(v)) \\
&= \frac{4}{1+a}\int_0^\pi \cos(uh)^2\, dH_\pi(u) - \frac{4}{1+a}\int_0^\pi \cos(uh)\, dH_\pi(u) \int_0^\pi \cos(vh)\, dH_\pi(v) \\
&= \frac{1}{1+a}\big[2H_\pi(\pi) + R_\pi(2h) - R_\pi(h)^2\big] \quad \text{for } h = 0, 1, \ldots.
\end{aligned}$$

Again we have the same problem as in Example 2.3, the variance tends to a finite limit as $h \to \infty$, if we have the simple structure as in the previous example, for example $R(h) = \rho^{h^k}$, where $k = 1, 2, \ldots$ and $h = 0, 1, 2, \ldots$, we can use the same trick where we combine a random and a deterministic covariance functions.

In these two examples we have shown how we in general can construct prior distributions covariance and correlation functions. We did also have a concrete solution to the problem with the constant variance. Unfortunately I do not have a general solution to this problem, depending on the situation we can always use the trick presented in the examples, but this will not always work and it is not difficult to come up with examples where this solution is not possible, see also the concluding remarks of the next chapter.

2.3. Posterior distributions. In this section we will investigate the properties of the posterior distribution and try to make inference about the dependency structure after a sample from a stationary time series is observed. In the case of finite parametric problems the inference from the posterior distribution is obtained by using the fact that the posterior density is proportional to the product of the prior density and the likelihood, i.e.

$$\pi(\boldsymbol{\theta}|\text{data}) \propto \pi(\boldsymbol{\theta}) \times \text{Lik}(data|\boldsymbol{\theta}).$$

As a introduction we will first complete the the discussion of the simple example we introduced in the previous section, this falls into the class of finite parameter problems and we may obtain the posterior distribution as described above.

EXAMPLE 2.5. Let $Y(0), \ldots, Y(n-1)$ be a sample from a stationary Gaussian time series with expectation zero where the covariance structure and the true spectral measure is unknown. We will assume that the unknown spectral measure $F(u)$ is a step function with a finite number of jumps located at points $0 < u_1 < \cdots < u_{M-1} < \pi$, where M is much smaller than n. We can then use the result from Example 2.2 where we defined a prior density for $F(u)$ by the function $\pi(F) = \pi_1(\Delta F(u_1)) \cdots \pi_M(\Delta F(u_M))$, where $\pi_i(\Delta F(u_i))$ is the prior density for the unknown

and random variable $\Delta F(u_i) = F(u_i) - F(u_{i-1})$, for $i = 1, 2, \ldots, M$. The posterior distribution is now proportional to

$$\pi(F|\text{data}) \propto \prod_{i=1}^{m} \pi_i(\Delta F(u_i)) \times \text{Lik}(\text{data}|F).$$

From log-likelihood given by equation (2.6) below we have that the log-posterior density of the spectral measure is given by

$$\log(\pi(F|\text{data})) = \sum_{i=1}^{m} \log(\pi(\Delta F(u_i))) - \frac{1}{2}\log(|\Sigma(F)|) - \frac{1}{2}(\boldsymbol{Y})^T\Sigma(F)^{-1}(\boldsymbol{Y}) + c^*,$$

where c^* is a constant, $\boldsymbol{Y}^T = (Y(0), \ldots, Y(n-1))$ and Σ_F is the covariance matrix that with elements of the form

$$\Sigma(F)_{k,l} = 2\sum_{i=1}^{m} \cos(u_i|k-l|)\Delta F(u_i).$$

It is not easy or straightforward to determine the properties of the posterior distribution, the parameters under interest are hidden within the covariance matrix and it is not clear how the observed data will affect them. The easiest and perhaps most common solution in order to solve such complicated problems and to make posterior inference is to use MCMC methods, see Appendix A.

Let $Y(0), \ldots, Y(n-1)$ be a sample of size n from a real valued stationary Gaussian time series $Y(t)$ with expectation $E[Y(t)] = \mu$ and true covariance function $C_0(h)$, where $h = 0, \pm 1, \pm 2, \ldots$, determined by the spectral measure $F_0(u)$ with fundamental domain $[0, \pi]$. To be concrete we will assume that the time series $Y(t)$ has expectation zero and that the covariance function is isotropic. The likelihood function of $Y(0), \ldots, Y(n-1)$ is a function of the unknown spectral measure $F(u)$ and is given by the multivariate Gaussian distribution as follows

$$\text{Lik}(\text{data}|F) = (2\pi)^{-n/2}|\Sigma(F)|^{-1/2}\exp\left(-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{0})^T\Sigma(F)^{-1}(\boldsymbol{Y}-\boldsymbol{0})\right)$$

where $\boldsymbol{Y}^T = (Y(0), \ldots, Y(n-1))$ and the elements of the covariance matrix $\Sigma(F)$ depends on the spectral measure through

$$\Sigma(F)_{k,l} = C(|k-l|) = \int_{-\pi}^{\pi} \exp(iu|k-l|)\, dF(u) = 2\int_0^{\pi} \cos(u|k-l|)\, dF(u),$$

where $k, l = 0, 1, \ldots, n$. Further the logarithm of the likelihood function is now determine by the function

$$\begin{aligned} L_n(F) = \text{lik}(\text{data}|F) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma(F)|) - \frac{1}{2}(\boldsymbol{Y})^T\Sigma(F)^{-1}(\boldsymbol{Y}) \\ &= -\frac{1}{2}\log(|\Sigma(F)|) - \frac{1}{2}(\boldsymbol{Y})^T\Sigma(F)^{-1}(\boldsymbol{Y}) + c^* \end{aligned} \tag{2.6}$$

where c^* is a constant.

In the nonparametric situation we are concerned with processes and we have in a sense infinitely many unknown parameters. From **?** Chapter 1 we know that in order to specify the posterior distribution for the whole process it is sufficient to specify the distribution for an arbitrary finite set of increments of the process. Further, also from **?** Chapter 1, we have that in order to specify the process for any finite set of increments it is actually sufficient to specify the process for any finite partition of its domain.

Let k be any positive integer and $0 = u_0 < u_1 < \ldots < u_k = \pi$ be a partition of the interval $[0, \pi]$ and let $(v_1, \ldots, v_k)$ be the set of increments, where v_i is defined as $v_i = \Delta F(u_i) = F(u_i) - F(u_{i-1})$, then the posterior distribution of $v_1, \ldots, v_k$ given a sample of observation has density proportional to

$$\pi(v_1, \ldots v_k|\text{data}) \propto \prod_{i=1}^{k} \pi(v_i) \times \text{Lik}(\text{data}|v_1, \ldots, v_k). \tag{2.7}$$

If the $Y(0), \ldots, Y(n-1)$ is a sample from a stationary Gaussian time series with expectation zero and true spectral measure $F_0(u)$, then the logarithm of the posterior density for any finite set of increments is given as

$$\log(\pi(v_1, \ldots v_k|\text{data})) = \sum_{i=1}^{k} \pi(v_i) - \frac{1}{2}\log(|\Sigma(v_1, \ldots, v_k)|) - \frac{1}{2}(\boldsymbol{Y})^T \Sigma(v_1, \ldots, v_k)^{-1}(\boldsymbol{Y}) + c^*$$

where c^* is a constant, $\boldsymbol{Y} = (Y(0), \ldots, Y(n-1))$ and $\Sigma_{\boldsymbol{v}}$ is the covariance matrix that with elements of the form

$$\Sigma(v_1, \ldots, v_k)_{k,l} = 2\sum_{i=1}^{m} \cos(u_i|k-l|)v_i, \quad \text{where } k, l = 0, 1, \ldots, n.$$

From the expression (2.7) it is not obvious how the posterior process is updated, but it will certainly not be as refined as with the Dirichlet process and distribution function estimation in section 2.1. For any finite set of increments it will always be possible to use MCMC methods to make inference about the unknown process, see Appendix A. In several situations the approximative inference based on the MCMC simulations will be satisfying and accurate enough, and we are also able to produce estimated functions with approximative confidence intervals for both the spectral measure and covariance function. We can also predict the outcome at future time points and for large simulated data sets we can investigate the approximative the large-sample properties.

There is one problem with the setup above, if n is large we have to determine the inverse of a large matrix Σ in order to calculate the likelihood function, where all the elements (or $[n^2 - n]/2 + 1$ elements) in the matrix Σ are determined by a unique sum. In a normal simulation routine we will need to do this calculation several times (perhaps millions) and even with a fast computer this can take very long time even with a reasonable large set of observations. In Appendix A there is a explanation of the simulation routine and some tricks we have used in order to try to speed things up.

The complexity of multivariate Gaussian likelihood has in some sense become the crux of this thesis, the intricate structure of the inverse and determinant of the covariance matrix makes it hard to make exact inference or investigate asymptotic properties of the posterior process. For some simple time series models, such as autoregressive model, there do exist explicit expressions for the full log-likelihood, see **?**, even with this simple model the expressions for the inverse and determinant are still not easy to work with.

When we are studying the asymptotic or large-sample properties of an estimator we are only interested in the limit case, when the numbers of observation approaches infinity. In the next chapter we will se that there exists a approximation of the full log-likelihood that under fairly weak conditions of the power spectrum will become arbitrarily close to the full log-likelihood in

the limit. Let $\tilde{L}_n(F)$ be the approximation to the log-likelihood and assume that the true power spectrum is a parametric function with parameters $\theta_1, \ldots, \theta_p$, then under some mild conditions we will specify later, the following two results will be satisfied as $n \to \infty$

$$n^{-1/2}(L_n(F) - \tilde{L}_n(F)) \to 0,$$

where $L_n(F)$ is the full Gaussian log-likelihood given by (2.6), and

$$\frac{1}{n}E\left[\frac{\partial}{\partial\theta_k}L_n(F)\frac{\partial}{\partial\theta_l}L_n(F)\right] \to \Gamma(\theta)_{k,l} \text{ and } -\frac{1}{n}\frac{\partial^2}{\partial\theta_k\partial\theta_l}\tilde{L}_n(F) \to \Gamma(\theta)_{k,l}, \text{ where } k,l = 1,\ldots,p.$$

In other words the approximation and the full log-likelihood become arbitrary close as n gets large, and the second derivatives, needed to estimate the Fisher's information matrix, converge towards the same limit.

We will conclude this section with an example based on simulated data that will illustrate most of the concepts discussed here in this section.

EXAMPLE 2.6. Let $Y(0), \ldots, Y(24)$ be a sample from a stationary Gaussian time series with expectation zero and unknown covariance function, see Figure 2.1. Suppose we fore some reason are convinced that these data are actually are realization from a stationary Gaussian time series with covariance function

$$C_\pi(h) = \sigma_\pi^2 \rho^h, \text{ where } h = 0, 1, \ldots, 24.$$

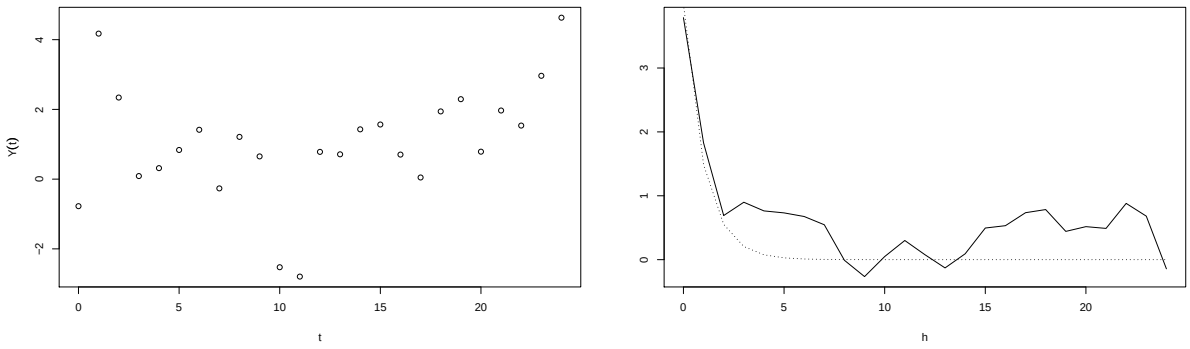

Figure 2.1: A simulated stationary Gaussian time series $Y(t)$, for $t = 0, \ldots 24$, with $\mu_0 = 0$ and true covariance function $C_0(h) = \sigma_0^2 \exp(-\phi_0|h|)$, where $\sigma_0 = 2$ and $\phi_0 = 0.99$ (left panel). The estimated covariance (solid line, right panel) for the time series $Y(t)$ (left panel) and the true covariance function (dotted line).

From Example 1.11 we know that this covariance function is valid and from the same example we have an exact expression of the power spectrum. Let $f_\pi(u)$ be the power spectrum from Example 1.11 that corresponds to $C_\pi(h) = \rho_\pi^{|h|}$, we will then apply the method described in the sections above to some. Our prior distribution is given by a Gamma process which is in turn determined by its increment distribution,

$$dG(u) \sim \text{Ga}(\alpha(u),\ \beta(u)),$$

where $u \in (0, \pi]$ and we will concentrate our prior around our a priori guess $\alpha(u) = b(u) f_\pi(u)\, du$ and $\beta(u) = b_0 + b_1 u$. We have chosen a simple function for $\beta(u)$ instead of a constant, there is at least two reasons for doing this. First note that the high end of the frequencies, i.e. when u is close to π, corresponds short and fast oscillations of the covariance function and a decreasing variance might smooth this effect, also a increasing $\beta(u)$ function might help to stabilize the simulations

In order to implement this construction it is necessary to do a numerical approximation, this is typically done by dividing the domain of the power spectrum, i.e. the interval $[0, \pi]$, into equidistant subintervals. Note that we will use the midpoints of the intervals in order to make the numerical approximation of the integral more accurate. Note that if the interval $[0, \pi]$ is divided into M subintervals of the length π/M, then the approximated covariance function become periodic with period $4M$ and also symmetric, see Figure 2.2. This is not very surprising since under the approximation, given a spectral measure F, we have that the covariance function is given by

$$C(h) = \sum_{m=1}^{M} \cos(\pi m h/M)\, [F(\pi m/M) - F(\pi(m-1)/M)] = \sum_{m=1}^{M} \cos(\pi m h)\, \Delta F(\pi m/M),$$

where $h = 0, 1, 2 \ldots$, which is obviously a periodic function.

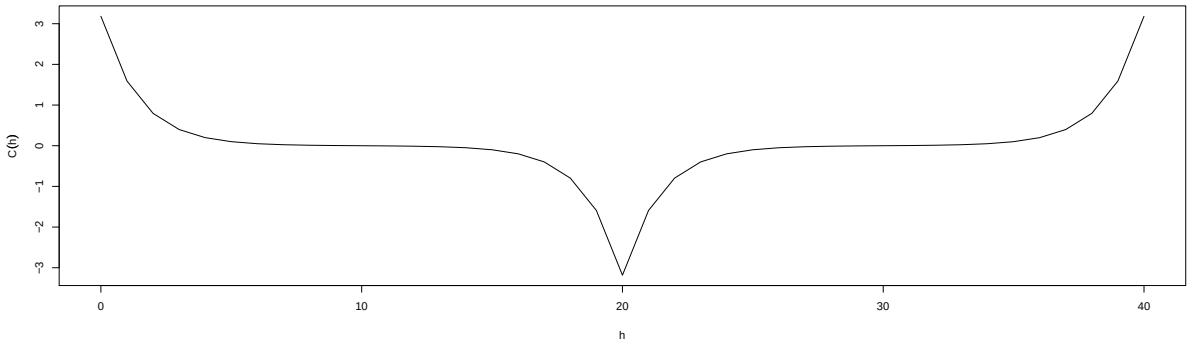

Figure 2.2: Plot of $C(h) = \sigma^2 \rho^h$, where $\rho = 0.5$, $\sigma = 1$, $M = 10$ and $h = 0, \ldots, 40$.

This means that it is important and necessary to be careful then we shall decide how many subintervals we will use. Because of the symmetry we should actually have more subintervals than the largest lag distance, unfortunately an increase in the number of subintervals will make simulation more unstable. In this example where $Y(t)$ has largest lag of $h = 24$ are we going to use 25 subintervals of the interval $[0, \pi]$, i.e $0 = u_0 < u_1 < \cdots < u_{24} < u_{25} = \pi$, where $u_i = \pi i/25$ for $i = 1, \ldots, 25$.

To be specific we will concentrate our a priori guess around $C_\pi(h) = \sigma_\pi^2 \exp(-\phi|h|)$, where $\sigma_\pi = 2$ and $\phi_\pi = 2$, and $\beta(u) = 7 + 7u$, se Figure 2.3 for the simulated exception of the random $C(h)$ and one single realization.

In Figure 2.4 is a plot of the expected posterior covariance function together with the the expected covariance function from the prior and the covariance estimated from data, it is clear from this figure that the a posterior covariance function is a weighted average.

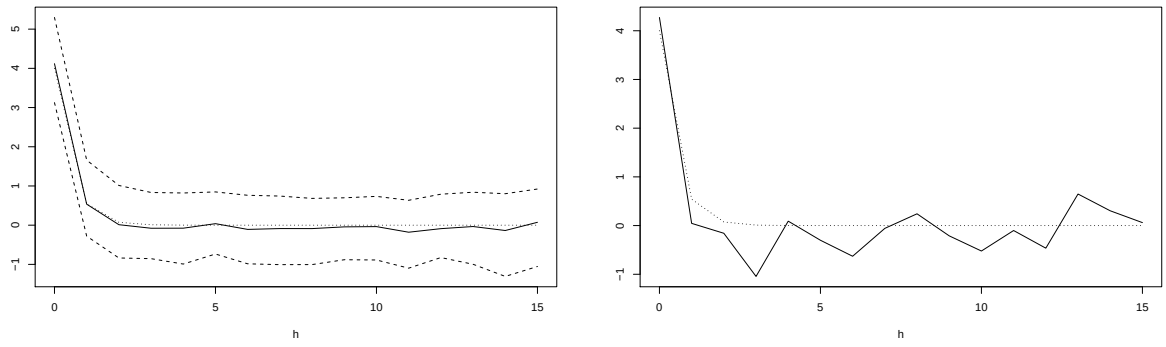

Figure 2.3: Estimated expected prior covariance function, i.e. $E[C(h)] = \sigma_\pi^2 \exp(-\phi_\pi h)$, where $\sigma_\pi = 2$ and $\phi_\pi = 2$, together with upper and lower 0.95-bound from the simulations (left panel). A single realization of a the random $C(h)$ (right panel). In both panels the dotted line is the true a priori expectation.

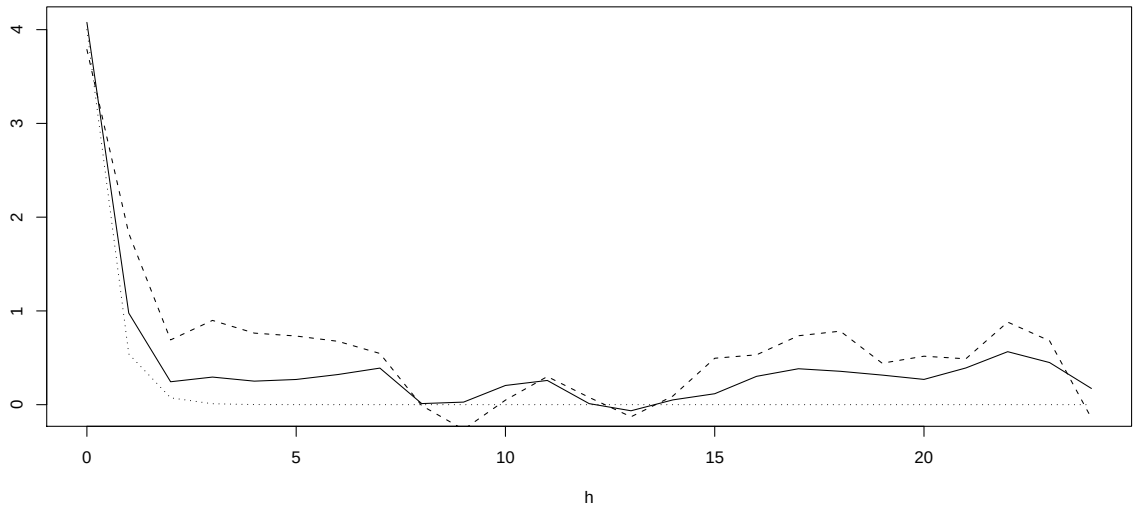

Figure 2.4: Estimated expected posterior covariance function, i.e. $E[C(h)|\text{data}]$ (solid line), estimated covariance from data (dashed line) and expected prior covariance function (dotted line).

We have already discussed that we are not complete satisfied with this construction, and the natural extension is to consider the a priori function $C(h) = \exp(-a\phi_\pi|h|)C(h)$, where $C(h)$ is a random covariance function centered around $C_\pi(h) = \sigma^2 \exp(-(1-a)\phi_\pi|h|)$, where ϕ_π and σ_π as above and $a = 0.05$. In Figure 2.5 is the estimated a priori expectation based on simulations and in Figure 2.6 is the estimated posterior covariance function. It is clear that this reduces the noise in the tail of the covariance function and force it to be close to zero for large h, but from Figure 2.6 it looks like we have "lost" some information from data by doing this. In Figure 2.7 are the estimated density for $Y(25)|\text{data}$ for both construction discussed in this example.

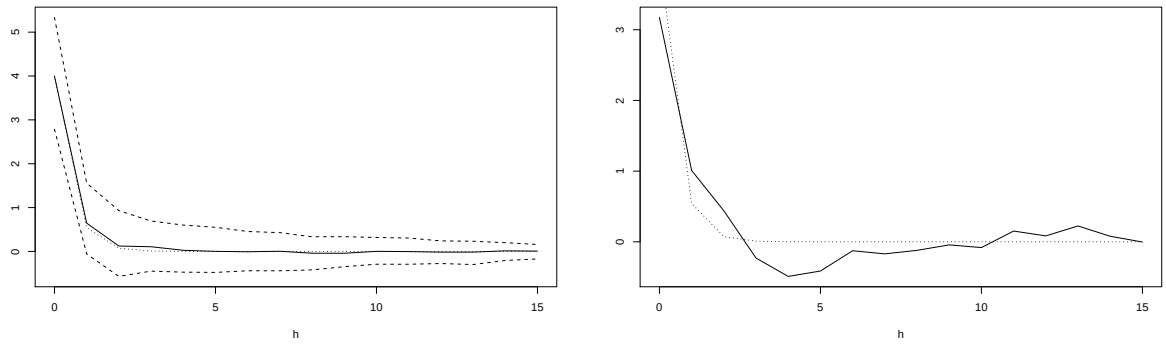

Figure 2.5: Estimated expected prior covariance function, i.e. $C^*(h) \times E[C(h)] = \exp(-a\phi_\pi h) \times \sigma_\pi^2 \exp(-(1-a)\phi_\pi h) = \sigma_\pi^2 \exp(-\phi_\pi h)$, where $\sigma_\pi = 2$, $\phi_\pi = 2$ and $a = 0.05$, together with upper and lower 0.95-bound from the simulations (left panel). A single realization of a random the $C(h)$ (right panel). In both panels the dotted line the is true a priori expectation.

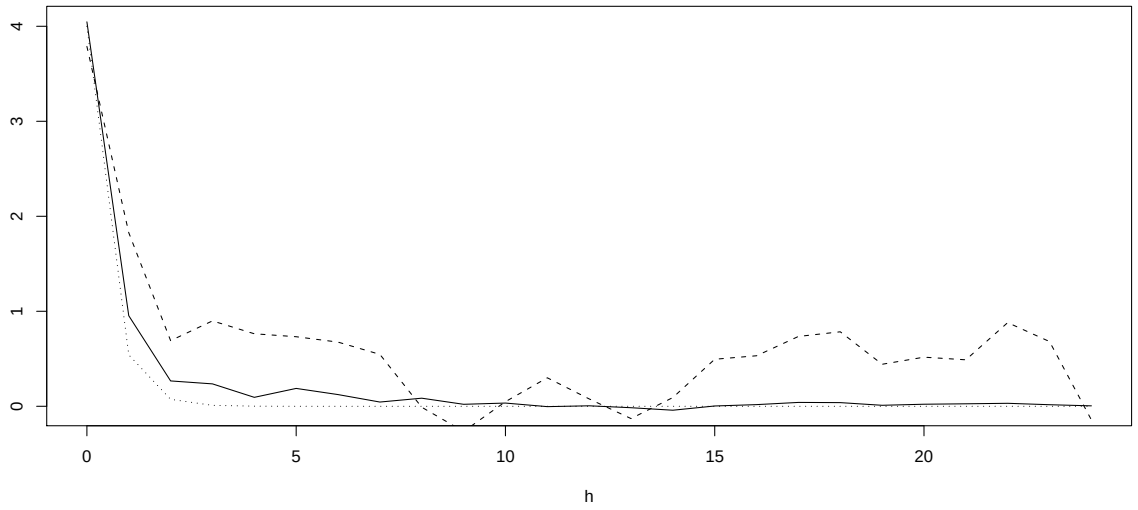

Figure 2.6: Estimated expected posterior covariance function, i.e. $E[C(h)|\text{data}]$, from the deterministic times random covariance function construction (solid line), estimated covariance from data (dashed line) and expected prior covariance function (dotted line).

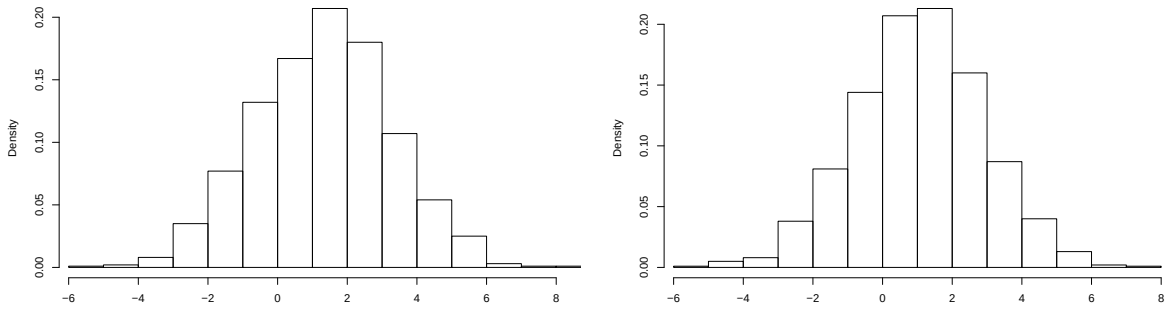

Figure 2.7: Estimated distribution for $Y(25)|$data for the random covariance function construction (left panel) and the deterministic times random covariance function construction

CHAPTER 3

Asymptotic properties

In the previous section we have already mentioned the need for an approximation to the exact multivariate Gaussian log-likelihood. This approximation should be fairly simple to work with and in the limit the approximation should become arbitrary close to the exact log-likelihood. In Section 1.1 we will introduce an asymptotic approximation to the full log-likelihood given in **?**. This approximation is known as the 'principal part' to the log-likelihood and satisfies both of the desired properties, it become arbitrary close for large n and it is sufficiently simple to work with. In Section 1.2 we will discuss a related and discrete version of the principal part which is also known as the Whittle approximation. We will also study some of the large-sample properties of the spectral measure after a sample is observed in this simple construction. We will continue the discussion of the large sample properties for more general spectral measures after a sequence of data is observed in Section 2 and derive the main properties for the posterior spectral measure and covariance function. Note that we sometimes will refer to the multivariate Gaussian likelihood (2.6) as the full or exact log-likelihood rather than multivariate Gaussian log-likelihood.

1. Approximations

1.1. The "principal part". In the book by **?** he suggest an approximation to the exact multivariate Gaussian log-likelihood for stationary Gaussian time series with expectation zero. The approximation is throughout the text referred to as the *'principal part'* of the log-likelihood and we will therefore also use this name. It is defined as a function of the power spectrum and is given by

$$\tilde{L}_n(F) = -\frac{n}{2}\left[\log(2\pi) + \frac{1}{2\pi}\int_{-\pi}^{\pi}\log(2\pi f(u))\,du + \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{I_n(u)}{f(u)}\,du\right]. \tag{1.1}$$

From equation (1.1) it is clear that the principal part of the log-likelihood will fit quite good to our nonparametric approach. It will make all the computations much easier and also speed up the numerical simulations since we do not need to invert any large matrices. The principal part is an approximation of the real log-likelihood, therefore before we start we have to establish how good the approximation is. Also note that in this section we are only interested in the limit situations, as the number of observations approaches infinity, it is therefore sufficient to check that the approximation is good enough in the situations where the number of observations is large. At the end of Section 2.3 we mentioned two properties a good approximation should satisfy. The approximation should at least become close to the real log-likelihood in the limit and both expressions for the observed information should converge towards the same limit.

The following two results can be found in the first two chapters of **?** and are exactly what we need to verify that the principal part is a suitable approximation. Theorem 1.1 first shows that

the difference between the approximation and the exact expression becomes small as the number of observations increases.

THEOREM 1.1. *Let $Y(t)$, where $t = 0, \pm 1, \pm 2, \ldots$, be a stationary Gaussian process with expectation zero, true covariance function $C_0(h)$, where $h = 0, \pm 1, \ldots$, and spectral density $f_0(u)$, where $u \in [-\pi, \pi]$. Assume that the process $Y(t)$ satisfies the following conditions*

i) $f_0(u) \geq m > 0$, *for* $-\pi < u < \pi$, *and*
ii) $\sum_{h=1}^{\infty} h|C(h)|^2 < \infty$,

then the "principal part" of the log-likelihood $\tilde{L}_n(F_0)$ (1.1) and the exact log-likelihood $L_n(F_0)$ (2.6) satisfies the following limit as $n \to \infty$

$$n^{-1/2}(L_n(F_0) - \tilde{L}_n(F_0)) \to 0.$$

PROOF. See Chapter 1 of **?** for a proof. Note that the assumption that $f(u) \geq m$ on the interval $[0, \pi]$, for a positive number m, is not necessary, in **?** it is shown that it is sufficient to require that $f(u)$ is positive on the same interval. □

The next result establishes exactly what we need in order to be able to show that the observed information matrices from the principal part and the full log-likelihood converges to the same limit.

THEOREM 1.2. *Let $Y(0), \ldots, Y(n-1)$ be a sample from a stationary Gaussian times series with expectation zero and power spectrum $f_0(u)$. Assume that the power spectrum is a smooth parametric function with parameters $\theta_1, \ldots, \theta_p$ where all second-order mixed partial derivatives exist, then as $n \to \infty$ we have that*

$$\frac{1}{n} I(\theta)_{k,l} = \frac{1}{n} E\left[\frac{\partial}{\partial \theta_k} L_n(F) \frac{\partial}{\partial \theta_l} L_n(F) \right] \to \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_k} \log(f_0(u)) \frac{\partial}{\partial \theta_l} \log(f_0(u))\, du = \Gamma_{k,l} \quad (1.2)$$

for every choice of θ_k and θ_l, where $k, l = 1, 2, \ldots, p$.

COROLLARY 1.3. *Let $Y(0), \ldots, Y(n-1)$ be a sample from a stationary Gaussian times series with expectation zero and power spectrum $f_0(u)$, where $f_0(u) \geq m > 0$ for $u \in [-\pi, \pi]$. Assume that the power spectrum is a smooth parametric function with parameters $\theta_1, \ldots, \theta_p$ where all the second-order mixed partial derivatives exist and is bounded, then as $n \to \infty$ we have that*

$$-\frac{1}{n}\left[\frac{\partial^2}{\partial \theta_k \partial \theta_l} \tilde{L}_n(F) \right] \xrightarrow{a.s} \Gamma_{k,l} \quad or \quad -\frac{1}{n}\left| I(\theta)_{k,l} - \frac{\partial^2}{\partial \theta_k \partial \theta_l} \tilde{L}_n(F) \right| \xrightarrow{a.s} 0. \quad (1.3)$$

for every choice of θ_k and θ_l, where $k, l = 1, 2, \ldots, p$ and $\Gamma_{k,l}$ is the limit (1.2).

PROOF. (Sketch) The first thing we need is an expression for the partial derivatives of $\tilde{L}_n(F)$,

$$\begin{aligned}
&\frac{\partial^2}{\partial \theta_k \partial \theta_l} \tilde{L}_n(F) \\
&\quad = -\frac{n}{2} \frac{\partial^2}{\partial \theta_k \partial \theta_l} \left[\log(2\pi) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(2\pi f_0(u))\, du + \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_n(u)}{f_0(u)}\, du \right] \\
&\quad = -\frac{n}{4\pi} \int_{-\pi}^{\pi} \left[\frac{f_0^{k,l}(u) f_0(u) - f_0^{k,l}(u) I_n(u)}{f_0(u)^2} + \frac{2 f_0^k(u) f_0^l(u) I_n(u) - f_0^k(u) f_0^l(u) f_0(u)}{f_0(u)^3} \right] du.
\end{aligned}$$

where $f_0^k(u)$ and $f_0^{k,l}(u)$ are the partial derivatives of $f_0(u)$ with respect to θ_k and/or θ_l. We will divide the problem into two parts and show that the first fraction approaches zero and that the second converges towards $\Gamma_{k,l}$. Since all the partial derivatives are bounded there exists a constant M so large that $f_\theta^k(u)$, $f_\theta^{k,l}(u) < M$ for $u \in [-\pi, \pi]$ and $l, k = 1, \ldots, p$, also from the conditions we have that $f_0(u) \geq m > 0$ for $u \in [-\pi, \pi]$. Then from Theorem 1.22 we do now have that

$$\left| \int_{-\pi}^{\pi} \frac{f_0^{k,l}(u) f_0(u) - f_0^{k,l}(u) I_n(u)}{f_0(u)^2} \, du \right| \leq \frac{M}{m^2} \left| \int_{-\pi}^{\pi} f_0(u) - I_n(u) \, du \right| \xrightarrow{a.s.} 0.$$

If we work out the expression $\Gamma_{k,l}$ given in (1.2), we find that

$$\begin{aligned}
&\left| \int_{-\pi}^{\pi} \frac{2 f_0^k(u) f_0^l(u) I_n(u) - f_0^k(u) f_0^l(u) f_0(u)}{f_\theta(u)^3} \, du - \Gamma_{k,l} \right| \\
&\qquad = \left| \int_{-\pi}^{\pi} \frac{2 f_0^k(u) f_0^l(u) I_n(u) - f_0^k(u) f_0^l(u) f_0(u)}{f_0(u)^3} \, du - \int_{-\pi}^{\pi} \frac{f_0^k(u) f_0^l(u)}{f_0(u)^2} \, du \right| \\
&\qquad\qquad \leq \frac{2M^2}{m^3} \left| \int_{-\pi}^{\pi} I_n(u) - f_0(u) \, du \right| \xrightarrow{a.s.} 0.
\end{aligned}$$

We have now shown that $-[\partial^2/(\partial\theta_k \partial\theta_l)\tilde{L}_n(F)]/n$ is a sum of two parts that converges almost surely towards zero and $\Gamma_{k,l}$, this completes the proof and we have shown that

$$-\frac{1}{n} \frac{\partial^2}{\partial\theta_k \partial\theta_l} \tilde{L}_n(F) \xrightarrow{a.s.} \Gamma_{k,l}, \quad \text{for every} \quad k, l = 1, \ldots, p.$$

□

From **?** we know that the two functions $f_0(u)$ and $I_n(u)$ share some of the same properties, especially they are nonnegative, symmetric, and they are both periodic with period 2π. This means essentially that if we know how $f_0(u)$ and $I_n(u)$ behave on interval $[0, \pi]$ we know everything we need to know about the two functions and we will therefore as a standard use this interval as the fundamental domain. From these properties it is now possible to rewrite the principal part of the log-likelihood (1.1)

$$\begin{aligned}
\tilde{L}_n(F) &= n \log(2\pi) - \frac{n}{2\pi} \left[\int_0^{\pi} \log(f(u)) \, du + \int_0^{\pi} \frac{I_n(u)}{f(u)} \, du \right] \\
&= n \log(2\pi) - \lim_{m \to \infty} \frac{n}{2\pi} \left[\sum_{i=1}^{m} \log(f(u_i)) \, \Delta_i + \sum_{i=1}^{m} \frac{I_n(u_i)}{f(u_i)} \, \Delta_i \right]
\end{aligned}$$

where $\Delta_i = u_i(m) - u_{i-1}(m)$ and $u_i \in [u_i(m), u_{i-1}(m)]$. The reason we use the Riemann definition of the integral is that this will become useful in the later sections. We can now further rewrite principal part and find a new expression for $\tilde{L}_n(F)$ with respect on $\Delta F(u_i)$

$$\begin{aligned}
\tilde{L}_n(F) &= n \log(2\pi) - \lim_{m \to \infty} \frac{n}{2\pi} \left[\sum_{i=1}^{m} \log(f(u_i)\Delta_i/\Delta_i) \, \Delta_i + \sum_{i=1}^{m} \frac{I_n(u_i)\Delta_i}{f(u_i)\Delta_i} \, \Delta_i \right] \\
&= n \log(2\pi) - \lim_{m \to \infty} \frac{n}{2\pi} \left[\sum_{i=1}^{m} \log(f(u_i)\Delta_i) \, \Delta_i - m \log(\Delta_i) \, \Delta_i + \sum_{i=1}^{m} \frac{I_n(u_i)\Delta_i}{f(u_i)\Delta_i} \, \Delta_i \right] \\
&= \lim_{m \to \infty} -\frac{n}{2\pi} \left[\sum_{i=1}^{m} \log(\Delta F(u_i)) \, \Delta_i + \sum_{i=1}^{m} \frac{I_n^*(u_i)}{\Delta F(u_i)} \, \Delta_i \right] + c^*,
\end{aligned}$$

where c^* is a constant, $I_n^*(u_i) = I_n(u_i)\,\Delta_i$. Finally we will define

$$\tilde{L}_n^*(F) = \frac{n}{2\pi}\int_0^\pi \left[\log(dF(u_i)) + \frac{I_n^*(u_i)}{dF(u_i)}\right] du \equiv \lim_{m\to\infty}\frac{n}{2\pi}\sum_{i=1}^m \left[\log(\Delta F(u_i)) + \frac{I_n^*(u_i)}{\Delta F(u_i)}\right]\Delta_i. \quad (1.4)$$

The expression for $L_n^*(F)$ is constructed to fit our nonparametric Bayesian approach and its meaning will become clear in the next sections. We will also introduce a *likelihood element* of $\tilde{L}_n^*(F)$ that will be denoted by $d\tilde{L}_n^*(u)$ and is defined such that

$$\tilde{L}_n^*(F) = \int_0^\pi d\tilde{L}_n^*(v)\,dv = \lim_{m\to\infty}\sum_{i=1}^m d\tilde{L}_n^*(u_i),$$

where u_i is as defined above.

REMARK 1.4. Let $Y(t)$, where $t = 0, \pm1, \pm2, \ldots$, be a stationary time series that satisfies the conditions of Theorem 1.1 and assume that the true power spectrum $f_0(u)$ is constant on given subintervals of the interval $[0, \pi]$, i.e. $f_0(u) = f_0(u_i)$, for $u \in [u_i, u_{i-1}]$ and all $i = 1, 2, \ldots, M$, where $0 = u_0 < u_1 < \cdots < u_{M-1} < u_M = \pi$. Define $\Delta_i = u_i - u_{i-1}$ and $\Delta F_0(u_i) = F_0(u_i) - F_0(u_{i-1}) = f_0(u_i)\,\Delta_i$, then for a sample of size n from $Y(t)$ it is possible to rewrite the principal part of the log-likelihood as

$$\begin{aligned}
\tilde{L}_n(F_0) &= -\frac{n}{2}\left[\log(2\pi) + \frac{1}{2\pi}\int_{-\pi}^{\pi}\log(2\pi f_0(u))\,du + \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{I_n(u)}{f_0(u)}\,du\right] \\
&= -\frac{n}{2}\left[\log(2\pi) + \frac{1}{\pi}\sum_{i=1}^M \log(2\pi\Delta F(u_i)/\Delta_i)\,\Delta_i + \frac{1}{\pi}\sum_{i=1}^M \frac{\Delta_i}{\Delta F(u_i)}\int_{u_{i-1}}^{u_i} I_n(v)\,dv\right] \\
&= -\frac{n}{2\pi}\sum_{i=1}^M\left[\log(\Delta F_0(u_i))\,\Delta_i + \frac{\Delta_i}{\Delta F_0(u_i)}\int_{u_{i-1}}^{u_i} I_n(v)\,dv\right] + c^*.
\end{aligned}$$

where c^* is a constant.

Before we continue the discussion of the principal part of the log-likelihood and derive some asymptotic properties for the posterior spectral measure and covaraince function, we will discuss the discrete version of the approximation.

1.2. The Whittle approximation. In this section will we introduce a discrete approximation of the multivariate Gaussian log-likelihood. This discrete approximation was first suggested by Whittle in the early fifties and is therefor often referred to as the *Whittle approximation*. The easiest way to obtain the Whittle approximation is to derive it from the discrete version of the already established principal part approximation. We can write expression (1.1) as

$$\begin{aligned}
\tilde{L}_n(F) &= \lim_{m\to\infty} -\frac{n}{2}\left[\log(2\pi) + \log(2\pi) + \frac{1}{\pi}\sum_{i=1}^m \log(f(\pi i/m))\,\frac{\pi}{m} + \frac{1}{\pi}\sum_{i=1}^m \frac{I_n(\pi i/m)}{f(\pi i/m)}\,\frac{\pi}{m}\right] \\
&= \lim_{m\to\infty} -\left[2n\log(2\pi) + \frac{n}{2m}\sum_{i=1}^m \log(f(u_i)) + \frac{n}{2m}\sum_{i=1}^m \frac{I_n(u_i)}{f(u_i)}\right].
\end{aligned} \quad (1.5)$$

where $u_i = \pi i/m$. The Whittle approximation is now obtained from equation (1.5) if we replace m with n, the number of observation, we denoted the approximation by $L_W(F)$ and it is defined

as the expression

$$L_W(F) = -n\log(2\pi) - \frac{1}{2}\Big[\sum_{i=1}^{n}\log(f(u_i)) + \sum_{i=1}^{n}\frac{I_n(u_i)}{f(u_i)}\Big] \tag{1.6}$$

where $u_i = \pi i/n$, for $i = 1, \ldots, n$. The next lemma establishes that the Whittle approximation is also close enough to the full multivariate Gaussian log-likelihood for a stationary Gaussian time series.

LEMMA 1.5. *Under the same conditions as in Theorem 1.1 the Whittle approximation (1.6) satisfies*

$$n^{-1/2}|L_W(F) - L_n(F)| \to 0 \quad \textit{as } n \to \infty,$$

where $L_n(F)$ is the full multivariate Gaussian log-likelihood.

PROOF. (Sketch) Observe that it is possible to write

$$n^{-1/2}|L_W(F) - L_n(F)| = n^{-1/2}|L_W(F) - \tilde{L}_n(F) + \tilde{L}_n(F) - L_n(F)| \leq$$
$$n^{-1/2}|L_W(F) - \tilde{L}_n(F)| + n^{-1/2}|\tilde{L}_n(F) - L_n(F)|$$

from Theorem 1.1 we know that $n^{-1/2}|\tilde{L}_n(F) - L_n(F)| \to 0$ as $n \to \infty$, therefore the remaining part is to show that $n^{-1/2}|L_W(F) - \tilde{L}_n(F)|$ approaches zero as $n \to \infty$. From the definitions (1.1) and (1.6) we find that to show that $n^{-1/2}|L_W(F) - \tilde{L}_n(F)| \to 0$ it is equivalent to proving that

$$n^{-1/2}\bigg|\sum_{i=1}^{n}\bigg[\log(f_0(u_i)) + \frac{I_n(u_i)}{f_0(u_i)}\bigg]\Delta - \int_0^{\pi}\bigg[\log(f_0(u)) + \frac{I_n(u)}{f_0(u)}\bigg]\,du\bigg| \to 0,$$

where $u_i = \pi i/n$ and $\Delta = \pi/n$, as $n \to \infty$. Now since $f_0(u)$ is integrable $\log(f_0(u))$ must also be integrable and therefor there exist an integer N_1 such that for $n \geq N_1$ we have that

$$n^{-1/2}|L_W(F) - L_n(F)| \leq n^{-1/2}\bigg(\delta + m^{-1}\bigg|\sum_{i=1}^{n} I_n(u_i)\Delta - \int_0^{\pi} I_n(u)\,du\bigg|\bigg).$$

From Theorem 1.22 and 1.20 we have that

$$\sum_{i=1}^{n} I_n(u_i)\Delta \xrightarrow{P} \int_0^{\pi} f_0(u)\,du \quad \text{and} \quad \int_0^{\pi} I_n(u)\,du \xrightarrow{a.s} \int_0^{\pi} f_0(u)\,du \tag{1.7}$$

as $n \to \infty$. There exist now N_2 such that for $n \geq N_2$ both convergences in (1.7) is satisfied and N_3 such that for $n \geq N_3$

$$m^{-1}\bigg|\sum_{i=1}^{n} f_0(u_i)\Delta - \int_0^{\pi} f_0(u)\,du\bigg| \leq \delta'$$

and for $n \geq N$ where $N = \max(N_1, N_2, N_3)$ we have now that

$$|L_W(F) - L_n(F)| \to (\delta + \delta') < \infty \Rightarrow n^{-1/2}|L_W(F) - L_n(F)| \to 0.$$

which completes the proof. □

In order to make the Whittle approximation more suitable for a Bayesian nonparametric approach we are going to rewrite expression (1.6). Let $\Delta F(u_i) = F(u_i) - F(u_{i-1}) = f(u_i)\,\Delta_i$ where $\Delta_i = u_i - u_{i-1} = \pi/n$, then the new version of $L_W(F)$ is given by

$$L_W(F) = -n\log(2\sqrt{n\pi}) - \frac{1}{2}\Big[\sum_{i=1}^{n}\log(\Delta F(u_i)) + \sum_{i=1}^{n}\frac{I_n^*(u_i)}{\Delta F(u_i)}\Big] \tag{1.8}$$

where $I_n^*(u_i) = I_n(u_i)\,\Delta_i$. The next example illustrates a somehow natural approach to how we can define a prior distribution for the unknown spectral measure in this discrete approach.

EXAMPLE 1.6. Suppose the time series $Y(t)$, where $t = 0, \pm 1, \ldots$, satisfy the assumption of Lemma 1.5, then the Whittle Approximation given by (1.8) is a satisfying approximation to the full likelihood (2.6). Let $v_i = \Delta F(u_i) = F(u_i) - F(u_{i-1})$ for $i = 1, \ldots, n$ and where $\Delta = u_i - u_{i-1} = \pi/n$, for the finite vector $\boldsymbol{v} = (v_1, \ldots, v_n)$ let $\pi(\boldsymbol{v}) = \pi(v_1)\cdots\pi(v_n)$ be a prior density for $\boldsymbol{v}$ where $\pi(v_i) = \text{Inv-Gamma}(\alpha(u_i) + c, \beta(u_i))$, where c is a number chosen such that the desired order of moments exist, see Appendix B. The posterior distribution is then given in the usual way as

$$\begin{aligned}\pi(\boldsymbol{v}|\text{data}) &\propto \pi(\boldsymbol{v}) \times L_W(F)\\ &\propto \prod_{i=0}^{n-1}\Big\{v_i^{-[\alpha(u_i)+c+1/2]-1}\exp\Big[-\frac{I_n^*(u_i)/2 + \beta(u_i)}{v_i}\Big]\Big\}.\end{aligned} \tag{1.9}$$

From (1.9) it is easy to verify that $\boldsymbol{v}|$data is a product of Inverse-Gamma densities which means that the elements in the vector $\boldsymbol{v}$ are asymptotically independent after the data are observed. The updated parameters for $v_i|$data are $\alpha'(u_i) = \alpha(u_i) + c + 1/2$ and $\beta(u_i) = I_n(u_i)\,\Delta/2 + \beta(u_i)$. The expectation and variance of the posterior density for a single $v_i|$data are now found from the properties of the Inverse-Gamma distribution and are given by

$$E[v_i] = \frac{I_n^*(u_i)}{2\alpha(u_i) + 2c - 1} + \frac{2\beta(u_i)}{2\alpha(u_i) + 2c - 1}$$

and

$$\text{Var}(v_i) = \frac{2(I_n^*(u_i) + 2\beta(u_i))^2}{(2\alpha(u_i) + 2c - 1)^2(2\alpha(u_i) + 2c - 3)} = \frac{2[(I_n^*(u_i))^2 + 4I_n^*(u_i)\beta(u_i) + \beta(u_i)^2]}{(2\alpha(u_i) + 2c - 1)^2(2\alpha(u_i) + 2c - 3)}.$$

Assume we have chosen $\alpha(u_i) = \Delta$ and $\beta(u_i) = f_\pi(u_i)\Delta$, where $f_\pi(u)$ is the power spectrum that corresponds to our a priori beliefs about the covariance function for the time series $Y(t)$. Motivated from the asymptotic independency of the parameters and from the definition of Riemann sum, we have that for the estimator $\hat{F}$ the expectation and variance are given by

$$\begin{aligned}E[\hat{F}(u)|\text{data}] = E\big[\sum_{\pi i/n<u} v_j\big] &= \frac{1}{2\alpha(u_i) + 2c - 1}\sum_{\pi i/n<u} I_n(u_i)\,du + \frac{2}{2\alpha(u_i) + 2c - 1}\sum_{\pi i/n<u}\beta(u_i)\\ &\to \frac{1}{2c-1}F_0(u) + \frac{2}{2c-1}F_\pi(u),\end{aligned}$$

and

$$\begin{aligned} n\operatorname{Var}(\hat{F}(u)|\text{data}) &= n\operatorname{Var}\Big(\sum_{\pi i/n<u} v_i\Big) \\ &= \frac{2}{(\alpha(u_i)+3)^2(\alpha(u_i)+1)} \\ &\quad\times\Big[\pi\sum_{\pi i/n<u} I_n(u_i)^2\,du + 2\pi\sum_{\pi i/n<u} I_n(u_i)\beta(u_i) + n\sum_{\pi j/n<u}\beta(u_i)^2\Big] \\ &\to \frac{2\pi}{(2c-1)^2(2c-3)} \\ &\quad\times\Big[\int_0^u f_0(v)^2\,dv + 2\int_0^u f_0(v)f_\pi(v)\,dv + \int_0^u f_\pi(v)^2\,dv\Big]. \end{aligned}$$

A reasonable choice for c might be $c = 2$, as this will make sure that the prior density for v_i does have existing expectation and variance.

We will now derive an equivalent expression for the Whittle approximation as we did for the principal part of the log-likelihood in Remark 1.4.

REMARK 1.7. Let $Y(t)$, where $t = 0, \pm1, \pm2, \ldots$, be a stationary time series that satisfies the conditions of Theorem 1.1 and assume that the true power spectrum $f(u_i)$, where $u_i = \pi i/n$ for $i = 0, \ldots, n$, is constant on equidistant subintervals of length π/M of the interval $[0, \pi]$, where $M \in \mathbb{N}$ and $M < n$. Then there exist integers $m_1, \ldots, m_M$, m and index sets $U_1, \ldots, U_M$ such that $\sum_j m_j = n$ and for every $j = 1, \ldots, M$ we have that $m_j \geq m > 0$ and for $i \in U_j$ we have that $u_{j-1} < u_i < u_j$ and $f(u_i) = f(u_j)$. Define $\Delta_j = u_j - u_{j-1}$ and $\Delta F(u_j) = F(u_j) - F(u_{j-1}) = f(u_j)\Delta_i$ then it is possible to rewrite the Whittle approximation given by (1.8) as

$$\begin{aligned} L_W(F) &= -n\log(2\sqrt{n\pi}) - \frac{1}{2}\Big[\sum_{j=1}^M m_j\log(\Delta F(u_j)) + \sum_{j=1}^M \frac{\sum_{i\in U_j} I_n^*(u_j)}{\Delta F(u_j)}\Big] \\ &= -\frac{1}{2}\sum_{j=1}^M\Big[m_j\log(\Delta F(u_j)) + \frac{1}{\Delta F(u_j)}\sum_{i\in U_j} I_n^*(u_i)\Big] + c^* = L_W^*. \end{aligned} \tag{1.10}$$

where $c^*$is a constant, $I_n^*(u_i) = I_n(u_i)\,\Delta_j$. Note that we might refer to expression (1.10) as the *modified Whittle approximation* and we will also sometimes write it as $L_W^*(F) = \sum_j \Delta L_W^*(u_j)$.

Note that $\Delta L_W^*(u_j)$ from Remark 1.7 has the same shape as the Inverse-Gamma density, it is therefor tempting to try to use a product of Inverse-Gamma densities as a prior distribution on F since this will become the conjugate prior for the modified Whittle approximation. This idea is in some sense related to the work of **?**, he uses a different starting point but his conclusions are similar to those we derive here. Note that since the independent increment process defied by the Inverse-Gamma distribution does not exist, see Appendix B, it is impossible to generalize this idea to the limit situation. In the following example we will show how the Inverse-Gamma distribution will work as the a priori distribution for a finite product set of variables.

EXAMPLE 1.8. Suppose the time series $Y(t)$, where $t = 0, \pm1, \ldots$, satisfies the assumption of Lemma 1.5 and that the true spectral measure, $F_0(u)$, is a step function, then the modified Whittle approximation given by (1.10) is a satisfying approximation to the full likelihood. Given

a sample $Y(0), \ldots, Y(n-1)$ of size n, let $M < n$ be an integer that is not too large and such that $m_i > m > 0$ for all $i = 1, \ldots, M$. Define $\Delta = u_i - u_{i-1} = \pi/M$, $v_i = \Delta F(u_i) = F(u_i) - F(u_{i-1})$ and assume that $\pi(\boldsymbol{v}) = \pi(v_1) \cdots \pi(v_M)$ is a product of Inverse-Gamma densities with respective shape and scale parameters $\alpha(u_i) + c$ and $\beta(u_i)$ where $i = 1, \ldots, M$. From equation (1.10) we see that the posterior distribution $\pi(\boldsymbol{v}|\text{data})$ is proportional to

$$\pi(\boldsymbol{v}|\text{data}) \propto \prod_{i=1}^{M} v_i^{-[m_i/2+\alpha(u_i)+c]-1} \exp\left[-\sum_{i=1}^{M} \frac{\frac{1}{2}\sum_{u_j \in U_i} I_n^*(u_j) + \beta(u_i)}{v_i} \right]$$

which is proportional to a product of Inverse-Gamma densities that suggest that the parameters, $(v_1, \ldots, v_M)$, are asymptotically independency after the data are observed. The a posterior moments are now easily found from the properties of the Inverse-Gamma distribution and Theorem 1.19. For $i = 1, \ldots, M$ the expectation of $v_i|\text{data}$ is

$$\begin{aligned} E[v_i|\text{data}] &= \frac{\sum_{u_j \in U_i} I_n^*(u_i) + 2\beta(u_i)}{m_i + 2\alpha(u_i) + 2c - 2} = \frac{m_i \, \Delta \hat{F}_{m_i}(u_i)}{m_i + 2\alpha(u_i) + 2c - 2} + \frac{2\beta(u_i)}{m_i + 2 \, \alpha(u_i) + 2c - 2} \\ &\to \Delta F_0(u_i), \end{aligned}$$

where $\Delta \hat{F}_{m_i}(u_i) = \hat{f}_{m_i}(u_i) \, \Delta$, as $n \to \infty$, since $n \to \infty$ implies that $m_i \to \infty$ for all $i = 1, \ldots, M$. The variance is further given by the expression

$$\begin{aligned} n \operatorname{Var}(v_i|\text{data}) &= \frac{2[m_i \, \Delta \hat{F}_{m_i}(u_i) + 2\beta(u_i)]^2}{(m_i + 2\alpha(u_i) + 2c - 2)^2 (m_i + 2\alpha(u_i) - 2c - 4)} \\ &= k(m_i) \left[\frac{2\pi n}{m_i M} \frac{\Delta \hat{F}_{m_i}(u_i)^2}{du} + \frac{8\pi n}{m_i^2 M} \frac{\Delta \hat{F}_m(u_i)\beta(u_i)}{du} + \frac{8\pi n}{m_i^3 M} \frac{\beta(u_i)^2}{du} \right] \end{aligned}$$

since $k(m_i) = 1/[(1 + 2\alpha(u_i)/m_i + (2c-2)/m_i)^2 (1 + 2\alpha(u_i)/m_i + (2c-4)/m_i)] \to 1$ as $n \to \infty$ we find that

$$n \operatorname{Var}(v_i|\text{data}) \to \frac{2\pi \Delta F_0(u_i)^2}{\Delta}$$

and in the case where $F_0(u)$ is differentiable we have that $n \operatorname{Var}(v_i) \to 2\pi f_0(u_i) \, \Delta$.

In Example 1.8 we saw that as the amount of observed data increases the posterior parameters approach the estimates from Theorem 1.19. This is in general a desirable property for a Bayesian estimator, that the prior information should become negligible as the amount of observations become large. This means that no matter which prior density we choose all solutions should become equal in the limit. The next lemma proves that this is exactly the case for spectral measure and the modified Whittle approximation.

In order to prove the next lemma we need a result regarding the remainder of Taylor expansions from **?**. We will repeat the general definition of the Taylor expansion.

Let $f(x)$ be a smooth function of x that is infinitely differentiable in a neighborhood of a number a. Then the following infinite sum is known as the *Taylor expansion* of $f(x)$ about a

$$f(x) = f(a) + \frac{1}{1!}\frac{d}{dx} f(x)(x-a) + \frac{1}{2!}\frac{d^2}{dx^2} f(x)(x-a)^2 + \cdots + \frac{1}{k!}\frac{d^k}{dx^k} f(x)(x-a)^k + R_k(x)$$

where $R_k(x)$ is the remainder and $R_k(x)$ satisfy

$$R_k(x) = \frac{1}{(k+1)!}\frac{d^{k+1}}{dx^{k+1}} f(\zeta)(x-a)^{k+1}, \quad \text{where } |\zeta - a| \leq |x - a|. \tag{1.11}$$

In order to prove the result we will show that the Taylor expansion of the log-posterior density for a single $\Delta F(u_j)$ converges to the log-density of a Gaussian distributed random variable as n becomes large. We will also have to use property (1.11) for the remainder in order to complete the proof. The technique suggested here is a well known method and is described in detail in several textbooks in statistics.

LEMMA 1.9. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a process with true power spectrum $f_0(u)$ which satisfies the conditions of Lemma 1.5 and is constant on the subintervals of $[0, \pi]$ such that the assumptions of Remark 1.7 is satisfied. Given a sample $Y(0), \ldots, Y(n-1)$ of size n from $Y(t)$ let $\pi_j(\Delta F(u_j))$ be any prior density for the unknown quantity $\Delta F(u_j)$, such that $\pi_j(\Delta F(u_j))$ is bounded an has bounded derivative in a neighborhood of $\Delta \hat{F}_{m_j}(u_j)$. Then $\Delta F(u_k)\big|\text{data}$ and $\Delta F(u_l)\big|\text{data}$ are asymptotically independent, for $k, l = 1, \ldots, M$ and $k \neq l$, also $\Delta F(u_j)\big|\text{data}$ converges in distribution to a Gaussian as $n \to \infty$, i.e.*

$$\sqrt{n}[\Delta F(u_j) - \Delta \hat{F}_{m_j}(u_j)]\big|\text{data} \xrightarrow{d} N(0, 2\pi f_0(u_j)^2 \, \Delta_j), \ a.s.$$

where $\Delta \hat{F}_{m_j}(u_j) = 1/m_j \sum_{i \in U_j} I_n(u_i) \, \Delta_i$ and $\Delta \hat{F}_{m_j}(u_j) \xrightarrow{P} f_0(u_j) \, \Delta_j$.

PROOF. Let $v_j = \Delta F(u_j)$, $\hat{v}_j = \Delta \hat{F}_{m_j}(u_j)$ and $w_j = \sqrt{n}(v_j - \hat{v}_j)$, where $j = 1, \ldots, M$, the prior density of the scaled and centered variable w_j is proportional to the density $\pi_j(w_j/\sqrt{n} + \hat{v}_j)$ and the log-posterior density is a constant away from

$$\begin{aligned}
\log(\pi_{\boldsymbol{w}}(w_1, \ldots, w_M\big|\text{data})) &= \log(\pi(w'_1, \ldots, w'_M\big|\text{data})) + c^* \\
&= \sum_{j=1}^{M} \log(\pi_j(w'_j)) + \log(\text{Lik}(w'_1, \ldots, w'_M\big|\text{data})) + c^* \\
&\approx \sum_{j=1}^{M} \left\{ \log(\pi_j(w'_j)) - \frac{1}{2}\left[m_j \log(w'_j) + \frac{1}{w'_j} \sum_{i \in U_j} I^*_n(u_i) \right] \right\}.
\end{aligned}$$

where c^* is a constant, $w'_j = w_j/\sqrt{n} + \hat{v}_j$, for $j = 1, \ldots, M$. From the structure of the log-posterior density it is clear that the the unknown variables $w_1, \ldots, w_M$ will become asymptotically independent after the data are observed. It is therefore sufficient, in order to prove the lemma, to show that the result holds for an arbitrary w_j, where $j = 1, \ldots, m$. Since we are able to split the log-posterior density into log-prior and log-likelihood, the Taylor expansion of the log-posterior density about zero is

$$\begin{aligned}
\log(\pi_{w_j}(w_j|\text{data})) &= \log(\pi_j(w_j/\sqrt{n} + \hat{v}_j)) + c^* \\
&= \log(\pi_j(\hat{v}_j|\text{data})) + \sum_{k=1}^{\infty} w_j^k \frac{1}{k!} \frac{d^k}{dw_j^k} \log(\Delta L^*_W(w'_j))\big|_{w_j=0} + R_0^\pi(w_i) + c^*
\end{aligned}$$

where c^* is a constant, $\Delta L^*_W(u)$ is defined in Remark 1.7 and $R_0^\pi(w_j)$ is the reminder of the log-prior part of the Taylor expansion. From property (1.11) we know that the exists a number ξ where $|\xi| < |w_j|$ such that the following is satisfied

$$\begin{aligned}
R_0^\pi(w_j) = w_j \left[\frac{d}{dw_j} \log(\pi_{w_j}(w_j))\big|_{w_j=\xi} \right] &= w_j \left[n^{-1/2} \frac{d}{dw_j} \log(\pi_j(w_j))\big|_{w_j=\xi/\sqrt{n}+\hat{v}_j} \right] \\
&= \frac{w_i}{n^{1/2}\pi(\xi/\sqrt{n} + \hat{v}_j)} \frac{d}{dw_j} \pi(w_j)\big|_{w_j=\xi/\sqrt{n}+\hat{v}_j}.
\end{aligned} \tag{1.12}$$

We are also able to obtain a general expression of the derivatives of the log-likelihood, then

$$\begin{aligned}\frac{d^k}{dw_j^k}\log(\Delta L_W^*(w_j'))\big|_{w_j=0} &= n^{-k/2}\frac{d^k}{dw_j^k}\log(\Delta L_W^*(w_j))\big|_{w_j=\hat{v}_j}\\ &= \frac{1}{2n^{k/2}}\left[\frac{(-1)^{k-1}(k-1)!m_j}{\hat{v}_j^k} - \frac{(-1)^k k!\sum_{i\in U_j} I_n(u_i)\,\Delta_j}{\hat{v}_j^{k+1}}\right]\\ &= \frac{(k-1)!}{2n^{k/2}}\left[\frac{(-1)^k(k-1)m_i}{(1/m_j\sum_{i\in U_j} I_n(u_i)\,\Delta_j)^k}\right]\\ &= \frac{(-1)^{k-1}(k-1)!(k-1)m_j}{2n^{k/2}}\left[\left(\frac{1}{m_j}\sum_{i\in U_j} I_n(u_i)\,\Delta_j\right)^k\right]^{-1},\end{aligned}$$

for $k = 1, 2, \ldots$. From this expression it is clear that the derivative of the log-likelihood becomes zero when evaluated in $w_j = 0$. Since we known that $\Delta_j = \pi/M$ we can now write the Taylor expansion of the log-posterior density as

$$\begin{aligned}\log(\pi_{w_j}(w_j|\text{data})) &= \log(\pi_j(\hat{v}_j|\text{data}))\\ &\quad - \frac{1}{2}w_j^2\left(\frac{m_j M}{n}\left[2\pi\left(\frac{1}{m_j}\sum_{i\in U_j} I_n(u_i)\right)^2\Delta_j\right]^{-1}\right) + R_3^{\text{lik}}(w_i) + R_0^{\pi}(w_i) + c_i^*\end{aligned}$$

where c_i^* is a constant and $R_3^{\text{lik}}(w_i)$ is the reminder of log-likelihood part of the Taylor expansion. The first term in the Taylor expansion is a constant and in order to prove the result it is sufficient to show that both $R_3^{\text{lik}}(w_i)$ and $R_0^{\pi}(w_i)$ become arbitrarily small for large n. From the assumption that the prior is bounded and has bounded derivative it is clear that (1.12) approaches zero as $n \to \infty$ as long as w_j is bounded for all $j = 1, \ldots, M$. From the derivatives of the log-likelihood and from property (1.11) we know that there exist a number ξ', where $|\xi'| < |w_i|$, such that

$$\begin{aligned}n^{k/2-1}R_k^{\text{lik}}(w_i) &= w_i^k\frac{m_j}{2nk!}\left[\frac{(-1)^{k-1}(k-1)!}{(\xi'/\sqrt{n}-\hat{v}_j)^k} - \frac{(-1)^k k!\sum_{i\in U_j} I_n(u_i)\,\Delta_j}{m_j(\xi'/\sqrt{n}-\hat{v}_j)^{k+1}}\right]\\ &\to w_i^k\frac{1}{2k!}\left[\frac{(-1)^{k-1}(k-1)!}{(f_0(u_j)\,\Delta_j)^k} - \frac{(-1)^k k!}{(f_0(u_j)\,\Delta_j)^k}\right]\\ &= w_j^k\frac{(-1)^{k-1}(k-1)M^k}{2\pi^k k f_0(u_j)^k} \le w_j^k\frac{(-1)^{k-1}(k-1)M^k}{2\pi^k k m^k} < \infty\end{aligned}$$

for $k = 2, 3, \ldots$ as long as w_j is bounded, since $\hat{v}_j \xrightarrow{P} f_0(u_j)$ and $\Delta_j = \pi/M$, also from the conditions of Lemma 1.5 we know that $f_0(u) \ge m > 0$, for $u \in [0, \pi]$. Especially this means that $R_3^{\text{lik}}(w_i) \to 0$ as $n \to \infty$ if w_j is bounded and all that remains now is to show that there exist a constant c such that $\Pr\{|w_i| < c\} = 1 - \epsilon$ as $n \to \infty$.

Under the assumption that the modified Whittle approximation is good enough, we have that the posterior density for w_j is proportional to

$$\begin{aligned}\pi_{w_j}(w_j|\text{data}) &\propto \pi_j(w_j') \times \left[{w'}_j^{-m_j/2}\exp\left(-\frac{1}{2w_j'}\sum_{i\in U_j} I_n^*(u_i)\right)\right]\\ &= \pi_j(w_j/\sqrt{n}+\hat{v}_j) \times \left[(w_j/\sqrt{n}+\hat{v}_j)^{-m_j/2}\exp\left(-\frac{m_j\hat{v}_j}{2(w_j/\sqrt{n}+\hat{v}_j)}\right)\right]\end{aligned}$$

The first term will become almost constant for large n and m_j so all the "action" be in the last term. Let M_n be the greatest integer such that $(w_j/\sqrt{n} + \hat{v}_j) > 0$ for $w_i \in [-M_n, M_n]$, then $M_n \to \infty$ as $n \to \infty$ and we have that for large n

$$\int_{-M_n}^{M_n} \left[(w_j/\sqrt{n} + \hat{v}_j)^{-m_j/2} \exp\left(- \frac{m_j \hat{v}_j}{2(w_j/\sqrt{n} + \hat{v}_j)} \right) \right] dw_j = \frac{\sqrt{n}\,\Gamma(m_j/2 - 1)}{(m_j \hat{v}_j/2)^{m_j/2-1}}.$$

Then for a given constant $c > 0$ we have that as $n \to \infty$

$$\frac{(m_j \hat{v}_j/2)^{m_j/2-1}}{\sqrt{n}\,\Gamma(m_j/2-1)} \int_{-c}^{c} \left[(w_j/\sqrt{n} + \hat{v}_j)^{-m_j/2} \exp\left(- \frac{m_j \hat{v}_j}{2(w_j/\sqrt{n} + \hat{v}_j)} \right) \right] dw_j$$
$$= \frac{\Gamma(m_j/2 - 1,\ 2/m_j + \delta_j) + \gamma(m_j/2 - 1,\ 2/m_j - \delta_j)}{\Gamma(m_j/2 - 1)} \to 1$$

where $\delta_j = 2c/[m_j \hat{v}_j \sqrt{n}]$ and $\Gamma(\alpha,\ t)$ and $\gamma(\alpha,\ t)$ is the upper and lower incomplete Gamma functions. This completes the proof since we have shown that the log posterior density of w_j converges towards

$$\log(\pi(w_j/\sqrt{n} + \hat{v}_j|\text{data})) = \text{ const.} - \frac{1}{2} w_i^2 \left[2\pi \left(\frac{1}{m_j} \sum_{i \in U_j} I_n(u_i) \right)^2 \Delta_j \right]^{-1} + \text{ small}$$
$$\to \text{ const.} - \frac{1}{2} w_j^2 \left[2\pi f_0(u_j)^2\, \Delta_j \right]^{-1},$$

as $n \to \infty$, which is the log-density a Gaussian distribution with expectation $\mu_j = 0$ and variance $\sigma_j^2 = f_0(u_j)^2\, \Delta_j$. □

The next example illustrates Lemma 1.9.

EXAMPLE 1.10. Assume that the same assumptions as in Example 1.8 is satisfied. But instead of using prior based on a product of Inverse-Gamma densities, we will assume that the prior density for $\boldsymbol{v} = (v_1, \ldots, v_J)$ is given by a product of independent $\pi_i(v_i)$ such that $\pi(\boldsymbol{v}) = \prod_i \pi_i(v_i)$, where $\pi_i(v_i)$ follows a gamma distribution with shape parameter $\alpha(u_i)$ and rate parameter $\beta(u_i)$. The posterior distribution has density given by

$$\begin{aligned} \pi(\boldsymbol{v}|\text{data}) &\propto \prod_{i=1}^{M} \pi_i(v_i) \times dL_W^*(u_i) \\ &= \prod_{i=1}^{M} v_i^{-[m_i/2 - \alpha(u_i) + 1] - 1} \exp\left[- \sum_{i=1}^{M} \left(\frac{\frac{1}{2} \sum_{u_j \in U_i} I_n^*(u_j)}{v_i} + \beta(u_i) v_i \right) \right] \end{aligned} \tag{1.13}$$

this implies that $v_i|$data follows a distribution that is proportional to the product of a Inverse-Gamma density and a Gamma density, see Appendix B. Let $\alpha'(u) = m_i/2 - \alpha(u_i) + 1$, $\beta'(u_i) = \beta'(u_i)$ and $\gamma'(u) = 1/2 \sum_{u_j \in U_i} I_n^*(u_j)$, then if $2\sqrt{\beta'(u_i)\gamma'(u_i)}$ is small enough we can use the approximative version of the expectation and variance given by

$$E[v_i|\text{data}] \approx \frac{m\, \Delta \hat{F}_{m_i}(u_i)}{m_i - 2\alpha(u_i) + 2} \to \Delta F_0(u_i)$$

and

$$n \operatorname{Var}(v_i|\text{data}) \approx \frac{2n(m_i\, \Delta \hat{F}_{m_i}(u_i))^2}{(m_i - 2\alpha(u_i) + 2)^2 (m_i - 2\alpha(u_i))} \to 2\pi\, \Delta F_0(u_i)^2/\Delta$$

as $n \to \infty$ as long as the numbers of intervals is fixed, see Appendix B.

2. Asymptotic properties

We will now return to the principal part approximation and motivated from the previous section we will now study some of the large sample properties of the posterior spectral measure and covariance function. In the first lemma we will establish the equivalent result to Lemma 1.9 for some more general situations. We will still assume that the true power spectrum is constant on subintervals of $[0, \pi]$, this is a somehow unnatural assumptions, but a sometimes a necessary conditions in for example a discrete approximations. In the following two results we will extend the results from Lemma 2.1 below to the general situation with smooth power spectrum and general finite Lévy processes. From theses results it will become fairly straightforward to extend the properties to covariance functions.

We will first establish the asymptotically distribution for the posterior spectral measures. We will use the same technique as we did in Lemma 1.9 and apply the Taylor expansion on the log-posterior density to show that this converges towards the log-density of a Gaussian random variable.

LEMMA 2.1. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a time series with true power spectrum $f_0(u)$ that satisfies the conditions of Theorem 1.1 and is constant on the subintervals $[u_i, u_{i-1}]$, where $i = 1, \ldots, M$ and $0 = u_0 < u_1 < \cdots < u_{M-1} < u_M = \pi$. Given a sample $Y(0), \ldots, Y(n-1)$ of size n from $Y(t)$ let the prior distribution for the spectral measure be given by a Lévy process, i.e. F is a Lévy process. Let $F(u) = \int_0^u dF(\omega)$ and define $\Delta_i = u_i - u_{i-1}$ and $\Delta F(u_i) = F(u_i) - F(u_{i-1})$ where $\pi_i(\Delta F(u_i))$ is the prior density for $\Delta F(u_i)$ specified by the Lévy process and assume that $\pi_i(\Delta F(u_i))$ is bounded with bounded derivative in a neighborhood of $\Delta\tilde{F}(u_i)$. Then for $i, j = 1, \ldots, M$, we have that $\Delta F(u_i)\big|\text{data}$ and $\Delta F(u_j)\big|\text{data}$ are asymptotically independent for $i \neq j$ and that $\Delta F(u_i)|\text{data}$ converges in distribution to a Gaussian distribution as $n \to \infty$, i.e.*

$$\sqrt{n}[\Delta F(u_i) - \Delta\tilde{F}(u_i)]\big|\text{data} \xrightarrow{d} N(0, 2\pi f_0(u_i)^2\, \Delta_i), \ a.s.$$

and where $\Delta\tilde{F}(u_i) = \displaystyle\int_{u_{i-1}}^{u_i} I_n(\omega)\, dv$ *and* $\Delta\tilde{F}(u_i) \xrightarrow{a.s.} F_0(u_i) - F_0(u_{i-1}) = \Delta F_0(u_i)$.

PROOF. Let $v_i = \Delta F(u_i)$, $\hat{v}_i = \Delta\tilde{F}(u_i)$ and $\Delta = \min(\Delta_1, \ldots, \Delta_M)$, also from the conditions of Theorem 1.1 we have that the spectral density is positive, i.e $f_0(u) \geq m > 0$, for $u \in [0, \pi]$. From Theorem 1.22 we do already know that

$$\hat{v_i} = \int_{u_{i-1}}^{u_i} I_n(v)\, dv \xrightarrow{a.s} \int_{u_{i-1}}^{u_i} f_0(v)\, dv = \Delta F_0(u_i), \quad \text{for all } u_i, \text{ where } i = 1, \ldots, M,$$

further under the assumption of constant spectral density we have that $\Delta F_0(u_i) = f_0(u_i)\, \Delta_i \geq m\Delta$, for all $i = 1, \ldots, M$. Let $w_i = \sqrt{n}(v_i - \hat{v}_i)$ be a scaled and centered version of v_i, for all $i = 1, \ldots, M$, then the posterior density of the vector $(w_1, \ldots w_M)\big|\text{data}$ is proportional to the product of the prior density times the likelihood, i.e.

$$\begin{aligned}\pi_w(w_1, \ldots w_M|\text{data}) &\propto \prod_{i=1}^{M} \pi_i'(w_i) \times \text{Lik}(\text{data}|w_1, \ldots, w_M) \\ &\propto \prod_{i=1}^{M} \pi_i(w_i/\sqrt{n} + \hat{v_i}) \times \text{Lik}(\text{data}|w_1, \ldots, w_M),\end{aligned}$$

where $\text{Lik}(\text{data}|w_1,\ldots,w_M)$ is the multivariate Gaussian likelihood. Since we have assumed that $Y(t)$ satisfies the conditions of Theorem 1.1 we can use the principal part as an approximation to the full the multivariate Gaussian log-likelihood. Under the assumption that the power spectrum is constant on subintervals of a given partition of the interval $[0,\pi]$ we can use the result from Remark 1.4. Therefore the logarithm of the posterior density is now given by the following approximation

$$\begin{aligned}
\log(\pi(w_1,\ldots,w_M\big|\text{data}) &= \sum_{i=1}^{M}\log(\pi_i(w_i')) + \log(\text{Lik}(\text{data}|w_1',\ldots,w_M')) + c^* \\
&\approx \sum_{i=1}^{M}\log(\pi_i(w_i')) - \frac{n}{2\pi}\sum_{i=1}^{M}\left[\log(w_i')\,\Delta_i + \frac{\Delta_i}{w_i'}\int_{u_{i-1}}^{u_i} I_n(v)\,dv\right] + c^* \\
&= \sum_{i=1}^{M}\left\{\log(\pi_i(w_i')) - \frac{n}{2\pi}\left[\log(w_i')\,\Delta_i + \frac{\Delta_i}{w_i'}\int_{u_{i-1}}^{u_i} I_n(v)\,dv\right]\right\} + c^*
\end{aligned}$$

where c^* is a constant and $w_i' = w_i/\sqrt{n} + \hat{v}_i$. We know that the principal part approximation of the log-likelihood is satisfactory for large n and it follows from the structure of the asymptotic log-posterior density that w_i and w_j are asymptotically independent for all $i,j = 1,\ldots M$ where $i \neq j$. Since the sequence of variables $w_1,\ldots,w_M$ are asymptotically independent and have the same marginal posterior densities, it is sufficient to show that $w_i|\text{data}$ converges towards a Gaussian distribution, for an arbitrary $i = 1,\ldots,M$, in order to prove the result. As mentioned we will use the Taylor expansion of the log-posterior distribution of $w_i|\text{data}$ and show that this converges towards the log-density of a Gaussian random variable.

From the proof of Lemma 1.9 we have that the derivatives of the log-likelihood part of $\log(\pi(w_i/\sqrt{n}+\hat{v}_i))$ with respect to w_i evaluated at zero are given by

$$n^{-k/2}\frac{d^k}{dw_i^k}\log(\Delta\tilde{L}_n(w_i))\big|_{w_i=\hat{v}_i} = \frac{(-1)^{k-1}(k-1)!(k-1)\Delta_i}{2\pi n^{k/2-1}\hat{v}_i^k}, \quad \text{for } k = 1,2,3,\ldots. \tag{2.1}$$

From equation (2.1) above it is now easy to verify that the principal part of the log likelihood reaches it maximum at $\hat{v}_i$. The Taylor expansion of $\log(\pi(w_i/\sqrt{n}+\hat{v}_i))$ around zero can now be expressed by

$$\begin{aligned}
&\log(\pi_i(w_i/\sqrt{n}+\hat{v}_i)) \\
&\quad= \log(\pi_i(\hat{v}_i|\text{data})) - \frac{1}{2}w_i^2\left[\frac{2\pi}{\Delta_i}\left(\int_{u_{i-1}}^{u_i} I_n(v)\,dv\right)^2\right]^{-1} \\
&\qquad+ \sum_{k=3}^{\infty} w_i^k\frac{(-1)^{k-1}(k-1)\Delta_i}{2\pi k n^{k/2-1}\hat{v}_i^k} + R_0^{\pi}(w_i) + c_i \\
&\quad= \log(\pi_i(\hat{v}_i|\text{data})) - \frac{1}{2}w_i^2\left[\frac{2\pi}{\Delta_i}\left(\int_{u_{i-1}}^{u_i} I_n(v)\,dv\right)^2\right]^{-1} + R_3^{\text{lik}}(w_i) + R_0^{\pi}(w_i) + c_i^*
\end{aligned}$$

where c_i^* is a constant. The first term is a constant and all we have to do is to show that the remainder $R_0^{\pi}(w_i)$ and $R_3^{\text{lik}}(w_i)$ converge to zero as n approaches infinity. From equation (1.11)

we know that there exist a number ξ, where $|\xi|, \leq |w_i|$, such that

$$\begin{aligned} R_0^{\pi}(w_i) &= w_i \left[n^{-1/2} \frac{d}{dv_i} \log(\pi_i(v_i)) \big|_{v_i = \xi/\sqrt{n} + \hat{v}_i} \right] \\ &= \frac{w_i}{n^{1/2} \pi(\xi/\sqrt{n} + \hat{v}_i)} \frac{d}{dv_i} \pi(\xi/\sqrt{n} + \hat{v}_i) \big|_{v_i = \xi/\sqrt{n} + \hat{v}_i}, \end{aligned} \tag{2.2}$$

and for a general integer k we know that there exists a number ξ', where $|\xi'| \leq |w_i|$, such that

$$R_k^{\text{lik}}(w_i) = \frac{1}{k!} \frac{\Delta_i}{2\pi n^{k/2-1}} w_i^k \left[\frac{(-1)^{k-1}(k-1)!}{(\xi/\sqrt{n} + \hat{v}_i)^k} - \frac{(-1)^k k!\, \hat{v}_i}{(\xi/\sqrt{n} + \hat{v}_i)^k} \right]. \tag{2.3}$$

Since the prior density is bounded and has bounded derivative, it is easy to see from (2.2) that $R_0^{\pi}(w_i) \to 0$ as $n \to \infty$ as long as w_i is bounded. Further we know from Theorem 1.22 that $\hat{v}_i \xrightarrow{a.s} F_0(u_i) - F_0(u_{i-1}) = \Delta F_0(u_i)$ and from the conditions we know tha $\Delta F_0(u_i) \geq m\Delta_i$, where $m > 0$, therefore it is now easy to verify that for (2.3) we have that

$$\lim_{n \to \infty} n^{k/2-1} R_k^{\text{lik}}(w_i) = w_i^k \frac{(-1)^{k-1}(k-1)\Delta_i}{2\pi k\, \Delta F_0(u_i)^k} \leq w_i^k \frac{(-1)^{k-1}(k-1)}{2\pi k m^k \Delta_i^{k-1}} < \infty$$

as long as w_i is bounded and $\Delta_i > 0$. This implies that for all $k > 2$ we will have that $R_k(w_i) \to 0$ as $n \to \infty$, if we are able to show that there exists a number c such that $\Pr\{|w_i| < c\} = 1 - \epsilon$ for all $i = 1, \ldots, M$. Under the assumption that the principal part approximation is good enough as an approximation to the multivariate Gaussian log-likelihood, the asymptotic density for $w_i|\text{data}$ is proportional to

$$\begin{aligned} \pi_{w_i}(w_i|\text{data}) &\propto \pi_i(w_i') \times \exp\left(-\frac{n}{2\pi} \left[\log(w_i')\, \Delta_i + \frac{\Delta_i}{w_i'} \int_{u_{i-1}}^{u_i} I_n(v)\, dv \right] \right) \\ &= \pi_i(w_i/\sqrt{n} + \hat{v}_i) \times \left[(w_i/\sqrt{n} + \hat{v}_i)^{-n\,\Delta_i/2\pi} \exp\left(-\frac{n\,\Delta_i \hat{v}_i}{2\pi(w_i/\sqrt{n} + \hat{v}_i)} \right) \right]. \end{aligned}$$

It follows now from Lemma 1.9 that for a given constant $c > 0$ we have that as $n \to \infty$

$$\begin{aligned} &\frac{(n\Delta_i \hat{v}_i/2\pi)^{n\Delta_i/2\pi - 1}}{\sqrt{n}\,\Gamma(n\Delta_i/2\pi - 1)} \int_{-c}^{c} \left[(w_i/\sqrt{n} + \hat{v}_j)^{-n\Delta_i/2\pi} \exp\left(-\frac{n\Delta_i \hat{v}_i}{2\pi(w_j/\sqrt{n} + \hat{v}_j)} \right) \right] dw_i \\ &\qquad = \frac{\Gamma(n\Delta_i/2\pi - 1,\ n\Delta_i/2\pi + \delta_j) + \gamma(n\Delta_i/2\pi - 1,\ n\Delta_i/2\pi - \delta_j)}{\Gamma(n\Delta_i/2\pi - 1)} \to 1 \end{aligned}$$

where $\delta_j = 2c/[n\Delta_i \hat{v}_i \sqrt{n}]$ and $\Gamma(\alpha,\ t)$ and $\gamma(\alpha,\ t)$ is the upper and lower incomplete Gamma functions. This completes the proof since we have shown that for large n

$$\begin{aligned} \log(\pi_i(w_i/\sqrt{n} + \hat{v}_i)) = \log(\pi_i(\hat{v}_i|\text{data})) - \frac{1}{2} w_i^2 \left[\frac{2\pi}{\Delta_i} \left(\int_{u_{i-1}}^{u_i} I_n(v)\, dv \right)^2 \right]^{-1} + c_i + \text{ small} \\ \to \text{const.} - \frac{1}{2} w_i^2 \left[2\pi f(u_i)^2 \Delta_i \right]^{-1} \end{aligned}$$

as $n \to \infty$ since $2\pi \hat{v}_i^2/\Delta_i \xrightarrow{a.s} 2\pi F_0(u_i)^2\, \Delta_i = f_0(u_i)^2\, \Delta_i$, which is the logarithm of a Gaussian density with expectation zero and variance σ^2, where $\sigma^2 = f_0(u_i)^2\, \Delta_i$. □

The next corollary is a necessary extension of Lemma 2.1 in order to prove the final result.

COROLLARY 2.2. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a time series with true power spectrum $f_0(u)$ that satisfies the conditions of Theorem 1.1. Given a sample $Y(0), \ldots, Y(n-1)$ of size n from $Y(t)$ we will assume the same construction as in Lemma 2.1 and let $0 = u_1' < u_2' < \cdots < u_{M'-1}' < u_{M'}' = \pi$ be an arbitrary partition of the interval $[0, \pi]$. Then for $i, j = 1, \ldots, M'$,*

$\Delta F(u_i')|\text{data}$ *and* $\Delta F(u_j')|\text{data}$ *are asymptotically independent for* $i \neq j$ *and* $\Delta F(u_i')|\text{data}$ *converges in distribution to a Gaussian distribution as* $n \to \infty$*, i.e*

$$\sqrt{n}[\Delta F(u_i') - \Delta\tilde{F}(u_i')] \xrightarrow{d} N(0, \int_{u_{i-1}'}^{u_i'} f_0(v)^2\, dv), \quad a.s. \tag{2.4}$$

PROOF. Let $0 = v_0 < v_1 < \cdots < v_{K-1} < v_K = \pi$ be a partition of the interval $[0, \pi]$ such that for every choice of u_i', where $i = 1, \ldots, M'$, there exist a subset U_i' of elements from $\{v_1, \ldots v_K\}$ such that

$$u_i' - u_{i-1}' = \sum_{v_j \in U_i'} v_j.$$

Suppose we are able to show that (2.4) is true for all v_i, where $i = 1, \ldots, K$, then (2.4) will automatically become true for all u_i', where $i = 1, \ldots, M'$, since we can construct the variables for larger subintervals as sums of independent variables over small subintervals.

First note that under the assumption that the true power spectrum is constant on subintervals we have that for any partition $0 = v_0 < v_1 < \cdots < v_{K-1} < v_K = \pi$ of the interval $[0, \pi]$ that includes the discontinuity points of the true power spectrum, $\{u_1, \ldots, u_{M-1}\}$, it is possible to rewrite the approximation of the log-likelihood given in Remark 1.4 as

$$\begin{aligned}
\tilde{L}_n(F) &= -\frac{n}{2\pi}\sum_{i=1}^{M}\left[\log(\Delta F(u_i))\,\Delta_i + \frac{\Delta_i}{\Delta F(u_i)}\int_{u_{i-1}}^{u_i} I_n(\omega)\, d\omega\right] \\
&= -\frac{n}{2\pi}\sum_{i=1}^{M}\left[\sum_{v_j \in U_i}\left[\log(\Delta F(v_j))\,\Delta_j + \frac{\Delta_j}{\Delta F(v_j)}\int_{v_{j-1}}^{v_j} I_n(\omega)\, d\omega\right]\right] \\
&= -\frac{n}{2\pi}\sum_{i=1}^{K}\left[\log(\Delta F(v_i))\,\Delta_j + \frac{\Delta_j}{\Delta F(v_j)}\int_{v_{j-1}}^{v_j} I_n(\omega)\, d\omega\right]
\end{aligned}$$

where $U_i = \{v_j \mid v_j \in \{v_0, v_1, \ldots, v_K\} \text{ and } u_{i-1} < v_j < u_i\}$ and $\Delta_j = v_j - v_{j-1}$. Let now $0 = v_0 < v_1 < \cdots < v_{K-1} < v_K = \pi$ be the partition of $[0, \pi]$ given by the two set of cut points of $\{u_0', u_1', \ldots, u_{M'}'\}$ and $\{u_0, u_1, \ldots, u_M\}$ and define $\Delta F(v_i) = F(v_i) - F(v_{i-1})$, $\Delta_i = v_i - v_{i-1}$ and $\Delta = \min(\Delta_1, \ldots, \Delta_K)$. The posterior distribution of $w_1, \ldots, w_K$ have density that is proportional to

$$\pi(w_1, \ldots, w_K|\text{data}) \propto \prod_{i=1}^{K} \pi_i(w_i/\sqrt{n} + \hat{v}_i) \times \text{Lik}(data|w_1, \ldots, w_K)$$

and by approximation of the log-posterior density we obtain the following

$$\begin{aligned}
&\log(\pi(w_1, \ldots, w_K|\text{data})) \\
&\quad = \sum_{i=1}^{K} \log(\pi_i(w_i/\sqrt{n} + \hat{v}_i)) + \log(\text{Lik}(data|w_1, \ldots, w_K)) + c^* \\
&\quad \approx \sum_{i=1}^{K}\left\{\log(\pi_i(w_i/\sqrt{n} + \hat{v}_i)) - \frac{n}{2\pi}\left[\log(\Delta F(v_i))\,\Delta_j + \frac{\Delta_j}{\Delta F(v_j)}\int_{v_{j-1}}^{v_j} I_n(\omega)\, d\omega\right]\right\} + c^*
\end{aligned}$$

where c^* is a constant and the result follows now from Lemma 2.1 as long as $\Delta > 0$. □

For the following result, let $W(\cdot)$ be a Brownian motion, i.e. a zero-mean normal process with independent increments and $\text{Var}\{W(t) - W(s)\} = t - s$, for all intervals (s, t). In the next master theorem we shall argue that the limit process of the standardized spectral measure given data has a Brownian motion as the limit process, note that this is a type of Bernstein-von Mises result, see **?**, in the sense that an exact mirror result to Theorem 1.20 is reached. If a stochastic process Z_n has Z as a limit process, as $n \to \infty$, it means that the process Z_n converges towards Z in in the Skorokhod topology on the space $D[0, \pi]$ of all right-continuous functions on $[0, \pi]$ with at most a finite number of discontinuities, see Chapter 3 in **?**.

In order to show that our process has the right limit it is sufficient to show two things. We have to show finite-dimensional convergence and tightness, see Chapter 3 in **?**. It turns out that finite-dimensional convergence follows almost directly from our earlier work, but it is much more complicated to show tightness and we shall only give an argument why this should hold if F is a Gamma process.

THEOREM 2.3. *Let $Y(t)$, where $t = 0, \pm 1, \ldots$, be a time series with true power spectrum $f_0(u)$ that satisfies the conditions of Theorem 1.1 and assume that $f_0(u)$ is smooth for $u \in [0, \pi]$. Given a sample $Y(0), \ldots, Y(n-1)$ of size n from $Y(t)$ let the prior distribution for the spectral measure be defined by a Gamma process. Then*

$$Z_n(t) = \sqrt{n}[F(t) - \tilde{F}(t)]\big|\text{data} \xrightarrow{d} W(2\pi \int_0^t f_0(\omega)^2 \, d\omega), \ \textit{a.s.} \tag{2.5}$$

where $\tilde{F}(t) = \int_0^t I_n(\omega) \, d\omega$, $\tilde{F}(t) \xrightarrow{a.s.} F_0(t)$ and $W(\cdot)$ is a Brownian motion.

PROOF. We will first show that we have finite-dimensional convergence. We know that in order to prove that the finite-dimensional convergence property is satisfied it is sufficient to show that that for any subinterval $[u'_{i-1}, u'_i]$ of an arbitrary partition $0 = u'_0 < u'_1 < \ldots < u'_{M'-1} < u'_{M'} = \pi$ of the interval $[0, \pi]$ we have that

$$Z_n(u'_i) - Z_n(u'_{i-1}) = \sqrt{n}(v_i - \hat{v})\big|\text{data} \xrightarrow{d} N(0, \int_{u'_{i-1}}^{u'_i} f_0(\omega)^2 \, d\omega)$$

Let $0 = u_0 < u_1 < \cdots < u_{M-1} < u_M = \pi$ be the coarsest partition of the interval $[0, \pi]$ such that there exist w_i, where $w_i \in [u_{i-1}, u_i]$, that satisfies

$$|f_0(w_i) - f_0(u)| < \epsilon \text{ for all } u \in [u_{i-1}, u_i]$$

and that $u_i - u_{i-1} > 0$ for all $i = 1, \ldots, K$. It is now acceptable to assume that power spectrum is constant on subintervals of the interval $[0, \pi]$, i.e. $f_0(u) = f_0(w_i)$ for $u \in [u_{i-1}, u_i]$. Let $0 = v_0 < v_1 < \cdots < v_{K-1} < v_K = \pi$ be the partition of the interval $[0, \pi]$ that is defined by the two set of cut points $\{u'_0, u'_1, \ldots, u'_{M'}\}$ and $\{u_0, u_1, \ldots, u_M\}$. Let $\Delta = \min(\Delta_1, \ldots, \Delta_K)$, where $\Delta_i = v_i - v_{i-1}$, then result follows from Lemma 2.1 as long as $\Delta > 0$.

We will now give an argument why the the tightness property is satisfied. From **?** we have that in order to prove tightness it is sufficient to show that

$$E[(Z_n(t) - Z_n(s)^2 (Z_n(u) - Z_n(t))^2] \leq K[G_n(u) - G_n(s)]^2,$$

for all choices of $t, s, u \in \mathbb{R}$ and integers n. Under the assumption that the principal part approximation is good enough this problems simplifies to show that $E[(Z_n(t) - Z_n(s)^2] < K[G_n(t) - G_n(s)]$ and from Lemma 2.1 we have that if $t > s$, then

$$E[(Z_n(t) - Z_n(s))^2] = n\,\mathrm{Var}(F(t) - F(s)|\mathrm{data}) = n\,\mathrm{Var}(dF(u_i)|\mathrm{data}).$$

Further from Appendix B we know that if $\alpha(u_i)$ and $\beta(u_i)$ are the respective parameters from the Gamma process, then

$$n\,\mathrm{Var}(dF(u_i)|\mathrm{data}) = \frac{n^2\Delta_i\hat{v}_i}{2\pi\beta(u_i)}\left[\frac{K_{\nu-1}(\kappa)}{K_{\nu+1}(\kappa)} - \left(\frac{K_\nu(\kappa)}{K_{\nu+1}(\kappa)}\right)^2\right]$$

where $\nu = n\Delta_i/2\pi - \alpha(u_i) + 1$, $\kappa = (2n\Delta_i\hat{v}_i/\pi\beta(u_i))^{1/2}$, $\hat{v}_i = \int_s^t I_n(w)\,dw$ and $\Delta_i = t - s$. In order to find a explicit expression for the G_n function we will set $s = 0$, then

$$G_n(t) = \frac{n^2 t}{2\pi\beta(t)}\int_0^t I_n(w)\,dw\left[\frac{K_{\nu'-1}(\kappa')}{K_{\nu'+1}(\kappa')} - \left(\frac{K_{\nu'}(\kappa')}{K_{\nu'+1}(\kappa')}\right)^2\right]$$

and $\nu' = nt/2\pi - \alpha(t) + 1$ and $\kappa' = (2nt\int_0^t I_n(w)\,dw/\pi\beta(t))^{1/2}$. From Appendix B we know that if the following is satisfied for large n

$$0 < \left[nt\int_0^t I_n(w)\,dw/2\pi\beta(u_i)\right]^{1/2} \le \left[nt^2 M/2\pi\beta(t)\right]^{1/2} \ll \sqrt{nt/2\pi - \alpha(t)}$$

which implies that it is sufficient to assume that $\beta(u) > K't$, where $K' \gg 4M$, $M = \max_{0\le u\le\pi}(f_0(u))$ and $f_0(u)$ is true power spectrum, then

$$G_n(t) \to \frac{2\pi M^2}{K'}t,$$

for all t as $n \to \infty$, which completes the argument. □

The next master corollary follows directly from the master theorem above.

COROLLARY 2.4. *Let $Y(t)$, where $t = 0, \pm1, \ldots$, be a time series with true power spectrum $f_0(u)$ that satisfies the conditions of Theorem 1.1. Given a sample $Y(0), \ldots, Y(n-1)$ of size n from $Y(t)$ let $\tilde{C}(h)$ be the estimator defined by*

$$\tilde{C}(h) = 2\int_0^\pi \cos(uh)\,d\tilde{F}(u).$$

Then

$$\begin{aligned}&\sqrt{n}[C(h) - \tilde{C}(h)]\big|\mathrm{data}\\ &\qquad = 2\int_0^\pi \cos(uh)\,\sqrt{n}[dF(u) - d\tilde{F}(u)]\big|\mathrm{data} \xrightarrow{d} 2\int_0^\pi \cos(uh)\,dZ(u),\ a.s.\end{aligned} \tag{2.6}$$

where $Z(t)$ is the limit process form Corollary 2.3, i.e. $Z(t) = W(2\pi\int_0^t f_0(\omega)^2\,d\omega)$.

PROOF. From **?** we have that for random processes Z_n the following is always true

$$Z_n(\cdot) \xrightarrow{d} Z(\cdot) \quad \Rightarrow \quad h(Z_n(\cdot)) \xrightarrow{d} h(Z(\cdot)) \tag{2.7}$$

if h is a functional that is continuous in supremum norm. Especially if the functional h is the stochastic integral given by

$$h(Z) = \int_0^\pi g(t)\,dZ(t)$$

then property (2.7) is true for all functions $g(t)$ if $g(t)$ is of bounded variation. Let $g(t) = \cos(th)$, which is obviously of bounded variation, this is exactly the situation in the corollary and the result follows now directly from Corollary 2.3. □

We will conclude this section with some examples that will illustrates the main ideas of the thesis and the asymptotic properties.

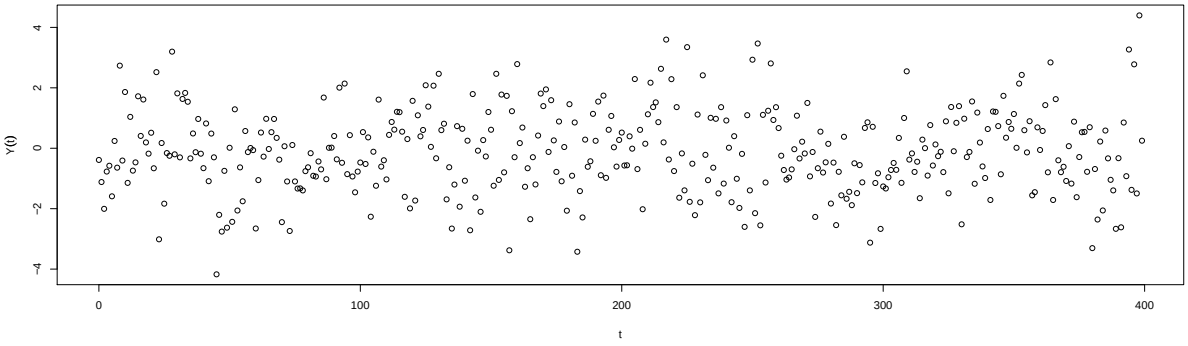

Figure 2.1: The whole odd-even time series $Y(t)$, where $t = 0, \ldots, 399$, $\rho_0 = 0.5$ and $\sigma_0 = 2$

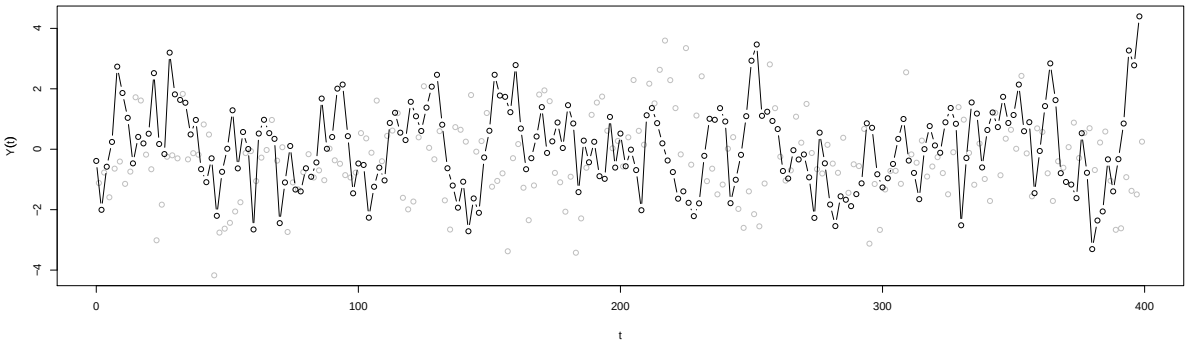

Figure 2.2: The same series as in Figure 2.1, where the even part of the series is marked, i.e $Y(t)$, where $t = 0, 2, \ldots, 398$.

EXAMPLE 2.5. Let $Y(t)$, where $t = 0, 1, 2, \ldots$ be a time series defined by

$$Y(k) \sim N(\rho Y(k-2), \sigma^2(1-\rho^2)), \quad \text{for } k \geq 2.$$

We can construct such a series by sampling to AR(1), both with $\rho = \rho_0$ and $\sigma_0 = \sigma$, and let the sequence of realizations $Y(0), Y(2), Y(4), \ldots$ be from the first and sample $Y(1), Y(3), Y(5), \ldots$ from the second. The idea is to use our nonparametric construction to estimate the predictive density for the variable $Y(n+1)|\text{data}$ and compare our solution with an AR(1) analysis. Since $Y(k)$ and $Y(k+1)$ is uncorrelated the AR(1) solution will result in $\rho = 0$ and the distribution for $Y(n+1)|\text{data}$ is therefore given gy

$$Y(n+1)|\text{data} \sim N(0, \hat{\sigma}^2)$$

where $\hat{\sigma}$ is estimated from data. In this specific example we know from well-known theory that the true distribution of $Y(n+1)$ is given by

$$Y(n+1) \sim N(\rho_0 Y(n-1), \sigma_0^2(1-\rho_0^2)).$$

We will sample two sequences of length $n_1 = n_2 = 200$, see Figure 2.1 and 2.1, and with prior distribution centered around $C(h) = \rho_\pi^{|h|}$, where $\rho_\pi = 0.1$, with this many observations we will expect that the a priori information will become almost completely forgotten and the answer will therefore not in particular depend on the choice of ρ_π.

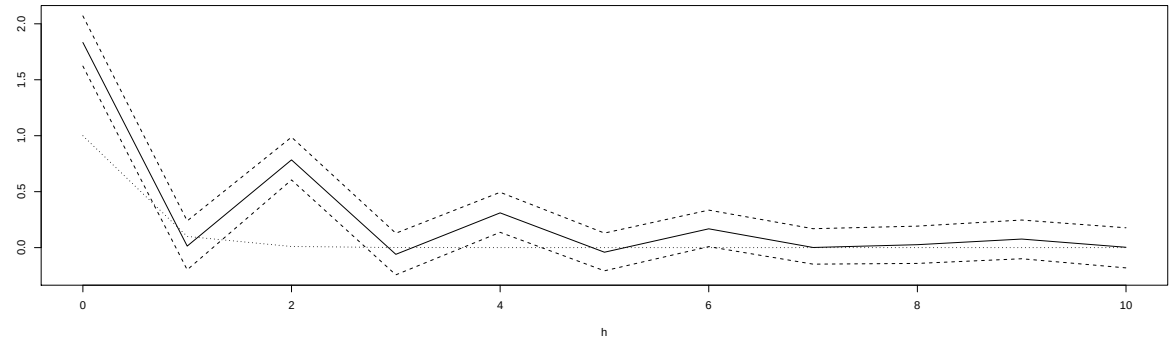

Figure 2.3: The posterior covariance function with upper and lower 0.95-bound estimated form the simulations, the dotted line is the a priori expectation

The posterior covariance function has clearly adopted the information from data, see Figure 2.3. From Figure 2.4 we see that the AR(1) analysis fails and that our nonparametric succeeds in finding the right density for $Y(n+1)|$data.

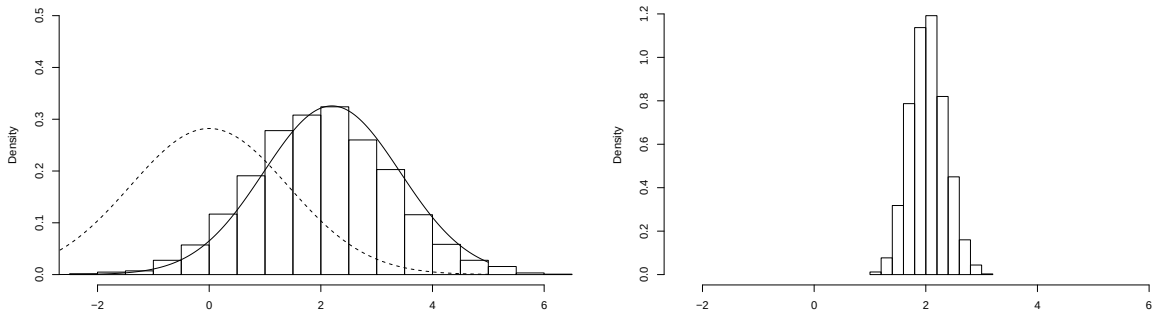

Figure 2.4: Estimated density for the observation $Y(400)|$data (left panel) with true density (solid line) and the wrong density (dotted line) based on the AR(1) model. In the left panel is the estimated density for $E[Y(n+1)]|$data.

EXAMPLE 2.6. In this example we will try to illustrate some of the large sample properties for the posterior covariance function, in the three figures below we have estimated the density for different combinations of specific values of the covariance functions given data, for three different

length of observed series. These densities should become approximate Gaussian distributed in the limit.

We have simulated samples from the stationary Gaussian time series $Y(t)$, where $t = 0, \ldots, n-1$, with expectation zero and covariance function $C_0(h) = \sigma^2 \rho^{|h|}$, where $\sigma = 2.50$ and $\rho = 0.5$, for $n = (10, 100, 1000)$. We have centered our prior distribution around $C_\pi(h) = \sigma_\pi^2 \rho_\pi^{|h|}$, where $\sigma_\pi = 0.95$, $\rho_\pi = 0.5$ and $\beta(u) = 5 + 10u$, we will also here assume that the effect from the prior distribution vanish for the long sequences of observations.

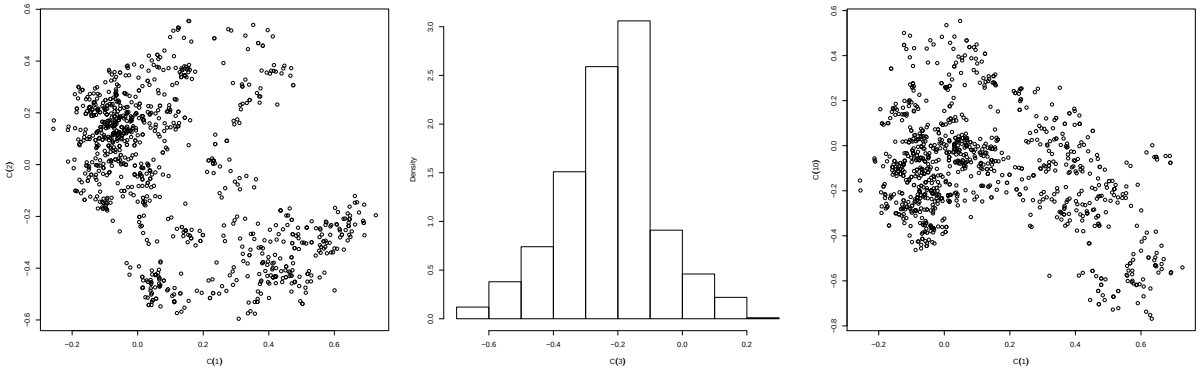

Figure 2.5: True value for $C(1) \approx 3.13$, $C(3) \approx 0.78$ and $C(10) \approx 0.061$, the length of the simulated sequence is $n = 10$.

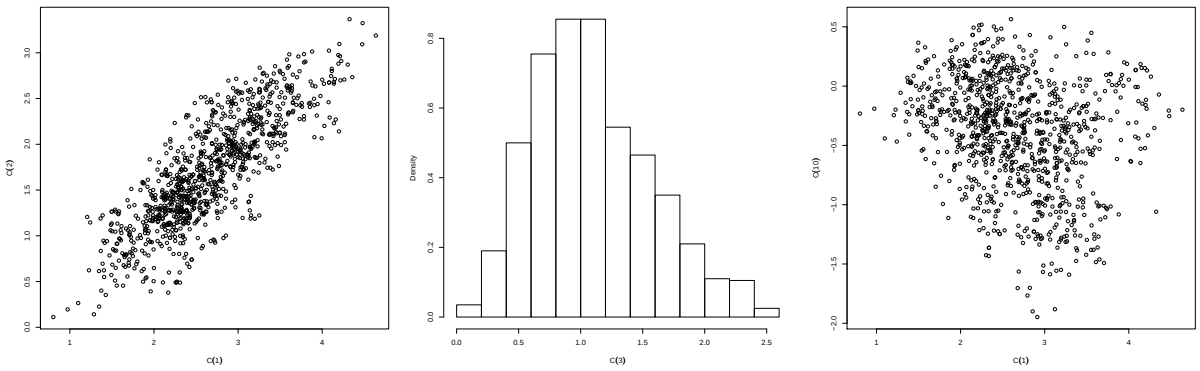

Figure 2.6: True value for $C(1) \approx 3.13$, $C(3) \approx 0.78$ and $C(10) \approx 0.061$, the length of the simulated sequence is $n = 100$.

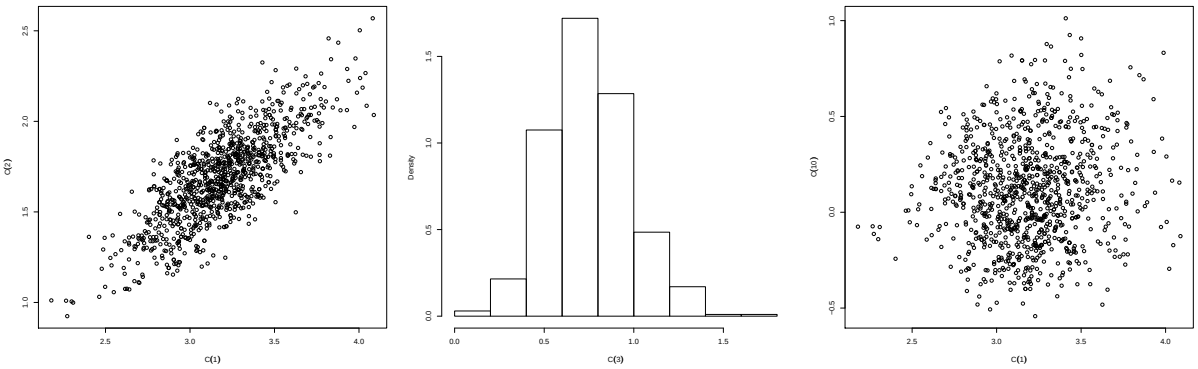

Figure 2.7: True value for $C(1) \approx 3.13$, $C(3) \approx 0.78$ and $C(10) \approx 0.061$, the length of the simulated sequence is $n = 1000$.

3. Concluding remarks

Through these chapters we have shown that it is possible to define model and estimate the covariance and the correlation function in stationary Gaussian time series by using the nonparametric Bayesian approach. We have introduced a large class of prior distributions and we have shown that it is possible to derive approximative posterior inference by the use of simulations. We have also derived some of the large-sample properties for the covariance function given data and shown that in the end/limit the estimates from the Bayesian approach becomes arbitrary close to the estimates given in the end of Section 1 in Chapter 2.

We have not used to much effort to discuss the correlation function and the prior construction with the Dirichlet process. Anyhow we do believe that there exists a similar result as Corollary 2.4 for the correlation function with prior distribution on the normalized spectral measure given by the Dirichlet process. This will involve finding the limit of the process

$$Z_n(t) = \sqrt{n}\left(\frac{F(t)}{F(\pi)} - \frac{\tilde{F}(t)}{\tilde{F(\pi)}}\right)\Big|\text{data}$$

that probably (from the work above) will have Brownian bridge as the limit.

The natural extension to the work presented in this thesis would be to assume that our process is stationary with unknown expectation different from zero and place a prior distribution on the unknown expectation μ. This will result in a semiparametric construction and there is not necessary ant problem involved in placing a prior distribution on μ, most of the general framework should hold, even the principal part approximation if we replace y_i with $y_i^* = y_i - \mu$. Things will become a bit more complicated and this construction will probably not lead to independency between the parameter μ and the spectral measure. Another idea is to place a prior distribution on some or all of the hyper parameters in the set $\{a, b, \sigma, \rho\}$ from Examples 2.3 and 2.4 or a similar construction.

In the work of **?** he mentions in the appendix to the first chapter that there exist similar results as the principal part and Whittle approximation for the Gaussian random field over a discrete domain $D \subset \mathbb{R}^d$. There was unfortunately not enough time to study these references properly

but at this point I do believe that it should be possible to extend some of the general results to from Chapter 3 to second-order stationary random fields over discrete domains.

At the end of Section 2 of Chapter 2 we showed that there was some unsatisfactory properties with our original construction, we showed that the variation for a given random covariance function $C(\cdot)$ will approach a finite limit for increasing lag distances, i.e.

$$\text{Var}(C(h)) \to c > 0, \quad \text{as } h \to \infty$$

where c is a positive constant. This is not a desirable property, since we wish to be able to construct prior distributions that are able reflect all of our a priori belief about the covariance function. For some classes of covariance functions we had a natural solution by combining a random covariance function with a deterministic covariance function from the same family in a way that we did not change original structure. This is of course always possible to do, but in most situations we will probably change the starting point and it is note a priori easy to find suitable pairs of functions. Although we are always able to search for the optimal combination, this topic need some more work to find good solutions to the most common models.

In Chapter 4 we derived some of the large sample properties for the posterior covariance function and spectral measure. In all these constructions we assumed that the approximation provided in **?** is good and close enough to the real log-likelihood and there is still some work left to prove that this is the case for almost any given situation. Note also that all the results were developed are defined for the random covariance function alone and not the construction with the a deterministic part. I will guess that it is still possible extend some of the work, but since this necessarily will involve the use of the convulsion between tow power spectrum functions, everything becomes more complicated.

An important topic in nonparametric Bayesian analysis is the support of the prior distribution. We wish that our prior distribution has the largest support possible. This means that given a covariance function $C^*(\cdot)$ we wish that there exist a positive probability that our random $C(h)$ will come close arbitrary close to this specific covariance function. In order to show this we have to define a metric and a ϵ-neighborhood and prove that our random covariance function have a positive probability to be within the ϵ-neighborhood of $C^*(\cdot)$ in our given metric, i.e.

$$\Pr\left\{\sum_{h=1}^{\infty} |C^*(h) - C(h)|(1/2)^h < \epsilon\right\} > 0$$

for any $\epsilon > 0$. We do believe that our nonparametric construction has full support, i.e. that for any covariance function $C^*(\cdot)$ and ϵ-neighborhood there is a positive probability that our $C(h)$ is in within this neighborhood.

CHAPTER 4

Continuous time

This chapter will be a generalization of the ideas presented in the earlier chapters. In Chapter 2 we mainly concerned with discrete time processes, in this chapter we will try to extend some of the concepts to general continuous time processes over a d-dimensional domain. Such processes are often referred to as random fields or spatial processes. To be specific we will for the most part restrict ourselves to the study of second order stationary Gaussian processes over a d-dimensional domain $D \subset \mathbb{R}^d$, where $d = 1, 2$, with isotropic covariance function. The main difference between this and the previous chapters is that everything becomes much more complicated for random fields and that we do not have such a nice approximations for the likelihood for Gaussian random fields as we have for stationary Gaussian time series, see Chapter 3.

The level of ambition is accordingly not as high as in Chapter 3, because of this and since much of the presentation is already made in the previous chapters the work here will be somehow less formal. We will introduce most of the general theory for Spatial Statistics and Random Fields in Section 1, the notation and results are mainly from the books **?** and **?**, both are excellent and complete introductions to the topics of Spatial Statistics. We will use some extra time to explain the concept of spatial prediction and Kriging. Both topics are large and important and are fully discussed in **?** and **?**, for a more detailed treatment see **?**. In Section 2 we continue the work of Section 2 of Chapter 2 and carry the nonparametric Bayesian ideas developed for stationary time series over to second order stationary Gaussian random fields. Since we do not have any well-behaved approximation we have to rely entirely on use of MCMC simulations to derive posterior inference.

1. Spatial data analysis

In spatial statistic, or spatial data analysis, the location of the observation is as important as the amount observed at the specific location. We will denote the output of a spatial process by $Z(\boldsymbol{s})$, where $\boldsymbol{s} \in D$ is a vector of spatial coordinates and D is a subset of $\mathbb{R}^d$. Note that if the domain is discrete and $D \subset \mathbb{R}$ then $Z(\boldsymbol{s})$ will become the time series discussed in the Chapter 2. If the domain D is one dimensional and continuous we will refer to the $Z(\boldsymbol{s})$ as a *continues time process* and we will use the more common notation $Y(t)$, where t will be referred to as time. If the domain D is of dimension d, where $d > 1$, the process is usually called a *random field*, or *spatial process*. The perhaps most common type of spatial processes are the continuous time process and random field over a two-dimensional plane.

There are three main types of spatial data and they are all characterized by their domain, see **?** or **?**. If the domain D is continuous and a fixed the data are said to be *geostatistical*, where continuous is continuous in the normal Euclidian sense and fixed means that the points

in the domain do not change, i.e., they are nonrandom. Note that the definition does not say anything about the attribute $Z(\boldsymbol{s})$, which can be either continuous or discrete. An example of geostatistical data can be measures of air temperature done at fixed locations, in theory the temperature could be measured everywhere and is therefore continuous, note that in this case the attribute (temperature) is also continuous.

If the domain of the process is discrete and fixed we will say that the spatial model is *lattice*, then D consist of a finite, or countable, set of nonrandom points or locations. An example might be a country divided into regions where a certain type of event is recorded, this event could be a disease, a special type of crime or some other incident.

The third type of data are *point patterns*, the main difference from the other two is that in point patters the points in the domain are random. Often in a point pattern study the domain itself is the main and interesting subject, not the size or amount observed. Examples of point patterns are the locations of a certain type of plant in a forest, the locations of lightning strikes in Norway or earthquakes around the world.

Our focus in this chapter will be the geostatistical data, the discrete or lattice type of models is mentioned in the concluding remarks of the previous chapter. We will now give a little more detailed discussion of random fields and establish some of the main concepts and theory needed throughout this chapter.

A *random field* or *spatial process* over a domain of dimension d will be determined by

$$\{Z(\boldsymbol{s})|\boldsymbol{s} \in D \subset \mathbb{R}^d\},$$

where $Z(\boldsymbol{s})$ is the output and $\boldsymbol{s}$ is a d-dimensional vector of coordinates, where D is either continuous, or a set of discrete spatial locations. As mention earlier if $d = 1$ we call the process a time process and we will write $Y(t)$ rather than $Z(\boldsymbol{s})$, and if $d = 2$ we will think D as a subset of the Cartesian plane, which $\boldsymbol{s} = (x, y)$ as Cartesian coordinates.

It is sometimes more instructive to think of a spatial process as a *random function* produced by a random experiment, rather than a sequence of random experiments indexed by their location. We will then associate $Z(\boldsymbol{s})$ with the value of the random function Z at position $\boldsymbol{s}$ where Z is the complete surface generated by a random experiment. It is important to realize that a collection of n spatial data are not a sample of size n, but rather an incomplete observation of one random experiment, i.e. a sample of size one from a n-dimensional distribution. In many situations it is unfortunately impossible to observe more than one independent realization of the same process.

In order to make inference we need to assume some restrictions about the random process, this compensates to some degree for incomplete observation and the impossibility of making repeated measures of the process. A random field is said to be *strictly* or *strongly stationary* if the distribution is invariant under translation, i.e.,

$$\Pr\{Z(\boldsymbol{s}_1) < z_1, \ldots, Z(\boldsymbol{s}_k) < z_k\} = \Pr\{Z(\boldsymbol{s}_1 + \boldsymbol{h}) < z_1, \ldots, Z(\boldsymbol{s}_k + \boldsymbol{h}) < z_k\},$$

for all $\mathbf{h}$ and k, where $k \in \mathbb{N}$, $\boldsymbol{h} \in \mathbb{R}^d$ and $\boldsymbol{s} \in D \subset \mathbb{R}^d$. The strict assumption is sometimes too harsh and not realistic in real life and in those cases it is more reasonable and common to require stationarity of the moments rather than of the distribution itself. We will say that a

random field is *second-order* or *weakly stationary* if the distribution of the field satisfies the two conditions

$$E[Z(\boldsymbol{s})] = \mu \text{ and } \operatorname{Cov}(Z(\boldsymbol{s}),\ Z(\boldsymbol{s}+\mathbf{h})) = C(\boldsymbol{h}), \quad \text{for } \boldsymbol{h} \in \mathbb{R}^d, \boldsymbol{s} \in D \subset \mathbb{R}^d. \tag{1.1}$$

Note that the spatial separation $\mathbf{h}$ will sometimes be referred to as a *lag-vector*. The function $C(\boldsymbol{h})$ is the *covariance function* and if $C(\boldsymbol{h})$ does not depend on the direction, but only the absolute distance between points, i.e. $C(\boldsymbol{h}) = C'(||\boldsymbol{h}||)$, we will say that the covariance function is *isotropic*. If a random field do not itself satisfy second-order stationary property, but have increments, $Z(\boldsymbol{s}) - Z(\boldsymbol{s}+\boldsymbol{h})$, which satisfy is second-order stationary property, we will say that the random field is *intrinsic* stationarity. The intrinsic property of a random field is sometimes defined by the property

$$E[Z(\boldsymbol{s}))] = \mu \text{ and } \frac{1}{2}\operatorname{Var}(Z(\boldsymbol{s}) - Z(\boldsymbol{s}+\boldsymbol{h})) = \gamma(\boldsymbol{h}), \quad \text{for } \boldsymbol{h} \in \mathbb{R}^d, \boldsymbol{s} \in D \subset \mathbb{R}^d.$$

The function $\gamma(\boldsymbol{h})$ is also known as the *semivariogram* of a spatial process. Note that for a second-order stationary random field we have that $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\boldsymbol{h})$ and we are in some sense free to choose which one to work with, also $\operatorname{Var}(Z(\boldsymbol{s})) = C(\mathbf{0}) = \sigma^2$ and the *correlation function* is defined as $R(\boldsymbol{h}) = C(\boldsymbol{h})/C(\mathbf{0})$. As we did for time series we will define a covariance function to be valid if it satisfies the positive definiteness property,

$$\sum_{i,j=1}^{k} a_i a_j \operatorname{Cov}(Z(\boldsymbol{s}_i),\ Z(\boldsymbol{s}_j)) \geq 0 \tag{1.2}$$

for any set of locations $(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k)$ and real numbers $(a_1, \ldots, a_k)$. Note that for second-order stationary random fields all the properties from Lemma 1.1 of Chapter 2 are satisfied for $C(\boldsymbol{h})$. Also if $C(\boldsymbol{h})$ is a valid covariance function in $\mathbb{R}^{d_2}$ then $C(\boldsymbol{h})$ will become valid covariance function in $\mathbb{R}^{d_1}$, where $d_2 > d_1$, but the opposite is not necessarily always true. A covariance function in $\mathbb{R}^d$ is said to be *separable* if $C(\boldsymbol{h})$ can be written as

$$C(\boldsymbol{h}) = \prod_{i=1}^{d} C_i(h_i), \tag{1.3}$$

where $\boldsymbol{h} = (h_1, \ldots, h_d) \in \mathbb{R}^d$ and $C(h_i)$, for $i = 1, \ldots, d$, is a covariance function in $\mathbb{R}$. Note that the separable property is usually not assumed unless the domain is separable in a natural way, such as space and time. Before we continue with some results for the covariance and correlation function we will define a random field as a *Gaussian field* if for any choice of positive integer k the cumulative distribution function

$$\Pr\{Z(\boldsymbol{s}_1) < z_1, \ldots, Z(\boldsymbol{s}_k) < z_n\} \tag{1.4}$$

is the same as for a k-variate Gaussian random variable.

The next result is known as Bochner's Theorem and it provides us with the necessary conditions we need to satisfy in order to construct prior distributions for the set of covariance functions.

THEOREM 1.1. *(Bochner's Theorem) A function $C(\boldsymbol{h})$ is positive definite if and only if it has a spectral representation*

$$C(\boldsymbol{h}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(i\boldsymbol{u}^T\boldsymbol{h}) dF(\boldsymbol{u}), \tag{1.5}$$

where $dF(\boldsymbol{u})$ is a bounded symmetric measure.

We will sometimes refer to $F(u)$ as the spectral measure as we did in Chapter 2, but in the case where $F(u)$ is differentiable, i.e. $dF(u) = f(u)\,du$ it is common to use the name *spectral density* for the function $f(u)$, not power spectrum as we did in Chapter 2. Also note that from Bochner's Theorem it is easy to verify that if $C(\boldsymbol{h})$ is a separable covariance function, it is sufficient to check if $C(h_i)$ is a valid covariance function in $\mathbb{R}$ for all $i = 1, \ldots, d$, in order to make sure that $C(\boldsymbol{h})$ is valid in $\mathbb{R}^d$.

In the case of isotropic covariance functions Bochner's Theorem simplifies to the following, see **?**.

THEOREM 1.2. *For $d \geq 2$, a function $C(h)$ is a valid isotropic covariance function for a random field on $\mathbb{R}^d$ if and only if $C(h)$ has the representation*

$$C(h) = 2^{d/2-1}\Gamma(d/2) \int_0^\infty (uh)^{-(d/2-1)} J_{d/2-1}(uh)\, dF(u), \quad \text{for } h \in \mathbb{R}^+$$

where $J_\nu(t)$ is the Bessel function of the first kind and F is nondecreasing, bounded on $[0, \infty)$ and satisfies $F(0) = 0$.

Since we only will be concerned with second-order stationary continuous time processes, which have isotropic covariance function by construction, and random processes with isotropic covariance functions. The version of Bochner's Theorem for isotropic covariance functions above will become our fundamental and main result for the rest of this chapter.

From **?** we do also have that if $\int_0^\infty h^{d-1}|C(h)|\, dh < \infty$ is satisfied, there exists an even more simplified connection between the isotropic covariance function and the spectral density given by the two equations

$$f(u) = (2\pi)^{-d/2} \int_0^\infty (uh)^{-(d/2-1)} J_{d/2-1}(uh) h^{d-1} C(h)\, dh \tag{1.6}$$

and

$$C(h) = (2\pi)^{d/2} \int_0^\infty (uh)^{-(d/2-1)} J_{d/2-1}(uh) u^{d-1} f(u)\, dh \tag{1.7}$$

for $u, h \in \mathbb{R}^+$.

For continuous time processes Bochner's Theorem is known as The Wiener-Khintchine Theorem, see **?**.

THEOREM 1.3. *(The Wiener-Khintchine Theorem) A necessary and sufficient condition for $R(h)$ to be a correlation function of some stochastically continuous stationary process $Y(t)$, where $t \in \mathbb{R}$, is that the exist a function $F(u)$ having the properties of a distribution function on $(-\infty, \infty)$ (i.e. $H(-\infty) = 0$, $H(+\infty) = 1$, and $H(u)$ non-decreasing), such that, for all h, $R(h)$ may be expressed in the form,*

$$R(h) = \int_{-\infty}^\infty \exp(iuh)\, dH(u).$$

It follows directly from The Wiener-Khintchine Theorem that we have a similar representation for the covariance function as Corollary 1.10. We will not state this as a separate result here,

but only observe that we have the following relation between the covariance function and the spectral density

$$f(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-uh)C(h)\,dh = \frac{1}{\pi}\int_{0}^{\infty} \cos(uh)C(h)\,dh, \quad \text{for } -\infty < u < \infty, \tag{1.8}$$

and

$$C(h) = \int_{-\infty}^{\infty} \exp(uh)f(u)\,du = \int_{0}^{\infty} \cos(uh)f(u)\,du, \quad \text{for } 0 \leq h < \infty, \tag{1.9}$$

see **?** for details. These two equation are very similar to those derived in Chapter 2, the only difference is that we integrate over a larger domain. Again note that equation (1.8) provides us with a method for how we always can find the corresponding spectral measure given a covariance function. We will now introduce some basic and common choices of covariance functions and their spectral measures for continuous time processes.

EXAMPLE 1.4. Let $Y(t)$, where $t \in [0, \infty)$, be a stationary Gaussian continuous time process with expectation zero and covariance function $C(h)$, where $h \in [0, \infty)$. In **?** the spectral densities for several common covariance functions is given, we will present some and rewrite them inspired from the results of the previous chapter such that they will fit our nonparametric approach presented in the next section.

Let $C(h) = \rho^{-ah}$ if $c = a\log(\rho)$ and $\rho > 1$ then the spectral density is given by

$$f(u) = \frac{\sigma^2}{c\pi(1 + (u/c)^2)}, \tag{1.10}$$

where the special case is to choose $c = a\phi$ then $C(h) = \exp(-a\phi h)$ which is known as the *exponential covariance function.*

Let $C(h) = \rho^{-ah^2}$ if $c = \sqrt{a\log(\rho)}$ and $\rho > 1$ then we have spectral density

$$f(u) = \frac{\sigma^2}{2c\sqrt{\pi}} \exp(-(u/2c)^2), \tag{1.11}$$

where the special case is $c = \sqrt{a\phi}$ and the covariance function become $C(h) = \sigma^2 \exp(-a\phi h^2)$, which is often referred to as the *Gaussian covariance function* from the structural similarity with the Gaussian distribution.

Let $C(h) = \sigma^2(1 - h\phi)I_{[0,\phi]}(h)$, where $I_{[a,b]}(h)$ is a indicator function, this covariance function is known as the *tent model*, apparently from its shape, and the spectral density is given by

$$f(u) = \frac{\sigma^2}{\phi\pi}\frac{1 - \cos(u/\phi)}{(u/\phi)^2}. \tag{1.12}$$

The next example introduces the classical solution of isotropic parametric covariance function for second order stationary random fields.

EXAMPLE 1.5. The Matérn class of covariance functions is an important and large class of isotropic covariance functions in spatial data analysis. The Matérn class is also a common choice of parametric class for second order stationary spatial processes with isotropic covariance function in higher dimension, i.e. $d > 1$, the Matérn class is also often used for the time continuous processes. There exist several different parameterizations for the Matérn functions, we will use the following given in **?**,

$$C(\boldsymbol{h}) = \phi(\alpha h)^{\nu} K_{\nu}(\alpha h), \quad \text{where } \phi > 0, \alpha > 0, \nu > 0 \text{ and } h = ||\boldsymbol{h}||, \tag{1.13}$$

where $K_\nu(t)$ is the modified Bessel function of the second kind. The corresponding spectral density is determined by

$$f(u) = \frac{2^{\nu-1}\phi\Gamma(\nu + d/2)\alpha^{2\nu}}{\pi^{d/2}}(\alpha^2 + u^2)^{-(\nu+d/2)}, \tag{1.14}$$

for $u \in \mathbb{R}^+$, and the relation between them are given by the equations (1.6) and (1.7).

REMARK 1.6. It is possible to obtain the Gaussian covariance function as a special case from the Matérn class by taking the limit $\nu \to \infty$, see **?**. It is more interesting for us that we are also able to obtain the exponential covariance function from the Matérn class. This gives us an explicit expression for the spectral density for a the exponential covariance function for all dimensions.

From **?** we have that the $K_{1/2}(t) = (\pi/2t)^{1/2}\exp(-t)$, then if $\phi = \sigma^2(2/\pi)^{1/2}$ we find that

$$C(\boldsymbol{h}) = \phi(\alpha h)^{1/2}K_{1/2}(\alpha h) = \sigma^2(2/\pi)^{1/2}(\alpha h)^{1/2}(\pi/[2\alpha h])^{1/2}\exp(-\alpha h) = \sigma^2\exp(-\alpha h)$$

where $h = ||\boldsymbol{h}||$. Let $\alpha = \log(\rho)$ then we have the first covariance from Example 1.4. It is now straightforward to derive the corresponding spectral density, from Example 1.5 we do now have that

$$f(u) = \sigma^2\left[\frac{\Gamma(1/2 + d/2)\alpha}{\pi^{1/2+d/2}}\right](\alpha^2 + u^2)^{-(1/2+d/2)}, \quad \text{where } u \in \mathbb{R}.$$

This last remark is important for us since we would like to use the results from Chapter 2 to construct prior distributions for the covariance function for random processes. From this result we do now have everything we need in order to make make flexible prior distributions around the class of exponential covariance functions in all dimensions. Before we extend the nonparametric ideas from Chapter 2 we will briefly discuss how we can make predictions about future outcomes in spatial data analysis.

1.1. Spatial prediction and Kriging. Let $Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n)$ be a sample of geostatistical data. As we have already mentioned earlier this sample can be interpreted as an incomplete observation from the surface $Z(\boldsymbol{s}, \omega)$, where ω represent a realization of an experiment.

A important topic in spatial data analysis is to predict the rest of the unobserved surface, or a specific unobserved value $Z(\boldsymbol{s}_0)$, where $\boldsymbol{s}_0 \in D$ is a new location, from the observed sample. If the distribution of the random filed is known the optimal predictor $p(\mathbf{Z}; \boldsymbol{s}_0)$ of $Z(\boldsymbol{s}_0)$ under squared error loss is the conditional mean as in Section 1 of Chapter 2. It is not unusual that the distribution of the random field is not known or that it is unrealistic to assume a specific distribution, in such situations it is common to restrict the search for optimal prediction to those that are linear, this method is known as Kriging after the South African mining engineer D. G. Krige.

We will first discuss the spatial prediction for second order stationary Gaussian processes and show how a simple example of Bayesian posterior prediction. Then we will give a short introduction to the simplest from of Kriging, also known as simple Kriging.

1.1.1. *Spatial prediction.* Let $\{Z(\boldsymbol{s})|\boldsymbol{s} \in D \subset \mathbb{R}^d\}$ be a random field, or random process, and suppose that we observe the realization $\mathbf{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))$ from the true process, where the locations $(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)$ are known. The idea is to use the information in the observed data to make inference about, or predict, the unknown outcome at a new location $\boldsymbol{s}_0 \in D$, this is sometimes referred to as *point prediction.* A predictor of $Z(\boldsymbol{s}_0)$ at the new location is denoted by $p(\mathbf{Z}; \boldsymbol{s}_0)$, in the case of an optimal predictor, relative to some loss function, we will use the notation $p_0(\mathbf{Z}; \boldsymbol{s}_0)$ To be specific we will assume that we are working under squared-error loss, i.e.

$$L(Z(\boldsymbol{s}_0), p(\mathbf{Z}; \boldsymbol{s}_0)) = (Z(\boldsymbol{s}_0) - p(\mathbf{Z}; \boldsymbol{s}_0))^2,$$

and the optimal predictor is the one that minimizes the expected loss, also known as Bayes risk. From Section 1.0.1 of Chapter 2 we know that under squared-error loss the optimal predictor is the conditional mean

$$p_0(\mathbf{Z}; \boldsymbol{s}_0) = E[Z(\boldsymbol{s}_0)|\mathbf{Z}]. \tag{1.15}$$

In order to compute the conditional expectation we need the conditional distribution given the observed data and the new unknown observation and in the case of a Gaussian random field this can be done as we explained in Example 1.5 in Chapter 2 which is straightforward to extend to random fields. To illustrate a future needed concept we will assume a slightly different model and show how a simple Bayesian prediction scenario works out. First we have to introduce two concepts form Bayesian theory.

In a Bayesian prediction setup suppose data, $\mathbf{y}$, are realizations from a model with density given by $f(y|\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$ is a unknown vector of parameters with priori density $\pi(\boldsymbol{\theta})$. If no data are observed, or before the data are collected, the distribution of an unknown y_0 is given by

$$p(y) = \int_\Theta f(y, \boldsymbol{\theta}) d\boldsymbol{\theta} = \int_\Theta f(y|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{1.16}$$

and is often called the *Prior Predictive Distribution* of y. After we have observed data $\mathbf{y}$ we would like to use the new information in the prediction of y_0, then

$$p(y|\mathbf{y}) = \int_\Theta f(y, \boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} = \int_\Theta f(y|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}, \tag{1.17}$$

is called the *Posterior Predictive Distribution*, or just the *Predictive Distribution.* Under squared-error loss the optimal predictor, the one that minimizes expected loss (Bayes risk), is the conditional mean of the posterior predicative distribution.

Example 1.7. Suppose $Z(\boldsymbol{s})$ is a Gaussian random field that satisfies the decomposition

$$Z(\boldsymbol{s}) = X(\boldsymbol{s})^T \boldsymbol{\beta} + \epsilon(\boldsymbol{s}), \quad \text{for } \boldsymbol{s} \in D \subset \mathbb{R}^d,$$

where $X(\boldsymbol{s})$ is a p-dimensional vector of known functions of s, $\boldsymbol{\beta}$ is a p-dimensional vector of unknown parameters with known prior density $\pi(\boldsymbol{\beta})$. This decomposition of the response as trend or deterministic part and random noise is quite common in spatial statistic. Suppose $Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n)$ is an observed sample and assume the following hierarchical structure:

$$\mathbf{Z}|\boldsymbol{\beta} \sim N(X\boldsymbol{\beta}, \Sigma_{\mathbf{Z}}) \text{ and } \boldsymbol{\beta} \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}}),$$

where $\Sigma_{\mathbf{Z}}$ is defined by a known covariance function $C(\boldsymbol{h})$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\Sigma_{\boldsymbol{\beta}}$ are both known parameters that reflect our a priori beliefs about the unknown parameter $\boldsymbol{\beta}$. In order to able to make predictions about an unobserved location $Z(\boldsymbol{s}_0)$ it is necessary to determine the posterior

predictive distribution. From Remark 1.6 in Chapter 2 it is clear that $f(Z(\boldsymbol{s}_0)|\boldsymbol{\beta}, \mathbf{Z})$ in (1.17) will become Gaussian and we need to determine the density of the posterior distribution $\pi(\boldsymbol{\beta}|\mathbf{Z})$,

$$\begin{aligned}
\pi(\boldsymbol{\beta}|\mathbf{Z}) &\propto f(\mathbf{Z}|\boldsymbol{\beta})\pi(\boldsymbol{\beta}) \\
&\propto \exp(-\frac{1}{2}[(\mathbf{Z}-X\boldsymbol{\beta})^T\Sigma_{\mathbf{Z}}^{-1}(\mathbf{Z}-X\boldsymbol{\beta}) + (\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}})^T\Sigma_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}})]) \\
&\propto \exp(-\frac{1}{2}[\boldsymbol{\beta}^T(\Sigma_{\boldsymbol{\beta}}^{-1} - X^T\Sigma_{\mathbf{Z}}^{-1}X)\boldsymbol{\beta} \\
&\qquad - \boldsymbol{\beta}^T(\Sigma_{\boldsymbol{\beta}}^{-1}\mu_{\boldsymbol{\beta}} + X^T\Sigma_{\mathbf{Z}}^{-1}\mathbf{Z}) \\
&\qquad - (\Sigma_{\boldsymbol{\beta}}^{-1}\mu_{\boldsymbol{\beta}} + X^T\Sigma_{\mathbf{Z}}^{-1}\mathbf{Z})^T\boldsymbol{\beta}] \\
&\qquad - \frac{1}{2}[\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{Z}}^T\Sigma_{\boldsymbol{\beta}|\mathbf{Z}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{Z}}]) \\
&\propto \exp(-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{Z}})^T\Sigma_{\boldsymbol{\beta}|\mathbf{Z}}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{Z}}))
\end{aligned}$$

which shows that the posterior distribution is Gaussian, with mean and covariance given by

$$\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{Z}} &= (\Sigma_{\boldsymbol{\beta}}^{-1} + X^T\Sigma_{\mathbf{Z}}^{-1}X)^{-1}(X^T\Sigma_n^{-1}\mathbf{Z} + \Sigma_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}) \\
\Sigma_{\boldsymbol{\beta}|\mathbf{Z}} &= (\Sigma_{\boldsymbol{\beta}}^{-1} + X^T\Sigma_{\mathbf{Z}}^{-1}X)^{-1}.
\end{aligned}$$

where X is the matrix $X = (X(\boldsymbol{s}_1), \ldots, X(\boldsymbol{s}_n))^T$. Under the assumption of square-error loss we are interested in the conditional mean, the optimal prediction. From the property of the double expectation the mean could be written as

$$E[Z(\boldsymbol{s}_0)|\mathbf{Z}] = E_{\boldsymbol{\beta}}[E[Z(\boldsymbol{s}_0)|\boldsymbol{\beta}, \mathbf{Z}]] = \int_{\boldsymbol{\beta}} (X\boldsymbol{\beta} + \boldsymbol{\sigma}_{0\mathbf{Z}}^T\Sigma_{\mathbf{Z}}^{-1}(\mathbf{Z}-X\boldsymbol{\beta}))\pi(\boldsymbol{\beta})d\boldsymbol{\beta} \tag{1.18}$$

and another way to write the covariance is

$$\mathrm{Var}(Z(\boldsymbol{s}_0)|\mathbf{Z}) = E_{\boldsymbol{\beta}}[\mathrm{Var}(Z(\boldsymbol{s}_0)|\boldsymbol{\beta}, \mathbf{Z})] + \mathrm{Var}_{\boldsymbol{\beta}}(E[Z(\boldsymbol{s}_0)|\boldsymbol{\beta}, \mathbf{Z}]). \tag{1.19}$$

The exact solution to (1.18) and (1.19) can be found in **?** who also treats other related problems where the assumptions are relaxed and prior distributions are defined on the remaining parameters in the model.

$$\begin{aligned}
E[Z(\boldsymbol{s}_0)|\mathbf{Z}] &= (X(\boldsymbol{s}_0) - \boldsymbol{\sigma}_{0\mathbf{Z}}^T\Sigma_{\mathbf{Z}}^{-1}X)(\Sigma_{\boldsymbol{\beta}}^{-1} + X^T\Sigma_{\mathbf{Z}}^{-1}X)^{-1}\Sigma_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}} \\
&\quad + [\boldsymbol{\sigma}_{0\mathbf{Z}}^T\Sigma_{\mathbf{Z}}^{-1} + (X(\boldsymbol{s}_0) - \boldsymbol{\sigma}_{0\mathbf{Z}}^T\Sigma_{\mathbf{Z}}^{-1}X)(\Sigma_{\boldsymbol{\beta}}^{-1} + X^T\Sigma_{\mathbf{Z}}^{-1}X)^{-1}X^T\Sigma_{\mathbf{Z}}^{-1}]\mathbf{Z} \\
\mathrm{Var}(Z(\boldsymbol{s}_0)|\mathbf{Z}) &= (\sigma_0 - \boldsymbol{\sigma}_{0\mathbf{Z}}^T\Sigma_{\mathbf{Z}}^{-1}\boldsymbol{\sigma}_{0\mathbf{Z}}) \\
&\quad + (X(\boldsymbol{s}_0) - \boldsymbol{\sigma}_{0\mathbf{Z}}^T\Sigma_{\mathbf{Z}}^{-1}X)(\Sigma_{\boldsymbol{\beta}}^{-1} + X^T\Sigma_{\mathbf{Z}}^{-1}X)^{-1}(X(\boldsymbol{s}_0) - \boldsymbol{\sigma}_{0\mathbf{Z}}^T\Sigma_{\mathbf{Z}}^{-1}X)^T.
\end{aligned}$$

The two solutions are not very informative in the way they are stated here, but on the other hand the exact expressions do exist, in the models we will introduce later in this chapter the exact solutions do not exist and we will have to use equations similar to (1.18) and (1.19) together with MCMC simulation.

1.1.2. *Kriging.* Let $\{Z(\boldsymbol{s})|\boldsymbol{s} \in D \subset \mathbb{R}^d\}$ be a random field where D is fixed and continuous and let $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))$ be a sample of size n. Assume that the mean $E[Z(\boldsymbol{s})] = \mu(\boldsymbol{s})$ and the covariance structure is known and therefore will also both $E[\boldsymbol{Z}] = \boldsymbol{\mu}_{\boldsymbol{Z}}$ and $\mathrm{Var}(\boldsymbol{Z}) = \Sigma_{\boldsymbol{Z}}$ be known. The goal is to find the predictor $p(\boldsymbol{Z}, Z(\boldsymbol{s}_0))$ of $Z(\boldsymbol{s}_0)$ that minimizes $E[(p(\boldsymbol{Z}, Z(\boldsymbol{s}_0)) - Z(\boldsymbol{s}_0))^2]$ and is linear, i.e. $p(\boldsymbol{Z}, Z(\boldsymbol{s}_0)) = \lambda_0 + \boldsymbol{\lambda}^T\boldsymbol{Z}$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^T$. From **?** we have

that the unique solution to the stated problem is given by $\lambda_0 = \mu(\boldsymbol{s}_0) - \boldsymbol{\lambda}^T \boldsymbol{\mu}_{\boldsymbol{Z}}$ and $\boldsymbol{\lambda} = \Sigma_{\boldsymbol{Z}}^{-1}\boldsymbol{\sigma}$, where $\boldsymbol{\sigma} = \text{Cov}(\boldsymbol{Z},\ Z(\boldsymbol{s}_0))$ and the optimal linear predictor is given by

$$p_0(\boldsymbol{Z}, Z(\boldsymbol{s}_0)) = \mu(\boldsymbol{s}_0) + \boldsymbol{\sigma}^T \Sigma_{\boldsymbol{Z}}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}) \tag{1.20}$$

and is known as *simple Kriging*. The simple kriging solution is also the best linear predictor under squared-error loss and is an unbiased estimator. Also from **?** we have that the simple kriging variance is given by

$$\sigma_0^2 = \sigma^2 - \boldsymbol{\sigma}^T \Sigma_{\boldsymbol{Z}}^{-1}\boldsymbol{\sigma}, \tag{1.21}$$

where $\sigma^2 = \text{Var}(Z(\boldsymbol{s}_0))$. From the equations above it is possible to show that simple kriging will predict the exact observed value at the known locations, it is said to "honor the data", which is a desirable property for many predictors. The situation with constant but unknown mean is known as *ordinary Kriging* and there exists an explicit and optimal linear unbiased predictor as a solution, see **?**, **?** or **?** for more about Kriging, Ordinary Kriging and other general types of prediction in more general situations.

2. Bayesian methods

This section will in some sense be the natural extension of the ideas from Section 2 in Chapter 2 to the case of continuous time processes. We will show that the framework developed for stationary time series will become fairly easy to extend and use in order to make prior distributions for isotropic covariance functions for random processes over a d-dimensional domain. The discussion will be less detailed since the main ideas have already been introduced and discussed in Section 2 of Chapter 2. As we have already mentioned we do not have the nice approximations for the multivariate Gaussian likelihood for continuous time processes such as we did for the time series. We will therefore not be able to derive the explicit asymptotic behavior for the posterior distribution and the only solution is to use simulation to study the posterior process properties.

The section is divided into two parts, we will in the first discuss some choices of prior distributions and in the second section we will discuss how we can obtain posterior inference based on MCMC simulations.

2.1. Prior distributions. Let $\{Z(\boldsymbol{s}) | \boldsymbol{s} \in D \subset \mathbb{R}^d\}$ be a second-order stationary random process with expectation zero and unknown isotropic covariance function. We will follow the same idea as in Section 2 of Chapter 2 and construct the prior distribution for the covariance function by placing a prior distribution on the set of spectral measurers, i.e. view the spectral measure F as a random process. The reason for this is the same as in the discrete time chapter, that this is a simple and straightforward method to always be sure that we are working with valid covariance functions. Therefore we will again be most concerned with the prior and posterior distributions for the spectral measures which we will use to make inference about the covariance function.

From Theorem 1.2 it is clear that all the random measure F needs to satisfy is to be a finite and positive increment process. It would therefore be sufficient to assume that F is a Lévy process

that also satisfies $\Pr\{F(\infty) < \infty\} = 1$. In order to more specific we will assume that F is a Gamma process defined on $[0, \infty)$, i.e. F is a independent positive increment process, where

$$dF(u) \sim \mathrm{Ga}(\alpha(u)\, du, \beta(u))$$

and $\mathrm{Ga}(\alpha, \beta)$ is the Gamma distribution with shape parameter α and rate parameter β. We will define $F(0) = 0$ and the common choice for the parameters will be $\alpha(u) = b(u) f_0(u)\, du$ and $\beta(u) = b(u)$, where $f_0(u)$ is the spectral density determined by our a priori information. As in Chapter 2 we will refer to the following function as a *random isotropic covariance function*

$$C(h) = 2^{d/2-1}\Gamma(d/2) \int_0^\infty (uh)^{-(d/2-1)} J_{d/2-1}(uh)\, dF(u), \tag{2.1}$$

defined for $h \in [0, \infty)$, where F will be called the *random spectral measure*. Also we have that

$$E[C(h)] = 2^{d/2-1}\Gamma(d/2) \int_0^\infty (uh)^{-(d/2-1)} J_{d/2-1}(uh) f_0(u)\, du = C_0(h)$$

where $h \in [0, \infty)$, and it is clear that we should choose the spectral density $f_0(u)$ that corresponds to our a priori guess for the covariance function. The variance of the random $C(h)$ becomes

$$\mathrm{Var}(C(h)) = 2^{d-2}\Gamma(d/2)^2 \int_0^\infty (uh)^{-(d-2)} J_{d/2-1}(uh)^2 f_0(u)/b(u)\, du,$$

where $h \in [0, \infty)$. It is not obvious how different choices of $\beta(u)$ will affect the variation in $C(h)$ for a general dimension d.

In the one-dimensional situation we can replace Theorem 1.2 with the simpler covariance extension of the Wiener-Khintchine Theorem 1.3. It is now fairly easy see how we should specify $\alpha(u)$ and $\beta(u)$ in order to reflect our prior beliefs from the expressions for the expectation and variance as we did for time series. It is straightforward to verify that the expectation is given by

$$E[C(h)] = 2 \int_0^\infty \cos(uh)\, E[dF(u)] = 2 \int_0^\infty \cos(uh) \alpha(u)/\beta(u)\, du, \tag{2.2}$$

and the variance

$$\mathrm{Var}(C(h)) = 4 \int_0^\infty \cos^2(uh)\ \mathrm{Var}(dF(u)) = 2 \int_0^\infty [1 + \cos(u2h)] \alpha(u)/\beta(u)^2\, du. \tag{2.3}$$

Our common choice is to choose $\alpha(u) = b(u) f_0(u)\, du$ and $\beta(u) = b(u)$ and the simple solution will be to assume that b is a positive constant and $f_0(u)$ is the spectral density that corresponds to our prior belief for the covariance function. As in Chapter 2 it is not obvious what the exact expression for the variance becomes unless $b(u)$ is a constant and we still get the separation of the variance into two parts, one that becomes a constant and one that approaches zero for large separation, as we did in the situation with stationary time series.

Note that there is at least one obvious reason why it might be desirable to let $b(u)$ be an increasing function and not a constant. The high end of the spectrum corresponds to short oscillations in the covariance, then by forcing the posterior process to be close to the prior process for high frequencies we might smooth out some of the noise from the observation. In the two following examples we will show how we can obtain explicit solutions for two simple classes of covariance functions for continuous time processes.

EXAMPLE 2.1. In Example 1.4 we introduced among others the class of covariance functions of the form $C(h) = \rho^{-ah}$, for $\rho > 1$ and $h = [0, \infty)$, for continuous time processes. The spectral density is given by

$$f(u) = \frac{\sigma^2}{c\pi(1 + (u/c)^2)}, \text{ for } u \in [0, \infty),$$

where $c = a\log(\rho)$ and we will define a prior distribution on covariance function with expectation $E[C(h)] = \rho^h$ by defining a suitable Gamma process as a prior distribution for the spectral measure. We will use what we already know from the time series chapter on how to solve the problem with the constant variance, i.e. we will combine a random and stochastic covariance function. Let $\alpha(u) = b(u)f_0(u)\,du > 0$ and $\beta(u) = b(u) > 0$, where $f_0(u)$ is the spectral density above where ρ is replaced with ρ^a, then

$$E[\rho^{(1-a)h}C(h)] = 2\rho^{-(1-a)h}\int_0^\infty \cos(uh)\,E[dF(u)] = \sigma^2\rho^{-ah}$$

and

$$\text{Var}(\rho^{(1-a)h}C(h)) = 2\rho^{-2(1-a)h}\int_0^\infty (1 + \cos(u2h))\ \text{Var}(dF(u))$$

where both are defined for $h \in [0, \infty)$. If $b(u) = b$ is a constant, then

$$\text{Var}(\rho^{(1-a)h}C(h)) = \frac{\sigma^2}{b}[2\rho^{-2(1-a)h}F(\infty) + \rho^{-2h}] = \frac{\sigma^2}{b}[\rho^{-2(1-a)h} + \rho^{-2h}],$$

where $h \in [0, \infty)$ and $F_0(\infty) = 1/2$ since $f_0(u)$ is the Cauchy density. The random covariance function is now defined by the four parameters (a, b, ρ, σ) where (ρ, σ) controls the expected structure of the covariance function and reflects our a priori beliefs. The parameter a determines how fast the random covariance function will converges towards our a priori expectation and b determines the amount of variation and reflects the precision in the prior guess.

It is not necessary to let $b(u)$ be a constant, but we will not always get an explicit expression for the variance of $C(h)$ for every choice of $b(u)$. We mentioned above why we perhaps might want to let $b(u)$ be a function and we will now show one possible choice. Assume that $b(u)$ is of the same form as the spectral density, $b(u) = b + (u/d)^2$ where b, d is positive numbers, then it is possible to show that

$$\text{Var}(\rho^{(1-a)h}C(h)) = \frac{\sigma^2\rho^{-2(1-a)h}}{c^2 - bd^2}\left[\frac{cd}{\sqrt{b}}(1 + \rho'^{-2\sqrt{b}h}) - d^2(1 + \rho^{-2ah})\right],$$

where $h \in [0, \infty)$, $c = a\log(\rho)$ and $\rho' = \exp(d)$.

EXAMPLE 2.2. From the discussion in Example 2.1 it is clear that the we can use the same construction to make random covariance functions that are close to the Gaussian, i.e. where $C(h) = \rho^{h^2}$. From Example 1.4 we have that the spectral density is given by

$$f(u) = \frac{\sigma^2}{2c\sqrt{\pi}}\exp(-(u/2c)^2),$$

where $u \in [0, \infty)$ and $c = \sqrt{a\log(\rho)}$. Let F be a gamma process defined by the shape and rate parameters $\alpha(u) = b(u)f_0(u)\,du > 0$ and $\beta(u) = b(u) > 0$, where $f_0(u)$is the spectral density above with ρ replaced by ρ^a, then

$$\rho^{(1-a)h^2}C(h) = 2\rho^{(1-a)h^2}\int_0^\infty \cos(uh)\,dF(u), \tag{2.4}$$

where $h \in [0, \infty)$, is a random covariance function which is always valid. The parameters (a, b, ρ, σ) have the same interpretation as in Example 2.1. It is also possible to show that

$$E[\rho^{(1-a)h^2} C(h)] = \rho^{ah^2} \quad \text{and} \quad \text{Var}(\rho^{(1-a)h^2} C(h)) = \frac{\sigma^2}{b}[\rho^{(1-a)h^2} + \rho^{a2h^2}]$$

for $h \in [0, \infty)$ and $b(u) = b$ since $F_0(\infty) = 1$. If it is needed we could choose $b(u) = b \exp((u/2d)^2)$, where b, d are positive numbers which gives us a new expression for the variance

$$\text{Var}(\rho^{(1-a)h^2} C(h)) = \sigma^2 \rho^{(1-a)h^2} \frac{d}{\sqrt{c^2 + d^2}} \left[1 + \exp(-(2hcd)^2/[c^2 + d^2] \right]$$

for $h \in [0, \infty)$.

REMARK 2.3. From Remark 1.6, Theorem 1.2 and what we have already discussed in this section, it is clear that we are always able to make quite flexible prior distributions that are close to the exponential isotopic covariance function for random fields over a general domain $D \subset \mathbb{R}^d$, where $d = 1, 2, \ldots$. Let F be a Gamma process as discussed above with parameters $\alpha(u) = b(u) f_0(u)\, du$ and $\beta(u) = b(u)$. The spectral density $f_0(u)$ is given by

$$f_0(u) = \sigma^2 \left[\frac{\Gamma(1/2 + d/2)\alpha}{\pi^{1/2+d/2}} \right] (\alpha^2 + u^2)^{-(1/2+d/2)}, \quad \text{for } -\infty < u < \infty,$$

where $\alpha = \log(\rho^a)$. Then

$$\rho^{(1-a)h} C(\boldsymbol{h}) = \rho^{(1-a)h} 2^{d/2-1} \Gamma(d/2) \int_0^\infty (uh)^{-(d/2-1)} J_{d/2-1}(uh)\, dF(u),$$

where $h = ||\boldsymbol{h}||$, will be a random covariance function with expectation $E[\rho^{(1-a)h} C(\boldsymbol{h})] = \rho^{||h||}$. It is clearly not straightforward to evaluate how different choices of $b(u)$ functions will affect the variation of $C(\boldsymbol{h})$, but $b(u)$ can always be set to a constant or be chosen so that it reduces high frequencies variation of F.

2.1.1. *Alternative Gamma Process Representation.* In the Examples above we have for the most assumed that the parameters for the Gamma Process are of the form $\alpha(u) = b(u) f_0(u)\, du > 0$ and $\beta(u) = b(u) > 0$, where $b(u)$ often for simplicity is assumed to be a constant and $f_0(u)$ is the spectral density that corresponds to our a priori guess for the covariance function $C_0(h)$. For some reason it turns out that this construction is not very stable to simulate from by the standard methods in R and for this reason we introduce the alternative parameterization for the Gamma process, let $\alpha(u) = b(u)\, du > 0$ and $\beta(u) = b(u)/f_0(u) > 0$.

$$E[C(h)] = 2 \int_0^\infty \cos(uh)\, E[dF(u)] = 2 \int_0^\infty \cos(uh) f_\pi(u)\, du, \tag{2.5}$$

and

$$\text{Var}(C(h)) = 4 \int_0^\infty \cos^2(uh)\ \text{Var}(dF(u)) = 2 \int_0^\infty [1 + \cos(u2h)] f_\pi(u)^2 / b(u)\, du. \tag{2.6}$$

This alternative parameterization turns out to much more stable to simulate from, but since the expression for the variance involves $f_0(u)^2$ we do not get such a simple expression as we did with the first parameterization. We do not necessary need an explicit expression, but in the next example we show a possible solution for our favorite example.

EXAMPLE 2.4. Assume that $f_0(u)$ is the spectral density for the exponential model given in Example 2.1 and we will use the alternative representation of the Gamma process given above. Let $\beta(u) = b$, the expectation of the random covariance function is clearly not changed, if we try to evaluate the variance we find that

$$\begin{aligned}\mathrm{Var}(C(h)) &= 4\int_0^\infty \cos^2(uh)\ \mathrm{Var}(dF(u)) \\ &= 4\int_0^\infty \cos^2(uh)\frac{\sigma^2 b}{(\pi\phi(1+(u/\phi)^2))^2 b^2}\,du = 2\frac{\sigma^2}{b\pi^2\phi^2}\int_0^\infty [1+\cos(2uh)]\frac{1}{(1+(u/\phi)^2)^2}\,du,\end{aligned}$$

which is not straight forward to evaluate. If we instead assume that

$$\beta(u)^{-1} = (1+(u/\phi)^2)^k \Big/ \Big[\prod_{i=1}^{l}(b_i + c_i u^2)\Big],$$

where $k = 1, 2$ and normally $l \geq k$, it is possible to solve analytically, suppose $k = l = 2$ then

$$\mathrm{Var}(C(h)) = \frac{\sigma^2}{\pi\phi^2(b_2c_1 - b_1c_2)}\sum_{i=1}^{2}(-1)^{i+1}\sqrt{\frac{c_i}{b_i}}\Big[1+\exp(-2\sqrt{b_i/c_i}h)\Big].$$

As a final remark to this section, assume that $\{Y(t)|t \in D \subset \mathbb{R}^d\}$ is a second-order stationary Gaussian continuous time process with expectation zero and unknown covariance function $C(\boldsymbol{h})$, where $C(\boldsymbol{h})$ is separable, i.e.

$$C(\boldsymbol{h}) = \prod_{i=1}^{d} C(h_i), \quad \text{where } \boldsymbol{h}^T = (h_1, \ldots, h_d).$$

Then it is straightforward to generalize the work we have done for one-dimensional random covariance to produce prior distributions for the separable covariance function.

2.2. Posterior simulation. In Chapter 2 we concluded that the full multivariate Gaussian likelihood was too complicated to use in order to make exact inference about the posterior process. The solution to this problem was to introduce an approximation that was easier to work with and that become arbitrarily close to the real log-likelihood in the limit. As far as I know there does not exist such nice approximations for the likelihood of second-order stationary Gaussian random field over a d-dimensional domain. For this reason we have to base our posterior inference on the MCMC simulations. The Markov chain Monte Carlo simulation routine was also used in Chapter 2 and there is a very brief explanation of the method and how this was implemented in R in Appendix A and Appendix C.

Let $Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n)$ be a sample from a second-order stationary Gaussian random field on a d-dimensional domain with unknown isotropic covariance function. Following the ideas from the previous section define a prior distribution for the spectral measure by defining a suitable Gamma process, we do then know from the discussion in Chapter 2 that in order to specify the posterior process it is sufficient to specify the posterior distribution for any finite collection of increments. Let $k \in \mathbb{N}$ then the posterior distribution for the finite set of increments $\{dF(u_1), \ldots, dF(u_k)\}$ is determined by the product of prior density and likelihood, i.e.

$$(dF(u_1), \ldots, dF(u_k))|\text{data} \propto \prod_{i=1}^{k}\pi(dF(u)) \times \text{Lik(data)}, \tag{2.7}$$

where Lik(data) is the multivariate Gaussian likelihood.

Before we continue and end this section with a large example we need to introduce an important topic known as the *aliasing effect*. This discussion is necessary before we try to use our model on some real/simulated data since we have to make everything discrete in order to do numerical approximations.

Let $Y(t)$, where $t \in [0, \infty)$, be a continuous time random process and assume that we observe $Y(t)$ at the time points $t_1, t_2, \ldots$, where $t_i = \Delta t \times i$ for $i = 0, 1, 2, \ldots$. From this construction it is clear that we will lose some information about the dependency structure since do not observe any observations that are closer than Δt in time. We know that the spectral measure is closely related to the covariance or correlation and it is not surprising that this construction will result in some loss of information about the spectral properties as well. Actually we will only lose information about the end high order frequencies of spectrum, see **?**. We will not go into to too much details about this, for a more complete discussion see **?**, see also **?** for methods regarding *alias free sampling*. But there is one result we will need in order to avoid some possible mistakes.

Again let $Y(t)$, where $t \in [0, \infty)$, be a continues time random process and suppose that we only observe $Y(t)$ at the time points $t_1, t_2, \ldots$, where $t_i = \Delta t \times i$ for $i = 0, 1, 2, \ldots$. Assume that the covariance function of the real $Y(t)$ is given by $C_0(h)$ with spectral measure $F_0(u)$, where $h \in [0, \infty)$ and $u \in (-\infty, \infty)$. Then the observed discrete process have covariance function given by

$$C_1(h) = \int_{-\pi/\Delta t}^{\pi/\Delta t} \exp(iuh\Delta t)\, dF_0(u), \quad \text{for } h = 1, 2, \ldots, \tag{2.8}$$

where

$$dF_1(u) = \sum_{j=-\infty}^{\infty} dF_0(u + 2j\pi/\Delta t), \quad \text{for } -\pi/\Delta t \leq u \leq \pi/\Delta t. \tag{2.9}$$

Our main concern regarding the aliasing effect is in the construction of a priori distributions for the spectral measure. We have already seen that in the case of a continuous time process the spectral density is defined on the interval $[0, \infty)$ and for time series the corresponding process where defined on the interval $[0, \pi]$. If a continuous time process is observed with equidistant intervals of length Δt, then from above, the random spectral measure must be specified for the interval $[0, \pi/\Delta t]$ or be redefined according to (2.9).

Remember that if we do a discrete approximation of the spectral density by dividing its domain into an equidistant grid of length π/M, then the covariance function will become periodic with period $2M$. This means that we have to choose M so large that the covariance function explains everything that we are interested in.

Example 2.5. This is a simple illustration of the basic concepts. Assume we have observed a continuous time Gaussian process $Y(t)$, for $t = 0, 0.5, 1, \ldots, 10$, where the observations $t = \{4, 4.5, 5, 5.5, 6\}$ are missing. Assume also that the series is second order stationary and has expectation zero and unknown covariance structure. From these observation we would try to estimate the posterior covariance function and predict the the unobserved values.

The true process has covariance function given by $C_0(h) = \sigma_0^2 \exp(\phi_0|h|)$, where $\sigma_0^2 = 1$ and $\phi = 0.7$. We will follow the idea from Example 2.2 and center our a priori distribution around

$C_\pi(h) = \sigma_\pi^2 \exp(-\phi_\pi |h|)$, where $\sigma_0^2 = 1$, $\phi = 0.5$, $a = 0.1$ and $\beta(u) = 2 + 5u$. Note that we have to use the result from the aliasing effect paragraph to find the right spectral density.

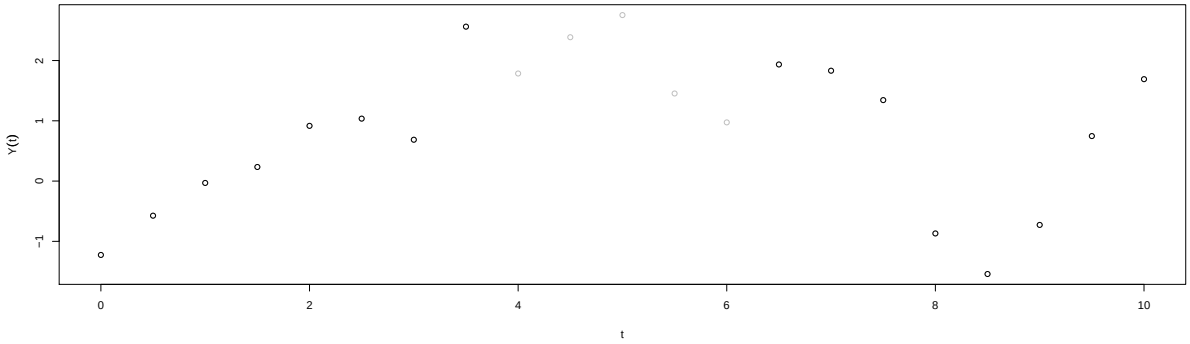

Figure 2.1: Simulated sequence $Y(t)$, for $t = 0, 0.5, \ldots, 10$, the grey points are the missing values.

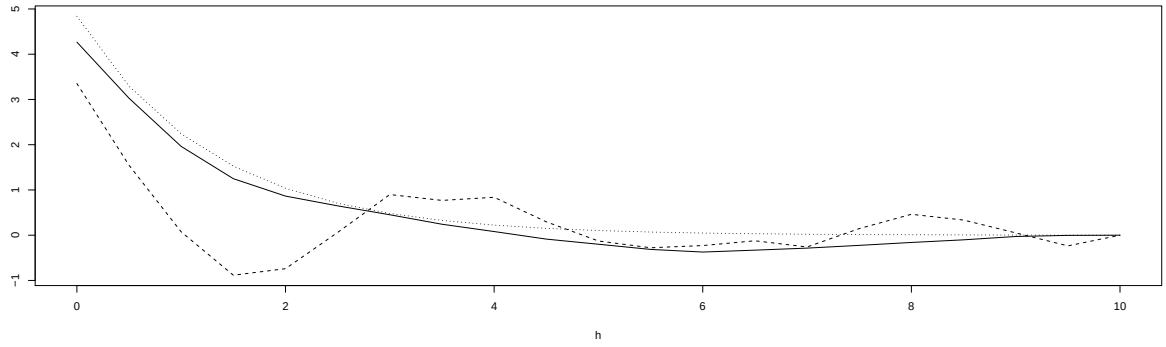

Figure 2.2: Estimated expected posterior covariance function (solid line), expected prior covariance function (dotted) and covariance estimated from data (dashed line).

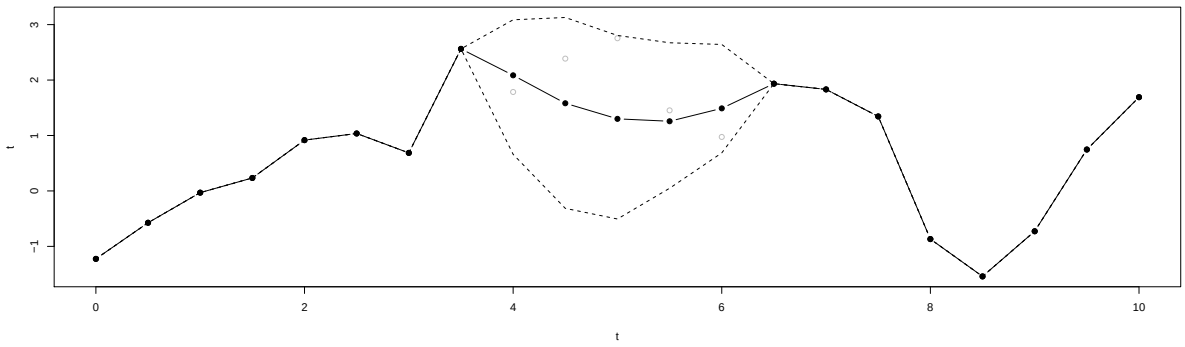

Figure 2.3: The true data sequence with predicted values together with upper and lower 0.95-bound estimated from the simulations.

3. Concluding remarks

In this chapter we have shown that the ideas from the stationary time series model can easily be extended to the spatial models defined on a general domain $D \subset \mathbb{R}^d$, where $d \geq 1$. The main goal for this chapter was to show that it was possible to do this and also to provide some simple and useful models. We have in a sense solved both problems, we have extended the framework from Chapter 2 and we have created a model setup that makes it possible to construct prior distributions centered around the exponential for nonparametric isotropic covariance functions defined for any random field on a general domain $D \subset \mathbb{R}^d$, where $d \geq 1$.

There is at least three natural unfinished problems associated with this chapter. The first is to construct a flexible model for the prior for the whole class of Matérn functions, such as the general exponential model given in Remark 2.3. At this point the only solution we have for the general Matérn class is to combine different types of deterministic and random covariance functions as explained in Chapter 2. The reason this is a natural extension is that the Matérn class is the most common choice of class of parametric isotropic covariance functions for higher order dimensional spatial analysis.

The second problem is the Metropolis-Hastings algorithm, the routine written in R works, but it is awfully slow since we do have to invert a large matrix at every step. A solution to this problem is to write some of the routine in C or Python (both can be incorporated in R) which will speed up the calculations for the algorithm.

A third topic that needs a more complete discussion than what we were able to give here, is the consequence of the aliasing effect together with the numerical approximation of the spectral measure. Suppose we observe a continuous time process at equidistant time points with distance Δ_{data}, and that we have approximated our spectral measure by using the method described in the aliasing effect paragraph. Then if we have divided the domain of the spectral density function into M piece then our covariance function will become periodic with period $2M\Delta_{\text{data}}$ and also $C(h) = C(M\Delta_{\text{data}} - h)$, for $M\Delta_{\text{data}} < h < 2M\Delta_{\text{data}}$. There exist work where people have developed *alias-free* sampling methods, the problem is that this usually involves sampling at a random rate which in turn will lead to very few equally spaced observations and less information

about the dependency structure. Another solutions is of course to choose M very large, but then we will have a problem getting our simulation routine to converge. A complete discussion of nonparametric prior distributions for the spectral measure for continuous time process should include a detailed treatment of this topic.

Note that some of the extensions suggested in the concluding remarks of the previous chapter are natural suggestions here as well and will therefore not be repeated.

APPENDIX A

Markov chain Monte Carlo simulation

In Bayesian data analysis is it not unusual that the posterior distribution become so complex that it is impractical to make inference or derive large-sample properties for the parameters under study. In such situations it is quite common to use simulation technique to make approximate inference about the posterior distribution. The two main types of simulation algorithm that are used is the *Gibbs sampler* and the *Metropolis-Hastings algorithm*, for a complete introduction to these two methods see **?** and **?**, or for a more general introduction to posterior simulation in Bayesian data analysis see Chapter 11 in **?**.

In this thesis we will only use the Metropolis-Hastings algorithm which we will introduce very briefly in next paragraph. The reason for this is that the Metropolis-Hastings algorithm in general is easy and straightforward to implement, we do not necessarily need to calculate anything, it works, but in its raw form and compared to the Gibbs sampler it might a very slow and time-consuming algorithm.

Suppose the posterior distribution is determined by $\pi(\theta|\text{data}) \propto \pi(\theta) \times \text{Lik}(\text{data}|\theta)$, where θ is a p-dimensional parameter vector. Then according to the Metropolis-Hastings algorithm we can obtain a sample $\theta^1, \theta^2, \ldots$ from the posterior distribution by the following algorithm, see **?**.

i) Draw a starting point θ^0 from a *starting distribution* or choose the starting point θ^0 based on some a priori information.

ii) Then for $t = 1, 2, \ldots$:

   (a) Draw θ^* from a *jumping distribution* or *proposal distribution* $J_t(\theta^*|\theta^{t-1})$.

   (b) Calculate the ratio of densities

$$r = \frac{\pi(\theta^*|\text{data})J_t(\theta^{t-1}|\theta^*)}{\pi(\theta^{t-1}|\text{data})J_t(\theta^*|\theta^{t-1})}.$$

   Note that in the case of a symmetric jumping distribution the jump densities cancel out of the fraction, this is also known as the *Metropolis algorithm*.

   (c) Set

$$\theta^t = \begin{cases} \theta^* & \text{with probability } \min(r, 1) \\ \theta^{t-1} & \text{otherwise.} \end{cases}$$

A common solution to determine if the routine has reached its target distribution is to start the Metropolis-Hastings algorithm at several different locations and monitor the convergence of each sequence. Hopefully all of them should converge towards the same limit, in **?** page 296-297 they suggest a measure based on a comparison of the within and between variation of each parameter.

As mentioned we will only use the Metropolis-Hastings algorithm in this thesis. The idea was to write a quite general Metropolis-Hastings sampler (the basic framework), mostly consisting of

nested loops, that were also able to perform some of the simplest strategies in MCMC simulation, such as determine the *jump variance* based on accept ratio, use *log-probability* and do *burn-in*, to make it more user friendly. But also leave all the cleverness to the user who has to define a smart jumping distribution and also an accept/reject ratio that is fast enough to compute.

Figure 0.1 illustrates the Metropolis-Hastings algorithm, where the routine makes three attempt to draw samples from a known bivariate Gaussian distribution.

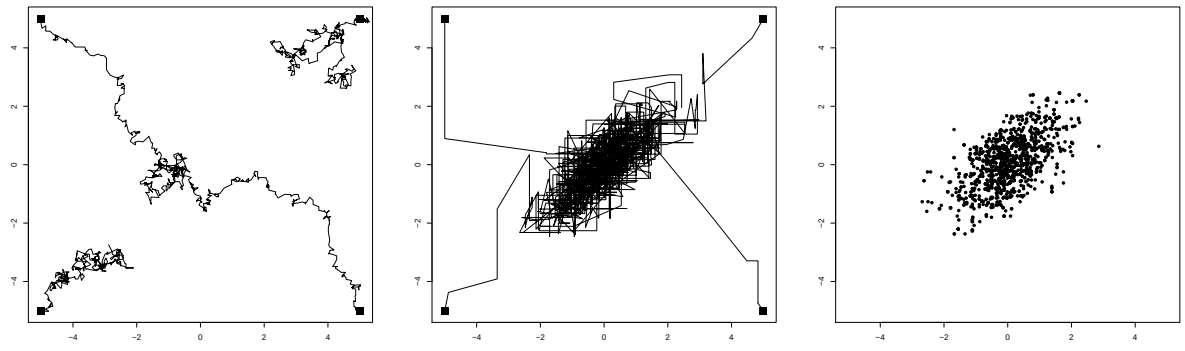

Figure 0.1: Three times four independent sequences of the Metropolis sampler for the bivariate Gaussian distribution with correlation $\rho = 0.7$. i) The jump variance is on purpose set too low. ii) The same as in i), but where the routine has searched for the jump variance that correspond to an accept ratio of 0.44. iii) The same as in ii) but with 100 burn-in steps.

The Metropolis-Hastings algorithm and the related routines are all written as functions in the software package R and the source code will not be included here, the main reason for this is that it consist of several hundred lines of code and it would not be that interesting to read. There is a little more complete explanation in Appendix C and for those who are interested in a copy of all the source and examples files used in the thesis can send me an email at `gudmunhh@student.matnat.uio.no`.

The main use of the Metropolis-Hastings algorithm will be to make approximative posterior inference about the random spectral measure. The most time consuming task in this routine will be to compute the covariance matrix, its determinant and inverse in the multivariate Gaussian likelihood. The covariance matrix is an $n \times n$ - matrix, where n is the number of observations, also every element is determined by an integral. In order to make posterior inference about the random spectral measure F we have to approximate the covariance matrix $\Sigma_n(F)$ by doing numerical integration. Let $\hat{\Sigma}_n(F)$ be the the numerical approximation of $\Sigma_n(F)$, where the elements in $\Sigma_n(F)$ is approximated by the numerical integration, i.e.

$$\Sigma_n(F)_{k,l} = 2\int_0^\pi \cos(u|k-l])dF(u) \approx 2\sum_{i=1}^{m} \cos(u_i|k-l])\Delta F(u_i) = \hat{\Sigma}_n(F)_{k,l}$$

where m is a positive integer such that the partition $0 = u_0 < u_1 < \cdots < u_{m-1} < u_m = \pi$ is dense enough to make the approximation suitably accurate. In order to speed up the algorithm we will use two small tricks, first observe that

$$\Sigma_n(F) \approx \hat{\Sigma}_n(F) = \hat{\Sigma}_n(F(u_i)) + \cdots + \hat{\Sigma}_n(F(u_m)) = \Delta F(u_1)\hat{\Sigma}_n(u_1) + \cdots + \Delta F(u_m)\hat{\Sigma}_n(u_m)$$

where $\Delta F(u_1), \ldots, \Delta F(u_m)$ are the increments and $\hat{\Sigma}_n(u_i)$ is the matrix with elements given by

$$\hat{\Sigma}_n(u_i)_{k,l} = 2\cos(u_i|k-l|).$$

If the sequence of matrixes $\hat{\Sigma}_n(u_1), \ldots, \hat{\Sigma}_n(u_m)$ are computed and stored in an array at the beginning of the routine we will save a lot of computing time since it is faster to find and obtain an element in a register than to compute the whole sequence at every iteration. Also if we always keep track of the updates we can save the covariance matrices at every iteration, its determinant and inverses, and use them to save computing time in the next iteration.

Let $\Delta F(u_1), \ldots, \Delta F(u_m)$ be the increments in the discrete approximation to the random process and that the prior distribution is the Gamma process discussed in Chapter 2 and 4. We will then use two different methods in the MCMC simulation routine, a 'safe' method where we update one increment at the time and a 'fast' method where we will update the whole sequence of increments at every iteration.

If we are updating the increments one by one, then for every iteration in the Metropolis-Hastings algorithm we will have to run through every element in the vector $(\Delta F(u_1), \ldots, \Delta F(u_m))$. This is of course very time consuming, but it is stable, easy to implement and 'safe' in the sense that not many things can go wrong. The jumping distribution will be given by

$$J_j(\Delta F(u_j)^*|\Delta F(u_j)^{t-1}) \sim \begin{cases} U[\Delta F(u_j)^{t-1} - d_i,\ \Delta F(u_j)^{t-1} + d_i] & \text{if } \Delta F(u_j)^{t-1} - d_i < 0 \\ U[0,\ \Delta F(u_j)^{t-1} + 2d_i] & \text{else} \end{cases}$$

where $U[a,b]$ is the uniform distribution on the interval $[a,b]$. The variation in the jump distribution is given by the parameter d_i and can be determined by a search for a desired accept ratio. Note that the jumping distribution cancels out of the accept ratio and we can therefore use the simpler Metropolis algorithm to obtain samples from posterior processes. In order to be precise the logarithm of the accept ratio is given by the expression

$$\begin{aligned} \log(r) = (\alpha_j - 1)\log(\Delta F(u_j)^*/\Delta F(u_j)^{t-1}) - \beta_j[\Delta F(u_j)^* - \Delta F(u_j)^{t-1}] \\ + \text{lik}\left(\Delta F(u_1), \ldots, \Delta F(u_j)^*, \ldots, \Delta F(u_m)\right) \\ - \text{lik}\left(\Delta F(u_1), \ldots, \Delta F(u_j)^{t-1}, \ldots, \Delta F(u_m)\right). \end{aligned}$$

where $\text{lik}\left(\Delta F(u_1), \ldots, \Delta F(u_m)\right)$ is the logarithm of the multivariate Gaussian likelihood and the first line that depends on the parameters α_j and β_j is what is left from the prior density.

We will also do MCMC simulation runs where we will update all the increments in the vector $(\Delta F(u_1), \ldots, \Delta F(u_m))$ at every iteration. This strategy is much faster than the method described above, but everything becomes a bit more complicated. In this case we will use the jump distribution that is in the same family as the prior distribution and the density is given by

$$J(F^*|\Delta F^{t-1}) = \prod_{i=1}^{m} \text{Ga}(b_i\, dF(u_i),\ b_i)$$

where $\text{Ga}(\alpha,\ \beta)$ is the Gamma distribution with shape parameter α and rate parameter β. The variation can be determined by the search for a desired accept ratio. In this situation we have to use the Metropolis-Hastings routine since the jump distribution is obviously not symmetric.

The logarithm of the accept ratio is now given by

$$\begin{aligned}\log(r) = \sum_{j=1}^{m} \Big[& \log(f_j(\Delta F(u_j)^*|\alpha_j,\beta_j) - \log(f_j(\Delta F(u_j)^{t-1}|\alpha_j,\beta_j) \\ & + \log(f_j(\Delta F(u_j)^{t-1}|b_j\,\Delta F(u_j)^*,b_j) - \log(f_j(\Delta F(u_j)^*|b_j\,\Delta F(u_j)^{t-1},b_j)\Big] \\ & + \text{lik}\left(\Delta F(u_1)^*,\ldots,\Delta F(u_m)^*\right) - \text{lik}\left(\Delta F(u_1)^{t-1},\ldots,\Delta F(u_m)^{t-1}\right).\end{aligned}$$

where $f_j(u|\alpha,\beta)$ is the Gamma density with shape parameter α and rate parameter β. The first line in the expression above is the prior, the second line is the correction for the asymmetry in the jumping rule and the last line is the log-likelihood of the multivariate Gaussian distribution.

We will end this short introduction with some final remarks. As we have mentioned it is very time consuming to determine the inverse of a large matrix, another problem is that the covariance matrix in the multivariate Gaussian likelihood becomes singular as a result to the numerical approximation. A solution to this is to add a small number to the diagonal, this helps the numerical calculation and will not make a critical change in the results.

A second problem is that when we ask the computer for a Gamma distributed random variable with expectation close to zero and moderate variance, the software often returns a numerical zero. This should not be possible and causes the Metropolis-Hastings algorithm to fail, see the accept ratio. A acceptable solution turns out to be to replace the numerical zero value with a small positive number, this does not seem to alter the solution in any way.

Also we have to consider the aliasing effect, see Section 2 in Chapter 4, the point is that there is a connection between the numerical approximation of the spectral measure and the distance between the observed sample. Another numerical problem is that the covariance function will become periodic if the partition of the domain of the power spectrum is not dense enough. These problems will also be pointed out in the examples.

At last we will of course in the case of a stationary Gaussian time series, where the number of observations is large enough, use the approximations to the full log-likelihood we refer to as the principal part. This approximation does not involve any matrixes and is therefore several times faster and safer to use than the routine based on the full multivariate Gaussian likelihood.

APPENDIX B

Calculations

1. A density based on the Gamma and Inverse-Gamma density

Let Y be a random variable distributed according to the Inverse-Gamma distribution with shape parameter ν and scale parameter γ , then Y has density function given by

$$f(y) = \frac{\gamma^\nu}{\Gamma(\nu)} y^{-\nu-1} \exp(-\gamma/y), \quad \text{for } y > 0. \tag{1.1}$$

The expectation of Y is given by $E[Y] = \gamma/(\nu - 1)$ and if $\nu > 2$ then the variance of Y exists and is determine by $\mathrm{Var}(Y) = \gamma^2/[(\nu-1)^2(\nu-2)]$.

We will now derive a density function that is proportional to the product of an Inverse-Gamma density and a Gamma density. Let Z be the random variable with the density described above and assume that the Gamma density has has shape parameter α' and rate parameter β, and that the Inverse-Gamma density is as described in equation (1.1). Then the random variable Z has density function given by

$$f_Z(z) = \frac{\gamma^{\alpha/2}}{2\beta^{\alpha/2} K_\alpha(2\sqrt{\beta\gamma})} z^{-\alpha-1} \exp\left[-\beta z - \frac{\gamma}{z} \right] \tag{1.2}$$

where $K_\kappa(t)$ is the modified Bessel function of the second kind and $\alpha = \nu - \alpha' + 1$. This follows directly from the expression of the moment generating function of a Inverse-Gamma distributed random variable, i.e.

$$E[\exp(tY)] = \int_0^\infty \exp(ty) f(y)\, dy = \frac{2(-\gamma t)^{\nu/2}}{\Gamma(\nu)} K_\nu(\sqrt{-4\gamma t}), \quad \text{for } t \le 0.$$

It is now easy to find expressions for the expectation and the variance of Z,

$$E[Z] = \left(\frac{\gamma}{\beta}\right)^{1/2} \frac{K_{\alpha-1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} \quad \text{and} \quad \mathrm{Var}(Z) = \frac{\gamma}{\beta} \frac{K_{\alpha-2}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} - \frac{\gamma}{\beta} \left(\frac{K_{\alpha-1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} \right)^2. \tag{1.3}$$

In order to make the expressions more informative we will derive some approximations, but first we need some properties of the modified Bessel function of the second kind. From **?** we have that

i) $K_\kappa(t) = K_{-\kappa}(t)$,
ii) $K_{\kappa+1}(t) = K_{\kappa-1}(t) + \frac{2\kappa}{t} K_\kappa(t)$,
iii) $K_\kappa(t) \approx \frac{\Gamma(\kappa)}{2} \left(\frac{t}{2}\right)^{-\kappa}$, if $\kappa > 0$ and $0 < t \ll \sqrt{\kappa+1}$.

If the difference $\sqrt{\kappa+1} - t$ become large or $t \to 0$ then we can in a sense replace the approximation in property iii) with a limit. From these properties we can now derive limit versions of the

expectation and variance. Assume that $\nu > \alpha$, we can find limit expressions for the expectation and variance from 1.3 as

$$E[Z] \approx \left(\frac{\gamma}{\beta}\right)^{1/2} \frac{\Gamma(\alpha-1)(\beta\gamma)^{-\alpha/2+1/2}}{\Gamma(\alpha)(\beta\gamma)^{-\alpha/2}} = \frac{\gamma}{\alpha} = \frac{\gamma}{\nu-\alpha'+1}, \text{ if } 0 < \sqrt{4\beta\gamma} \ll \sqrt{\alpha}.$$

and

$$\begin{aligned}\mathrm{Var}(Z) &\approx \frac{\gamma}{\beta}\left\{\frac{\Gamma(\alpha-2)(\beta\gamma)^{-\alpha/2+1}}{\Gamma(\alpha)(\beta\gamma)^{-\alpha/2}} - \left[\frac{\Gamma(\alpha-1)(\beta\gamma)^{-\alpha/2+1/2}}{\Gamma(\alpha)(\beta\gamma)^{-\alpha/2}}\right]^2\right\} \\ &= \frac{\gamma^2}{\alpha^2(\alpha-1)} = \frac{\gamma^2}{(\nu-\alpha'+1)^2(\nu-\alpha')}\end{aligned}$$

if $0 < \sqrt{4\beta\gamma} \ll \sqrt{\alpha-2}$.

2. Inverse-Gamma process

We will in this section argue that the Inverse-Gamma process, i.e. the independent positive Inverse-Gamma distributed increment process, does not exist, by showing that a very simple type of Inverse-Gamma process becomes a nonrandom constant process. Let Y_t be a Inverse-Gamma process define by

$$Y_t = \sum_{j=1}^{m} Y_j$$

where $Y_j \sim$ Inv-Gamma(α, β) and $\alpha = a/m + 2$ and $\beta = b/m$, then

$$E[Y_t] = \frac{b}{a/m+1} \to b \text{ and } \mathrm{Var}(Y_t) = \frac{b^2}{(a/m+1)^2 a} \to \frac{b^2}{a}.$$

In order to show that the Inverse-Gamma process exist we need to show that has a Lévy representation, and the first step towards this is to show that Y_t has Laplace transformation of the form

$$E[\exp(-\theta Y_t)] = E[\exp(-\theta Y_j)]^m = \exp(-M(\theta)), \text{ as } m \to \infty. \tag{2.1}$$

The easiest way to show (2.1) is to work with the core $\mathcal{L}(f_j(y)) = E[\exp(-\theta Y_j)]$ and show that it is possible to write the expression (2.1) as

$$E[\exp(-\theta Y_t)]^m = \left[1 - M(\theta)/m + O(1/m^2)\right]^m \to \exp(-M(\theta)), \text{ as } m \to \infty.$$

In this case the process do not exist, and we will prove this by showing that $E[\exp(-\theta Y_t)]^m \to \exp(-\theta)$, which is the Laplace transformation for a plain nonrandom constant.

Let Y_j follow an Inverse-Gamma distribution, then Y_j has density $f_j(u)$ given by (1.1) with parameters $\alpha = a/m + 2$ and $\beta = b/m$. The expectation and variance of Y_j do then exist and we have that the Laplace transformation of Y_j is determined by

$$\begin{aligned}\mathcal{L}(f_j(y)) = E[\exp(-\theta Y_j)] &= \frac{2(\beta\theta)^{\alpha/2}}{\Gamma(\alpha)} K_\alpha(\sqrt{4\beta\theta}) \\ &= \frac{2(\theta b/m)^{a/2m+1}}{\Gamma(a/2m+2)} K_{a/m+2}(2(\theta b/m)^{1/2}), \text{ for } \theta \geq 0\end{aligned}$$

where $K_\nu(t)$ is again the modified Bessel function of the second kind. Because of the Bessel function it is not easy to verify the limit of $E[\exp(-\theta Y_j)]^m$, it is therefore necessary to simplify the expression. From 11.3.27 in **?** we have that

$$\int_0^t u^v K_{\nu-1}(u)\,du = -t^\nu K_\nu(t) + 2^{\nu-1}\Gamma(\nu) \quad \Rightarrow \quad K_\nu(t) = 2^{\nu-1}t^{-\nu}\Gamma(\nu) - t^{-\nu}\int_0^t u^v K_{\nu-1}(u)\,du.$$

We will also need a series representation for the modified Bessel function of the second kind given by

$$K_\nu(t) = \frac{\pi}{2\sin(\pi\nu)}\left(\sum_{k=0}^{\infty}\frac{1}{\Gamma(k-\nu+1)k!}(t/2)^{2k-\nu} - \sum_{k=0}^{\infty}\frac{1}{\Gamma(k+\nu+1)k!}(t/2)^{2k+\nu}\right)$$

for $\nu \notin \mathbb{Z}$. Let $z = \sqrt{4\beta\theta}$ and $\mu = a/m + 2$, then by using 11.3.27 from **?** we have that

$$\begin{aligned}\mathcal{L}(f_j(y)) = \frac{2(z/2)^\mu}{\Gamma(\mu)}K_\mu(z) &= 1 - \frac{2^{1-\mu}}{\Gamma(\mu)}\int_0^z u^\mu K_{\mu-1}(u)\,du \\ &= 1 - \frac{2^{-(a/m+1)}}{\Gamma(a/m+2)}\int_0^{2(\theta b/m)^{1/2}} u^{a/m+2}K_{a/m+1}(u)\,du\end{aligned}$$

and from the series representation we find further that

$$\begin{aligned}\mathcal{L}(f_j(y)) &= 1 - \frac{2^{1-\mu}}{\Gamma(\mu)}\int_0^z u^\mu K_{\mu-1}(u)\,du \\ &= 1 - \frac{2^{1-\mu}}{\Gamma(\mu)}\int_0^z u^\mu \frac{\pi}{2\sin(\pi\nu)}\left(\sum_{k=0}^{\infty}\frac{1}{\Gamma(k-\nu+1)k!}(u/2)^{2k-\nu}\right. \\ &\qquad \left. - \sum_{k=0}^{\infty}\frac{1}{\Gamma(k+\nu+1)k!}(u/2)^{2k+\nu}\right)du\end{aligned}$$

where $\nu = \mu - 1$. If u is now integrated out we find that

$$\begin{aligned}\mathcal{L}(f_j(y)) &= 1 - \frac{2^{1-\mu}}{\Gamma(\mu)}\frac{\pi}{2\sin(\pi\nu)}\left(\sum_{k=0}^{\infty}\frac{1}{\Gamma(k-\nu+1)k!}(1/2)^{2k-\nu}\int_0^z u^{2k-\nu+\mu}\,du\right. \\ &\qquad \left. - \sum_{k=0}^{\infty}\frac{1}{\Gamma(k+\nu+1)k!}(1/2)^{2k+\nu}\int_0^z u^{2k+\nu+\mu}\,du\right) \\ &= 1 - \frac{2^{1-\mu}}{\Gamma(\mu)}\frac{\pi}{2\sin(\pi\nu)}\left(\sum_{k=0}^{\infty}\frac{1}{\Gamma(k-\nu+1)k!}(1/2)^{2k-\nu}\frac{z^{2k-\nu+\mu+1}}{2k-\nu+\mu+1}\right. \\ &\qquad \left. - \sum_{k=0}^{\infty}\frac{1}{\Gamma(k+\nu+1)k!}(1/2)^{2k+\nu}\frac{z^{2k+\nu+\mu+1}}{2k+\nu+\mu+1}\right) \\ &= 1 - \frac{2^{1-\mu}}{\Gamma(\mu)}\frac{\pi z^{\mu+1}}{2\sin(\pi\nu)}\left(\sum_{k=0}^{\infty}\frac{1}{\Gamma(k-\nu+1)k!}(z/2)^{2k-\nu}\frac{1}{2k-\nu+\mu+1}\right. \\ &\qquad \left. - \sum_{k=0}^{\infty}\frac{1}{\Gamma(k+\nu+1)k!}(z/2)^{2k+\nu}\frac{1}{2k+\nu+\mu+1}\right).\end{aligned}$$

Since $z = \theta b/m$, $\mu = a/m + 2$ and $\nu = a/m + 1$ we are now able to simplify the expression above and obtain the following expression for the Laplace transformation of Y_j,

$$\begin{aligned}\mathcal{L}(f_j(y)) &= 1 - \frac{2^{1-\mu}}{\Gamma(\mu)} \frac{\pi z^{\mu+1}}{2\sin(\pi\nu)} \left(\frac{1}{\Gamma(1-\nu)} (z/2)^{-\nu} \frac{1}{1-\nu+\mu} \right) + O(1/m^2) \\ &= 1 - \frac{1}{m} \frac{\pi}{\Gamma(-a/m)\sin(\pi(1+a/m))} \frac{\theta b}{\Gamma(2+a/m)} + O(1/m^2).\end{aligned}$$

The final thing we need is a formula from **?** also known as *Euler's reflection formula* which stats that

$$\frac{\pi}{\Gamma(-a/m)\sin(\pi(1+a/m))} = \Gamma(1+a/m).$$

It now follows from the above that since $\Gamma(1+a/m) \to 1$, $\Gamma(2+a/m) \to 1$ as $m \to \infty$ and by Euler's reflection formula that

$$\begin{aligned}\lim_{m\to\infty} \mathcal{L}(f_j(y))^m &= \lim_{m\to\infty} E[\exp(-\theta Y)]^m \\ &= \lim_{m\to\infty} \left[1 - \frac{1}{m} \frac{\pi}{\Gamma(a/m)\sin(\pi(1+a/m))} \frac{\theta b}{\Gamma(2+a/m)} + O(1/m^2) \right]^m = \exp(-\theta b).\end{aligned}$$

3. Properties of the Dirichlet distribution

Let $(Y_1, \ldots, Y_k)$ be distributed according to a Dirichlet distribution with parameters $(a_1, \ldots, a_k)$, we will write $(Y_1, \ldots, Y_k) \sim \text{Dir}(a_1, \ldots, a_k)$. Let $a_0 = \sum_{i=1}^k a_i$ then for every i where $i = 1, \ldots, k$, we have that $E[Y_i] = a_i/a_0$ and $\text{Var}(Y_i) = [a_i(a_0 - a_i)]/[a_0^2(a_0+1)]$, for $i, j = 1, \ldots, p$ and $i \neq j$ we have $\text{Cov}(Y_i, Y_j) = -a_i a_j/[a_0^2(\alpha_0+1)]$. Now the expectation and variance for the sum $b_1 Y_1 + \cdots + b_k Y_k$, for any set of real numbers $b_1, \ldots b_k$ is given by

$$E[\sum_{i=1}^k b_i Y_i] = \sum_{i=1}^k b_i E[Y_i] = \frac{a_1 b_1 + \ldots + a_k b_k}{a_0}$$

and

$$\begin{aligned}&\text{Var}(\sum_{i=1}^k b_i Y_i) \\ &\quad = \sum_{i=1}^k \sum_{j=1}^k b_i b_j \,\text{Cov}(Y_i,\ Y_j) \\ &\quad = \sum_{i=1}^k b_i b_i \,\text{Var}(X_i) + \sum_{i\neq j} b_i b_j \,\text{Cov}(Y_i,\ Y_j) \\ &\quad = \sum_{i=1}^k b_i b_i \frac{a_i(a_0 - a_i)}{a_0(a_0+1)} - \sum_{i\neq j} b_i b_j \frac{a_i a_j}{a_0^2(a_0+1)} = \frac{1}{a_0(a_0+1)} \left[\sum_{i=1}^k b_i^2 a_i - \sum_{i=1}^k b_i a_i \sum_{j=1}^k b_j a_j \right].\end{aligned}$$

APPENDIX C

R

```
# This is just a short overview of some of the main functions
# written in R, for a complete explanation see the source files.

# function.R
#

# Functions:

# Calculate the Covariance Function From the Power Spectrum
#
covariance.f <- function(h, power) {}

# Estimate the Covariance From Data
#
covariance.e <- function(t, data, mu = NULL) {}

# Estimate the Covariance Based on the Periodogram
#
covariance.e.p <- function(u, data, time) {}

# Calculate the Covariance Matrix From a Covariance function or
# power spectrum
#
covariance.m <- function(u, cov.func = NULL, power.func = NULL) {}

# Calculate the Exact, Whittle or Principal Part of the log-likelihood
#
likelihood <- function(time, data, power.s = NULL, spectral.m = NULL,
        method = c("Dzhaparidze", "Whttle", "Gaussian")) {}

# Calculate the Smoothed Periodogram
#
periodogram.s <- function(u, data, time, m) {}

# Calculate the Periodogram
#
periodogram <- function(u, data, time) {}
```

```
# MCMC:

# Calculate the Covariance Function for a Single Spectral Measure From a
# Metropolis-Hastings Routine
#
covariance.mcmc.f <- function(h, u, v) {

# Calculate the Covariance Matrix for a Single Spectral Measure From a
# Metropolis-Hastings Routine
#
covariance.mcmc.m <- function(t, u, v) {

# Returns Mean, Variance and Upper and Lower 0.95-Bound for any Type of Data
# From a  Metropolis-Hastings routine
#
mcmc.summary <- function(mcmc, conf.int = 0.95, cut = 0) {}

# Returns Mean, Variance and Upper and Lower 0.95-Bound for the Covariance
# Function From a  Metropolis-Hastings routine
#
mcmc.covariance.summary <- function(mcmc, lag, conf.int = 0.95,
        cut = 0, cov.d = NULL) {}


# Prediction:

# Returns the Estimated Densities and Mean (for Kriging) From a Single
# New Observation
#
prediction.time <- function(mcmc, new, data, time, nr = 100) {}

# Returns the Estimated Densities and Mean (for Kriging) From a Vector
# of New Observations
#
prediction.times <- function(mcmc, new, data, time, nr = 100, conf.int = 0.95) {}



# plot.R
#

# Plot the Spectral Measure From a Metropolis-Hastings Routine
#
mcmc.measure.plot <- function(mcmc, conf.int = 0.95, gamma.para = NULL, cut = 0) {}

# Plot the Covariance Function From a Metropolis-Hastings routine
#
mcmc.covariance.plot <- function(mcmc, lag, conf.int = 0.95, cut = 0,
```

```
        cov.true = NULL, cov.d = NULL) {}



# prior.R
#
# Jump Distributions and Accept Ratio for the Metropolis-Hastings Algorithm
# Needed in Order to Sample From the Prior Distribution



# post.R
#
# Jump Distributions and Accept Ratio for the Metropolis-Hastings Algorithm
# Needed in Order to Sample From the Posterior Distribution



# metropolis.0.1.0.R
#
# The Metropolis-Hastings algorithm and a Test for Covergence, See
# the Source File for Details
```