Aalborg University
Department of Mathematical Sciences

MSc Thesis

Central limit theorems for weakly dependent stochastic processes

– An application within communication technology

June 2007

Ege Rubak

Department of Mathematical Sciences, Aalborg University,
Fredrik Bajers Vej 7 G, 9220 Aalborg East, Denmark

AALBORG UNIVERSITET

Department of Mathematical Sciences
Fredrik Bajers Vej 7G
9220 Aalborg Ø
Telephone: +45 96 35 88 02
Fax: +45 98 15 81 29
http://www.math.aau.dk

Title:
Central limit theorems for weakly dependent stochastic processes
– An application within communication technology

Project period:
Feburary 1st - June 18th, 2007

Author:
Ege Rubak

Supervisors:
Martin Bøgsted Hansen
Persefoni Kyritsi

Semester:
MAT6, spring 2007

Number of copies: 10

Report – number of pages: 78

Deadline: June 18th, 2007

Abstract

This thesis investigates the performance of a wireless communication system using a so-called equalizer. The specific environment in which the wireless communication system is used is called the communication channel. Since wireless systems usually are mobile the channel is a random quantity changing according to the dynamic environment. The channel distorts the transmitted signals, and it is necessary to compensate for this distortion in the design of the system. One way of reducing the effect of the distortion is to use an equalizer in the receiver, which attempts to reverse the effect of the channel and thereby restore the original signal. As a measure of performance of the system the minimum means square error (MMSE) is used. The MMSE value is determined for each channel making it a random quantity. It is interesting to describe the distribution of the MMSE for a given channel model, since it then would be possible to give confidence bands on the performance of the system. The MMSE depends on the bandwidth B used by the communication system. In modern ultra wide band (UWB) systems very large bandwidths are used making asymptotic results for increasing bandwidth interesting. The problem of an asymptotic characterization of the MMSE is the starting point of this thesis.

First a mathematical model for the communication system is presented. In this model two different equalizers are introduced: The infinite and the finite equalizer, which give rise to two different MMSEs denoted respectively MMSE_∞ and MMSE_N. The communication channel can be described via the channel response or equivalently the Fourier transform of this response called the transfer function H. Due to the random nature of the channel this is a stochastic process, and it turns out that this process determines MMSE_∞ uniquely as

$$\mathrm{MMSE}_\infty = \frac{1}{B}\int_0^B \frac{\sigma_n^2}{|H(f)|^2 + \frac{1}{\rho}} df,$$

where σ_n^2 is the noise power, ρ is the signal to noise ratio, and B denotes the bandwidth.

A brief simulation study indicates a central limit theorem (CLT) holds for MMSE_∞, and the rest of the thesis is devoted to the investigation of this. As a background for this investigation a wide range of topics within the field of stochastic processes is presented. Especially a detailed proof of a CLT for continuous time stochastic processes is given. This establishes that under suitable conditions the normalized integral for a stochastic process is asymptotically Gaussian, when the integration limit grows to infinity. For this to hold a key assumption is weak dependence of the stochastic process. The concept of weak dependence is defined via mixing properties, and this topic is also studied. In this classical central limit theory the process is not allowed to depend on the integration limit, which the process H does. To deal with this problem it is necessary to consider the limiting behavior of H as B grows. A powerful tool for studying asymptotics of stochastic processes is weak convergence in metric spaces, and a brief introduction to this topic is provided.

Finally the properties of MMSE_∞, MMSE_N, and H are studied, and a weak limit of H is determined. Furthermore it is proven that under suitable conditions a CLT holds for MMSE_∞, and finally it is proven that MMSE_N approximates MMSE_∞ for $N \to \infty$, which is used to give a condition ensuring the CLT behavior is inherited by MMSE_N.

Acknowledgements

This project was jointly proposed by Persefoni Kyritsi, Department of Electronic Systems Aalborg University and Martin Bøgsted Hansen, Department of Mathematical Sciences, Aalborg University. They have served as supervisors on the project and I will like to thank them both for their thorough supervision and guidance during the project period.

I will also like to thank the group at Standford University working with Persefoni Kyritsi on the paper Pereira et al. (2006). They have granted me access to their unpublished work, which has provided the basis of Chapter 1, where the communication system model is presented. Furthermore a number of conjectures in their work served as a guideline for the present work.

Ege Rubak
Aalborg, 2007

Contents

CHAPTER 1
Introduction

1.1 Introduction

In recent years ultra wide band (UWB) techniques have gained considerable interest as a novel approach to high speed wireless communications both over short range and for longer range systems as e.g. positioning systems. One of the effects of the large bandwidth in UWB communication systems is the possibility of using short time intervals for symbol transmission and of thereby increasing the data rate. An estimate of the bandwidth B used by a communication system is $B = 1/T$, where T is the symbol time of the system. Therefore a bandwidth of e.g. 1 GHz corresponds to a symbol time of $T = 1$ ns. When the transmitting antenna in a wireless communication system transmits a given sequence of symbols modulated into radio waves it will be reflected on various objects present in the propagation medium. This leads to several different propagation paths between the transmitter and receiver as illustrated in Figure 1.1. The waves always travel at the constant speed of light, which means that the signal at the receiver will be a superposition of delayed versions of the symbol sequence corresponding to different propagation paths. This type of signal corruption is called inter symbol interference (ISI) since previously transmitted symbols interfere with the new symbols at the receiver due to the various propagation delays.

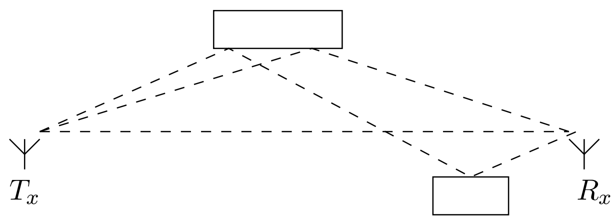

Figure 1.1: Schematic representation of different propagation paths from the transmitter antenna T_x to the receiver antenna R_x leading to ISI.

A common way of reducing the effect of the ISI is to include an equalizer in the receiver part of the communication system. There are several types of both linear and non-linear equalizers that are used today, but here we will only treat linear equalizers and specifically we will study characteristics of the minimum mean square error (MMSE) linear equalizer.

1.1.1 System model

It is assumed that we wish to transmit a sequence of information symbols $\boldsymbol{x} = \{x_i\}_{-\infty}^{\infty}$ over a communication channel. We assume that $x_i \in S$, where $S = \{s_1, \ldots, s_n\}$ and

$s_i \in \mathbb{C}$ for $i = 1, \ldots, n$. The information symbols are unknown a priori and they are assumed to be independent identically distributed (iid) on S. The task of modulating the information symbols into continuous time signals and the demodulation of these signals is the topic of digital modulation and it will not be treated here. A text book introduction to various digital modulation techniques can be found in Proakis (2001) or in the lecture notes Cioffi (2005). One of the main results is that the continuous time communication system can be modeled by an equivalent discrete time system, which often will simplify the analysis of the system. The transmission time of each discrete information symbol is called the symbol time T and we will assume throughout the thesis that the transmission uses the bandwidth $B = 1/T$.

We assume the equivalent discrete time system model specifically is as illustrated in Figure 1.2. Here it is assumed that the effect of the communication channel can be divided into two separate parts. Firstly the channel introduces ISI modeled by a convolution with the so called channel response $\boldsymbol{h}_L^B = (h_0^B, \ldots, h_{L-1}^B)$, $h_l^B \in \mathbb{C}$, such that given $\boldsymbol{h}_L^B$ and the input $\boldsymbol{x}$ the channel output $\boldsymbol{y} = \{y_i\}_{-\infty}^{\infty}$, due to this effect alone, would be given by

$$\boldsymbol{y} = \boldsymbol{h}_L^B * \boldsymbol{x}.$$

The output is thus explicitly given by

$$y_i = \sum_{l=0}^{L-1} h_l^B x_{i-l}, \quad i \in \mathbb{Z}.$$

The second part of the channel is assumed to add white complex Gaussian noise to the signal, and it is thus called an Additive White Gaussian Noise (AWGN) channel.

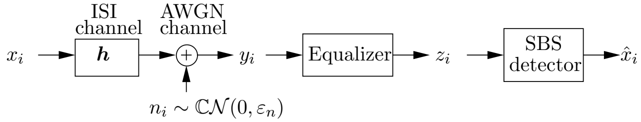

Figure 1.2: Discrete time equivalent model of a communication system with symbol time T operating over the bandwidth $B = 1/T$.

The final output is then given by

$$\boldsymbol{y} = \boldsymbol{h}_L^B * \boldsymbol{x} + \boldsymbol{n},$$

where the elements of the sequence $\boldsymbol{n} = \{n_i\}_{-\infty}^{\infty}$ are assumed to be iid with $n_i \sim \mathbb{C}\mathcal{N}(0, \sigma_n^2)$. For each i, y_i is given by

$$y_i = \sum_{l=0}^{L-1} h_l^B x_{i-l} + n_i, \quad i \in \mathbb{Z}. \tag{1.1}$$

It is clear from this model that the output at time j only depends on the input values in the time interval $(-\infty, j]$, and we say that the system is causal. It is straight forward to extend the model to non-causal systems where we have $h_l^B \neq 0$ for some $l < 0$.

At the receiver an equalizer is used to reduce the effect of the ISI, which is done by convolving the channel output $\boldsymbol{y}$ with the equalizer response $\boldsymbol{w} = \{w_i\}_0^\infty$, $w_i \in \mathbb{C}$ to obtain the equalizer output $\boldsymbol{z} = \{z_i\}_{-\infty}^\infty$ given by

$$z_i = \sum_{j=0}^{\infty} w_j y_{i-j}, \quad i \in \mathbb{Z}.$$

Finally each $z_i \in \mathbb{C}$ is mapped to an estimate $\hat{x}_i \in S$ by a symbol-by-symbol (SBS) detector. In order to achieve good performance of the system the equalizer response $\boldsymbol{w}$ must be chosen according to some criterion based on knowledge of the channel response $\boldsymbol{h}_L^B$. Here we will focus on the MMSE linear equalizer, which we elaborate on in the following.

1.1.2 MMSE equalization

This short introduction to MMSE equalization is based on Cioffi (2005) and a more thorough exposition can be found there. We assume the channel response $\boldsymbol{h}_L^B$ is known at the receiver such that the equalizer response $\boldsymbol{w}$ is chosen given $\boldsymbol{h}_L^B$. Furthermore we always assume the mean noise power σ_n^2 (the variance of the noise distribution) to be known along with the signal power, which we denote σ_x^2. Often we are interested in the relative size of the signal compared to the noise called the signal to noise ratio (SNR), which is defined by

$$\rho = \frac{\sigma_x^2}{\sigma_n^2}. \tag{1.2}$$

With this information available at the receiver we need a criterion of optimality to choose the equalizer response by, and in this thesis we will use the MMSE defined below.

Definition 1.1.1. Let a communication system as shown in Figure 1.2 use bandwidth B to communicate over a channel of length L. If an equalizer with response $\boldsymbol{w} = \{w_i\}_0^\infty$ is used, the mean square error, $\mathrm{MSE}_\infty^{L,B}$, and the minimum mean square error, $\mathrm{MMSE}_\infty^{L,B}$, respectively are given by

$$\mathrm{MSE}_\infty^{L,B} = \mathbb{E}(\|x_i - z_i\|^2),$$

and

$$\mathrm{MMSE}_\infty^{L,B} = \min_{\boldsymbol{w}} \mathbb{E}(\|x_i - z_i\|^2),$$

where z_i is the equalizer output and x_i is the input symbol. □

The subscript ∞ refers to the response of the equalizer having infinite length, which in another setting discussed later will be finite. Since both the input process $\boldsymbol{x}$ and the noise process $\boldsymbol{n}$ are assumed stationary the choice of index i in Definition 1.1.1 does not matter and the definition is consistent.

Intuitively it is clear that this is a sensible optimality criterion since a small MSE means the equalizer output is close to the actual transmitted information symbols suggesting that fewer estimation errors will be made. Furthermore it is intuitively reasonable that the MMSE value can be used as a parameter describing the performance of a communication system. A system with a small MMSE is preferable compared to a system with a high MMSE.

In Cioffi (2005) it is shown that the value of the MMSE can be calculated by

$$\mathrm{MMSE}_{\infty}^{L,B} = \frac{1}{B}\int_0^B \frac{\sigma_n^2}{|H^{L,B}(f)|^2 + \frac{1}{\rho}} df, \tag{1.3}$$

where ρ is the SNR defined in (1.2), and $H^{L,B}$ is the discrete time Fourier transform of $\boldsymbol{h}_L^B$ given by

$$H^{L,B}(f) = \sum_{l=0}^{L-1} h_l^B \exp\left(-2\pi i f \tfrac{l}{B}\right), \qquad f \in \mathbb{R}. \tag{1.4}$$

This is known as the transfer function or frequency response of the channel and it is used to describe the properties of a communication channel in the frequency domain.

The deterministic expression for $\mathrm{MMSE}_{\infty}^{L,B}$ in (1.3) relies on the transfer function $H^{L,B}$ being known, but as we will discuss in the following section the channel response $\boldsymbol{h}_L^B$ and thereby also $H^{L,B}$ is actually modeled by a stochastic process.

1.1.3 Channel model

The channel response is determined by the physical nature of the channel, which is dynamically changing for a wireless channel and therefore the channel response $\boldsymbol{h}_L^B$ is a stochastic process. We will however assume that it is constant for each period of use. It is assumed the mean power of the channel over time can be described by a non-negative continuous time function p called the power delay profile (PDP). We will assume the channel is causal with normalized power and therefore:

(i) $p(t) = 0$ for $t < 0$

(ii) $\int_0^\infty p(t)dt = 1$,

which means that p is a probability density function (pdf). The PDP is assumed to determine the mean power of the discrete time channel response by

$$\mathbb{E}(|h_l^B|^2) = \nu_l^B = \int_{lT}^{(l+1)T} p(t)dt = \int_{\frac{l}{B}}^{\frac{l+1}{B}} p(t)dt. \tag{1.5}$$

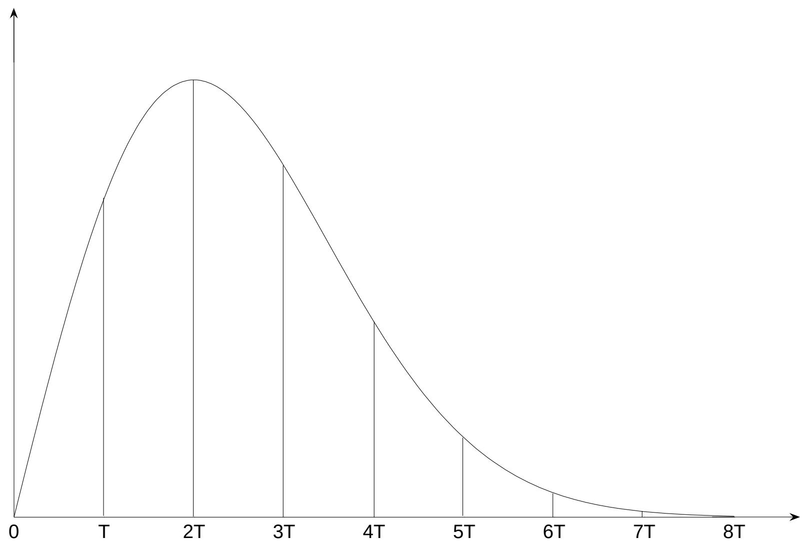

Figure 1.3: Example of a power delay profile. The variance of $h_0^B, \ldots, h_{L-1}^B$ is determined by the integral of p over the corresponding sampling interval.

Furthermore we assume $(h_0^B, \ldots, h_{L-1}^B)$ are obtained from iid complex valued stochastic variables $X_0, \ldots, X_{L-1}$ with $\mathbb{E}(X_l) = 0$ and $\text{Var}(X_l) = 1$ as $h_l^B = \sigma_l^B X_l$, where $\sigma_l^B = \sqrt{\nu_l^B}$. The distribution of X_l is called the channel distribution and we will often assume this to be the complex Gaussian distribution, which is common in the literature (Jakes, 1994).

Figure 1.3 is a schematic representation of p and the relation to the mean power of $h_0^B, \ldots, h_{L-1}^B$. The PDP is an idealized model of how the power in the channel is distributed. If we were to use measuring equipment with unlimited sensitivity we would on average observe an amount of power at all times. This limit corresponds to the channel length $L = \infty$, and it is seen in this case the mean total power of the channel response is 1, since

$$\begin{aligned}\mathbb{E}\left(\sum_{l=0}^{\infty}|h_l^B|^2\right) &= \sum_{l=0}^{\infty}\mathbb{E}(|h_l^B|^2)\\ &= \sum_{l=0}^{\infty}\int_{\frac{l}{B}}^{\frac{l+1}{B}} p(t)dt\\ &= \int_0^{\infty} p(t)dt = 1.\end{aligned}$$

It is also remarked that if B is increased it will usually be natural to increase L simultaneously. It is clear that if L is fixed, we integrate over a smaller and smaller interval when B is increased, implying that we incorporate less and less of the channel power in the model. If L/B is kept constant during the increase of L, B we keep the channel power constant, which can be interpreted physically as fixing the sensitivity of the measuring equipment. Another type of asymptotics that is reasonable to consider is when the ratio L/B increases as L and B are increased. This setting corresponds to both using larger bandwidth and more sensitive equipment, and it has the mathematically attractive property that in the limit we incorporate all the power of the channel in the model.

1.1.4 Finite MMSE equalization

While the previous setup is nice from a theoretical point of view there are some issues regarding the practical implementation that need to be considered. It is very difficult to implement an equalizer with an infinite response and in practice the equalizer will have a response of finite length N. In this section we will describe finite MMSE equalization, which for a fixed N aims to find the choice of $\boldsymbol{w}_N = (w_1, \ldots, w_{N-1})$ that minimizes the MSE of the system.

We define the vector $\boldsymbol{y}_i = (y_i, \ldots, y_{i-(N-1)})^\top$ which is the equalizer input for the finite equalizer with response of length N (often called the length of the equalizer). A standard method to calculate a convolution is a matrix-vector product, which we will do in the following. Using (1.1) we have

$$\begin{aligned}\boldsymbol{y}_i &= \begin{bmatrix} h_0^B & \cdots & h_{L-1}^B & & 0 \\ 0 & \ddots & \ddots & \ddots & \\ & & h_0^B & \cdots & h_{L-1}^B \end{bmatrix}\begin{pmatrix} x_i \\ \vdots \\ x_{i-(L+N-2)}\end{pmatrix} + \begin{pmatrix} n_i \\ \vdots \\ n_{i-(N-1)}\end{pmatrix} \\ &= H\boldsymbol{x}_i + \boldsymbol{n}_i,\end{aligned} \tag{1.6}$$

where we have defined the $N\times(L+N-1)$ dimensional channel matrix H, the input vector $\boldsymbol{x}_i$ and the noise vector $\boldsymbol{n}_i$. The equalizer output at time i is then $z_i = \boldsymbol{w}_N\boldsymbol{y}_i$. From (1.6) it is seen that z_i contains information of the symbols $x_{i-(L+N-2)}, \ldots, x_i$

and we could choose the equalizer response $\boldsymbol{w}_N$ in order to obtain an estimate on any of these. We thus use z_i as an estimate of the input symbol $x_{i-\Delta}$, where Δ is called the delay of the equalizer, and depending on the structure of the channel it may be advantageous to choose one delay rather than the other. Accordingly we have an extra parameter involved in the design of the equalizer leading to the following modification of the MMSE in Definition 1.1.1.

Definition 1.1.2. Let a communication system as shown in Figure 1.2 use bandwidth B to communicate over a channel of length L. If an equalizer with response $\boldsymbol{w}_N = (w_0, \ldots, w_{N-1})$ and delay Δ is used the mean square error, $\text{MSE}_N^{L,B}$, and the minimum mean square error, $\text{MMSE}_N^{L,B}$, respectively are defined as

$$\text{MSE}_N^{L,B} = \mathbb{E}(\|x_{i-\Delta} - z_i\|^2),$$

and

$$\text{MMSE}_N^{L,B} = \min_{\boldsymbol{w}_N} \min_{\Delta} \mathbb{E}(\|x_{i-\Delta} - z_i\|^2),$$

where z_i is the equalizer output and $x_{i-\Delta}$ is the input symbol. □

Given the channel response $\boldsymbol{h}_L^B$ along with the signal and noise power σ_x^2 and σ_n^2 respectively, Cioffi (2005) finds an expression for both the optimal choice of $\boldsymbol{w}_N$ and the corresponding value of $\text{MMSE}_N^{L,B}$. For our purposes only the latter is necessary and it is given by

$$\text{MMSE}_N^{L,B} = \sigma_n^2 \min \operatorname{diag} \left(H^* H + \tfrac{1}{\rho} I \right)^{-1}. \tag{1.7}$$

It should be noticed that for a given channel we always have $\text{MMSE}_\infty^{L,B} \leq \text{MMSE}_N^{L,B}$ since the finite equalizer is a sub-model of the infinite equalizer. Obviously if the optimal finite response is $\boldsymbol{w}_N = (w_0, \ldots, w_{N-1})$ we could simply choose the infinite response as $\boldsymbol{w} = (\ldots, 0, 0, w_0, \ldots, w_{N-1}, 0, 0, \ldots)$ to obtain the same MSE.

1.2 Simulations

To illustrate the behavior of the performance measure $\text{MMSE}_\infty^{L,B}$ introduced in Section 1.1 we present a simulation study of the distribution of $\text{MMSE}_\infty^{L,B}$ in a specific setup. For simplicity we fix the signal and noise power to 1, i.e. $\sigma_x^2 = \sigma_n^2 = 1$, corresponding to a signal to noise ratio of $\rho = 1$. We need to choose a probability distribution on the positive real half line as the PDP. Several different distributions have been used in the literature, and Pereira et al. (2006) suggests that either the exponential distribution or the Rayleigh distribution can be used. As an illustration we use the Rayleigh distribution illustrated in Figure 1.3, which has density and distribution function given by

$$p(x) = 2x \exp(-x^2) \qquad \text{and} \qquad F(x) = 1 - \exp(-x^2).$$

We wish to study the distribution of $\text{MMSE}_{\infty}^{L,B}$ as the bandwidth B tends to infinity corresponding to the usage of large bandwidth in UWB systems. In the analytical study of this problem in Chapter 3 it is necessary to assume the channel length L grows such that $L/B \to \infty$, which is done here by setting $B = B(L) = \sqrt{L}$ (or equivalently $L = B^2$). Finally we assume the channel distribution to be complex Gaussian.

Given this assumptions we can simulate a channel response $h_0^B, \ldots, h_{L-1}^B$ by drawing L iid complex Gaussian variables $X_0, \ldots, X_{L-1}$ and set

$$h_l^B = \sigma_l^B X_l,$$

where

$$\sigma_l^B = \int_{\frac{l}{B}}^{\frac{l+1}{B}} p(t)dt = \exp\left(-\left(\tfrac{l}{B}\right)^2\right) - \exp\left(-\left(\tfrac{l+1}{B}\right)^2\right).$$

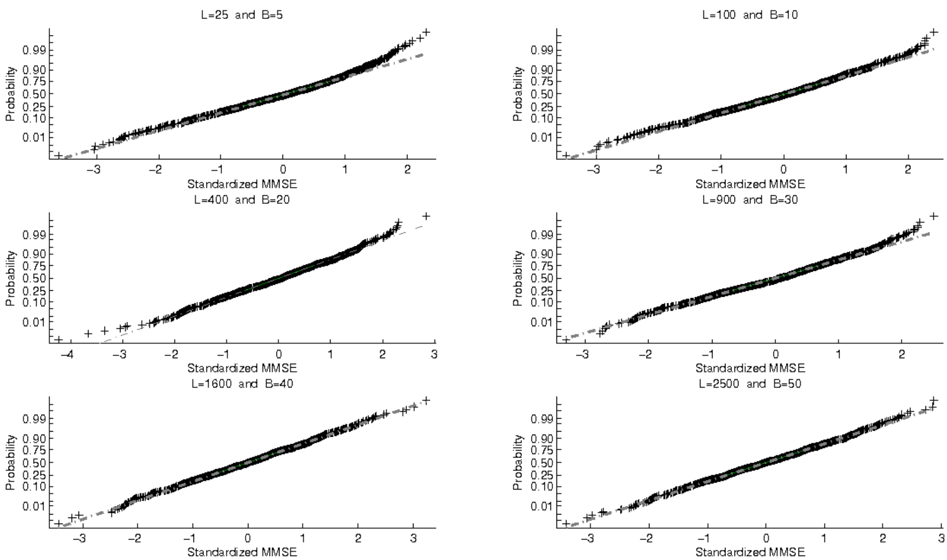

Figure 1.4: Quantile-quantile plot of 1000 standardized simulated MMSEs for six different values of L, B.

Given the channel response we have a deterministic expression for the transfer function $H^{L,B}$, and correspondingly a deterministic integral expression for $\text{MMSE}_{\infty}^{L,B}$ in (1.3), which can be evaluated by standard quadrature methods. The results given here are obtained using MATLAB, where the standard quadrature algorithm is an adaptive Simpson sum algorithm. For each bandwidth $B = 5, 10, 20, 30, 40, 50$ we have simulated 1000 channel responses and correspondingly calculated 1000 values of $\mathrm{MMSE}_\infty^{L,B}$, which we for each L, B have standardized by the empirical mean and standard deviation. The Q-Q plots for the standardized values are shown in Figure 1.4. For small values of L the distribution looks non-Gaussian, whereas the Gaussian assumption seems to be reasonable for large L. It is on the basis of these observations we set forth to study the asymptotic properties of the random variable $\mathrm{MMSE}_\infty^{L,B}$ and the transfer function $H^{L,B}$ used to calculate it.

If a finite equalizer is used the performance measure is denoted $\mathrm{MMSE}_N^{L,B}$, which we also wish to characterize. In this case we rely on a result in Chapter 3, where it is proven that for fixed L, B the distribution of $\mathrm{MMSE}_N^{L,B}$ converges to the distribution of $\mathrm{MMSE}_\infty^{L,B}$. We would thus obtain results arbitrarily close in distribution to the results given here by choosing N sufficiently large, and therefore the simulations are not carried out in this case.

1.3 Problem delimitation

As mentioned in the preface the subject of this thesis was inspired by the unpublished work in Pereira et al. (2006). They have conjectured that it is possible to prove asymptotic normality of $\mathrm{MMSE}_\infty^{L,B}$ given by (1.3) when we let the bandwidth B and the channel length L tend to infinity. Furthermore they have realized that the expression (1.7) for $\mathrm{MMSE}_N^{L,B}$ corresponds to a Riemann sum approximation of $\mathrm{MMSE}_\infty^{L,B}$, such that the pointwise convergence $\mathrm{MMSE}_N^{L,B} \to \mathrm{MMSE}_\infty^{L,B}$ holds for $N \to \infty$, which we verify in detail in Section 3.3. This result leads us to conjecture that the asymptotic normality of $\mathrm{MMSE}_\infty^{L,B}$ is inherited by $\mathrm{MMSE}_N^{L,B}$.

The two conjectures are:

Conjecture 1.3.1. *Under suitable conditions there exist $\mu, \sigma \in \mathbb{R}$ such that for $L, B \to \infty$ we have*

$$\frac{\sqrt{B}}{\sigma}\left(\mathrm{MMSE}_\infty^{L,B} - \mu\right) \xrightarrow{D} \mathcal{N}(0,1),$$

where $\xrightarrow{D}$ denotes convergence in distribution.

Conjecture 1.3.2. *Under suitable conditions exist $\mu, \sigma \in \mathbb{R}$ such that for $L, B, N \to \infty$ we have*

$$\frac{\sqrt{B}}{\sigma}\left(\mathrm{MMSE}_N^{L,B} - \mu\right) \xrightarrow{D} \mathcal{N}(0,1).$$

The aim of this thesis is to study these conjectures and outline how they can be proved. In order to do this we will study the transfer function $H^{L,B}$ given by (1.4), which is a complex-valued stochastic process. We wish to find a limit of this process when $L, B \to \infty$, denoted H^∞, and determine the properties of this limiting transfer function.

1.3.1 Outline

In the following we give an outline of how the rest of the thesis is structured.

Chapter 2 contains the necessary topics from the theory of stochastic processes. Sections 2.1 and 2.2 briefly review basic properties of random variables and stochastic processes, and they mainly serve as a way to introduce a consistent notation and to provide a reference on some necessary results. Section 2.3 covers mixing properties of stochastic processes, which is an essential assumption to prove a continuous CLT, which is done in detail in Section 2.4. This CLT ensures that the integral of a stochastic process is asymptotically Gaussian under suitable conditions and normalizations, and it is thus a natural step in proving the conjectures. Finally Section 2.5 finishes the chapter with a brief description of weak convergence in metric spaces and provides results used to determine the limiting transfer function H^{∞}.

Chapter 3 presents the specific results obtained for $\text{MMSE}_{\infty}^{L,B}$, $\text{MMSE}_{N}^{L,B}$, $H^{L,B}$, and H^{∞} in this thesis. In Section 3.1 it is shown that the limit H^{∞} must be a complex Gaussian process with marginal distribution $\mathbb{C}\mathcal{N}(0,1)$ and autocovariance function $\hat{p}$, where $\hat{p}$ denotes the Fourier transform of the power delay profile of the channel. In Section 3.2 it is proven that if $H^{L,B}$ is substituted by the limiting process H^{∞} in (1.3) asymptotic normality holds. Furthermore conditions are given to ensure the asymptotic normality also holds for $\text{MMSE}_{\infty}^{L,B}$. Section 3.3 presents the results for $\text{MMSE}_{N}^{L,B}$. First for fixed L, B it is shown that $\text{MMSE}_{N}^{L,B} \to \text{MMSE}_{\infty}^{L,B}$ pointwise for $N \to \infty$. Then extensions to ensure the distributional properties of $\text{MMSE}_{\infty}^{L,B}$ are inherited by $\text{MMSE}_{N}^{L,B}$ are considered. Finally Section 3.4 discusses the results obtained in the chapter and especially the imposed assumptions are reviewed.

Chapter 4 rounds off the thesis with a conclusion and a discussion of possible future work within this area.

Appendix A consists of three sections. Sections A.1 and A.2 serve as a reference on some necessary results and no proofs for these are given. In Section A.3 some auxiliary lemmas used in Chapters 2 and 3 are presented. These lemmas constitute important steps in some of the proofs given in the thesis, and they are all proved in detail. The results have merely been moved to the appendix to smooth the flow of the main exposition.

CHAPTER 2

Selected topics in stochastic processes

To treat the problem of Chapter 1 rigorously some results on stochastic processes are needed. These will be outlined in this chapter.

2.1 Random Variables

A real random variable is a measurable mapping $X : \Omega \to \mathbb{R}$, where $(\Omega, \mathcal{E}, P)$ is a probability space and we use $(\mathbb{R}, \mathbb{B}(\mathbb{R}), \mu)$ as the measurable space X maps to, where $\mathbb{B}(\cdot)$ denotes the Borel σ-algebra and μ is the Lebesgue measure. The distribution F of X is the measure induced on $\mathbb{B}(\mathbb{R})$ by P. Thus for any Borel set A, $F(A) = P(X \in A)$. The distribution of X is also referred to as the law of X, denoted $\mathcal{L}(X)$. The distribution function, which we also denote F, is the function from $\mathbb{R}$ to $[0, 1]$ given by $F(x) = P(X \leq x)$. The distribution function will also be used notationally when we integrate with respect to a probability distribution. I.e. if F is the distribution function of the random variable X and we wish to integrate a function f with respect to this probability distribution we use either of the following notations:

$$\int f(x)dP \qquad \int f(x)dF(x) \qquad \int f(x)dP(X \leq x).$$

If X follows a standard normal distribution denoted $X \sim \mathcal{N}(0, 1)$ we denote the distribution function Φ rather than F.

2.1.1 Conditional expectation

We will need a few results on conditional expectations, which are presented here.

Definition 2.1.1. Let X be a random variable on the probability space $(\Omega, \mathcal{E}, P)$. If $\mathcal{F}$ is a σ-algebra contained in $\mathcal{E}$ we define the conditional expectation of X given $\mathcal{F}$ denoted $\mathbb{E}(X|\mathcal{F})$ as the random variable satisfying

(i) $\mathbb{E}(X|\mathcal{F})$ is measurable with respect to $\mathcal{F}$.

(ii) $\int_A \mathbb{E}(X|\mathcal{F})dP = \int_A XdP$ for all $A \in \mathcal{F}$. □

The existence of such a random variable is proved in Billingsley (1986) to which we refer for more details on conditional expectation. From the definition it is clear that

$$\mathbb{E}[\mathbb{E}(X|\mathcal{F})] = \int_\Omega \mathbb{E}(X|\mathcal{F})dP = \int_\Omega XdP = \mathbb{E}(X), \tag{2.1}$$

which is referred to as the law of total expectation.

Another useful property is that if X is measurable with respect to the conditional σ-algebra it behaves like a constant with respect to the conditional expectation, i.e. if X is measurable with respect to $\mathcal{F}$, and if Y and XY are integrable, then

$$\mathbb{E}(XY|\mathcal{F}) = X\mathbb{E}(Y|\mathcal{F}) \tag{2.2}$$

with probability 1 (Billingsley, 1986, Theorem 34.3.).

2.1.2 Stochastic convergence

Since a large part of this thesis is concerned with asymptotic properties of stochastic processes and variables it is important first to clarify what is meant by the various types of convergence that are used. We let (X_n) denote an entire sequence of random variables

$$\ldots, X_{-1}, X_0, X_1, \ldots$$

whereas X_n refers specifically to the n'th random variable of this sequence. This notation will also be carried over to the case of stochastic processes as we shall see later.

Definition 2.1.2. A sequence of random variables (X_n) is said to converge weakly to the random variable X if the distribution function F_n of X_n converges pointwise to the distribution function F of X at all continuity points of F. I.e. if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

at all continuity points of F. This is also called convergence in distribution or law and it is denoted either

$$X_n \xrightarrow{\mathcal{D}} X \qquad \text{or} \qquad \lim_{n\to\infty} \mathcal{L}(X_n) = \mathcal{L}(X).$$ □

Definition 2.1.3. A sequence of random variables (X_n) is said to converge in probability to the random variable X if for all $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0,$$

and convergence in probability is denoted $X_n \xrightarrow{p} X$. □

Definition 2.1.4. A sequence of random variables (X_n) is said to converge almost surely (a.s.) to the random variable X if

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

for all $\omega \in \Omega \backslash N$, where $P(N) = 0$. This is also called convergence almost everywhere or convergence with probability 1, and we denote it $X_n \xrightarrow{\text{a.s.}} X$.
If $N = \emptyset$ we say X_n converges surely, pointwise or everywhere to X, which is denoted $X_n \to X$. □

These definitions of convergence have been listed in ascending order of strength meaning

$$X_n \to X \quad \Longrightarrow \quad X_n \xrightarrow{\text{a.s.}} X \quad \Longrightarrow \quad X_n \xrightarrow{p} X \quad \Longrightarrow \quad X_n \xrightarrow{D} X.$$

There are several useful equivalent characterizations of convergence in distribution, which are known as the Portmanteau Lemma, but we shall only need one of these.

Lemma 2.1.5. *$X_n \xrightarrow{D} X$ if and only if $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all bounded, continuous functions f.*

Another fundamental property of weak convergence is that it is conserved under continuous mappings as stated below (van der Vaart and Wellner, 1996, Theorem 1.3.6).

Theorem 2.1.6. *If $g : \mathbb{R} \to \mathbb{R}$ is continuous and $X_n \xrightarrow{D} X$, then $g(X_n) \xrightarrow{D} g(X)$.*

In the case of convergence in distribution to a constant c we have the following result due to Slutsky.

Lemma 2.1.7. *If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} c$ then*

$$X_n + Y_n \xrightarrow{D} X + c.$$

Since convergence in probability implies convergence in distribution the lemma also holds with $Y_n \xrightarrow{p} c$, which is a often used version of the result. In fact it can be shown, that when the limiting random variable is a constant, convergence in distribution and in probability are equivalent. I.e. $Y_n \xrightarrow{D} c$ if and only if $Y_n \xrightarrow{p} c$ (van der Vaart, 1998, Theorem 2.7(iii)).

A powerful tool in the theory of convergence of random variables is the characteristic function defined below.

Definition 2.1.8. For a random variable X the characteristic function $\phi : \mathbb{R} \to \mathbb{C}$ of X is given by

$$\phi(t) = \mathbb{E}\big(\exp(itX)\big) = \int_{-\infty}^{\infty} \exp(itx)dP. \qquad \square$$

The importance of the characteristic function is apparent in the following theorem (Billingsley, 1986, Theorem 26.3.)

Theorem 2.1.9. *Let (X_n) be a sequence of random variables with corresponding characteristic functions ϕ_n and let the random variable X have characteristic function ϕ. Then the following two statements are equivalent.*

(i) $X_n \xrightarrow{D} X$

(ii) $\phi_n(t) \to \phi(t)$ *for all* $t \in \mathbb{R}$.

2.1.3 Central limit theorem for triangular arrays

A classical result on convergence of stochastic variables is the CLT for triangular arrays, which we for convenience recall here (Billingsley, 1986, p. 368-369). Suppose that for each $n \in \mathbb{N}$ we have k_n independent random variables and let $k_n \to \infty$ for $n \to \infty$ then we have the following triangular array

$$\begin{array}{l} X_{11}, \ldots, X_{1k_1} \\ X_{21}, \ldots\ldots\ldots, X_{2k_2} \\ \quad\vdots \qquad\qquad\qquad \ddots \\ X_{n1}, \ldots\ldots\ldots\ldots\ldots\ldots, X_{nk_n} \\ \quad\vdots \qquad\qquad\qquad\qquad\qquad \ddots \end{array}$$

Let

$$S_{k_n} = \sum_{j=1}^{k_n} X_{nj} \qquad \text{and} \qquad \sigma_{k_n}^2 = \operatorname{Var}\big(S_{k_n}\big) = \sum_{j=1}^{k_n} \operatorname{Var}(X_{nj}).$$

Then the CLT for triangular arrays states as follows.

Theorem 2.1.10. *Suppose that for each $n \in \mathbb{N}$ the random variables $X_{n1}, \ldots, X_{nk_n}$ are dependent with $\mathbb{E}(X_{nj}) = 0$, $j = 1, \ldots, k_n$. If for all $\varepsilon > 0$*

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} \frac{1}{\sigma_{k_n}^2} \int_{|x|\geq \varepsilon\sigma_{k_n}} x^2 dP(X_{nj} \leq x) = 0, \tag{2.3}$$

then

$$\frac{S_{k_n}}{\sigma_{k_n}} \xrightarrow{D} \mathcal{N}(0,1) \qquad \textit{as } n \to \infty.$$

The condition (2.3) is called the Lindeberg condition, which we will refer to later.

2.1.4 Complex random variables

In the following the concept of a random variable is extended to be complex-valued rather than real-valued.

Definition 2.1.11. A mapping $Z : \Omega \to \mathbb{C}$ is called a complex random variable if there exist real random variables X, Y defined on Ω, such that

$$Z = X + iY. \qquad \square$$

For complex random variables we let $\mathbb{E}(Z) = \mathbb{E}(X) + i\mathbb{E}(Y)$ and $\operatorname{Cov}(Z_1, Z_2) = \mathbb{E}\big[\big(Z_1 - \mathbb{E}(Z_1)\big)\big(\overline{Z_2 - \mathbb{E}(Z_2)}\big)\big]$, such that the variance $\operatorname{Var}(Z) = \operatorname{Cov}(Z, Z) = \mathbb{E}(Z\overline{Z}) -$

$\mathbb{E}(Z)\overline{\mathbb{E}(Z)}$ is a non-negative real number if it exists.

One of the most important examples of a complex random variable is the complex normal distribution, which we define through the two dimensional real normal distribution as in Andersen et al. (1995). If we let $[\cdot] : \mathbb{C}^p \to \mathbb{R}^{2p}$ denote the natural bijection, between complex and real vectors, given by

$$[\boldsymbol{x}] = \begin{pmatrix} \mathrm{Re}(\boldsymbol{x}) \\ \mathrm{Im}(\boldsymbol{x}) \end{pmatrix}, \quad \boldsymbol{x} \in \mathbb{C}^p,$$

then the complex normal distribution is defined as

Definition 2.1.12. A complex random variable X follows a complex normal distribution, with mean $\mu \in \mathbb{C}$ and variance $\sigma^2 \in \mathbb{R}^+$ if $[X] \sim \mathcal{N}_2\Big([\mu], \frac{\sigma^2}{2} I_2\Big)$. This is denoted $X \sim \mathbb{C}\mathcal{N}(\mu, \sigma^2)$. □

A useful property directly inherited from the usual two dimensional Gaussian distribution with independent marginals is isotropy, i.e. the complex normal distribution is invariant to rotation. If we consider the magnitude of a complex Gaussian stochastic variable we obtain the Rayleigh distribution, which is treated in the following.

The Rayleigh Distribution

Let $X \sim \mathbb{C}\mathcal{N}(0, 2\sigma^2)$, then the density of $[X]$ is

$$f_{[X]}(x_1, x_2) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(x_1^2 + x_2^2)}{2\sigma^2}\right].$$

Transforming this into modulus/argument representation ($X = R_\Theta$) using the transformation theorem yields

$$f_{R,\Theta}(r, \theta) = \frac{r}{2\pi\sigma^2} \exp\left[-\frac{r^2}{2\sigma^2}\right], \tag{2.4}$$

since the inverse Jacobian of the transform between Cartesian coordinates and modulus/argument representation is r. The expression (2.4) is seen to be independent of the phase θ which means that $p(r, \theta) = p(r)p(\theta)$. We furthermore have

$$\begin{aligned} f_\Theta(\theta) &= \int_0^\infty \frac{r}{2\pi\sigma^2} \exp\left[-\frac{r^2}{2\sigma^2}\right] dr = \frac{1}{2\pi}, \quad 0 \le \theta \le 2\pi \\ f_R(r) &= \int_0^{2\pi} \frac{r}{2\pi\sigma^2} \exp\left[-\frac{r^2}{2\sigma^2}\right] d\theta = \frac{r}{\sigma^2} \exp\left[-\frac{r^2}{2\sigma^2}\right], \quad 0 \le r. \end{aligned} \tag{2.5}$$

Thus if $X \sim \mathbb{C}\mathcal{N}(0, 2\sigma^2)$ or equivalently $[X] \sim \mathcal{N}_2(0, \sigma^2 I_2)$, then the distribution of $R = |X|$ is independent of the phase and the density is given by (2.5). A distribution with this density is called a Rayleigh distribution with parameter σ. This is denoted $R \sim \text{Rayleigh}(\sigma)$, and the mean and variance are given by

$$\mathbb{E}(R) = \sigma\sqrt{\frac{\pi}{2}}$$
$$\text{Var}(R) = \frac{(4-\pi)\sigma^2}{2}.$$

Since we often work with squared magnitudes we wish to describe the distribution of $Y = R^2$ as well. Using the transformation theorem for the transformation $t(r) = r^2$ yields

$$\begin{aligned} f_Y(y) &= f_R(t^{-1}(y))\left|\frac{d}{dy}t^{-1}(y)\right| \\ &= \frac{\sqrt{y}}{\sigma^2}\exp\left[-\frac{\sqrt{y}^2}{2\sigma^2}\right]\frac{1}{2\sqrt{y}} \\ &= \frac{1}{2\sigma^2}\exp\left[-\frac{y}{2\sigma^2}\right]. \end{aligned}$$

This is seen to be the density of the exponential distribution with mean $2\sigma^2$, such that if $X \sim \mathbb{C}\mathcal{N}(0, 2\sigma^2)$ then $|X|^2 \sim \text{Exp}(\frac{1}{2\sigma^2})$.

2.2 Stochastic processes

We will work with stochastic processes with both real and complex values, and therefore the majority of definitions and results are done for the complex case. Thus in this thesis the term stochastic process allows complex values unless stated otherwise. The section is based on Doob (1953).

A stochastic (or random) process (X_t) is a family of complex random variables X_t, $t \in T \subseteq \mathbb{R}$, defined on the same probability space $(\Omega, \mathcal{E}, P)$. If T is countable the process is said to be a discrete time process and otherwise it is called a continuous time process. For discrete time processes we usually work with $T = \mathbb{Z}$ or $T = \mathbb{N}_0$ while continuous time processes usually have $T = \mathbb{R}$ or $T = \mathbb{R}_+ \cup \{0\}$. The subscript t is used both in general and when we only consider continuous time processes, whereas a discrete time process often will be denoted (X_n) or (X_j). A stochastic process (X_t) is called strict sense stationary if for all h and all finite collections $\{t_1, \ldots, t_n\}$

$$\mathcal{L}\left((X_{t_1}, \ldots, X_{t_s})\right) = \mathcal{L}\left((X_{t_1+h}, \ldots, X_{t_s+h})\right).$$

The process is called wide sense stationary if $\mathbb{E}(|X_t|^2) < \infty$ for all $t \in T$ and if $\mathbb{E}(X_s)$ and the autocovariance function $R_X(s,t) = \mathbb{E}(X_{s+t}X_s^*)$ does not depend on s.

While strict sense stationarity implies wide sense stationarity the opposite is not true in general, but it does hold for Gaussian processes. When the process is wide sense stationary the autocovariance function is considered as a function of only one variable $R_X : T \to \mathbb{C}$. It can be shown to be a continuous and positive definite function, which implies it has spectral representation (Ibragimov and Linnik, 1971, p. 291)

$$R_X(t) = \int_{-\infty}^{\infty} \exp(it\lambda) dF(\lambda), \tag{2.6}$$

where the spectral function F is bounded and non-decreasing. The totally finite Borel measure μ on $\mathbb{R}$ induced by F is called the spectral measure of (X_t). If μ is absolutely continuous with respect to the Lebesgue measure it has a density f, which we call the spectral density of (X_t).

A stochastic process (X_t) is a function of two variables (T, Ω) and (X_t) is called measurable, if it is measurable with respect to the product σ-algebra generated by the Lebesgue measure on T and the probability measure P on Ω. The concept of measurable stochastic processes is very convenient when working with integrals as the following theorem illustrates (Doob, 1953, Theorem 2.7).

Theorem 2.2.1. *Let (X_t) be a measurable stochastic process. Then for almost all $\omega \in \Omega$ the sample function $X_.(\omega) : T \to \mathbb{C}$ is measurable. If $\mathbb{E}(X_t(\omega))$ exists for $t \in T$, it defines a measurable function of t. If $A \subseteq T$ is Lebesgue measurable and*

$$\int_A \mathbb{E}(|X_t(\omega)|) dt < \infty$$

then almost all sample functions are Lebesgue integrable over A.

For a given stochastic process (X_t) it is natural to consider the minimal σ-algebra $\mathcal{M}(X)$ on Ω making every random variable X_t measurable. This is the σ-algebra generated by events of the form

$$A = \left\{ \omega \in \Omega : \left[X_{t_1}(\omega), \ldots, X_{t_n}(\omega) \right] \in \tilde{A} \right\}, \tag{2.7}$$

where $\{t_1, \ldots, t_n\} \in T$ and $\tilde{A}$ is a n-dimensional Borel set. Later we will need some smaller σ-algebras $\mathcal{M}_a^b(X)$ generated by the sets of the form (2.7), where $\{t_1, \ldots, t_n\} \in (a, b) \cap T$. I.e. $\mathcal{M}_a^b(X)$ is the smallest σ-algebra such that each random variable X_t of the stochastic process (X_t) truncated to the interval (a, b) is measurable.

When we study continuous time processes we shall always assume that the process is stochastically continuous such that for almost all t we have $X_{t+s} \xrightarrow{p} X_t$ for $s \to 0$. This includes most processes of interest and especially it covers all the cases needed here, but it is worth noticing, that white noise processes are not stochastically continuous and thus not covered by the analysis done here. An important implication of stochastic continuity is that the process almost surely is measurable in the following sense (Doob, 1953, Theorem 2.6, p. 61).

Theorem 2.2.2. *Let (X_t), $t \in T$ be a stochastic process with a Lebesgue measurable parameter set T. Suppose that for all $\varepsilon > 0$*

$$\lim_{s \to 0} P(|X_{t+s} - X_t| > \varepsilon) = 0,$$

for all $t \in T \backslash N$ where $\mu(N) = 0$. Then there is a process $(\tilde{X}_t)$, $t \in T$ defined on the same measurable space $(\Omega, \mathcal{E}, P)$ which is measurable, and for which

$$P\big(\tilde{X}_t(\omega) = X_t(\omega)\big) = 1, \qquad t \in T.$$

This result ensures that to every stochastically continuous process (X_t) corresponds a measurable process $(\tilde{X}_t)$, which is almost surely equal to the original process (X_t). In this thesis the measurable version of (X_t) will always be used such that integrals of stochastically continuous processes will be meaningful. In Section 2.4 the limiting distribution of such integrals is studied, but first we need to introduce the concept of mixing for stochastic processes.

2.3 Mixing

In this section we will clarify what is meant by weakly dependent stochastic processes using the concept of mixing. This introduction only covers what is necessary in connection with the CLT in Section 2.4, and it is based on Ibragimov and Linnik (1971).

As usual the underlying probability space is denoted $(\Omega, \mathcal{E}, P)$. We denote by $L^2(\mathcal{E})$ the set of real functions measurable with respect to $\mathcal{E}$, which are square integrable. For any two σ-algebras $\mathcal{F}, \mathcal{G} \subseteq \mathcal{E}$ we define the two measures of dependence α and ρ by

$$\alpha(\mathcal{F}, \mathcal{G}) = \sup \big\{ |P(A)P(B) - P(A \cap B)| : A \in \mathcal{F}, B \in \mathcal{G} \big\} \tag{2.8}$$

and

$$\rho(\mathcal{F}, \mathcal{G}) = \sup \big\{ |\mathrm{Corr}(X, Y)| : X \in L^2(\mathcal{F}), Y \in L^2(\mathcal{G}) \big\},$$

where the correlation between two random variables X and Y as usual is given by

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

In the literature several other measure of dependence of σ-algebras are studied (for reviews see Doukhan, 1994; Bradley, 2005), but in this exposition only the two mentioned above are necessary.

For a given stochastic process (X_t) with $T=\mathbb{Z}$ or $T=\mathbb{R}$ we define the two mixing coefficients

$$\alpha_X(\tau) = \sup_{t\in T} \left\{ \alpha\Big(\mathcal{M}^t_{-\infty}(X), \mathcal{M}^\infty_{t+\tau}(X)\Big) \right\}$$

and

$$\rho_X(\tau) = \sup_{t\in T} \left\{ \rho\Big(\mathcal{M}^t_{-\infty}(X), \mathcal{M}^\infty_{t+\tau}(X)\Big) \right\}.$$

Clearly for strictly stationary processes the value is independent of t and thus

$$\alpha_X(\tau) = \alpha\Big(\mathcal{M}^0_{-\infty}(X), \mathcal{M}^\infty_{\tau}(X)\Big) \tag{2.9}$$

and

$$\rho_X(\tau) = \rho\Big(\mathcal{M}^0_{-\infty}(X), \mathcal{M}^\infty_{\tau}(X)\Big). \tag{2.10}$$

Since we will only consider mixing properties of stationary processes (2.9) and (2.10) will be taken as definitions of α_X and ρ_X.

Definition 2.3.1. A stationary process (X_t) is said to be strongly mixing or α-mixing if $\alpha_X(\tau) \to 0$ for $\tau \to \infty$ and (X_t) is called completely regular or ρ-mixing if $\rho_X(\tau) \to 0$ for $\tau \to \infty$. Furthermore ρ is called the maximal correlation coefficient. □

A general property of the mixing coefficients is that $\alpha_X(\tau) \leq \rho_X(\tau)$ for all τ, such that complete regularity implies strong mixing. The converse is not true in general, but for Gaussian processes the Kolmogorov-Rozanov Theorem (Doukhan, 1994, Theorem 1, page 57) states that

$$\alpha_X(\tau) \leq \rho_X(\tau) \leq 2\pi\alpha_X(\tau),$$

and in this case the two mixing conditions are equivalent since if $\alpha_X(\tau) \to 0$ for $\tau \to \infty$, then $\rho_X(\tau) \to 0$ for $\tau \to \infty$ and vice versa.

The following theorem is essential in the proof of the CLT in Section 2.4 (Ibragimov and Linnik, 1971, Theorem 17.2.1.).

Theorem 2.3.2. *Let the real stationary stochastic process (X_t) be strongly mixing. If ξ is measurable with respect to $\mathcal{M}^t_{-\infty}(X)$ and η with respect to $\mathcal{M}^\infty_{t+\tau}(X)$, and if $|\xi| \leq C_1$, $|\eta| \leq C_2$, then*

$$|\mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\mathbb{E}(\eta)| \leq 4C_1C_2\alpha_X(\tau). \tag{2.11}$$

Furthermore if (X_t) is a complex-valued stochastic process (2.11) *holds with 4 replaced by 16.*

Proof. To ease the notation we let $\mathcal{M}_{-\infty}^0 = \mathcal{M}_{-\infty}^0(X)$ and $\mathcal{M}_\tau^\infty = \mathcal{M}_\tau^\infty(X)$ in the following. By stationarity it is suffices to consider $t = 0$. The law of total expectation (2.1) and (2.2) yields

$$\mathbb{E}(\xi\eta) = \mathbb{E}\big[\mathbb{E}\big(\xi\eta|\mathcal{M}_{-\infty}^0\big)\big] = \mathbb{E}\big[\xi\mathbb{E}\big(\eta|\mathcal{M}_{-\infty}^0\big)\big],$$

whereby

$$\begin{aligned}|\mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\mathbb{E}(\eta)| &= \big|\mathbb{E}\big(\xi\big[\mathbb{E}\big(\eta|\mathcal{M}_{-\infty}^0\big) - \mathbb{E}(\eta)\big]\big)\big| \\ &\le C_1\mathbb{E}\big(\big|\mathbb{E}\big(\eta|\mathcal{M}_{-\infty}^0\big) - \mathbb{E}(\eta)\big|\big).\end{aligned}$$

Introducing the random variable

$$\xi_1 = \operatorname{sign}\Big\{\mathbb{E}\big(\eta|M_{-\infty}^0\big) - \mathbb{E}(\eta)\Big\},$$

which is measurable with respect to $\mathcal{M}_{-\infty}^0$, we have

$$\begin{aligned}|\mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\mathbb{E}(\eta)| &\le C_1\mathbb{E}\big(\xi_1\big[\mathbb{E}\big(\eta|\mathcal{M}_{-\infty}^0\big) - \mathbb{E}(\eta)\big]\big) \\ &\le C_1|\mathbb{E}(\xi_1\eta) - \mathbb{E}(\xi_1)\mathbb{E}(\eta)|\end{aligned}$$

If we in a similar manner condition on $\mathcal{M}_\tau^\infty$ and define

$$\eta_1 = \operatorname{sign}\Big\{\mathbb{E}(\xi_1|M_\tau^\infty) - \mathbb{E}(\xi_1)\Big\},$$

which is measurable with respect to $\mathcal{M}_\tau^\infty$, we arrive at

$$|\mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\mathbb{E}(\eta)| \le C_1C_2|\mathbb{E}(\xi_1\eta_1) - \mathbb{E}(\xi_1)\mathbb{E}(\eta_1)|. \tag{2.12}$$

For the final part of the proof we introduce the events

$$\begin{aligned}A &= \{\omega : \xi_1(\omega) = 1\} \in \mathcal{M}_{-\infty}^0 \\ A^c &= \{\omega : \xi_1(\omega) = -1\} \in \mathcal{M}_{-\infty}^0 \\ B &= \{\omega : \eta_1(\omega) = 1\} \in \mathcal{M}_\tau^\infty. \\ B^c &= \{\omega : \eta_1(\omega) = -1\} \in \mathcal{M}_\tau^\infty.\end{aligned}$$

Since the random variables ξ_1, η_1 only take the values 1 and -1 it is easy to express the expected values

$$\begin{aligned}\mathbb{E}(\xi_1\eta_1) &= P(A\cap B) + P(A^c\cap B^c) - P(A\cap B^c) - P(A^c\cap B) \\ \mathbb{E}(\xi_1)\mathbb{E}(\eta_1) &= P(A)P(B) + P(A^c)P(B^c) - P(A)P(B^c) - P(A^c)P(B).\end{aligned}$$

Using the definition of the strong mixing coefficient (2.8) we have

$$|P(A \cap B) - P(A)P(B)| \leq \alpha_X(\tau),$$

which also holds for the other combinations of A, A^c, B, B^c and therefore

$$|\mathbb{E}(\xi_1\eta_1) - \mathbb{E}(\xi_1)\mathbb{E}(\eta_1)| \leq 4\alpha_X(\tau).$$

Inserting into (2.12) yields the result

$$|\mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\mathbb{E}(\eta)| \leq 4C_1C_2\alpha_X(\tau).$$

■

Another useful property of mixing is that it is conserved under measurable mappings as we shall see in the following. We will only need this result for a function $f : \mathbb{C} \to \mathbb{R}$, so it is solely stated for this case.

Lemma 2.3.3. *Let (X_t) be a complex-valued stationary stochastic process defined on the probability space $(\Omega, \mathcal{E}, P)$, and let $f : \mathbb{C} \to \mathbb{R}$ be a $\mathbb{B}(\mathbb{C})$–$\mathbb{B}(\mathbb{R})$ measurable map. Then the minimal σ-algebras $\mathcal{M}_a^b(X)$ and $\mathcal{M}_a^b(f(X))$ generated by respectively X_t and $f(X_t)$ for $t \in (a, b) \cap T$ satisfy*

$$\mathcal{M}_a^b(f(X)) \subseteq \mathcal{M}_a^b(X).$$

Proof. The σ-algebra $\mathcal{M}_a^b(f(X))$ is generated by sets of the form

$$A = \{\omega \in \Omega : [f(X_{t_1}(\omega)), \dots, f(X_{t_n}(\omega))] \in \tilde{A}\},$$

where $\{t_1, \dots, t_n\} \in (a, b) \cap T$ and $\tilde{A} = \tilde{A}_1 \times \cdots \times \tilde{A}_n$ is a n-dimensional Borel set. This can be written as

$$A = \{\omega \in \Omega : [X_{t_1}(\omega), \dots, X_{t_n}(\omega)] \in f^{-1}(\tilde{A})\},$$

where $f^{-1}(\tilde{A}) = f^{-1}(\tilde{A}_1) \times \cdots \times f^{-1}(\tilde{A}_n)$. Since f is $\mathbb{B}(\mathbb{C})$–$\mathbb{B}(\mathbb{R})$ measurable $f^{-1}(\tilde{A})$ is a n-dimensional Borel set, and thus element of the generating class for $\mathcal{M}_a^b(X)$. This shows that every set used to generate $\mathcal{M}_a^b(f(X))$ also is used to generate $\mathcal{M}_a^b(X)$ and thus $\mathcal{M}_a^b(f(X)) \subseteq \mathcal{M}_a^b(X)$. ■

This lemma implies that if α_Y and ρ_Y denote the mixing coefficients of the process Y, given by $Y_t = f(X_t)$ then $\alpha_Y(\tau) \leq \alpha_X(\tau)$ and $\rho_Y(\tau) \leq \rho_X(\tau)$ for all τ.

The next lemma uses big-O and little-o notation, which we briefly recap before presenting the lemma. We say $f = O(g)$, if there exists $x_0, M > 0$ such that

$$|f(x)| \leq M|g(x)| \text{ for } x > x_0,$$

and $f = o(g)$ if

$$\lim_{x \to \infty} \left| \frac{f(x)}{g(x)} \right| = 0.$$

Now we give some properties of the mixing coefficients used to prove the CLT later.

Lemma 2.3.4. *Let (X_j) be a strongly mixing stochastic process with mixing coefficient α_X. If*

$$\sum_{n=1}^{\infty} \alpha_X(n) < \infty,$$

then

(i) $\alpha_X(n) = o(n^{-1})$,

(ii) $\sum_{j=1}^{n} j\alpha_X(j) = o(n)$,

(iii) $\sum_{j=1}^{k} \alpha_X(nj) = O(n^{-1})$.

Proof.

Ad (i):

Since $\alpha_X(n)$ is decreasing we have

$$\frac{2}{n} \sum_{j=\lfloor \frac{n}{2} \rfloor}^{n} \alpha_X(j) \geq \frac{2}{n}\Big(\frac{n}{2}+1\Big)\alpha_X(n) \geq \alpha_X(n),$$

such that

$$n\alpha_X(n) \leq 2 \sum_{j=\lfloor \frac{n}{2} \rfloor}^{n} \alpha_X(j).$$

The result then follows by the assumed convergence of $\sum \alpha_X(n)$.

Ad (ii):

This result follows by splitting the sum at the $\sqrt{n}$'th term, such that

$$\frac{1}{n}\sum_{j=1}^{n} j\alpha_X(j) \leq \frac{1}{n}\sqrt{n} \sum_{j \leq \sqrt{n}} \alpha_X(j) + \sum_{j > \sqrt{n}} \alpha_X(j) = o(1).$$

Ad (iii):

Again using $\alpha_X(n)$ is decreasing we have for all $j \in \mathbb{N}$

$$\alpha_X(nj) \leq \frac{1}{n} \sum_{i=(j-1)n}^{jn-1} \alpha_X(i),$$

implying that

$$\sum_{j=1}^{k} \alpha_X(nj) \leq \frac{1}{n}\sum_{j=1}^{k}\sum_{i=(j-1)n}^{jn-1} \alpha_X(i) \leq \frac{1}{n}\sum_{j=1}^{\infty} \alpha_X(j) = O(n^{-1}). \qquad \blacksquare$$

2.4 Central limit theorems

In this section a CLT will be proved for both discrete and continuous time stochastic processes based on the results in Ibragimov and Linnik (1971). The continuous time CLT will turn out to be obtained as a corollary to the discrete time case, and therefore we start by proving the CLT for discrete time processes. First we define what we understand by a CLT in both cases.

Definition 2.4.1. A real stationary discrete time stochastic process (X_j) with $\mathbb{E}(X_j) = 0$ and $\mathrm{Var}(X_j) < \infty$ is said to satisfy the central limit theorem if

$$\frac{S_n}{\sigma_n} \xrightarrow{D} \mathcal{N}(0,1), \qquad \text{for} \quad n \to \infty,$$

where

$$S_n = \sum_{j=1}^{n} X_j, \qquad \sigma_n^2 = \mathrm{Var}(S_n). \qquad \square$$

Definition 2.4.2. A real stationary continuous time stochastic process (X_t) with $\mathbb{E}(X_t) = 0$, $\mathrm{Var}(X_t) < \infty$ is said to satisfy the central limit theorem if

$$\frac{S_T}{\sigma_T} \xrightarrow{D} \mathcal{N}(0,1), \qquad \text{for} \quad T \to \infty,$$

where

$$S_T = \int_0^T X_t dt, \qquad \sigma_T^2 = \mathrm{Var}(S_T). \qquad \square$$

From these definitions it is seen that the variance of the sum or integral plays an important role as a standardization factor, and the following lemma states some useful formulas for calculating this variance.

Lemma 2.4.3. *Let the real-valued stationary processes (X_j) and (X_t) be respectively a discrete time and a continuous time process with $\mathbb{E}(X_n) = \mathbb{E}(X_t) = 0$, $\mathrm{Var}(X_n) < \infty$, and $\mathrm{Var}(X_t) < \infty$. Then*

$$\sigma_n^2 = \mathrm{Var}(S_n) = \sum_{|j| \leq n} (n - |j|) R_X(j) = nR_X(0) + 2\sum_{j=1}^{n-1} (n-j) R_X(j)$$

and

$$\sigma_T^2 = \mathrm{Var}(S_T) = \int_{-T}^{T} (T - |t|) R_X(t) dt = 2\int_0^T (T-t) R_X(t) dt$$

Proof. We only prove the continuous time result since the discrete time analogue follows in the same manner by interpreting the integration to be with respect to the counting measure. By direct calculations we have

$$\begin{aligned}\sigma_T^2 &= \mathbb{E}\Big[\Big(\int_0^T X_t dt\Big)^2\Big] \\ &= \mathbb{E}\Big(\int_0^T \int_0^T X_t X_s dt ds\Big) \\ &= \int_0^T \int_0^T R_X(t-s) dt ds \\ &= \int_{-T}^T (T - |t|) R_X(t) dt.\end{aligned}$$

■

We will only need the CLT for bounded stochastic processes and therefore this will be required in the following lemmas. Also we will need the process to be strongly mixing with summable or integrable mixing coefficients to show the CLT, which also will be assumed in the lemmas.

Lemma 2.4.4. *Let the real-valued stationary process (X_j) be strongly mixing, with*

$$\sum_{n=1}^{\infty} \alpha_X(n) < \infty, \tag{2.13}$$

and let there exist a constant $C < \infty$ such that $P(|X_j| < C) = 1$. Then

$$\sigma^2 = R_X(0) + 2\sum_{j=1}^{\infty} R_X(j)\infty, \tag{2.14}$$

and if $\sigma \neq 0$, then

$$\sigma_n^2 = \sigma^2 n(1 + o(1)). \tag{2.15}$$

Proof. From Theorem 2.3.2 we have

$$|R_X(j)| = |\mathbb{E}(X_0 X_j)| = |\mathbb{E}(X_0 X_j) - \mathbb{E}(X_0)\mathbb{E}(X_j)| \leq 4C^2 \alpha_X(j).$$

The convergence of (2.14) now follows by the assumption (2.13). To verify the asymptotic variance formula (2.15) we use Lemma 2.4.3, which yields

$$
\begin{aligned}
\mathrm{Var}(S_n) &= nR_X(0) + 2\sum_{j=1}^{n}(n-j)R_X(j) \\
&= n\Big(R_X(0) + 2\sum_{j=1}^{n} R_X(j)\Big) - 2\sum_{j=1}^{n} jR_X(j) \\
&= n\Big(\sigma^2 - 2\sum_{n+1}^{\infty} R_X(j)\Big) - 2\sum_{j=1}^{n} jR_X(j) \\
&= n\sigma^2\left(1 - \frac{2}{\sigma^2}\sum_{n+1}^{\infty} R_X(j) - \frac{2}{n\sigma^2}\sum_{j=1}^{n} jR_X(j)\right).
\end{aligned}
$$

We notice that

$$\frac{2}{\sigma^2}\sum_{n+1}^{\infty} R_X(j) = O\Big(\sum_{n+1}^{\infty} \alpha_X(j)\Big) = o(1),$$

and using Lemma 2.3.4 we have

$$\frac{2}{n\sigma^2}\sum_{j=1}^{n} jR_X(j) = \frac{1}{n}O\Big(\sum_{j=1}^{n} j\alpha_X(j)\Big) = \frac{1}{n}o(n) = o(1). \qquad \blacksquare$$

Lemma 2.4.5. *Let the real stationary continuous time process (X_t) be strongly mixing, with*

$$\int_0^\infty \alpha_X(\tau)d\tau < \infty,$$

and let there exist a constant $C < \infty$ such that $P(|X_t| < C) = 1$. Then

$$\sigma^2 = 2\int_0^\infty R_X(t)dt$$

converges and if $\sigma \neq 0$, then

$$\sigma_T^2 = \sigma^2 T(1 + o(1)).$$

Proof. Using the same strategy as in the discrete time case above we have

$$|R_X(t)| \leq 4C^2\alpha_X(t),$$

which ensures convergence. Furthermore

$$\sigma_T^2 = T\sigma^2\Big(1 - \frac{2}{\sigma^2}\int_T^\infty R_X(t)dt - \frac{2}{T\sigma^2}\int_0^T tR_X(t)dt\Big).$$

Finally

$$\frac{2}{\sigma^2}\int_T^\infty R_X(t)dt = O\Big(\int_T^\infty \alpha_X(t)dt\Big) = o(1),$$

and

$$\begin{aligned}\frac{2}{T\sigma^2}\int_0^T tR_X(t)dt &= O\Big(\frac{1}{T}\Big[\int_0^{\sqrt{T}} t\alpha_X(t)dt + \int_{\sqrt{T}}^T t\alpha_X(t)dt\Big]\Big)\\ &= O\Big(\frac{\sqrt{T}}{T}\int_0^{\sqrt{T}}\alpha_X(t)dt + \frac{T}{T}\int_{\sqrt{T}}^T \alpha_X(t)dt\Big)\\ &= o(1).\end{aligned}$$

■

2.4.1 Discrete time central limit theorem

Now we prove the discrete time CLT for strongly mixing real-valued bounded stochastic processes with summable mixing coefficients.

Theorem 2.4.6. *Let the real-valued stationary process (X_j) be strongly mixing, with*

$$\sum_{n=1}^\infty \alpha_X(n) < \infty, \tag{2.16}$$

and let there exist a constant $C < \infty$ such that $P(|X_j| < C) = 1$. Then

$$\sigma^2 = R_X(0) + 2\sum_{j=1}^\infty R_X(j) < \infty \tag{2.17}$$

and, if $\sigma \neq 0$, then

$$\lim_{n\to\infty}\mathcal{L}\Big(\frac{1}{\sigma_n}\sum_{j=1}^n X_j\Big) = \lim_{n\to\infty}\mathcal{L}\Big(\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^n X_j\Big) = \mathcal{N}(0,1).$$

Proof. Using Lemma 2.4.4 we know (2.17) holds and if $\sigma^2 \neq 0$, then

$$\sigma_n^2 = \sigma^2 n\big(1 + o(1)\big), \tag{2.18}$$

and we conclude

$$\lim_{n\to\infty}\mathcal{L}\Big(\frac{S_n}{\sigma\sqrt{n}}\Big) = \lim_{n\to\infty}\mathcal{L}\Big(\frac{S_n}{\sigma_n}\Big).$$

The main idea of the proof builds on a clever decomposition of the sum S_n in multiple blocks of length p_n and q_n:

$$S_n = \underbrace{X_1 + \cdots + X_{p_n}}_{\xi_0} + \underbrace{X_{p_n+1} + \cdots + X_{p_n+q_n}}_{\eta_0} + \underbrace{X_{p_n+q_n+1} + \cdots + X_{2p_n+q_n}}_{\xi_1} + \cdots + X_n.$$

I.e. we make the decomposition

$$S_n = \sum_{i=0}^{k_n-1} \xi_i + \sum_{i=0}^{k_n} \eta_i = S_n' + S_n'',$$

where

$$\xi_i = \sum_{ip_n+iq_n+1}^{(i+1)p_n+iq_n} X_j, \qquad \eta_i = \sum_{(i+1)p_n+iq_n+1}^{(i+1)p_n+(i+1)q_n} X_j, \qquad \text{for} \quad i = 0, 1, \ldots, k_n - 1$$

and

$$\eta_{k_n} = \sum_{k_np_n+k_nq_n+1}^{n} X_j,$$

with $k_n = \lfloor \frac{n}{p_n+q_n} \rfloor$.

The motivation for this decomposition is, that if q_n is large the strong mixing condition ensures approximate independence of the ξ's. If furthermore p_n is large compared to q_n the η-terms become negligible and the sum is then over approximately independent terms yielding a CLT. The construction of the sequences p_n and q_n with properties ensuring a limiting behavior as outlined above is lengthy and therefore it is stated separately as a lemma in the appendix (Lemma A.3.2). In the following we will just sum up the properties of the sequences constructed in the lemma. For $n \to \infty$:

(i) $p_n \to \infty$, $p_n = o(n)$,

(ii) $q_n \to \infty$, $q_n = o(p_n)$,

(iii) $k_n = \frac{n}{p_n}\Big(1 + o(1)\Big)$,

(iv) $\frac{q_n^2 k_n}{np_n} = o(1)$,

(v) $k_n \alpha_X(q_n) = o(1)$,

(vi) $\frac{k_n}{\sigma_n^4}\mathbb{E}\Big[\Big(\sum_{j=1}^{p_n} X_j\Big)^4\Big] = o(1)$.

The corresponding decomposition of the normalized sum is

$$Z_n = \sigma_n^{-1} S_n = \sigma_n^{-1} S_n' + \sigma_n^{-1} S_n'' = Z_n' + Z_n''.$$

We will show that with the given properties of (p_n) and (q_n), the term Z_n'' converges to zero in probability as n grows. Then we can use the result by Slutsky (Lemma 2.1.7) to conclude that $Z_n \xrightarrow{\mathcal{D}} Z_n'$.

Since Z''_n is a sum of zero mean random variables it has mean zero and Chebyshev's inequality (A.2) yields

$$P(|Z''_n| \geq \varepsilon) \leq \frac{\mathbb{E}(|Z''_n|^2)}{\varepsilon^2}. \tag{2.19}$$

Therefore $Z''_n \xrightarrow{p} 0$ if

$$\lim_{n\to\infty} \mathbb{E}(|Z''_n|^2) = 0,$$

which we will show in the following by decomposing $\mathbb{E}(|Z''_n|^2)$ into four terms and showing each term is $o(1)$.

$$\begin{aligned}
\mathbb{E}(|Z''_n|^2) &= \mathbb{E}\Big[\Big(\frac{1}{\sigma_n}\sum_{i=0}^{k_n}\eta_i\Big)^2\Big] \\
&= \frac{1}{\sigma_n^2}\mathbb{E}\Big[\Big(\sum_{i=0}^{k_n-1}\eta_i + \eta_{k_n}\Big)^2\Big] \\
&= \frac{1}{\sigma_n^2}\Big[\mathbb{E}\Big(\sum_{i=0}^{k_n-1}\eta_i\Big)^2 + \mathbb{E}(\eta_{k_n}^2) + 2\sum_{i=0}^{k_n-1}\mathbb{E}(\eta_i\eta_{k_n})\Big] \\
&= \frac{1}{\sigma_n^2}\Big[k_n\mathbb{E}(\eta_0^2) + 2\sum_{j=1}^{k_n-1}(k_n-j)\mathbb{E}(\eta_0\eta_j) + \mathbb{E}(\eta_{k_n}^2) + 2\sum_{i=0}^{k_n-1}\mathbb{E}(\eta_i\eta_{k_n})\Big].
\end{aligned}$$

Using (2.18) we have

$$\frac{1}{\sigma_n^2} = \frac{1}{\sigma^2 n(1+o(1))} = \frac{\frac{1}{1+o(1)}}{\sigma^2 n} = \frac{1+o(1)}{\sigma^2 n}, \tag{2.20}$$

and we thus need to show, that

(I) $\frac{k_n}{n}\mathbb{E}(\eta_0^2) = o(1)$,

(II) $\frac{1}{n}\mathbb{E}(\eta_{k_n}^2) = o(1)$,

(III) $\frac{1}{n}\sum_{j=1}^{k_n-1}(k_n-j)\mathbb{E}(\eta_0\eta_j) = o(1)$,

(IV) $\frac{1}{n}\sum_{i=0}^{k_n-1}\mathbb{E}(\eta_i\eta_{k_n}) = o(1)$.

This relies mostly on the properties $(i)-(vi)$ listed above and we will refer the reader to this list in general rather than making a reference every time one of the properties is used.

Ad I:

Since η_0 is a sum of q_n X_j's we have

$$\mathbb{E}(\eta_0^2) = \mathbb{E}(S_{q_n}^2) = \sigma^2 q_n(1+o(1)),$$

and thus

$$\frac{k_n}{n}\mathbb{E}(\eta_0^2) = \frac{\frac{n}{p_n}}{n}\sigma^2 q_n(1+o(1)) = O\Big(\frac{q_n}{p_n}\Big) = o(1).$$

Ad II:

Since η_{k_n} is a sum of at most p_n+q_n X_j's we have

$$\mathbb{E}(\eta_{k_n}^2) \le \mathbb{E}(S_{p_n+q_n}^2) = \sigma^2(p_n+q_n)(1+o(1)),$$

such that

$$\frac{1}{n}\mathbb{E}(\eta_{k_n}^2) \le \frac{p_n+q_n}{n}\sigma^2(1+o(1)) = o(1).$$

Ad III:

For $j = 0, \ldots, k_n-1$ the variables η_j and η_{j+1} are separated by p_n terms, and since $P(|\eta_j| < Cq_n) = 1$, Theorem 2.3.2 yields

$$\begin{aligned}\big|\mathbb{E}(\eta_0\eta_j)\big| &= \big|\mathbb{E}(\eta_0\eta_j) - \mathbb{E}(\eta_0)\mathbb{E}(\eta_j)\big| \\ &\le 4C^2 q_n^2 \alpha_X(jp_n).\end{aligned}$$

Hereby we conclude

$$\begin{aligned}\Big|\frac{1}{n}\sum_{j=1}^{k_n-1}(k_n-j)\mathbb{E}(\eta_0\eta_j)\Big| &\le \frac{k_n}{n}\sum_{j=1}^{k_n-1}\big|\mathbb{E}(\eta_0\eta_j)\big| \\ &\le \frac{4C^2q_n^2k_n}{n}\sum_{j=1}^{k_n-1}\alpha_X(jp_n).\end{aligned}$$

Using (iii) in Lemma 2.3.4 yields

$$\begin{aligned}\Big|\frac{1}{n}\sum_{j=1}^{k_n-1}(k_n-j)\mathbb{E}(\eta_0\eta_j)\Big| &\le \frac{4C^2q_n^2k_n}{n}O(p_n^{-1}) \\ &= O\Big(\frac{q_n^2k_n}{np_n}\Big) = o(1).\end{aligned}$$

Ad IV:

This follows along very similar lines to the above. First Theorem 2.3.2 yields

$$\big|\mathbb{E}(\eta_j\eta_{k_n})\big| \le 4C^2 q_n(p_n+q_n)\alpha_X\big((k_n-j)p_n\big).$$

Consequently

$$\begin{aligned}\Big|\frac{1}{n}\sum_{j=1}^{k_n-1}\mathbb{E}(\eta_j\eta_{k_n})\Big| &\le \frac{4C^2q_n(p_n+q_n)}{n}\sum_{j=1}^{k_n-1}\alpha_X\big((k_n-j)p_n\big).\\ &\le O\Big(\frac{q_n(p_n+q_n)}{n}\Big)\sum_{j=1}^{k_n-1}\alpha_X(jp_n).\\ &\le O\Big(\frac{q_n(p_n+q_n)}{np_n}\Big).\\ &= O(\frac{q_n}{n}+\frac{q_n}{n}\frac{q_n}{p_n})\\ &= o(1).\end{aligned}$$

This finishes the verification of (2.19) such that $Z_n'' \xrightarrow{p} 0$ and by Slutsky's result (Lemma 2.1.7) we have

$$\lim_{n\to\infty}\mathcal{L}(Z_n) = \lim_{n\to\infty}\mathcal{L}(Z_n').$$

The rest of the proof is concerned with showing

$$\lim_{n\to\infty}\mathcal{L}(Z_n') = \mathcal{N}(0,1).$$

First we prove that

$$\big|\mathbb{E}[\exp(itZ_n')] - \phi_n(t)^{k_n}\big| \to 0 \qquad \text{for } n\to\infty, \tag{2.21}$$

where ϕ_n denotes the characteristic function of $\sigma_n^{-1}\xi_0$. For fixed t the complex random variable

$$\mathbb{E}[\exp\big(itZ_{n-1}'\big)] = \exp\Big(\frac{it}{\sigma_n}\sum_{j=0}^{k_n-2}\xi_j\Big)$$

is measurable with respect to $\mathcal{M}_{-\infty}^{(k_n-1)p_n+(k_n-2)q_n}(X)$ and bounded by 1. Similarly

$$\exp\left(\frac{it}{\sigma_n}\xi_{k_n-1}\right)$$

is measurable with respect to $\mathcal{M}^{\infty}_{(k_n-1)p_n+(k_n-1)q_n+1}(X)$ and is also bounded by 1. The complex-valued version of Theorem 2.3.2 yields

$$\Big|\mathbb{E}\big[\exp(itZ_n')\big] - \mathbb{E}\big[\exp(itZ_{n-1}')\big]\mathbb{E}\big[\exp\big(\tfrac{it}{\sigma_n}\xi_{k_n-1}\big)\big]\Big| \le 16\alpha_X(q_n+1) \le 16\alpha_X(q_n).$$

Using this recursively yields

$$
\begin{aligned}
\Big|\mathbb{E}[\exp(itZ_n')] - \phi_n(t)^{k_n}\Big| = &\Big|\mathbb{E}\Big[\exp\Big(itZ_n'\Big)\Big] - \mathbb{E}\big[\exp\big(itZ_{n-1}'\big)\big]\mathbb{E}\big[\exp\big(\tfrac{it}{\sigma_n}\xi_{k_n-1}\big)\big] \\
&+ \mathbb{E}\Big[\exp\big(itZ_{n-1}'\big)\big]\mathbb{E}\big[\exp\big(\tfrac{it}{\sigma_n}\xi_{k_n-1}\big)\big] - \phi_n(t)^{k_n}\Big| \\
\le &\Big|\mathbb{E}\big[\exp\big(itZ_n'\big)\big] - \mathbb{E}\big[\exp\big(itZ_{n-1}'\big)\big]\mathbb{E}\big[\exp\big(\tfrac{it}{\sigma_n}\xi_{k_n-1}\big)\big]\Big| \\
&+ \Big|\mathbb{E}\big[\exp\big(itZ_{n-1}'\big)\big]\phi_n(t) - \phi_n(t)^{k_n}\Big| \\
\le &16\alpha_X(q_n) + \Big|\mathbb{E}\big[\exp\big(itZ_{n-1}'\big)\big] - \phi_n(t)^{k_n-1}\Big|\Big|\phi_n(t)\Big| \\
\le &16\alpha_X(q_n) + \Big|\mathbb{E}\big[\exp\big(itZ_{n-1}'\big)\big] - \phi_n(t)^{k_n-1}\Big| \\
\le &2\cdot 16\alpha_X(q_n) + \Big|\mathbb{E}\big[\exp\big(itZ_{n-2}'\big)\big] - \phi_n(t)^{k_n-2}\Big| \\
&\vdots \\
\le &k_n 16\alpha_X(q_n).
\end{aligned}
$$

According to property (v) on page 29 this tends to zero for $n \to \infty$.

For the final part of the proof consider the triangular array given by the iid random variables

$$\xi_{nj}' \qquad (j = 1, \dots, k_n),$$

where $n = 1, 2, \dots$ and $\mathcal{L}(\xi_{nj}') = \mathcal{L}(\sigma_n^{-1}\xi_0)$. Then (2.21) yields

$$\lim_{n\to\infty} \mathcal{L}(Z_n') = \lim_{n\to\infty} \mathcal{L}(\widetilde{S}_{k_n}),$$

where $\widetilde{S}_{k_n} = \xi_{n1}' + \cdots + \xi_{nk_n}'$. By Theorem 2.1.10

$$\frac{\widetilde{S}_{k_n}}{\tilde{\sigma}_{k_n}} = \frac{\widetilde{S}_{k_n}}{\sqrt{\mathrm{Var}\big(\widetilde{S}_{k_n}\big)}} \xrightarrow{D} \mathcal{N}(0,1),$$

if the Lindeberg condition (2.3) is satisfied. In the following we will show this condition indeed is satisfied, and furthermore we will show $\tilde{\sigma}_{k_n} \to 1$ for $n \to \infty$ such that we can conclude

$$\lim_{n\to\infty} \mathcal{L}(\widetilde{S}_{k_n}) = \lim_{n\to\infty} \mathcal{L}\Big(\frac{\widetilde{S}_{k_n}}{\tilde{\sigma}_{k_n}}\Big) = \mathcal{N}(0,1).$$

First we observe that

$$\tilde{\sigma}_{k_n}^2 = \sum_{j=1}^{k_n} \mathrm{Var}(\xi_{nj}') = \frac{k_n}{\sigma_n^2}\mathrm{Var}\Big(\sum_{j=1}^{p_n} X_j\Big) = \frac{k_n\sigma_{p_n}^2}{\sigma_n^2}.$$

Using (2.18) and (iii) on page 29 yields

$$\tilde{\sigma}_{k_n}^2 = \frac{k_n p_n}{n}\big(1 + o(1)\big) = 1 + o(1).$$

We have thus proved the theorem if we can verify the Lindeberg condition

$$\sum_{j=1}^{k_n} \frac{1}{\tilde{\sigma}_{k_n}^2} \int_{|x| \geq \varepsilon \tilde{\sigma}_{k_n}} x^2 dP(X_{nj} \leq x) = o(1), \qquad \text{as} \ \ n \to \infty, \tag{2.22}$$

for all $\varepsilon > 0$. We introduce the notation $\tilde{\varepsilon} = \varepsilon(1 + o(1))$, and by direct calculation we have

$$\begin{aligned}
\sum_{j=1}^{k_n} \frac{1}{\tilde{\sigma}_{k_n}^2} \int_{|x| \geq \varepsilon \tilde{\sigma}_{k_n}} x^2 dP(\xi'_{nj} \leq x) &= k_n \int_{|x| \geq \varepsilon(1+o(1))} x^2 dP(\sigma_n^{-1}\xi_0 \leq x) \\
&= k_n \int_{|x| \geq \tilde{\varepsilon}} x^2 dP(\sigma_n^{-1}\xi_0 \leq x) \\
&= k_n \int_{|x| \geq \tilde{\varepsilon}} x^2 dP(\xi_0 \leq \sigma_n x) \\
&= k_n \int_{|x| \geq \tilde{\varepsilon}\sigma_n} \Big(\frac{x}{\sigma_n}\Big)^2 dP(\xi_0 \leq x) \\
&\leq \frac{k_n}{\sigma_n^2} \int_{|x| \geq \tilde{\varepsilon}\sigma_n} \frac{x^2}{(\tilde{\varepsilon}\sigma_n)^2} x^2 dP(\xi_0 \leq x) \\
&\leq \frac{k_n}{\tilde{\varepsilon}^2 \sigma_n^4} \int_{-\infty}^{\infty} x^4 dP(\xi_0 \leq x) \\
&= \frac{k_n}{\tilde{\varepsilon}^2 \sigma_n^4} \mathbb{E}\Big[\Big(\sum_{j=1}^{p_n} X_j\Big)^4\Big] \\
&= o(1),
\end{aligned}$$

where the last equality is given by property (vi) on page 29. We have thus verified the Lindeberg condition and the theorem is proved. ■

2.4.2 Continuous time central limit theorem

The following theorem extends Theorem 2.4.6 to continuous time.

Theorem 2.4.7. *Let the real stationary process* (X_t) *be strongly mixing, with*

$$\int_0^{\infty} \alpha_X(\tau) d\tau < \infty, \tag{2.23}$$

and let there exist a constant $C < \infty$ *such that* $P(|X_t| < C) = 1$*. Then*

$$\sigma^2 = 2\int_0^\infty \mathbb{E}(X_0 X_t)dt < \infty,$$

and if $\sigma \neq 0$*, then*

$$\frac{1}{\sigma\sqrt{T}}\int_0^T X_t dt \xrightarrow{D} \mathcal{N}(0,1).$$

Proof. From Lemma 2.4.5 we know that

$$\sigma_T^2 = \sigma^2 T(1 + o(1)).$$

such that

$$\lim_{T\to\infty} \mathcal{L}\Big(\frac{S_T}{\sigma\sqrt{T}}\Big) = \lim_{T\to\infty} \mathcal{L}\Big(\frac{S_T}{\sigma_T}\Big).$$

To prove, that this limiting distribution is standard Gaussian we introduce the stationary process (ξ_j) given by

$$\xi_j = \int_{j-1}^{j} X_t dt, \tag{2.24}$$

which we will show is strongly mixing in the following.

We consider the process (X_t), $t \in (-\infty, j]$, which is assumed to be stochastically continuous and by Theorem 2.2.2 and the remarks thereafter we can assume it is measurable. Thus this process is measurable with respect to the product σ-algebra $\mathcal{M}_{-\infty}^j(X) \otimes \mathbb{B}\big((-\infty, j]\big)$. By the extension to the first part of Tonelli's Theorem (Theorem A.2.1) we have for all $i \leq j$ that the function $\xi_i : \Omega \to \mathbb{R}$ defined by (2.24) is measurable with respect to $\mathcal{M}_{-\infty}^j(X)$. Since $\mathcal{M}_{-\infty}^j(\xi)$ denotes the minimal σ-algebra which ξ_i, $i \leq j$ all are measurable with respect to, we have

$$\mathcal{M}_{-\infty}^j(\xi) \subseteq \mathcal{M}_{-\infty}^j(X).$$

Using similar arguments we conclude

$$\mathcal{M}_k^\infty(\xi) \subseteq \mathcal{M}_k^\infty(X).$$

Denoting the strong mixing coefficient of (ξ_j) by α_ξ we have

$$\alpha_\xi(n) \leq \alpha_X(n),$$

for all n. This implies that strong mixing of (X_t) ensures strong mixing of (ξ_j) and by the integral test for convergence of series, condition (2.23) implies

$$\sum_{n=1}^{\infty} \alpha_\xi(n) \leq \sum_{n=1}^{\infty} \alpha_X(n) < \infty.$$

Obviously $P(|\xi_j| < C) = 1$ and by Theorem 2.4.6 (ξ_j) satisfies the CLT. I.e. for $T \to \infty$ we have

$$Z_{\lfloor T \rfloor} = \frac{S_{\lfloor T \rfloor}}{\sigma_{\lfloor T \rfloor}} = \sigma_{\lfloor T \rfloor}^{-1} \sum_{j=0}^{\lfloor T \rfloor} \xi_j = \sigma_{\lfloor T \rfloor}^{-1} \int_0^{\lfloor T \rfloor} X_t dt \xrightarrow{D} \mathcal{N}(0,1),$$

where

$$\sigma_{\lfloor T \rfloor}^2 = \mathrm{Var}\Big(S_{\lfloor T \rfloor}\Big) = \mathrm{Var}\Big(\int_0^{\lfloor T \rfloor} X_t dt\Big).$$

For the final part of the proof we make the decomposition

$$Z_T = Z_{\lfloor T \rfloor} + (Z_T - Z_{\lfloor T \rfloor}),$$

where

$$Z_T = \sigma_T^{-1} \int_0^T X_t dt \qquad \text{and} \qquad \sigma_T^2 = \mathrm{Var}\Big(\int_0^T X_t dt\Big).$$

By Slutsky's result (Lemma 2.1.7) the theorem is proved if $|Z_{\lfloor T \rfloor} - Z_T| \xrightarrow{p} 0$. As in the proof of Theorem 2.4.6 Chebyshev's inequality ensures, that it is sufficient to show mean square convergence to 0 since

$$P(|Z_{\lfloor T \rfloor} - Z_T| \geq \varepsilon) \leq \frac{\mathbb{E}(|Z_{\lfloor T \rfloor} - Z_T|^2)}{\varepsilon^2}. \tag{2.25}$$

In the following we use the general rule $(a-b)^2 \leq 2a^2 + 2b^2$ and $P(|X_t| < C) = 1$, which yields

$$\begin{aligned}
\mathbb{E}(|Z_{\lfloor T \rfloor} - Z_T|^2) &= \mathbb{E}\Big(\Big|\sigma_{\lfloor T \rfloor}^{-1} \int_0^{\lfloor T \rfloor} X_t dt - \sigma_T^{-1} \int_0^T X_t dt\Big|^2\Big) \\
&= \mathbb{E}\Big(\Big|(\sigma_{\lfloor T \rfloor}^{-1} - \sigma_T^{-1}) \int_0^{\lfloor T \rfloor} X_t dt - \sigma_T^{-1} \int_{\lfloor T \rfloor}^T X_t dt\Big|^2\Big) \\
&\leq 2(\sigma_{\lfloor T \rfloor}^{-1} - \sigma_T^{-1})^2 \mathbb{E}\Big[\Big(\int_0^{\lfloor T \rfloor} X_t dt\Big)^2\Big] + 2\sigma_T^{-2} \mathbb{E}\Big[\Big(\int_{\lfloor T \rfloor}^T X_t dt\Big)^2\Big] \\
&\leq 2\Big(1 - \frac{\sigma_{\lfloor T \rfloor}}{\sigma_T}\Big)^2 + 2\sigma_T^{-2} C^2.
\end{aligned}$$

To ensure convergence of this expression to 0 it suffices to show

$$\sigma_T^2 = \sigma_{\lfloor T \rfloor}^2 \big(1 + o(1)\big),$$

which is verified by direct calculation

$$\begin{aligned}\sigma_T^2 &= \mathbb{E}\Big[\Big(\int_0^{\lfloor T\rfloor} X_t dt + \int_{\lfloor T\rfloor}^T X_t dt\Big)^2\Big]\\ &= \sigma_{\lfloor T\rfloor}^2 + 2\mathbb{E}\Big(\int_0^{\lfloor T\rfloor} X_t dt \int_{\lfloor T\rfloor}^T X_t dt\Big) + \mathbb{E}\Big[\Big(\int_{\lfloor T\rfloor}^T X_t dt\Big)^2\Big]\\ &= \sigma_{\lfloor T\rfloor}^2\big(1+o(1)\big).\end{aligned}$$

■

The final result of this section extends the theorem above to hold for a transformation of a strongly mixing stochastic process.

Corollary 2.4.8. *Let the stationary stochastic process (X_t) be strongly mixing, with mixing coefficient α_X, and let the stationary process (Y_t) be defined by $Y_t = f(Y_t)$, where $f : \mathbb{C} \to \mathbb{R}$ is measurable. If there exists a constant $C < \infty$, such that $P(|Y_0| < C) = 1$, and if*

$$\int_0^\infty \alpha_X(\tau)d\tau < \infty.$$

Then

$$\sigma^2 = 2\int_0^\infty \mathbb{E}(Y_0 Y_t)dt < \infty,$$

and if $\sigma \neq 0$, then

$$\frac{1}{\sigma\sqrt{T}}\int_0^T Y_t dt \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

for $T \to \infty$.

Proof. Using Lemma 2.3.3 we know mixing properties are conserved under measurable mappings and all the conditions of Theorem 2.4.7 are thus satisfied. ■

Before we can apply the central limit theorem to the problem of Chapter 1 we will need a concept of convergence for stochastic processes rather than for single random variables as described in Section 2.1.2, and the following generalizes weak convergence to cover entire stochastic processes.

2.5 Weak convergence

The classical reference on weak convergence is Billingsley (1968) where an in-depth description of weak convergence can be found, and it is the basis of this section along with van der Vaart and Wellner (1996). The latter has a very general approach to weak convergence in the sense that non-measurable sequences of random variables are allowed. This complicates the topic unnecessarily for our purposes, and when referring to results we will always give a simplified version only covering measurable random variables. Furthermore van der Vaart and Wellner (1996) is mainly focused on empirical distribution functions, and therefore the results are only stated for real-valued stochastic processes, we will however state the results for complex-valued processes.

We let $\mathbb{D}$ be a metric space with metric d and denote the set of all continuous, bounded functions $f : \mathbb{D} \to \mathbb{R}$ by $C_b(\mathbb{D})$.

Definition 2.5.1. Let $(\Omega_n, \mathcal{E}_n, P_n)$, $n = 1, 2, \dots$ be a sequence of probability spaces and $X_n : \Omega_n \to \mathbb{D}$ measurable maps. The sequence (X_n) converges weakly to a Borel measure L if

$$\mathbb{E}[f(X_n)] \to \int f dL, \qquad \text{for every } f \in C_b(\mathbb{D}).$$

This is denoted $X_n \xrightarrow{D} L$, and if $L = \mathcal{L}(X)$ for some random variable X we also write $X_n \xrightarrow{D} X$. □

Using Lemma 2.1.5 it is noticed that for $\mathbb{D} = \mathbb{R}$ with metric $d(x, y) = |x - y|$ this is equivalent to conventional weak convergence of random variables (Definition 2.1.2). An important concept in the study of weak convergence is tightness.

Definition 2.5.2. A Borel probability measure P is called tight if for every $\varepsilon>0$ there exists a compact set K with $P(K) \geq 1 - \varepsilon$. Correspondingly we say that $X : \Omega \to \mathbb{D}$ is tight if $\mathcal{L}(X)$ is tight. The sequence (X_n) is asymptotically tight if for every $\varepsilon>0$ there exists a compact set K such that

$$\liminf_{n\to\infty} P(X_n \in K^\delta) \geq 1 - \varepsilon, \qquad \text{for every } \delta > 0,$$

where $K^\delta = \{y \in \mathbb{D} \mid d(y, K) < \delta\}$ is the δ-enlargement around K. □

A useful result on asymptotic tightness is (van der Vaart and Wellner, 1996, Lemma 1.3.8).

Lemma 2.5.3. *If $X_n \xrightarrow{D} X$, then (X_n) is asymptotically tight if and only if X is tight.*

A related concept is separability and we say that X or $L = \mathcal{L}(X)$ is separable if there exists a separable, measurable set A with $P(X \in A) = 1$. An important lemma states that tightness and separability are equivalent in complete metric spaces (van der Vaart and Wellner, 1996, Lemma 1.3.2).

Lemma 2.5.4. *On a complete metric space separability and tightness are equivalent.*

As mentioned earlier a stochastic process is a map $X : T \times \Omega \to \mathbb{C}$, where T is the index set. In this context it is convenient to denote a stochastic process simply by X and write $X(t, \omega)$ for a realization at time t. As before we usually let the ω-dependence be implicit and write $X(t)$ for the marginal of X at time t. For fixed ω we have a sample path $X(\cdot,\omega) : T \to \mathbb{C}$, and often we have prior knowledge of some properties of the sample paths. In our applications it will turn out that the sample paths are uniformly bounded such that the stochastic process is a mapping $X : \Omega \to \ell^\infty_{\mathbb{C}}(T)$, where $\ell^\infty_{\mathbb{C}}(T)$ denotes the set of all uniformly bounded complex-valued functions on T. In the real case the corresponding space is denoted $\ell^\infty_{\mathbb{R}}(T)$, and when we state results covering both cases we simply write $\ell^\infty(T)$. Equipped with the metric

$$d(x_1, x_2) = \|x_1 - x_2\|_T = \sup_{t\in T} |x_1(t) - x_2(t)| \tag{2.26}$$

$\ell^\infty(T)$ is a metric space, and we can apply the weak convergence of definition 2.5.1 to these processes. Before studying weak convergence in this space we introduce a much weaker concept of convergence.

Definition 2.5.5. A sequence of stochastic processes (X_n) is said to converge in finite dimensional distribution to a process X if for all $t_1, \dots, t_k$

$$\big(X_n(t_1), \dots, X_n(t_k)\big) \xrightarrow{\mathcal{D}} \big(X(t_1), \dots, X(t_k)\big).$$

This is denoted $X_n \overset{\text{fidi}}{\to} X$. □

If we have convergence in finite dimensional distribution and the process is suitably regular we also have weak convergence in $\ell^\infty(T)$, which the next theorem formalizes (van der Vaart and Wellner, 1996, Theorem 1.5.4).

Theorem 2.5.6. *Let (X_n) be a sequence of stochastic processes $X_n : \Omega_n \to \ell^\infty(T)$. Then (X_n) converges weakly to a tight limit X if and only if (X_n) is asymptotically tight and $X_n \overset{fidi}{\to} X$. If (X_n) is asymptotically tight and $X_n \overset{fidi}{\to} X$ for a stochastic process X, then there is a version of X with uniformly bounded sample paths and $X_n \xrightarrow{\mathcal{D}} X$.*

Asymptotic tightness is thus essential for weak convergence in $\ell^\infty(T)$ and the following theorem establishes a method for checking asymptotic tightness (van der Vaart and Wellner, 1996, Theorem 1.5.7).

Theorem 2.5.7. *Let (X_n) be a sequence of stochastic processes $X_n : \Omega_n \to \ell^\infty(T)$. Then (X_n) is asymptotically tight if and only if $X_n(t)$ is asymptotically tight for every t and there exists a semimetric ρ on T such that the semimetric space (T, ρ) is totally bounded and (X_n) is asymptotically uniformly ρ-equicontinuous in probability.*

Asymptotic uniform ρ-equicontinuity in probability is defined as follows.

Definition 2.5.8. Let (T, ρ) be a semimetric space. A sequence of stochastic processes $X_n : \Omega_n \to \ell^\infty(T)$ is asymptotically uniformly ρ-equicontinuous in probability if for every $\varepsilon, \eta > 0$ there exists a $\delta > 0$ such that

$$\lim_{n\to\infty} P\Big(\sup_{\rho(s,t)<\delta} |X_n(s) - X_n(t)| > \varepsilon \Big) < \eta.$$

□

It will turn out that it is more convenient to study the convergence in a slightly different space for our applications. We let $T_1 \subseteq T_2, \dots$ be compact sets such that $\cup_{i=1}^{\infty} T_i = T$. The set of all real- or complex-valued functions uniformly bounded on each T_i is denoted $\ell^\infty(T_1, T_2, \dots)$. This is a complete metric space if we equip it with the metric

$$d(x_1, x_2) = \sum_{i=1}^{\infty} \min\{\|x_1 - x_2\|_{T_i}, 1\} 2^{-i},$$

where $\|\cdot\|_{T_i}$ denotes the supremum norm on T_i as defined in (2.26). The following theorem links convergence in $\ell^\infty(T_1, T_2, \dots)$ to convergence in $\ell^\infty(T_i)$ for all i (van der Vaart and Wellner, 1996, Theorem 1.6.1).

Theorem 2.5.9. *Let $X_n : \Omega_n \to \ell^\infty(T_1, T_2, \dots)$, $n = 1, 2, \dots$ be a sequence of stochastic processes. Then X_n converges weakly to a tight limit if and only if for all i the sequence $X_{n|T_i} : \Omega_n \to \ell^\infty(T_i)$ converges weakly to a tight limit.*

Combining these three theorems with Lemma 2.5.3 we obtain the following corollary.

Corollary 2.5.10. *Let $X_n : \Omega_n \to \ell^\infty(T_1, T_2, \dots)$, $n = 1, 2, \dots$ be a sequence of stochastic processes. Then (X_n) converges weakly to a tight limit X if and only if $X_n \overset{fidi}{\to} X$, $X(t)$ is tight for all $t \in T$, and for all T_i, $i \in \mathbb{N}$ there exists a semimetric ρ on T_i such that the semimetric space (T_i, ρ) is totally bounded and (X_n) is asymptotically uniformly ρ-continuous in probability on T_i.*

In our applications we let $T_i = [-i, i]$ and $\rho(s,t) = |s - t|$. In this setup (T_i, ρ) is always bounded, and we need not check this condition. We conclude the results on weak convergence with the following theorem due to Skorokhod relating weak convergence to almost sure convergence.

Theorem 2.5.11. *Let $X_n : \Omega_n \to \mathbb{D}$, $n \in \mathbb{N}$ be measurable maps. If $X_n \overset{D}{\to} X$ and X is separable, then there exists measurable maps $\tilde{X}_n : \tilde{\Omega} \to \mathbb{D}$ and $\tilde{X} : \tilde{\Omega} \to \mathbb{D}$ all defined on some probability space $(\tilde{\Omega}, \tilde{\mathcal{E}}, \tilde{P})$ with*

(i) $\tilde{X}_n \overset{a.s.}{\to} \tilde{X}$

(ii) $\mathcal{L}(\tilde{X}_n) = \mathcal{L}(X_n)$ *for all* $n \in \mathbb{N}$ *and* $\mathcal{L}(\tilde{X}) = \mathcal{L}(X)$.

CHAPTER 3

Results

In this chapter the specific results obtained for the problem introduced in Chapter 1 are presented. In Section 3.1 the relevant stochastic processes are presented and studied. Section 3.2 treats the asymptotic properties of the MMSE for an infinite equalizer for $L, B \to \infty$ whereas Section 3.3 presents the results for the finite equalizer MMSE for $L, B, N \to \infty$. Finally Section 3.4 discusses these results and especially the imposed assumptions are reviewed.

When we let L and B grow simultaneously we will consider $B = B(L)$ as a function of L, and analogously we let $N = N(L)$. Usually it will be too cumbersome to write $B(L)$ in the notation and we will simply write B.

3.1 Weak convergence

First we recall the basic concepts from Chapter 1. A power delay profile (PDP) is a non-negative real valued function $p \in L_2(\mathbb{R})$ satisfying $p(t) = 0$ for $t < 0$ and

$$\int_{-\infty}^{\infty} p(t)dt = 1.$$

Let $X_0, X_1, \ldots, X_{L-1}$ be iid complex-valued stochastic variables with $\mathbb{E}(X_j) = 0$ and $\mathrm{Var}(X_j) = 1$. Then the channel response of a communication system using bandwidth B consists of the independent variables $h_l^B = \sigma_l^B X_l$, $l = 0, \ldots, L-1$, where

$$\sigma_l^B = \sqrt{\nu_l^B} \qquad \text{and} \qquad \nu_l^B = \mathrm{Var}(h_l^B) = \int_{\frac{l}{B}}^{\frac{l+1}{B}} p(t)dt, \tag{3.1}$$

for a given PDP p. Furthermore $\mathcal{L}(X_l)$ is called the channel distribution and often $\mathbb{C}\mathcal{N}(0,1)$ is used. With these definitions we can introduce two of the central stochastic processes of this chapter. The first is the discrete time Fourier transform of the channel response called the transfer function of the channel. That is, given a channel response $(h_0^B, \ldots, h_{L-1}^B)$ sampled with bandwidth B we introduce the complex-valued stochastic process $H^{L,B}$ determined by

$$H^{L,B}(f) = \sum_{l=0}^{L-1} h_l^B \exp\left(-2\pi i f \tfrac{l}{B}\right), \tag{3.2}$$

for all $f \in \mathbb{R}$. To define the other process we recall the expression for $\mathrm{MMSE}_\infty^{L,B}$ given in Chapter 1,

$$\mathrm{MMSE}_\infty^{L,B} = \tfrac{1}{B} \int_0^B \frac{\sigma_n^2}{|H^{L,B}(f)|^2 + \frac{1}{\rho}} df. \tag{3.3}$$

From this we define the stochastic process $Y^{L,B}$ as

$$Y^{L,B}(f) = \frac{\sigma_n^2}{|H^{L,B}(f)|^2 + \frac{1}{\rho}} \tag{3.4}$$

for all $f \in \mathbb{R}$.

In the following we derive the autocovariance function of $H^{L,B}$, denoted $R_H^{L,B}$. Using the independence of h_l^B and h_k^B for $l \neq k$, we have

$$\begin{aligned} R_H^{L,B}(\tilde{f}, f) &= \mathbb{E}[H^{L,B}(\tilde{f}+f)\overline{H^{L,B}(\tilde{f})}] \\ &= \mathbb{E}\left(\left[\sum_{l=0}^{L-1} h_l^B \exp\left(-2\pi i(\tilde{f}+f)\tfrac{l}{B}\right)\right]\left[\sum_{l=0}^{L-1} \overline{h_l^B} \exp\left(2\pi i\tilde{f}\tfrac{l}{B}\right)\right]\right) \\ &= \sum_{l=0}^{L-1} \mathbb{E}(h_l^B\overline{h_l^B}) \exp\left(-2\pi i f\tfrac{l}{B}\right) \\ &= \sum_{l=0}^{L-1} \int_{\frac{l}{B}}^{\frac{l+1}{B}} p(t) \exp\left(-2\pi i f\tfrac{l}{B}\right) dt. \end{aligned} \tag{3.5}$$

Since $R_H^{L,B}(\tilde{f}, f) = R_H^{L,B}(f)$ only depends on f, $H^{L,B}$ is stationary in the wide sense, and we will show that this autocovariance function has a pointwise limit R_H^{∞} as $L \to \infty$. From (3.5) we conjecture that $R_H^{\infty}(f) = \hat{p}(f)$, where $\hat{p}$ is the Fourier transform of the PDP.

Lemma 3.1.1. *Let $\hat{p}$ denote the Fourier transform of the PDP, p, such that*

$$\hat{p}(f) = \int_0^{\infty} p(t) \exp(-i2\pi f t) dt,$$

and let

$$R_H^{L,B}(f) = \sum_{l=0}^{L-1} \int_{\frac{l}{B}}^{\frac{l+1}{B}} p(t) \exp\left(-2\pi i f\tfrac{l}{B}\right) dt.$$

If $L, B(L) \to \infty$ such that $L/B \to \infty$, then $R_H^{L}(f) \to \hat{p}(f)$ for all f.

Proof. First we decompose $R_H^{L,B}$ as

$$\begin{aligned} \lim_{L\to\infty} R_H^{L,B}(f) = \lim_{L\to\infty} \sum_{l=0}^{L-1} \int_{\frac{l}{B}}^{\frac{l+1}{B}} p(t) \exp(-i2\pi f t) dt \\ + \lim_{L\to\infty} \sum_{l=0}^{L-1} \int_{\frac{l}{B}}^{\frac{l+1}{B}} p(t)[\exp\left(-2\pi i f\tfrac{l}{B}\right) - \exp(-i2\pi f t)] dt \end{aligned}$$

For the first limiting term we have

$$\lim_{L\to\infty}\sum_{l=0}^{L-1}\int_{\frac{l}{B}}^{\frac{l+1}{B}} p(t)\exp(-i2\pi ft)dt = \lim_{L\to\infty}\int_0^{\frac{L}{B}} p(t)\exp(-i2\pi ft)dt = \hat{p}(f),$$

where we used $L/B \to \infty$. We thus only need to verify the second limit is 0. In each integral we have a difference of the form $[\exp(-i2\pi fs) - \exp(-i2\pi ft)]$, where $|s-t| \leq 1/B$. The Lipschitz continuity of the complex exponential function (see Lemma A.3.3) ensures that for all $\varepsilon{>}0$ exists L_0 such that for $L{>}L_0$

$$|\exp\left(-2\pi i f\tfrac{l}{B}\right) - \exp(-i2\pi ft)| < \varepsilon \qquad \text{for all } l.$$

We thus have for $L > L_0$

$$\begin{aligned}\sum_{l=0}^{L-1}\int_{\frac{l}{B}}^{\frac{l+1}{B}} p(t)|\exp\left(-2\pi i f\tfrac{l}{B}\right) - \exp(-i2\pi ft)|dt &\leq \varepsilon\sum_{l=0}^{L-1}\int_{\frac{l}{B}}^{\frac{l+1}{B}} p(t)dt \\ &\leq \varepsilon\int_0^{\infty} p(t)dt \\ &= \varepsilon.\end{aligned}$$

This verifies that the limit of this term is 0 and the result is proved. ■

We use this lemma to show weak convergence of the finite dimensional marginals in the following proposition, if the channel distribution is complex Gaussian, which we will assume in the rest of the thesis.

Proposition 3.1.2. *Assume the channel distribution is $\mathbb{C}\mathcal{N}(0,1)$. Let $H^{L,B}$ be given by* (3.2) *and let H^∞ be a stationary zero-mean complex Gaussian stochastic process with autocovariance function $R_H^\infty = \hat{p}$, then $H^{L,B} \overset{fidi}{\to} H^\infty$.*

Proof. First we remark that the existence of a stationary Gaussian process with autocovariance function $\hat{p}$ follows from Kolmogorov's extension theorem (Ibragimov and Linnik, 1971, Page 292).

For all finite collections $f_1, \ldots, f_k$ we have by the definition of $H^{L,B}$

$$\begin{bmatrix} H^{L,B}(f_1) \\ \vdots \\ H^{L,B}(f_k)\end{bmatrix} = \begin{bmatrix} 1 & \exp\left(-2\pi i f_1\frac{1}{B}\right) & \ldots & \exp\left(-2\pi i f_1\frac{L-1}{B}\right) \\ \vdots & \vdots & \ldots & \vdots \\ 1 & \exp\left(-2\pi i f_k\frac{1}{B}\right) & \ldots & \exp\left(-2\pi i f_k\frac{L-1}{B}\right)\end{bmatrix}\begin{bmatrix} h_0^B \\ h_1^B \\ \vdots \\ h_{L-1}^B\end{bmatrix}.$$

Since the channel distribution is assumed to be complex Gaussian the entries of the vector $(h_0^B, \ldots, h_{L-1}^B)^\top$ are independent complex Gaussians and the vector is thus multivariate complex Gaussian. As all finite dimensional marginals of $H^{L,B}$ are a linear transformation of a multivariate complex Gaussian vector the finite dimensional marginals are themselves multivariate Gaussian. Notationally we write

$$\big(H^{L,B}(f_1),\dots,H^{L,B}(f_k)\big)^\top \sim \mathbb{C}\mathcal{N}_k(\mathbf{0},\Sigma^{L,B}), \tag{3.6}$$

where $\Sigma^{L,B}=[\sigma_{ij}^{L,B}]$. The elements of the covariance matrix are determined from the autocovariance function by $\sigma_{ij}^{L,B}=R_H^{L,B}(f_i-f_j)$. The characteristic function for the random vector given in (3.6) is (Andersen et al., 1995, Theorem 2.7)

$$\phi^{L,B}(\boldsymbol{w})=\exp\Big(-\frac{\boldsymbol{w}^*\Sigma^{L,B}\boldsymbol{w}}{4}\Big),\qquad \boldsymbol{w}\in\mathbb{C}^n.$$

Since $R_H^{L,B}(f)\to R_H^{\infty}(f)$ for all f we have element-wise convergence of the matrices $\Sigma^{L,B}\to\Sigma$, where $\Sigma=[\sigma_{ij}]$, with $\sigma_{ij}=R_H^{\infty}(f_i-f_j)$. This implies

$$\lim_{L\to\infty}\phi^{L,B}(\boldsymbol{w})=\phi(\boldsymbol{w}),\qquad \text{for all } \boldsymbol{w}\in\mathbb{C}^n,$$

where

$$\phi(\boldsymbol{w})=\exp\Big(-\frac{\boldsymbol{w}^*\Sigma\boldsymbol{w}}{4}\Big),\qquad \boldsymbol{w}\in\mathbb{C}^n.$$

This is the characteristic function for

$$\big(H^{\infty}(f_1),\dots,H^{\infty}(f_k)\big)^\top.$$

We have thus proved convergence of the characteristic function for the finite dimensional marginals of $H^{L,B}$ to the characteristic function for the finite dimensional marginals of H^∞. Since convergence of the characteristic function implies weak convergence we have established $H^{L,B}\overset{\text{fidi}}{\to}H^\infty$. ■

It is remarked that $\hat{p}(0)=\int_0^\infty p(t)dt=1$ such that for all $f\in\mathbb{R}$ we have $H^\infty(f)\sim\mathbb{C}\mathcal{N}(0,1)$. Since weak convergence is conserved under continuous mappings we have $Y^{L,B}\overset{\text{fidi}}{\to}Y^\infty$, where

$$Y^\infty(f)=\frac{\sigma_n^2}{|H^\infty(f)|^2+\frac{1}{\rho}}.$$

The processes H^∞ and Y^∞ are candidates for weak limits of $H^{L,B}$ and $Y^{L,B}$ respectively in the sense of Section 2.5, which we will investigate in the following. The sample paths of $Y^{L,B}$ are real-valued, bounded, and continuous functions, whereas the sample paths of $H^{L,B}$ are seen to be complex-valued, continuous, and also bounded since

$$\begin{aligned}\sup_{f\in\mathbb{R}}|H^{L,B}(f)| &\le \sup_{f\in\mathbb{R}}\sum_{l=0}^{L-1}|h_l^B||\exp\big(-2\pi i f\tfrac{l}{B}\big)| \\ &=\sum_{l=0}^{L-1}|h_l^B|<\infty.\end{aligned}$$

The two stochastic processes are thus mappings from Ω to respectively $\ell^\infty_{\mathbb{C}}(T_1, T_2, \dots)$ and $\ell^\infty_{\mathbb{R}}(T_1, T_2, \dots)$, where $T_i = [-i, i]$. From Section 2.5 we know the key to showing weak convergence in these spaces is asymptotic uniform ρ-equicontinuity in probability, which we give conditions for in the following two propositions.

Proposition 3.1.3. *Let $H^{L,B} : \Omega_L \to \ell^\infty_{\mathbb{C}}(T_1, T_2, \dots)$, $L = 1, 2, \dots$ be a sequence of stochastic processes defined by (3.2), and let ρ be the usual metric on $\mathbb{R}$ defined by $\rho(s,t) = |s - t|$. If*

$$\lim_{L\to\infty} \sum_{l=1}^{L-1} \frac{l}{B(L)} \sigma_l^B = C < \infty, \tag{3.7}$$

where σ_l^B is defined in (3.1), then $H^{L,B}$ is asymptotically uniformly ρ-equicontinuous in probability on all T_i. I.e. for every $\varepsilon, \eta > 0$ there exists a $\delta > 0$ such that

$$\lim_{L\to\infty} P\Big(\sup_{\rho(s,t)<\delta} |H^{L,B}(s) - H^{L,B}(t)| > \varepsilon \Big) < \eta, \qquad s, t \in T_i.$$

Proof. Firstly we use the Lipschitz continuity of the complex exponential function (Lemma A.3.3) to show

$$\begin{aligned}
\sup_{\rho(s,t)<\delta} |H^{L,B}(s) - H^{L,B}(t)| &= \sup_{|s-t|<\delta} \Big| \sum_{l=0}^{L-1} h_l^B \Big[\exp\big(-2\pi i s \tfrac{l}{B}\big) - \exp\big(-2\pi i t \tfrac{l}{B}\big) \Big] \Big| \\
&\le \sup_{|s-t|<\delta} \sum_{l=0}^{L-1} \big|h_l^B\big| \Big| \exp\big(-2\pi i s \tfrac{l}{B}\big) - \exp\big(-2\pi i t \tfrac{l}{B}\big) \Big| \\
&\le \sum_{l=0}^{L-1} \big|h_l^B\big| \frac{2\pi l}{B(L)} \delta.
\end{aligned}$$

From this we conclude

$$P\Big(\sup_{\rho(s,t)<\delta} |H^{L,B}(s) - H^{L,B}(t)| > \varepsilon \Big) \le P\Big(\sum_{l=0}^{L-1} \big|h_l^B\big| \frac{2\pi l}{B(L)} \delta > \varepsilon \Big),$$

and by Markov's inequality (A.1) we have

$$\begin{aligned}
P\Big(\sum_{l=0}^{L-1} \big|h_l^B\big| \frac{2\pi l}{B(L)} \delta > \varepsilon \Big) &\le \frac{1}{\varepsilon} \mathbb{E}\Big(\sum_{l=0}^{L-1} \big|h_l^B\big| \frac{2\pi l}{B(L)} \delta \Big) \\
&= \frac{2\pi\delta}{\varepsilon} \sum_{l=0}^{L-1} \frac{l}{B(L)} \mathbb{E}\big(\big|h_l^B\big|\big).
\end{aligned}$$

We recall from the definition of h_l^B, that $h_l^B = \sigma_l^B X_l$, where $X_0, X_1, \dots$ are iid complex-valued stochastic variables with $\mathbb{E}(X_l) = 0$ and $\mathrm{Var}(X_l) = 1$. Introducing the notation $M = \mathbb{E}(|X_l|)$ we thus have

$$P\bigg(\sum_{l=0}^{L-1} |h_l^B| \frac{2\pi l}{B(L)} \delta > \varepsilon\bigg) \le \frac{2M\pi\delta}{\varepsilon} \sum_{l=0}^{L-1} \frac{l}{B(L)} \sigma_l^B.$$

It is clear that if (3.7) is satisfied, then the condition for asymptotic uniform ρ-equicontinuity in probability holds by choosing δ such that

$$\frac{2MC\pi\delta}{\varepsilon} <$$ [redacted]

Proposition 3.1.4. *Let $Y^{L,B} : \Omega_L \to \ell^\infty_{\mathbb{R}}(T_1, T_2, \dots)$, $L = 1, 2, \dots$ be a sequence of stochastic processes defined by* (3.4)*, and let ρ be the usual metric on $\mathbb{R}$. If*

$$\lim_{L\to\infty} \sum_{l=1}^{L-1} \sum_{k=1}^{L-1} \frac{|l-k|}{B(L)} \sigma_l^B \sigma_k^B = C < \infty, \tag{3.8}$$

then $Y^{L,B}$ is asymptotically uniformly ρ-equicontinuous in probability on all T_i. I.e. for every $\varepsilon, \eta > 0$ there exists a $\delta > 0$ such that

$$\lim_{L\to\infty} P\bigg(\sup_{\rho(s,t)<\delta} |Y^{L,B}(s) - Y^{L,B}(t)| > \varepsilon\bigg) < \eta, \qquad s, t \in T_i.$$

Proof. The proof is similar to that of Proposition 3.1.3 so we do the different parts in detail and skip quickly through the parts completely analogous to the previous proof.

$$\begin{aligned}
|Y^{L,B}(s) - Y^{L,B}(t)| &= \bigg| \frac{\sigma_n^2}{|H^{L,B}(s)|^2 + \frac{1}{\rho}} - \frac{\sigma_n^2}{|H^{L,B}(t)|^2 + \frac{1}{\rho}} \bigg| \\
&= \bigg| \frac{\sigma_n^2 \big(|H^{L,B}(t)|^2 - |H^{L,B}(s)|^2\big)}{(|H^{L,B}(s)|^2 + \frac{1}{\rho})(|H^{L,B}(t)|^2 + \frac{1}{\rho})} \bigg| \\
&\le \sigma_n^2 \rho^2 \Big| |H^{L,B}(t)|^2 - |H^{L,B}(s)|^2 \Big|.
\end{aligned}$$

Since

$$\begin{aligned}
|H^{L,B}(t)|^2 &= \bigg(\sum_{l=0}^{L-1} h_l^B \exp\big(-2\pi i t \tfrac{l}{B}\big)\bigg) \bigg(\sum_{k=0}^{L-1} \overline{h_k^B} \exp\big(2\pi i t \tfrac{k}{B}\big)\bigg) \\
&= \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} h_l^B \overline{h_k^B} \exp\big(-2\pi i t \tfrac{l-k}{B}\big),
\end{aligned}$$

we have

$$|Y^{L,B}(s) - Y^{L,B}(t)| \le \sigma_n^2\rho^2 \Big| \sum_{l=0}^{L-1}\sum_{k=0}^{L-1} h_l^B \overline{h_k^B}\big[\exp\big(-2\pi i t\tfrac{l-k}{B}\big) - \exp\big(-2\pi i s\tfrac{l-k}{B}\big)\big]\Big|$$
$$\le \sigma_n^2\rho^2 \sum_{l=0}^{L-1}\sum_{k=0}^{L-1} |h_l^B||\overline{h_k^B}|\big|\exp\big(-2\pi i t\tfrac{l-k}{B}\big) - \exp\big(-2\pi i s\tfrac{l-k}{B}\big)\big|.$$

Using Lipschitz continuity, Markov's inequality, and the definition of h_l^B as before we conclude

$$P\Big(\sup_{\rho(s,t)<\delta} |Y^{L,B}(s) - Y^{L,B}(t)| > \varepsilon\Big) \le \frac{2\pi M^2\delta}{\varepsilon}\sum_{l=0}^{L-1}\sum_{k=0}^{L-1}\frac{|l-k|}{B(L)}\sigma_l^B\sigma_k^B.$$

This yields the result in same way as in Proposition 3.1.3. ■

We thus have a condition for asymptotic uniform ρ-equicontinuity in probability when working with either $H^{L,B}$ or $Y^{L,B}$, which has to be verified for the specific PDP used in the model. Under this assumption we can establish weak convergence in $\ell^\infty_{\mathbb{C}}(T_1, T_2, \dots)$ and $\ell^\infty_{\mathbb{R}}(T_1, T_2, \dots)$ respectively for $H^{L,B}$ and $Y^{L,B}$.

Proposition 3.1.5. *If the sequence of stochastic processes $(H^{L,B})$ is asymptotically uniformly ρ-equicontinuous in probability, then $H^{L,B} \xrightarrow{D} H^\infty$ in $\ell^\infty_{\mathbb{C}}(T_1, T_2, \dots)$.*

Proof. We have already established $H^{L,B} \xrightarrow{\text{fidi}} H^\infty$ and assuming asymptotic uniform ρ-equicontinuity in probability the result follows from Corollary 2.5.10 if we can show the limit $H^\infty(t)$ is tight for all $t \in \mathbb{R}$. We know that $H^\infty(t) \sim \mathbb{C}\mathcal{N}(0,1)$ for all $t \in \mathbb{R}$, and thus for all $\varepsilon > 0$ there exists a sufficiently large r such that $P\big(H^\infty(t) \in B(r)\big) \ge 1-\varepsilon$, where $B(r) = \{z \in \mathbb{C} \mid |z| \le r\}$. Since $B(r)$ is closed and bounded it is compact, and the tightness is thus verified. ■

This result also holds for the transformed process.

Proposition 3.1.6. *If the sequence of stochastic processes $Y^{L,B}$ is asymptotically uniformly ρ-equicontinuous in probability, then $Y^{L,B} \xrightarrow{D} Y^\infty$ in $\ell^\infty_{\mathbb{R}}(T_1, T_2, \dots)$.*

Proof. We have already established $Y^{L,B} \xrightarrow{\text{fidi}} Y^\infty$ and asymptotic uniform ρ-equicontinuity in probability, so the result follows from Corollary 2.5.10 if we can show the limit $Y^\infty(t)$ is tight for all $t \in \mathbb{R}$. We recall $Y^\infty(t) = g\big(H^\infty(t)\big)$, where $g : \mathbb{C} \to \mathbb{R}$ is continuous, and by Proposition 3.1.5 there exists a sufficiently large r such that $P\big(H^\infty(t) \in B(r)\big) \ge 1-\varepsilon$. Since compact sets are mapped to compact sets by continuous mappings the image $g\big(B(r)\big)$ is compact and $P\Big(Y^\infty(t) \in g\big(B(r)\big)\Big) = P\big(H^\infty(t) \in B(r)\big) \ge 1-\varepsilon$, which yields the tightness. ■

The final results we will prove utilizes Skorokhod's Theorem (Theorem 2.5.11) to establish almost sure convergence for representations $\widetilde{H}^{L,B}$, $\widetilde{H}^{\infty}$, $\widetilde{Y}^{L,B}$, and $\widetilde{Y}^{\infty}$ of the original processes $H^{L,B}$, H^{∞}, $Y^{L,B}$, and Y^{∞}.

Proposition 3.1.7. *If the sequence of stochastic processes $(H^{L,B})$ is asymptotically uniformly ρ-equicontinuous in probability there exists measurable maps $\widetilde{H}^{L,B} : \widetilde{\Omega} \to \mathbb{D}$ and $\widetilde{H}^{\infty} : \widetilde{\Omega} \to \mathbb{D}$ defined on some probability space $(\widetilde{\Omega}, \widetilde{\mathcal{E}}, \widetilde{P})$ with*

(i) $\widetilde{H}^{L,B} \xrightarrow{a.s.} \widetilde{H}^{\infty}$ *in* $\ell^{\infty}_{\mathbb{C}}(T_1, T_2, \dots)$,

(ii) $\mathcal{L}(\widetilde{H}^{L,B}) = \mathcal{L}(H^{L,B})$ *for all* $L \in \mathbb{N}$ *and* $\mathcal{L}(\widetilde{H}^{\infty}) = \mathcal{L}(H^{\infty})$.

Proof. This follows from Theorem 2.5.11, and we need only verify the conditions of this theorem are satisfied. We have already established $H^{L,B} \xrightarrow{D} H^{\infty}$ and it remains to verify separability of H^{∞}. Since $\ell^{\infty}_{\mathbb{C}}(T_1, T_2, \dots)$ is a complete metric space, separability is equivalent to tightness (Lemma 2.5.4). From Corollary 2.5.10 we know the limit is tight, which yields the result. ■

The exact same argument establishes the result for the transformed process.

Proposition 3.1.8. *If $Y^{L,B}$ is asymptotically uniformly ρ-equicontinuous in probability there exists measurable maps $\widetilde{Y}^{L,B} : \widetilde{\Omega} \to \mathbb{D}$ and $\widetilde{Y}^{\infty} : \widetilde{\Omega} \to \mathbb{D}$ defined on some probability space $(\widetilde{\Omega}, \widetilde{\mathcal{E}}, \widetilde{P})$ with*

(i) $\widetilde{Y}^{L,B} \xrightarrow{a.s.} \widetilde{Y}^{\infty}$ *in* $\ell^{\infty}_{\mathbb{R}}(T_1, T_2, \dots)$,

(ii) $\mathcal{L}(\widetilde{Y}^{L,B}) = \mathcal{L}(Y^{L,B})$ *for all* $L \in \mathbb{N}$ *and* $\mathcal{L}(\widetilde{Y}^{\infty}) = \mathcal{L}(Y^{\infty})$.

This concludes the results on weak convergence, and in the next section we will show that under suitable conditions the convergence is inherited by $\mathrm{MMSE}^{L,B}_{\infty}$ defined in (3.3).

3.2 The infinite equalizer

Before studying convergence and asymptotic properties of $\mathrm{MMSE}^{L,B}_{\infty}$ we define

$$\mathrm{MMSE}^{\infty,B}_{\infty} = \int_0^B Y^{\infty}(f)df = \int_0^B \frac{\sigma_n^2}{|H^{\infty}(f)|^2 + \frac{1}{\rho}} df.$$

For this process we have a central limit theorem.

Proposition 3.2.1. *If the PDP, p, is such that the stationary complex Gaussian process H^{∞} with autocovariance function $\hat{p}$ is strongly mixing with mixing coefficient α satisfying*

$$\int_0^{\infty} \alpha(\tau)d\tau < \infty,$$

and if

$$\sigma^2 = 2\int_0^\infty \mathbb{E}\big[Y_c^\infty(0)Y_c^\infty(f)\big]df \neq 0, \tag{3.9}$$

where

$$Y_c^\infty(f) = Y^\infty(f) - \mu \qquad \textit{and} \qquad \mu = \mathbb{E}\big[Y^\infty(f)\big].$$

Then

$$\frac{\sqrt{B}}{\sigma}\Big(\mathrm{MMSE}_\infty^{\infty,B} - \mu\Big) \xrightarrow{D} \mathcal{N}(0,1)$$

for $L \to \infty$ (and thereby $B = B(L) \to \infty$).

Proof. Since $Y_c^\infty(f)$ is a measurable transformation of $H^\infty(f)$ for $f \in \mathbb{R}$, all the conditions of Corollary 2.4.8 are fulfilled. Consequently

$$\frac{1}{\sigma\sqrt{B}}\int_0^B Y_c^\infty(f)df \xrightarrow{D} \mathcal{N}(0,1),$$

and the result follows from

$$\frac{1}{\sigma\sqrt{B}}\int_0^B Y_c^\infty(f)df = \frac{\sqrt{B}}{\sigma}\Big[\frac{1}{B}\int_0^B Y^\infty(f) - \mu df\Big] = \frac{\sqrt{B}}{\sigma}\Big(\mathrm{MMSE}_\infty^{\infty,B} - \mu\Big). \qquad \blacksquare$$

In the following we will show that the CLT behavior of $\mathrm{MMSE}_\infty^{\infty,B}$ is inherited by $\mathrm{MMSE}_\infty^{L,B}$, which is the quantity we wish to study. To prove this we make use of the Skorokhod representations obtained in Proposition 3.1.8. Using these representations we define

$$\widetilde{\mathrm{MMSE}}_\infty^{L,B} = \frac{1}{B}\int_0^B \widetilde{Y}^{L,B}(f)df \qquad \text{and} \qquad \frac{1}{B}\widetilde{\mathrm{MMSE}}_\infty^{\infty,B} = \int_0^B \widetilde{Y}^\infty(f)df.$$

The following lemma ensures it suffices to show the CLT for these representations.

Lemma 3.2.2. *If*

$$\frac{\sqrt{B}}{\sigma}\Big(\widetilde{\mathrm{MMSE}}_\infty^{L,B} - \mu\Big) \xrightarrow{D} \mathcal{N}(0,1)$$

for $L \to \infty$. Then

$$\frac{\sqrt{B}}{\sigma}\Big(\mathrm{MMSE}_\infty^{L,B} - \mu\Big) \xrightarrow{D} \mathcal{N}(0,1)$$

for $L \to \infty$.

Proof. Since $\mathcal{L}(\widetilde{\mathrm{MMSE}}_\infty^{L,B}) = \mathcal{L}(\mathrm{MMSE}_\infty^{L,B})$ for all L, we have

$$P\Big(\frac{\sqrt{B}}{\sigma}\Big(\mathrm{MMSE}_\infty^{L,B} - \mu\Big) \leq x\Big) = P\Big(\frac{\sqrt{B}}{\sigma}\Big(\widetilde{\mathrm{MMSE}}_\infty^{L,B} - \mu\Big) \leq x\Big) \to \Phi(x). \qquad \blacksquare$$

Theorem 3.2.3. *Let the PDP p be such that the conditions of Proposition 3.2.1 are satisfied, and let $Y^{L,B}$ be asymptotically uniformly ρ-equicontinuous in probability. If the first moment of the PDP is finite and $\frac{L}{B^{3/2}} \to \infty$ for $L \to \infty$, then*

$$\frac{\sqrt{B}}{\sigma}\Big(\mathrm{MMSE}_\infty^{L,B} - \mu\Big) \xrightarrow{D} \mathcal{N}(0,1)$$

for $L \to \infty$.

Proof. Asymptotic uniform ρ-equicontinuity in probability implies the existence of Skorokhod representations of the process as shown in Proposition 3.1.8. By Lemma 3.2.2 we know it suffices to show the result for the Skorokhod representations. First we make the decomposition

$$\frac{\sqrt{B}}{\sigma}\Big(\widetilde{\mathrm{MMSE}}_\infty^{L,B} - \mu\Big) = \frac{\sqrt{B}}{\sigma}\Big(\widetilde{\mathrm{MMSE}}_\infty^{\infty,B} - \mu\Big) + \frac{\sqrt{B}}{\sigma}\Big(\widetilde{\mathrm{MMSE}}_\infty^{L,B} - \widetilde{\mathrm{MMSE}}_\infty^{\infty,B}\Big),$$

and notice that by Slutsky's result (Lemma 2.1.7) it suffices to show

$$\frac{\sqrt{B}}{\sigma}\Big(\widetilde{\mathrm{MMSE}}_\infty^{L,B} - \widetilde{\mathrm{MMSE}}_\infty^{\infty,B}\Big) \xrightarrow{p} 0.$$

To show this we use Markov's inequality, which yields

$$\begin{aligned} P\Big(\big|\tfrac{\sqrt{B}}{\sigma}\big(\widetilde{\mathrm{MMSE}}_\infty^{L,B} - \widetilde{\mathrm{MMSE}}_\infty^{\infty,B}\big)\big| \geq \varepsilon\Big) &\leq \frac{1}{\varepsilon}\mathbb{E}\Big(\Big|\tfrac{\sqrt{B}}{\sigma}\big(\widetilde{\mathrm{MMSE}}_\infty^{L,B} - \widetilde{\mathrm{MMSE}}_\infty^{\infty,B}\big)\Big|\Big) \\ &= \frac{\sqrt{B}}{\varepsilon\sigma}\mathbb{E}\Big(\frac{1}{B}\Big|\int_0^B \big[\widetilde{Y}^{L,B}(f) - \widetilde{Y}^{\infty}(f)\big]df\Big|\Big). \end{aligned}$$

The theorem is thus proved if we can verify

$$\frac{1}{B}\int_0^B \sqrt{B}\,\mathbb{E}\big(\big|\widetilde{Y}^{L,B}(f) - \widetilde{Y}^{\infty}(f)\big|\big)df \to 0 \qquad \text{for } L \to \infty. \tag{3.10}$$

Intuitively this seems plausible since we have convergence in the supremum norm on compact sets for the Skorokhod representations. This also clarifies why we need to work in term of the representations. The original processes are only similar in distribution, which does not imply a difference as in (3.10) goes to 0. The remaining difficulty is of course we need a rate of the convergence due to the factor $\sqrt{B}$. In the following we prove in detail the (3.10) does hold for the Skorokhod representations.

We use a construction from the proof of Skorokhod's Theorem for real random variables (see e.g. Billingsley, 1986, Theorem 25.6.). Let $Z \sim \mathrm{Unif}(0,1)$ and define for each $L \in \mathbb{N}$ and $f \in \mathbb{R}$ the random variables

$$\begin{aligned} \widehat{Y}^{L,B}(f) &= F^{-1}_{\widetilde{Y}^{L,B}}(Z) \\ \widehat{Y}^{\infty}(f) &= F^{-1}_{\widetilde{Y}^{\infty}}(Z). \end{aligned}$$

Since these new representations have same distribution as the old representations, we have

$$\mathbb{E}\big(\big|\widetilde{Y}^{L,B}(f) - \widetilde{Y}^{\infty}(f)\big|\big) = \mathbb{E}\big(\big|\widehat{Y}^{L,B}(f) - \widehat{Y}^{\infty}(f)\big|\big) = \int_0^1 |F^{-1}_{\widetilde{Y}^{L,B}}(z) - F^{-1}_{\widetilde{Y}^{\infty}}(z)| dz.$$

The inverse distribution functions are (see Lemma A.3.4):

$$F^{-1}_{\widetilde{Y}^{L,B}}(z) = \frac{\sigma_n^2}{\rho^{-1} - \nu^{L,B}\ln(z)} \qquad \text{and} \qquad F^{-1}_{\widetilde{Y}^{\infty}}(z) = \frac{\sigma_n^2}{\rho^{-1} - \ln(z)}$$

where

$$\nu^{L,B} = \int_0^{\frac{L}{B}} p(t)dt.$$

Consequently we have

$$\begin{aligned}
\mathbb{E}\big(\big|\widetilde{Y}^{L,B}(f) - \widetilde{Y}^{\infty}(f)\big|\big) &= \int_0^1 \Big|\frac{\sigma_n^2}{\rho^{-1} - \nu^{L,B}\ln(z)} - \frac{\sigma_n^2}{\rho^{-1} - \ln(z)}\Big| df \\
&= \sigma_n^2 \int_0^1 \Big|\frac{\nu^{L,B}\ln(z) - \ln(z)}{\rho^{-2} - (\nu^{L,B}+1)\rho^{-1}\ln(z) + \nu^{L,B}\ln(z)^2}\Big| df \\
&= (1-\nu^{L,B})\sigma_n^2 \int_0^1 \Big|\frac{\ln(z)}{\rho^{-2} - (\nu^{L,B}+1)\rho^{-1}\ln(z) + \nu^{L,B}\ln(z)^2}\Big| df \\
&\leq (1-\nu^{L,B})\sigma_n^2 C,
\end{aligned}$$

where C is a constant bounding the integral. Now we return to the verification of (3.10), and observe

$$\frac{1}{B}\int_0^B \sqrt{B}\,\mathbb{E}\big(\big|\widetilde{Y}^{L,B}(f) - \widetilde{Y}^{\infty}(f)\big|\big) df \leq \sqrt{B}(1-\nu^{L,B})\sigma_n^2 C.$$

If we let X be a stochastic variable with density p (the PDP) we have

$$\begin{aligned}
\frac{1}{B}\int_0^B \sqrt{B}\,\mathbb{E}\big(\big|\widetilde{Y}^{L,B}(f) - \widetilde{Y}^{\infty}(f)\big|\big) df &\leq \sqrt{B}\Big(1 - P(X \leq \tfrac{L}{B})\Big)\sigma_n^2 C \\
&\leq \sqrt{B} P\big(X \geq \tfrac{L}{B}\big)\sigma_n^2 C \\
&\leq \sqrt{B}\frac{\mathbb{E}(X)}{L/B}\sigma_n^2 C.
\end{aligned}$$

This converges to 0 for $L \to \infty$ by the assumptions in the theorem, and the proof is thus complete. ■

3.3 The finite equalizer

Given a channel response $(h_0^B, \dots, h_{L-1}^B)$ sampled using bandwidth $B = 1/T$ we wish to establish a relation between $\mathrm{MMSE}_N^{L,B}$ given by (1.7) in Chapter 1 and $\mathrm{MMSE}_\infty^{L,B}$ studied in the previous section. At first we will keep L, B fixed and only study asymptotics for growing equalizer length N, which is done in Proposition 3.3.2. To do this we need a lemma providing an upper bound on $\mathrm{MMSE}_N^{L,B}$.

Lemma 3.3.1. *Let $L, N \in \mathbb{N}$ and $B \in \mathbb{R}_+$. Then for all realizations of $(h_0^B, \dots, h_{L-1}^B)$*

$$\mathrm{MMSE}_N^{L,B} \leq \frac{1}{N+L-1} \sum_{j=1}^{N+L-1} Y^{L,B}\left(\tfrac{jB}{N+L-1}\right) + \frac{(L-1)\sigma_x^2}{L+N-1}, \tag{3.11}$$

where σ_x^2 is the noise power and $\mathrm{MMSE}_N^{L,B}$ is defined in (1.7).

Proof. Since $\mathrm{MMSE}_N^{L,B}$ is given by the minimum diagonal element of a matrix we can upper bound it by the average diagonal element, i.e.

$$\mathrm{MMSE}_N^{L,B} \leq \frac{\sigma_n^2}{D} \mathrm{tr}[(H^*H + \tfrac{1}{\rho}I)^{-1}],$$

where $D = L+N-1$ is the number of columns in the matrix H defined in (1.6) on page 6. Any matrix of the form H^*H is Hermitian and using the spectral theorem (Theorem A.1.1) we have $H^*H = U\Lambda U^*$, where U is a unitary matrix and Λ is a diagonal matrix consisting of the eigenvalues of H^*H. Furthermore H^*H is positive semi-definite since for all $\boldsymbol{x}$ we have $\boldsymbol{x}^*H^*H\boldsymbol{x} = \|H\boldsymbol{x}\|^2 \geq 0$, which also implies the eigenvalues $\lambda_1, \dots, \lambda_D$ of H^*H are non-negative. The matrix

$$H^*H + \tfrac{1}{\rho}I = U(\Lambda + \tfrac{1}{\rho}I)U^*$$

has eigenvalues $\lambda_j + \frac{1}{\rho}$, which are strictly positive due to the non-negativity of λ_j. Since all eigenvalues are non-zero the matrix is invertible and the eigenvalues of the inverse matrix are

$$\frac{1}{\lambda_j + \frac{1}{\rho}}, \quad j = 1, \dots, D.$$

Using that the trace of a matrix is the sum of the eigenvalues yields

$$\mathrm{MMSE}_N^{L,B} \leq \frac{\sigma_n^2}{D} \mathrm{tr}[(H^*H + \tfrac{1}{\rho}I)^{-1}] = \frac{\sigma_n^2}{D} \sum_{j=1}^{D} \frac{1}{\lambda_j + \frac{1}{\rho}}. \tag{3.12}$$

This gives an upper bound on $\mathrm{MMSE}_N^{L,B}$ in terms of the eigenvalues of H^*H, but we do not have an explicit expression for these. In the following we will approximate the Toeplitz matrix H by a circulant matrix C for which we can calculate the eigenvalues explicitly and achieve an alternative upper bound on $\mathrm{MMSE}_N^{L,B}$. Let C be the $D\times D$ circulant matrix obtained by right-shifting the first row of H:

$$C=\left[\begin{array}{cccccccc} h_0^B & h_1^B & \cdots & h_{L-1}^B & 0 & 0 & \cdots & 0 \\ 0 & h_0^B & h_1^B & \cdots & h_{L-1}^B & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_0^B & h_1^B & \cdots & h_{L-1}^B & 0 \\ 0 & 0 & \cdots & 0 & h_0^B & h_1^B & \cdots & h_{L-1}^B \\ \hdashline h_{L-1}^B & 0 & \cdots & 0 & 0 & h_0^B & \cdots & h_{L-2}^B \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ h_1^B & \cdots & h_{L-1}^B & 0 & \cdots & 0 & 0 & h_0^B \end{array}\right]=\begin{bmatrix} H \\ E \end{bmatrix}$$

It is seen that C is simply obtained by appending $L-1$ rows to H, and the matrix consisting of these rows is called E. Using the block matrix notation from above we see that

$$C^*C=[H^*E^*]\begin{bmatrix} H \\ E \end{bmatrix}=H^*H+E^*E.$$

We denote the eigenvalues of H^*H, C^*C and E^*E by λ_i, μ_i and ν_i respectively for $i=1,\dots,D$. The matrices H^*H, C^*C and E^*E are all Hermitian and positive semidefinite, and we assume the eigenvalues to be in non-increasing order such that $\lambda_1\geq\cdots\geq\lambda_D$ and analogously for μ_i and ν_i. Since E only has $L-1$ rows we conclude that $\mathrm{rank}(E^*E)\leq\min(\mathrm{rank}(E^*),\mathrm{rank}(E))\leq L-1$, and therefore $\nu_i=0$ for $i\geq L$. Weyl's inequalities (Theorem A.1.2) imply, that

$$\mu_{i+j-1}\leq\lambda_i+\nu_j,$$

and for $j=L$ we get

$$\mu_{i+L-1}\leq\lambda_i. \tag{3.13}$$

We split the sum in (3.12) in two parts and apply (3.13), which yields

$$\begin{aligned}\sum_{i=1}^{D}\frac{1}{\lambda_i+\frac{1}{\rho}}&=\sum_{i=1}^{N}\frac{1}{\lambda_i+\frac{1}{\rho}}+\sum_{i=N+1}^{N+L-1}\frac{1}{\lambda_i+\frac{1}{\rho}}\\&\leq\sum_{i=1}^{N}\frac{1}{\mu_{i+L-1}+\frac{1}{\rho}}+(L-1)\rho\\&\leq\sum_{i=1}^{D}\frac{1}{\mu_i+\frac{1}{\rho}}+(L-1)\rho.\end{aligned}$$

Inserting this into (3.12) yields the alternative upper bound

$$\mathrm{MMSE}_N^{L,B}\leq\frac{1}{D}\sum_{i=1}^{D}\frac{\sigma_n^2}{\mu_i+\frac{1}{\rho}}+\frac{(L-1)\sigma_x^2}{D}, \tag{3.14}$$

where we used the definition of the signal to noise ratio $\rho = \frac{\sigma_x^2}{\sigma_n^2}$. Now we exploit that Theorem A.1.3 provides an explicit expression for the eigenvalues of a circulant matrix. Since the first row of C in our case is $[h_0^B \cdots h_{L-1}^B]$ appended with $N-1$ zeros the eigenvalues of C, denoted α_j, $j = 1, \ldots, D$, are

$$\alpha_j = \sum_{l=0}^{L-1} h_l^B \exp(-2\pi i j \tfrac{l}{D}) = H^{L,B}(\tfrac{jB}{D}),$$

where $H^{L,B}$ is the transfer function given by (3.2).

Using an eigenvalue decomposition (Theorem A.1.3) $C = U\Lambda U^*$ we notice that $C^* = U\Lambda^* U^*$ and therefore $C^*C = U\Lambda^*\Lambda U^*$. Consequently the eigenvalues of C^*C can be expressed as $\mu_j = |\alpha_j|^2 = |H^{L,B}(\frac{jB}{D})|^2$, $j = 1, \ldots, D$. Inserting this into (3.14) yields

$$\text{MMSE}_N^{L,B} \leq \frac{1}{D} \sum_{j=1}^{D} \frac{\sigma_n^2}{|H^{L,B}(\frac{jB}{D})|^2 + \frac{1}{\rho}} + \frac{(L-1)\sigma_x^2}{D}.$$

This is seen to be the bound (3.11) we set out to prove, since $D = L+N-1$ and

$$Y^{L,B}(f) = \frac{\sigma_n^2}{|H^{L,B}(f)|^2 + \frac{1}{\rho}}$$

for all $f \in \mathbb{R}$. ■

Using the upper bound provided in Lemma 3.3.1 we can prove $\text{MMSE}_N^{L,B} \to \text{MMSE}_\infty^{L,B}$. Intuitively this holds since the sum term in the upper bound is seen to be a Riemann sum approximation to $\text{MMSE}_\infty^{L,B}$. We prove this in detail in the following proposition.

Proposition 3.3.2. *Let $L \in \mathbb{N}$ and $B \in \mathbb{R}_+$. Then*

$$\text{MMSE}_N^{L,B} \to \text{MMSE}_\infty^{L,B} \quad \text{as} \quad N \to \infty,$$

for all realizations of $(h_0^B, \ldots, h_{L-1}^B)$.

Proof. We let $S_N^{L,B}$ denote the sum term of the upper bound (3.11) such that

$$\begin{aligned} S_N^{L,B} &= \frac{1}{N+L-1} \sum_{j=1}^{N+L-1} Y^{L,B}(\tfrac{jB}{N+L-1}) \\ &= \frac{1}{B} \sum_{j=1}^{N+L-1} \frac{B}{N+L-1} Y^{L,B}(\tfrac{jB}{N+L-1}). \end{aligned}$$

We can write this sum as the integral

$$S_N^{L,B} = \frac{1}{B}\int_0^B Y_N^{L,B}(f)df, \tag{3.15}$$

where $Y_N^{L,B} : \mathbb{R} \to [0, \rho\sigma_n^2]$ is the simple function defined by

$$Y_N^{L,B}(f) = \sum_{j=1}^{L+N-1} \mathcal{I}\big(f \in \big[\tfrac{(j-1)B}{L+N-1}, \tfrac{jB}{L+N-1}\big]\big) Y^{L,B}\big(\tfrac{jB}{L+N-1}\big), \tag{3.16}$$

for all $f \in \mathbb{R}$.

As mentioned in Chapter 1 the finite equalizer is a sub model of the infinite equalizer implying $\mathrm{MMSE}_\infty^{L,B} \leq \mathrm{MMSE}_N^{L,B}$. Combining this with the upper bound on $\mathrm{MMSE}_N^{L,B}$ yields

$$\mathrm{MMSE}_\infty^{L,B} \leq \mathrm{MMSE}_N^{L,B} \leq S_N^{L,B} + \frac{(L-1)\sigma_x^2}{L+N-1},$$

which leads to

$$\big|\mathrm{MMSE}_N^{L,B} - \mathrm{MMSE}_\infty^{L,B}\big| \leq S_N^{L,B} - \mathrm{MMSE}_\infty^{L,B} + O\Big(\frac{L}{N}\Big). \tag{3.17}$$

It thus only remains to verify that

$$S_N^{L,B} - \mathrm{MMSE}_\infty^{L,B} \to 0 \quad \text{for} \quad N \to 0.$$

From (3.15) it follows that

$$\begin{aligned}\big|S_N^{L,B} - \mathrm{MMSE}_\infty^{L,B}\big| &= \Big|\frac{1}{B}\int_0^B \big[Y_N^{L,B}(f) - Y^{L,B}(f)\big]df\Big| \\ &\leq \frac{1}{B}\int_0^B \big|Y_N^{L,B}(f) - Y^{L,B}(f)\big|df.\end{aligned} \tag{3.18}$$

From (3.16) it is clear that $Y_N^{L,B}$ is a simple function approximation of $Y^{L,B}$ on the interval $[0, B]$ such that for all $f \in [0, B]$ we have $Y_N^{L,B}(f) \to Y^{L,B}(f)$ or equivalently $\big|Y_N^{L,B}(f) - Y^{L,B}(f)\big| \to 0$ for $N \to \infty$. Since $\big|Y_N^{L,B}(f) - Y^{L,B}(f)\big| \leq \rho\sigma_x^2\mathcal{I}(f \in [0, B])$ we can use Lebesgue's theorem on dominated convergence to conclude

$$\begin{aligned}\lim_{N\to\infty} \frac{1}{B}\int_0^B \big|Y_N^{L,B}(f) - Y^{L,B}(f)\big|df &= \lim_{N\to\infty} \frac{1}{B}\int_0^\infty \mathcal{I}(f \in [0, B])\big|Y_N^{L,B}(f) - Y^{L,B}(f)\big|df \\ &= \frac{1}{B}\int_0^\infty \lim_{N\to\infty} \mathcal{I}(f \in [0, B])\big|Y_N^{L,B}(f) - Y^{L,B}(f)\big|df \\ &= 0.\end{aligned}$$

■

We have already exploited this result in Section 1.2, where we only carried out simulations of $\text{MMSE}_\infty^{L,B}$. Proposition 3.3.2 ensures that for $N \to \infty$ we have the pointwise convergence

$$\text{MMSE}_N^{L,B} \to \text{MMSE}_\infty^{L,B},$$

for all realizations of the channel response, which implies the weak convergence

$$\text{MMSE}_N^{L,B} \xrightarrow{D} \text{MMSE}_\infty^{L,B}.$$

Consequently for a number of simulated channel responses the results obtained by calculating the integral expression for $\text{MMSE}_\infty^{L,B}$ are arbitrarily close to the results obtained by inverting a matrix and finding the minimum diagonal element as described by (1.7), when N is sufficiently large. This has the advantage that the matrix inversion, which is computer intensive for large N, can be avoided.

3.3.1 Central limit theorem

We will briefly consider the situation when L, B, N grow simultaneously. We let $B = B(L)$ and $N = N(L)$ depend on L such that $B, N \to \infty$ for $L \to \infty$. Analogously to the result for $\text{MMSE}_\infty^{L,B}$ we have a condition ensuring asymptotic normality of $\text{MMSE}_N^{L,B}$ in the following theorem.

Theorem 3.3.3. *Let*

$$\frac{\sqrt{B}}{\sigma}\Big(\text{MMSE}_\infty^{L,B} - \mu\Big) \xrightarrow{D} \mathcal{N}(0,1).$$

If $\frac{L\sqrt{B}}{N} \to 0$ *for* $L \to \infty$ *and*

$$\mathbb{E}\Big(\Big|\frac{1}{B}\int_0^B \big[Y_N^{L,B}(f) - Y^{L,B}(f)\big]\,df\Big|\Big) = o\big(\tfrac{1}{\sqrt{B}}\big), \tag{3.19}$$

then

$$\frac{\sqrt{B}}{\sigma}\Big(\text{MMSE}_N^{L,B} - \mu\Big) \xrightarrow{D} \mathcal{N}(0,1)$$

for $L \to \infty$.

Proof. The proof is similar to the beginning of that of Theorem 3.2.3, and we write

$$\frac{\sqrt{B}}{\sigma}\Big(\text{MMSE}_N^{L,B} - \mu\Big) = \frac{\sqrt{B}}{\sigma}\Big(\text{MMSE}_\infty^{L,B} - \mu\Big) + \frac{\sqrt{B}}{\sigma}\Big(\text{MMSE}_N^{L,B} - \text{MMSE}_\infty^{L,B}\Big).$$

By Slutsky's result (Lemma 2.1.7) it suffices to show

$$\frac{\sqrt{B}}{\sigma}\Big(\text{MMSE}_N^{L,B} - \text{MMSE}_\infty^{L,B}\Big) \xrightarrow{p} 0.$$

Markov's inequality yields

$$P\Big(\big|\tfrac{\sqrt{B}}{\sigma}\Big(\mathrm{MMSE}_N^{L,B} - \mathrm{MMSE}_\infty^{L,B}\Big)\big| \geq \varepsilon\Big) \leq \frac{1}{\varepsilon}\mathbb{E}\Big(\tfrac{\sqrt{B}}{\sigma}\Big|\mathrm{MMSE}_N^{L,B} - \mathrm{MMSE}_\infty^{L,B}\Big|\Big).$$

From (3.17) and (3.18) we know

$$\tfrac{\sqrt{B}}{\sigma}\Big|\mathrm{MMSE}_N^{L,B} - \mathrm{MMSE}_\infty^{L,B}\Big| \leq \tfrac{\sqrt{B}}{\sigma}\Big|\frac{1}{B}\int_0^B \big[Y_N^{L,B}(f) - Y^{L,B}(f)\big]df\Big| + \frac{\sqrt{B}}{\sigma}O\Big(\tfrac{L}{N}\Big),$$

such that

$$\tfrac{\sqrt{B}}{\sigma}\mathbb{E}\Big(\Big|\mathrm{MMSE}_N^{L,B} - \mathrm{MMSE}_\infty^{L,B}\Big|\Big) \leq \tfrac{\sqrt{B}}{\sigma}\mathbb{E}\Big(\Big|\frac{1}{B}\int_0^B \big[Y_N^{L,B}(f) - Y^{L,B}(f)\big]df\Big|\Big) + O\Big(\tfrac{L\sqrt{B}}{N}\Big).$$

By the imposed assumptions the two terms converge to 0, which finishes the proof. ■

3.4 Discussion

In this section we will discuss some of the conditions we have imposed in the results in Sections 3.2 and 3.3. We start with the first condition in Proposition 3.2.1.

Condition 1. *Let the PDP, p, be such that the stationary complex Gaussian process H^∞ with autocovariance function $\hat{p}$ is strongly mixing with mixing coefficient α satisfying*

$$\int_0^\infty \alpha(\tau)d\tau < \infty.$$

Discussion:
A difficulty regarding mixing properties of stochastic processes is that mixing coefficients are hard to calculate, and it is thus desirable to characterize mixing processes in another way. It would be useful to have a sufficient condition on the rate of decay of the autocovariance function to ensure a mixing property. This is however not possible (except in the case when the autocovariance is zero from some point) since no matter how fast the autocovariance decays it is always possible to construct a Gaussian process with this rate of decay that does not exhibit strong mixing (Piterbarg, 1996, page 43).

A more fruitful approach to this problem is to study the spectral density of the process, which e.g. is done in Dym and McKean (1976). The spectral representation given in (2.6) can be written as

$$\mathbb{E}(X_{t_1}\overline{X_{t_2}}) = \int_{-\infty}^{\infty} \exp(it_1\lambda)\overline{\exp(it_2\lambda)}d\mu(\lambda).$$

The left hand side of this expression is the inner product of the elements X_{t_1} and X_{t_2} on the space $L^2(\Omega, P)$ consisting of all complex random variables ξ on $(\Omega, \mathcal{E}, P)$

with $\mathbb{E}(|\xi|^2) < \infty$. The right hand side is the inner product on $L^2(\mathbb{R}, \mu)$ of the elements e_{t_1} and e_{t_2}, where $e_{t_j} : \mathbb{R} \to \mathbb{C}$ is defined by $e_{t_j}(\lambda) = \exp(it_j\lambda)$ for all $\lambda \in \mathbb{R}$. In this way we have an operator $\mathcal{T} : (X_t) \to L^2(\mathbb{R}, \mu)$ taking the element X_t to the element e_t. We now consider the entire process (X_t) and the corresponding collection (e_t). We denote the closed linear span of these collections by respectively $A \subseteq L^2(\Omega, P)$ and $B \subseteq L^2(\mathbb{R}, \mu)$. The operator $\mathcal{T} : A \to B$ is called the trigonometric isomorphism and this relation allows probabilistic problems concerning (X_t) to be treated by trigonometrical approximation in $L^2(\mathbb{R}, \mu)$, which is used in Domínguez (1992) to study completely regular (ρ-mixing) processes. Then it can be exploited that for Gaussian processes strong mixing and complete regularity are equivalent with proportional mixing coefficients. The more detailed study of this does however turn out to be a study of complex analysis, which is outside the scope of this thesis. □

In the following we consider the second condition in Proposition 3.2.1.

Condition 2. *Let the PDP, p, be such that*

$$\sigma^2 = 2\int_0^\infty \mathbb{E}\big[Y_c^\infty(0)Y_c^\infty(f)\big]df \neq 0, \tag{3.20}$$

where

$$Y_c^\infty(f) = Y^\infty(f) - \mu \qquad \text{and} \qquad \mu = \mathbb{E}\big[Y^\infty(f)\big].$$

Discussion:
First we recall the marginal variance of H^∞ is $R_H^\infty(0) = \hat{p}(0) = 1$. In Section 2.1.4 it was shown that if $X \sim \mathbb{C}\mathcal{N}(0, 1)$ then $|X|^2 \sim \text{Exp}(1)$, and thus

$$\mu = \mathbb{E}\big[Y^\infty(f)\big] = \int_0^\infty \frac{\sigma_n^2}{u + \frac{1}{\rho}} \exp(-u)du.$$

This integral cannot be evaluated analytically, but it is well suited for numerical integration since the integrand vanishes quickly for large u. The expression for μ is used when checking the condition (3.20), which can be rewritten as

$$2\int_0^\infty \mathbb{E}\big[Y^\infty(0)Y^\infty(f)\big] - \mu^2 df \neq 0.$$

The mean value in this expression is

$$\mathbb{E}\big[Y^\infty(0)Y^\infty(f)\big] = \int_{\mathbb{C}^2} \frac{\sigma_n^2}{|x_1|^2 + \frac{1}{\rho}} \frac{\sigma_n^2}{|x_2|^2 + \frac{1}{\rho}} q(x_1, x_2)dx_1dx_2,$$

where q is the density of $(X_1, X_2)^\top \sim \mathbb{C}\mathcal{N}_2(\mathbf{0}, \Sigma)$ with

$$\Sigma = \begin{bmatrix} 1 & \hat{p}(f) \\ \hat{p}(-f) & 1 \end{bmatrix}.$$

Therefore the verification of (3.20) for specific PDPs is a multidimensional integration problem, which we will not discuss further here. □

The final condition we will mention appears in Theorem 3.3.3.

Condition 3. *Let*

$$\mathbb{E}\Big(\Big|\frac{1}{B}\int_0^B \big[Y_N^{L,B}(f) - Y^{L,B}(f)\big]\,df\Big|\Big) = o\big(\tfrac{1}{\sqrt{B}}\big).$$

Discussion:
Intuitively there is some hope showing convergence to 0 since $Y_N^{L,B}$ is a simple function approximating $Y^{L,B}$, but we also need a rate of the convergence, and it remains as an open problem to study this condition in detail. □

CHAPTER 4

Conclusion and future work

4.1 Conclusion

In this thesis we have studied the behavior of the minimum mean square error (MMSE) as a performance measure for a communication system using either a finite of infinite equalizer. The communication system model has been parametrized by the power delay profile p, the system bandwidth B, the channel length L, and in the case of a finite equalizer also the equalizer length N. The performance measures were denoted respectively $\mathrm{MMSE}_\infty^{L,B}$ and $\mathrm{MMSE}_N^{L,B}$ for the infinite and finite equalizer, and a fundamental process involved in calculating $\mathrm{MMSE}_\infty^{L,B}$ is the transfer function $H^{L,B}$. We have proven asymptotic uniform ρ-equicontinuity in probability for this process, under suitable conditions, and thereby established a weak limit H^∞. Using these results a central limit theorem (CLT) was proven for $\mathrm{MMSE}_\infty^{L,B}$ as the main result in the thesis. Furthermore it was proven that $\mathrm{MMSE}_N^{L,B} \to \mathrm{MMSE}_\infty^{L,B}$ as $N \to \infty$ for fixed L, B. This makes approximate simulations of $\mathrm{MMSE}_N^{L,B}$ possible by simulating $\mathrm{MMSE}_\infty^{L,B}$, which is less computer intensive. Finally a condition was given to extend the CLT to hold for $\mathrm{MMSE}_N^{L,B}$.

4.2 Future work

The discussion in Section 3.4 points out some interesting open questions to address in the future. A more clear description of the class of power delay profiles ensuring the CLT behavior is desirable. Especially some conditions on the power delay profile ensuring strong mixing with mixing coefficients decaying at a specified rate would be beneficial. Furthermore a study of the condition ensuring the CLT for $\mathrm{MMSE}_N^{L,B}$ is needed.

Another interesting issue is to relate the results obtained here to the results presented in Barriac and Madhow (2004). Here a similar quantity to $\mathrm{MMSE}_\infty^{L,B}$ called the spectral efficiency, I_B, is studied. It is given by

$$I_B = \frac{1}{B} \int_{-\frac{B}{2}}^{\frac{B}{2}} \log(1 + \rho |H(f)|^2) df,$$

which is a very similar structure to the one studied in the present thesis. Barriac and Madhow (2004) proves a CLT for the spectral efficiency using Serfling (1968). It is interesting investigate if this setup could be used to obtain alternative conditions ensuring the CLT in our case.

APPENDIX A

Miscellaneous results

A.1 Linear Algebra

A.1.1 The Spectral Theorem

A fundamental result in linear algebra is the spectral theorem, which can be found in Hogben (2006).

Theorem A.1.1. *Let A be a Hermitian matrix. Then there is a unitary matrix U such that $U^*AU = D$, where D is a real diagonal matrix whose diagonal entries are the eigenvalues of A.*

A.1.2 Weyl's inequalities

The following result is due to Weyl (Bhatia, 1997, Theorem III.2.1 and (III.8)).

Theorem A.1.2. *Let A, B be Hermitian matrices with eigenvalues $\lambda_1^A \geq \cdots \geq \lambda_n^A$ and $\lambda_1^B \geq \cdots \geq \lambda_n^B$ respectively. The eigenvalues $\lambda_1^{A+B} \geq \cdots \geq \lambda_n^{A+B}$ of $A + B$ satisfy the inequalities*

$$\lambda_{i+j-1}^{A+B} \leq \lambda_i^A + \lambda_j^B$$

for all $i, j \in \{1, \dots, n\}$ such that $j + k - 1 \leq n$.

A.1.3 Circulant matrices

An important property of circulant matrices is that they can be diagonalized and there is a closed form expression for the eigenvalues (Davis, 1994, Theorem 3.2.2.).

Theorem A.1.3. *Let $[c_0 \cdots c_{n-1}]$ be the first row of a $n \times n$ circulant matrix C. Then there exists a unitary matrix U, such that $C = U^*DU$, where D is a diagonal matrix whose diagonal entries are the eigenvalues $\alpha_1, \dots, \alpha_n$ of C. Furthermore*

$$\alpha_j = \sum_{l=0}^{n-1} c_l \exp(-2\pi i j \tfrac{l}{n}), \qquad j = 1, \dots, n.$$

A.2 Probability and measure theory

A.2.1 Fundamental inequalities

Markov's, Chebyshev's, and Lapynov's inequalities are used in this thesis and for convenience we state them here (Billingsley, 1986, p. 74-76):

$$P(|X| \geq \alpha) \leq \frac{1}{\alpha}\mathbb{E}(|X|). \tag{A.1}$$

$$P(|X-\mathbb{E}(X)|\geq\alpha)\leq\frac{1}{\alpha^2}\mathrm{Var}(X). \tag{A.2}$$

$$\mathbb{E}(|X|^s)^{1/s}\geq\mathbb{E}(|X|^t)^{1/t},\quad 0<s\leq t. \tag{A.3}$$

A.2.2 Tonelli's Theorem

A classical result in measure theory is Tonelli's Theorem (Billingsley, 1986, Theorem 18.3) presented below.

Theorem A.2.1. *Let $(X,\mathcal{F},\mu)$ and $(Y,\mathcal{G},\nu)$ be σ-finite measure spaces and let the function $f:X\times Y\to\mathbb{R}_+$ be measurable with respect to the product σ-algebra $\mathcal{F}\otimes\mathcal{G}$. Then*

(i) the function defined by $g(x)=\int_Y f(x,y)d\nu(y)$ is measurable with respect to $\mathcal{F}$.

(ii) $\int_{X\times Y} f(x,y)d\mu\otimes\nu(x,y)=\int_X\big[\int_Y f(x,y)d\nu(y)\big]d\mu(x)$.

Extension:
If f takes values in $\mathbb{R}$ then (i) still holds, which can be verified by using (i) on the the positive and negative part of f separately.

A.3 Auxiliary lemmas

The following lemma is used to prove Lemma A.3.2, which is part of the proof of Theorem 2.4.6.

Lemma A.3.1. *Under the conditions of Theorem 2.4.6 we have*

$$\mathbb{E}(S_n^4)=\mathbb{E}[(\sum_{j=1}^n X_j)^4]=o(n^3).$$

Proof. We start by splitting the product in to several parts and for notational convenience let $N=\{1,\ldots,n\}$, such that

$$\begin{aligned}\mathbb{E}\Big[\Big(\sum_{j=1}^n X_j\Big)^4\Big] &= \sum_{i,j,k,l\in N}\mathbb{E}(X_iX_jX_kX_l)\\ &=\sum_{i=1}^n\mathbb{E}(X_i^4)+\sum_{\substack{i,j\in N\\ i\neq j}}\mathbb{E}(X_i^3X_j)+\sum_{\substack{i,j\in N\\ i\neq j}}\mathbb{E}(X_i^2X_j^2)\\ &\quad+\sum_{\substack{i,j,k\in N\\ i\neq j\neq k}}\mathbb{E}(X_i^2X_jX_k)+\sum_{\substack{i,j,k,l\in N\\ i\neq j\neq k\neq l}}\mathbb{E}(X_iX_jX_kX_l).\end{aligned}$$

Since $P(|X_j| < C) = 1$ it is clear that we can bound each of the mean values in the expression above by C^4. The first sum only has n terms and therefore it is $o(n^2)$. The number of terms in the second and third sum is $O(n^2)$, so that these are $o(n^3)$, and we thus only need to verify the fourth and fifth sums are $o(n^3)$. We start with the fourth and split the index set $\{(i,j,k) \in N^3 \mid i \neq j \neq k\}$ into the six index sets

$$\begin{aligned} N_1 &= \{(i,j,k) \in N^3 \mid i < j < k\}, & N_2 &= \{(i,j,k) \in N^3 \mid i < k < j\}, \\ N_3 &= \{(i,j,k) \in N^3 \mid j < i < k\}, & N_4 &= \{(i,j,k) \in N^3 \mid j < k < i\}, \\ N_5 &= \{(i,j,k) \in N^3 \mid k < i < j\}, & N_6 &= \{(i,j,k) \in N^3 \mid k < j < i\}, \end{aligned}$$

which leads to

$$\sum_{\substack{i,j,k \in N \\ i \neq j \neq k}} \mathbb{E}(X_i^2 X_j X_k) = \sum_{s=1}^{6} \sum_{i,j,k \in N_s} \mathbb{E}(X_i^2 X_j X_k).$$

In the following we show that the sum over N_1 is $O(n^2)$.

Clearly

$$\sum_{i,j,k \in N_1} \mathbb{E}(X_i^2 X_j X_k) = O\Big(\sum_{i,j,k \in N_1} |\mathbb{E}(X_i^2 X_j X_k)| \Big).$$

For each of the terms in this sum we use $i<j<k$ and Theorem 2.3.2 to obtain

$$|\mathbb{E}(X_i^2 X_j X_k)| = |\mathbb{E}(X_i^2 X_j X_k) - \mathbb{E}(X_i^2 X_j)\mathbb{E}(X_k)| \leq 4C^4 \alpha_X(k-j), \tag{A.4}$$

such that

$$\begin{aligned} \sum_{i,j,k \in N_1} \mathbb{E}(X_i^2 X_j X_k) &= O\Big(\sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{k=j+1}^{n} \alpha_X(k-j) \Big) \\ &= O\Big(\sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{l=1}^{n} \alpha_X(l) \Big) \\ &= O\Big(\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{\infty} \alpha_X(l) \Big) \\ &= O(n^2). \end{aligned} \tag{A.5}$$

The other five sums follow in the exact same manner with the appropriate change of indices in (A.4) and (A.5). Hereby we have established

$$\sum_{\substack{i,j,k \in N \\ i \neq j \neq k}} \mathbb{E}(X_i^2 X_j X_k) = o(n^3),$$

as desired.

In the same manner as before we have

$$\sum_{\substack{i,j,k,l \in N \\ i \neq j \neq k \neq l}} \mathbb{E}(X_i X_j X_k X_l) = \sum_{s=1}^{24} \sum_{i,j,k,l \in N_s} \mathbb{E}(X_i X_j X_k X_l),$$

where $N_1, \ldots, N_{24}$ are the different orderings of the four indices. In this case we notice that the different index sets merely correspond to a permutation of the factors X_i, X_j, X_k, and X_j such that we actually have

$$\sum_{\substack{i,j,k,l \in N \\ i \neq j \neq k \neq l}} \mathbb{E}(X_i X_j X_k X_l) = 24 \sum_{i,j,k,l \in N_1} \mathbb{E}(X_i X_j X_k X_l),$$

where $N_1 = \{(i,j,k,l) \in N^4 \mid i<j<k<l\}$. Theorem 2.3.2 yields two different bounds for the terms in the sum:

$$|\mathbb{E}(X_i X_j X_k X_l)| = |\mathbb{E}(X_i X_j X_k X_l) - \mathbb{E}(X_i)\mathbb{E}(X_j X_k X_l)| \leq 4C^4 \alpha_X(j-i),$$

and

$$|\mathbb{E}(X_i X_j X_k X_l)| = |\mathbb{E}(X_i X_j X_k X_l) - \mathbb{E}(X_i X_j X_k)\mathbb{E}(X_l)| \leq 4C^4 \alpha_X(l-k).$$

Since both of these holds for all configurations of $i<j<k<l$ we can use the minimum as an upper bound such that

$$\begin{aligned} \sum_{\substack{i,j,k,l \in N \\ i \neq j \neq k \neq l}} \mathbb{E}(X_i X_j X_k X_l) &= O\Big(\sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{k=j+1}^{n} \sum_{l=k+1}^{n} \min\{\alpha_X(j-i), \alpha_X(l-k)\}\Big) \\ &= O\Big(n^2 \sum_{j=1}^{n} j\alpha_X(j)\Big). \end{aligned}$$

From Lemma 2.3.4 we have $\sum_{j=1}^{n} j\alpha_X(j) = o(n)$, which establishes the result. ■

Using the above result on the fourth moment of X_j we can establish the following result, which is used in the proof of Theorem 2.4.6.

Lemma A.3.2. *Under the conditions of Theorem 2.4.6 we can construct sequences (q_n) and (p_n) such that for $n \to \infty$:*

(i) $p_n \to \infty$, $p_n = o(n)$,

(ii) $q_n \to \infty$, $q_n = o(p_n)$,

(iii) $k_n = \frac{n}{p_n}\Big(1 + o(1)\Big)$,

(iv) $\frac{q_n^2 k_n}{n p_n} = o(1)$,

(v) $k_n \alpha_X(q_n) = o(1)$,

(vi) $\frac{k_n}{\sigma_n^4}\mathbb{E}\Big[\Big(\sum_{j=1}^{p_n} X_j\Big)^4\Big] = o(1)$.

Proof. From Lemma A.3.1 we know $\mathbb{E}[S_n^4] = o(n^3)$ such that

$$\mathbb{E}[S_n^4] = n^3\gamma(n) \leq n^3\tilde{\gamma}(n), \tag{A.6}$$

where

$$\gamma(n) \to 0 \qquad \text{and} \qquad \tilde{\gamma}(n) = \sup_{j\geq n}\gamma(j).$$

We define (p_n) and (q_n) by

$$p_n = \min\{p \in \mathbb{N} : p \geq \sqrt{n}\,|\log\tilde{\gamma}(p)|\} \qquad \text{and} \qquad q_n = \Big\lfloor \frac{n}{p_n} \Big\rfloor.$$

Finally we recall $k_n = \Big\lfloor \frac{n}{p_n + q_n} \Big\rfloor$.

<u>Ad (i):</u>

Clearly $p_n \to \infty$ for $n \to \infty$. Lapynov's inequality (A.3) with $s = 2$ and $t = 4$ yields

$$\mathbb{E}\Big[\Big(\sum_{j=1}^{n} X_j\Big)^4\Big] \geq \Big\{\mathbb{E}\Big[\Big(\sum_{j=1}^{n} X_j\Big)^2\Big]\Big\}^2 = \sigma^4 n^2(1 + o(1)).$$

Combining this with (A.6) leads to

$$\tilde{\gamma}(n) \geq \frac{\sigma^4}{n}(1 + o(1)),$$

such that

$$\lim_{n\to\infty} n\tilde{\gamma}(n) > 0. \tag{A.7}$$

From this we conclude there exists a $c > 0$ such that for sufficiently large n we have $\frac{c}{n} < \tilde{\gamma}(n) < 1$. Then we have

$$\begin{aligned}\sqrt{n}\,|\log\tilde{\gamma}(n)| &\leq \sqrt{n}\,|\log\big(\tfrac{c}{n}\big)| \\ &\leq \sqrt{n}\,|\log(c) - \log(n)| \\ &\leq \sqrt{n}\big(|\log(c)| + |\log(n)|\big)\end{aligned}$$

Thus for sufficiently large n we have $\sqrt{n}\,|\log\tilde{\gamma}(n)| \leq n$, which is the condition in the definition of p_n. Since p_n is the least integer satisfying this we have $p_n \leq n$ for sufficiently large n, and thus

$$\begin{aligned} p_n &\leq \sqrt{n}\,|\log\tilde{\gamma}(p_n)| + 1 \\ &\leq \sqrt{n}\,|\log\tilde{\gamma}(n)| + 1 \\ &\leq \sqrt{n}\big(|\log(c)| + |\log(n)|\big) + 1 \end{aligned}$$

From this we conclude

$$\frac{p_n}{n} \leq \frac{\sqrt{n}(|\log(n)| + |\log(c)|) + 1}{n} = o(1),$$

such that $p_n = o(n)$.

Ad (ii):

From the definition of q_n it is clear that since $p_n = o(n)$ then $q_n \to \infty$. Furthermore we have from the definition of p_n that

$$\frac{n}{p_n^2} \leq \frac{n}{n\big|\log\big(\tilde{\gamma}(p_n)\big)\big|^2} = o(1). \tag{A.8}$$

From this we conclude

$$\frac{q_n}{p_n} \leq \frac{n}{p_n^2} = o(1),$$

such that $q_n = o(p_n)$.

Ad (iii):

Directly from the definition of k_n we have

$$k_n = \Big\lfloor \frac{n}{p_n + q_n} \Big\rfloor = \frac{\frac{n}{p_n}}{1 + \frac{q_n}{p_n}}\Big(1 + o(1)\Big) = \frac{n}{p_n}\Big(1 + o(1)\Big). \tag{A.9}$$

Ad (iv):

Using $q_n = O(\frac{n}{p_n})$, $k_n = O(\frac{n}{p_n})$ and (A.8) yields

$$\frac{q_n^2 k_n}{n p_n} = O\Big(\frac{n^2}{p_n^4}\Big) = o(1).$$

Ad (v):

From Lemma 2.3.4 we know $q_n \alpha_X(q_n) = o(1)$. Since $q_n = \left\lfloor \frac{n}{p_n} \right\rfloor = \frac{n}{p_n}\big(1+o(1)\big)$ we have

$$k_n \alpha_X(q_n) = \frac{k_n}{q_n} q_n \alpha_X(q_n) = \frac{\frac{n}{p_n}\big(1+o(1)\big)}{\frac{n}{p_n}\big(1+o(1)\big)} o(1) = o(1). \tag{A.10}$$

Ad (vi):

Recalling (2.20) we have $\sigma_n^{-4} = \frac{1}{\sigma^4 n^2}\big(1+o(1)\big)$. Using this and (A.6) yields

$$\begin{aligned} \frac{k_n}{\sigma_n^4} \mathbb{E}\Big[\Big(\sum_{j=1}^{p_n} X_j\Big)^4\Big] &\leq k_n \sigma_n^{-4} p_n^3 \tilde{\gamma}(p_n) \\ &= \frac{n}{p_n}\big(1+o(1)\big)\frac{1}{\sigma^4 n^2}\big(1+o(1)\big) p_n^3 \tilde{\gamma}(p_n) \\ &= \frac{p_n^2 \tilde{\gamma}(p_n)}{n}\big(1+o(1)\big) \end{aligned}$$

From the definition of p_n we know, that $p_n < n^{\frac{1}{2}} |\log\big(\tilde{\gamma}(p_n)\big)| + 1$, which yields

$$\begin{aligned} \frac{k_n}{\sigma_n^4} \mathbb{E}\Big[\Big(\sum_{j=1}^{p_n} X_j\Big)^4\Big] &= O\Big(\frac{n|\log\big(\tilde{\gamma}(p_n)\big)|^2 \tilde{\gamma}(p_n)}{n}\Big) \\ &= O\Big(|\log\big(\tilde{\gamma}(p_n)\big)|^2 \tilde{\gamma}(p_n)\Big). \end{aligned}$$

This concludes the proof since $\tilde{\gamma}(p_n) \to 0$ for $n \to \infty$ and $\lim_{x \to 0} x\big(\log(x)\big)^2 = 0$. ■

In Chapter 3 we use the Lipschitz continuity of the complex exponential function several times, and it is derived below.

Lemma A.3.3. *The function $f : \mathbb{R} \to \mathbb{C}$ defined by $f(t) = \exp(iat)$, for some $a \in \mathbb{R}$ is Lipschitz continuous with Lipschitz constant $\sqrt{2}|a|$, i.e. for all $s, t \in \mathbb{R}$*

$$|\exp(ias) - \exp(iat)| \leq \sqrt{2}|a||s-t|.$$

Proof. For a real-valued differentiable function $f : \mathbb{R} \to \mathbb{R}$ we have in general

$$|f(s) - f(t)| \leq \max_{t \in [s,t]} |f'(t)||s-t|,$$

and in particular for all $s, t \in \mathbb{R}$ we have

$$|\cos(as) - \cos(at)| \leq |a||s-t| \qquad \text{and} \qquad |\sin(as) - \sin(at)| \leq |a||s-t|.$$

The result follows by direct calculation

$$\begin{aligned}|\exp(ias)-\exp(iat)| &= \sqrt{|\cos(as)-\cos(at)|^2+|\sin(as)-\sin(at)|^2}\\ &\leq \sqrt{2}|a||s-t|.\end{aligned}$$ ■

The final lemma in this appendix is used in the proof of Theorem 3.2.3.

Lemma A.3.4. *Using the notation in Theorem 3.2.3 we have*

$$F^{-1}_{\widetilde{Y}^{L,B}}(z) = \frac{\sigma_n^2}{\rho^{-1}-\nu^{L,B}\ln(z)} \qquad \textit{and} \qquad F^{-1}_{\widetilde{Y}^{\infty}}(z) = \frac{\sigma_n^2}{\rho^{-1}-\ln(z)}$$

where ρ is the signal to noise ratio, σ_n^2 is the noise power, and

$$\nu^{L,B} = \int_0^{\frac{L}{B}} p(t)dt.$$

Proof. First we recall that

$$H^{\infty}(f) \sim \mathbb{C}\mathcal{N}(0,1), \quad \text{for all } f \in \mathbb{R},$$

as remarked after the proof of Proposition 3.1.2. From Section 2.1.4 (page 18) we then know $|H^{\infty}(f)|^2 \sim \text{Exp}(1)$.

From the definition of $H^{L,B}$ we have

$$H^{L,B}(f) = \sum_{l=0}^{L-1} h_l^B \exp\left(-2\pi i f \tfrac{l}{B}\right).$$

Since this is a linear transformation of the complex Gaussian variables $h_1^B, \dots, h_{L-1}^B$ we conclude that $H^{L,B}(f)$ is complex Gaussian for all $f \in \mathbb{R}$. The mean value is

$$\begin{aligned}\mathbb{E}\big[H^{L,B}(f)\big] &= \mathbb{E}\Big(\sum_{l=0}^{L-1} h_l^B \exp\left(-2\pi i f \tfrac{l}{B}\right)\Big)\\ &= \sum_{l=0}^{L-1} \exp\left(-2\pi i f \tfrac{l}{B}\right)\mathbb{E}\big(h_l^B\big) = 0.\end{aligned}$$

To calculate the variance we use the autocovariance function (3.5), which yields

$$\begin{aligned}\text{Var}\big[H^{L,B}(f)\big] &= R_H^{L,B}(0)\\ &= \sum_{l=0}^{L-1} \int_{\frac{l}{B}}^{\frac{l+1}{B}} p(t)\exp\left(-2\pi i 0 \tfrac{l}{B}\right)dt\\ &= \int_0^{\frac{L}{B}} p(t)dt\\ &= \nu^{L,B}.\end{aligned}$$

We have thus shown that $H^{L,B}(f) \sim \mathbb{C}\mathcal{N}(0, \nu^{L,B})$, which as before implies

$$|H^{L,B}(f)|^2 \sim \text{Exp}\Big(\frac{1}{\nu^{L,B}}\Big).$$

Now we define $t : [0, \infty) \to [0, \rho\sigma_n^2]$ by

$$t(x) = \frac{\sigma_n^2}{x + \frac{1}{\rho}}.$$

Then for all $f \in \mathbb{R}$ we have

$$Y^{L,B}(f) = t\big(|H^{L,B}(f)|^2\big) \quad \text{and} \quad Y^{\infty}(f) = t\big(|H^{\infty}(f)|^2\big).$$

Our goal is to find the inverse distribution function for these random variables. Both cases have the same structure. We wish to find the inverse distribution function of $Y = t(X)$, where $X \sim \text{Exp}(\lambda)$. For $Y^{L,B}$ we have $\lambda = \nu^{L,B}$, whereas $\lambda = 1$ for Y^{∞}. To calculate the density of Y we remark

$$t^{-1}(y) = \frac{\sigma_n^2}{y} - \frac{1}{\rho} \quad \text{and} \quad \Big|\frac{d}{dy} t^{-1}(y)\Big| = \frac{\sigma_n^2}{y^2}.$$

Consequently the density of Y is

$$\begin{aligned} f_Y(y) &= \frac{1}{\lambda}\frac{\sigma_n^2}{y^2} \exp\Big(-\frac{\sigma_n^2}{\lambda y} + \frac{1}{\lambda\rho}\Big) \\ &= \exp\Big(\frac{1}{\lambda\rho}\Big)\frac{\sigma_n^2}{\lambda y^2} \exp\Big(-\frac{\sigma_n^2}{\lambda y}\Big). \end{aligned}$$

The distribution function of Y is then

$$F_Y(y) = \int_0^y \exp\Big(\frac{1}{\lambda\rho}\Big)\frac{\sigma_n^2}{\lambda z^2} \exp\Big(-\frac{\sigma_n^2}{\lambda z}\Big) dz.$$

Making the change of variables $x = \frac{\sigma_n^2}{\lambda z}$ yields

$$\begin{aligned} F_Y(y) &= \exp\Big(\frac{1}{\lambda\rho}\Big) \int_{\frac{\sigma_n^2}{\lambda y}}^{\infty} \exp(-x) dx \\ &= \exp\Big(\frac{1}{\lambda\rho}\Big) \exp\Big(\frac{-\sigma_n^2}{\lambda y}\Big) \end{aligned}$$

From this we express the inverse distribution function as

$$F_Y^{-1} = \frac{\sigma_n^2}{\frac{1}{\rho} - \lambda \ln(z)},$$

and the result follows by inserting $\lambda = 1$ and $\lambda = \nu^{L,B}$. ■

Notation

The list below provides an insight to the basic notational conventions used in the thesis.

$^\top$	Transpose of matrix or vector (e.g. $X^\top$).
$\overline{}$	Complex conjugation (e.g. $\overline{X}$).
$\hat{}$	Fourier transform (e.g. $\hat{f}$).
*	Conjugate transpose of matrix or vector (e.g. A^*)).
$*$	Convolution (e.g. $f * g$).
$\lfloor\ \rfloor$	Integer part of a real number (e.g. $\lfloor x \rfloor$).
i	Imaginary unit $i = \sqrt{-1}$.
$\mathcal{I}(\cdot)$	Indicator function.
$L_2(\mathbb{R})$	The set of square integrable functions on $\mathbb{R}$.
$L_2(\Omega)$	The set of functions on Ω square integrable with respect to P.
$L_2(X, \mu)$	The set of functions on X square integrable with respect to μ.
$\mathbb{N}$	The set of natural numbers $\{1, 2, \dots\}$.
$\mathbb{Z}$	The set of integer numbers $\{\dots, -1, 0, 1, \dots\}$.
$\mathbb{R}$	The set of real numbers.
$\mathbb{R}_+$	The set of positive real numbers.
$\mathbb{C}$	The set of complex numbers.
$\mathrm{Exp}(\cdot)$	Exponential distribution.
$\mathcal{N}_d(\cdot, \cdot)$	Normal distribution of dimension d.
$\mathbb{C}\mathcal{N}_d(\cdot, \cdot)$	Complex normal distribution of dimension d.
$\mathrm{Rayleigh}(\cdot)$	Rayleigh distribution.

Bibliography

Andersen, H., Højbjerre, M., Sørensen, D., and Eriksen, P. (1995). *Linear and Graphical Models for the Multivariate Complex Normal Distribution*. Springer.

Barriac, G. and Madhow, U. (2004). Characterizing Outage Rates for Space-Time Communication Over Wideband Channels. IEEE Transactions on Communications, 52(12): 2198–2208.

Bhatia, R. (1997). *Matrix Analysis*. Springer.

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons.

Billingsley, P. (1986). *Probability and Measure*. John Wiley & Sons.

Bradley, R. C. (2005). Basic Properties of Strong Mixing Conditions - A Survey and Some Open Questions. Probability Surveys, 2: 107–144.

Cioffi, J. M. (2005). *EE379A course reader*. Stanford University. . http://www.stanford.edu/class/ee379a/.

Davis, P. J. (1994). *Circulant Matrices*. Chelsea Publishing.

Domínguez, M. (1992). Mixing Coefficient, Generalized Maximal Correlation Coefficients, and Weakly Positive Measures. Journal of Multivariate Analysis, 43: 110–124.

Doob, J. L. (1953). *Stochastic Processes*. John Wiley & Sons.

Doukhan, P. (1994). *Mixing - Properties and Examples*. Springer.

Dym, H. and McKean, H. P. (1976). *Gaussian Processes, Function Theory, and the Inverse Spectral Problem*. Academic Press.

Hogben, L. (2006). *Handbook of Linear Algebra*. CRC Press.

Ibragimov, I. A. and Linnik, V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff Publishing.

Jakes, W. C. (1994). *Microwave Mobile Communications*. IEEE Press.

Pereira, S., Kyritsi, P., Papanicolaou, G., and Paulraj, A. (2006). Asymptotic Properties of Richly Scattering Ultra Wideband Channels. Working paper.

Piterbarg, V. I. (1996). *Asymptotic Methods in the Theory of Gaussian Processes and Fields*. American Mathematical Society.

Proakis, J. G. (2001). *Digital Communications*. McGraw-Hill.

Serfling, R. J. (1968). Contributions to Central Limit Theory for Dependent Variables. The Annals of Mathematical Statistics, 39(4): 1158–1175.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press.

van der Vaart, A. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.