Charles University in Prague
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

DOCTORAL THESIS

RNDr. Šárka Došlá

Stationary processes with negatively correlated random variables

Supervisor: Prof. RNDr. Jiří Anděl, DrSc.
Branch: 4M4 - Probability and Mathematical Statistics

Prague, 2010

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta
katedra pravděpodobnosti a matematické statistiky

DISERTAČNÍ PRÁCE

RNDr. Šárka Došlá

Stacionární posloupnosti negativně korelovaných náhodných veličin

Školitel: Prof. RNDr. Jiří Anděl, DrSc.

Obor: 4M4 Pravděpodobnost a matematická statistika

Praha 2010

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. RNDr. Jiří Anděl, DrSc., for his patient guidance, encouragement, and valuable advice during my research. I am extremely grateful for having such supportive and inspiring supervisor. His comments, suggestions, and careful corrections of all my manuscripts contributed enormously to the production of this thesis.

I owe my deepest gratitude to my family for their support and patience during my whole studies.

This work was supported by the grant 201/05/H007 of the Grant Agency of the Czech Republic, the grant 56808 of the Grant Agency of Charles University in Prague, the grant 201/08/0486 of the Grant Agency of the Czech Republic, and the research project MSM 0021620839 financed by MŠMT.

Prague March 4, 2010 Šárka Došlá

Contents

Abstract

Title: Stationary processes with negatively correlated random variables
Author: RNDr. Šárka Došlá
Department: Department of Probability and Mathematical Statistics
Supervisor: Prof. RNDr. Jiří Anděl, DrSc.
Supervisor's e-mail address: *jiri.andel@mff.cuni.cz*
Abstract: We deal with weakly stationary processes with negatively correlated random variables. First, we provide an introduction to the studied topic. We explain the considered model and relate it to other concepts of negative dependence among random variables. Then we investigate general processes with negatively correlated variables. Their properties and characteristics of their autocorrelation functions are derived. A possible construction in the time domain as well as in the spectral domain is discussed. Finally, we restrict our attention to sequences of 0-1 valued (Bernoulli) negatively correlated variables. Several models for this situation are considered.
Keywords: negatively correlated random variables, autocorrelation function, clipping model, 0-1 valued variables

Název práce: Stacionární posloupnosti negativně korelovaných náhodných veličin
Autor: RNDr. Šárka Došlá
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí disertační práce: Prof. RNDr. Jiří Anděl, DrSc.
E-mail vedoucího: *jiri.andel@mff.cuni.cz*

Abstrakt: V práci se zabýváme slabě stacionárními posloupnostmi negativně korelovaných náhodných veličin. Nejdříve je představen a blíže popsán uvažovaný model. Poté jsou zkoumány obecné procesy negativně korelovaných veličin. Jsou odvozeny některé jejich důležité vlastnosti a vlastnosti jejich autokorelační funkce. Dále jsou diskutovány možnosti konstrukce takových procesů, a to jak v časové doméně, tak i ve spektrální doméně. V poslední kapitole se omezíme na zkoumání posloupností binárních negativně korelovaných veličin. Uvažujeme zde několik obecných modelů pro tuto situaci.
Klíčová slova: negativně korelované náhodné veličiny, autokorelační funkce, clipping model, 0-1 veličiny

Chapter 1

Introduction

Let $\mathbf{X} = \{X_t, t \in \mathbb{Z}\}$ be a real *(weakly) stationary* random process. By this we mean that $\mathbf{X}$ has finite second order moments, $\mathsf{E}X_t^2 < \infty$ for all $t \in \mathbb{Z}$, and satisfies $\mathsf{E}X_t = \mu \in \mathbb{R}$ for all $t \in \mathbb{Z}$ and $\mathsf{cov}\,(X_t, X_{t+k}) = \mathsf{cov}\,(X_0, X_k)$ for all $k, t \in \mathbb{Z}$. Let $\{r_k, k \in \mathbb{Z}\}$ denote the autocorrelation function of $\mathbf{X}$, i.e. $r_k = \mathsf{cor}\,(X_t, X_{t+k})$. In the following we always consider the nontrivial case $\mathsf{var}\, X_t > 0$ for all $t \in \mathbb{Z}$, and then we can assume that the autocorrelation function $\{r_k\}$ is well-defined.

It is well-known that the autocorrelation function $\{r_k\}$ satisfies the relation $r_{-k} = r_k$, $k \in \mathbb{Z}$, and it is always positive semidefinite. We refer to Doob (1953) and Brockwell and Davis (1991) for some other properties of general weakly stationary processes and their autocorrelation functions.

Suppose that the autocorrelations r_k satisfy

$$r_k = \mathsf{cor}\,(X_t, X_{t+k}) \leq 0 \quad \text{for all} \quad k = 1, 2, \ldots. \tag{1.1}$$

Then we call $\mathbf{X}$ the process with *non-positively correlated variables*. If at least one of the inequalities in (1.1) is sharp then we call $\mathbf{X}$ the process with *negatively correlated variables*. Such processes arise in some practical situations, and therefore it is desirable to investigate their properties, see Bondesson (2003). The objective of this thesis is to study properties of processes with negatively correlated random variables and their autocorrelation functions.

This thesis is structured as follows. Chapter 1 provides an introduction to the studied topic. Section 1.1 explains the studied model and Section 1.2 relates it to other concepts of negative dependence among random variables. Some examples of processes with negatively correlated variables are presented in Section 1.3. They illustrate the broadness of the class of negatively correlated variables.

In Chapter 2 we investigate general processes with negatively correlated variables. Their properties and characteristics of their autocorrelation functions are presented in Sections 2.1 and 2.2. In Section 2.3 we show how to generate processes with negatively correlated variables.

In Chapter 3 we restrict our attention to sequences of 0-1 valued (Bernoulli) negatively correlated variables. Models for 1-dependent variables are investigated in Section 3.1. Estimators of some characteristics in these models are constructed in Section 3.2. Some more general models for negatively correlated Bernoulli variables are studied in Sections 3.3 and 3.4.

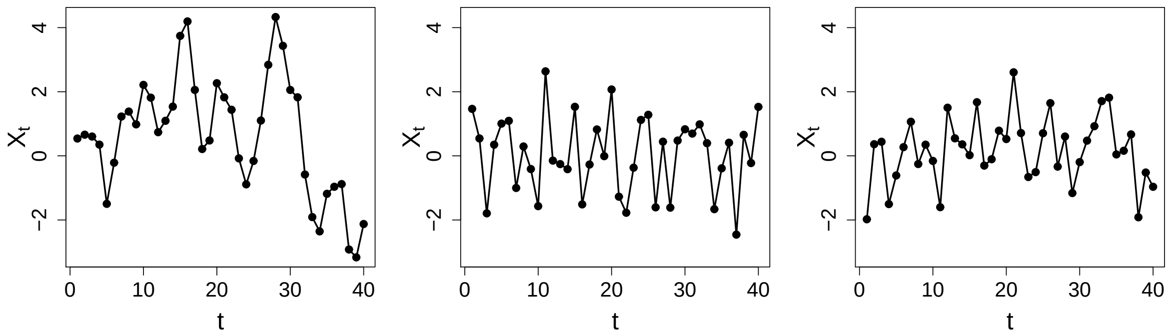
New results in this thesis include:

- properties of general negatively correlated variables in Theorems 2.5 and 2.7,
- properties of a spectral density and an autocorrelation function of a general process with negatively correlated variables in Theorem 2.1, Theorem 2.13, and Remark 2.6,
- sufficient conditions for variables being negatively correlated
  - expressed in the time domain in Theorem 2.10,
  - expressed in the spectral domain in Theorem 2.19,
- models for sequences of negatively correlated variables with given autocorrelations in Section 2.3.3,
- properties of the Bondesson's model from Bondesson (2003) for some particular distributions in Section 3.1.1,
- generalization of the Bondesson's model in Section 3.1.2,
- estimation of parameters of the generalized Bondesson's model in Section 3.2,
- numerical studies,
- a number of illustrative examples and figures.

1.1 Motivation

It is generally accepted that a correlation coefficient between two variables U, V describes somehow their dependence. The sign of this correlation gives us an idea whether the variables are related in a "positive" or "negative" sense, and the absolute value of the correlation reflects the "strength" of this dependence. This intuition is completely correct if the two variables U, V have a joint Gaussian distribution. However, the correlation reflects only the dependence in second order moments, and therefore the described conception can fail in a non-Gaussian case. (It is very easy to construct an example of dependent variables with a vanishing correlation coefficient.) For this reason, other measures of dependence between two variables have been proposed, and several definitions of positive and negative dependence have been offered, see Lehmann (1966). Nevertheless, the correlation coefficient still remains to be the most common and widespread measure of dependence, and the intuition described in the beginning of this paragraph is generally "tolerated" and used. In the following we sometimes use this conception keeping in mind that it is only an informal illustration providing an insight.

What is the intuition behind the condition (1.1)? The negative sign of r_k reflects some kind of negative dependence between the variables X_t and X_{t+k} for all $t \in \mathbb{Z}$. Similarly, (1.1) reflects negative dependence among all the variables $\{X_t,\ t \in \mathbb{Z}\}$. Hence, the



(a) $a = -0.5$, $b = 0.5$ (b) $a = -0.2$, $b = -1$ (c) $a = -0.1$, $b = -0.2$

Figure 1.1: Realizations of an ARMA(1,1) process $X_t + aX_{t-1} = \varepsilon_t + b\varepsilon_{t-1}$ for several choices of a, b

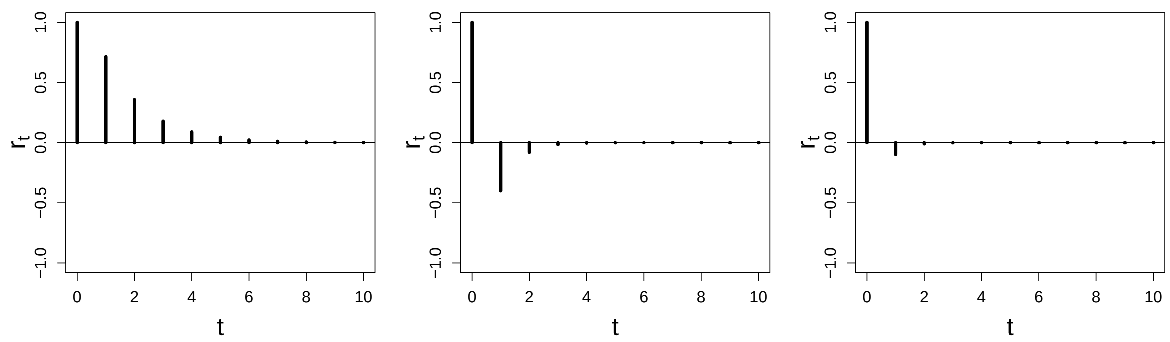
condition (1.1) provides us a qualitative description whether we consider the variables $\{X_t,\ t \in \mathbb{Z}\}$ as "negatively dependent" or not. Moreover, we will see in Chapter 2 that the inequality $\sum_{k=1}^{\infty} r_k \geq -1/2$ holds whenever (1.1) is satisfied. Hence, the quantity $\sum_{k=1}^{\infty} r_k$ might reflect a "strength" of the dependence. As stated before, this idea is only informal, and some more discussion on this topic is provided in Section 1.2.

Example 1.1. Figure 1.2 shows simulated realizations of three ARMA$(1, 1)$ processes $X_t + aX_{t-1} = \varepsilon_t + b\varepsilon_{t-1}$ for different choices of the parameters a, b. ARMA processes are analyzed in more detail in Section 1.3.2. At this point we just give a few comments on the plotted series as a motivation for the future work.

The case (a) corresponds to a process with all autocorrelations positive, the two cases (b) and (c) correspond to processes with negatively correlated variables. We can see that the process (a) seems to have a smoother realization than (b) and (c). This is due to the fact that in (a) the neighbouring variables are positively correlated, and thus X_t tends to have a similar value as X_{t-1}. On the other hand, we have purely negative correlation between the neighbouring variables in the cases (b) and (c), which results in rather oscillating realizations. Some slight differences in the "level of the oscillation" can be found between the graphs (b) and (c) as well.

The autocorrelations $\{r_t\}$ of the series (a), (b), (c) are plotted in Figure 1.1. Note that we have $r_1 = -0.400$ and $\sum_{t=1}^{\infty} r_t = -0.500$ in (b) and $r_1 = -0.098$ and $\sum_{t=1}^{\infty} r_t = -0.109$ in (c). Hence, the 1-lag autocorrelation r_1 and the sum $\sum_{t=1}^{\infty} r_t$ can be viewed as some "measures of the oscillation", i.e. measures of "negativeness" of the dependence among the variables $\{X_t,\ t \in \mathbb{Z}\}$.

All the three series were simulated in the program `R` version `2.10.1`, see R Development Core Team (2009), using the function `arima.sim` with setting `set.seed=2009` for each of the series and a Gaussian white noise.



(a) $a = -0.5$, $b = 0.5$ (b) $a = -0.2$, $b = -1$ (c) $a = -0.1$, $b = -0.2$

Figure 1.2: Autocorrelation function $\{r_t\}$ of an ARMA(1,1) process $X_t + aX_{t-1} = \varepsilon_t + b\varepsilon_{t-1}$ for several choices of a, b

1.2 Other Concepts of Negative Dependence

There have been several attempts to introduce a “good” qualitative measure of positive and negative dependence among random variables.

The concept of *positively associated* (PA) random variables was presented first by Esary et al. (1967). The random variables $X_1, \dots, X_n$ are said to be positively associated if

$$\mathsf{cov}\,[f(X_1, \dots, X_n), g(X_1, \dots, X_n)] \geq 0 \tag{1.2}$$

holds for all nondecreasing functions f and g for which the covariance in (1.2) exists. (Here, a function $f : \mathbb{R}^n \to \mathbb{R}$ is called nondecreasing if it is nondecreasing in each variable when the other variables are held fixed.) Esary et al. (1967) show that the positive association is preserved under taking subsets, forming unions of independent sets, forming unions of nondecreasing functions, and taking limits in distribution. These properties together with some other desirable characteristics of PA variables, listed in Esary et al. (1967), explain why the concept of positive association (1.2) has become an accepted and used measure of positive dependence.

The negative counterpart, the *negative association* (NA), was introduced 16 years later by Joag-Dev and Proschan (1983). Random variables $X_1, \dots, X_n$ are said to be *negatively associated* if for every pair of disjoint subsets A_1, A_2 of $\{1, \dots, n\}$

$$\mathsf{cov}\,[f_1(X_i, i \in A_1), f_2(X_j, j \in A_2)] \leq 0 \tag{1.3}$$

holds whenever f_1, f_2 are nondecreasing (in each coordinate) and the covariance in (1.3) exists. An infinite family $\{X_n, n \in \mathbb{N}\}$ is negative associated if every finite subfamily is negative associated. Negative association is related to, but it is not simply the dual of the positive association (1.2).

Joag-Dev and Proschan (1983) show that NA variables have a number of desirable properties. In particular, a subset of two or more NA random variables is NA, non-decreasing functions defined on disjoint subsets of NA random variables are NA, and a union of independent sets of NA random variables is NA. Furthermore, a number of well-known multivariate distributions possesses the NA property such as multinomial, multivariate hypergeometric, Dirchlet, negatively correlated Gaussian variables, permutation distribution, random sampling without replacement, joint distribution of ranks, and others.

The negative association (1.3) is not the only one possible qualitative measure of negative dependence among random variables that can be found in the literature. We refer to Lehmann (1966), Block et al. (1982), and Karlin and Rinott (1980) for some other approaches. However, compared to these types of negative dependence NA has one important advantage because only the NA class enjoys the property of being closed under formation of increasing functions of disjoint sets of random variables. This type of closure property is sometimes important in applications and it does not hold for the other types of negative dependence. This is the main reason why none of the other classes of dependent variables received so much attention as the NA class.

Even though the concept of negatively associated variables is quite attractive from the theoretical point of view, it is rather difficult to examine whether the condition (1.3) is satisfied in a given situation. Some sufficient conditions for NA can be found in Joag-Dev and Proschan (1983) and Hu and Hu (1999), but unfortunately, they do not bring much insight. In practice, one rather looks at the correlations among the given variables $X_1, \ldots, X_n$ in order to judge whether they are "positively or negatively dependent". If all the correlations have a positive sign then the variables are considered as "positively dependent" and if all the correlations have a negative sign then the variables are considered as "negatively dependent". Although this procedure has several drawbacks, it is widely used in practice mainly due to its simplicity. For this reason it makes sense to study negatively correlated variables, i.e. random variables with the correlation structure (1.1). Besides, as stated in Section 1.1, the sum $\sum_{k=1}^{\infty} r_k$ can serve as a quantitative measure of the mentioned dependence.

It is natural to ask the following question: How is the concept of NA related to stationary sequences of negatively correlated variables? Obviously, if $\mathbf{X} = \{X_n, n \in \mathbb{N}\}$ is a sequence of NA variables then all the correlations $\mathsf{cor}\,(X_k, X_l)$, $k \neq l$, are non-positive. In addition, if $\mathbf{X}$ is stationary then it satisfies the condition (1.1), and it is a process with negatively correlated variables. On the other hand, Joag-Dev and Proschan (1983) give an example of four variables with non-positive correlations which do not satisfy the NA condition (1.3), see our Example 1.2. This means that negatively correlated variables do not have to be negatively associated, and thus the concept of NA is more general. However, it is clear that the condition (1.1) is easier to work with in practice than the condition (1.3).

Example 1.2. This example comes from Joag-Dev and Proschan (1983). Let X_1, X_2, X_3, X_4 be 0-1 valued random variables such that $\mathsf{P}(X_i = 1) = 1/2$ for $i = 1, 2, 3, 4$, and let (X_1, X_2) and (X_3, X_4) have the same bivariate distributions. Assume that the joint distribution of (X_1, X_2, X_3, X_4) is given in Table 1.1. Then

$$\begin{aligned}&\mathsf{cov}\,(X_1, X_2) = \mathsf{cov}\,(X_3, X_4) = -0.01,\\ &\mathsf{cov}\,(X_1, X_3) = \mathsf{cov}\,(X_1, X_4) = \mathsf{cov}\,(X_2, X_4) = \mathsf{cov}\,(X_2, X_3) = 0.\end{aligned}$$

On the other hand, $\mathsf{P}(X_1 = X_2 = X_3 = X_4 = 1) > \mathsf{P}(X_1 = X_2 = 1)\mathsf{P}(X_3 = X_4 = 1)$, and this violates the NA condition.

	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,0)$	0.0577	0.0623	0.0623	0.0577
$(0,1)$	0.0623	0.0677	0.0677	0.0623
$(1,0)$	0.0623	0.0677	0.0677	0.0623
$(1,1)$	0.0577	0.0623	0.0623	0.0577

Table 1.1: Joint distribution of (X_1, X_2, X_3, X_4)

1.3 Examples of Negatively Correlated Variables

In the following we briefly introduce some examples of processes with the correlation structure (1.1). These examples illustrate that the class of negatively correlated variables is quite broad and involves a number of various well-known and widely used processes.

1.3.1 Clipping Model

Let $\{X_t,\ t \in \mathbb{Z}\}$ be independent identically distributed (iid) random variables and $c \in \mathbb{R}$. Define

$$\xi_t = \begin{cases} 1 & \text{if } X_t - X_{t-1} < c,\\ 0 & \text{otherwise.}\end{cases}$$

Then $\{\xi_t,\ t \in \mathbb{Z}\}$ is a process with 0-1 valued 1-dependent negatively correlated variables. This model, introduced in Bondesson (2003), is a special case of so called *clipping model*, cf. Kedem (1980a). We describe its properties and discuss some generalizations in Chapter 3.

1.3.2 ARMA Models

Let $\mathbf{X} = \{X_t,\ t \in \mathbb{Z}\}$ follow a stationary ARMA(m, n) model

$$\sum_{k=0}^{m} a_k X_{t-k} = \sum_{j=0}^{n} b_j \varepsilon_{t-j}, \tag{1.4}$$

where $a_0 = 1$, $b_0 = 1$, and $\boldsymbol{\varepsilon} = \{\varepsilon_t,\ t \in \mathbb{Z}\}$ is a *white noise* with $\mathsf{var}\ \varepsilon_t = \sigma^2$. By this we mean that the variables ε_t are uncorrelated with zero mean and a common variance $\sigma^2 \in (0, \infty)$, and we use the notation $\mathsf{WN}(0, \sigma^2)$. Recall that $\mathbf{X}$ is stationary whenever the roots of $1 + a_1 z + \cdots + a_m z^m = 0$ lie outside the unit circle. We refer to Hamilton (1994) for a summary of properties of ARMA processes.

The form of the autocorrelation function $\{r_t\}$ of the process $\mathbf{X}$ is well known, see for example Hamilton (1994): The sequence $\{r_t\}_{t=n+1}^{\infty}$ is a solution of a mth-order difference equation governed by the autoregressive parameters $a_1, \dots, a_m$, and $r_1, \dots, r_n$ are obtained from n initial conditions (equations) derived from (1.4). If the roots of the characteristic polynomial of the difference equation are real and distinct then r_t takes the form of a sum of m decaying exponentials. However, the form of r_t can be more complicated, for example an arithmetic-geometric sequence or a sinusoidal pattern with decaying amplitude etc. Since the form of $\{r_t\}$ is known, it is possible to investigate under which conditions $\mathbf{X}$ given by (1.4) is a process with negatively correlated variables.

Example 1.3. Let $\mathbf{X} = \{X_t,\ t \in \mathbb{Z}\}$ follow an ARMA(1,1) model

$$X_t + aX_{t-1} = \varepsilon_t + b\varepsilon_{t-1}$$

with $a \in (-1, 1)$, $b \in [-1, 1]$, $a \neq b$, and $a, b \neq 0$. Calculations give

$$r_1 = \frac{(1 - ab)(b - a)}{1 - 2ab + b^2}, \quad r_k = -ar_{k-1} = (-a)^{k-1} r_1 \quad \text{for } k \geq 2.$$

The condition (1.1) is fulfilled if and only if $-1 \leq b < a < 0$ holds. In this case $\mathbf{X}$ is a process with negatively correlated variables.

In Example 1.3 we were able to state directly the conditions for the parameters a, b under which the corresponding ARMA(1,1) process consists of negatively correlated variables. However, the situation becomes considerably more complicated for higher order ARMA models, see Example 1.4. Hence, a different approach is used in Chapter 2.

Example 1.4. Let $\mathbf{X} = \{X_t,\ t \in \mathbb{Z}\}$ follow an ARMA$(1, 2)$ model

$$X_t + aX_{t-1} = \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2}$$

with $a \in (-1, 1)$, $a \neq 0, b_2 \neq 0$, and $a^2 - b_1 a + b_2 \neq 0$. The last condition ensures that the model cannot be simplified to an ARMA model of a lower order. Assume further that

$$1 + b_1 z + b_2 z^2 \neq 0 \quad \text{for all} \quad |z| < 1 \tag{1.5}$$

holds. The latter condition ensures the invertibility of $\mathbf{X}$. The process $\mathbf{X}$ is stationary because we have $|a| < 1$.

The calculations give

$$\begin{aligned} \mathsf{var}\, X_t &= \frac{1 - 2ab_1 + b_1^2 + 2a^2b_2 - 2ab_1b_2 + b_2^2}{(1-a)(1+a)}\,\sigma^2, \\ r_1 &= -\frac{a - b_1 - a^2b_1 + ab_1^2 + ab_2 + a^3b_2 - b_1b_2 - a^2b_1b_2 + ab_2^2}{1 - 2ab_1 + b_1^2 + 2a^2b_2 - 2ab_1b_2 + b_2^2}, \\ r_2 &= \frac{(a^2 - ab_1 + b_2)(1 - ab_1 + a^2b_2)}{1 - 2ab_1 + b_1^2 + 2a^2b_2 - 2ab_1b_2 + b_2^2}, \end{aligned}$$

and $r_k = (-a)^{k-2} r_2$ for $k \geq 3$. For which a, b_1, b_2 the autocorrelation function $\{r_k\}$ satisfies (1.1)?

Let us first have a look at the condition (1.5). It holds if and only if the roots z_1, z_2 of the equation $z^2 + b_1 z + b_2 = 0$ satisfy $|z_i| < 1$, $i = 1, 2$. It is possible to show after some algebraic manipulations that this is equivalent to the conditions $b_1 + b_2 \geq -1$, $b_1 - b_2 \leq 1$, $b_2 \leq 1$. If in addition $b_2 < 1/(4b_1^2)$ then the roots of $z^2 + b_1 z + b_2 = 0$ are both real and distinct.

Now let us study under which conditions (1.1) holds. In view of our assumptions the considered model cannot be simplified to a model of a lower order, and $r_2 \neq 0$. This implies the constraint $a < 0$. Furthermore, it follows from (1.5) that $1 - b_1 a + b_2 a^2 > 0$, and thus $r_2 < 0$ if and only if $a^2 - ab_1 + b_2 < 0$. Hence, $r_2 < 0$ if and only if the roots z_1, z_2 are both real and $z_1 < -a < z_2$. For a given a we have obtained the constraints for b_1, b_2 (triangle in $\mathbb{R}^2$): $b_1 + b_2 \geq -1$, $b_1 - b_2 \leq 1$, $b_2 < ab_1 - a^2$. Finally, the condition $r_1 \leq 0$ is equivalent to

$$a - b_1 - a^2b_1 + ab_1^2 + ab_2 + a^3b_2 - b_1b_2 - a^2b_1b_2 + ab_2^2 \geq 0.$$

This further restricts the possible values of a, b_1, b_2 in a quite nontrivial way.

1.3.3 Fractional Gaussian Noise

An example of a process with the correlation structure (1.1) is a fractional Gaussian noise with the Hurst exponent $H \in (0, 1/2)$. We briefly introduce the definition of this term.

Definition 1.5. A zero mean Gaussian process $\{B_H(t),\ t \geq 0\}$ is called *fractional Brownian motion* (FBM) if it satisfies $B_H(0) = 0$, and $\mathsf{E}[B_H(t) - B_H(s)]^2 = \sigma^2 |t - s|^{2H}$ for some $\sigma > 0$ and $0 < H < 1$. A discrete time process $\mathbf{X} = \{X_t,\ t \in \mathbb{Z}\}$ is called *fractional Gaussian noise* (FGN) if it is an increment process of a fractional Brownian motion, i.e.

$$X_t = B_H(t) - B_H(t-1) \quad \text{for} \quad t = 1, 2, \ldots.$$

The number $H \in (0, 1)$ is referred to as the *Hurst exponent*.

The fractional Brownian motion is a generalization of the (ordinary) Brownian motion, which is obtained as a special case of FBM with $H = 1/2$. It is possible to show that the fractional Gaussian noise $\mathbf{X}$ does exist, and it is a stationary Gaussian process with the autocorrelation function

$$r_t = \frac{1}{2}\left[(t+1)^{2H} + |t-1|^{2H} - 2t^{2H}\right] \text{ for } t \geq 0, \tag{1.6}$$

see Samorodnitsky (2007). It follows from (1.6) that if $0 < H < 1/2$ then $r_t < 0$ for all $t \geq 1$. Therefore, FGN with $0 < H < 1/2$ is an example of a process with negatively correlated variables. Moreover, we get $\sum_{t=-\infty}^{\infty} r_t = 0$, which can be further equivalently rewritten as $\sum_{t=1}^{\infty} r_t = -1/2$. It is possible to show that the inequality $\sum_{t=1}^{\infty} r_t \geq -1/2$ holds for all correlation functions satisfying (1.1), see (2.1) in Chapter 2. This means that the sum of correlations of a FGN always reaches the lower bound $-1/2$.

Remark 1.6. Beran (1994) claims that FGN with $0 < H < 1/2$ is encountered in practice rather rarely, mainly because the condition $\sum_{t=-\infty}^{\infty} r_t = 0$ is unstable in the sense that arbitrarily small disturbances destroy this property. On the other hand, he admits that FGN with $0 < H < 1/2$ may occur after overdifferencing, cf. Samorodnitsky (2007).

Remark 1.7. The summability of correlations, i.e. the condition $\sum_{t=1}^{\infty} |r_t| < \infty$, is often taken as an indication of so called *short memory* while the divergence of $\sum_{t=1}^{\infty} |r_t|$ is taken as a definition of so called *long range dependence* or *long memory*, see Samorodnitsky (2007) or Beran (1994). It follows from (1.6) that if $1/2 < H < 1$, then $r_t > 0$ for all $t \geq 1$ and $\sum_{t=1}^{\infty} r_t = \infty$. Hence, FGN with $1/2 < H < 1$ is a popular example of a stationary process with a long memory. For this reason, FBM with $H \in (1/2, 1)$ is an attractive tool in some financial applications, and the interest has centered in this case in the empirical work, see Robinson (2003).

1.3.4 Stationary Increments of Self-Similar Processes.

The fractional Gaussian noise is not the only one random process with the autocorrelation function given in (1.6). There exists a number of non-Gaussian processes with the autocorrelations specified by (1.6) with $0 < H < 1/2$. These processes are therefore other examples of negatively correlated variables.

Definition 1.8. A process $\mathbf{Y} = \{Y(t),\ t \geq 0\}$ is called a *self-similar* (SS) process with a self-similarity parameter $H \in \mathbb{R}$ if $\{Y(ct),\ t \geq 0\} \overset{D}{=} \{c^H Y(t),\ t \geq 0\}$ for all $c > 0$.

Let $\mathbf{Y}$ be a self-similar process with stationary increments (denoted as SSSI) and with $0 < H < 1$. Suppose that the process $\mathbf{Y}$ is centered and its variance is finite. Define the increments $X_t = Y(t) - Y(t-1)$ for $t \in \mathbb{N}$. Then $\mathbf{X} = \{X_t,\ t \in \mathbb{N}\}$ is a stationary process whose autocorrelation function coincides with that of the fractional

Gaussian noise, given by (1.6), see Samorodnitsky (2007). In particular, if $0 < H < 1/2$ then the process $\mathbf{X}$ has the desirable correlation structure (1.1).

It is possible to show that the fractional Brownian motion is a special case of SSSI processes. Samorodnitsky (2007) further shows how to construct SSSI processes with a finite variance, different from the FBM, via multiple Wiener-Itô integrals.

Chapter 2

Negatively Correlated Variables

In this chapter we deal with general stationary random processes with negatively correlated random variables. Hence, we assume that $\mathbf{X} = \{X_t, t \in \mathbb{Z}\}$ is a real stationary process with the autocorrelation function $\{r_t\}$ satisfying (1.1). We can assume without loss of generality that $\mathbf{X}$ is centered, i.e. $\mathsf{E}X_t = 0$ for all $t \in \mathbb{Z}$.

Some properties of these processes and their autocorrelation functions are investigated in Sections 2.1 and 2.2. In Section 2.3 we show how to model processes with negatively correlated variables. This chapter contains results from Došlá and Anděl (2010).

2.1 Autocorrelation Function

It is easy to characterize the autocorrelation function of negatively correlated Gaussian variables. It is shown in Bondesson (2003) that a sequence of real numbers $\{r_k, k \in \mathbb{Z}\}$ such that $r_0 = 1$, $r_{-k} = r_k$ for $k \in \mathbb{Z}$, and $r_k \leq 0$ for $k \geq 1$, is the autocorrelation function of a stationary Gaussian process $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ if and only if

$$\sum_{k=1}^{\infty} r_k \geq -\frac{1}{2} \tag{2.1}$$

holds.

Moreover, the inequality (2.1) holds for any weakly stationary process $\mathbf{X}$ with negatively correlated variables. Indeed, if $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ is a stationary process with the autocorrelation function $\{r_t\}$ satisfying (1.1) and $\mathsf{var}\, X_t = \sigma^2 > 0$ then

$$\mathsf{var}\left(\sum_{k=1}^{N} X_k\right) = N\sigma^2 + 2\sigma^2 \sum_{k=1}^{N} (N-k) r_k \geq 0$$

holds for all $N \in \mathbb{N}$. Dividing the latter inequality by N and letting $N \to \infty$ implies that $\sum_{k=1}^{\infty} r_k \geq -1/2$ holds.

The inequality (2.1) is crucial and has several consequences as we will see later in this text. At this point note that (2.1) implies that if the autocorrelation function $\{r_t\}$ satisfies (1.1) then necessarily $0 \geq r_t \geq -1/2$ holds for all $t \geq 1$. In particular, we always have $r_1 \geq -1/2$. This gives us a strong limitation for the value of the correlation between two neighbouring variables of the process $\mathbf{X}$. On the other hand, if the condition (1.1) does not hold for $\{r_t\}$ then the 1-lag autocorrelation r_1 of a stationary random process can be arbitrarily close to 1 or -1. Furthermore, if (1.1) holds for $\{r_t\}$ then (2.1) implies $\sum_{k=1}^{\infty} |r_k| < \infty$. This means that a process with negatively correlated variables has always a short memory (see Remark 1.7 for the definition of this notion).

Recall that the negativeness of the correlations r_k indicate that the variables $\{X_t\}$ are somehow negatively dependent. The quantity $\sum_{k=1}^{\infty} r_k$ further reflects "the strength" of the dependence. The following theorem shows how the minimal value $-1/2$ in (2.1) can be attained.

Theorem 2.1. *Let $\mathbf{X} = \{X_t, t \in \mathbb{Z}\}$ be a stationary process with the autocorrelation function $\{r_k, k \in \mathbb{Z}\}$ such that $r_k \leq 0$ for $k \geq 1$. Then there exists a continuous spectral density f of the process X_t and $\sum_{k=1}^{\infty} r_k = -1/2$ holds if and only if $f(0) = 0$.*

Proof. Let $\{R_k\}$ be the autocovariance function of $\mathbf{X}$. The inequality (2.1) implies that $\sum_{k=-\infty}^{\infty} |R_k| < \infty$. In this case it is well-known that the spectral density f of $\mathbf{X}$ is given as $f(\lambda) = [1/(2\pi)] \sum_{k=-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}k\lambda} R_k$. This Fourier series is absolutely summable, and therefore f is continuous on the interval $[-\pi, \pi]$. Furthermore, we have $\sum_{k=1}^{\infty} R_k = \pi f(0) - R_0/2$, or equivalently $\sum_{k=1}^{\infty} r_k = \pi f(0)/R_0 - 1/2 \geq -1/2$. Obviously, the bound $-1/2$ is reached if and only if $f(0) = 0$. □

Example 2.2. We have seen in Section 1.3.3 that a fractional Gaussian noise $\mathbf{X}$ with the Hurst exponent $0 < H < 1/2$ is an example of negatively correlated variables such that $\sum_{t=1}^{\infty} r_t = -1/2$ holds. In view of (2.1) this means that the sum $\sum_{t=1}^{\infty} r_t$ reaches its minimal value $-1/2$. Keeping in mind the intuition that $\sum_{k=1}^{\infty} r_k$ reflects "the strength" of the dependence among variables X_t we can say that FGN is a collection of negatively correlated random variables which are "negatively dependent as much as possible".

The spectral density $f(\lambda)$ of a fractional Gaussian noise can be derived in the usual way, see Samorodnitsky (2007). If $H \neq 1/2$ then

$$f(\lambda) = \frac{\sigma^2}{2} C(H)(1 - \cos\lambda) \sum_{j=-\infty}^{\infty} |2\pi j + \lambda|^{-(1+2H)}, \tag{2.2}$$

where

$$C(H) = \frac{2H(1-2H)}{\Gamma(2-2H)} \frac{1}{\cos \pi H} = \frac{2}{\pi} \Gamma(2H+1) \sin(\pi H)$$

and $\lambda \in (-\pi, \pi)$, cf. Geweke and Porter-Hudak (1983). It follows from (2.2) that $f(0) = 0$ whenever $0 < H < 1/2$. This corresponds to the conclusion of Theorem 2.1.

Example 2.3. The proof of Theorem 2.1 shows that the statement

$$\sum_{t=1}^{\infty} r_t = -1/2 \text{ if and only if } f(0) = 0$$

holds for a general autocorrelation function $\{r_t\}$ such that $\sum_{t=1}^{\infty} |r_t| < \infty$. Conditions under which the equality $\sum_{t=1}^{\infty} r_t = -1/2$ holds for ARMA processes introduced in Section 1.3.2 can be then easily derived.

Let $\mathbf{X}$ follow a stationary ARMA(m, n) model $\sum_{k=0}^{m} a_k X_{t-k} = \sum_{j=0}^{n} b_j \varepsilon_{t-j}$, where $\{\varepsilon_t\}$ is a white noise with $\mathsf{var}\ \varepsilon_t = \sigma^2$. The spectral density f of $\mathbf{X}$ is given as

$$f(\lambda) = \frac{\sigma^2}{2\pi} \Big| \sum_{j=0}^{n} b_j \mathrm{e}^{-\mathrm{i}j\lambda} \Big|^2 \Big| \sum_{k=0}^{m} a_k \mathrm{e}^{-\mathrm{i}k\lambda} \Big|^{-2}, \quad \lambda \in [-\pi, \pi].$$

Hence, the equality in (2.1) is reached if and only if $\sum_{j=0}^{n} b_j = 0$. This equality holds if and only if 1 is a root of the MA polynomial $\sum_{j=0}^{n} b_j z^j$. This demonstrates that the equality in (2.1) is reached neither for autoregressive sequences nor for an invertible ARMA process.

Example 2.4. Let $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ be a white noise $\mathsf{WN}(0, \sigma^2)$. Consider an MA(1) process $X_t = \varepsilon_t - \varepsilon_{t-1}$. The autocorrelations of $\mathbf{X}$ are given as $r_0 = 1$, $r_1 = -1/2$, and $r_k = 0$ for $k \geq 2$. Hence, $\mathbf{X}$ is a simple example of a process with negatively correlated variables such that $\sum_{t=1}^{\infty} r_t = -1/2$ holds.

2.2 General Properties

In Example 2.3 we have obtained conditions for the parameters of an ARMA model such that $\sum_{t=1}^{\infty} r_t = -1/2$. However, $\sum_{t=1}^{\infty} r_t = -1/2$ does not imply the condition (1.1), i.e. $r_t \leq 0$ for all $t \geq 1$. The process $X_t = \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}$ is an example. It is much more difficult to specify conditions under which (1.1) holds for an ARMA process, see Section 1.3.2. Since each stationary ARMA model can be represented as a linear process, see Hamilton (1994), one can equivalently investigate conditions for the MA(∞) parameters which ensure that the resulting process satisfies (1.1). This approach turns out to be more convenient, and the following Theorem 2.5 justifies its use for general stationary processes with negatively correlated variables.

Theorem 2.5. *Let $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ be a real weakly stationary process with the autocorrelation function $\{r_k,\, k \in \mathbb{Z}\}$ such that $r_k \leq 0$ for all $k \geq 1$. Then $\mathbf{X}$ can be represented as a linear process*

$$X_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \tag{2.3}$$

where $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ is a white noise with $\mathsf{var}\,\varepsilon_t = 1$, and $\{c_k\}_0^{\infty}$ is a sequence of real constants such that $\sum_{k=0}^{\infty} c_k^2 < \infty$.

Proof. Without loss of generality assume that $\mathsf{var}\, X_t = 1$. Then $\{r_k\}$ is the autocovariance function of $\mathbf{X}$. It is shown in Gichman and Skorochod (1965) (cf. our Theorem A.3) that a weakly stationary process can be represented as a linear process (2.3) if and only if its spectral distribution function is absolutely continuous and its spectral density f satisfies the condition

$$\int_{-\pi}^{\pi} \ln f(\lambda)\, \mathrm{d}\lambda > -\infty. \tag{2.4}$$

Remind that the inequality (2.1) holds, and assume first that $\sum_{t=1}^{\infty} r_t > -1/2$. Since the spectral density is given as $f(\lambda) = [1/(2\pi)]\,[1 + 2\sum_{t=1}^{\infty} r_t \cos(t\lambda)]$, we obtain

$$2\pi f(\lambda) = 1 + 2\sum_{t=1}^{\infty} r_t \cos(t\lambda) = 1 - 2\sum_{t=1}^{\infty} |r_t| \cos(t\lambda) \geq 1 - 2\sum_{t=1}^{\infty} |r_t| > 0 \tag{2.5}$$

for all $\lambda \in [-\pi, \pi]$. In this case $f(\lambda)$ is positive and continuous on $[-\pi, \pi]$, and therefore $\ln[f(\lambda)]$ is continuous on $[-\pi, \pi]$. Hence, the integral in (2.4) is finite.

Now let $\sum_{t=1}^{\infty} r_t = -1/2$. Then we can write

$$2\pi f(\lambda) = 1 - 2\sum_{t=1}^{\infty} |r_t| \cos(t\lambda) = 2\sum_{t=1}^{\infty} |r_t|[1 - \cos(t\lambda)]. \tag{2.6}$$

Since $|r_t| \geq 0$ and $1 - \cos(t\lambda) \geq 0$ for all $t \geq 1$, the equality (2.6) implies that $f(\lambda) = 0$ if and only if $|r_t|[1 - \cos(t\lambda)] = 0$ for all $t \geq 1$. This is always satisfied for $\lambda = 0$, and thus $\lambda = 0$ is always a zero point of $f(\lambda)$ (cf. Theorem 2.1). Furthermore, $\lambda \in [0, \pi]$ is a root of $f(\lambda) = 0$ if and only if $\cos(t\lambda) = 1$ for all $t \geq 1$ such that $r_t \neq 0$. Remind that we assume that $\sum_{t=1}^{\infty} |r_t| = 1/2$, and therefore the trivial case $r_t = 0$ for all $t \geq 1$ is eliminated.

Suppose that $f(\lambda) > 0$ for all $\lambda \in (0, \pi]$, i.e. $\lambda = 0$ is the only root of $f(\lambda) = 0$ in the interval $[0, \pi]$ (this assumption is removed in the next step). Since f is even, we can write $\int_{-\pi}^{\pi} \ln[f(\lambda)]\, \mathrm{d}\lambda = 2\int_0^{\pi} \ln[f(\lambda)]\, \mathrm{d}\lambda$. The function $\ln[f(\lambda)]$ is continuous on $(0, \pi]$, and therefore it suffices to investigate its behavior in a neighborhood of the point 0. Let $\delta > 0$ be sufficiently small, and take $T \in \mathbb{N}$ such that $T\delta < \pi/2$. We can assume that there exists $r_t \neq 0$ for some $1 \leq t \leq T$ (otherwise, we would choose smaller δ). Denote $A(\lambda) = \sum_{t=1}^{T} |r_t|[1 - \cos(t\lambda)]$ and $B(\lambda) = \sum_{t=T+1}^{\infty} |r_t|[1 - \cos(t\lambda)]$. Then $\pi f(\lambda) = A(\lambda) + B(\lambda)$. Obviously, $B(\lambda) \geq 0$ for all $\lambda \in [0, \pi]$. Furthermore, $(2t^2\lambda^2)/\pi^2 < 1 - \cos(t\lambda)$ holds for all $\lambda \in (0, \delta)$. Hence, we get

$$\frac{2\lambda^2}{\pi^2} \sum_{t=1}^{T} |r_t|\, t^2 < A(\lambda).$$

Denote $C = \sum_{t=1}^{T} |r_t|\, t^2$. Then $C > 0$ and we have obtained the inequality $f(\lambda) > 2C\lambda^2/\pi^3$ which implies that

$$\ln[f(\lambda)] > 2\ln\lambda + \ln\left[\frac{2C}{\pi^3}\right] \qquad \text{for all } \lambda \in (0, \delta).$$

Hence, $\int_0^\delta \ln[f(\lambda)]\,\mathrm{d}\lambda$ converges and so does $\int_0^\pi \ln[f(\lambda)]\,\mathrm{d}\lambda$.

Finally, let $\lambda_0 \neq 0$ be a root of $f(\lambda) = 0$ in the interval $[0, \pi]$. It follows from (2.6) that $\cos(t\lambda_0) = 1$ for all t such that $r_t \neq 0$. In other words, $r_t = 0$ for all t such that $t\lambda_0$ is not a multiple of 2π. Obviously, λ_0 must be of the form $\lambda_0 = l\pi/s$ for some $l, s \in \mathbb{N}$, $1 \le l \le s$ such that the greatest common divisor of l and s equals 1. (Otherwise we would get $r_t = 0$ for all $t \ge 1$, and this is not possible.) It further follows that if l is odd then $r_t = 0$ for all t such that $t \neq 2sk$ for all $k \in \mathbb{N}$, $k \ge 1$. Similarly, if l is even then $r_t = 0$ for all t such that $t \neq sk$ for all $k \in \mathbb{N}$, $k \ge 1$. Assume that l is odd (the situation for even l is analogous). We can then write

$$2\pi f(\lambda) = 2\sum_{k=1}^{\infty} |r_{2sk}|[1 - \cos(2sk\lambda)].$$

Since $\cos[2sk(l\pi/s - \lambda)] = \cos(2sk\lambda) = \cos[2sk(l\pi/s + \lambda)]$, the behavior of $f(\lambda)$ in a neighborhood of the point $l\pi/s$ is the same as in a neighborhood of the point 0. The previous conclusions imply that $\int_0^\pi \ln[f(\lambda)]\,\mathrm{d}\lambda$ converges. □

Remark 2.6. Let us point out several properties of the spectral density f from the proof of Theorem 2.5.

(i) In (2.5) we have seen that

$$2\pi f(\lambda) = 1 + 2\sum_{t=1}^{\infty} r_t \cos(t\lambda) \ge 1 + 2\sum_{t=1}^{\infty} r_t = 2\pi f(0),$$

and thus $f(\lambda) \ge f(0)$ holds for all $\lambda \in [0, \pi]$.

(ii) It is shown in the proof of Theorem 2.5 that if $\sum_{t=1}^\infty r_t > -1/2$ then f has no zero points. Furthermore, if $\sum_{t=1}^\infty r_t = -1/2$ and $r_1 \neq 0$ then $\lambda = 0$ is the only root of $f(\lambda) = 0$ on the interval $[-\pi, \pi]$.

(iii) We have shown that if $\lambda_0 \in (0, \pi]$ is a root of $f(\lambda) = 0$ then $r_t = 0$ for all t such that $t\lambda_0$ is not a multiple of 2π. In particular, if $\sum_{t=1}^\infty r_t = -1/2$ and $f(\pi/s) = 0$ for some $s \ge 1$, $s \in \mathbb{N}$, then $f(l\pi/s) = 0$ for all $l = 1, \dots, s$ and $r_t = 0$ for all $t \ge 1$ such that $t \neq 2ks$ for all $k \in \mathbb{Z}$. For $s = 1$ it follows that if $f(\pi) = 0$ then $r_t = 0$ for all odd $t \in \mathbb{Z}$.

The representation (2.3) of the process $\mathbf{X}$ is not unique in general. However, there exists a unique linear process (2.3) such that $c_0 > 0$ and the zeros of the function $\sum_{j=0}^\infty c_j z^j$ do not lie in the interior of the unit disc, see Remark A.4 in Appendix. In the following we always consider such representation due to the possible invertibility of the linear process. The coefficients c_k from this representation can be obtained in the following way (see Remark A.4 in Appendix for more details). Define

$$\begin{aligned} d_k &= \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}k\lambda} \ln[f(\lambda)]\,\mathrm{d}\lambda = 2\int_0^{\pi} \ln[f(\lambda)]\cos[k\lambda]\,\mathrm{d}\lambda, \quad k = 0,1,2\ldots, \\ P &= \exp\Big\{\frac{d_0}{4\pi}\Big\} = \exp\Big\{\frac{1}{4\pi}\int_{-\pi}^{\pi} \ln f(\lambda)\,\mathrm{d}\lambda\Big\}, \\ h(z) &= \exp\Big\{\frac{1}{2\pi}\sum_{k=1}^{\infty} d_k z^k\Big\}. \end{aligned} \tag{2.7}$$

The function $h(z)$ belongs to the Hardy space H_2, and therefore $h(z) = \sum_{k=0}^{\infty} a_k z^k$ with $\sum_{k=0}^{\infty} a_k^2 < \infty$ and $a_0 = h(0) = 1$, see Remark A.2. The coefficients $\{c_k\}$ from (2.3) are given as $c_k = \sqrt{2\pi}Pa_k$ for $k \geq 0$.

The linear process (2.3) in Theorem 2.5 is standardized in the way that $\operatorname{var}\varepsilon_t = 1$. However, it is sometimes useful to work with the representation (2.3) where c_0 is set to be 1. The following theorem presents some properties of the coefficients from such representation.

Theorem 2.7. *Let $\{X_t,\, t \in \mathbb{Z}\}$ satisfy the assumptions of Theorem 2.5. Then $\{X_t\}$ can be represented as a linear process $X_t = \sum_{k=0}^{\infty} a_k\varepsilon_{t-k}$ such that $a_0 = 1$, $\sum_{k=0}^{\infty} a_k^2 < \infty$, and the zeros of $\sum_{k=0}^{\infty} a_k z^k$ do not lie in the interior of the unit disc. The process $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ is a white noise that belongs to the space generated by the values of $\{X_t,\, t \in \mathbb{Z}\}$, and $\operatorname{var}\varepsilon_t = \sigma^2 = 2\pi\exp\left\{[1/(2\pi)]\int_{-\pi}^{\pi}\ln f(\lambda)\,\mathrm{d}\lambda\right\}$, $0 < \sigma^2 < \infty$.*

The coefficients $\{a_k\}_{k=0}^{\infty}$ satisfy $0 \leq \sum_{k=0}^{\infty} a_k \leq 1$, and the equality

$$\left(\sum_{k=0}^{\infty} a_k\right)^2 = \left(\sum_{k=0}^{\infty} a_k^2\right)\left(1 + 2\sum_{k=1}^{\infty} r_k\right) \tag{2.8}$$

holds. In particular, $\sum_{k=0}^{\infty} a_k = 0$ if and only if $\sum_{k=1}^{\infty} r_k = -1/2$, and $\sum_{k=0}^{\infty} a_k = 1$ if and only if $r_k = 0$ for all $k \geq 1$ (and so $a_k = 0$ for all $k \geq 1$).

Proof. Recall that $\{X_t\}$ can be represented as (2.3) with $c_k = \sqrt{2\pi}Pa_k$, where a_k, P are defined in (2.7), and the zeros of $\sum_{k=0}^{\infty} c_k z^k$ do not lie in the interior of the unit disc. Equivalently, we can write $X_t = \sum_{k=0}^{\infty} a_k\tilde{\varepsilon}_{t-k}$, where $\{\tilde{\varepsilon}_t\}$ is a white noise such that $\tilde{\varepsilon}_t = \sqrt{2\pi}P\varepsilon_t$. Then $\operatorname{var}\tilde{\varepsilon}_t = \sigma^2 = 2\pi P^2 = 2\pi\exp\left\{[1/(2\pi)]\int_{-\pi}^{\pi}\ln f(\lambda)\,\mathrm{d}\lambda\right\}$.

The property $\sum_{k=0}^{\infty} a_k \geq 0$ follows from the fact that $a_0 = 1$ and the zeros of $\sum_{k=0}^{\infty} a_k z^k$ do not lie in the interior of the unit disc. The continuous spectral density f of the process $\{X_t\}$ satisfies $f(\lambda) = [\sigma^2/(2\pi)]\left|\sum_{k=0}^{\infty} a_k\mathrm{e}^{-\mathrm{i}k\lambda}\right|^2$ for all $\lambda \in [-\pi,\pi]$. Since $\operatorname{var}X_t = \sigma^2\left(\sum_{k=0}^{\infty} a_k^2\right)$, we get

$$f(\lambda) = \frac{\sigma^2\left(\sum_{k=0}^{\infty} a_k^2\right)}{2\pi}\left(1 + 2\sum_{t=1}^{\infty} r_t\cos(t\lambda)\right) = \frac{\sigma^2}{2\pi}\left|\sum_{k=0}^{\infty} a_k\mathrm{e}^{-\mathrm{i}k\lambda}\right|^2$$

for all $\lambda \in [-\pi, \pi]$. The equality (2.8) is obtained for $\lambda = 0$. Since $\sum_{k=0}^{\infty} a_k^2 = 1 + \sum_{k=1}^{\infty} a_k^2 \geq 1$, it then follows from (2.8) that $\sum_{k=0}^{\infty} a_k = 0$ if and only if $\sum_{k=1}^{\infty} r_k = -1/2$.

To see that $\sum_{k=0}^{\infty} a_k \leq 1$ holds we show that $f(0) \leq \sigma^2/(2\pi)$. If $\sum_{k=1}^{\infty} r_k = -1/2$ then $f(0) = 0$, and the inequality holds trivially. If $\sum_{k=1}^{\infty} r_k > -1/2$ it follows from (2.5) that $f(\lambda) \geq f(0) > 0$ holds for all $\lambda \in [-\pi, \pi]$, and

$$\frac{\sigma^2}{2\pi} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln f(\lambda)\, \mathrm{d}\lambda\right\} \geq \exp\{\ln f(0)\} = f(0).$$

Hence, $f(0) = [\sigma^2/(2\pi)] \left(\sum_{k=0}^{\infty} a_k\right)^2 \leq \sigma^2/(2\pi)$, and thus $\sum_{k=0}^{\infty} a_k \leq 1$. The equality holds if and only if $f(\lambda) = f(0)$ for all $\lambda \in [-\pi, \pi]$. This is the case only for $r_t = 0$ for all $t \geq 1$ (the process $\{X_t\}$ is a white noise). □

Obviously, the relationship between the autocorrelations r_k and the coefficients c_k (or a_k) is not trivial, and it is not possible to express c_k (or a_k) directly from r_k only using algebraic operations.

We have shown that except the trivial case $r_t = 0$ for all $t \geq 1$ the coefficients $\{a_k\}$ from Theorem 2.7 always satisfy

$$-1 \leq \sum_{k=1}^{\infty} a_k < 0. \tag{2.9}$$

The first inequality is due to the fact that we always choose the representation such that the zeros of $\sum_{k=0}^{\infty} a_k z^k$ do not lie inside the unit disc. The second inequality follows from the properties of a spectral density of a process with negatively correlated variables. The constraints in (2.9) are thus necessary for $\{a_k\}_{k=0}^{\infty}$ to define $\sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$ a process with negative autocorrelations.

Example 2.8. Let $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ be a centered stationary process with $\mathsf{var}\, X_t = 4$ and let the correlations of $\mathbf{X}$ be $r_0 = 1$ and

$$r_t = \begin{cases} -\dfrac{1}{t^2 - 1} & \text{for even } t,\ |t| \geq 2, \\ 0 & \text{for odd } t. \end{cases}$$

Then the spectral density f of $\mathbf{X}$ is given as $f(\lambda) = |\sin(\lambda)|$ for $\lambda \in [-\pi, \pi]$. Let us compute the coefficients $\{a_k\}$ from the representation in Theorem 2.7. Calculations give $\sigma^2 = \pi$ and

$$h(z) = \sqrt{1 - z^2} = \sum_{n=0}^{\infty} (-1)^n \binom{1/2}{n} z^{2n} \quad \text{for all} \quad |z| \leq 1.$$

Define

$$a_k = \begin{cases} 0, & k \text{ odd}, \\ (-1)^{k/2} \dbinom{1/2}{k/2}, & k \text{ even}. \end{cases}$$

The process $\mathbf{X}$ can be represented as $X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$ with $\mathsf{var}\, \varepsilon_t = \pi$, $a_0 = 1$. Notice that $a_k \le 0$ for all $k \ge 1$. Furthermore, we have $\sum_{t=1}^{\infty} r_t = -1/2$ and $\sum_{k=0}^{\infty} a_k = 0$ which illustrates the conclusion of Theorem 2.7.

Remark 2.9. Recall the well-known Wold's decomposition formulated in Theorem A.5. In view of this statement Theorem 2.5 claims that a process with negatively correlated variables is *always purely non-deterministic*.

2.3 Construction

In this section we dwell on the following question: How processes with negatively correlated variables can be modeled? Section 2.3.1 deals with a construction in the time domain, and the spectral domain construction is studied in Section 2.3.2. In Section 2.3.3 we investigate some special models of time series with a given autocorrelation function.

2.3.1 Time Domain

Theorems 2.5 and 2.7 show that the class of weakly stationary processes with non-positive autocorrelations is a subset of a class of all linear processes. Let $\{a_k\}_0^{\infty}$ be a sequence of real numbers such that $a_0 = 1$ and $\sum_{k=0}^{\infty} a_k^2 < \infty$. Let $\boldsymbol{\varepsilon} = \{\varepsilon_t,\, t \in \mathbb{Z}\}$ be a white noise with $\mathsf{var}\, \varepsilon_t = \sigma^2 > 0$. In the following we provide sufficient conditions for the coefficients $\{a_k\}$ under which the process $X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$ is a process with negatively correlated variables, i.e. it has the desirable property $r_k \le 0$ for all $k \ge 1$.

Theorem 2.10. *Let $\{a_k\}_{k=0}^{\infty}$ be a sequence of real numbers such that*

1. *$a_0 = 1$ and $a_k \le 0$ for all $k \ge 1$,*
2. *$\{a_k\}_{k=1}^{\infty}$ is a non-decreasing sequence,*
3. *$\sum_{k=1}^{\infty} a_k \ge -1$.*

Then the autocorrelation function $\{r_t\}$ of the process $X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$ satisfies $r_t \le 0$ for all $t \ge 1$. Moreover, $\sum_{t=1}^{\infty} r_t = -1/2$ holds if and only if $\sum_{k=1}^{\infty} a_k = -1$.

Proof. The autocovariance function of the process $\mathbf{X}$ is given as $R_t = \sigma^2 \sum_{k=0}^{\infty} a_k a_{t+k}$. Hence, $R_t \le 0$ for all $t \ge 1$ if and only if

$$-a_t \ge \sum_{k=1}^{\infty} a_{t+k} a_k \qquad \text{for all } t \ge 1. \tag{2.10}$$

In view of the assumptions we have $-a_k = |a_k|$, the sequence $\{|a_k|\}_{k=1}^{\infty}$ is non-increasing, and thus $|a_k| \ge |a_{t+k}|$ holds for all $k \ge 1$. It follows that

$$\sum_{k=1}^{\infty} a_k a_{t+k} = \sum_{k=1}^{\infty} |a_k||a_{t+k}| \le |a_t| \sum_{k=1}^{\infty} |a_k| \le |a_t|$$

for all $t \geq 1$ and (2.10) holds. Since $\sum_{k=0}^{\infty} a_k = 0$ if and only if $\sum_{k=1}^{\infty} a_k = -1$, the rest of the assertion follows from Theorem 2.7. □

Remark 2.11. Let us list several remarks to Theorem 2.10.

(i) We have stressed out that we always consider only linear processes $\sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$ such that the zeros of the function $\sum_{k=0}^{\infty} a_k z^k$ do not lie inside the unit disc. Under this condition it is shown in Theorem 2.7, see also (2.9), that $\sum_{k=1}^{\infty} a_k \geq -1$ must hold.

(ii) We know that $\sum_{k=1}^{\infty} a_k \leq 0$ holds for all processes with non-positive correlations, see (2.9). Theorem (2.10) works with a stronger assumption $a_k \leq 0$ for all $k \geq 1$.

(iii) The monotonicity of the sequence $\{a_k\}_{k=1}^{\infty}$ is only a sufficient condition as well. In Example 2.8 we have already presented an example where $a_k = 0$ for all odd k and $a_k \neq 0$ otherwise.

(iv) Note that the condition $a_k \leq 0$ for all $k \geq 1$ alone in general does not imply $r_t \leq 0$ for all $t \geq 1$. Consider an invertible MA(3) process $\mathbf{X}$ with $a_1 = -5/66$, $a_2 = -13/22$ and $a_3 = -5/33$ as an example (here $R_1 = 85\sigma^2/1452 > 0$).
However, for invertible MA(2) processes the condition $r_t \leq 0$ for $t = 1, 2$ is equivalent to $a_t \leq 0$ for $t = 1, 2$. Indeed, for MA(2) we have $R_2 = a_2$ and $R_1 = a_1 + a_1 a_2 = a_1(1 + a_2)$. The process is invertible and therefore, $a(z) = 1 + a_1 z + a_2 z^2 \neq 0$ for all $|z| \leq 1$. In particular, $a(1) = 1 + a_1 + a_2 > 0$ and $a(-1) = 1 - a_1 + a_2 > 0$. It follows that $1 + a_2 > 0$, and therefore $R_t \leq 0$ if and only if $a_t \leq 0$, $t = 1, 2$.

Example 2.12. Let $a > 0$, $0 < q < 1$ such that $q \leq 1/(1 + a)$. Let $\{\varepsilon_t, t \in \mathbb{Z}\}$ be a white noise with $\mathsf{var}\, \varepsilon_t = \sigma^2 > 0$. Define $a_0 = 1$ and $a_k = -aq^k$ for $k \geq 1$. Then $\{a_k\}_1^{\infty}$ is a non-decreasing sequence and the condition $\sum_{k=1}^{\infty} a_k \geq -1$ is satisfied. Hence, it follows from Theorem 2.10 that the linear process $X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$ is a process with negatively correlated variables. If $q < 1/(1 + a)$ then this process is invertible.

Indeed, the autocorrelations r_t of the process $\mathbf{X}$ are of the form

$$r_0 = 1 \quad \text{and} \quad r_t = -aq^t \frac{1 - (1 + a)q^2}{1 + (a^2 - 1)q^2}, \quad t \geq 1.$$

It is easy to see that r_t is negative for all $t \geq 1$ if and only if $q < 1/\sqrt{1 + a}$. This is satisfied since we assume $q < 1/(1 + a)$. Furthermore, we have

$$\sum_{t=1}^{\infty} r_t = -\frac{\tilde{a}q[1 - (1 + \tilde{a})q^2]}{[1 + (\tilde{a}^2 - 1)q^2](1 - q)}.$$

It can be shown that the equality $\sum_{t=1}^{\infty} r_t = -1/2$ occurs if and only if $q = 1/(1 + a)$, i.e. if $\sum_{k=1}^{\infty} a_k = -1$.

2.3.2 Spectral Domain

The requirement $r_t \leq 0$ for all $t \geq 1$ can be expressed in terms of the spectral density of the process $\mathbf{X}$ as well. Remember that we have shown that there always exists a continuous spectral density, see Theorem 2.1. Let us first summarize some of its general properties.

Theorem 2.13. *Let $\{r_t\}_0^\infty$ be a sequence of real numbers such that $r_0 = 1$, $r_t \leq 0$ for all $t \geq 1$, and $\sum_{t=1}^\infty r_t \geq -1/2$. Define $f(\lambda) = [1/(2\pi)]\,[1 + 2\sum_{t=1}^\infty r_t \cos(t\lambda)]$ for $\lambda \in [-\pi, \pi]$.*

1. *The function f is continuous, non-negative, and has a global minimum at the point $\lambda = 0$. If $\sum_{t=1}^\infty r_t \neq 0$ then 0 is also a strong local minimum.*
2. *If $f(0) = f(\pi) = 0$ and $f(\lambda) > 0$ for all $\lambda \in (0, \pi)$ then f is symmetric around the point $\pi/2$, i.e. $f(\lambda) = f(\pi - \lambda)$ for all $\lambda \in [0, \pi]$.*
3. *If $r_k = \mathcal{O}(k^{-2-\varepsilon})$ for some $\varepsilon > 0$ then f is differentiable, its derivative f' is continuous on $[-\pi, \pi]$ and is given as $f'(\lambda) = -\sum_{k=1}^\infty k r_k \sin(k\lambda)$. In particular, $f'(0) = 0$.*

Proof. 1. The inequality $f(\lambda) \geq f(0)$ holds for all $\lambda \in [-\pi, \pi]$, see Remark 2.6. If there exists $t_0 \geq 1$ such $r_{t_0} \neq 0$ then there exists $\varepsilon > 0$ such that $-|r_{t_0}|\cos(t_0\lambda) > -|r_{t_0}|$ for all $\lambda \in (-\varepsilon, \varepsilon) \smallsetminus \{0\}$. Hence, it follows from (2.5) that $f(\lambda) > f(0)$ holds for $\lambda \in (-\varepsilon, \varepsilon) \smallsetminus \{0\}$.

2. It follows from Remark 2.6 that $r_t = 0$ for all odd $t \in \mathbb{Z}$, and thus $f(\lambda) = [1/(2\pi)]\,[1 + \sum_{k=1}^\infty r_{2k}\cos(2k\lambda)]$. The assertion follows from the equality $\cos[2k(\pi - \lambda)] = \cos(2k\lambda)$.

3. If $r_k = \mathcal{O}(k^{-2-\varepsilon})$ then $\sum_{k=1}^\infty k r_k$ converges, and therefore $\sum_{k=1}^\infty k r_k \sin(k\lambda)$ converges uniformly for all $\lambda \in [-\pi, \pi]$. It follows from the theory of Fourier series, see for instance Hobson (1927), that $f'(\lambda) = -\sum_{k=1}^\infty k r_k \sin(k\lambda)$ for $\lambda \in [-\pi, \pi]$. □

Example 2.14. The continuous spectral density corresponding to the process with negatively correlated variables in Example 2.12 is given as

$$f(\lambda) = \frac{1}{2\pi}\frac{1 - 2(1+a)q\cos\lambda + (1 + 2a + a^2)q^2}{1 - 2q\cos\lambda + q^2}$$

for $\lambda \in [-\pi, \pi]$. This spectral density is plotted in Figure 2.1 for $q = 1/4$ and various choices of $a > 0$. The properties listed in Theorem 2.13 are illustrated. Recall that we have $r_k = -aq^k$, and therefore f is always differentiable on $[-\pi, \pi]$ with $f'(0) = 0$. In addition, $f(0) = 0$ if and only if $q = 1/(a+1)$.

Example 2.15. Revisit the fractional Gaussian noise from Section 1.3.3. It is possible to show that $r_k \sim H(2H-1)k^{-2(1-H)}$ as $k \to \infty$, see Samorodnitsky (2007). In particular, $r_k = \mathcal{O}(k^{-2+\varepsilon})$ for $\varepsilon = 2H > 0$, and the assumption of the statement 4. in Theorem 2.13

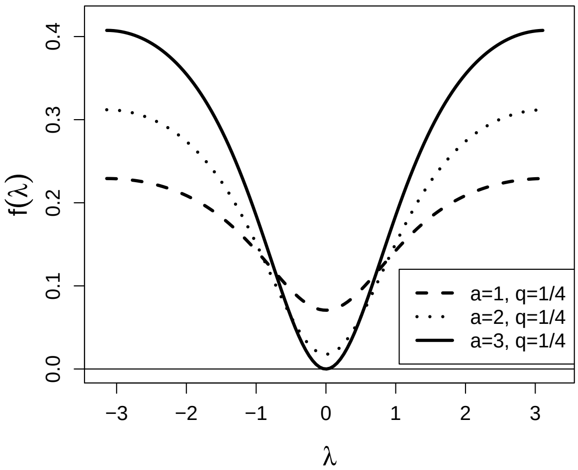


Figure 2.1: Spectral density of the process from Example 2.12 for $q = 1/4$ and various choices of a

is not satisfied. The formula for the spectral density of a FGN is given in (2.2) in Example 2.2. This function is plotted in Figure 2.2 for various choices of $H \in (0, 1/2)$. We can see that f is not differentiable at the point $\lambda = 0$.

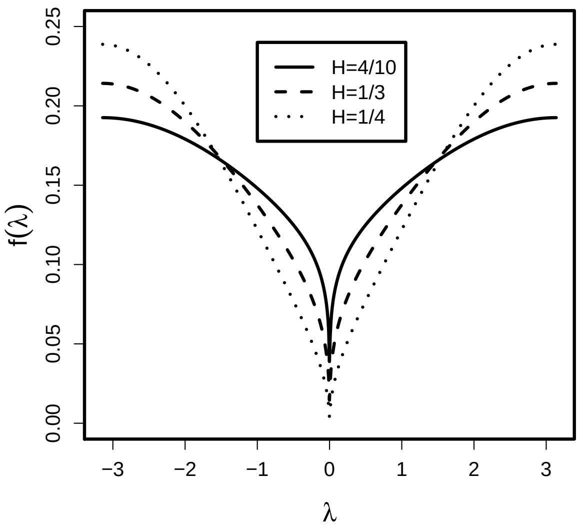


Figure 2.2: Spectral density (2.2) of the FGN for several choices of $H \in (0, 1/2)$

In the following we provide some sufficient conditions for the spectral density f such that the corresponding autocorrelations satisfy (1.1), i.e. $r_t \leq 0$ for all $t \geq 1$.

Lemma 2.16. *Let g be a measurable non-negative and non-increasing function on $[0, \pi]$, and let $n \in \mathbb{N}$, $n \geq 1$. Then $\int_0^\pi g(x)\sin(nx)\,\mathrm{d}x \geq 0$.*

Proof. Assume first that n is even. Then $\int_0^\pi g(x)\sin(nx)\,\mathrm{d}x = \sum_{i=0}^{n/2-1} I_i$, where $I_i = \int_{2i\pi/n}^{(2i+2)\pi/n} g(x)\sin(nx)\,\mathrm{d}y$. The substitution $y = nx$ gives

$$\begin{aligned} I_i &= \frac{1}{n}\int_{2i\pi}^{(2i+1)\pi} g(y/n)\sin(y)\,\mathrm{d}y + \frac{1}{n}\int_{(2i+1)\pi}^{(2i+2)\pi} g(y/n)\sin(y)\,\mathrm{d}y \\ &= \frac{1}{n}\int_{2i\pi}^{(2i+1)\pi} \Big[g\Big(\frac{y}{n}\Big) - g\Big(\frac{y+\pi}{n}\Big)\Big]\sin(y)\,\mathrm{d}y. \end{aligned}$$

Since $\sin(y) \geq 0$ for $y \in (0, \pi)$ and g is non-increasing the inequality $I_i \geq 0$ holds for all $i = 0, \ldots, n/2 - 1$. Hence, $\int_0^\pi g(x)\sin(nx)\,\mathrm{d}x \geq 0$.

For odd n we can write $\int_0^\pi g(x)\sin(nx)\,\mathrm{d}x = \sum_{i=0}^{(n-3)/2} I_i + J$, where I_i are defined as before and $J = \int_{(n-1)\pi/n}^{\pi} g(x)\sin(nx)\,\mathrm{d}x$. We have shown that $I_i \geq 0$ holds. The inequality $J \geq 0$ follows from the non-negativeness of g and due to the positive sign of $\sin(nx)$ on the interval $\big((n-1)\pi/n, \pi\big)$. This finishes the proof. □

Corollary 2.17. *Let f be a differentiable function on $(0, \pi]$ such that its derivative f' is non-negative and non-increasing on $(0, \pi)$. Then $\int_0^\pi f(x)\cos(nx)\,\mathrm{d}x \leq 0$ holds for any $n \in \mathbb{N}$, $n \geq 1$.*

Proof. Integration by parts gives

$$\int_0^\pi f(x)\cos(nx)\,\mathrm{d}x = \frac{1}{n} f(x)\sin(nx)\Big|_0^\pi - \frac{1}{n}\int_0^\pi f'(x)\sin(nx)\,\mathrm{d}x.$$

The first term is equal to 0, and the assertion follows from Lemma 2.16 □

Corollary 2.18. *Let f be a measurable function on $(0, \pi)$ such $f(\lambda) = f(\pi - \lambda)$ for all $\lambda \in [0, \pi]$. Let f be differentiable on $(0, \pi/2)$ such that the derivative f' is non-negative and non-increasing on $(0, \pi/2)$. Then for any $n \in \mathbb{N}$, $n \geq 1$, $\int_0^\pi f(x)\cos(nx)\,\mathrm{d}x \leq 0$ holds.*

Proof. The property $f(\lambda) = f(\pi - \lambda)$ implies that $\int_0^\pi f(x)\cos(nx)\,\mathrm{d}x = 0$ for all odd n and $\int_0^\pi f(x)\cos(nx) = 2\int_0^{\pi/2} f(x)\cos(nx)\,\mathrm{d}x$ for even n. The proof of the inequality $\int_0^{\pi/2} f(x)\cos(nx)\,\mathrm{d}x \leq 0$ for even n is analogous to the proof of Theorem 2.17. □

Theorem 2.19. *Let $\tilde{f} \geq 0$, $\tilde{f} \not\equiv 0$, satisfy the assumptions of Corollary 2.17 or 2.18. Then the function f defined as $f(\lambda) = \tilde{f}(|\lambda|)$ for $\lambda \in [-\pi, \pi]$ is a spectral density of a process with non-positive autocorrelations.*

Proof. The function f is non-negative on $[-\pi, \pi]$ and symmetric around 0, and therefore f is a spectral density of a real weakly stationary process $\mathbf{X}$. It follows from Corollary 2.17 and 2.18 that the autocovariances R_t of $\mathbf{X}$ satisfy

$$R_0 = \int_{-\pi}^{\pi} f(\lambda)\,\mathrm{d}\lambda = 2\int_0^{\pi} \tilde{f}(\lambda)\,\mathrm{d}\lambda > 0,$$
$$R_t = \int_{-\pi}^{\pi} f(\lambda)\mathrm{e}^{it\lambda}\,\mathrm{d}\lambda = 2\int_0^{\pi} \tilde{f}(\lambda)\cos(t\lambda)\,\mathrm{d}\lambda \leq 0 \quad \text{for } t \geq 1.$$

The autocorrelations of $\{X_t\}$ are well-defined, and they satisfy the condition (1.1). □

Theorem 2.19 gives a clue how a stationary process with the desirable property $r_t \leq 0$ for all $t \geq 1$ can be constructed in a spectral domain. The choice of $\tilde{f}$ such that $\tilde{f}(0) = 0$ further ensures that $\sum_{t=1}^{\infty} r_t = -1/2$.

Example 2.20. A simple example of a spectral density f constructed using Theorem 2.19 (with $\tilde{f}$ satisfying the assumptions of Corollary 2.17) is the function $f(\lambda) = k|\lambda|^{\alpha}$ for $k > 0$ and $\alpha \in (0, 1]$. For instance, if $f(\lambda) = |\lambda|$ then $R(0) = \pi^2$ and $r_k = -4/(\pi^2 k^2)$ for odd k and $r_k = 0$ for even k.

An example of f constructed from $\tilde{f}$, which satisfies the assumptions of Corollary 2.18, is the function $f(\lambda) = k|\sin\lambda|$, $k > 0$. The case $k = 1$ has been already discussed in Example 2.8.

Remark 2.21. (i) Let $\tilde{f} \geq 0$ satisfy the assumptions of Corollary 2.17. This means that $\tilde{f}$ is differentiable, non-decreasing and concave on $(0, \pi]$. If g is a non-decreasing concave function such that $g \circ \tilde{f}$ is differentiable on $(0, \pi]$ then $g \circ \tilde{f}$ satisfies the assumptions of Corollary 2.17 as well. This is the case for example for $g = \log$ if $\tilde{f}(\lambda) \neq 0$ for all $\lambda \in (0, \pi]$. In particular, if $\tilde{f}(\lambda) \geq 1$ for all $\lambda \in [0, \pi]$ then $\log(\tilde{f}(|\lambda|))$ is a spectral density of a process with non-positive autocorrelations as well. This is the case for example for $g = \log$ if $\tilde{f}(\lambda) \neq 0$ for all $\lambda \in (0, \pi]$. In particular, if $\tilde{f}(\lambda) \geq 1$ for all $\lambda \in [0, \pi]$ then $\log(\tilde{f}(|\lambda|))$ is a spectral density of a process with non-positive autocorrelations as well.

(ii) Corollary 2.18 considers a function symmetric around $\pi/2$ and non-decreasing on $(0, \pi/2)$. It is shown in the proof that $\int_0^{\pi} f(\lambda)\cos(n\lambda) \neq 0$ if and only if t is even. One could proceed even further and divide the interval $[0, \pi]$ into more pieces. For instance consider a function $\tilde{f}$ defined on $[0, \pi/3]$ and differentiable on $(0, \pi/3)$ with $\tilde{f}'$ non-negative and non-increasing. Take $f(\lambda) = \tilde{f}(\lambda)$ for $\lambda \in [0, \pi/3]$, $f(\lambda) = \tilde{f}(|\lambda - 2\pi/3|)$ for $\lambda \in (\pi/3, \pi]$. Then $\int_0^{\pi} f(\lambda)\cos(n\lambda) \leq 0$ for all $n \geq 1$ and $\int_0^{\pi} f(\lambda)\cos(n\lambda) \neq 0$ if and only if $n = 3k$ for some $k \in \mathbb{Z}$. Similarly, we could consider $\tilde{f}$ defined on $[0, \pi/4]$ etc. Using this approach we are able to construct for a fixed $j \in \mathbb{N}$ a process $\mathbf{X}$ with autocorrelations $r_t \leq 0$ for all $t \geq 1$ such that $r_t \neq 0$ only for $t = kj$, $k \in \mathbb{Z}$.

(iii) Finally, see that if $\tilde{f}$ satisfies the assumptions of Corollary 2.17 then f is concave and increasing on $(0, \pi)$ and concave and decreasing on $(-\pi, 0)$. Hence, $f'(0)$ does not exist. It then follows from Theorem 2.13 that $r_k = \mathcal{O}(k^{-\beta})$ for some $1 < \beta \leq 2$. The same conclusion is obtained for $\tilde{f}$ satisfying the assumptions of Theorem 2.18. Hence, if one wants to obtain a process with autocorrelations $\{r_k\}$ such that $r_k \leq 0$ for all $k \geq 1$ and $r_k = \mathcal{O}(k^{-\beta})$ for $\beta > 2$ then the generating procedures cannot be based on Theorem 2.19.

2.3.3 Processes with Given Autocorrelations

In this section we assume that a sequence of negative numbers $\{r_t\}_{t=0}^{\infty}$ satisfying (1.1) is given to be of some specified parametric form, for instance $r_t = -aq^t$, $t \geq 1$, for some $a > 0$ and $0 < q < 1$. We ask whether there exists a stationary process $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ with this autocorrelation function $\{r_t\}$, and we investigate how to generate such process. In particular, we investigate some models of negative exponentially decaying autocorrelations.

Example 2.22. Revisit the ARMA process $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ introduced in Example 1.3. We have shown that the autocorrelation function $\{r_t\}$ of $\mathbf{X}$ is given as

$$r_t = \frac{(b-a)(1-ab)}{1-2ab+b^2}(-a)^{t-1}, \quad t \geq 1. \tag{2.11}$$

Hence, $\{r_t\}$ follows a pattern of geometric decay of the form $r_t = \tilde{c}q^t$ for some $\tilde{c}, q \in \mathbb{R}$. Furthermore, under the conditions stated in Example 1.3 the autocorrelations satisfy the condition (1.1).

Now let $c > 0$ and $q \in (0,1)$, and consider a sequence $\{r_t\}$ such that $r_0 = 1$ and $r_t = -cq^t$ for $t \geq 1$, $r_t = r_{-t}$ for $t \leq -1$. Let us investigate under which conditions there exists a stationary ARMA$(1,1)$ with this autocorrelation function $\{r_t\}$ and how to generate such a process. We compare (2.11) with $-cq^t$. It follows that $q = -a$ and b is a solution of a quadratic equation $b^2q(c+1) + b(1+q^2+2cq^2) + cq + q = 0$. This equation has two distinct real roots $b_1 < b_2$ if and only if $q < 1/(2c+1)$. It is possible to show that $b_1 < -1 < b_2 < 0$. Thus, we choose $b = b_2$ in order to have the resulting ARMA process stationary and invertible. Furthermore, if $q = 1/(2c+1)$ then the equation has one double root $b_1 = b_2 = -1$. Choosing $b = -1$ results in a stationary non-invertible ARMA process.

In Example 2.22 it was easy to find an ARMA process with the specified autocorrelation function $r_t = -cq^t$ if the parameters c, q satisfy some additional conditions. In the following we provide a more general approach to this problem.

General Model of Weighted Exponentials

Let $I \geq 1$ and let $a_1, \dots, a_I > 0$ and $0 < q_1, \dots, q_I < 1$. Consider a sequence $\{r_t\}_{t=-\infty}^{\infty}$ of the form

$$r_0 = 1, \quad r_t = -\sum_{i=1}^{I} a_i q_i^t \quad \text{for } t \geq 1, \quad \text{and} \quad r_t = r_{-t} \quad \text{for } t \leq -1. \tag{2.12}$$

We are interested under which conditions the model (2.12) defines an autocorrelation function of a stationary process.

Theorem 2.23. *The model* (2.12) *defines an autocorrelation function of a stationary process* $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ *with* $\mathsf{var}\, X_t > 0$ *if and only if the condition*

$$\sum_{i=1}^{I} \frac{a_i q_i}{1 - q_i} \leq \frac{1}{2}. \tag{2.13}$$

is satisfied.

Proof. Recall that it is stated in Section 2.1 that a sequence of real numbers $\{r_k\}$ such that $r_0 = 1$, $r_{-k} = r_k$ for $k \in \mathbb{Z}$, and $r_k \leq 0$ for all $k \geq 1$ is the autocorrelation function of a centered stationary process $\{X_t, t \in \mathbb{Z}\}$ if and only if (2.1) holds. For the model (2.12) we have

$$\sum_{k=1}^{\infty} r_k = \sum_{t=1}^{\infty} \sum_{i}^{I} a_i q_i^t = -\sum_{i=1}^{I} a_i \frac{q_i}{1 - q_i}.$$

It is then easy to see that the condition (2.1) is equivalent to (2.13). □

Theorem 2.23 shows that the model (2.12) can be considered as a model for an autocorrelation function only under the condition (2.13). In this case (2.12) corresponds to a process with negatively correlated variables. Note that we have restricted ourselves only to the cases $q_i > 0$ for all $i = 1, \dots, I$. In addition, one can assume that $q_i \neq q_j$ for all $i \neq j$, $i, j = 1, \dots I$. If $q_i = q_j$ for some $i \neq j$ or $q_i = 0$ for some i then r_t can be written in the form (2.12) with a reduced number of parameters. The special case $q_1 = \dots = q_I = 0$ corresponds to the autocorrelation function of a white noise.

Let $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ be a centered process the autocorrelation function r_t given by (2.12) such that (2.13) holds. Suppose that $\mathsf{var}\, X_t = 1$. Then the spectral density of $\mathbf{X}$ is given as

$$f(\lambda) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} r_t \mathrm{e}^{-\mathrm{i}t\lambda} = \frac{1}{2\pi} \left[1 - \sum_{i=1}^{I} \frac{2a_i q_i (\cos\lambda - q_i)}{1 - 2q_i \cos\lambda + q_i^2} \right]. \tag{2.14}$$

This function can be rewritten as

$$f(\lambda) = \frac{1}{2\pi} \frac{R(\cos\lambda)}{\prod_{i=1}^{I} (1 - 2q_i \cos\lambda + q_i^2)}, \tag{2.15}$$

where R is a polynomial function of degree I.

Suppose that all the roots $y_1, \dots, y_I$ of the equation $R(y) = 0$ are real. For this case we can formulate the following statement.

Theorem 2.24. *Let the condition* (2.13) *be satisfied. Suppose that all the roots* $y_1, \dots, y_I$ *of the equation* $R(y) = 0$ *are real, and define* $\alpha_i = y_i - \sqrt{y_i^2 - 1}$ *if* $y_i > 0$ *and* $\alpha_i = y_i + \sqrt{y_i^2 - 1}$ *if* $y_i < 0$, $i = 1, \dots, I$. *Let* $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ *be a white noise with* $\mathsf{var}\,\varepsilon_t = \sigma^2 > 0$.

If $\mathbf{X}$ *follows a stationary ARMA*(I, I) *model*

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_I X_{t-I} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_I \varepsilon_{t-I}, \tag{2.16}$$

where

$$(1 - \phi_1 z - \phi_2 z \cdots - \phi_I z) = (1 - q_1 z) \dots (1 - q_I z),$$
$$(1 + \theta_1 z + \theta_2 z \cdots + \theta_I z) = (1 - \alpha_1 z) \dots (1 - \alpha_I z),$$

then the autocorrelation function $\{r_t\}$ *of* $\mathbf{X}$ *is given by* (2.12).

If the inequality (2.13) *is sharp then the ARMA process* $\mathbf{X}$ *is invertible.*

Proof. The spectral density f corresponding to the autocovariance function (2.12) is given in (2.14). Let us investigate whether there exists an ARMA process with the same spectral density f.

The condition (2.13) implies $f \geq 0$ on $(-\pi, \pi)$. It is not difficult to show that f has the global sharp minimum on $(-\pi, \pi)$ at the point 0, and thus we have $f(\lambda) > 0$ for all $(-\pi, 0) \cup (0, \pi)$. The denominator $\prod_{i=1}^{I} (1 - 2q_i \cos\lambda + q_i^2)$ in (2.15) is always positive because for any $i = 1, \dots, I$ we can write $1 - 2q_i \cos\lambda + q_i^2 \geq 1 - 2q_i + q_i^2 = (1 - q_i)^2 > 0$. Hence, the property $f(\lambda) > 0$ for all $(-\pi, 0) \cup (0, \pi)$ implies that $R(y) > 0$ for all $|y| < 1$. In particular, $|y_i| \geq 1$ for all $i = 1, \dots, I$ roots of $R(y) = 0$. Since all the roots $y_1, \dots, y_I$ of the equation $R(y) = 0$ are real, it follows from (2.14) after some algebraic manipulations that $R(y) = A(y_1 - y) \dots (y_I - y)$, where

$$A = 2^I \left(\prod_{i=1}^{I} q_i \right) \left[1 + \sum_{i=1}^{I} a_i \right], \quad A y_1 \dots y_I = \prod_{i=1}^{I} (1 + q_i^2) \left[1 + 2 \sum_{i=1}^{I} a_i \frac{q_i^2}{1 + q_i^2} \right]. \tag{2.17}$$

Hence,

$$f(\lambda) = \frac{1}{2\pi} \frac{A \prod_{i=1}^{I} (y_i - \cos\lambda)}{\prod_{i=1}^{I} (1 - 2q_i \cos\lambda + q_i^2)}.$$

The spectral density of an ARMA(p, q) process is shown in (A.5) in Lemma A.6. We can see that f is of the same form. Comparing f with (A.5) we immediately obtain that $p = q = I$ and $\beta_j = q_j$, $j = 1, \dots, I$. Furthermore, $y_i - \cos\lambda = 2\alpha_i[(1+\alpha_i^2)/(2\alpha_i) - \cos\lambda]$, and thus α_i is a solution of the equation $\alpha_i^2 - 2\alpha_i y_i + 1 = 0$. This equation has roots $\alpha_i^1 = y_i - \sqrt{y_i^2 - 1}$ and $\alpha_i^2 = y_i + \sqrt{y_i^2 - 1}$. We have shown that $|y_i| \geq 1$ for all $i = 1, \dots, I$. This implies that α_i^1, α_i^2 are always real, and they are distinct if and only if $|y_i| > 1$. If $y_i > 1$ then $y_i + \sqrt{y_i^2 - 1} > 1$ and $0 < y_i - \sqrt{y_i^2 - 1} < 1$. The converse holds for $y_i < -1$. Define $\alpha_i = y_i - \operatorname{sgn}(y_i)\sqrt{y_i^2 - 1}$ for $i = 1, \dots, I$, and set $\sigma_0^2 = A/[2^I \prod_{i=1}^I \alpha_i]$. Then $\sigma_0^2 = A \prod_{i=1}^I y_i (1 + \alpha_i)^2$, and it follows from (2.17) that $\sigma_0^2 > 0$. Hence, the spectral density f coincides with the spectral density of an ARMA process (2.16) with $\{\varepsilon_t, t \in \mathbb{Z}\}$ being $\mathsf{WN}(0, \sigma_0^2)$ (i.e. the choice $\sigma^2 = \sigma_0^2$ ensures $\mathsf{var}\, X_t = 1$). We see that if $\mathbf{X}$ follows (2.16) with a general $\sigma^2 > 0$ then the autocorrelations $\{r_t\}$ are given by (2.12).

It follows from the proof of Theorem 2.23 that the equality in (2.13) holds if and only if $\sum_{t=1}^\infty r_t = -1/2$. We have seen in Example 2.3 that for an ARMA process this condition is equivalent to the requirement that the MA part contains a unit root. This finishes the proof. □

Theorem 2.7 states that every process with negatively correlated variables can be expressed as a linear process. Hence, we are interested in the coefficients of this representation for the model (2.12).

Theorem 2.25. *Let y_i and α_i, $i = 1, \dots, I$, be defined as in Theorem 2.24. Define*

$$M_i = \frac{q_i - \alpha_i}{q_i} \prod_{j \neq i, j=1}^{I} \frac{q_i - \alpha_j}{q_i - q_j}, \quad i = 1, \dots, I, \qquad M_0 = 1 - \sum_{i=1}^{I} M_i, \tag{2.18}$$

and

$$c_0 = 1, \qquad c_k = \sum_{i=1}^{I} M_i q_i^k \quad \textit{for} \quad k \geq 1. \tag{2.19}$$

Let $\boldsymbol{\varepsilon} = \{\varepsilon_t, t \in \mathbb{Z}\}$ be a white noise with $\mathsf{var}\, \varepsilon_t = \sigma^2 > 0$. Then the process

$$X_t = \sum_{k=0}^{\infty} c_k \varepsilon_{t-k} \tag{2.20}$$

has the autocorrelation function given by (2.12).

Proof. Recall that f denotes the spectral density corresponding to the autocovariance function $\{r_t\}$. The spectral density of $\mathbf{X}$ is given as $\tilde{f}(\lambda) = [\sigma^2/(2\pi)] |\sum_{k=0}^\infty c_k \mathrm{e}^{-\mathrm{i}k\lambda}|^2$. Let us find $\{c_k\}_0^\infty$ such that $f \equiv \tilde{f}$ holds. It is shown in the proof of Theorem 2.24 that f can be rewritten as

$$f(\lambda) = K \cdot \prod_{i=1}^{I} \frac{|1 - \alpha_i \mathrm{e}^{-\mathrm{i}\lambda}|^2}{|1 - q_i \mathrm{e}^{-\mathrm{i}\lambda}|^2}, \tag{2.21}$$

where $K = A \left[2\pi \prod_{i=1}^I (2\alpha_i) \right]^{-1}$. Remind that we assume $q_i \neq q_j$ for all $i \neq j$. The partial fraction expansion gives

$$\prod_{i=1}^{I} \frac{1 - \alpha_i \mathrm{e}^{-\mathrm{i}\lambda}}{1 - q_i \mathrm{e}^{-\mathrm{i}\lambda}} = M_0 + M_1 \frac{1}{1 - q_1 \mathrm{e}^{-\mathrm{i}\lambda}} + \cdots + M_I \frac{1}{1 - q_I \mathrm{e}^{-\mathrm{i}\lambda}},$$

where $M_0, \ldots, M_I$ are given by (2.18). The coefficients $\{c_k\}_0^\infty$ can be obtained after expressing (2.21) as I geometrical series and comparing this expression with $\tilde{f}$. We get

$$f(\lambda) = K \cdot \left| M_0 + \sum_{i=1}^{I} M_i \sum_{k=0}^{\infty} q_i^k \mathrm{e}^{-\mathrm{i}k\lambda} \right|^2 = K \cdot \left| 1 + \sum_{k=1}^{\infty} \left(\sum_{i=1}^{I} M_i q_i^k \right) \mathrm{e}^{-\mathrm{i}k\lambda} \right|^2.$$

Hence, $\{c_k\}_0^\infty$ must satisfy (2.19) and $\sigma^2 = A \prod_{i=1}^{I} (2\alpha_i)^{-1} > 0$. This choice of σ^2 corresponds to the process with the autocorrelations (2.12) and $\mathsf{var}\, X_t = 1$. For general $\sigma^2 > 0$ we get a process with the autocorrelations (2.12) and $\mathsf{var}\, X_t > 0$. □

Remark 2.26. (i) Notice that the coefficients $\{c_k\}$ from the representation (2.20) are of a similar form (weighted exponentials) as the correlations r_k in (2.12). This should not be surprising. It is well-known that for a general ARMA(p, q) process the coefficients $\{c_k\}_{k>q}$ from the representation (2.20) are a solution of the same pth order difference equation as $\{r_k\}_{k>q}$, and the two sequences $\{c_k\}$ and $\{r_k\}$ differ only in the inial conditions, see Hamilton (1994). For instance for the ARMA$(1, 1)$ process from Example 2.22 we have $c_0 = 1$, $c_k = (b - a)(-a)^{k-1}$, $k \geq 1$.

(ii) Let $0 < q_1, \ldots, q_I < 1$, and $M_1, \ldots, M_I \in \mathbb{R}$. Define the coefficients $c_0 = 1$ and $c_k = \sum_{i=1}^{I} M_i q_i^k$ for $k \geq 1$. Let $\boldsymbol{\varepsilon} = \{\varepsilon_t,\, t \in \mathbb{Z}\}$ be a white noise with $\mathsf{var}\, \varepsilon_t = \sigma^2 > 0$, and consider the linear process $\mathbf{X}$ given by (2.20). It can be easily shown from its spectral density that $\mathbf{X}$ follows an ARMA(I, I) process. The autocovariances $\{R_t\}$ of the process $\mathbf{X}$ are given as

$$R_0 = \sigma^2 \sum_{k=0}^{\infty} c_k^2 = \sigma^2 \left[1 + \sum_{i=1}^{I} \sum_{j=1}^{I} M_i M_j \frac{q_i q_j}{1 - q_i q_j} \right],$$

$$R_t = \sigma^2 \sum_{k=0}^{\infty} c_k c_{t+k} = \sigma^2 \sum_{i=1}^{I} M_i \left[1 + \sum_{j=1}^{I} M_j \frac{q_i q_j}{1 - q_i q_j} \right] q_i^t, \quad t \geq 1.$$

The autocorrelations $r_t = R_t / R_0$ are of the form (2.12) for

$$a_i = -M_i \left[1 + \sum_{j=1}^{I} M_j \frac{q_i q_j}{1 - q_i q_j} \right].$$

If the choice of the parameters $M_1, \ldots, M_I$ and $q_1, \ldots, q_I$ is such that $a_i > 0$ for all $i = 1, \ldots, I$ then all the autocorrelations r_t are negative, and the condition (1.1) is satisfied.

Example 2.27. (Geometric Sequence)

Consider the model (2.12) with $I = 1$, $a > 0$, and $q \in (0, 1)$, i.e. assume that $r_t = -aq^t$ for $t \geq 1$. This model has been already discussed in Example 2.22. Let us apply Theorem 2.24 to this situation. The condition (2.13) is equivalent to the inequality $0 < q \leq 1/(1+2a)$. The spectral density f is given as

$$f(\lambda) = \frac{1}{2\pi} \frac{1 + (1+2a)q^2 - 2q(a+1)\cos\lambda}{1 + q^2 - 2q\cos\lambda} = \frac{1}{2\pi} \frac{R(\cos\lambda)}{1 + q^2 - 2q\cos\lambda},$$

where R is a linear function with the root $y_1 = (1 + q^2 + 2aq^2)/[2q(a+1)]$. Hence, the desired process $\mathbf{X}$ follows a model

$$X_t = q\,X_{t-1} + \varepsilon_t - \alpha\varepsilon_{t-1},$$

where $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ is a white noise $\mathsf{WN}(0, \sigma^2)$ and

$$\alpha = \frac{1 + q^2 + 2aq^2 - \sqrt{(1-q^2)(1-(q+2aq)^2)}}{2q(1+a)}.$$

The coefficients $\{c_k\}$ from the representation (2.20) are given as $c_0 = 1$ and

$$c_k = \frac{\sqrt{1-q^2}\sqrt{1-(1+2a)^2q^2} - (1-q^2)}{2(1+a)q^2}\, q^k \quad \text{for } k \geq 1. \tag{2.22}$$

It can be shown that c_k are negative for all $k \geq 1$. Hence, one can interpret the variable X_t as the white noise ε_t minus a linear combination of the past noises ε_{t-k}, each one standing with a positive coefficient.

Theorem 2.1 states that the equality $\sum_{t=1}^{\infty} r_t = -1/2$ in (2.1) is reached if and only if $f(0) = 0$. This is equivalent to $q = 1/(2a+1)$. In view of Theorem 2.24 the process $\mathbf{X}$ is not invertible in this case.

As an illustration, for the special case $a = 1$ and $q = 1/3$ we get $c_0 = 1$ and $c_k = -2/3^k$ for $k \geq 1$. Let $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ be a white noise $\mathsf{WN}(0, \sigma^2)$. Then the linear process

$$X_t = \varepsilon_t - 2\sum_{k=1}^{\infty} \frac{1}{3^k}\,\varepsilon_{t-k}$$

has the autocorrelation function $r_0 = 1$ and $r_t = -1/3^t$ for $t \geq 1$. In this case we get $\sum_{t=1}^{\infty} r_t = -1/2$.

Example 2.28. (Sum of Two Geometric Sequences)

Consider the model (2.12) with $I = 2$, $a_1 > 0, a_2 > 0$, and $0 < q_1, q_2 < 1$, i.e. assume that $r_t = -(a_1 q_1^t + a_2 q_2^t)$, for $t \geq 1$. The condition (2.13) can be rewritten as

$$0 < q_1 < \frac{1}{1+2a_1}, \quad 0 < q_2 \leq \frac{(1+2a_1)q_1 - 1}{(1+2a_1+2a_2)q_1 - (2a_2+1)}. \tag{2.23}$$

The set of all (q_1, q_2) satisfying (2.23) is plotted in Figure 2.3. The area is bordered by axes x and y and by the hyperbola $y = [(1+2a_1)x - 1]/[(1+2a_1+2a_2)x - (2a_2+1)]$.

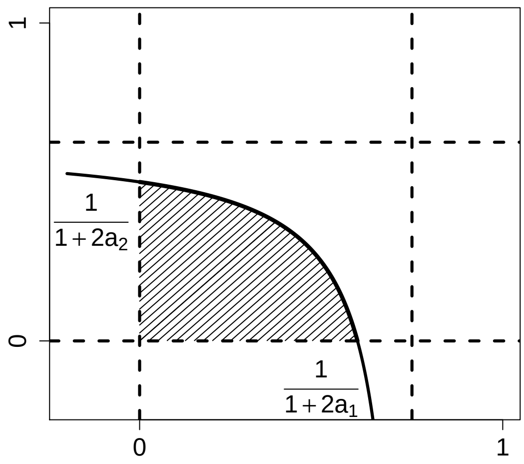


Figure 2.3: Set of (q_1, q_2) satisfying (2.23)

This hyperbola crosses the axis x at $1/(2a_1+1)$ and the axis y at $1/(2a_2+1)$, and it has asymptotes $x = (1+a_2)/(1+a_1+a_2)$ and $y = (1+a_2)/(1+a_1+a_2)$.

The spectral density f corresponding to $\{r_t\}$ is given as

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi}\left[1 - \frac{2a_1q_1(\cos\lambda - q_1)}{1+q_1^2-2q_1\cos\lambda} - \frac{2a_2q_2(\cos\lambda - q_2)}{1+q_2^2-2q_2\cos\lambda}\right] \\ &= \frac{1}{2\pi}\frac{A\cos^2\lambda - B\cos\lambda + C}{(1+q_1^2-2q_1\cos\lambda)(1+q_2^2-2q_2\cos\lambda)}, \end{aligned}$$

where

$$\begin{aligned} A &= 4q_1q_2 + 4a_1q_1q_2 + 4a_2q_1q_2, \\ B &= 2(q_1 + a_1q_1 + q_2 + a_2q_2 + q_1^2q_2 + 2a_1q_1^2q_2 + a_2q_1^2q_2 + q_1q_2^2 + a_1q_1q_2^2 + 2a_2q_1q_2^2), \\ C &= 1 + q_1^2 + 2a_1q_1^2 + q_2^2 + 2a_2q_2^2 + q_1^2q_2^2 + 2a_1q_1^2q_2^2 + 2a_2q_1^2q_2^2. \end{aligned}$$

It follows after some computation that the two roots y_1, y_2 of the equation $Ay^2 - By + C = 0$ are always real and greater than 1. Hence, $\{r_t\}$ is the autocorrelation function of a stationary ARMA$(2,2)$ process $\mathbf{X}$ with the parameters specified in Theorem 2.24. The coefficients c_k, $k \geq 0$, from the MA(∞) representation (2.20) are given as

$$c_k = \frac{\alpha_1\alpha_2 + q_1^2 - \alpha_1q_1 - \alpha_2q_1}{q_1(q_1-q_2)}\, q_1^k + \frac{\alpha_1\alpha_2 + q_2^2 - \alpha_1q_2 - \alpha_2q_2}{q_2(q_2-q_1)}\, q_2^k \quad \text{for} \quad k \geq 1.$$

Finally, it can be easily shown that $f(0) = 0$ if and only if

$$q_2 = \frac{(1+2a_1)q_1 - 1}{(1+2a_1+2a_2)q_1 - (2a_2+1)}. \tag{2.24}$$

Hence, the process $\mathbf{X}$ is invertible if and only if (2.24) does not hold.

For a numerical illustration consider the case $a_1 = a_2 = 1$, $q_1 = 1/4$, and $q_2 = 1/7$. Then q_1, q_2 satisfy (2.23) and (2.24), and thus $\sum_{t=1}^{\infty} r_t = -1/2$. We get

$$\alpha_1 = 1, \quad \alpha_2 = \frac{41 - 3\sqrt{165}}{14},$$
$$M_1 = 75 - 6\sqrt{165} \doteq -2.071, \quad M_2 = 12(\sqrt{165} - 13) \doteq -1.857.$$

Let $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ be a white noise $\mathsf{WN}(0, \sigma^2)$. The process

$$X_t = \Big(\frac{1}{4} + \frac{1}{7}\Big)X_{t-1} - \frac{1}{28}X_{t-2} + \varepsilon_t - (1 + \alpha_2)\varepsilon_{t-1} + \alpha_2\varepsilon_{t-2}$$

has the autocorrelation function $r_t = -1/4^t - 1/7^t$ for $t \geq 1$. The process $\mathbf{X}$ can be also represented as (2.20) with

$$c_0 = 1, \quad c_k = \frac{75 - 6\sqrt{165}}{4^k} + \frac{12(\sqrt{165} - 13)}{7^k} \quad \text{for } k \geq 1.$$

Note that $c_k < 0$ for all $k \geq 1$.

Remark 2.29. Results for some considered special cases of the general model (2.12) indicate that the coefficients c_k, $k \geq 1$, from (2.19) might be always negative. However, this stays an open problem for the general situation.

Arithmetic–Geometric Sequence

Let $0 < q < 1$ and $a > 0$. Consider a sequence $\{r_t\}_{t=-\infty}^{\infty}$ of the form

$$r_0 = 1, \quad r_t = -atq^{t+1} \quad \text{for } t \geq 1. \tag{2.25}$$

We are interested under which conditions the model (2.25) defines an autocorrelation function of a stationary process.

Theorem 2.30. *The model* (2.25) *defines an autocorrelation function of a stationary process* $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ *if and only if the condition*

$$0 < a \leq \frac{(1-q)^2}{2q^2} \tag{2.26}$$

is satisfied.

The proof of the foregoing statement is analogous to the proof of Theorem 2.5. The condition (2.26) can be further rewritten as

$$q \leq \frac{1}{1 + \sqrt{2a}} = \frac{1 - \sqrt{2a}}{1 - 2a}.$$

Theorem 2.31. *Let the condition* (2.26) *be satisfied. Define*

$$y_{1,2} = \frac{1}{4q^2}\left[q(2+aq)(1+q^2) \pm \sqrt{aq^3[4(q^2-1)^2 + aq(1+q^2)^2]}\right] \tag{2.27}$$

and $\alpha_i = y_i - \sqrt{y_i^2 - 1}$ *for* $i = 1,2$. *Let* $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ *be a white noise with* $\mathsf{var}\,\varepsilon_t > 0$. *Then a stationary ARMA*$(2,2)$ *process* $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ *defined as*

$$X_t = 2qX_{t-1} - q^2 X_{t-2} + \varepsilon_t - (\alpha_1 + \alpha_2)\varepsilon_{t-1} + \alpha_1\alpha_2\varepsilon_{t-2}$$

has the autocorrelation function (2.25).

Proof. The spectral density corresponding to the autocovariance function (2.25) is given as

$$f(\lambda) = \frac{1}{2\pi}\frac{1 + 2q^2 + 4aq^3 + q^4 - 2q(2+aq)(1+q^2)\cos\lambda + 4q^2\cos^2\lambda}{(1 - 2q\cos\lambda + q^2)^2}. \tag{2.28}$$

The numerator in (2.28) can be rewritten as $A\cos^2\lambda - B\cos\lambda + C$, where $A = 4q^2$, $B = 2q(2+aq)(1+q^2)$, $C = 1 + 2q^2 + 4aq^3 + q^4$. The equation $Ay^2 - By + C = 0$ has two real roots y_1, y_2 given in (2.27). It is possible to show that both y_1, y_2 are real and greater than 1. The rest of the proof is analogous to that of Theorem 2.24. □

Theorem 2.32. *Let* α_1, α_2 *be defined as in Theorem 2.31. Let* $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ *be a white noise with* $\mathsf{var}\,\varepsilon_t > 0$. *Define*

$$M_0 = \frac{\alpha_1\alpha_2}{q^2}, \quad M_1 = \frac{\alpha_1 + \alpha_2}{q} - M_0, \quad M_2 = 1 - M_0 - M_1$$

and

$$c_0 = 1, \quad c_k = \big[(M_1 + M_2) + M_2 k\big]q^k, \quad k \geq 1. \tag{2.29}$$

Then the linear process $X_t = \sum_{k=0}^{\infty} c_k\varepsilon_{t-k}$ *has the autocorrelation function* (2.25).

Proof. We proceed in a similar way as in the proof of Theorem 2.25. It follows from (2.28) that

$$f(\lambda) = \frac{A}{8\pi\alpha_1\alpha_2}\left|\frac{(1-\alpha_1 e^{-i\lambda})(1-\alpha_2 e^{-i\lambda})}{(1-qe^{-i\lambda})^2}\right|^2.$$

Partial fraction expansion gives

$$\frac{(1-\alpha_1 e^{-i\lambda})(1-\alpha_2 e^{-i\lambda})}{(1-qe^{-i\lambda})^2} = M_0 + M_1\frac{1}{1-qe^{-i\lambda}} + M_2\frac{1}{(1-qe^{-i\lambda})^2},$$

where M_0, M_1, M_2 are given above. We get

$$\begin{aligned} f(\lambda) &= \frac{A}{8\pi\alpha_1\alpha_2}\left|M_0 + \frac{M_1}{1-qe^{-i\lambda}} + \frac{M_2}{(1-qe^{-i\lambda})^2}\right|^2 \\ &= \frac{A}{8\pi\alpha_1\alpha_2}\left|1 + \sum_{k=1}^{\infty}\big[(M_1+M_2) + M_2 k\big]q^k e^{-ik\lambda}\right|^2. \end{aligned}$$

The coefficients $\{c_k\}$ satisfy $f(\lambda) = [\sigma^2/(2\pi)]|\sum_{k=0}^{\infty} c_k \mathrm{e}^{-\mathrm{i}\lambda}|^2$ for $\sigma^2 = A(4\alpha_1\alpha_2)^{-1} > 0$. This choice of σ^2 corresponds to the process $\mathbf{X}$ such that $\mathsf{var}\, X_t = 1$. A different value of σ^2 leads to a general process $\mathbf{X}$ with the autocorrelations (2.25) and $\mathsf{var}\, X_t > 0$. □

Example 2.33. Consider the model (2.25) with $a = 2$ and $q = 1/3$. We obtain $y_1 = 1$, $y_2 = 31/9$, $\alpha_1 = 1$, $\alpha_2 = (31 - 4\sqrt{55}\,)/9$. It follows from Theorem 2.31 that a stationary ARMA(2,2) process $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ defined as

$$X_t = \frac{2}{3} X_{t-1} - \frac{1}{9} X_{t-2} + \varepsilon_t - \frac{40 - 4\sqrt{55}}{9}\,\varepsilon_{t-1} + \frac{31 - 4\sqrt{55}}{9}\,\varepsilon_{t-2}$$

has the autocorrelation function $r_t = -(2t)/(3^{t+1})$.

In Theorem 2.32 we have

$$M_1 = \frac{2}{3}\,(-73 + 10\sqrt{55}\,) \doteq 0.77, \quad M_2 = -\frac{8}{3}\,(-7 + \sqrt{55}\,) \doteq -1.11,$$
$$c_0 = 1, \quad c_k = -\frac{2}{3^{k+1}}\left(45 - 6\sqrt{55} - 4k(7 + \sqrt{55})\right), \quad k \ge 1.$$

Let $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ be a white noise with $\mathsf{var}\,\varepsilon_t = \sigma^2 > 0$. Then the linear process $X_t = \sum_{k=0}^{\infty} c_k \varepsilon_{t-k}$ has the autocorrelation function $r_t = -(2t)/(3^{t+1})$. Let us point out that all the coefficients c_k, $k \ge 1$, are all negative.

The models (2.12) and (2.25) can be further generalized in various ways. For instance, let $a > 0$, $q > 0$, and $x \in \mathbb{R}$, and define

(a) $r_t = -at^k q^{t+1}$,
(b) $r_t = -a[1 - \cos(tx)]q^t$,
(c) $r_t = -a[1 - \cos(tx)]tq^t$.

All the models (a)—(c) correspond to autocorrelation functions of some ARMA models (cf. Section 1.3.2). Hence, results analogous to those for models (2.12) and (2.25) can be obtained.

Investigation of models that do not lead to ARMA processes is more complicated. For an example of such a model see Example 2.8.

Chapter 3

Negatively Correlated Bernoulli Variables

Dependent 0-1 valued variables (so called Bernoulli variables) have attained a lot of consideration in the statistical literature. Remind that a random variable ξ is called *Bernoulli variable* if $\xi = 1$ with probability $p \in (0, 1)$ and $\xi = 0$ with probability $q = 1 - p$. The problem of simulating a binary n-dimensional vector with a specified vector of success probabilities and a given correlation matrix has been studied by many authors, see for instance Lee (1993) or Quaqish (2003). The compatible range of the correlation matrix as a function of success probabilities for $n \geq 3$ is investigated in Chaganty and Joe (2006). The obtained restrictions for $n \geq 3$ are quite complex, and it is therefore natural that the autocorrelation function of an infinite sequence of binary variables has even more complicated structure, see Masry (1972) for the stationary case. Hence, generating a sequence of binary variables with specified probabilities and correlations is extremely difficult in general, and some special models are needed.

In this chapter we study stationary sequences of Bernoulli variables with the autocorrelation structure (1.1). Such processes have some special practical applications mentioned in Meister and Bondesson (2001), see our Example 3.1. In particular, it is sometimes desirable to have a model for generating Bernoulli variables which are "negatively correlated as much as possible".

Example 3.1. Consider a problem of real time sampling from a passing population of units. Let ξ_t denote the indicator whether the tth unit is chosen to be inspected. In so called *Bernoulli sampling* the population units are selected independently of each other. This means that $\{\xi_t,\, t \in \mathbb{Z}\}$ is a sequence of iid variables. However, sometimes the neighbouring units may have similar properties, and one would like to avoid sampling units close to each other too often. Thus, sampling according to a stationary process of negatively correlated variables $\{\xi_t,\, t \in \mathbb{Z}\}$ might be more efficient. Indeed, Meister and Bondesson (2001) describe situations in which some gain is achieved in the efficiency comparing this approach with the classical Bernoulli sampling.

A common model for generating dependent 0-1 valued variables is based on the operation called *clipping* (also *hard limiting* or *hard clipping*). Variables $\{\xi_t,\, t \in \mathbb{Z}\}$ are obtained by clipping a random sequence $\mathbf{Z} = \{Z_t,\, t \in \mathbb{Z}\}$ at the level $c \in \mathbb{R}$ if

$$\xi_t = \mathrm{I}[Z_t \geq c] = \begin{cases} 1 & \text{if } Z_t \geq c, \\ 0 & \text{if } Z_t < c, \end{cases} \tag{3.1}$$

where $\mathrm{I}[\cdot]$ stands for the indicator function, see Kedem (1980a), cf. our Section 1.3.1. This means that the variable ξ_t is an indicator whether the process $\mathbf{Z}$ crosses the level c at time t. Such discretization by a threshold of a continuous-valued process is a common phenomenon in biology, engineering and other areas. Properties of such *clipped processes* are studied extensively in Kedem (1980a).

For the case of 1-dependent variables $\{\xi_t,\, t \in \mathbb{Z}\}$ with a negative 1-lag autocorrelation r_1 Bondesson (2003) proposed a model with the clipped process $\mathbf{Z}$ formed by differences of some iid variables X_n. In such a case the autocorrelation function depends solely on the distribution of X_n and on the clipping parameter c. Properties of this model are investigated in Section 3.1.1. In Section 3.1.2 we slightly generalize the model proposed by Bondesson (2003), and derive an explicit formula for r_1 in some special cases. The influence of the clipping parameters is analyzed and discussed. Estimators of some parameters in this model are constructed and investigated in Section 3.2. Some special models of general dependent Bernoulli variables with the autocorrelation structure (1.1) are treated in Section 3.4.

Remark 3.2. Consider the clipping model (3.1). If $\mathbf{Z}$ is a strictly stationary process then $\boldsymbol{\xi} = \{\xi_t,\, t \in \mathbb{Z}\}$ is stationary as well, and $\boldsymbol{\xi}$ inherits a certain structure from the underlying process $\mathbf{Z}$. In particular, the success probability $p = \mathsf{P}(\xi_t = 1)$ and the autocorrelations $r_t = \mathsf{cor}\,(\xi_k, \xi_{k+t})$ are uniquely determined by the choice of $\{Z_t,\, t \in \mathbb{Z}\}$ and the clipping parameter $c \in \mathbb{R}$. However, the correlation r_t depends not only on the autocorrelation function of the original process $\mathbf{Z}$, but on further properties of the corresponding multivariate probability distributions as well.

Properties of clipped processes introduced above have been intensively studied in the literature. Several authors have been concerned with the problem of estimation of the autocorrelations, spectral density, or some other parameters of the original process $\mathbf{Z}$ using the information about $\{\xi_t,\, t \in \mathbb{Z}\}$ only. Kedem (1980b) derives estimates of parameters of an AR(p) process based on the clipped data. Estimation of the Hurst exponent, see Section 1.3.3, based on the clipped data is described in Coeurjolly (2000). For other references see Lomnicki and Zaremba (1955), Kedem (1980a), Hinich (1967) and Keenan (1982).

However, all these works give a solution only for the special case where the process $\mathbf{Z}$ is Gaussian and $c = 0$. In fact, if $\mathbf{Z}$ is centered Gaussian then the covariances completely determine the multivariate distributions, and there is a closed form connection between the autocorrelations of $\mathbf{Z}$ and the clipped process $\{\xi_t,\, t \in \mathbb{Z}\}$, see Lemma A.7 in Appendix. This explains why the Gaussian case attracts so much attention.

Remark 3.3. Someone might also suggest to model dependent Bernoulli variables $\{\xi_t,\, t \in \mathbb{Z}\}$ as a stationary Markov chain. However, this approach can never result in a process with negatively correlated variables.

Let $\{\xi_t,\, t \in \mathbb{Z}\}$ be a Markov chain with states 0 and 1 and the matrix of transition probabilities

$$\mathbf{P} = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix},$$

where $a, b \in (0,1)$. Let the initial distribution be the stationary one $\boldsymbol{\pi} = 1/(a+b)\big(b, a\big)^\top$. Then the process $\boldsymbol{\xi}$ is strictly stationary with $\mathsf{E}\,\xi_t = a/(a+b)$ and $\mathsf{var}\,\xi_t = ab/(a+b)^2$. It is derived in Feller (1968) that for $n > 1$

$$\mathsf{E}\,\xi_0\xi_n = \frac{a + b(1-a-b)^n}{a+b}\frac{a}{a+b}$$

holds. The autocovariance $R_n = \mathsf{cov}\,(\xi_k, \xi_{n+k}) = \mathsf{cov}\,(\xi_0, \xi_n)$ is given as

$$R_n = \mathsf{E}\,\xi_0\xi_n - p^2 = \frac{a + b(1-a-b)^n}{a+b}\frac{a}{a+b} - \frac{a^2}{(a+b)^2} = \frac{ab(1-a-b)^n}{(a+b)^2}.$$

Hence, the autocorrelations are $r_n = (1-a-b)^n$, and the sequence $\{r_t\}$ cannot satisfy the condition (1.1).

3.1 1-Dependent Variables

The following model (3.2) for 1-dependent variables with a negative 1-lag autocorrelation is introduced in Bondesson (2003). Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid random variables and $c \in \mathbb{R}$. Define

$$\xi_t = \begin{cases} 1 & \text{if } X_t - X_{t-1} < c, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

The model (3.2) transforms an MA(1) process $Z_t = X_t - X_{t-1}$ to a sequence of 0-1 valued variables $\{\xi_t,\, t \in \mathbb{Z}\}$. These variables $\{\xi_t,\, t \in \mathbb{Z}\}$ are obviously 1-dependent, and the process $\boldsymbol{\xi}$ is (strictly) stationary. The process $\mathbf{Z}$ consists of negatively correlated variables and attains the minimal bound $-1/2$ for the 1-lag autocorrelation, see Example 2.4. One would then expect that the corresponding Bernoulli process $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ has a minimal possible autocorrelation r_1 (in some sense) as well. This motivates us to study the model (3.2) in more detail. We show in Corollary 3.15 that r_1 is always negative in the model (3.2).

Remark 3.4. The Bernoulli variables $\{\xi_t,\, t \in \mathbb{Z}\}$ defined in the model (3.2) are a special case of 0-1 valued *two-block-factors*. Recall that a two-block-factor is defined as

$f(X_n, X_{n+1})$ for some iid sequence $\{X_n\}$ and some function f. For this class of processes de Valk (1988) proved that the 1-lag autocorrelation r_1 always satisfies $r_1 \geq -1/3$.

Let $\min[r_1(p)]$ denote the minimal possible 1-lag autocorrelation for a given $p \in (0,1)$. For $p \in (0, 1/2)$ define $m = \left\lfloor \frac{1}{1-2p} \right\rfloor$ and $\delta = \sqrt{1 - 2[p(m+1)/m]}$, where $\lfloor x \rfloor$ stands for the integer part of x. It is derived in de Valk (1988) that in the class of all 0-1 valued two-block-factors we have

$$\min[r_1(p)] = \begin{cases} \dfrac{m(m-1)(1-2\delta)(1+\delta)^2}{6(m+1)^2 p(1-p)} - \dfrac{p}{1-p}, & 0 < p < \frac{1}{2}, \\ -\frac{1}{3}, & p = \frac{1}{2}, \end{cases} \tag{3.3}$$

and $\min[r_1(p)] = \min[r_1(1-p)]$ for $p \in (1/2, 1)$ due to the symmetry. In particular, if $p < 1/4$ then $m = 1$ and $\min[r_1(p)] = -p/(1-p)$, and by symmetry if $p > 3/4$ then $\min[r_1(p)] = -(1-p)/p$, cf. Figure 3.5 or Anděl and Došlá (2010).

The minimal bound (3.3) is obtained in de Valk (1988) as the infimum over the class of all indicator processes $\{Y_t, t \in \mathbb{Z}\}$ such that $Y_t = \mathrm{I}[(U_t, U_{t+1}) \in A]$, where $\{U_t,\, t \in \mathbb{Z}\}$ are iid variables uniformly distributed over the unit square, and A is a measurable subset of the unit square. Hence, for each $p \in (0,1)$ a different model is used to obtain the minimum bound $\min[r_1(p)]$. It follows from (3.3) that

$$r_1 \geq -1/3 \tag{3.4}$$

holds for all $p \in (0,1)$. In addition, $\min[r_1(p)] = -1/3$ if and only if $1/(1-2p)$ is an integer or $p = 1/2$.

"Optimal" processes which attain the bound $\min[r_1(p)]$ are described in de Valk (1988) as well, cf. Bondesson (2003). For $p < 1/2$ this minimum is reached by an indicator process corresponding to the set $A_p^{min} = \{(x,y) \in [0,1]^2 : \; y \leq a(x)\}$, where $a(x)$ is a step-function with m step-points on the diagonal $y = x$ and with all steps of an equal height. If $p = 1/2$ then the optimal set $A_{1/2}^{min}$ is the lower right quadrant of the unit square. Note also that Joe (1997) presents a simple model of 1-dependent Bernoulli variables for which the boundary (3.3) is reached whenever $p \leq 1/4$ or $p \geq 3/4$, cf. Došlá (2008).

Notice that (3.4) "improves" the general inequality (2.1). Gandolfi et al. (1989) conjectured that $r_1 \geq -1/3$ holds for general Bernoulli 1-dependent processes as well, but this question remains open. (It is shown in Aaronson et al. (1989), cf. Matúš (1996), that there exist 0-1 valued 1-dependent processes which are not two-block-factors.)

Remark 3.5. The model (3.2) is obviously closely related to the clipped process introduced in the beginning of this chapter. Indeed, if the iid variables $\{X_t,\, t \in \mathbb{Z}\}$ in (3.2) are continuously distributed then the process $\boldsymbol{\xi}$ defined by (3.2) is equal to the process $X_{t-1} - X_t$ clipped at the level $-c$. Equivalently, $\xi_t = 1 - \tilde{\xi}_t$ where $\tilde{\boldsymbol{\xi}}$ is the process $X_t - X_{t-1}$ clipped at the level c. It is clear that the autocorrelation function $\{r_t\}$ of the process $\boldsymbol{\xi}$ coincides with the autocorrelation function of the process $\tilde{\boldsymbol{\xi}}$, and the success probabilities are related by $p = 1 - \tilde{p}$, where $\tilde{p} = \mathsf{P}(\tilde{\xi}_t = 1)$.

3.1.1 Properties of the Bondesson's Model

Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid variables with an absolutely continuous distribution with a distribution function F. Let $\boldsymbol{\xi}$ be defined in the model (3.2) for some $c \in \mathbb{R}$. Then the success probability $p = \mathsf{P}(\xi_t = 1)$ and the 1 lag autocorrelation $r_1 = \operatorname{cor}(\xi_t, \xi_{t-1})$ are given as

$$p = \mathsf{P}(\xi_t = 1) = \mathsf{E}F(X + c), \tag{3.5}$$

$$r_1 = \operatorname{cor}(\xi_t, \xi_{t-1}) = 1 - \frac{\mathsf{E}F(X+c)F(X-c)}{\mathsf{E}F(X+c)\mathsf{E}F(X-c)} = 1 - \frac{Q}{p(1-p)}, \tag{3.6}$$

where $Q = \mathsf{E}F(X+c)F(X-c)$, see Lemma A.8 in Appendix. Hence, p and r_1 are completely determined by the parameter c and the choice of the distribution F.

Example 3.6. If $c = 0$ then the formulas (3.5) and (3.6) simplify to $p = 1/2$ and $r_1 = -1/3$. In view of Remark 3.4 the model (3.2) with $c = 0$ (i.e. $p = 1/2$) corresponds to the minimal possible value for r_1.

In the following we express the success probability p and the correlation r_1 as functions of the parameter $c \in \mathbb{R}$ for different distributions of X_t. All the calculations are based on the formulas (3.5) and (3.6), and therefore we present only the final expressions and omit the computational details.

It is clear that we can restrict ourselves only to the case $c \geq 0$ (corresponding to $p \in [1/2, 1]$) because the situation for $c < 0$ is symmetric. Indeed, if $c < 0$ then $\xi_t = \mathrm{I}[X_{t-1} - X_t > -c]$, and the rest follows from Remark 3.5.

Remark 3.7. Bondesson (2003) considered the model (3.2) with discrete random variables $\{X_t,\, t \in \mathbb{Z}\}$ as well. He shows that if $c = 0$ and X_t is a taking N different values then $r_1 = -1/3$ is reached if and only if X_t has a uniform distribution on $\{1, 2, \ldots, N\}$ and $p = (N-1)/(2N)$. This corresponds to the conclusion of de Valk (1988) mentioned in Remark 3.4.

Proposition 3.8. *Let X_t have the uniform distribution on the interval $[0,1]$. Consider the model (3.2) with $c \in [0,1]$. Then*

$$p(c) = 1 - \frac{(1-c)^2}{2}, \quad c \in [0,1], \tag{3.7}$$

$$r_1(c) = \begin{cases} -\dfrac{1 - 6c^2 + 4c^3 + 3c^4}{3(1-c)^2(1+2c-c^2)} & \text{if } \ 0 \leq c < \frac{1}{2}, \\[2ex] -\dfrac{(1-c)^2}{1+2c-c^2} & \text{if } \ \frac{1}{2} \leq c < 1. \end{cases} \tag{3.8}$$

It is clear that $p = 1$ for $c \geq 1$ and $p(c) = 0$ for $c \leq -1$. This means that $r_1(c)$ is not defined for $|c| \geq 1$, and it suffices to consider $c \in [0, 1)$ in Proposition 3.8.

It follows from (3.8) that $r_1(c) \geq r_1(0)$ holds in Proposition (3.8). Hence, the function $r_1(c)$ reaches its minimum $-1/3$ at the point $c = 0$. This corresponds to the success probability $p = 1/2$.

The formula (3.7) implies that $c = 1 - \sqrt{2(1-p)}$, and it is further possible to derive a formula describing the dependence of the autocorrelation r_1 directly on the success probability $p \in (0, 1)$. We get

$$r_1(p) = \begin{cases} -\dfrac{31 - 22\sqrt{2-2p} + 4(-9 + 4\sqrt{2-2p})p + 6p^2}{6(1-p)p}, & \dfrac{1}{2} \leq p < \frac{7}{8}, \\ 1 - \dfrac{1}{p}, & \frac{7}{8} \leq p < 1. \end{cases} \tag{3.9}$$

It follows from Remark 3.4 that the model (3.2) with X_t uniformly distributed on $[0, 1]$ leads to the minimal possible 1-lag autocorrelation r_1 for all $p \geq 7/8$ (and symmetrically for all $p \leq 1/8$), cf. Theorem 3.12.

Proposition 3.9. *Let X_t have the exponential distribution with the density $f(x) = \lambda e^{-\lambda x} I[x \geq 0]$ for some $\lambda > 0$. Consider the model (3.2) for $c \geq 0$. Then*

$$p(c) = 1 - \frac{\exp\{-\lambda c\}}{2}, \qquad r_1(c) = 1 - \frac{2}{3} \cdot \frac{3 - e^{-2\lambda c}}{2 - e^{-\lambda c}}.$$

It follows from the latter formulas that $c = -\log[2(1-p)]/\lambda$, and thus

$$r_1(p) = \frac{(1-p)(1-4p)}{3p}, \quad p \in [1/2, 1).$$

It is easy to show that the uniform distribution gives lower values of r_1 than the exponential one for all $p \in (0, 1)$, cf. Figure 3.1.

It is possible to derive an explicit formula for r_1 as a function of p for the uniform and exponential distributions. Unfortunately, this is not the case for all the continuous distributions.

Proposition 3.10. *Let X_t have the logistic distribution with the distribution function $F(x) = (1 + e^{-x})^{-1}$, $x \in \mathbb{R}$. Consider the model (3.2) with $c > 0$. Then*

$$p(c) = \frac{e^c(e^c - 1 - c)}{(e^c - 1)^2},$$
$$r_1(c) = \frac{ce^c(2 + c - 2e^c + ce^c)}{(1 + e^c)(e^c - 1 - c)(1 + (c-1)e^c)}.$$

For $c = 0$ we obtain the trivial case $p = 1/2$ and $r_1 = -1/3$ mentioned in Example 3.6. The functions $p(c), r_1(c)$ are both continuous at the point $c = 0$.

Proposition 3.11. *Let $A > 0$, and let X_t have the double exponential distribution with the density $f(x) = (A/2)\,\mathrm{e}^{-A|x|}$, $x \in \mathbb{R}$. Consider the model (3.2) with $c \geq 0$. Then*

$$p(c) = \frac{\mathrm{e}^{-Ac}(4\mathrm{e}^{Ac} - 2 - Ac)}{4},$$

$$r_1(c) = \frac{12Ac + 3A^2c^2 + 4\mathrm{e}^{-Ac}}{12 + 12Ac + 3A^2c^2 - 24\mathrm{e}^{Ac} - 12Ac\,\mathrm{e}^{Ac}}.$$

Clearly, the inverse formula for c as a function of p can be obtained neither for the logistic distribution nor for the double exponential. Hence, the values of r_1 for a given p need to be computed numerically. In addition, for X_t with the standard Gaussian distribution one even cannot express the success probability p in a closed form, and the values of p, and r_1 for a given c need to be computed numerically as well.

	Uniform	Exponential	Gaussian	Logistic	Double Exp.
p			$r_1(p)$		
0.5	−0.3333	−0.3333	−0.3333	−0.3333	−0.3333
0.55	−0.3316	−0.3273	−0.3294	−0.3287	−0.3269
0.6	−0.3258	−0.3111	−0.3175	−0.3150	−0.3090
0.65	−0.3148	−0.2872	−0.2979	−0.2928	−0.2825
0.7	−0.2971	−0.2571	−0.2710	−0.2632	−0.2498
0.75	−0.2702	−0.2222	−0.2372	−0.2274	−0.2129
0.8	−0.2306	−0.1833	−0.1973	−0.1867	−0.1732
0.85	−0.1753	−0.1412	−0.1521	−0.1425	−0.1316
0.9	−0.1111	−0.0963	−0.1029	−0.0960	−0.0888
0.95	−0.0526	−0.0491	−0.0514	−0.0483	−0.0452

Table 3.1: Comparison of r_1 in the model (3.2) for X_t having uniform, exponential, Gaussian, logistic, and double exponential distribution

Distributions with a finite support. The values of the autocorrelation r_1 in the model (3.2) were compared numerically for the uniform, exponential, and Gaussian distribution in Bondesson (2003). We performed an extended comparison with the logistic and double exponential distribution as well. It was found that for any given $p \in (0, 1)$ the uniform distribution gives the lowest (i.e. maximally negative) correlations, see Table 3.1, cf. Figure 3.1. Further considerations based on this comparison makes us believe that in order to obtain the lowest possible r_1 it is better to choose a distribution with a majority of the mass concentrated on a bounded interval.

In the following we consider some other continuous distributions, namely distributions with U-shaped density on a finite support, and we compare them with the uniform

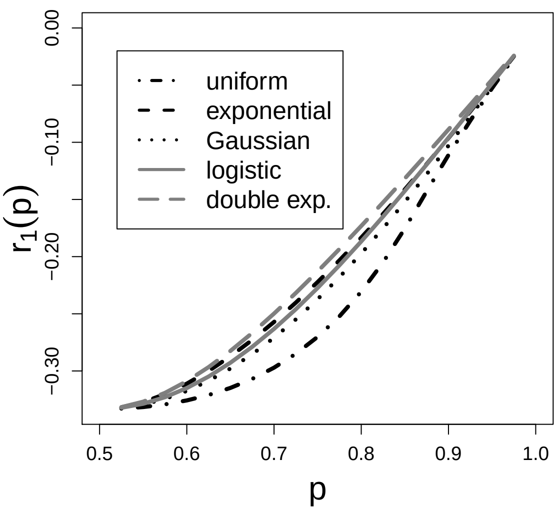


Figure 3.1: Comparison of r_1 in the model (3.2) for X_t having uniform, exponential, Gaussian, logistic, and double exponential distribution

distribution. We show that the uniform distribution is not generally "optimal" in sense that for some $p \in (0,1)$ there exists a continuous distribution leading to a lower r_1. However, for p close to $1/2$ (r_1 close to $-1/3$) the uniform distribution still gives the lowest possible values for r_1.

The following theorem describes a general property of r_1 in the model (3.2) with X_t continuously distributed over a finite interval.

Theorem 3.12. *Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid variables with an absolutely continuous distribution with a density f such that $f(x) = 0$ for $x \notin (0,d)$ for some $0 < d < \infty$. Let $\{\xi_t,\, t \in \mathbb{Z}\}$ be defined by (3.2) for some $c \in \mathbb{R}$. Let F be the cumulative distribution function of X_t. Then the 1-lag autocorrelation r_1 of $\boldsymbol{\xi}$ satisfies*

$$r_1(c) = 1 - \frac{1}{p(c)} \quad \text{for } d/2 \le c < d \qquad \text{and} \qquad r_1(c) = 1 - \frac{1}{1-p(c)} \quad \text{for } -d < c \le -d/2,$$

where $p(c) = \mathsf{E}F(X+c)$ is the success probability of ξ_t. In particular, $r_1(p) = -(1-p)/p$ for all $p \in [p(d/2), 1)$ and $r_1(p) = -p/(1-p)$ for all $p \in (0, p(-d/2)]$.

Proof. Recall the formulas (3.5) and (3.6). The cdf $F(x)$ is zero for $x \le 0$ and 1 for $x \ge d$. Hence, if $c \ge d/2$ then $F(x-c) = 0$ for all $x < d/2$ and $F(x+c) = 1$ for all

$x > d/2$. We get

$$1 - p(c) = \mathsf{E}F(X - c) = \int_0^d F(x - c)f(x)\,\mathrm{d}x = \int_{d/2}^d F(x - c)f(x)\,\mathrm{d}x,$$
$$Q(c) = \int_0^d F(x + c)F(x - c)f(x)\,\mathrm{d}x = \int_{d/2}^d F(x - c)f(x)\,\mathrm{d}x = 1 - p(c)$$

for all $c \in [d/2, d)$. This implies that $r_1(c) = 1 - 1/p(c)$ for all $c \geq d/2$. Since $Q(-c) = Q(c)$ and $p(-c) = 1 - p(c)$ we get $r_1(c) = 1 - 1/[1 - p(c)]$ for $c \leq -d/2$. □

In view of Theorem 3.12 and Remark 3.4 a distribution on a finite interval $(0, d)$ leads to the minimal possible r_1 for all $p \geq p(d/2)$ and $p \leq p(-d/2)$.

Let us now introduce the following three densities f_1, f_2, f_3 defined as

$$f_1(x) = \begin{cases} \frac{3}{2}(x-1)^2, & x \in [0,2], \\ 0 & \text{otherwise}, \end{cases}$$
$$f_2(x) = \begin{cases} |1-x|, & x \in [0,2], \\ 0 & \text{otherwise}, \end{cases}$$
$$f_3(x) = \begin{cases} \frac{5}{2}(x-1)^4, & x \in [0,2], \\ 0 & \text{otherwise}. \end{cases}$$

All the three densities have a minimum at the point 1, they are symmetric around this point, and they vanish outside the interval $[0, 2]$.

The computations of the probability p and the 1-lag autocorrelation r_1 for the densities f_1, f_2, and f_3 are based on the formulas (3.5) and (3.6). Since the final form of the correlation r_1 is quite complicated only the formulas for p and Q are reported. The value of r_1 can be then obtained from (3.6). For each of the four models we assume $c \in [0, 2]$ because $p = 1$ for $c > 2$ and $p = 0$ for $c < 0$.

Proposition 3.13. *Consider the model* (3.2).

If X_t has a distribution with the density f_1 then

$$p(c) = (40 + 72c - 90c^2 + 40c^3 - c^6)/80, \quad c \in [0, 2],$$
$$Q(c) = \begin{cases} \dfrac{(560 - 2430c^2 + 7560c^3 - 10584c^4 + 7560c^5 - 2331c^6 + 64c^9)}{1680}, & c \in [0, 1), \\ 1 - p(c), & c \in [1, 2]. \end{cases}$$

If X_t has a distribution with the density f_2 then

$$p = \begin{cases} (12 + 16c - 12c^2 + 3c^4)/24, & c \in [0, 1), \\ (24 - 16c + 12c^2 - c^4)/24, & c \in [1, 2], \end{cases}$$

$$Q = \begin{cases} (8 - 18c^2 + 32c^3 - 17c^4 - 2c^6)/24, & c \in [0, 1/2), \\ (10 - 16c + 30c^2 - 32c^3 + 15c^4 - 2c^6)/24, & c \in [1/2, 1), \\ 1 - p(c), & c \in [1, 2]. \end{cases}$$

If X_t has a distribution with the density f_3 then

$$p = (504 + 1400c - 3150c^2 + 3600c^3 - 2100c^4 + 504c^5 - c^{10})/1008, \quad c \in [0, 2],$$

$$Q = \begin{cases} 1/3 - 375c^2/104 + 125c^3/6 - 8375c^4/132 + 125c^5 - 3025c^6/18 \\ +925c^7/6 - 650c^8/7 + 100c^9/3 - 5503c^{10}/1008 + 64c^{15}/9009, & c \in [0, 1], \\ 1 - p(c), & c \in (1, 2]. \end{cases}$$

It is possible to show that the correlation r_1 reaches its global minimum $-1/3$ for $c = 0$ ($p = 1/2$) for each density f_1, f_2, f_3. Furthermore, besides this global minimum, r_1 has a local minimum equal to -0.3294 at the point $c = 0.7067$ ($p = 0.7491$) for the density f_1, -0.3172 at $c = 0.6230$ ($p = 0.7401$) for f_2, and -0.3332 at $c = 0.7576$ ($p = 0.749988$) for f_3.

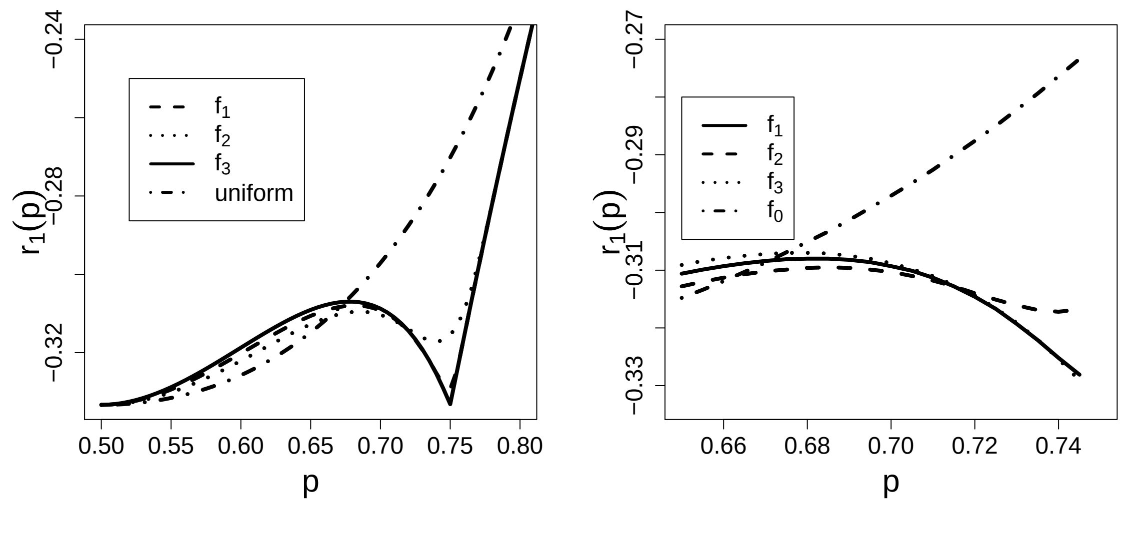


Figure 3.2: Comparison of r_1 in the model (3.2) with X_t having a distribution with densities f_1, f_2, f_3 and the uniform distribution

For each of the three distributions we get $Q(c) = 1 - p(c)$ for $c \geq 1$, and thus $r_1 = 1 - 1/p(c)$ for all $c \geq 1$. This illustrates the conclusion of Theorem 3.12. Hence,

	Density f_1	Density f_2	Density f_3	Uniform
p	$r_1(p)$			
0.500	−0.3333	−0.3333	−0.3333	−0.3333
0.525	−0.3323	−0.3325	−0.3321	−0.3329
0.550	−0.3294	−0.3301	−0.3288	−0.3316
0.575	−0.3252	−0.3266	−0.3241	−0.3292
0.600	−0.3202	−0.3221	−0.3187	−0.3258
0.625	−0.3151	−0.3172	−0.3134	−0.3210
0.650	−0.3106	−0.3128	−0.3091	−0.3148
0.675	−0.3081	−0.3099	−0.3070	−0.3069
0.700	−0.3093	−0.3103	−0.3088	−0.2971
0.725	−0.3167	−0.3151	−0.3166	−0.2850
0.750	−0.3293	−0.3155	−0.3332	−0.2702
0.775	−0.2903	−0.2899	−0.2903	−0.2522
0.800	−0.2500	−0.2500	−0.2500	−0.2306
0.825	−0.2121	−0.2121	−0.2121	−0.2050
0.850	−0.1765	−0.1765	−0.1765	−0.1753
0.875	−0.1429	−0.1429	−0.1429	−0.1429
0.900	−0.1111	−0.1111	−0.1111	−0.1111
0.925	−0.0811	−0.0811	−0.0811	−0.0811
0.950	−0.0526	−0.0526	−0.0526	−0.0526
0.975	−0.0256	−0.0256	−0.0256	−0.0256

Table 3.2: Comparison of r_1 in the model (3.2) with X_t having a distribution with densities f_1, f_2, f_3 and the uniform distribution

the model (3.2) with X_t distributed with densities f_1, f_2, f_3 is "optimal"—in the sense that it reaches the boundary $1 - 1/p$ for r_1—whenever $p \geq p(1)$. For f_1, f_2, f_3 we get $p(1)$ equal to $61/80 = 0.7625$, $19/24 = 0.7917$, $757/1008 = 0.7510$ respectively. For the uniform distribution on $[0, 1]$ we have $p(1/2) = 7/8 = 0.8750$.

Let us compare the correlation r_1 in the model (3.2) with X_t having the uniform distribution and the distribution with the densities f_1, f_2, f_3 respectively. Unfortunately, it is not possible to express r_1 as a function of p for the densities f_1, f_2, f_3, and therefore the comparison can be performed only numerically. All the numerical calculations were conducted in the program Mathematica 5.2, see Wolfram Research (2005). The computation accuracy was set to be 50 digits. For a given probability $p \in [1/2, 1)$ we obtained the corresponding $c > 0$ numerically and then the correlation r_1 was computed. The results are summarized in Table 3.2, a graphical illustration is provided in Figure 3.2. The smallest number in each row is underlined. The uniform distribution leads to the

	Density f_1	Density f_2	Density f_3	Uniform
p	$r_1(p)$			
0.650	−0.3106	−0.3128	−0.3091	−0.3148
0.655	−0.3099	−0.3120	−0.3084	−0.3134
0.660	−0.3093	−0.3113	−0.3079	−0.3119
0.665	−0.3088	−0.3107	−0.3075	−0.3103
0.670	−0.3084	−0.3102	−0.3072	−0.3087
0.675	−0.3081	−0.3099	−0.3070	−0.3069
0.680	−0.3080	−0.3096	−0.3070	−0.3051
0.685	−0.3080	−0.3095	−0.3071	−0.3033
0.690	−0.3082	−0.3096	−0.3075	−0.3013
0.695	−0.3086	−0.3099	−0.3080	−0.2992
0.700	−0.3093	−0.3103	−0.3088	−0.2971
0.705	−0.3101	−0.3110	−0.3098	−0.2949
0.710	−0.3113	−0.3118	−0.3110	−0.2926
0.715	−0.3128	−0.3128	−0.3126	−0.2901
0.720	−0.3146	−0.3140	−0.3144	−0.2876
0.725	−0.3167	−0.3151	−0.3166	−0.2850
0.730	−0.3193	−0.3161	−0.3191	−0.2822
0.735	−0.3221	−0.3169	−0.3220	−0.2794
0.740	−0.3252	−0.3172	−0.3254	−0.2764
0.745	−0.3281	−0.3168	−0.3291	−0.2734

Table 3.3: Comparison of $r_1(p)$ for $p \in [0.65, 0.75]$ in the model (3.2) with X_t having a distribution with densities f_1, f_2, f_3 and the uniform distribution

lowest values of r_1 for p closed to 1/2. However, the situation changes if $p > 0.65$. For $p > 0.7510$, the lowest correlation is equal to $1 - 1/p$ and it is obtained by the density f_3. If $p \in [0.875, 1]$ then $r_1 = 1 - 1/p$ for all the involved distributions, and the correlations are exactly the same. Results for p from the interval $[0.65, 0.75]$ are shown in more detail in Table 3.3. We can see that if p increases from 0.5 to 0.75 then the lowest value of r_1 is first obtained by the constant density then by the linear density f_2, consequently by the quadratic function f_1, and finally by the fourth degree polynomial f_3. Remark that the density for which the minimal r_1 is obtained changes in the same way as the maximum of f_i, $i = 1, 2, 3, 4$.

Hence, the performed comparison shows that the uniform distribution is not generally "optimal" in the sense that for some $p \in (0, 1)$ there exists a continuous distribution leading to a lower r_1. However, it still gives generally "good results" (in the same sense) compared to other distributions.

3.1.2 Generalization of the Bondesson's Model

Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid random variables with an absolutely continuous CDF F. Let $a > 0$ and $c \in \mathbb{R}$ be some constants. Define

$$\xi_t = \begin{cases} 1 & \text{if } X_t - aX_{t-1} < c, \\ 0 & \text{otherwise.} \end{cases} \tag{3.10}$$

The variables $\{\xi_t,\, t \in \mathbb{Z}\}$ are 1-dependent, and the process $\boldsymbol{\xi} = \{\xi_t,\, t \in \mathbb{Z}\}$ is (strictly) stationary. Their characteristics p and r_1 depend on the distribution F of X_t and the parameters a, c. We stress this dependence by writing $p(a, c)$ instead of p only if necessary. The particular variable ξ_t is an indicator whether the MA(1) process $X_t - aX_{t-1}$ stays below the level c at time t. See Figure 3.3 for an illustration.

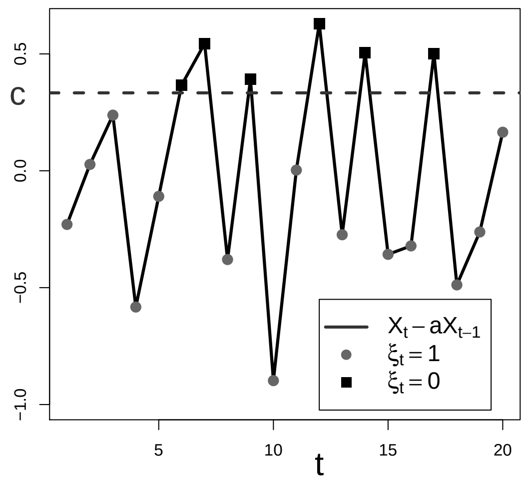


Figure 3.3: Construction of $\{\xi_t,\, t \in \mathbb{Z}\}$ from the MA(1) process $X_t - aX_{t-1}$ (simulation for the uniform distribution on $[0, 1]$ with $a = 1$, $c = 1/3$, setting `set.seed(2010)` in the program `R`)

The model (3.10) generalizes the model (3.2). We show that $\{\xi_t,\, t \in \mathbb{Z}\}$ defined by (3.10) are always negatively correlated as well, see Corollary 3.15.

We first describe some properties of the studied model.

Proposition 3.14. *Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid variables with an absolutely continuous distribution with a density f and a CDF F. Let $\boldsymbol{\xi} = \{\xi_t,\, t \in \mathbb{Z}\}$ be defined by* (3.10) *for some $a > 0$ and $c \in \mathbb{R}$. Define*

$$Q(a, c) = E\Big[F\Big(\frac{X_1 - c}{a}\Big)F(c + aX_1)\Big]. \tag{3.11}$$

Then

$$p(a,c) = \mathsf{E}F(aX_1 + c), \quad r_1(a,c) = 1 - \frac{Q(a,c)}{p(a,c)[1-p(a,c)]}, \tag{3.12}$$

and

$$p(a,c) = 1 - p\left(\frac{1}{a}, -\frac{c}{a}\right), \quad r_1(a,c) = r_1\left(\frac{1}{a}, -\frac{c}{a}\right). \tag{3.13}$$

Proof. We have

$$\mathsf{P}(\xi_t = 1) = \mathsf{P}(X_t - aX_{t-1} < c) = \int_{-\infty}^{\infty} F(c+ay)f(y)\ \mathrm{d}y = \mathsf{E}F(aX_1 + c),$$

$$\begin{aligned}\mathsf{P}(\xi_t = 1, \xi_{t-1} = 1) &= \mathsf{P}(X_t - aX_{t-1} < c,\ X_{t+1} - aX_t < c) \\ &= \int_{-\infty}^{\infty} \Big[1 - F\Big(\frac{y-c}{a}\Big)\Big] F(c+ay)f(y)\ \mathrm{d}y \\ &= \mathsf{E}F(c+aX_1) - \mathsf{E}\Big[F\Big(\frac{X_1 - c}{a}\Big)F(c+aX_1)\Big].\end{aligned}$$

The formula for r_1 in (3.12) then easily follows.

It follows from (3.10) that $\xi_t = \mathrm{I}[X_{t-1} - (1/a)X_t > -c/a]$, where $\mathrm{I}[\cdot]$ stands for the indicator function. Then $\xi_t = 1 - \tilde{\xi}_t$, where $\tilde{\boldsymbol{\xi}} = \{\tilde{\xi}_t,\, t \in \mathbb{Z}\}$ has the same distribution as the variables defined by (3.10) with $\tilde{a} = 1/a$ and $\tilde{c} = -c/a$. The autocorrelation function $\{r_t\}$ of the process $\boldsymbol{\xi}$ coincides with the autocorrelation function of the process $\tilde{\boldsymbol{\xi}}$, and the success probabilities are related by $p = 1 - \tilde{p}$, where $\tilde{p} = \mathsf{P}(\tilde{\xi}_t = 1)$. This proves (3.13). □

The formulas for p and r_1 in (3.12) generalize (3.5) and (3.6) derived by Bondesson (2003).

Corollary 3.15. *Let* $\{\xi_t,\, t \in \mathbb{Z}\}$ *be defined by* (3.10)*. Then* $\boldsymbol{\xi} = \{\xi_t,\, t \in \mathbb{Z}\}$ *is a strictly stationary process with 1-dependent negatively correlated variables.*

Proof. It suffices to show that $r_1(a,c) \le 0$ for all $a > 0$ and $c \in \mathbb{R}$. Formulas (3.12) and (3.13) imply

$$r_1(a,c) = 1 - \frac{E\Big[F\Big(\frac{X_1-c}{a}\Big)F(c+aX_1)\Big]}{\mathsf{E}F(aX_1+c)\mathsf{E}F\Big(\frac{X_1-c}{a}\Big)}.$$

Define $U = F(aX_1 + c)$ and $V = F\big[(X_1 - c)/a\big]$. Then $r_1 = -\mathsf{cov}\,(U,V)/(\mathsf{E}U\mathsf{E}V)$, and $r_1 \le 0$ if and only if $\mathsf{cov}\,(U,V) \ge 0$. The functions $F(ax+c)$ and $F[(x-c)/a]$ are increasing functions of x for all $a > 0$ and $c \in \mathbb{R}$. Hence, the inequality $\mathsf{cov}\,(U,V) \ge 0$ is a consequence of Chebyshev's inequality for similarly ordered functions, see Hardy et al. (1988) . A different justification is that a set consisting of a single random variable X_1 is positively associated, i.e. $\mathsf{cov}\,[f(X_1), g(X_1)] \ge 0$ whenever f, g are nondecreasing functions and $\mathsf{E}f(X_1)$, $\mathsf{E}g(X_1)$, $\mathsf{E}f(X_1)g(X_1)$ exist, see Esary et al. (1967). □

Some properties of the functions p, Q, and r_1 are summarized in the following Proposition 3.16.

Proposition 3.16. *Consider the same setting as in Proposition 3.14.*

1. *The functions $p(a,c)$ and $Q(a,c)$ are continuous on $(0,\infty)\times\mathbb{R}$. The function $r_1(a,c)$ is continuous in all $(a,c)\in(0,\infty)\times\mathbb{R}$ satisfying $p(a,c)\in(0,1)$.*
2. *Let $a>0$ be fixed. Define $C_1=\sup\{c\in\mathbb{R}:p(a,c)=0\}$, $C_2=\inf\{c\in\mathbb{R}:p(a,c)=1\}$ (with the classical convention $\sup\emptyset=-\infty$ and $\inf\emptyset=\infty$). The function $p(a,\cdot)$ is nondecreasing on $\mathbb{R}$ and increasing on (C_1,C_2). We have*

$$\lim_{c\to\infty}p(a,c)=1,\qquad \lim_{c\to\infty}Q(a,c)=0,\qquad \lim_{c\to C_1^+}r_1(a,c)=0,$$

$$\lim_{c\to-\infty}p(a,c)=0,\qquad \lim_{c\to-\infty}Q(a,c)=0,\qquad \lim_{c\to C_2^-}r_1(a,c)=0.$$

3. *If $c\in\mathbb{R}$ is fixed then*

$$\lim_{a\to\infty}p(a,c)=1-F(0),\qquad \lim_{a\to0^+}p(a,c)=F(c),$$

$$\lim_{a\to\infty}Q(a,c)=F(0)[1-F(0)],\qquad \lim_{a\to0^+}Q(a,c)=F(c)[1-F(c)].$$

If $f\equiv0$ on $(-\infty,0)$ then $p(\cdot,c)$ is nondecreasing on $(0,\infty)$ and increasing at all $a>0$ that satisfy $p(a,c)\in(0,1)$.

Proof. The functions $p(a,c)$ and $Q(a,c)$ are expressed in (3.12) and (3.11). Their continuity follows directly from the well-known theorem about continuity of an integral depending on a parameter, see (Zorich, 2004, chap. 17). It follows from the dominated convergence theorem that the limits in 2. and 3. can be passed under the integral sign. The limits of $p(a,c)$ and $Q(a,c)$ are then easily obtained. The inequality $Q(a,c)\le p(a,c)$ together with Corollary 3.15 imply that $0\ge r_1(a,c)\ge 1-1/[1-p(a,c)]$. Hence, $\lim_{c\to C_1^+}r_1(a,c)=0$. For $c\to C_2^-$ use the symmetry in (3.13). Finally, the monotonicity of p in 2. and 3. follows from the monotonicity of the CDF F. □

Remark 3.17. Proposition 3.16 deals with the continuity of the functions p, Q, and r_1. In view of the well-known theorem about differentiation of an integral depending on a parameter, see (Zorich, 2004, chap. 17), we are able to formulate some general additional conditions for the distribution F under which the functions p and r_1 are differentiable with respect to the variables a and c respectively. However, these conditions do not give much insight, and therefore the differentiability is investigated only in special cases using the formulas for p and r_1 as in Propositions 3.18 and 3.19.

The success probability $p(a,c)$ and the correlation $r_1(a,c)$ can be obtained for a given distribution F, $a>0$, and $c\in\mathbb{R}$ from the formulas in (3.12) after some integration and algebraic manipulations. We present results for two particular distributions F in order to illustrate the conclusions of Proposition 3.16.

Proposition 3.18. *Let* $\{X_t,\, t \in \mathbb{Z}\}$ *be iid random variables with the exponential distribution with the density* $f(x) = \lambda e^{-\lambda x} \mathrm{I}[x \geq 0]$ *for some* $\lambda > 0$. *Then*

$$p(a,c) = \begin{cases} 1 - \dfrac{e^{-\lambda c}}{a+1}, & c \geq 0, \\ \dfrac{a e^{\lambda c/a}}{a+1}, & c < 0, \end{cases} \tag{3.14}$$

and

$$r_1(a,c) = \begin{cases} \dfrac{e^{-\lambda a c}(1+a) - (1+a+a^2)}{(1+a+a^2)[(1+a)e^{\lambda c} - 1]}, & c > 0, \\ a\, \dfrac{e^{\lambda c/a}(1+a+a^2) - a(1+a)e^{\lambda c(1+a)/a^2}}{(1+a+a^2)[a e^{\lambda c/a} - 1 - a]}, & c \leq 0. \end{cases} \tag{3.15}$$

The function $p(a,c)$ in (3.14) illustrates the general conclusions of Proposition 3.16. In particular, $p(a,\cdot)$ is increasing on $\mathbb{R}$ for all $a > 0$, and $p(\cdot,c)$ is increasing on $(0,\infty)$ for all $c \in \mathbb{R}$. Moreover, it follows from (3.14) that $p(a,c)$ is continuously differentiable with respect to both a, c, and $\partial p/\partial c > 0$, $\partial p/\partial a > 0$. The range of $p(a,\cdot)$ is equal to $(0,1)$ for all $a > 0$. On the other hand, the range of $p(\cdot,c)$ is equal to $(0,1)$ for $c \leq 0$ and equal to $(F(c),1) = (1 - e^{-c}, 1)$ for $c > 0$.

The function $r_1(a,c)$ in (3.15) is plotted in Figure 3.4(a). It follows from (3.15) that $r_1(a,c)$ in (3.15) is continuously differentiable with respect to both a, c (the result for $c = 0$ follows after some algebraic manipulations). It has a global minimum equal to $-1/3$ at the point $(a,c) = (1,0)$ (this corresponds to $p = 1/2$). Indeed, if $a > 0$ is fixed then $\partial r_1/\partial c > 0$ for $c \in (0,\infty)$ and $\partial r_1/\partial c < 0$ for $c \in (-\infty,0)$. We get $r_1(a,c) \geq r_1(a,0) = -a(1+a+a^2)^{-1} \geq -1/3 = r_1(1,0)$.

It was stated in the previous section that the uniform distribution gives "good" (though not always the "best") results for the negative correlation r_1. Moreover, from the computational and practical points of view the uniform distribution is one of the most convenient to work with. Hence, we look at this case in more detail.

Proposition 3.19. *Let* $\{X_t,\, t \in \mathbb{Z}\}$ *be iid random variables with the uniform distribution on the interval* $[0,1]$. *Consider model* (3.10) *for* $a > 0$ *and* $c \in (-1,1)$. *Then*

$$p(a,c) = \begin{cases} 1 - \dfrac{(1-c)^2}{2a}, & a + c > 1 \text{ and } c \geq 0, \\ \dfrac{a}{2} + c, & a + c \leq 1 \text{ and } c \geq 0, \\ \dfrac{2a + 2c - 1}{2a}, & a + c > 1 \text{ and } c < 0, \\ \dfrac{(a+c)^2}{2a}, & 0 \leq a + c \leq 1 \text{ and } c < 0, \\ 0, & a + c < 0. \end{cases} \tag{3.16}$$

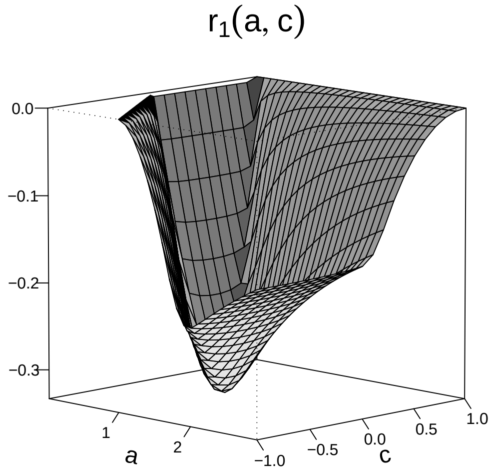
If $c \in [0,1)$ *then*

$$r_1(a,c) = \begin{cases} -\dfrac{a(-3a+2a^2-6c+6ac+6c^2)}{3(2c+a-2)(a+2c)}, & a+c<1, \\ 1-\dfrac{a}{3(1-c)^2}+\dfrac{1+a}{1-c} \\ \quad +\dfrac{-2-3a+10a^2+4a^3+2c+9ac+9a^2c+2a^3c}{3a(1-2a-2c+c^2)}, & 1-c \le a < \dfrac{1}{c}-1, \\ -\dfrac{(1-c)^2}{-1+2a+2c-c^2}, & a \ge \dfrac{1}{c}-1. \end{cases} \tag{3.17}$$

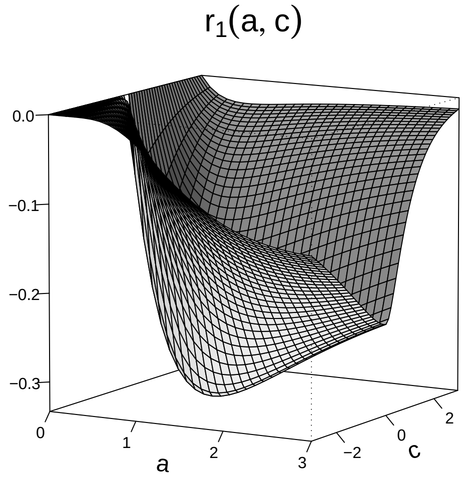
The correlation r_1 *for* $c \in (-1,0)$ *can be obtained from* (3.17) *using* (3.13).

It is clear that $p(a,c)=1$ for $c \ge 1$ and $p(a,c)=0$ for $c \le -1$. Hence, $r_1(a,c)$ is not defined for $|c| \ge 1$, and it suffices to consider only $c \in (-1,1)$ in Proposition 3.19. The symmetry (3.13) implies that $p(a,c)=0$ if $a+c \le 0$. The other expressions for $p(a,c)$ and $r_1(a,c)$ follow directly from (3.12). The function $p(a,\cdot)$ is nondecreasing on $\mathbb{R}$ and increasing on $(-1,1)$. The function $p(\cdot,c)$ in increasing on $(0,\infty)$ if $c>0$ and increasing on $(-c,\infty)$ if $c<0$. Furthermore, $p(a,c)$ is continuously differentiable with respect to both a,c. If $a>0$, $c \in (-1,1)$, and $a+c>0$ then $\partial p/\partial c > 0$, $\partial p/\partial a > 0$, and $r_1(a,c)$ is also continuously differentiable with respect to both a,c. The function $r_1(a,c)$ has a global minimum equal to $-1/3$ at the point $(a,c)=(1,0)$.

The function $r_1(a,c)$ from Proposition 3.19 is plotted in Figure 3.4(b).



(a) uniform



(b) exponential with $\lambda = 1$

Figure 3.4: Correlation r_1 in model (3.10) for the uniform and exponential distribution

For the special case $a = 1$ we get the results stated in Proposition 3.8. It was mentioned that for $a = 1$ the minimum autocorrelation is reached for $c = 0$ (i.e. $p = 1/2$). On the other hand, for the special case $c = 0$ we obtain

$$p(a) = \begin{cases} \dfrac{a}{2}, & a \le 1, \\ 1 - \dfrac{1}{2a}, & a > 1, \end{cases} \quad \text{and} \quad r_1(a) = \begin{cases} \dfrac{a(2a-3)}{3(2-a)}, & a \le 1, \\ \dfrac{2-3a}{3a(2a-1)}, & a > 1. \end{cases}$$

If $c = 0$ then the characteristics p and r_1 satisfy $p(a) = 1 - p(1/a)$, $r_1(a) = r_1(1/a)$, and $r_1(a) \ge r_1(1) = -1/3$ holds. We can express the autocorrelation r_1 as a function of the success probability $p \in (0, 1)$ as

$$r_1(p) = \begin{cases} \dfrac{p(4p-3)}{3(1-p)}, & p \in (0, 1/2], \\ \dfrac{(1-4p)(1-p)}{3p}, & p \in (1/2, 1). \end{cases}$$

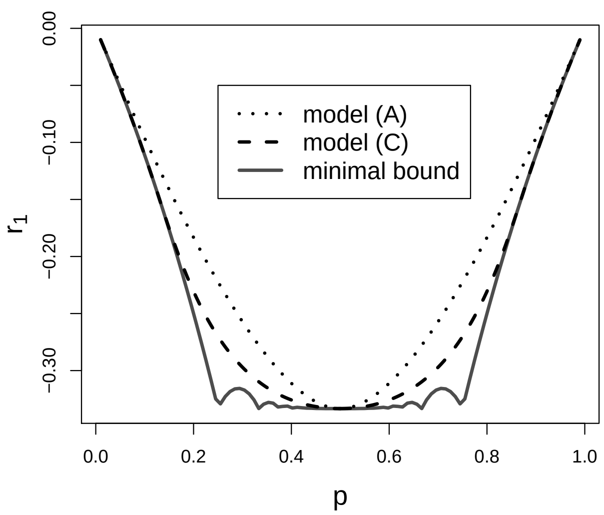


Figure 3.5: Comparison of the minimal bound $\min[r_1(p)]$ with r_1 from the model (3.10)

Let (A) denote the model (3.10) with $c = 0$ and $a > 0$, and let (C) denote the model (3.10) with $a = 1$ and $c \in [-1, 1]$. Figure 3.5 provides a comparison of the minimal bound $\min[r_1(p)]$, introduced in Remark 3.4, and the correlation r_1 corresponding to the models (A) and (C) respectively. The model (C) leads to a lower value of r_1 than the model (A) for all $p \in (0, 1)$ and therefore, we look at the model (C) in more detail.

Figure 3.5 shows that if p is close to 1/2 then the correlation r_1 is very close to $\min[r_1(p)]$. Similar conclusion is obtained for $p < 1/5$ and $p > 4/5$. Remind that the formula (3.3) implies that if $p < 1/4$ then $m = 1$ and $\min[r_1(p)] = -p/(1-p)$. It follows from (3.9) that r_1 and $\min[r_1(p)]$ coincide for $p \leq 1/8$ and $p \geq 7/8$, cf. Section 3.1.1 and Theorem 3.12.

Remark 3.20. It is stated in Remark 3.4 that the inequality $r_1 \geq -1/3$ holds for all 0-1 valued two-block-factors. In the case of 1-dependent variables taking three different values the correlation r_1 can be made arbitrary close to $-1/2$. This illustrates that the case of Bernoulli variables has a unique position in the class of all two-block-factors.

Let $\{X_t,\, t \in \mathbb{Z}\}$ be discrete iid random variables such that $\mathsf{P}(X_t = i) = u_i$ for $i = 1, 2, 3$ and $u_1 + u_2 + u_3 = 1$. Define $\xi_t = \mathrm{I}[X_t = X_{t-1}] + 2\mathrm{I}[X_t > X_{t-1}]$, $t \in \mathbb{Z}$. Then $\{\xi_t,\, t \in \mathbb{Z}\}$ is a stationary process with 1-dependent variables. We get $\mathsf{P}(\xi_t = 0) = \mathsf{P}(\xi_t = 2) = u_1u_2 + u_1u_3 + u_2u_3$, $\mathsf{P}(\xi_t = 1) = u_1^2 + u_2^2 + u_3^2$, and thus $\mathsf{E}\xi_t = 1$, $\mathsf{var}\ \xi_t = 1 - u_1^2 - u_2^2 - u_3^2$. Define $u_{ij} = \mathsf{P}(\xi_t = i, \xi_{t-1} = j)$ for $i = 0, 1, 2$. Then

$$
\begin{aligned}
u_{00} &= u_1u_2u_3, & u_{01} &= u_1u_2^2 + u_1u_3^2 + u_2u_3^2,\\
u_{10} &= u_1^2u_2 + u_1^2u_3 + u_2^2u_3, & u_{11} &= u_1^3 + u_2^3 + u_3^3,\\
u_{12} &= u_{01}, & u_{21} &= u_{10},\\
u_{02} &= u_{10} + 2u_1u_2u_3, & u_{20} &= u_{01} + 2u_1u_2u_3,\\
u_{22} &= u_1u_2u_3. & &
\end{aligned}
$$

These expressions lead to $\mathsf{cov}\,(\xi_t, \xi_{t-1}) = -(1-u_1)(u_1 + u_2u_3)$ and

$$
r_1 = -\frac{(1-u_1)(u_1 + u_2u_3)}{1 - u_1^2 - u_2^2 - u_3^2}.
$$

For $u_1 = u_2 = u_3 = 1/3$ we have $r_1 = -4/9$. If $u_2 \to 0$, $u_3 \to 0$, and $u_1 \to 1$, then $r_1 \to -\frac{1}{2}$.

3.2 Estimation of Parameters of the Generalized Bondesson's Model

In this section, we deal with the estimation problem in the model (3.10). Properties of the classical estimators of p and r_1 are investigated, and an alternative estimator of r_1 is proposed. Moreover, we construct estimators of the parameters a, c. In other words, we estimate the parameters of the underlying process $X_t - aX_{t-1} - c$ clipped at the level 0 from the clipped data $\xi_1, \ldots, \xi_n$. Several authors have been concerned with a similar estimation problem, see Lomnicki and Zaremba (1955), Kedem (1980b), Hinich (1967), Damsleth and El-Shaarawi (1988), cf. Remark 3.2. However, all these papers deal with the restrictive assumption that the original clipped process is Gaussian and the success probability p is equal to 1/2. This is not necessary the case in model (3.10), and therefore we use a different approach. We remark that this section is based on results from Došlá (2010).

Recall that a sequence of random variables $\{X_n,\ n \in \mathbb{N}\}$ is said to be *asymptotically normal* $\mathsf{AN}(\mu_n, \sigma_n^2)$ if $\sigma_n^2 > 0$ for sufficiently large n and $\sigma_n^{-1}(X_n - \mu_n)$ converges in law to the standard normal distribution $\mathsf{N}(0,1)$ as $n \to \infty$. In addition, we say that $\{X_n,\ n \in \mathbb{N}\}$ is $\mathsf{AN}(\mu_n, 0)$ if $X_n - \mu_n$ converges to 0 in probability as $n \to \infty$.

3.2.1 Estimation of p and r_1

The following assertion is a corollary of several well-known results for sequences of m-dependent strictly stationary variables.

Proposition 3.21. *Let $\{\xi_t,\ t \in \mathbb{N}\}$ be a sequence of strictly stationary 1-dependent Bernoulli variables with $p = \mathsf{P}(\xi_t = 1) \in (0,1)$. Define $Q_i = \mathsf{P}(\xi_t = 1, \xi_{t+1} = 1, \dots, \xi_{t+i} = 1)$ for $i = 1, 2, \dots$.*

1. *Define $T_n = \bar{\xi}_n = n^{-1}\sum_{t=1}^{n} \xi_t$. Then T_n is an unbiased and consistent estimator of the success probability p, and $\sqrt{n}(T_n - p)$ is $\mathsf{AN}(0, U)$, where $U = p + 2Q_1 - 3p^2$.*
2. *Define $S_n = [1/(n-1)]\sum_{t=1}^{n-1} \xi_t\xi_{t+1}$. Then S_n is an unbiased and consistent estimator of Q_1, and $\sqrt{n}(S_n - Q_1)$ is $\mathsf{AN}(0, V_{11})$, where $V_{11} = Q_1 + 2Q_2 + 2Q_3 - 5Q_1^2$.*
3. *Let*
$$R_n = \frac{S_n - T_n^2}{T_n(1 - T_n)}.$$
Then R_n is a consistent estimator of r_1, and $\sqrt{n}(R_n - r_1)$ is $\mathsf{AN}(0, w)$, where
$$\begin{aligned} w = \big[&p^5(1 + 4Q_1) - pQ_1(3Q_1 + 8Q_1^2 + 4Q_2) + p^4(5Q_1 - Q_1^2 + 6Q_2 + 2Q_3) \\ &- 2p^3(2Q_1 + 7Q_1^2 + 4Q_2 + 4Q_1Q_2 + 2Q_3) - 3p^6 + 2Q_1^3 \\ &+ p^2(Q_1 + 12Q_1^2 + 8Q_1^3 + 12Q_1Q_2 + 2Q_2 + 2Q_3)\big]\,[p^4(1-p)^4]^{-1}. \end{aligned}$$

Proof. The statements 1. and 2. follow immediately from the central limit theorem for strictly stationary m-dependent sequences, see (Brockwell and Davis, 1991, p. 213). Using the same theorem, it is easy to derive the asymptotic distribution of $\lambda_1 T_n + \lambda_2 S_{n+1}$ for $\lambda_1, \lambda_2 \in \mathbb{R}$, and consequently the joint asymptotic distribution of $(T_n, S_{n+1})'$. The joint asymptotic distribution of $(T_n, S_n)'$ then follows from Cramér-Slutsky theorem. We obtain that

$$\sqrt{n}\left[\begin{pmatrix} T_n \\ S_n \end{pmatrix} - \begin{pmatrix} p \\ Q_1 \end{pmatrix}\right] \quad \text{is} \quad \mathsf{AN}\left(\mathbf{0}, \begin{pmatrix} U & U_{12} \\ U_{12} & V_{11} \end{pmatrix}\right),$$

where $U_{12} = 2Q_1 + 2Q_2 - 4pQ_1$. The statement 3. follows consequently from the multivariate delta method, see (Brockwell and Davis, 1991, p. 211). □

Let $\{\xi_t,\, t \in \mathbb{Z}\}$ be defined by (3.10). The success probability $p = p(a,c)$ is given in (3.12). The probability Q_1 can be written as $Q_1 = Q_1(a,c) = p(a,c) - Q(a,c)$, where $Q(a,c)$ is defined in (3.11). The asymptotic variance of the sample mean $\bar{\xi}_n$ is equal to

$$\begin{aligned} U &= 3p(a,c)[1-p(a,c)] - 2Q(a,c) \\ &= 3\mathsf{E}F(aX+c)\mathsf{E}F\Big(\frac{X-c}{a}\Big) - 2\mathsf{E}\Big[F\Big(\frac{X-c}{a}\Big)F(c+aX)\Big]. \end{aligned}$$

Proposition 3.16 implies that the asymptotic variance U tends to 0 as c approaches $\pm\infty$ if $a > 0$ is fixed. If $c \in \mathbb{R}$ is fixed then $U \to F(0)[1-F(0)]$ as $a \to \infty$, and $U \to F(c)[1-F(c)]$ as $a \to 0^+$. Finally, we have $U(1,0) = 1/12$.

The probabilities Q_i from Proposition 3.21 can be derived in a similar way as we have derived p and Q in Proposition 3.14. We get

$$\begin{aligned} Q_2 =& Q_1 - \mathsf{E}\Big[F\Big(\frac{X_1-c}{a}\Big)F(c+aX_2)\mathrm{I}[X_2 < c+aX_1]\Big], \\ Q_3 =& Q_2 + \mathsf{E}\Big[F\Big(\frac{X_1-c}{a}\Big)F\Big(\frac{X_2-c}{a}\Big)F(c+aX_2)\mathrm{I}[X_2 < c+ca+a^2X_1]\Big] \\ &- \mathsf{E}\Big[F\Big(\frac{X_1-c}{a}\Big)F(c+aX_1)F(c+aX_2)\mathrm{I}[X_2 < c+ca+a^2X_1]\Big]. \end{aligned} \tag{3.18}$$

Hence, we are able to express the asymptotic variance w of the estimator R_n from Proposition 3.21. Set $\gamma = Q_1 - Q_2$, and $\delta = Q_3 - Q_2$. Then w satisfies

$$\begin{aligned} w =& \big(-2Q^3 - Q^2p + 8Q^3p - 4Q\gamma p + 3Qp^2 - 8Q^3p^2 + 12Q\gamma p^2 + 2\delta p^2 - 8Qp^3 \\ &+ 2Q^2p^3 - 8Q\gamma p^3 - 4\delta p^3 + 7Qp^4 - Q^2p^4 + 2\delta p^4 - 2Qp^5\big)[p(1-p)]^{-4}. \end{aligned} \tag{3.19}$$

The asymptotic variance w of the estimator R_n depends not only on the quantities p and Q, but on the higher order probabilities Q_2 and Q_3 as well. This means that computation of w requires to calculate several multiple integrals and then to evaluate the formula (3.19). This is illustrated in Example 3.22.

Example 3.22. Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid variables with the exponential distribution with the density $f(x) = \mathrm{e}^{-x}\mathrm{I}[x > 0]$. Consider model (3.10) with $c \geq 0$. It follows from Proposition 3.18 that $\sqrt{n}(\bar{\xi}_n - p)$ is $\mathsf{AN}(0, U)$, where

$$U = -\frac{3\mathrm{e}^{-2c}}{(1+a)^2} + \frac{\mathrm{e}^{-c}}{1+a} + \frac{2\mathrm{e}^{-(2+a)c}}{(1+a)(1+a+a^2)}. \tag{3.20}$$

The symmetry (3.13) can be used in order to obtain the results for $c < 0$. We get

$$\gamma = \frac{\mathrm{e}^{-(3+2a+a^2)c}\Big[1 + (1+a^2)\mathrm{e}^{(1+a+a^2)c}\big[-1 - a + (1+a+a^2)\mathrm{e}^{ac}\big(\mathrm{e}^c + a\mathrm{e}^c - 1\big)\big]\Big]}{(1+a)^2(1+a^2)(1+a+a^2)},$$

$$\begin{aligned}\delta = \Big[&a^5(10 - 11\mathrm{e}^c + 3\mathrm{e}^{-ac} - 2\mathrm{e}^{-(1+a)c}) + 2 - \mathrm{e}^c + \mathrm{e}^{-ac} - 2\mathrm{e}^{-(1+a)c} - \mathrm{e}^{-(1+a)^2c} \\ &+ a^4(12 - 11\mathrm{e}^c + 4\mathrm{e}^{-ac} - 4\mathrm{e}^{-(1+a)c} - \mathrm{e}^{-(1+a)^2c}) + \mathrm{e}^{-(1+a)(2+a+a^2)c} \\ &+ a^3(10 - 9\mathrm{e}^c + 4\mathrm{e}^{-ac} - 4\mathrm{e}^{-(1+a)c} - \mathrm{e}^{-(1+a)^2c}) + a^7(4 - 6\mathrm{e}^c + \mathrm{e}^{-ac}) \\ &+ a^2(8 - 6\mathrm{e}^c + 3\mathrm{e}^{-ac} - 4\mathrm{e}^{-(1+a)c} - \mathrm{e}^{-(1+a)^2c}) - a^9\mathrm{e}^c + a^8(2 - 3\mathrm{e}^c) \\ &+ a(4 - 3\mathrm{e}^c + 2\mathrm{e}^{-ac} - 2\mathrm{e}^{-(1+a)c} - \mathrm{e}^{-(1+a)^2c}) + a^6(8 - 9\mathrm{e}^c + 2\mathrm{e}^{-ac} - 2\mathrm{e}^{-(1+a)c})\Big] \\ &\times \mathrm{e}^{-2c}[(1+a)^2(1+a^2)(1+a+a^2)(1+a+a^2+a^3+a^4)]^{-1}.\end{aligned}$$

The results for $c < 0$ can be obtained from equalities

$$\begin{aligned}\gamma(a,c) &= p\left(\frac{1}{a}, -\frac{c}{a}\right)\left[1 - p\left(\frac{1}{a}, -\frac{c}{a}\right)\right] - \gamma\left(\frac{1}{a}, -\frac{c}{a}\right), \\ \delta(a,c) &= \delta\left(\frac{1}{a}, -\frac{c}{a}\right) + \left[2p\left(\frac{1}{a}, -\frac{c}{a}\right) - 1\right]Q\left(\frac{1}{a}, -\frac{c}{a}\right),\end{aligned}$$

which are derived from (3.18). If $a > 0$ and $c \in \mathbb{R}$ are given then the asymptotic variance w of the estimator R_n is obtained after evaluating the formula (3.19) for the particular p, Q, γ, δ. Note that a higher accuracy must be set when computing $w(a,c)$ for large values of c because of the term $[p(1-p)]^{-4}$ in (3.19). The graph of w as a function of a and c is plotted in Figure 3.6.

For the special case $a = 1$ we obtain

$$\begin{aligned}w(1,c) =& \big(9 - 36\mathrm{e}^c - 24\mathrm{e}^{2c} + 260\mathrm{e}^{3c} - 160\mathrm{e}^{4c} - 700\mathrm{e}^{5c} + 1080\mathrm{e}^{6c} + 90\mathrm{e}^{7c} \\ &- 1125\mathrm{e}^{8c} + 630\mathrm{e}^{9c}\big) \times 4\mathrm{e}^{-6c}[135(1 - 2\mathrm{e}^c)^4]^{-1}.\end{aligned} \tag{3.21}$$

The function $w(1,\cdot)$ has a global maximum equal to $32/45$ at $c = 0$, and $w(1,c)$ tends to 0 as c approaches $\pm\infty$.

Recall that the inequality $r_1 \geq -1/3$ always holds in the model (3.10), see Remark 3.4. However, the estimator R_n can lead to values below $-1/3$, and this might be a disadvantage in some applications.

Under some circumstances we are able to construct a simpler estimator of the autocorrelation r_1. Let us present one such situation. Suppose that $a > 0$ is known. It is stated in Proposition 3.16 that the function $p_a(c) = p(a,c)$ is continuous and increasing on (C_1, C_2), and $p_a\big((C_1, C_2)\big) = (0,1)$. Hence, there exists a continuous and increasing inverse $p_a^{-1} : (0,1) \to (C_1, C_2)$. We can then write $r_1 = r_1^*(p) = r_1\left(a, p_a^{-1}(p)\right)$, i.e. r_1 can be expressed as a function of p. An estimator of r_1 can be obtained after applying r_1^* to $\bar{\xi}_n$. Properties of such estimator are summarized in the following Proposition 3.23.

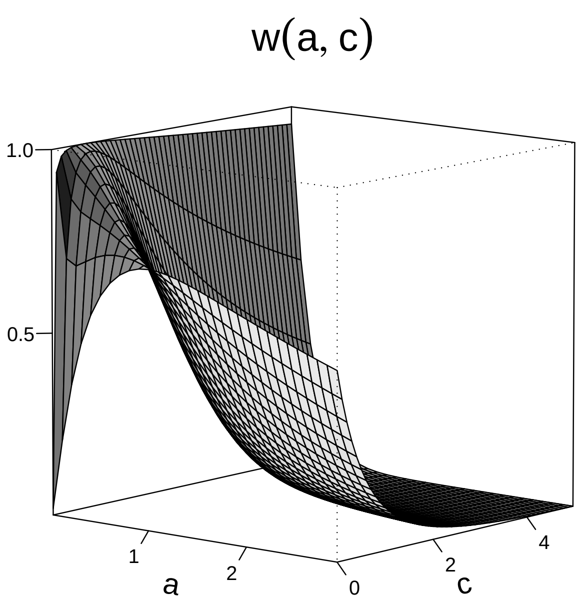


Figure 3.6: The asymptotic variance w for the exponential distribution

Proposition 3.23. *Let $a > 0$ be known. Consider the setting from Proposition 3.14. Define $\hat{r}_1 = r_1^*\left(\bar{\xi}_n\right)$. Then $\hat{r}_1$ is a consistent estimator of r_1.*

Assume further that the functions p_a and r_1 are differentiable with respect to c at all $c \in (C_1, C_2)$, and $\partial p_a(c)/\partial c \neq 0$. Then $\sqrt{n}(\hat{r}_1 - r_1)$ is $\mathsf{AN}(0, v)$, where

$$v = p(1-p)[2r_1^*(p) + 1][\partial r_1^*(p)/\partial p]^2.$$

The asymptotic variance v can be equivalently expressed as a function of c as

$$v = p_a(c)[1 - p_a(c)][2r_1(a, c) + 1]\left(\frac{\partial r_1(a, c)/\partial c}{\partial p_a(c)/\partial c}\right)^2.$$

Moreover, if $\partial r_1^*(p)/\partial p = 0$, and $\partial^2 r_1^*(p)/\partial p^2$ exists and is non-zero then $n(\hat{r}_1 - r_1)$ converges in law to the distribution of a random variable $(1/2) \cdot p(1-p)[2r_1^*(p) + 1][\partial^2 r_1^*(p)/\partial p^2] \cdot \chi_1^2$, where χ_1^2 follows the χ^2 distribution with 1 degree of freedom.

Proof. It follows from Proposition 3.16 that the function r_1^* is always continuous on $(0, 1)$, and the continuity implies the consistency of $\hat{r}_1$. Furthermore, r_1^* is differentiable on $(0, 1)$ under the additional assumptions stated in the proposition. The asymptotic distribution of $\sqrt{n}(\hat{r}_1 - r_1)$ follows from the classical delta method, see (Lehmann, 1998, p. 58–59)Lehmann, E. L. and from the asymptotic distribution of $\bar{\xi}_n$ stated in Proposition 3.21. $\square$

Note that the asymptotic variance v of the estimator $\hat{r}_1$ from Proposition 3.23 does not require computation of the higher order probabilities Q_2, Q_3. Furthermore, the inequality $r_1^*(p) \geq -1/3$ implies that $\hat{r}_1 \geq -1/3$ always holds.

Remark that a problem can appear in situations where $p_a(c)$ is very close to 0 or 1 and $(C_1, C_2) = (-\infty, \infty)$. In practice, we can then obtain $\bar{\xi}_n = 0$ or $\bar{\xi}_n = 1$ for small n. In such a case $p_a^{-1}(\bar{\xi}_n)$ is not defined, and consequently $\hat{r}_1$ cannot be defined as well. However, it follows from the consistency of $\bar{\xi}_n$ that this problem does not occur for sufficiently large n.

The estimator $\hat{r}_1$ in Proposition 3.23 is derived under the assumption that the parameter $a > 0$ is known. Similarly, we could consider the case where $c \in \mathbb{R}$ is known and $a > 0$ is unknown. We show in Section 4 that under some additional assumptions a can be expressed as a function of p, i.e. $a = (p^c)^{-1}(p)$, and therefore r_1 can be written as a function of p, $r_1 = r_1^{**}(p) = r_1\big((p^c)^{-1}(p), c\big)$. An assertion similar to Proposition 3.23 could be formulated about the properties of the corresponding estimator $\tilde{r}_1 = r_1^{**}(\bar{\xi}_n)$.

We illustrate the approach from Proposition 3.23 on two examples.

Example 3.24. Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid random variables with the exponential distribution with the density $f(x) = \mathrm{e}^{-x}\mathrm{I}[x \geq 0]$. Consider model (3.10) with $a = 1$. The relationship between p and c is given in (3.14). It follows from Proposition 3.18 that

$$r_1 = r_1^*(p) = \begin{cases} \dfrac{(1-4p)(1-p)}{3p}, & p \geq 1/2, \\ \dfrac{(4p-3)p}{3(1-p)}, & p < 1/2. \end{cases} \tag{3.22}$$

The formula (3.20) implies that $\sqrt{n}(\bar{\xi}_n - p)$ is asymptotically normal $\mathsf{AN}(0, U)$, where

$$U = \frac{\mathrm{e}^{-3|c|}(4 - 9\mathrm{e}^{|c|} + 6\mathrm{e}^{2|c|})}{12} = \begin{cases} \dfrac{(1-p)(2-7p+8p^2)}{3}, & p \geq 1/2, \\ p - 3p^2 + (8p^3)/3, & p < 1/2. \end{cases}$$

Note that U tends to 0 as p tends to 0 or 1 (i.e. c tends to $-\infty$ and ∞ respectively). Furthermore, U has two local maxima equal to $5/48$ at $p = 3/4$ and $p = 1/4$ (corresponding to $c = \pm \log 2$) and one local minimum equal to $1/12$ at $p = 1/2$ ($c = 0$).

The formula (3.22) suggests to estimate r_1 as

$$\hat{r}_1 = r_1^*(\bar{\xi}_n) = \begin{cases} \dfrac{(1-4\bar{\xi}_n)(1-\bar{\xi}_n)}{3\bar{\xi}_n} & \text{if } \bar{\xi}_n \geq 1/2, \\ \dfrac{(4\bar{\xi}_n-3)\bar{\xi}_n}{3(1-\bar{\xi}_n)} & \text{if } \bar{\xi}_n < 1/2. \end{cases}$$

The function $r_1^*(p)$ is continuously differentiable on $(0, 1)$. It follows from Proposition 3.23 that $\hat{r}_1$ is consistent, and $\sqrt{n}(\hat{r}_1 - r_1)$ is $\mathsf{AN}(0, v)$, where

$$v = \begin{cases} \dfrac{(1-p)(1-4p^2)^2(2-7p+8p^2)}{27p^4}, & p \geq 1/2, \\ \dfrac{p(3-8p+4p^2)^2(3-9p+8p^2)}{27(1-p)^4}, & p < 1/2. \end{cases}$$

Equivalently, $v = \left[4\mathrm{e}^{-3|c|}(1-4\mathrm{e}^{|c|}+3\mathrm{e}^{2|c|})^2(4-9\mathrm{e}^{|c|}+6\mathrm{e}^{2|c|})\right] \cdot \left[27(1-2\mathrm{e}^{|c|})^4\right]^{-1}$. The asymptotic variance v tends to 0 as p tends to 0 or 1. In addition, v has a local minimum equal to 0 at $p = 1/2$. This means that if $p = 1/2$ (i.e. $c = 0$) then $r_1 = -1/3$ and $\sqrt{n}(\hat{r}_1 + 1/3)$ tends to 0 in probability as $n \to \infty$. Moreover, it follows from Proposition 3.23 that $n(\hat{r}_1 + 1/3)$ converges in law to the distribution of a random variable $2/9 \cdot \chi_1^2$, where χ_1^2 follows the χ^2 distribution with 1 degree of freedom.

We can compare the asymptotic variance w of the estimator R_n given in (3.21) with the asymptotic variance v of the estimator $\hat{r}_1$. It follows that $v(c) < w(c)$ for all $c \in \mathbb{R}$, and therefore the estimator $\hat{r}_1$ might be preferable.

Example 3.25. Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid random variables with the uniform distribution on the interval $[0, 1]$. Consider model (3.10) with $a = 1$. It follows from Proposition 3.19 that

$$r_1 = r_1^*(p) = \begin{cases} 1 - \dfrac{1}{1-p}, & 0 < p \leq 1/8, \\ -\dfrac{1 - 6\sqrt{2p} + 24\,p - 16p\sqrt{2p} + 6p^2}{6(1-p)p}, & 1/8 < p \leq 1/2, \\ -\dfrac{31 - 22\sqrt{2-2p} + 4(-9+4\sqrt{2-2p})p + 6p^2}{6(1-p)p}, & 1/2 < p < 7/8, \\ 1 - \dfrac{1}{p}, & 7/8 \leq p < 1. \end{cases} \tag{3.23}$$

The standardized sample mean $\sqrt{n}(\bar{\xi}_n - p)$ is asymptotically normal $\mathsf{AN}(0, U)$, where

$$U = \begin{cases} p - 3p^2, & 0 < p \leq 1/8, \\ -1/3 + 2\sqrt{2p} - 7p + (16/3)p\sqrt{2p} - 3p^2, & 1/8 < p \leq 1/2, \\ [-31 + 22\sqrt{2-2p} + (39 - 16\sqrt{2-2p})p - 9p^2]/3, & 1/2 < p \leq 7/8, \\ -2 + 5p - 3p^2, & 7/8 < p < 1. \end{cases}$$

The asymptotic variance U tends to 0 as p approaches 0 or 1. As a function of p, U has two local maxima equal to 7/81 at $p = 7/9$ and $p = 2/9$, and one local minimum 1/12 at $p = 1/2$.

The formula (3.23) suggests to construct an estimator of r_1 given as $\hat{r}_1 = r_1^*(\bar{\xi}_n)$. Since r_1^* is continuously differentiable on $(0, 1)$, it follows from Proposition 3.23 that $\hat{r}_1$

is consistent, and $\sqrt{n}(\hat{r}_1 - r_1)$ is $\mathsf{AN}(0, v)$, where

$$v = \begin{cases} \dfrac{p - 3p^2}{(1-p)^4}, & 0 < p \leq 1/8, \\ -(-1 + 3\sqrt{2p} + 2p - 17p\sqrt{2p} + 30p^2 - 8p^2\sqrt{2p})^2 & \\ \quad \times (1 - 6\sqrt{2p} + 21p - 16p\sqrt{2p} + 9p^2)[108(1-p)^4 p^4]^{-1}, & 1/8 < p \leq 1/2. \end{cases}$$

The value of v for $p > 1/2$ is obtained using the symmetry $v(p) = v(1-p)$. The asymptotic variance v tends to 0 as p approaches 0 or 1. Furthermore, v has two local maxima and one local minimum equal to 0 at $p = 1/2$. Hence, if $p = 1/2$ then $r_1 = 1/3$ and $\sqrt{n}(\hat{r}_1 + 1/3)$ tends to 0 in probability as $n \to \infty$. In addition, $n(\hat{r}_1 + 1/3)$ converges in law to the distribution of a random variable $1/18 \cdot \chi_1^2$, where χ_1^2 follows the χ^2 distribution with 1 degree of freedom.

3.2.2 Estimation of a and c

Suppose that we would like to estimate the parameters a and c in model (3.10) from the observations $\xi_1, \dots, \xi_n$. Let us first remark that it is not possible to estimate both parameters a, c only from the statistics T_n and S_n defined in Proposition 3.21, see Example 3.26. In the following we consider only the situations where one of the parameters a, c is known and the other one is estimated.

Example 3.26. Consider the situation from Proposition 3.18 with $\lambda = 1$. There exist $a_1 \neq a_2$, $c_1 \neq c_2$ such that $1/3 = p(a_1, c_1) = p(a_2, c_2)$ and $0.285 = Q(a_1, c_1) = Q(a_2, c_2)$. The numerical computation gives $a_1 \doteq 0.48735$, $a_2 \doteq 0.94679$, $c_1 \doteq 0.00847$, $c_2 \doteq -0.35765$. This shows that different values of (a, c) lead to the same values of p and Q, i.e. values of p and Q_1. This demonstrates that the parameters a, c cannot be estimated from T_n and S_n only.

Let us start with the estimation of the parameter c. Assume that $a > 0$ is a known constant. Suppose that iid random variables $\{X_t,\, t \in \mathbb{Z}\}$ with an absolutely continuous distribution F are generated, and we observe whether the increment $X_t - aX_{t-1}$ exceeds some unknown threshold $c \in \mathbb{R}$. This information is expressed by 0-1 valued variables $\xi_1, \dots, \xi_n$ defined in model (3.10). We would like to estimate the parameter $c \in \mathbb{R}$ from the observations $\xi_1, \dots, \xi_n$.

The estimator $\hat{c}$ of the parameter c can be constructed in an analogous way as the estimator $\hat{r}_1$ in Proposition 3.23. We obtain the following statement.

Proposition 3.27. *Define $\hat{c} = p_a^{-1}(\bar{\xi}_n)$. Then $\hat{c}$ is a consistent estimator of the parameter c. Moreover, if p_a^{-1} is differentiable on $(0,1)$ then $\sqrt{n}(\hat{c} - c)$ is $\mathsf{AN}(0, V_c)$, where*

$$V_c = \frac{p_a(c)[1 - p_a(c)][2r_1(a,c) + 1]}{[\partial p_a(c)/\partial c]^2}.$$

Note that $\hat{c}$ is not defined if $\bar{\xi}_n = 0$ or $\bar{\xi}_n = 1$. This situation might occur in practice if $p_a(c)$ is close to 0 or 1 and the sample size n is small. However, it follows from consistency of $\bar{\xi}_n$ that the probability that $\hat{c}$ is defined tends to 1 as $n \to \infty$.

The properties of the estimator $\hat{c}$ are illustrated on two examples.

Example 3.28. Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid random variables with the exponential distribution with the density $f(x) = \mathrm{e}^{-x}\mathrm{I}[x \geq 0]$. Consider model (3.10). It follows from (3.14) that

$$c = p_a^{-1}(p) = \begin{cases} -\log\left[(a+1)(1-p)\right], & p > a/(a+1), \\ a\log\left[\frac{a+1}{a}p\right], & p < a/(a+1). \end{cases}$$

The function p_a^{-1} is continuously differentiable on $(0,1)$ for any $a > 0$. The estimator $\hat{c}$ is therefore consistent and asymptotically normal. The asymptotic variance V_c is given as $V_c = U(a,c)\left[(a+1)\mathrm{e}^{c}\right]^2$ if $c \geq 0$, and $V_c = U(a,c)\left[(a+1)\mathrm{e}^{-c/a}\right]^2$ if $c < 0$, where $U(a,c)$ is given in (3.20). For $a = 1$ we obtain $V_c = -3 + 2\mathrm{e}^{-|c|} + 4\mathrm{e}^{|c|}/3$. Then V_c tends to ∞ as c approaches $\pm\infty$, and V_c has a global minimum equal to $1/3$ at $c = 0$. This demonstrates that large values of c, which correspond to p closed to 1, are difficult to be estimated.

Example 3.29. Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid random variables with the uniform distribution on the interval $[0,1]$. Consider model (3.10) with $a = 1$. It follows from (3.16) that

$$p = p_1(c) = \mathsf{P}(\xi_t = 1) = \begin{cases} \frac{1}{2}(1 + 2c - c^2) & \text{if} \quad 0 \leq c \leq 1, \\ \frac{1}{2}(1 + 2c + c^2) & \text{if} \quad -1 \leq c < 0. \end{cases}$$

The estimator $\hat{c}$ is then given as

$$\hat{c}_n = p_1^{-1}(\bar{\xi}_n) = \begin{cases} \sqrt{2\bar{\xi}_n} - 1 & \text{if } \bar{\xi}_n \leq 1/2, \\ 1 - \sqrt{2(1-\bar{\xi}_n)}, & \text{if } \bar{\xi}_n > 1/2. \end{cases}$$

Since the function p_1^{-1} is continuously differentiable on $(0,1)$, the estimator $\hat{c}$ is consistent, asymptotically normal, and

$$V_c = V_c(c) = \begin{cases} \dfrac{1 + (4 - 9|c|)|c|^3}{12(1-|c|)^2}, & |c| \leq 1/2, \\ \dfrac{-1 + 6|c| - 3|c|^2}{4}, & c \in (-1,-1/2) \cup (1/2,1). \end{cases}$$

The function V_c has a global minimum equal to $1/12$ at $c = 0$, and V_c tends to $1/2$ as c approaches ± 1.

Consider now the problem of estimation of the parameter a in model (3.10) with a known $c \in \mathbb{R}$. Suppose that iid random variables $\{X_t,\, t \in \mathbb{Z}\}$ with an absolutely continuous distribution F are generated, and we observe whether the MA(1) process X_t-

aX_{t-1} exceeds some known threshold $c \in \mathbb{R}$ (for instance $c = 0$) or not. This information is expressed by the variables $\xi_1, \ldots, \xi_n$ defined by (3.10). We want to estimate the parameter $a > 0$ from the measurements $\xi_1, \ldots, \xi_n$.

We would like to construct an estimator $\hat{a}$ of the parameter a from the sample mean $\bar{\xi}_n$ in a similar way as we have constructed $\hat{c}$ in the previous paragraph. However, the function $p(\cdot, c)$ is not monotonous in general, see Example 3.30.

Example 3.30. Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid random variables with a distribution with a density $f(x) = \mathrm{e}^{-3x/2}\mathrm{I}[x \geq 0] + \mathrm{e}^{3x}\mathrm{I}[x < 0]$. It follows from (3.12) that

$$p(a, 1/2) = 1 - \frac{a^2\mathrm{e}^{-3/(2a)} - 4\mathrm{e}^{-3/4}}{6 + 3a - 3a^2}.$$

The function $p(a, 1/2)$ is increasing on $(0, a_0)$ and decreasing on (a_0, ∞), where $a_0 \doteq 0.7239$. Hence, $p(a, 1/2)$ is not monotonous on $(0, \infty)$.

Nevertheless, if the distribution F satisfies some additional assumptions (e.g. if $F' \equiv 0$ on $(-\infty, 0)$, see Proposition 3.16) then the monotonicity of $p(\cdot, c)$ follows. In the following we consider only such cases.

Proposition 3.31. *For a given $c \in \mathbb{R}$ define $A_c = \{a > 0 : \ p(a, c) \in (0, 1)\}$. Let the function $p^c(\cdot) = p(\cdot, c)$ be increasing on A_c. If $\bar{\xi}_n \in (F(c), 1 - F(0))$ define $\hat{a} = (p^c)^{-1}(\bar{\xi}_n)$. Then $\hat{a}$ is defined with probability that tends to 1 as the sample size n increases to ∞, and $\hat{a}$ is a consistent estimator of the parameter a. Moreover, if $(p^c)^{-1}$ is differentiable on $(F(c), 1 - F(0))$ then $\sqrt{n}(\hat{a} - a)$ is $\mathsf{AN}(0, V_a)$, where*

$$V_a = \frac{p^c(a)[1 - p^c(a)][2r_1(a, c) + 1]}{[\partial p^c(a)/\partial a]^2}.$$

Proof. If p^c is increasing on A_c then it follows from Proposition 3.16 that the image of A_c is equal to the interval $(F(c), 1 - F(0))$. The consistency of $\bar{\xi}_n$ implies that the probability that $\hat{a}$ is defined tends to 1 as $n \to \infty$, and $\hat{a}$ is consistent. The remaining part of the assertion follows from the delta method, see (Lehmann, 1998, p. 58–59). □

Example 3.32. Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid random variables with the uniform distribution on the interval $[0, 1]$. Consider model (3.10) with $c = 0$. It follows from (3.16) that

$$p^0(a) = \begin{cases} \dfrac{a}{2}, & a \leq 1, \\ 1 - \dfrac{1}{2a}, & a > 1. \end{cases}$$

The function p^0 is increasing on $(0, \infty)$, and its range is $(F(0), 1 - F(0)) = (0, 1)$. The estimator $\hat{a}$ is defined as

$$\hat{a} = (p^0)^{-1}(\bar{\xi}_n) = \begin{cases} 2\bar{\xi}_n, & \text{if } \bar{\xi}_n \leq 1/2, \\ \dfrac{1}{2(1 - \bar{\xi}_n)}, & \text{if } \bar{\xi}_n > 1/2. \end{cases}$$

The function $(p^0)^{-1}$ is continuously differentiable on $(0, 1)$. Hence, the estimator $\hat{a}$ is consistent, asymptotically normal, and

$$V_a = V_a(a) = \begin{cases} \dfrac{a(6 - 9a + 4a^2)}{3}, & a < 1 \\ \dfrac{a(4 - 9a + 6a^2)}{3}, & a > 1. \end{cases}$$

The asymptotic variance V_a tends to 0 as a approaches 0, and V_a tends to ∞ as $a \to \infty$. The function $V_a(a)$ has a local maximum equal to $5/12$ at $a = 1/2$ and a local minimum equal to $1/3$ at $a = 1$.

Example 3.33. Let $\{X_t, t \in \mathbb{Z}\}$ be iid random variables with the exponential distribution with the density $f(x) = \mathrm{e}^{-x}\mathrm{I}[x \geq 0]$. Let $c \in \mathbb{R}$ be a known given constant. If $c \leq 0$ then (3.14) implies $p^c(a) = a\mathrm{e}^{c/a}/(1+a)$, and the range of p^c is equal to the whole interval $(0, 1)$. The estimator $\hat{a}$ is a solution of the equation

$$\frac{1}{1+a} = 1 - \bar{\xi}_n \cdot \mathrm{e}^{c/a},$$

and can be found numerically.

If $c > 0$ then (3.14) implies $p^c(a) = 1 - \mathrm{e}^{-c}/(1 + a)$, and the range of p^c is equal to the interval $(F(c), 1) = (1 - \mathrm{e}^{-c}, 1)$. If $\bar{\xi}_n \in (1 - \mathrm{e}^{-c}, 1)$ then the estimator $\hat{a}$ is defined as

$$\hat{a} = \frac{\mathrm{e}^{-c}}{1 - p} - 1.$$

The probability that $\hat{a}$ is defined, i.e. $\bar{\xi}_n \in (1 - \mathrm{e}^{-c}, 1)$, increases with an increasing sample size n.

It is clear that the estimator $\hat{a}$ is consistent for any choice of $c \in \mathbb{R}$. Furthermore, $\hat{a}$ is asymptotically normal with the asymptotic variance $V_a = U(a, c) \times 1/[\partial p/\partial a]^2$, where $U(a, c)$ is given in (3.20).

Let us note that some other characteristics can be estimated in model (3.10) as well. For instance, in some situations the parameters a, c from (3.10) are known, but we would like to estimate some parameters of the distribution F of the variables $\{X_t, t \in \mathbb{Z}\}$. Such a situation is illustrated in the following example.

Example 3.34. Consider the setting of Proposition 3.18. Suppose that we would like to estimate the parameter $\lambda > 0$ of the exponential distribution, but we are not able to measure the variables $\{X_t, t \in \mathbb{Z}\}$ directly. However, we can observe whether the increment $X_t - aX_{t-1}$ exceeds some known level $c \in \mathbb{R}$. This information is expressed by the measurements $\xi_1, \ldots, \xi_n$ defined by (3.10). We assume that the parameter $a > 0$ is known as well. The formula (3.14) suggests to define an estimator $\hat{\lambda}$ as

$$\hat{\lambda} = l_{ac}(\bar{\xi}_n) = \begin{cases} -\dfrac{\log[(a+1)(1 - \bar{\xi}_n)]}{c}, & c \geq 0, \\ \dfrac{a \log[(a+1)\bar{\xi}_n/a]}{c}, & c < 0. \end{cases}$$

The function l_{ac} is continuously differentiable on $(0, 1)$ for any choice of $a > 0$, $c \in \mathbb{R}$. Hence, the estimator $\hat{\lambda}$ is consistent and asymptotically normal. In particular, $\sqrt{n}(\hat{\lambda}-\lambda)$ is $\mathsf{AN}(0, V_\lambda)$, where

$$V_\lambda = \begin{cases} \dfrac{p(a,c)}{1-p(a,c)} \dfrac{2r_1(a,c)+1}{c^2}, & c \geq 0, \\ \dfrac{1-p(a,c)}{p(a,c)} \dfrac{[2r_1(a,c)+1]a^2}{c^2}, & c < 0, \end{cases}$$

and $p(a, c)$ and $r_1(a, c)$ are given in (3.14) and (3.15) respectively.

Remark that it follows from (3.15) that it is not possible to estimate λ from $\xi_1, \dots, \xi_n$ if the threshold level c is not known.

3.2.3 Numerical Study

A numerical study was performed in order to illustrate the theoretical results presented in the previous text. We considered the problem of estimation of p, r_1, a, and c in model (3.10) respectively. For given $a > 0$, $c \in \mathbb{R}$, and the distribution F, we generated variables $\xi_1, \dots, \xi_n$ by (3.10) and computed the value of the estimator $\bar{\xi}_n$ (or $\hat{r}_1$, $\hat{c}$, $\hat{a}$ respectively). For each sample size n this procedure was repeated 1000 times. From these $N = 1000$ different realizations of the estimator $\bar{\xi}_n$ we computed the average (denoted as $\mathsf{aver}\,\bar{\xi}_n$) and the sample variance (denoted as $\widehat{\mathsf{var}}\,\bar{\xi}_n$). The same procedure was used for $\hat{r}_1$, $\hat{c}$, and $\hat{a}$.

A comparison of $\mathsf{aver}\,\bar{\xi}_n$ with the theoretical value p demonstrates the consistency of the estimator $\bar{\xi}_n$. Results for different values of n illustrate the corresponding rate of the convergence. A comparison of $n\,\widehat{\mathsf{var}}\,\bar{\xi}_n$ with the theoretical value U confirms our formula for the asymptotic variance of the estimator $\bar{\xi}_n$. Furthermore, we compared the sample quantiles of $\sqrt{n}(\bar{\xi}_n - p)$ and their theoretical counterparts in order to illustrate the asymptotic normality of $\bar{\xi}_n$. The sample skewness and kurtosis of $\bar{\xi}_n$ were calculated as well.

All the simulations were performed in the program `R`, see R Development Core Team (2009), with the setting `set.seed(1234)` (the initial state of the random number generator). The sample sizes were chosen to be $n = 100, 500, 1000, 5000, 10000$ respectively. We considered the following situations:

1. estimation of p for the exponential distribution and $a = 1$,
2. estimation of r_1 for the exponential distribution and $a = 1$,
3. estimation of c for the exponential distribution and various choices of $a > 0$,
4. estimation of p for the uniform distribution and $a = 1$,
5. estimation of r_1 for the uniform distribution and $a = 1$,
6. estimation of c for the uniform distribution and $a = 1$,
7. estimation of a for the uniform distribution and $c = 0$.

The parameter λ of the exponential distribution was set to be $\lambda = 1$, the uniform distribution was considered on the interval $[0, 1]$. Simulations for each of the situations were performed for several choices of c (or a respectively). The obtained results are summarized in the following paragraphs. Some particular settings are shown in Tables 3.4–3.8. The results for other settings are available upon request.

The simulations for the situations 1. and 4. show that the properties of the estimator $\bar{\xi}_n$ are good regardless the value of the parameters a and c. The convergence to the true success probability p is very fast, and the asymptotic normal approximation is good even for small sample sizes n ($n = 100$) if p is not "extreme". If the true value p is very close to 0 or 1 then the sample size n must be larger ($n \geq 500$) in order to obtain reasonable results. Table 3.4 provides an illustration of these conclusions for one particular setting.

n	p	$\mathsf{aver}\,\bar{\xi}_n$	U	$n\widehat{\mathsf{var}}\,\bar{\xi}_n$	skewness	kurtosis
100	0.8161	0.8146	0.0990	0.1018	-0.1172	0.0073
500	0.8161	0.8161	0.0990	0.1111	0.0510	0.1888
1000	0.8161	0.8156	0.0990	0.0983	0.0577	0.1508
5000	0.8161	0.8158	0.0990	0.0992	-0.1234	-0.2146
10000	0.8161	0.8160	0.0990	0.1030	0.0171	-0.0955

Table 3.4: Estimation of p for the exponential distribution, $c = 1$, $a = 1$

The behavior of the estimator $\hat{r}_1$ is more sensitive to the values of the parameters a and c. The simulations indicate that the convergence to the true correlation r_1 is quite fast because $\mathsf{aver}\,\hat{r}_1$ is close to r_1 even for small sample sizes ($n = 100$). The normal approximation is good if the values of a and c are not "extreme". If a and c are such that the corresponding success probability p is close to 0 or 1 then the normal distribution approximates the true distribution of $\hat{r}_1$ rather poorly for small sample sizes n, and one needs to consider large samples in order to obtain reasonable results. Recall that if $a = 1$ and $c = 0$ then $n(\hat{r}_1 + 1/3)$ converges in law to the distribution of a random variable $K \cdot \chi_1^2$, where χ_1^2 follows the χ^2 distribution with 1 degree of freedom and $K > 0$ is a constant specified in Proposition 3.23. The results for the situations 2. and 5. indicate that this asymptotic approximation is not very good for small sample sizes ($n = 100$). Reasonable results are obtained for $n > 1000$. Two particular settings are presented in Tables 3.5 and 3.6.

The estimation of the parameter c is considered in the situations 3. and 6. The behavior of $\hat{c}$ seems to resemble to the one of $\hat{r}_1$ described in the previous paragraph. The consistency of $\hat{c}$ is apparent even for small sample sizes ($n = 100$). The quality of the normal approximation depends on the value of the parameters a and c. If a and c are such that the corresponding probability p is close to 0 or 1 then the normal distribution does not approximate the true distribution of $\hat{c}$ well for small and moderate sample sizes n. On the other hand, if a and c are not "extreme" (in the previous sense)

n	r	aver $\hat{r}_1$	v	$n\widehat{\text{var}}\,\hat{r}_1$	skewness	kurtosis
100	-0.0660	-0.0653	0.0494	0.0478	-0.2894	0.1259
500	-0.0660	-0.0661	0.0494	0.0512	-0.1404	-0.1958
1000	-0.0660	-0.0663	0.0494	0.0506	0.0599	-0.1165
5000	-0.0660	-0.0662	0.0494	0.0484	0.1026	-0.0302
10000	-0.0660	-0.0660	0.0494	0.0518	-0.0065	0.0076

Table 3.5: Estimation of r_1 for the exponential distribution, $c = -2$, $a = 1$ (p= 0.0677)

n	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
100	0.0017	0.0017	0.0155	0.0155	0.0155	0.0437	0.0437	0.0866	0.1450
500	0.0003	0.0030	0.0084	0.0165	0.0274	0.0410	0.0573	0.0985	0.1509
1000	0.0015	0.0042	0.0082	0.0136	0.0284	0.0379	0.0609	0.0894	0.1631
5000	0.0008	0.0027	0.0075	0.0177	0.0281	0.0410	0.0637	0.0942	0.1599
10000	0.0008	0.0028	0.0074	0.0122	0.0229	0.0369	0.0582	0.0892	0.1518
$\frac{1}{18}\chi_1^2$	0.0009	0.0036	0.0082	0.0153	0.0253	0.0394	0.0597	0.0912	0.1503

Table 3.6: Sample and theoretical quantiles of $n(\hat{r_1} - r_1)$ for the uniform distribution, $c = 0$, $a = 1$

	c	aver $\hat{c}$	V_c	$n\widehat{\text{var}}\,\hat{c}$	skewness	kurtosis
100	1.0000	1.0070	2.1172	2.3144	0.3909	0.2055
500	1.0000	1.0005	2.1172	2.3165	0.1378	0.0453
1000	1.0000	0.9990	2.1172	2.0607	0.1762	0.0795
5000	1.0000	0.9995	2.1172	2.1796	-0.0573	-0.0543
10000	1.0000	1.0001	2.1172	2.2087	0.0662	-0.2331

Table 3.7: Estimation of c for the exponential distribution, $a = 1/2$, $c = 1$ ($p = 0.7547$)

then a good approximation is obtained even for $n = 100$. An example is presented in Table 3.7.

The simulations for the situation 7. show that the convergence of $\hat{a}$ to the true value a is quite fast independently of a. On the other hand, $n\widehat{\text{var}}\,\hat{a}$ is not very close to the asymptotic variance V_a if n is small and a is rather far from 1. A similar conclusion holds for the asymptotic normality: If a is close to 1 then the approximation is reasonable even for $n = 100$. If a is far from 1 then one needs to have a large sample size in order to obtain good results. See Table 3.8 for an example.

n	a	aver $\hat{a}$	V_a	$n\widehat{\text{var}}\,\hat{a}$	skewness	kurtosis
100	3.0000	3.1051	31.0000	42.7550	1.3379	3.2400
500	3.0000	3.0087	31.0000	32.7006	0.5419	0.2840
1000	3.0000	3.0057	31.0000	31.5092	0.4757	0.3059
5000	3.0000	2.9990	31.0000	31.7466	0.0355	0.0567
10000	3.0000	3.0050	31.0000	30.4402	0.0751	0.0156

Table 3.8: Estimation of a for the uniform distribution, $a = 3$, $c = 0$

3.3 2-Dependent Variables

In this section we consider 2-dependent Bernoulli variables that are obtained by a similar clipping procedure as we have seen in the model (3.10). Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid random variables and a, b, c some real constants. Define

$$\xi_t = \begin{cases} 1 & \text{if } X_t - aX_{t-1} - bX_{t-2} < c, \\ 0 & \text{otherwise.} \end{cases} \tag{3.24}$$

The previous findings motivate us to consider the case $c = 0$ and X_t with the uniform distribution on the interval $[0, 1]$. We show in this setting we get $r_1 \geq -1/3$ as well as $r_1 + r_2 \geq -1/3$.

Theorem 3.35. *Let $\{\xi_t,\, t \in \mathbb{Z}\}$ be defined by (3.24) where X_t has the uniform distribution on $(0, 1)$ and $c = 0$. Let $a \geq 0$, $b \geq 0$, and $a + b \leq 1$, $a + b > 0$. Let $r_1(a, b)$ and $r_2(a, b)$ be the 1-lag and 2-lag autocorrelations of $\boldsymbol{\xi}$, respectively. Define $s(a, b) = r_1(a, b) + r_2(a, b)$. Then*

$$p = \mathsf{P}(\xi_t = 1) = \frac{a+b}{2},$$

$$r_1 = r_1(a, b) = -\frac{a}{3}\frac{(2a+b)(1-a-b) + b(1-b) + a}{(a+b)(2-a-b)},$$

$$r_2 = r_2(a, b) = \frac{b}{3}\frac{(a+b)(2a+2b-3) - ab}{(a+b)(2-a-b)},$$

$$s(a, b) = \frac{a^2(2a-3) + b^2(2b-3) + 5ab(a+b-1)}{,}3(a+b)(2-a-b) \tag{3.25}$$

and the inequalities

$$-\frac{1}{3} \leq r_1(a, b) \leq 0, \quad -\frac{1}{3} \leq r_2(a, b) \leq 0, \quad -\frac{1}{3} \leq s(a, b) \leq 0$$

hold.

Proof. We have

$$p = \mathsf{P}(\xi_t = 1) = \iiint\limits_{\substack{0 \le x,y,z \le 1 \\ x < ay + bz}} \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z = \frac{a+b}{2},$$

$$Q_1 = \mathsf{P}(\xi_t = 1, \xi_{t-1} = 1) = \iiiint\limits_{\substack{0 \le x,y,z,w \le 1 \\ x < ay + bz \\ y < az + bw}} \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z\ \mathrm{d}w = \frac{a^3}{6} + \frac{a^2 b}{4} + \frac{ab^2}{6} + \frac{ab}{3} + \frac{b^2}{4},$$

$$Q_2 = \mathsf{P}(\xi_t = 1, \xi_{t-2} = 1) = \int\!\!\!\int\!\!\!\int\!\!\!\int\!\!\!\int\limits_{\substack{0 \le x,y,z,w,u \le 1 \\ x < ay + bz \\ z < aw + bu}} \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z\ \mathrm{d}w\ \mathrm{d}u = \frac{a^2}{4} + \frac{ab}{4} + \frac{a^2 b}{6} + \frac{ab^2}{4} + \frac{b^3}{6}.$$

For r_1 and r_2 we obtain

$$r_1(a,b) = \frac{Q_1 - p^2}{p(1-p)} = \frac{1}{3}\frac{2a^3 + 3a^2 b + 2ab^2 - 2ab - 3a^2}{(a+b)(2-a-b)},$$

$$r_2(a,b) = \frac{Q_2 - p^2}{p(1-p)} = \frac{b}{3}\frac{(a+b)(2a+2b-3) - ab}{(a+b)(2-a-b)}.$$

The inequalities $r_1(a,b) \le 0$, $r_2(a,b) \le 0$, and $s(a,b) \le 0$ are clear. The inequalities $r_1(a,b) \ge -\frac{1}{3}$, $r_2(a,b) \ge -\frac{1}{3}$, and $s(a,b) \ge -\frac{1}{3}$ follow from

$$r_1(a,b) + \frac{1}{3} = \frac{(1-a-b)[(2a+b)(1-a) + b(1-b)] + ab + b^2(2-b)}{3(a+b)(2-a-b)} \ge 0,$$

$$r_2(a,b) + \frac{1}{3} = \frac{ab(1-b) + a^2}{3(a+b)(2-a-b)} \ge 0,$$

$$s(a,b) + \frac{1}{3} = \frac{2(a+b)(1-a-b)^2 + ab(1-a-b)}{3(a+b)(2-a-b)} \ge 0.$$

□

Some other properties of the autocorrelation function of (3.24) can be further derived. For the 1-lag correlation r_1 we get $r_1(0,b) = 0$, $r_1(a,0) = r(a)$ and

$$\frac{\partial r_1(a,b)}{\partial b} = \frac{a[2a^2 + a^3 + 2b^2 + a(2 + 2b - b^2)]}{3(2-a-b)^2(a+b)^2} > 0.$$

Hence, the function $r_1(a,b)$ of the variable b is increasing for a fixed a. We have mentioned in the previous section that the function $r_1(a,0) = r(a)$ is decreasing for $a \in (0,1]$. Thus $r_1(a,b) > r_1(a,0) = r(a)$ and $r(a) > r(1)$ for $0 < a < 1$. Similarly, it can be shown that for a given a the functions $r_2(a,b)$ and $s(a,b)$ of the variable b are decreasing. For a fixed b the function $s(a,b)$ of the variable a is decreasing as well and $s(a, 1-a) = -1/3$.

Theorem 3.35 describes the dependence of r_1 on the parameters a, b. However, in practical situations one is more interested in the relation between the autocorrelation r_1 and the corresponding probability p. Let $p = (a+b)/2$ be given. Then $b = 2p - a$ and

$$r_1 = r_1^*(a,p) = -\frac{a}{12}\frac{(2p+a)(1-2p)+(2p-a)(1-2p+a)+a}{p(1-p)}.$$

The conditions $0 \le a \le 1$, $0 \le b \le 1$, $a+b \le 1$, $a+b > 0$ lead to $0 < p \le 1/2$ $0 \le a \le 2p$. We get

$$\frac{\partial r_1^*(a,p)}{\partial a} = -\frac{3a(2p-a)+4p(1-2p)+2a(1-p)}{12p(1-p)} \le 0.$$

For a given p, the function $r_1^*(a,p)$ of the variable a is decreasing and its minimum is reached at $a = 2p$. This minimum equals to $r_1^*(2p,p) = -[p\,(3-4p)]/[3(1-p)]$. The function $r_1^{**}(p) = r_1^*(2p,p)$ of the variable p is decreasing on $(0,1/2]$ because its derivative is negative for all $0 < p < 1/2$. Thus the minimum of $r_1^*(2p,p)$ is equal to $-1/3$ and it is reached at $p = 1/2$.

Finally, let us remark that even though we have shown that the model (3.24) satisfies $r_1 \le 0$, $r_2 \le 0$ and $r_1 + r_2 \ge -1/3$, it is possible to construct 2-dependent negatively correlated Bernoulli variables such that $r_1 + r_2 < -1/3$ holds, see Meister (2004), cf. Example 3.36 in our Section 3.4.

3.4 Clipping of a Gaussian Process

It is stated in Section 2.1 that an autocorrelation function $\{r_t\}$ of a general stationary process with negatively correlated variables satisfies the equality $\sum_{k=1}^{\infty} r_k \ge -1/2$. For 0-1 valued two-block-factors this inequality can be improved so that $r_1 \ge -1/3$ holds, see Remark 3.4. It is possible to show that for m-dependent Bernoulli variables (or Bernoulli $m+1$ block-factors), $m \ge 2$, the sum $\sum_{k=1}^{\infty} r_k$ can be below $-1/3$, see Example 3.36. It seems that the lower bound $-1/2$ for $\sum_{t=1}^{\infty} r_t$ might be improved even in this situation, but this problem remains open. Nevertheless, it is possible to find a lower bound for $\sum_{t=1}^{\infty} r_t$ for some classes of processes. For instance, Meister (2004) considers some models of $m+1$ block factors and studies the sum $\sum_{t=1}^{\infty} r_t$ for different sampling methods. He provides numerical examples of the minimal value of the sum in some special settings.

Example 3.36. Let $a, m \in \mathbb{N}$ and let $\{X_t,\, t \in \mathbb{Z}\}$ be iid random variables with the uniform distribution on the interval $[0,1]$. For $c \in (0,1)$ define

$$\xi_t = \begin{cases} 1 & \text{if } X_t < c, X_{t-1} < c, \dots, X_{t-a+1} < c, X_{t-a} > c, \dots X_{t-a-m+1} > c, \\ 0 & \text{otherwise.} \end{cases} \tag{3.26}$$

Then $\{\xi_t,\, t \in \mathbb{Z}\}$ is a sequence of $a+m-1$ dependent Bernoulli variables such that

$$\mathsf{P}(\xi_t = 1) = c^a(1-c)^m = p,$$
$$\mathsf{P}(\xi_t = 1,\ \xi_{t+k} = 1) = \begin{cases} 0 & k = 1, \dots, m+a-1, \\ p^2 & k > m+a-1. \end{cases}$$

We obtain

$$r_1 = r_2 = \cdots = r_m = \frac{-p}{1-p} = \frac{-c^a(1-c)^m}{1-c^a(1-c)^m},$$

and $r_k = 0$ for $k > m+a-1$. Hence, $r_t \le 0$ for all $t \ge 1$, and $\{\xi_t,\, t \in \mathbb{Z}\}$ are negatively correlated variables. Denote $s_m(c) = \sum_{t=1}^{\infty} r_t$. Then

$$s_m(c) = \sum_{t=1}^{\infty} r_t = \sum_{t=1}^{m+a-1} r_t = (a+m-1) r_1 = \frac{-(m+a-1)c^a(1-c)^m}{1-c^a(1-c)^m}.$$

Keeping a, m fixed, $s_m(c)$ reaches its minimum for $c = a/(m+a)$. Denote $s(a,m) = s_m(a/(m+a))$. Then we get

$$\lim_{m\to\infty} s(a,m) = \begin{cases} -1/\mathrm{e}, & a = 1, \\ 0, & a > 1. \end{cases}$$

The model (3.26) with $a = 1$ and $m \ge 2$ was proposed by Bondesson (2003) as an example of m-dependent variables with $\sum_{k=1}^{\infty} r_k$ below $-1/3$. It is left as an open problem whether $-1/\mathrm{e}$ is the highest possible lower bound for $\sum_{k=1}^{\infty} r_k$ for general negatively correlated Bernoulli variables.

In the following we describe $\sum_{t=1}^{\infty} r_t$ for variables obtained by clipping a Gaussian process. However, before formulating the main result in Theorem 3.38 we prove a helping lemma.

Lemma 3.37. *Let $\Phi(x)$ be the cumulative distribution function of a standard Gaussian distribution, and $f(x, y; \rho)$ be the probability density function of a bivariate Gaussian distribution with vanishing means, unit variances, and the correlation coefficient $\rho \in (-1, 1)$. Then the function $r(c) = -\left[\int_\rho^0 f(c, c, y)\,\mathrm{d}y\right] \{\Phi(c)[1-\Phi(c)]\}^{-1}$ has its global minimum at the point $c = 0$.*

Proof. Denote $H(c) = -\int_\rho^0 f(c, c, y)\,\mathrm{d}y = -\int_\rho^0 \exp\{-c^2/(1+y)\} \left(2\pi\sqrt{1-y^2}\right)^{-1} \mathrm{d}y$. Since $r(c) = r(-c)$, it suffices to investigate the behavior of the function $r(c)$ only for $c \ge 0$. Let φ denote the probability density function of a standard normal distribution. Then

$$r'(c) = \frac{H'(c)\Phi(c)[1-\Phi(c)] + H(c)[2\Phi(c)-1]\varphi(c)}{\{\Phi(c)[1-\Phi(c)]\}^2},$$

where $H'(c) = c\int_\rho^0 \mathrm{e}^{-c^2/(1+y)}\left(\pi(1+y)\sqrt{1-y^2}\right)^{-1} \mathrm{d}y$. It follows easily that $r'(0) = 0$. We show that $r'(c) > 0$ for $c > 0$.

Notice that $H'(c) > 2c[-H(c)]$ holds for all $c > 0$, and recall the well-known inequality $1 - \Phi(c) > c\varphi(c)/(1 + c^2)$ for all $c > 0$. If $c > 4/5$ then it follows after some easy algebraic manipulations that $2\Phi(c) - 1 < 2c^2\Phi(c)(1 + c^2)^{-1}$ holds. We get

$$H'(c)\Phi(c)[1-\Phi(c)] > 2c[-H(c)]\,\frac{c}{1+c^2}\,\varphi(c)\Phi(c) > [-H(c)]\varphi(c)[2\Phi(c)-1].$$

This implies that $r_1'(c) > 0$ for all $c > 4/5$. If $0 < c \leq 4/5$ then $\Phi(c)[1 - \Phi(c)] \geq \Phi(4/5)[1 - \Phi(4/5)] > 1/(2\pi) > \varphi(c)/\sqrt{2\pi}$. The function $2\Phi(c) - 1$ is concave for $c \geq 0$, and therefore the inequality $2\Phi(c) - 1 < 2c/(\sqrt{2\pi})$ holds. We obtain

$$[-H(c)]\varphi(c)[2\Phi(c)-1] < [-H(c)]\,\varphi(c)\frac{2}{\sqrt{2\pi}}c < \frac{H'(c)}{\sqrt{2\pi}}\varphi(c) < H'(c)\Phi(c)[1-\Phi(c)].$$

Hence, $r'(c) > 0$ holds for $0 < c \leq 4/5$ as well. □

Theorem 3.38. *Let $\{Z_t, t \in \mathbb{Z}\}$ be a stationary Gaussian random process with the autocorrelation function $\{\rho_t\}$ such that $\rho_t \leq 0$ for all $t \geq 1$. For $c \in \mathbb{R}$ define $\xi_t = \mathrm{I}[Z_t \geq c]$ and denote $\{r_t\}$ the autocorrelation function of the process $\{\xi_t, t \in \mathbb{Z}\}$. Then $r_t \leq 0$ for all $t \geq 1$ and*

$$\sum_{t=1}^{\infty} r_t \geq -\frac{1}{3}.$$

Proof. Without loss of generality assume that $\mathsf{E}Z_t = 0$ and $\mathsf{var}\, Z_t = 1$. Consider first the case $c = 0$. It is known that $\mathsf{P}(Z_t \geq 0, Z_{t+k} \geq 0) = 1/4 + (\arcsin \rho_k)/(2\pi)$ holds for $k \geq 1$, see Lemma A.7 in Appendix. Since $p = \mathsf{P}(\xi_t = 1) = 1/2$, the autocorrelations $\{r_k\}$ can be expressed as $r_k = (2/\pi) \arcsin \rho_k$. This implies that $r_k \leq 0$ for all $k \geq 1$. The autocorrelations $\{\rho_t\}$ of the original process $\mathbf{Z}$ satisfy the inequality $\sum_{t=1}^{\infty} \rho_t \geq -1/2$ and thus $-1/2 \leq \rho_t \leq 0$ for all $t \geq 1$. Since the function arcsin is concave on the interval $(-1/2, 0)$ the inequality $\arcsin x \geq \pi x/3$ holds for all $x \in [-1/2, 0]$. We get

$$\sum_{k=1}^{\infty} r_k = \frac{2}{\pi}\sum_{k=1}^{\infty} \arcsin \rho_k \geq \frac{2}{\pi}\frac{\pi}{3}\sum_{k=1}^{\infty} \rho_k \geq -\frac{1}{3}.$$

Now consider the general case $c \in \mathbb{R}$. Denote as $\Phi(x)$ the cumulative distribution function of a standard Gaussian distribution and $f(x, y; \rho)$ the probability density function of a bivariate Gaussian distribution with zero means, unit variances, and the correlation coefficient $\rho \in (-1, 1)$. The equality $\mathsf{P}(Z_t > c, Z_{t+k} > c) = -\int_{\rho_k}^0 f(c, c, y)\,\mathrm{d}y + [1 - \Phi(c)]^2$ holds, see Kedem (1980a). Hence, the correlation r_k is given as

$$r_k(c) = \frac{-\int_{\rho_k}^0 f(c,c,y)\,\mathrm{d}y}{\Phi(c)[1-\Phi(c)]}.$$

The function $r_k(c)$ has its global minimum at the point $c = 0$, see Lemma 3.37. Hence, $r_k(c) \geq r_k(0)$, and we get $\sum_{k=1}^{\infty} r_k(c) \geq \sum_{k=1}^{\infty} r_k(0) \geq -1/3$. □

Assume that $\mathsf{E} Z_t = 0$ and $\mathsf{var}\, Z_t = 1$ in Theorem 3.38. The proof of Theorem 3.38 shows that the lower bound $-1/3$ is reached for m-dependent variables if and only if $c = 0$, and $\rho_k = -1/2$ for some $k \in \{1, \dots, m\}$ and $\rho_t = 0$ otherwise. For any other setting, the sum of correlations lies strictly above $-1/3$.

It follows from the relationship $r_k = (2/\pi) \arcsin \rho_k$ that $|r_k| \leq |\rho_k|$ holds. In addition, ξ_t and ξ_{t+k} are independent if and only if they are uncorrelated. This demonstrates that the pairwise dependence in the clipped series $\boldsymbol{\xi}$ is always weaker than that in $\mathbf{Z}$.

In the beginning of this chapter we have specified that it is sometimes desirable to have a model for dependent Bernoulli variables which are "negatively correlated as much as possible". In other words the sum of the autocorrelations $\sum_{t=1}^{\infty} r_t$ should be as low as possible. Theorem 3.38 shows that generating dependent Bernoulli variables by clipping a Gaussian process leads to autocorrelations $\{r_t\}$ such that $\sum_{t=1}^{\infty} r_t \geq -1/3$. On the other hand, clipping Gaussian variables is probably the most common method of generating dependent 0-1 valued variables, cf. Došlá (2008). Theorem 3.38 demonstrated that there exist situations, e.g. negatively correlated variables, for which this approach is obviously not optimal.

Appendix

Let $\{a_k\}_{k=0}^{\infty}$ be a sequence of complex numbers such that $\sum_{k=0}^{\infty} |a_k|^2 < \infty$. Consider a linear process $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ defined as

$$X_t = \sum_{t=0}^{\infty} a_k \varepsilon_{t-k}, \quad \{\varepsilon_t,\, t \in \mathbb{Z}\} \text{ is a white noise } \mathsf{WN}(0,1). \tag{A.1}$$

The following result comes from Gichman and Skorochod (1965).

Lemma A.1. *A centered weakly stationary process* $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ *is a linear process* (A.1) *if and only if its spectral distribution function is absolutely continuous and its spectral density* f *satisfies*

$$f(\lambda) = |g(\mathrm{e}^{i\lambda})|^2, \tag{A.2}$$

where $g(\mathrm{e}^{i\lambda}) = \sum_{k=0}^{\infty} b_k \mathrm{e}^{ik\lambda}$ *and* $\sum_{k=0}^{\infty} |b_k|^2 < \infty$.

Proof. We present the proof from Gichman and Skorochod (1965). Let $\mathbf{X}$ be a linear process (A.1). It is well-known that $\mathbf{X}$ has an absolutely continuous spectrum, and the spectral density f is given as $f(\lambda) = [\sigma^2/(2\pi)]|\sum_{k=0}^{\infty} a_k \mathrm{e}^{-ik\lambda}|^2$ for $\lambda \in [-\pi, \pi]$. Hence, the function $g(\mathrm{e}^{i\lambda}) = [\sigma/(\sqrt{2\pi})]\sum_{k=0}^{\infty} \overline{a}_k \mathrm{e}^{ik\lambda}$ clearly satisfies (A.2).

Assume now that there exists a function g such that (A.2) holds. Since $\mathbf{X}$ is a centered weakly stationary process, we can consider its spectral representation $X_t = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\, \mathrm{d}Z(\lambda)$, where Z is an orthogonal random measure such that $\mathsf{E}Z(A)\overline{Z(B)} = \int_{A\cap B} f(\lambda)\, \mathrm{d}\lambda$ holds for any $A, B \in \mathcal{B}([-\pi, \pi])$, see Brockwell and Davis (1991). On $\mathcal{B}([-\pi, \pi])$ define a random measure ξ as

$$\xi(A) = \int\limits_A \frac{1}{\sqrt{2\pi}} \frac{1}{g(\mathrm{e}^{i\lambda})}\, \mathrm{d}Z(\lambda), \quad A \in \mathcal{B}([-\pi, \pi]).$$

Observe that $\mathsf{E}\xi(A)\overline{\xi(B)} = [1/(2\pi)]\int_{A\cap B} \mathrm{d}\lambda = \ell(A \cap B)/(2\pi)$, where ℓ is the Lebesque measure on $\mathbb{R}$. This means that ξ is an orthogonal random measure. Define $\varepsilon_t = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\, \mathrm{d}\xi(\lambda)$. Then it is easy to see that $\boldsymbol{\varepsilon} = \{\varepsilon_t,\, t \in \mathbb{Z}\}$ is a white noise with $\mathsf{var}\, \varepsilon_t = 1$.

Furthermore, we have

$$X_t = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\, \mathrm{d}Z(\lambda) = \int_{-\pi}^{\pi} \sqrt{2\pi}\mathrm{e}^{it\lambda}\, \overline{g(\mathrm{e}^{i\lambda})}\, \mathrm{d}\xi(\lambda) = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\sqrt{2\pi} \left(\sum_{n=0}^{\infty} \overline{b_n}\mathrm{e}^{-in\lambda}\right) \mathrm{d}\xi(\lambda)$$
$$= \sum_{n=0}^{\infty} \sqrt{2\pi}\, \overline{b_n} \int_{-\pi}^{\pi} \mathrm{e}^{i(t-n)\lambda}\, \mathrm{d}\xi(\lambda) = \sum_{n=0}^{\infty} a_n \varepsilon_{t-n},$$

where we have defined $a_n = \sqrt{2\pi}\, \overline{b_n}$ and used the definition of ε_t. Hence, $\mathbf{X}$ is a linear process (A.1). □

Remark A.2. Denote $D = \{z \in \mathbb{C} : \ |z| < 1\}$, and for $g : D \to \mathbb{C}$ define

$$||g||^2 = \lim_{r\to 1-} \frac{1}{2\pi} \int_{-\pi}^{\pi} |g(r\mathrm{e}^{i\lambda})|^2\, \mathrm{d}\lambda = \sup_{0<r<1} \frac{1}{2\pi} \int_{-\pi}^{\pi} |g(r\mathrm{e}^{i\lambda})|^2\, \mathrm{d}\lambda.$$

The Hardy space H_2 is defined as $H_2 = \{g : \ g \text{ is holomorphic in } D \text{ and } ||g||^2 < \infty\}$, see Katznelson (2004).

(i) Let us list some of the properties of the H_2 space. We refer to Katznelson (2004) for the proofs.

1. If $g(z) = \sum_{n=0}^{\infty} b_n z^n$ then $g \in H_2$ if and only if $\sum_{n=0}^{\infty} |b_n|^2 < \infty$.
2. The boundary value $g(\mathrm{e}^{i\lambda}) := \lim_{r\to 1-} g(r\mathrm{e}^{i\lambda})$ exists almost surely and

$$g(r\mathrm{e}^{i\lambda}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\mathrm{e}^{i\lambda}) P(r, \lambda - t)\, \mathrm{d}t,$$

where $P(r, t)$ is the Poisson kernel $P(r, t) = \dfrac{1 - r^2}{1 - 2r\cos t + r^2}$.

If $g(z) = \sum_{n=0}^{\infty} b_n z^n$ then $g(\mathrm{e}^{i\lambda}) = \sum_{n=0}^{\infty} b_n \mathrm{e}^{in\lambda} \in L_2$.

(ii) In view of (i) we can reformulate Lemma A.1 in an equivalent way as: *A centered weakly stationary process* $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ *is a linear process* (A.1) *if and only if there exists a function* $g \in H_2$ *such that the spectral density* f *of* $\mathbf{X}$ *satisfies* (A.2), *where* $g(\mathrm{e}^{i\lambda})$ *is the boundary value of* g.

It is not easy to verify the condition (A.2) from Lemma A.1 in practice. The following result from Gichman and Skorochod (1965) is more useful.

Theorem A.3. *A centered stationary process* $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ *with* $\operatorname{var} X_t > 0$ *is a linear process* (A.1) *if and only if its spectral distribution function is absolutely continuous and its spectral density* f *satisfies*

$$\int_{-\pi}^{\pi} |\ln f(\lambda)|\, \mathrm{d}\lambda < \infty. \tag{A.3}$$

Proof. Notice that it suffices to verify only the condition

$$\int_{-\pi}^{\pi} \ln f(\lambda)\,\mathrm{d}\lambda > -\infty \tag{A.4}$$

in (A.3) because $\int_{-\pi}^{\pi} \ln^+ f(\lambda)\,\mathrm{d}\lambda \le \int_{-\pi}^{\pi} f(\lambda)\,\mathrm{d}\lambda = \mathsf{var}\, X_t < \infty$. We briefly outline the proof from Gichman and Skorochod (1965). The objective is to show that (A.3) is equivalent to the condition (A.2) of Lemma A.1.

Assume that there exists $g \in H_2$ such that (A.2) holds. We want to show that (A.4) holds, or equivalently that $\ln^- |g(\mathrm{e}^{i\lambda})| \in L_1$. Without loss of generality assume that $g(0) \neq 0$. The Poisson-Jensen inequality, see Katznelson (2004), gives for $0 < r < 1$

$$\ln|g(0)| \le \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln|g(r\mathrm{e}^{i\lambda})|\,\mathrm{d}\lambda = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln^+|g(r\mathrm{e}^{i\lambda})|\,\mathrm{d}\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln^-|g(r\mathrm{e}^{i\lambda})|\,\mathrm{d}\lambda.$$

We get

$$\begin{aligned}\int_{-\pi}^{\pi} \ln^-|g(r\mathrm{e}^{i\lambda})|\,\mathrm{d}\lambda &\le \int_{-\pi}^{\pi} \ln^+|g(r\mathrm{e}^{i\lambda})|\,\mathrm{d}\lambda - 2\pi\ln|g(0)| \le \int_{-\pi}^{\pi} |g(r\mathrm{e}^{i\lambda})|^2\,\mathrm{d}\lambda - 2\pi\ln|g(0)| \\ &\le 2\pi||g||^2 - 2\pi\ln|g(0)| < \infty.\end{aligned}$$

The Fatou's lemma together with the latter inequality give $\int_{-\pi}^{\pi} \ln^-|g(\mathrm{e}^{i\lambda})|\,\mathrm{d}\lambda < \infty$.

Now assume that (A.3) holds. Define

$$u(r\mathrm{e}^{it}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln[f(\lambda)]P(r, t-\lambda)\,\mathrm{d}\lambda,$$

where P is the Poisson kernel defined in Remark A.2. Let $\varphi:\ \mathbb{C} \to \mathbb{C}$ be a function such that $\varphi(z)$ is analytic for $z \in D$, $\varphi(0) > 0$ and $\operatorname{Re}\varphi = u$. Define $g(z) = \mathrm{e}^{\varphi(z)/2}$. We show that $g \in H_2$ and (A.2) holds. In view of Poisson-Jensen inequality we get

$$u(r\mathrm{e}^{it}) \le \ln\Big[\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda)P(r, t-\lambda)\,\mathrm{d}\lambda\Big],$$

and thus

$$|g(r\mathrm{e}^{it})|^2 = \mathrm{e}^{\operatorname{Re}\varphi(r\mathrm{e}^{it})} = \mathrm{e}^{u(r\mathrm{e}^{it})} \le \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda)P(r, t-\lambda)\,\mathrm{d}\lambda.$$

Hence,

$$\int_{-\pi}^{\pi} |g(r\mathrm{e}^{it})|^2\,\mathrm{d}t \le \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda)\int_{-\pi}^{\pi} P(r, t-\lambda)\,\mathrm{d}t\,\mathrm{d}\lambda = \int_{-\pi}^{\pi} f(\lambda)\,\mathrm{d}\lambda = \mathsf{var}\, X_t < \infty,$$

where we have used $\int_{-\pi}^{\pi} P(r, t-\lambda)\,\mathrm{d}t = 2\pi$, see Zygmund (2002). Hence, $||g||^2 < \infty$ and $g \in H_2$. The boundary value of $|g(z)|^2$ is given as

$$\lim_{r\to 1-} |g(r\mathrm{e}^{i\lambda})|^2 = \exp\big\{\lim_{r\to 1-} u(r, \lambda)\big\} = \exp\{\ln f(\lambda)\} = f(\lambda) \text{ a.s.}$$

□

Remark A.4. (i) The function φ in the proof of Theorem A.3 can be taken for instance as

$$\varphi(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln[f(t)]\,\frac{\mathrm{e}^{it}+z}{\mathrm{e}^{it}-z}\,\mathrm{d}t.$$

This function have all the desirable properties, see Katznelson (2004).

(ii) The function g in Lemma A.1 is not unique. It follows from the proof of Theorem A.3 that it is always possible to choose g such that $g(0) > 0$ and $g(z) \neq 0$ for $z \in D$. Moreover, it is possible to show that g is unique under the latter condition and is equal to the function φ defined in (i).

(iii) If there exists a continuous spectral density f then $f(\lambda) = |g(\mathrm{e}^{i\lambda})|^2$ for all $\lambda \in [-\pi,\pi]$ in the proof of Theorem A.3. This follows from some general properties of Poisson integrals, see Zygmund (2002).

(iv) Recall that in the proof of Lemma A.1 we have shown that if $g(z) = \sum_{k=0}^{\infty} b_k z^k$ then $X_t = \sum_{k=0}^{\infty} c_k\varepsilon_{t-k}$, where $c_k = \sqrt{2\pi}\,\overline{b_k}$ and $\mathsf{var}\,\varepsilon_t = 1$. It follows from (ii) that the process $\mathbf{X}$ can be represented as $X_t = \sum_{k=0}^{\infty} a_k\tilde{\varepsilon}_{t-k}$ so that $a_0 = 1$, the zeros of $\sum_{k=0}^{\infty} a_k z^k$ do not lie inside the unit disc D and $\tilde{\boldsymbol{\varepsilon}}$ is a white noise with $\mathsf{var}\,\tilde{\varepsilon}_t > 0$.

(v) The proof of Theorem A.3 shows how to compute the coefficients $\{a_k\}$ from (A.1). Let $g(z) = \mathrm{e}^{\varphi(z)/2}$, where φ is defined in (i). Define $d_k = \int_{-\pi}^{\pi} \mathrm{e}^{ik\lambda}\ln[f(\lambda)]\,\mathrm{d}\lambda$ for $k = 0,1,2,\ldots$, and let $h(z) = \exp\Big\{\frac{1}{2\pi}\sum_{k=1}^{\infty} d_k z^k\Big\} = \sum_{k=0}^{\infty} a_k z^k$. Then $\overline{g(z)} = \exp\{d_0/(4\pi)\}\exp\{\frac{1}{2\pi}\sum_{k=1}^{\infty} d_k \bar{z}^k\} = \mathrm{e}^{d_0/(4\pi)}h(\bar{z})$ and

$$f(\lambda) = \mathrm{e}^{d_0/(2\pi)}\Big|\sum_{k=0}^{\infty} a_k \mathrm{e}^{-ik\lambda}\Big|^2.$$

Hence, $X_t = \sum_{k=0}^{\infty} a_k\varepsilon_{t-k}$, where $\mathsf{var}\,\varepsilon_t = 2\pi\mathrm{e}^{d_0/(2\pi)} = 2\pi\exp\Big\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln[f(\lambda)]\,\mathrm{d}\lambda\Big\}$.

The following theorem is fundamental in the theory of stationary time series. It dates back to Wold (1938).

Theorem A.5. *(Wold's decomposition.) Let $\mathbf{X} = \{X_t,\ t \in \mathbb{Z}\}$ be a centered (weakly) stationary time series. Define $\mathcal{H}_t = \overline{span}\{X_k,\ -\infty < k \le t\}$, and let P_tX_{t+k} denote the linear projection of X_{t+k} on the space $\mathcal{H}_t$. Define $\sigma_X^2 = \mathsf{E}|X_{t+1} - P_tX_{t+1}|^2$ the one-step squared error and $\mathcal{M}_{-\infty} = \cap_{t=-\infty}^{\infty}\mathcal{M}_t$ the closed linear subspace of the Hilbert space $\mathcal{M} = \overline{span}\{X_t,\ t \in \mathbb{Z}\}$. If $\sigma_X^2 > 0$ then X_t can be represented in the form*

$$X_t = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j} + \kappa_j,$$

where $\psi_0 = 1$, $\sum_{j=0}^{\infty}\psi_j^2 < \infty$, $\{\varepsilon_t,\ t \in \mathbb{Z}\}$ is a white noise $\mathsf{WN}(0,\sigma^2)$ such that $\varepsilon_t = X_t - P_{t-1}X_t$, ε_t and κ_s are uncorrelated for any $s,t \in \mathbb{Z}$, $\kappa_t \in \mathcal{M}_{-\infty}$ for all $t \in \mathbb{Z}$, $\{\kappa_t\}$ is deterministic, i.e. $\sigma_\kappa^2 = 0$, where σ_κ^2 is defined in an analogous way as σ_X^2.

The variable ε_t represents the error made in forecasting X_t on the basis of a linear function of lagged X. The term κ_t is called the linearly deterministic component *of X_t, and $\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ is called the* linearly non-deterministic component. *If $\kappa_t \equiv 0$, equivalently $\mathcal{M}_{-\infty} = \{0\}$, then the process is called* purely non-deterministic.

Proof. See e.g. Brockwell and Davis (1991). □

The following lemma summarizes basic properties of ARMA processes. We refer to Hamilton (1994) for the proof.

Lemma A.6. *Let $\mathbf{X} = \{X_t,\, t \in \mathbb{Z}\}$ follow an ARMA(p, q) process*

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_I X_{t-I} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_I \varepsilon_{t-I},$$

where $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ is a white noise with $\mathsf{var}\, \varepsilon_t = \sigma^2 > 0$. Let

$$1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q = (1 - \alpha_1 z) \ldots (1 - \alpha_q z),$$
$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = (1 - \beta_1 z) \ldots (1 - \beta_p z).$$

The process $\mathbf{X}$ is stationary if and only if the roots of the equation $1 - \phi_1 z - \phi_2 z^2 - \ldots \phi_p z^p = 0$ lie outside the unit circle, i.e. $|\beta_i| < 1$ for all $i = 1, \ldots, p$. In this case the spectral density of $\mathbf{X}$ is given as

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{\prod_{i=1}^{q} |1 - \alpha_i \mathrm{e}^{-\mathrm{i}\lambda}|^2}{\prod_{j=1}^{p} |1 - \beta_j \mathrm{e}^{-\mathrm{i}\lambda}|^2} = \frac{\sigma^2}{2\pi} \frac{\prod_{i=1}^{q} (1 + \alpha_i^2 - 2\alpha_i \cos\lambda)}{\prod_{j=1}^{p} (1 + \beta_j^2 - 2\beta_j \cos\lambda)}. \tag{A.5}$$

The process $\mathbf{X}$ is invertible if and only if $|\alpha_j| < 1$ for all $j = 1, \ldots, q$.

The following lemma describes some properties of a multivariate Gaussian distribution.

Lemma A.7. *Let $(Z_1, Z_2, Z_3, Z_4)'$ have a zero-mean multivariate Gaussian distribution such that $\mathsf{var}\, Z_i > 0$ for $i = 1, 2, 3, 4$. Let the correlation coefficient of Z_i and Z_j be ρ_{ij}, $i, j = 1, 2, 3, 4$. If $i \neq j$ then*

$$\mathsf{P}(Z_i \geq 0, Z_j \geq 0) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho_{ij}. \tag{A.6}$$

If $i \neq j$, $i \neq k$, $j \neq k$, then

$$\mathsf{P}(Z_i \geq 0, Z_j \geq 0, Z_k \geq 0) = \frac{1}{8} + \frac{1}{4\pi} \left(\arcsin \rho_{ij} + \arcsin \rho_{jk} + \arcsin \rho_{ik} \right). \tag{A.7}$$

Define $\rho = \max\{|\rho_{13}|, |\rho_{14}|, |\rho_{23}|, |\rho_{24}|\}$. Then

$$|\mathsf{P}(Z_1 \geq 0, Z_2 \geq 0, Z_3 \geq 0, Z_4 \geq 0) - \mathsf{P}(Z_1 \geq 0, Z_2 \geq 0)\mathsf{P}(Z_3 \geq 0, Z_4 \geq 0)| < C\rho, \tag{A.8}$$

where C is a function of ρ_{12} and ρ_{34} only.

Proof. See Kedem (1980a), cf. Lomnicki and Zaremba (1955). □

The probability $\mathsf{P}(Z_t \geq 0, Z_s \geq 0)$ is referred to as a bivariate orthant probability. The equality (A.6) is called *arcsine formula*, and its origin can be traced back to Sheppard (1899). Note that the formula (A.6) holds for general ellipsoidal distributions as well, see Kedem (1994).

Unfortunately, there is no simple extension of formulas (A.6), (A.7) to higher dimensions (when four and more variables are present), see David (1953). In general, only upper limits like (A.8) or some approximations are know. Some special cases are treated in Owen (1988).

In Chapter 3 we considered the Bondesson's model (3.2) for 1-dependent negatively correlated Bernoulli variables. The following lemma provides formulas for the basic characteristics of this model.

Lemma A.8. *Let $\{X_t,\, t \in \mathbb{Z}\}$ be iid variables with an absolutely continuous distribution with a density f and a distribution function F. Let $\boldsymbol{\xi} = \{\xi_t,\, t \in \mathbb{Z}\}$ be defined by the model* (3.2)*. Then*

$$p = \mathsf{P}(\xi_t = 1) = \mathsf{E}F(X+c),$$

$$r_1 = \mathsf{cor}\,(\xi_t, \xi_{t-1}) = 1 - \frac{\mathsf{E}F(X+c)F(X-c)}{\mathsf{E}F(X+c)\mathsf{E}F(X-c)} = 1 - \frac{Q}{p(1-p)},$$

where

$$Q = \mathsf{E}F(X+c)F(X-c).$$

Proof. We sketch the proof from Bondesson (2003). We have

$$\mathsf{P}(\xi_t = 1) = \iint\limits_{x-y<c} f(x)f(y)\ \mathrm{d}x\ \mathrm{d}y = \int_{-\infty}^{\infty} F(y+c)f(y)\ \mathrm{d}y = \mathsf{E}F(X+c),$$

$$\begin{aligned}
\mathsf{P}(\xi_t = 1, \xi_{t-1} = 1) &= \iiint\limits_{\substack{x-y<c\\ y-z<c}} f(x)f(y)f(z)\ \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z \\
&= \int_{-\infty}^{\infty} F(y+c)f(y)\ \mathrm{d}y - \int_{-\infty}^{\infty} F(y+c)F(y-c)f(y)\ \mathrm{d}y \\
&= \mathsf{E}F(X+c) - \mathsf{E}F(X+c)F(X-c).
\end{aligned}$$

Then

$$\begin{aligned}
r_1 &= \frac{\mathsf{E}F(X+c) - \mathsf{E}F(X+c)F(X-c) - \mathsf{E}F(X+c)^2}{\mathsf{E}F(X+c)[1-\mathsf{E}F(X+c)]} \\
&= 1 - \frac{\mathsf{E}F(X+c)F(X-c)}{\mathsf{E}F(X+c)\mathsf{E}F(X-c)},
\end{aligned}$$

because

$$p = \mathsf{E}F(X + c) = 1 - \mathsf{E}F(X - c) = 1 - (1 - p).$$

□

Bibliography

Aaronson, J., Gilat, D., Keane, M. S., and de Valk, V. (1989). An algebraic construction of a class of one-dependent processes. *Ann. Probab.*, 17:128–143.

Anděl, J. and Došlá, Š. (2010). Bernoulli processes with non-positive correlations. *To appear.*

Beran, J. (1994). *Statistics for Long-memory Processes.* Chapman & Hall, New York.

Block, H. W., Savits, T. H., and Shaked, M. (1982). Some concepts of negative dependence. *Ann. Probab.*, 10:765–772.

Bondesson, L. (2003). On a minimum correlation problem. *Statist. Probab. Lett.*, 62:361–370.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer, New York, second edition.

Chaganty, N. R. and Joe, H. (2006). Range of correlation matrices for dependent Bernoulli random variables. *Biometrika*, 93:197–206.

Coeurjolly, J.-F. (2000). Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study. *J. Stat. Softw.*, 5(7):1–53.

Damsleth, E. and El-Shaarawi, A. H. (1988). Estimation of autocorrelation in a binary time series. *Stochastic Hydrology Hydraul.*, 2(1):61–72.

David, F. N. (1953). A note on evaluation of the multivariate normal integral. *Biometrika*, 40:458–459.

de Valk, V. (1988). The maximal and minimal 2-correlation of a class of 1-dependent 0-1 valued processes. *Israel J. Math.*, 62:181–205.

Doob, J. (1953). *Stochastic Processes.* John Wiley & Sons, Inc., New York.

Došlá, Š. (2008). Processes with negative correlations. In Šafránková, J. and Pavlů, J., editors, *WDS'08 Proceedings of Contributed Papers: Part I – Mathematics and Computer Sciences*, pages 68–73, Prague. Matfyzpress.

Došlá, Š. (2010). Estimation of parameters of a clipped MA(1) process. *Submitted.*

Došlá, Š. and Anděl, J. (2010). Weakly stationary processes with non-positive autocorrelations. *To appear.*

Esary, J. D., Proschan, F., and Walkup, D. W. (1967). Association of random variables, with applications. *Ann. Math. Statist.*, 44:1466–1474.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, volume I. Wiley, New York.

Gandolfi, A., Keane, A., and de Valk, V. (1989). Extremal two-correlations of two-valued stationary one-dependent processes. *Probab. Theory Related Fields*, 80:475–480.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.*, 4:221–238.

Gichman, I. and Skorochod, A. V. (1965). *Vvedenije v teoriju slučajnych processov.* Moskva, Nauka.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press, Priceton, New Jersey.

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1988). *Inequalities.* Cambridge University Press, Cambridge, United Kingdom, second edition.

Hinich, M. (1967). Estimation of spectra after hard clipping of Gaussian processes. *Technometrics*, 9(3):391–400.

Hobson, E. W. (1927). *The Theory of Functions of a Real Variable and the Theory of Fourier's Series.* Dover Publications Inc., New York, third edition.

Hu, T. and Hu, J. (1999). Suffcient conditions for negative association of random variables. *Statist Probab. Lett.*, 45:167–173.

Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables, with applications. *Ann. Statist.*, 11(1):286–295.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman and Hall, London.

Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. II. multivariate reverse rule distributions. *J. Multivariate Anal.*, 10:499–516.

Katznelson, Y. (2004). *An Introduction to Harmonic Analysis.* Cambridge University Press, Cambridge, United Kingdom, third edition.

Kedem, B. (1980a). *Binary Time Series*, volume 52 of *Lecture Notes in Pure and Applied Mathematics*. Dekker, New York and Basel.

Kedem, B. (1980b). Estimation of the parameters in stationary autoregressive processes after hard limiting. *J. Amer. Statist. Assoc.*, 75(369):146–153.

Kedem, B. (1994). *Time Series Analysis by Higher Order Crossings*. IEEE Press, New York.

Keenan, D. M. (1982). A time series analysis of binary data. *J. Amer. Statist. Assoc.*, 77(380):816–821.

Lee, A. J. (1993). Generating random binary deviates having fixed marginal distributions and specified degrees of association. *Amer. Statist.*, 47:209–215.

Lehmann, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.*, 37:1137–1153.

Lehmann, E. L. (1998). *Theory of Point Estimation*. Springer, New York, second edition.

Lomnicki, Z. A. and Zaremba, S. K. (1955). Some applications of zero-one processes. *J. Roy. Statist. Soc. Ser. B*, 17:243–255.

Masry, E. (1972). On covariance functions of unit processes. *SIAM J. Appl. Math.*, 23:28–33.

Matúš, F. (1996). On two-block-factor sequences and one-dependence. *Proc. Amer. Math. Soc.*, 124:1237–1242.

Meister, K. (2004). Some different methods to get stationary Bernoulli sequences with negative correlations for sampling applications. Unpublished manuscript, paper C of PhD thesis at Umeå University.

Meister, K. and Bondesson, L. (2001). Some real time sampling methods. Technical Report 2, Dept. of Math. Statist., Umeå Univ.

Owen, D. B. (1988). Orthant probabilities. In *Kotz-Johnson Encyclopedia od Statistical Sciences*, pages 521–523. Wiley, New York.

Quaqish, B. F. (2003). A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika*, 90:455–63.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Robinson, P. M. (2003). *Time Series with Long Memory.* Oxford University Press.

Samorodnitsky, G. (2007). *Long Range Dependence.* Now Publishers Inc, Hanover, MA, USA.

Sheppard, W. F. (1899). On the application of the theory of error to the case of normal distribution and normal correlation. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 192:101–167.

Wold, H. (1938). *A Study in the Analysis of Stationary Time Series.* Almqvist and Wiksell, Uppsala, Sweden.

Wolfram Research, I. (2005). *Mathematica Edition: Version 5.2.* Wolfram Research, Inc., Champaign, Illinois.

Zorich, V. A. (2004). *Mathematical Analysis*, volume 2 of *Universitext Series.* Springer, New York. Translated by R. Cooke.

Zygmund, A. (2002). *Trigonometric Series.* Cambridge University Press, Cambridge, United Kingdom, third edition.

Author Index

Index