# Inaugural - Dissertation

**zur**
**Erlangung der Doktorwürde**
**der**
**Naturwissenschaftlich-Mathematischen Gesamtfakultät**
**der**
**Ruprecht-Karls-Universität**
**Heidelberg**

vorgelegt von

Diplom-Mathematiker Sebastian Schweer
aus Freiburg im Breisgau

Tag der mündlichen Prüfung: ____________

# Statistical Inference for Discrete-Valued Stochastic Processes

Betreuer: **Prof. Dr. Rainer Dahlhaus**

# Acknowledgements

I would like to thank Prof. Dr. Rainer Dahlhaus, my supervisor, for his great support during my studies at the University of Heidelberg. I have had the honor to be his student ever since my undergraduate years and I am grateful for everything he taught me during these years.

Dr. Cornelia Wichelhaus has contributed a lot to my education, and without her constant encouragement I would have never achieved as much as I did. I would like to thank her for all the time we spent working together, starting with my diploma thesis and culminating in our joint contribution discussed in this thesis.

I am grateful to have made the acquaintance of Prof. Dr. Christian Weiß during my PhD research. Discussing problems and new approaches with him has not only been enlightening but also very enjoyable, and I owe him special gratitude for his willingness to share his knowledge of count data time series models with me.

Finally, I had the pleasure to be a member of the Research Training Group (RTG) 1953, "Statistical Modeling of Complex Systems and Processes". I thank the *Deutsche Forschungsgemeinschaft* for this support.

# Zusammenfassung

Im Rahmen verschiedener Modelle diskretwertiger stochastischer Prozesse werden statistische Methoden und Hypothesentests behandelt. Im Falle von ganzzahligen autoregressiven Prozessen erster Ordnung können zugrundeliegende stochastische Eigenschaften genutzt werden, um geeignete Teststatistiken für diverse Szenarien herzuleiten. Drei verschiedene Tests werden eingeführt, welche die Abweichungen von empirischen Schätzern der Dispersion, der verallgemeinerten Autokovarianz und der Schiefe von den jeweiligen theoretischen Werten mit Hilfe der explizit berechneten asymptotischen Verteilungen der Schätzer bewerten. Für diese Tests werden Simulationstudien und Anwendungen auf echte Daten beschrieben, die das Verhalten der Test im Rahmen von kleinen Datensätzen veranschaulichen.

In einem allgemeineren Zusammenhang wird ein weiterer Ansatz verfolgt, der sich auf Erzeugendenfunktionen der Zufallsvariablen stützt. Die asymptotischen Eigenschaften der resultierenden Teststatistik werden für eine sehr allgemeine Klasse Markovscher Modelle, die eine Drift Bedingung erfüllen, hergeleitet. Darüberhinaus wird für einen nichtparametrischen Schätzer der stationären Verteilung ein funktionaler Grenzwertsatz bewiesen. Nachdem Zusammenhänge zwischen diesem Ansatz und demjenigen der vorhergehenden Kapitel aufgezeigt werden, hebt eine Simulationsstudie die gute Leistung dieser Tests in Anwendungen mit kleinen Anzahlen von Beobachtungen hervor.

Ein weiteres Kapitel hat einen speziellen nichtparametrischen Schätzer der Bedienzeitverteilung eines zeitdiskreten $GI/G/\infty$-Warteschlangenmodells zum Gegenstand, wobei angenommen wird, dass die verfügbaren Daten auf die Anzahlen der ankommenden und abgehenden Kunden pro Zeitabschnitt beschränkt sind. Es wird gezeigt, dass dieser sogenannte *sequence-of-difference estimator* einem funktionalen Grenzwertsatz auf einem geeignet gewählten zugrundeliegenden Folgenraum gehorcht. Eine *moving block bootstrap* Methode wird vorgeschlagen und die theoretischen Eigenschaften dieses Ansatzes genauer beleuchtet.

# Abstract

Statistical inference and hypothesis testing in the framework of several different models for discrete-valued stochastic processes is considered. In the case of integer-valued autoregressive (INAR) processes of the first order, underlying stochastic properties can be utilized to derive appropriate test statistics for certain scenarios. Three different tests are introduced, evaluating the deviation of empirical measures of dispersion, generalized autocovariance and skewness of the data set from the theoretical value using the explicitly calculated asymptotic distribution of the associated estimators. For each of these test statistics, simulation studies as well as real data applications are provided, showcasing the performance in small sample sizes.

In a more general setting, a different approach focusing on generating functions instead of moment-based estimators is pursued. The asymptotic characteristics of the resultant test statistic are derived for a very general class of Markovian models satisfying a drift condition. Furthermore, a nonparametric estimator of the stationary distribution is shown to obey a functional central limit theorem. After revealing the connections linking this approach with several methods of the preceding chapters, a simulation study highlights the strong performance of the tests in real data applications with a small number of observations.

As a further topic, one specific instance of a nonparametric estimator of the service time distribution of a discrete time $GI/G/\infty$ queueing system is presented, where the given information is assumed to be limited to the counts of arriving and departing customers of the queue. It is shown that this so-called *sequence of differences estimator* obeys a functional central limit theorem on an appropriately chosen underlying sequence space. Finally, a moving block bootstrap method is proposed and the theoretical features of this approach are investigated.

# Contents

# 1 Introduction

In this thesis, we examine properties of discrete-valued stochastic processes of various forms. As their name suggests, these models constitute a variation of stochastic processes, in which the time-dependent random variables are assumed to take on only nonnegative integer values, i.e., *counts*. In order to demonstrate some of the defining features, let us present an instructive example of such a data set. In Freeland (1998), the time series plotted in 1.1 is reported, the data originated from the Worker's Compensation Board (WCB) of British Columbia. This organization is tasked with managing the disability benefits for the workers in the heavy manufacturing industry of British Columbia. The data represents the counts of a special subclass of short-term wage loss claims. More precisely, the counts correspond to the number of workers each respective month who received short-term disability benefits due to burn injuries.

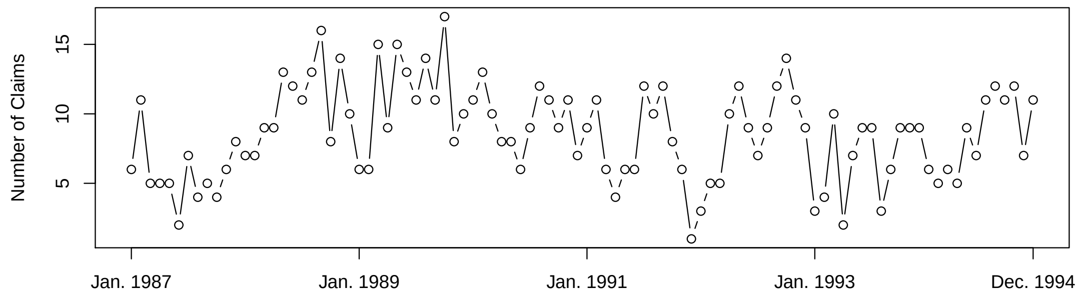


Figure 1.1: Plot of monthly claim counts of workers receiving short-term disability benefits after sustaining burn injuries, as reported by the WCB British Columbia.

For example, the first two data points show that in January 1987, 6 workers collected short-term disability benefits and in the following month, this number rose to 11. Since burn related injuries take some time to heal, it seems sensible to assume that some of the workers collecting benefits in February were among those counted in January. Thus, the count data should exhibit some form of dependency over time, i.e., it is unlikely that the counts are uncorrelated or even independent. Indeed, the empirical autocorrelation function (ACF) and the empirical partial autocorrelation function (PACF), plotted in Figure 1.2, uncover a significant deviation of the data from the null hypothesis of independence (dashed lines). For more details on both of these function, see Section 2.6.

We are thus faced with the problem of finding a suitable mathematical model for a sequence of counts which allows for serial dependence while its values remain in the range of the nonnegative integers $\mathbb{N}_0$. The approaches under consideration in this thesis can be roughly filed under two categories: The first approach is an *in*teger-valued analogon to the common *a*uto*r*egressive model, subsumed by the acronym INAR, the second employs elements of (discrete-time) queueing theory for modeling purposes.

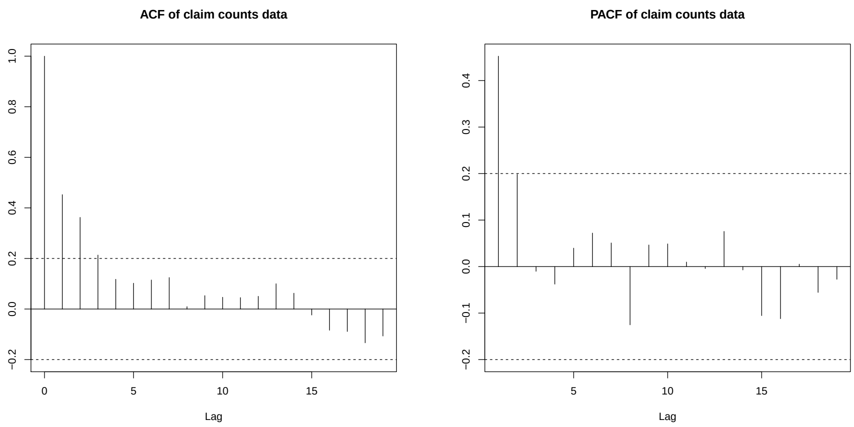


Figure 1.2: Claim counts for Figure 1.1: Plot of ACF and PACF.

## 1.1 Integer-Valued Autoregressive Processes

The first approach is motivated by finding similarities in the dependency on the past of the data at hand, $y_1, \dots, y_{96}$ say, with well-known stochastic processes. Looking at the ACF and the PACF in Figure 1.2, the ACF decays exponentially, and the deviation of the PACF from zero is only significant at lag 1. This pattern resembles that of the well known (continuous) autoregressive process of first order (AR(1)) process (see Section 2.6), thus a first approach would be to fit the data of Figure 1.1 to such a model.

However, there are major differences between the data plot in Figure 1.1 and that of an AR(1) process. The realizations $y_1, \dots, y_{96}$ are nonnegative, and they are integer valued. Both of these characteristics stem from the nature of the data and should be incorporated in an appropriate model. This motivates the introduction of the following integer-valued autoregressive process: Starting with an AR(1) process $(Y_t)_{t\in\mathbb{Z}}$ satisfying $Y_t = \alpha Y_{t-1} + \epsilon_t$, we need to first make sure that this process is nonnegative and that the realizations are integer-valued, so we assume that $\epsilon_t \in \mathbb{N}_0$ for all $t \in \mathbb{Z}$. However, the multiplication with a value of $\alpha \notin \mathbb{Z}$ would lead to values outside the integers, yet it is well known that an AR(1) process is only stationary if $\alpha \in [0, 1)$, where $\alpha = 0$ pertains to the trivial case of i.i.d. random variables.

Hence, we replace the multiplication with the parameter $\alpha \in (0, 1)$ with a different operation, the binomial thinning operator of Steutel and Van Harn (1979): If $Y$ is a discrete random variable with range $\mathbb{N}_0$ and if $\alpha \in (0, 1)$, then the random variable $\alpha \circ X := \sum_{i=1}^{Y} Z_i$ is said to arise from $Y$ by *binomial thinning*, and the $Z_i$ are referred to as the *counting series*, if they are independent and identically distributed (i.i.d.) Bernoulli random variables with $\mathbb{P}(Z_i = 1) = \alpha$, which are assumed to be independent of $Y$. So each $Z$ satisfies $Z \sim \text{Bin}(1, \alpha)$, and $\alpha \circ Y \sim \text{Bin}(Y, \alpha)$, where $\text{Bin}(n, \pi)$ abbreviates the binomial distribution with parameters $n \in \mathbb{N}$ and $\pi \in (0, 1)$. Using the random operator "$\circ$", let us define the integer-valued autoregressive process of first order (INAR(1)) in the following way.

**Definition 1.1.1** (INAR(1) Model)**.** *Let* $(\epsilon_t)_{t\in\mathbb{Z}}$ *be an i.i.d. process with range* $\mathbb{N}_0$*, let* $\mathrm{Var}[\epsilon_t] < \infty$ *and let* $\alpha \in (0,1)$*. A process* $(Y_t)_{t\in\mathbb{Z}}$*, which follows the recursion*

$$Y_t = \alpha \circ Y_{t-1} + \epsilon_t \quad \textit{for all } t \in \mathbb{Z} \tag{1.1}$$

*is said to be an* INAR(1) process *if all thinning operations are performed independently of each other and of* $(\epsilon_t)_{\mathbb{Z}}$*, and if the thinning operations at each time* $t$ *as well as* $\epsilon_t$ *are independent of* $(Y_s)_{s<t}$*.*

Let us give an overview of existing literature concerned with the process of Definition 1.1.1 and its extensions. The INAR(1) process owes its denomination to the contribution Al-Osh and Alzaid (1987), it was previously suggested as a discretized version of an AR(1) process in McKenzie (1985). Note, however, that very similar processes were already studied earlier in the context of branching processes, see, e.g., Pakes (1971), as the recursion (1.1) may be interpreted as the result of a special branching process, in which each member of the population at time $t-1$ gives birth to exactly one offspring with probability $\alpha$ and does not reproduce with probability $1-\alpha$. The older generation then dies out, and an external migration component, distributed as $\epsilon_t$, enters the population. This connection to branching process theory is employed in some arguments in this thesis, see for instance the proof of Theorem 4.2.6.

The "$\circ$" operator used in Definition 1.1.1 stems from Steutel and Van Harn (1979), the article which also uncovered the connection of INAR(1) processes and infinite divisibility of random variables, see Theorem 2.3.5. This connection is part of the reasons to consider the special case of Compound Poisson INAR(1) models in Chapters 4 and 5 in such detail, note that the Compound Poisson distributions correspond exactly to the infinitely divisible distributions on $\mathbb{N}_0$, cf. Theorem 2.3.3.

In the literature, the interest in INAR processes continued for some time after their introduction. Further results were discussed in Alzaid and Al-Osh (1988) and extension to both higher order INAR processes (see Alzaid and Al-Osh (1990) and Du and Li (1991) as well as Chapter 7) as well as integer-valued ARMA models (see McKenzie (1988)) were introduced. After a short hiatus, the interest began picking up again in the 2000s and the stream of contributions has not dried up since. Especially the application to real data sets has been discussed in a variety of contexts. The dissertation Freeland (1998) and the article Freeland and McCabe (2004) presented data obtained from short term wage loss benefits, one of these time series is presented in Figure 1.1. For a further discussion of these time series, the reader is referred to Section 5.1.3 and Section 6.4.3. In Jung et al. (2005), monthly strike counts published by the U.S. Bureau of Labor Statistics are fitted to an INAR(1) model, for further details we refer to Section 6.4. As a last example, the contribution Weiß (2008) considers this model in the context of download counts of a certain software package, where the number of downloads is counted daily. This list is far from being complete, yet it conveys the message of the versatile applicability of the model (1.1) quite well.

The basic INAR(1) model has been extended in various fashions to comply with additional considerations. In Zheng et al. (2007), the Random Coefficient INAR(1) model is introduced, which allows the parameter $\alpha$ in (1.1) to be a random variable on its own. Another possibility for generalizations consists in allowing for different thinning operations to replace the "$\circ$" operator, as discussed in Weiß (2008). The higher-order case was also considered in Dion et al. (1995), which emphasized the connection to branching process theory. For details on the approach and the results of the latter reference, the reader is referred to Section 7.2.2. Another topic of discussion is the so-called dispersion of the marginal distribution of the resulting processes as can be seen in Weiß (2009) (for overdispersion) and Weiß (2013) (for underdispersion). This concept is taken up again in Chapters 4 and 5.

In this thesis, a multitude of new research results for INAR(1) processes is presented. One recurring theme is the following question: Having identified the INAR(1) process of Definition 1.1.1 with a certain parametric choice such as $\epsilon_t \sim \text{Poi}(\lambda)$ as a *possible* model for a given data set $\{y_1, \ldots, y_T\}$, is it also an *appropriate* model? No less than four different statistical tests designed to answer this question (or variations of it) are put forward in this thesis. The basic approach remains the same in all four cases: First, an appropriate relation holding under the null hypothesis is identified. Then, an empirical test statistic is constructed which is able to detect deviations from the theoretical value. Finally, the asymptotic behavior of the test statistic is derived, allowing us to evaluate the significance of deviations of the test statistic.

Even though they share one basic approach, the mathematical tools involved in the derivations vary. The first three tests involve relations of joint moments of the processes, and the calculation of these expressions is facilitated by using joint cumulants as shown in Chapter 5. The necessary central limit theorems are derived via the classical notion of strong mixing. The last test of Chapter 6 stands alone, it is also more general than its predecessors. It utilizes empirical generating functions of random variables and is thus able to serve as a goodness-of-fit test for a very large class of processes. The derivation of the necessary central limit theorem does not rely on mixing, but on a more general result for ergodic processes. Furthermore, the convergence in distribution in this case is of a functional type, leading to a much more intricate structure of the proofs.

## 1.2 Discrete-Time Queueing Processes

A second approach for modeling data as given in Figure 1.1 considers discrete-time queueing models for the underlying process. The history of queueing theory dates back to Erlang (1909) and has since been established as a widely used approach in the assessment of real-time events. The main idea behind a single node queueing model is easily explained with the help of a sketch.

There is a certain arrival stream of customers coming from the outside, denoted by $A(t)$. Whenever a customer arrives and if at least one server out of a total of $N$ servers is available, the customer begins his service, the length of which is distributed according to some service time distribution $G$. Should all servers be busy at the time of arrival, the customer is placed in the queue and remains there until he receives service. Whenever a customer has finished his service, he leaves the queue via the departure process

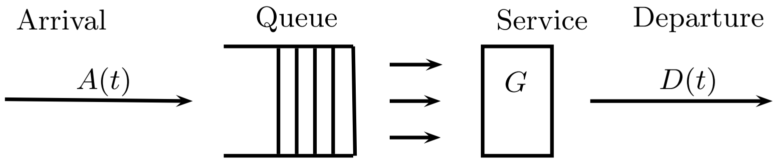


Figure 1.3: Sketch of a General Queueing Model

$D(t)$. Such a queueing model can be modeled both in continuous time and with discrete time slots, in this thesis we will always assume the time to be discrete. Concerning notation, the main characteristics of a queueing model are usually summarized in the so-called *Kendall notation* of the form $A/B/c/d$. The first entry "$A$" governs the arrival distribution, the entry "$B$" the service time distribution and the third, "$c$" reports the number of servers. The entry "$d$" denotes the size of the waiting room. This notation may be extended to include more information, for instance, the discipline of the queue (e.g., "first in, first out" or FIFO) can be added. However, such ramifications will not be necessary in this thesis.

Let us discuss how the data of Figure 1.1 fits into the framework of a discrete-time queueuing system: the time slots are the elapsing months, the arriving customers each month represent the number of workers who suffered an accident involving burn injuries in the course of the current month and now collect short-term wage loss benefits for the first time. The service time corresponds to the duration of the healing period, i.e., the time span in which the worker is ineligible for work due to his injuries. In this scenario, the number of servers is infinite, as there is no (theoretical) limit to the number of workers collecting short-term wage loss benefits. Hence, we are faced with a discrete-time queue in which the arrival is independent, yet general; the service time is general, as we have no preconception about the healing time of a worker and the number of servers is infinite. In Kendall's notation, this is a $GI/G/\infty$-queue.

The nonparametric estimation of the service time distribution $G$ within a discrete-time $GI/G/\infty$ queue has been previously considered in a number of publications. In the continuous time case, which can always serve as an orientation, there are the early contributions Brillinger (1974) and Brown (1970), where estimation techniques based on observations of the arrival and departure process is considered. Furthermore, the articles Bingham et al. (1989) and Hall and Park (2004) discuss the nonparametric analysis of the system based on consecutive sequences of busy and idle periods, a related concept and two additional approaches are the topic of Bingham and Pitts (1999). For a detailed literature overview, the reader is referred to Wichelhaus and Langrock (2012).

In discrete time, literature is more scant, notable exceptions are given by Pickands and Stine (1997) where knowledge about the queue length is assumed and Edelmann and Wichelhaus (2014). The results of the latter contribution are discussed in detail below, as they build the vantage point for the analysis of Chapter 3.

In particular, the idea of estimating the sequence of differences as a device to obtain an estimation of the service time distribution was first put forward in Brown (1970), refined in Blanghaps et al. (2013) and applied to the discrete time case in Edelmann and Wichelhaus (2014). The idea consist basically in circumventing the main matchmaking problem, i.e., that departures can not be matched to their respective arrivals, by estimating a different distribution first, for which simply (and falsely) any departure is assumed to have been caused by the latest possible arrival. Quite surprisingly, this distribution stands in a very simple relation to the sought after service time distribution. From a mathematical standpoint, both of these distributions may be represented as functionals on the general space of sequences $\mathbb{R}^{\mathbb{N}}$, and the main contribution of Chapter 3 is the proof of the functional convergence of appropriately chosen estimators of these distributions.

## 1.3 Bibliographic Notes

Large parts of this thesis are based on a number of published articles of the author in refereed journals. Two of these articles were co-authored by Prof. Dr. Christian Weiß: Schweer and Weiß (2014), Schweer and Weiß (2015). Another article, Schweer and Wichelhaus (2015a), was written in conjunction with Dr. Cornelia Wichelhaus. The published contributions Schweer (2015a) and Schweer (2015b), the latter of which appeared in a refereed conference proceedings volume, constitute solo efforts of the author. In order to clarify the origin of the results, the source of each theorem, proposition, lemma etc. is cited. If a theorem or a corollary remains unmarked, it states a new and unpublished result. Unmarked lemmata may state either new results or a well-known assertion, the difference should be clear from the context. In particular, unpublished work can be found in Section 5.3, Section 4.2.4 and Section 7.2.

Concerning notation, this thesis follows the usual mathematical conventions, e.g., denoting the normal distribution with mean $a$ and variance $b$ by $\mathcal{N}(a, b)$, and so forth. Furthermore, we set $\mathbb{N} = \{1, 2, 3, \dots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Denotations exceeding the scope of the "usual" are defined in the text and listed in the appendix of this thesis for easy reference. In the same place, the reader can find a list of the acronyms and an index of important keywords used in the text.

# 2 Mathematical Prerequisites

This chapter gathers together for easy reference a number of useful mathematical concepts used in this thesis. As the focus is put on the collection of established results, most of the proofs are omitted. They are included either if parts of the proof are used in this thesis or if they are slight generalizations of existing results.

The topics of the particular sections vary, beginning with the function-analytic aspects of sequence spaces. These abstract concepts provide the appropriate setting for parts of the analysis in Chapter 3. Next, we discuss some properties of (joint) cumulants, which simplify the explicit calculation of joint moments of INAR(1) processes significantly, see Chapters 4 and 5. Characteristics of Compound Poisson distributions are collected in the next section, these will turn out to be a natural candidate for consideration in the framework of INAR(1) processes. The fourth part of this chapter is concerned with the time-reversibility of stochastic processes, a concept which appears at numerous times throughout this thesis.

The last two sections collect some well-known results for dependent data, the first introduces the classical autocorrelation function and partial autocorrelation function. The second provides two different central limit theorems for dependent data due to Ibragimov (1962) and Billingsley (1999), both of which are employed at multiple times throughout this thesis.

## 2.1 Sequence Spaces

For the nonparametric estimation of a probability distribution on the real numbers, one usually considers convergence of these estimators in the so-called Skorokhod space $D[-\infty, \infty]$. In this thesis, however, the considered distributions are of a discrete type and it will therefore turn out to be convenient to consider convergence on sequence spaces instead. One such example which we will use extensively is the Banach space

$$c_0 = \left\{ x = (x_1, x_2, \dots) \in \mathbb{R}^{\mathbb{N}} \;\middle|\; \lim_{k \to \infty} x_k = 0 \right\}. \tag{2.1}$$

We equip this space with the norm $\|x\|_{c_0} = \sup_{k \in \mathbb{N}} |x_k|$. This space is used for proving functional central limit theorems of empirical distribution functions in Henze (1996), the following exposition parallels that of Section 2 of this article.

The $\sigma$-algebra of Borel sets of $c_0$ will be denoted by $\mathcal{B}'$, it is generated by the $\epsilon$-balls $S(x, \epsilon) := \{y \in c_0 \mid \|x - y\|_{c_0} < \epsilon\}$, where $x \in c_0$ and $\epsilon > 0$. Comparing this with the smallest $\sigma$-algebra on $c_0$ such that the projections $x \mapsto \pi_k \circ x := x_k$, $k \in \mathbb{N}, x \in c_0$ are measurable, denoted by $\mathcal{B}$, it is easily seen that $\mathcal{B} = \mathcal{B}'$. Let $\mathcal{X}_n$ be a mapping from the underlying sample space $\Omega$ (with associated $\sigma$-algebra $\mathcal{A}$) into $c_0$ for which $\pi_k \circ \mathcal{X}_n =: X_{n,k}$ is a random variable. It follows that $\mathcal{X}_n$ is $\mathcal{A}$-$\mathcal{B}$-measurable, implying that $\mathbb{P} \circ (\mathcal{X}_n)^{-1}$ is a Borel probability measure on $c_0$. In order to show convergence in distribution of a sequence to a random element in $c_0$, it is well known that we need to prove the weak convergence of finite-dimensional distributions and verify that the sequence is tight (see, e.g., Billingsley (1999)).

Recalling that a family of distributions $\Pi$ is *tight* if for every $\epsilon > 0$ there exists a compact set $K$ such that $\mathbb{P}(K) > 1 - \epsilon$ for every $\mathbb{P} \in \Pi$, it is clear that we need to classify the compact sets in $c_0$. It will turn out to be convenient to consider *relatively compact* sets $A$ of $c_0$, for which the closure, denoted by $A^-$, is compact. An $\epsilon$-cover is a cover of the space consisting of sets of diameter less than $\epsilon$, and a metric space is called *totally bounded* if it admits a finite $\epsilon$-cover for every $\epsilon > 0$. The relation between these concepts is the following (cf. Theorem M5 in Billingsley (1999)): Since $c_0$ is a Banach space, the closure of any set is complete, and thus every totally bounded set is relatively compact and vice versa. Let us formulate the key lemma in this context:

**Lemma 2.1.1** (Hanche-Olsen and Holden (2010), Lemma 1)**.** *Let $(X, d)$ be a metric space and assume that for every $\epsilon > 0$, there exists some $\delta > 0$ a metric space $(W, d')$ and a mapping $\Phi : X \to W$ such that $\Phi(X)$ is totally bounded. Furthermore, assume that whenever $x, y \in X$ are chosen such that $d'(\Phi(x), \Phi(y)) < \delta$, then $d(x, y) < \epsilon$. Then $X$ is totally bounded.*

This result allows for a very simple proof of the following assertion. Results of this kind are usually subsumed under the name Arzelà-Ascoli theorem in honor of Cesare Arzelà and Giulio Ascoli.

**Lemma 2.1.2** (Cp. Henze (1996), Sect.2)**.** *A set $A \subset c_0$ is totally bounded if and only if*

*(i) A is pointwise bounded, i.e., it holds that* $\sup_{x \in A} \|x\|_{c_0} < \infty$.

*(ii) for all $\epsilon > 0$ there exists an $N$ such that*

$$\sup_{x \in A} \sup_{k \geq N} |x_k| < \epsilon.$$

*Proof.* Let us first assume that $A$ is totally bounded. The existence of a finite $\epsilon$-cover for any $\epsilon > 0$ implies the uniform boundedness of $A$ and therefore the pointwise boundedness of $(i)$. Now, let $\epsilon > 0$ and choose a corresponding cover $\{U(1), \ldots, U(R)\}$ with center points $y(j) \in U(j)$. Now choose $N$ large enough so that

$$\sup_{k \geq N} |y(j)_k| < \epsilon \quad \text{for } j = 1, \ldots, R,$$

which is possible since $y(j) \in c_0$. Letting $x \in A$ be arbitrary, it follows that there exists a $y(j)$ with $\|x - y(j)\|_{c_0} < \epsilon$. Using this, we have

$$\sup_{k \geq N} |x_k| \leq \sup_{k \geq N} |x_k - y(j)_k| + \sup_{k \geq N} |y(j)_k| \leq 2\epsilon,$$

proving condition (ii). Now, assume (i) and (ii) hold. Let $\epsilon > 0$ and choose $N$ as in (ii). Consider the map $\Phi : A \to \mathbb{R}^N$, which maps $x$ to the vector $(x_1, \ldots, x_N)$. Employing the uniform norm $\| \cdot \|_\infty$ on $\mathbb{R}^N$, by (i), $\Phi(A)$ is totally bounded. Now, assuming that $x, y \in A$ such that

$$\|\Phi(x) - \Phi(y)\|_{\mathbb{R}^N} = \sup_{k \leq N} |x_k - y_k| < \epsilon,$$

we find that

$$\|x-y\|_{c_0} \leq \sup_{k\geq N} |x_k - y_k| + \sup_{k\leq N} |x_k - y_k| < 2\epsilon,$$

and an appeal to Lemma 2.1.1 concludes the proof. □

This can be translated into the following condition for tightness in the measurable space $(c_0, \mathcal{B})$. We repeat the proof shown in Henze (1996) on the one hand to highlight the usefulness of the considerations above and on the other hand to clear the pathway towards possible generalizations.

**Lemma 2.1.3** (Henze (1996), Lemma 2.1)**.** *Let $\mathcal{X}_n = (X_{n,k})_{k\in\mathbb{N}}$ be a sequence of random elements of $(c_0, \mathcal{B})$. Then $\mathcal{X}_n$ is tight if the following two conditions hold:*

*(i) For each positive $\delta$ and $l \in \mathbb{N}$ there is a finite constant $M$, such that*

$$\mathbb{P}\left(|X_{n,l}| \leq M\right) \geq 1-\delta, \quad n \geq 1. \tag{2.2}$$

*(ii) For each positive numbers $\delta, \epsilon$ there exist integers $n_0$ and $l_0$ such that*

$$\mathbb{P}\left(\sup_{k\geq l_0} |X_{n,k}| > \epsilon\right) \leq \delta \text{ for all } n \geq n_0. \tag{2.3}$$

*Proof.* The necessity of the conditions is easily seen. Suppose (i) and (ii) hold and let $\epsilon > 0$. Let us define the sets

$$A_j := \left\{ x \in c_0 \;\middle|\; \sup_{k\geq l(j)} |x_k| \leq \frac{1}{j} \right\},$$

where $l(j)$ is an integer chosen (depending on $j$ and $\epsilon$) such that $\mathcal{Q}_n(A_j) \geq 1-\epsilon\cdot 2^{-(j+1)}$ for all $n$. Condition (i) ensures the existence of constants $M_1, \ldots, M_{l(1)-1}$ such that $\mathcal{Q}_n(B_k) \geq 1-\epsilon/(2(l(1)-1))$ for all $n$, where $B_k := \{x \in c_0 | \; |x_k| \leq M_k\}$ for all $k = 1, \ldots, l(1)-1$. Now, we define

$$A := B_1 \cap B_2 \cap \cdots \cap B_{l(1)-1} \cap \bigcap_{j=1}^{\infty} A_j,$$

and with Lemma 2.1.2 it immediately follows from this construction that the closure $K$ of $A$ is compact. We calculate

$$\mathcal{Q}_n(K) \geq 1-\mathcal{Q}_n(A^C) \geq \sum_{i=1}^{l(1)-1} (1-\mathcal{Q}_n(B_i)) + \sum_{i=1}^{\infty} (1-\mathcal{Q}_n(A_i)) \geq 1-\epsilon,$$

concluding the proof. □

These considerations extend *mutatis mutandis* to a similar sequence space $\ell_1$, the space of all absolutely summable sequences, i.e., all sequences $(x_k)_{k\in\mathbb{N}}$ with $\sum_{k=1}^{\infty}|x_k| < \infty$. More precisely, it is easily seen that the results also hold in the space

$$(\ell_1)^2 := \left\{(x,y) \in \mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}} \;\middle|\; \sum_{k=1}^{\infty}(|x_k| + |y_k|) < \infty\right\},$$

equipped with the norm $\|(x,y)\|_{(\ell_1)^2} = \sum_{k=1}^{\infty}(|x_k| + |y_k|)$. The corresponding Borel $\sigma$-algebra will be denoted by $\mathcal{B}_1$. We find the following analogon of Lemma 2.1.3:

**Lemma 2.1.4.** *A sequence $\{\mathcal{Q}_n\}_{n\geq 1}$ of probability measures on $((\ell_1)^2, \mathcal{B}_1)$ is tight if and only if these two conditions hold:*

*(i) For each positive $\delta$ and each $l_1, l_2 \in \mathbb{N}$, there exists a finite constant $M$ such that*

$$\mathcal{Q}_n\left(\left\{x \in (\ell_1)^2 \;\mid\; |x_1(l_1)\,| + |x_2(l_2)| \leq M\right\}\right) \geq 1 - \delta.$$

*(ii) For each positive $\delta, \eta > 0$, there exists a $l_0 \in \mathbb{N}$ such that*

$$\mathcal{Q}_n\left(\left\{x \in (\ell_1)^2 \;\middle|\; \sum_{k\geq l_0}(|x_1(k)| + |x_2(k)|) \leq \eta\right\}\right) \geq 1 - \delta.$$

## 2.2 Cumulants

The following exposition resembles that of Section 2.3 in Brillinger (1981) closely. Let us consider a $r$-variate random variable $(Y_1, \ldots, Y_r)$ with $\mathbb{E}[Y_i^r] < \infty$ for all $i \in \{1, \ldots, r\}$. We define the $r$-th order joint cumulant $\mathrm{cum}(X_1, \ldots, X_r)$ as follows:

$$\mathrm{cum}(Y_1, \ldots, Y_r) = \sum_{\pi\in\Pi_r}(-1)^{|\pi|-1}(|\pi|-1)!\prod_{i=1}^{|\pi|}\mathbb{E}\left[\prod_{l\in B_i(\pi)} Y_l\right], \tag{2.4}$$

where $\Pi_r$ is the set of all possible partitions $\pi$ of the set $\{1, \ldots, r\}$, $|\pi|$ denotes the number of blocks in a partition $\pi$ and $B_i(\pi)$ denotes the $i$-th block in said partition, i.e., $\bigcup_i B_i(\pi) = \{1, \ldots, r\}$. A special case of (2.4) is given for $Y_i = Y$, then the definition reduces to that of the $r$th order cumulant of a random variable $Y$, which we will denote by $\kappa_r(X)$ throughout this thesis. We now summarize the following properties:

**Lemma 2.2.1** (Brillinger (1981), Theorem 2.3.1)**.** *Let $(Y_1, \ldots, Y_r)$ be a $r$-variate random variable with $\mathbb{E}(Y_i^r) < \infty$ for all $i \in \{1, \ldots, r\}$. Then*

*(i) If $Y_i = Y$ for all $i \in \{1, \ldots, r\}$, then* $\mathrm{cum}(Y, \ldots, Y) = \kappa_r(Y)$.

*(ii) If any group of $Y_i$'s is independent of the remaining $Y_j$'s, then* $\mathrm{cum}(Y_1, \ldots, Y_r) = 0$.

*(iii) The cumulant function is additive, i.e.,* $\mathrm{cum}(Y_1 + Z_1, \ldots, Y_r) = \mathrm{cum}(Y_1, \ldots, Y_r) + \mathrm{cum}(Z_1, \ldots, Y_r)$.

*(iv)* $\operatorname{cum}(a_1Y_1, \ldots, a_rY_r) = a_1 \cdots a_r \operatorname{cum}(Y_1, \ldots, Y_r)$ *for* $a_1, \ldots, a_r$ *constant.*

*(v)* *If* $r \geq 2$*, then the* $r$*-th order joint cumulant is shift-invariant, i.e., for constants* $c_1, \ldots, c_r \in \mathbb{R}$ *it holds that* $\operatorname{cum}(Y_1 + c_1, \ldots, Y_r + c_r) = \operatorname{cum}(Y_1, \ldots, Y_r)$.

*(vi)* *For any* $r \in \mathbb{N}$,

$$\mathbb{E}\left[Y_1 \cdots Y_r\right] = \sum_{\pi \in \Pi_r} \prod_{i=1}^{|\pi|} \operatorname{cum}(Y_{B_i(\pi)}),$$

*where* $\Pi_r$ *and* $B_i(\pi)$ *are as in* (2.4) *and where* $Y_{B_i(\pi)} = (Y_{b_i(\pi,1)}, \ldots, Y_{b_i(\pi,p)})$*, with* $B_i(\pi) = \{b_i(\pi, 1), \ldots, b_i(\pi, p)\}$.

*(vii)* *Let* $X_{1,i}, X_{2,i}, \ldots, X_{r,i}$ *be sequences of nonnegative random variables, such that* $X_{i,j}$ *and* $X_{k,l}$ *are independent for* $j \neq l$ *for all* $i, k \in \{1, \ldots, r\}$ *and such that* $\mathbb{E}[(\sum_{i=0}^{\infty} X_{j,i})^r] < \infty$ *for* $j \in \{1, \ldots, r\}$*. Then*

$$\operatorname{cum}\left(\sum_{i=0}^{\infty} X_{1,i}, \sum_{i=0}^{\infty} X_{2,i}, \ldots, \sum_{i=0}^{\infty} X_{r,i}\right) = \sum_{i=0}^{\infty} \operatorname{cum}\left(X_{1,i}, X_{2,i}, \ldots, X_{r,i}\right).$$

*Proof.* For (i) through (v), we refer to (Brillinger, 1981, Theorem 2.3.1), relation (vi) is well-known. Concerning (vii), as the condition $\mathbb{E}[(\sum_{i=0}^{\infty} X_{j,i})^r] < \infty$ ensures that the expression is well-defined (Brillinger, 1981, Definition 2.3.1), we prove this statement via application of the defining equation (2.4) of joint cumulants. In order to allow for the changing of the order of integration (i.e., taking the mean) and summation, we apply Lesbesgue's monotone convergence theorem. Since the $X_{ji}$'s are nonnegative and since the arising expectations have an upper bound in $\max_{1 \leq j \leq r} \{\mathbb{E}\left[(\sum_{i=0}^{\infty} X_{ji})^r\right]\} < \infty$, this theorem is applicable. We first find with (iii) that

$$\begin{aligned}\operatorname{cum}\left(\sum_{i_1=0}^{\infty} X_{1,i}, \sum_{i_2=0}^{\infty} X_{2,i}, \ldots, \sum_{i_r=0}^{\infty} X_{r,i}\right) &= \sum_{i_1=0}^{\infty} \operatorname{cum}\left(X_{1,i_1}, \sum_{i_2=0}^{\infty} X_{2,i}, \ldots, \sum_{i_r=0}^{\infty} X_{r,i}\right) \\ &= \ldots = \sum_{i_1=0}^{\infty} \cdots \sum_{i_r=0}^{\infty} \operatorname{cum}\left(X_{1,i_1}, X_{2,i_2}, \ldots, X_{r,i_r}\right),\end{aligned}$$

which, using (ii), concludes the proof. □

To give a better feeling for what cumulants do, notice that $\operatorname{cum}(X, X) = \operatorname{Var}(X)$ as well as $\operatorname{cum}(X, Y) = \operatorname{Cov}(X, Y)$ provided the respective moments exist. Thus, cumulants are close relatives of joint moments with rather nice mathematical aspects such as multilinearity. This makes their application in situations where the random variables in question carry a lot of structure quite advantageous in comparison to the calculation of raw joint moments. This is used, e.g., in the proof of Lemma 5.2.2 below.

Let us further record that the concept of overdispersion, which will be of great importance throughout this thesis, can also be stated in terms of cumulants. The *index of* *dispersion* of a random variable $X$ is defined as

$$I_X := \frac{\mathrm{Var}(X)}{\mathbb{E}[X]}. \tag{2.5}$$

If $I_X = 1$ (for instance in Poisson distributions), $X$ is *equidispersed*, if it is larger than 1 then $X$ is *overdispersed*. For $I_X < 1$ the term *underdispersed* is used. In terms of cumulants, this may be expressed via the fraction $\kappa_2(X)/\kappa(X)$.

After this side note, let us continue with the law of total cumulance. This was introduced in Brillinger (1969) and is restated below. It extends the law of total expectation $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$ and law of total variance $\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathbb{E}[X|Y])$ (where $X, Y$ are random variables on the same probability space with finite moments of appropriate order) to cumulants:

**Theorem 2.2.2** (Brillinger (1969), Theorem 1)**.** *Let $(Y_1, \ldots, Y_r)$ be a $r$-variate random variable with $\mathbb{E}(Y_i^r) < \infty$ for all $i \in \{1, \ldots, r\}$ and let $X$ be a random variable defined on the same probability space as the $Y_i$'s. Then*

$$\mathrm{cum}\left(Y_1, \ldots, Y_r\right) = \sum_{\pi \in \Pi_r} \mathrm{cum}\left(\mathrm{cum}(Y_{B_1(\pi)}|X), \ldots, \mathrm{cum}\left(Y_{B_p(\pi)}|X\right)\right),$$

*where $\Pi_r$, $B_i(\pi)$ and $Y_{B_i(\pi)}$ are as in Lemma 2.2.1 (vi).*

## 2.3 Compound Poisson Distributions

The following exposition is very closely modeled after the corresponding section of the Appendix of Schweer and Weiß (2014). For the Compound Poisson distribution, we adapt the notations and definitions from Chapter XII in Feller (1968).

**Definition 2.3.1** (Compound Poisson Distribution)**.** *Let $X_1, X_2, \ldots$ be i.i.d. random variables with range $\mathbb{N}$ and probability generating function (pgf) $H(z)$, the* compounding distribution*, where $\nu := \deg(H(z))$. Let $N$ be Poisson distributed with mean $\lambda > 0$, i.e., $N \sim \mathrm{Poi}(\lambda)$, independently of $X_1, X_2, \ldots$ Then $\epsilon := X_1 + \ldots + X_N$ is* Compound Poisson distributed*, denoted by $\epsilon \sim \mathrm{ComPoi}_\nu(\lambda, H)$. The pgf of $\epsilon$ is given by*

$$\mathrm{pgf}_\epsilon(z) = \exp\left(\lambda(H(z) - 1)\right). \tag{2.6}$$

The distribution of Definition 2.3.1 has also been referred to as Poisson-stopped sum distribution, stuttering Poisson distribution, multiple Poisson distribution (Johnson et al., 2005, Sections 4.11 and 9.3), and as extended Poisson distribution of order $\nu$ if $\nu < \infty$ in Aki (1985). The $\mathrm{ComPoi}_\nu$-distribution includes several well-known distributions as a special case. It is equidispersed if and only if $\nu = 1$, and overdispersed otherwise. Furthermore, the Compound Poisson distributions are closely linked to the infinitely divisible distributions, defined as follows.

**Definition 2.3.2** (Infinite Divisibility)**.** *Let $H$ be a probability distribution on $\mathbb{N}_0$ with $\mathrm{pgf}_X(z) = H(z)$. Then $H$ is* infinitely divisible*, if for each positive integer $n$, the $n$-th root, $\sqrt[n]{H(z)}$, is a pgf again.*

Whereas this definition uses pgfs for its formulation, one could equivalently define: $H$ is infinitely divisible if for each $n \in \mathbb{N}$ there exist $n$ i.i.d. random variables $X_{1n}, \dots, X_{nn}$ such that $X_{1n} + \cdots + X_{nn} \sim H$. The link between such distributions and Compound Poisson distributions is given by the next result.

**Theorem 2.3.3** (Feller (1968), Sect. XII.2)**.** *Let $G$ be a probability distribution on $\mathbb{N}_0$ with pgf $G(z)$. Then the following statements are equivalent:*

*(i)* *$G$ is infinitely divisible,*

*(ii)* *there exist $\lambda > 0$ and a pgf $H(z)$ such that $G \sim \mathrm{ComPoi}(\lambda, H)$,*

*(iii)* *$G(1) = 1$ and*

$$\log \frac{G(z)}{G(0)} = \sum_{i=1}^{\infty} a_i z^i, \quad \textit{where } a_k \geq 0, \sum_{i=1}^{\infty} a_k < \infty.$$

There is another classification for these distributions that we want to mention here. Let us first introduce the notion of discrete self-decomposability.

**Definition 2.3.4** (Discrete Self-Decomposability)**.** *Let $H$ be a probability distribution on $\mathbb{N}_0$ with pgf $H(z)$. Then $H$ is* discrete self-decomposable (DSD) *, if for each $\alpha \in (0, 1)$ there exists a pgf $H_\alpha(z)$ such that*

$$H(z) = H(1 - \alpha + \alpha z) H_\alpha(z). \tag{2.7}$$

In Steutel and Van Harn (1979), the following deep result is shown.

**Theorem 2.3.5** (Steutel and Van Harn (1979), Theorem 2.2)**.** *Let $G$ be a probability distribution on $\mathbb{N}_0$ with pgf $G(z)$. Then $G$ is discrete self-decomposable if and only if $G$ is infinitely divisible with $G \sim \mathrm{ComPoi}(\lambda, H)$, where the sequence $(n \cdot h_n)_{n \in \mathbb{N}}$ is nonincreasing.*

*Proof.* The entire assertion is contained in the cited reference, we only show that for the canonical measure $(r_n)_{n \in \mathbb{N}}$ we have $r_n = \lambda(n+1)h_{n+1}$ for all $n \in \mathbb{N}$. We calculate

$$-\int_z^1 \sum_{n=0}^{\infty} r_n u^n \mathrm{d}u = \sum_{n=0}^{\infty} \frac{r_n}{n+1} \left(-1 + z^{n+1}\right) = \lambda(H(z) - 1).$$

The latter equality, together with elementary characteristics of power series concludes the proof. □

**Example 1** (Poisson Distribution of Order $\nu$, Negative Binomial Distribution)**.**
*If $\nu < \infty$ and if the compounding distribution is the uniform distribution on $\{1, \dots, \nu\}$, i.e., if $h_x = 1/\nu$ for all $x = 1, \dots, \nu$, then the resulting distribution is also known as the* Poisson distribution of order $\nu$*, see Sections 9.3 and 10.7.4 in Johnson et al. (2005). This distribution is abbreviated hereafter as* $\mathrm{Poi}_\nu(\lambda)$*, where* $\mathrm{Poi}_1 = \mathrm{Poi}$.

*Also the* negative binomial distribution NegBin$(n, \pi)$ *with* $n > 0$ *and* $\pi \in (0,1)$ *is Compound Poisson, with* $\lambda := -n \ln \pi$ *and* $h_k = (1-\pi)^k/(-k \ln \pi)$ *for all* $k \in \mathbb{N}$*. Here, the* $X_i$ *of Definition 2.3.1 follow the* logarithmic series distribution LSD$(\pi)$ *(Johnson et al., 2005, Chapter 5).*

The requirement in Definition 2.3.1 that the $X_1, X_2, \ldots$ have the range $\mathbb{N} = \{1, 2, \ldots\}$ guarantees a unique representation of the associated Compound Poisson distribution. This follows from (2.6) and

$$\begin{aligned}\exp\left(\lambda(H(z)-1)\right) &= \exp\left(\lambda(h_0 - 1 + h_1 z + h_2 z^2 + \ldots)\right)\\ &= \exp\left(\lambda(1-h_0)\left(-1 + \frac{h_1}{1-h_0} z + \frac{h_2}{1-h_0} z^2 + \ldots\right)\right),\end{aligned}$$

which implies that for each choice of $\lambda' > 0$ and $H'(z)$ with $h'_0 > 0$ there exists a $\lambda > 0$ and a $H(z)$ with $h_0 = 0$ such that $\text{ComPoi}(\lambda', H') \stackrel{\mathcal{D}}{=} \text{ComPoi}(\lambda, H)$, as equality of the pgfs implies equality in distribution. Thus, without loss of generality, we may assume $H(0) = 0$ throughout this thesis, cp. p. 389 in Johnson et al. (2005).

The following two assertions are concerned with the raw moments of Compound Poisson distributions, so we first introduce some notation. Throughout this thesis, the moments about the origin of a random variable $\epsilon$ are abbreviated as $\mu_{\epsilon,k} := \mathbb{E}[\epsilon^k]$ with $\mu_\epsilon := \mu_{\epsilon,1}$. The central moments are denoted as $\bar{\mu}_{\epsilon,k} := \mathbb{E}[(\epsilon - \mu_\epsilon)^k]$, with $\sigma_\epsilon^2 := \bar{\mu}_{\epsilon,2}$.

**Proposition 2.3.6** (Schweer and Weiß (2014), Proposition B.1)**.** *Let* $\epsilon$ *be* ComPoi$(\lambda, H)$*-distributed according to Definition 2.3.1.*

*(i) A recursive scheme for the computation of* $\mathbb{P}(\epsilon = s)$ *is given by*

$$\mathbb{P}(\epsilon = 0) = e^{-\lambda}, \quad s\mathbb{P}(\epsilon = s) = \lambda \sum_{j=0}^{s-1} (s-j) h_{s-j} \mathbb{P}(\epsilon = j) \quad \textit{for} \quad s \geq 1.$$

*(ii) If the moments* $\mu_X, r$ *of* $X_i$ *exist, then the cumulants of* $\epsilon$ *are given by*

$$\kappa_{\epsilon,r} = \lambda \mu_X, r.$$

Part (i) was derived by Kemp (1967); an explicit expression for $\mathbb{P}(\epsilon = s)$ is given on p. 288 in Feller (1968). Part (ii) is considered in Section 2 of Aki et al. (1984), it follows immediately from the cumulant generating function (cgf) $\text{cgf}_\epsilon(z) = \lambda\left(H(e^z) - 1\right)$, also see relation (2.6), where $H(e^z)$ is just the moment generating function (mgf) of $X_i$. Note that part (ii) implies that the index of dispersion of $\epsilon$ equals $\mathbb{E}[X_1^2]/\mathbb{E}[X_1]$, i.e., the ComPoi$_\nu$-distribution is equidispersed if and only if $\nu = 1$, and overdispersed otherwise.

**Example 2** (Poisson Distribution of Order $\nu$)**.** *Let* $\epsilon \sim$ Poi$_\nu$ *as given in Example 1. Since the* $X_i$ *are uniformly distributed, the raw moments* $\mu_{\epsilon,r} = \frac{1}{\nu}\sum_{x=1}^{\nu} x^r$ *are easily* *computed, and from Proposition 2.3.6 (ii), we immediately obtain the cumulants of* $\epsilon$*. In particular,*

$$\bar{\mu}_{\epsilon,1} = \frac{\lambda(\nu+1)}{2}, \quad \bar{\mu}_{\epsilon,2} = \frac{\lambda(\nu+1)(2\nu+1)}{6}, \quad \bar{\mu}_{\epsilon,3} = \frac{\lambda\nu(\nu+1)^2}{4},$$
$$\bar{\mu}_{\epsilon,4} = \frac{\lambda(\nu+1)(2\nu+1)(3\nu^2+3\nu-1)}{30} + 3\sigma_\epsilon^4. \tag{2.8}$$

*For further details on the relations between moments and cumulants, the reader is referred to Appendix 7 in Douglas (1980).*

**Example 3** (Negative Binomial Distribution)**.** *The negative binomial distribution with parameter* $(n, \pi)$ *of Example 1 satisfies*

$$\bar{\mu}_{\epsilon,1} = \frac{n(1-\pi)}{\pi}, \quad \bar{\mu}_{\epsilon,2} = \frac{n(1-\pi)}{\pi^2}, \quad \bar{\mu}_{\epsilon,3} = \frac{n(1-\pi)(2-\pi)}{\pi^3},$$
$$\bar{\mu}_{\epsilon,4} = \frac{3n^2(1-\pi)^2 + n(1-\pi)(\pi^2-6\pi+6)}{\pi^4}. \tag{2.9}$$

*For these and further properties of the* $\mathrm{NegBin}(n, \pi)$*-distribution, we refer to Chapter 5 in Johnson et al. (2005).*

Another popular member of the $\mathrm{ComPoi}_\infty$-family is Consul's *generalized Poisson distribution* (also Lagrangian Poisson distribution), $\mathrm{GP}(\theta, \eta)$, see Zhu and Joe (2003) and Section 7.2.6 in Johnson et al. (2005) for details. Here, the compounding probabilities are given by $h_x = \eta(\eta x)^{x-1}\, e^{-\eta x}/x!$, see Example 5.5. in Zhu and Joe (2003).

## 2.4 Time-Reversibility of Stochastic Processes

The concept of time-reversibility of stochastic processes will be of interest during various stages of this thesis. Formally, a process is time-reversible if it obeys the following definition.

**Definition 2.4.1** (Time-Reversibility)**.** *A stochastic process* $(Y_t)_{t\in\mathbb{Z}}$ *is* time-reversible *if the vector* $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n})$ *has the same distribution as* $(Y_{\tau-t_1}, Y_{\tau-t_2}, \ldots, Y_{\tau-t_n})$ *for all* $t_1, t_2, \ldots, t_n, \tau \in \mathbb{Z}$*.*

In order for a process to be time-reversible it should exhibit the same behavior whether time passes normally or "backwards", or, to put it more eloquently,

> Speaking intuitively, if we take a film of such a process and then run the film backwards the resulting process will be statistically indistinguishable from the original process.
>
> Kelly (1979), p. 5.

The criterion given in the definition is formulated for general stochastic processes. If the underlying structure of the process is simpler, e.g., Markovian, the necessary and sufficient conditions for time-reversibility can be stated much more succinctly, as shown by Kolmogorov's criterion.

**Theorem 2.4.2** (Kelly (1979), Theorem 1.7)**.** *A stationary Markov chain $(Y_t)_{t\in\mathbb{Z}}$ on a discrete state space $S$ is time-reversible if and only if the transition probabilities, given by $\mathbb{P}(Y_t = l \mid Y_{t-1} = k) := p_Y(l|k)$ satisfy the relation*

$$p_Y(j_1|j_n)p_Y(j_2|j_1)\cdots p_Y(j_n|j_{n-1}) = p_Y(j_1|j_2)\cdot p_Y(j_2|j_3)\cdots p_Y(j_{n-1}|j_n)\cdot p_Y(j_n|j_1),$$

*for any finite sequence of states $j_1, j_2, \ldots, j_n \in S$.*

It should be noted that time-reversibility of a process is a very rare characteristic among stochastic processes. On the other hand, the time-reversible models are often the most popular models. As an example, Weiss (1975) shows that the continuous AR(p) process is time-reversible if and only if it is Gaussian, a very widely used model. Since the question whether a given process is Gaussian is a crucial one in time series analysis, this observation has lead to the following specification test by Ramsey and Rothman (1996): Suppose that the process is time-reversible, then it necessarily holds that $\mathbb{E}[Y_t^i Y_{t-k}^j] = \mathbb{E}[Y_t^j Y_{t-k}^i]$ for any $i, j, k \in \mathbb{N}$. In particular, in later parts of this thesis, (see Section 5.2) we will be interested in *generalized autocovariance function* $\beta(\cdot)$ defined by

$$\beta(k) := \mathbb{E}\left[Y_t^2 Y_{t-k}\right] - \mathbb{E}\left[Y_t Y_{t-k}^2\right] \quad \text{for } k \in \mathbb{N}. \tag{2.10}$$

If $(Y_t)_{t\in\mathbb{Z}}$ is time-reversible, it follows that $\beta(k) = 0$ for all $k \in \mathbb{N}_0$. The empirical counterpart of $\beta(\cdot)$, computed from an observation $Y_1, \ldots, Y_T$ of a stationary process $(Y_t)_{t\in\mathbb{Z}}$, is defined as

$$\hat{\beta}_T(k) := \frac{1}{T-k}\sum_{t=k+1}^{T}\left(Y_t^2 Y_{t-k} - Y_{t-k}^2 Y_t\right) \quad \text{for } k \in \{1, \ldots, T_1\}. \tag{2.11}$$

The resultant test for time-reversibility of $(Y_t)_{t\in\mathbb{Z}}$ checks the deviation of $\hat{\beta}_T(k)$ from 0.

## 2.5 Central Limit Theorems for Dependent Data

At various instances throughout this thesis, we will be interested in deriving central limit theorems (CLTs) for diverse functionals of discrete time series. The usual approach for such theorems is the following: Supposing that the random variables $Y_1, \ldots, Y_T$ are identically distributed and independent, have mean zero and a finite variance $\sigma^2$, it is easily shown using characteristic functions (or similar techniques), that the convergence $\frac{1}{\sqrt{T}}\sum Y_i \to \mathcal{N}(0, \sigma^2)$ holds in distribution. For the data we study in this thesis, however, the assumption of independence is too strong as we want to allow the processes to depend on their past realizations, hence the need for central limit theorems for dependent random variables. Note that the processes studied in this thesis are stationary as a rule, hence the assumption of identically distributed random variables is not violated.

The general idea behind central limit theorems for dependent data usually boils down to the assumption that the data, while dependent over time, become "more and more" independent the further the data points are apart. In order to measure the level of dependency within the data for a given stationary stochastic process $(Y_t)_{t\in\mathbb{Z}}$, we define the notion of *strong mixing*.

**Definition 2.5.1** ($\alpha$-Mixing). *Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary stochastic process and define the $\sigma$-algebras $\mathcal{F}_k(Y) := \sigma(Y_i\ ; -\infty < i \le k)$ and $\mathcal{F}^l(Y) := \sigma(Y_i\ ; l \le i < \infty)$. Then $(Y_t)_{t\in\mathbb{Z}}$ is* $\alpha$-mixing, *if*

$$\alpha_Y(n) := \sup_{A\in\mathcal{F}_k(Y);B\in\mathcal{F}^{k+n}(Y)} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \to 0 \quad \textit{as } n\to\infty. \tag{2.12}$$

For each $n \in \mathbb{N}$, the coefficient appearing in (2.12) are called *$\alpha$-mixing* or *strong mixing coefficients* (or weights). The notion of strong mixing is usually credited to Rosenblatt (1956), and it should be pointed out that the number of different types of mixing formulations has increased since then, necessitating specializations such as $\alpha$-mixing or $\beta$-mixing. Indeed, the excellent survey paper on mixing properties Bradley (2005) lists no less than eight different types of strong mixing conditions. In this thesis, we will make frequent use of the following central limit theorem for $\alpha$-mixing stationary processes.

**Theorem 2.5.2** (Ibragimov (1962), Theorem 1.7). *Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary and $\alpha$-mixing process with $\mathbb{E}[|Y_0|^{2+\delta}] < \infty$ for some $\delta > 0$ and where*

$$\sum_{j=1}^{\infty} (\alpha_Y(j))^{\frac{\delta}{2+\delta}} < \infty. \tag{2.13}$$

*Then the series $\sigma^2 = \sum_{j\in\mathbb{Z}} \operatorname{Cov}(Y_0, Y_j)$ converges absolutely, and*

$$\sqrt{T}\left(\frac{1}{T}\sum_{i=1}^{T} Y_i - \mathbb{E}[Y_0]\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2).$$

One of the reasons why $\alpha$-mixing has gained such an importance is that it is invariant under measurable maps, i.e., if $(Y_t)_{t\in\mathbb{Z}}$ is an $\alpha$-mixing process and $f$ is a measurable map, then the process $(f(Y_t))_{t\in\mathbb{Z}}$ is $\alpha$-mixing again. On the other hand, mixing conditions are often difficult to verify directly. An alternative criterion is used in the following result. For the first assertion we provide a sketch of the proof given in the cited reference, as some of the arguments will be used again at later stages of this thesis.

**Theorem 2.5.3** (Billingsley (1999), Theorem 19.1)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary and ergodic process with $\mathbb{E}[Y_0^2] < \infty$. Let*

$$\sum_{j=1}^{\infty} \|\mathbb{E}\left[Y_j - \mathbb{E}[Y_0] \mid \mathcal{F}_0(Y)\right]\|_{L^2} < \infty, \tag{2.14}$$

*where $\|X\|_{L^2} = (\mathbb{E}[|Y|^2])^{\frac{1}{2}}$ denotes the $L^2$ norm. Then the series $\sigma^2 = \sum_{j\in\mathbb{Z}} \operatorname{Cov}(Y_0, Y_j)$ converges absolutely, and*

$$\sqrt{T}\left(\frac{1}{T}\sum_{i=1}^{T} Y_i - \mathbb{E}[Y_0]\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2).$$

*Proof.* without loss of generality, $\mathbb{E}[Y_0] = 0$. By the Cauchy-Schwarz inequality,

$$|\mathbb{E}\left[Y_0 Y_i\right]| \leq \mathbb{E}\left[|Y_0| \cdot |\mathbb{E}\left[Y_i|\mathcal{F}_0(Y)\right]|\right] \leq \|Y_0\|_{L^2} \cdot \|\mathbb{E}\left[Y_i|\mathcal{F}_0(Y)\right]\|_{L^2}, \tag{2.15}$$

thus (2.14) shows that the series $\sigma^2$ converges absolutely. The stationarity of $(Y_t)_{t\in\mathbb{Z}}$ implies that $\mathbb{E}[(\sum_{i=1}^{n} Y_i)^2] = n\mathbb{E}[Y_0^2] + 2\sum_{i=1}^{n-1}(n-i)\mathbb{E}[Y_0 Y_i]$. Direct calculations yield

$$\left|\sum_{i=-\infty}^{\infty} \mathbb{E}[Y_0 Y_i] - \frac{1}{n}\mathbb{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2\right]\right| \leq 2\sum_{i=n}^{\infty} |\mathbb{E}\left[Y_0 Y_i\right]| + \frac{2}{n}\sum_{i=1}^{n-1}\sum_{l=i}^{\infty} |\mathbb{E}\left[Y_0 Y_l\right]|, \tag{2.16}$$

which implies $\frac{1}{n}\mathbb{E}[(\sum_{i=1}^{n} Y_i)^2] \to \sigma^2$. For the former series in (2.16) the convergence to zero is obvious, for the latter notice that we may rearrange the series due to absolute convergence as follows,

$$\frac{1}{n}\sum_{i=1}^{n-1}\sum_{l=i}^{\infty} |\mathbb{E}\left[Y_0 Y_l\right]| = \sum_{l=n-1}^{\infty} |\mathbb{E}\left[Y_0 Y_l\right]| + \sum_{j=1}^{n-1}\left(1 - \frac{n-j}{n}\right) |\mathbb{E}\left[Y_0 Y_j\right]|,$$

and notice that $\lim_{n\to\infty} \frac{n-j}{n} = 1$ for each $j \in \mathbb{N}$ which, due to the dominated convergence theorem, yields the result.

□

The condition (2.14) can be seen as an alternative form of the mixing condition (2.12), in the sense that, under this condition, the dependence of the process on a given reference state $Y_0$ decays fast enough to 0 to be summable in the $L^2$ norm. The advantage of this condition in comparison to mixing conditions is that the conditional expectation is more easily accessible if the structure of the process is simple. A useful inequality in the context of (2.14) is given in the following Lemma, a consequence of Jensen's inequality.

**Lemma 2.5.4** (Billingsley (1999), eq. (19.25))**.** *Let $X$ be a random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ with a finite second moment and let $\mathcal{M}$ and $\mathcal{N}$ be $\sigma$-algebras satisfying $\mathcal{M} \subset \mathcal{N} \subset \mathcal{A}$. Then*

$$\|\mathbb{E}\left[X|\mathcal{M}\right]\|_{L^2} \leq \|\mathbb{E}\left[X|\mathcal{N}\right]\|_{L^2}, \quad \|\mathbb{E}\left[X|\mathcal{M}\right]\|_{L^2} \leq \|X\|_{L^2}. \tag{2.17}$$

## 2.6 Autocovariance Function and Related Concepts

In order to assess the dependence of a given stochastic process on the past, there is a multitude of functions available. In this section, we consider the three most popular functions, the autocovariance function $\gamma(\cdot)$, the autocorrelation function $\rho(\cdot)$ and the partial autocorrelation function $\rho_{part}(\cdot)$. For a given stationary process $(Y_t)_{t\in\mathbb{Z}}$, the first two functions are easily defined:

$$\gamma(k) := \mathrm{Cov}\left(Y_t, Y_{t+k}\right) \quad \text{and} \quad \rho(k) := \frac{\gamma(k)}{\gamma(0)} \quad \text{for } k \in \mathbb{Z}.$$

Denoting $\overline{Y} := \frac{1}{T}\sum_{i=1}^{T} Y_i$, the empirical counterparts are also easily established:

$$\widehat{\gamma}(k) := \frac{1}{T-k}\sum_{i=1}^{T-k}\left(Y_i - \overline{Y}\right)\left(Y_{i+k} - \overline{Y}\right) \quad \text{and} \quad \widehat{\rho}(k) := \frac{\widehat{\gamma}(k)}{\widehat{\gamma}(0)} \quad \text{for } k \in \{0,\ldots,T-1\}.$$

Let us give an example. Let $p \in \mathbb{N}$ and let $\alpha_1,\ldots,\alpha_p \in (-1,1)$ denote appropriately chosen parameters, then a stationary stochastic process is called a continuous *autoregressive process of order* $p$ if it satisfies the recursion

$$Y_t = \sum_{i=1}^{p} \alpha_i Y_{t-i} + \epsilon_t, \tag{2.18}$$

where $(\epsilon_t)_{t\in\mathbb{Z}}$ is a sequence of white noise random variables (note that in this case, $\epsilon_t$ can take on values in $\mathbb{R}$). Since $\mathbb{E}[\epsilon_0] = 0$ it follows that $\mathbb{E}[Y_0] = 0$, and multiplication of (2.18) with $Y_{t-k}$ on both sides and taking expectations $\gamma(k) = \sum_{i=1}^{p} \alpha_i\gamma(k-i)$ for any $k \in \mathbb{N}$, the Yule-Walker equations. For instance, if $(Y_t)_{t\in\mathbb{Z}}$ is an AR(1) process with parameter $\alpha \in (-1,1)$, these immediately imply $\gamma(k) = \alpha^k\gamma(0)$ for all $k \in \mathbb{N}_0$ and thus $\rho(k) = \alpha^k$. The partial autocorrelation function is a little harder to define. Essentially, $\rho_{part}(k)$ evaluates the correlation of the random variables $Y_t$ and $Y_{t+k}$ adjusted for the intermediate values $Y_{t+1},\ldots,Y_{t+k-1}$.

**Definition 2.6.1** (Partial Autocorrelation Function)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary process with $\mathbb{E}[Y_0^2] < \infty$ and autocorrelation function $\rho(\cdot)$. Then $\rho_{part}(\cdot)$ is called the* partial autocorrelation function *if it satisfies $\rho_{part}(1) := \rho(1)$ and $\rho_{part}(k) = \det(\mathbf{U}_k)/\det(\mathbf{L}_k)$, where the matrices (for $k \geq 2$) are given by*

$$\mathbf{U}_k := \begin{pmatrix} 1 & \rho(1) & \ldots & \rho(k-2) & \rho(1) \\ \rho(1) & 1 & \ldots & \rho(k-3) & \rho(2) \\ \vdots & & \ddots & \vdots & \vdots \\ \rho(k-1) & & \ldots & \rho(1) & \rho(k) \end{pmatrix}, \mathbf{L}_k := \begin{pmatrix} 1 & \ldots & \rho(k-1) \\ \rho(1) & \ddots & \rho(k-2) \\ \vdots & & \vdots \\ \rho(k-1) & \ldots & 1 \end{pmatrix}.$$

It is quite obvious that ergodicity of the process together with appropriate moment assumptions already yields consistency of the estimators $\widehat{\gamma}(k)$ for any $k \in \mathbb{N}_0$, and the continuous mapping theorem immediately implies the consistency of $\widehat{\rho}(k)$ and $\widehat{\rho}_{part}(k)$. for any $k \in \mathbb{N}$.

The asymptotic normality of the estimators can be treated similarly. If suffices to show that the estimators for the autocovariance function $\gamma(\cdot)$ are jointly asymptotically normal to provide inference for the asymptotic behavior of the ACF and PACF estimators. We record this result together with two important consequences.

**Theorem 2.6.2** (Romano and Thombs (1996), Theorem 3.2). *Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary process with*

$$\sqrt{T}\left(\widehat{\gamma}(0)-\gamma(0),\dots,\widehat{\gamma}(K)-\gamma(K)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0},\mathbf{T}+\mathbf{U}\right),$$

*where the entries $\tau_{i,j}$ of the matrix $\mathbf{T}$ are given by*

$$\tau_{i+1,j+1} = \sum_{d=-\infty}^{\infty} \left[\gamma(d)\gamma(d+j-i)+\gamma(d+j)\gamma(d-i)\right],$$

*and where the entries $u_{i,j}$ of the matrix $\mathbf{U}$ are given by*

$$u_{i+1,j+1} = \sum_{d=-\infty}^{\infty} \operatorname{cum}(Y_0,Y_i,Y_d,Y_{d+j}).$$

*Let $q \geq 1$ and denote $K := p+q$. Then*

$$\sqrt{T}\left(\widehat{\rho}(1)-\rho(1),\dots,\widehat{\rho}(K)-\rho(K)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0},\mathbf{T}'+\mathbf{U}'\right),$$

*where the entries of the matrix $\mathbf{T}'$ are calculated from $\mathbf{T}$ via*

$$\tau'_{i,j} = \frac{1}{\gamma(0)^2}\left(\tau_{i+1,j+1}-\rho(i)\tau_{1,j+1}-\rho(j)\tau_{i+1,1}+\rho(i)\rho(j)\tau_{1,1}\right),$$

*analogously for the matrix $\mathbf{U}'$.*

In the special case of autoregressive processes, we are able to explicitly calculate the resulting covariance matrix of the estimator of the partial autocorrelation function.

**Theorem 2.6.3** (Ku and Seneta (1996), Theorem 1). *Let $(Y_t)_{t\in\mathbb{Z}}$ satisfy the conditions of Theorem 2.6.2 and let the ACF satisfy the Yule-Walker equations $\sum_{j=0}^{p}\alpha_j\rho(i-j)=0$, where $\alpha_0 := -1$ and where $\sum_{i=1}^{p}|\alpha_i|<1$. Then*

$$\sqrt{T}\left(\widehat{\rho}_{part}(p+1),\dots,\widehat{\rho}_{part}(K)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0},\mathbf{1}_{q\times q}+\mathbf{\Lambda}\right),$$

*where $\mathbf{1}_{q\times q}$ is the unity matrix and the entries of $\mathbf{\Lambda}$ are given by*

$$d_{i,j} = \frac{1}{\gamma(0)^2\left(\prod_{i=1}^{p}1-\rho_{part}(i)^2\right)^2}\sum_{u=0}^{2p}\sum_{m+n=u}\alpha_m\alpha_n\sum_{r=0}^{2p}\sum_{s+t=r}\alpha_s\alpha_t u_{p+1+i-r,p+1+j-u}.$$

*Proof.* In (Ku and Seneta, 1996, Theorem 1), it is shown that if the random vector $\sqrt{T}(\widehat{\rho}(1) - \rho(1), \ldots, \widehat{\rho}(K) - \rho(K))$ converges in distribution to some random vector $(V(1), \ldots, V(K))$, then the random vector $\sqrt{T}(\widehat{\rho}_{\text{part}}(p+1), \ldots, \widehat{\rho}_{\text{part}}(K))$ converges in distribution to the vector $(W(p+1), \ldots, W(K))$, where

$$W(k) = \frac{\sum_{m=0}^{p} \sum_{n=0}^{p} \alpha_i \alpha_j V(k-m-n)}{\prod_{i=1}^{p}(1 - \rho_{part}(i)^2)}. \tag{2.19}$$

Here, $V(0) := 0$ and $V(-k) := V(k)$ for $k \geq 1$. The conditions are obviously fulfilled in our case, and we further have $\mathbb{E}[V(k)] = 0$ for each $k \in \mathbb{Z}$. Using the approach of Ku and Seneta (1996), we first write

$$\mathbf{V} := (V(p+q), \ldots, V(-p+1))^\top, \mathbf{W} := (W(p+1), \ldots, W(p+q))^\top$$

and the $q \times (2p+q)$ matrix $\mathbf{A} := \mathbf{A}'/(\prod_{i=1}^{p}(1 - \rho_{part}(i)^2))$, where

$$\mathbf{A}' := \begin{pmatrix} 0 & \ldots & 0 & 1 & \sum_1 \alpha_m\alpha_n & \ldots & \ldots & \sum_{2p} \alpha_m\alpha_n \\ 0 & \ldots & 1 & \sum_1 \alpha_m\alpha_n & \ldots & \ldots & \sum_{2p} \alpha_m\alpha_n & 0 \\ \vdots & & & & & & & \vdots \\ 1 & \sum_1 \alpha_m\alpha_n & \ldots & \ldots & \sum_{2p} \alpha_m\alpha_n & 0 & \ldots & 0 \end{pmatrix},$$

where summing from 1, i.e., means for $m + n = 1$. With these notations, we may write (2.19) as $\mathbf{W} = \mathbf{AV}$, noticing that

$$\sum_{m=0}^{p} \sum_{n=0}^{p} \alpha_i \alpha_j V(k-m-n) = \sum_{u=0}^{2p} \sum_{m+n=u} \alpha_i \alpha_j V(k-u).$$

As $\mathbf{V}$ has a multivariate normal distribution with mean 0 so does $\mathbf{W}$. Let the covariance matrix of $\mathbf{W}$ be denoted by $\Sigma_{\mathbf{W}}$, then Theorem 2.6.2 implies $\Sigma_{\mathbf{W}} = \mathbf{AT}'\mathbf{A}^\top + \mathbf{AU}'\mathbf{A}^\top$. In Ku and Seneta (1996) it is shown that $\mathbf{AT}'\mathbf{A}^\top = \mathbf{1}_{q \times q}$, the entries of $\mathbf{\Lambda}$ remain to be calculated. For this, we introduce the convenient functions (cp. eq. (3.4) in Ku and Seneta (1996) and Property 3 in Choi (1990))

$$h(k) = \sum_{u=0}^{2p} \sum_{m+n=u} \alpha_m \alpha_n \rho(k-u), \tag{2.20}$$

with $h(k) = 0$ for $k = p, p+1, \ldots$. Theorem 2.6.2 implies that $d_{i,j}$ equals, omitting the constant factors,

$$\begin{aligned} &= \sum_{u=0}^{2p} \sum_{m+n=u} \alpha_m \alpha_n \sum_{r=0}^{2p} \sum_{s+t=r} \alpha_s \alpha_t u_{p+1+i-r, p+1+j-u} + h(p+i)h(p+j) \\ &- h(p+i) \sum_{u=0}^{2p} \sum_{m+n=u} \alpha_m \alpha_n u_{1,p+1+j-u} - h(p+b) \sum_{r=0}^{2p} \sum_{s+t=r} \alpha_s \alpha_t u_{p+1+i-r,1}, \end{aligned}$$

cp. p. 627 in Ku and Seneta (1996), concluding the proof. □

# 3 Nonparametric Estimation in Discrete-Time Queueing Processes

The problem under consideration in this chapter is the estimation of the cumulative distribution function (cdf) of the service time distribution $G$ in a discrete-time $GI/G/\infty$-queue, i.e., a queueing model with an infinite number of servers, a general service time distribution and a general i.i.d. batch arrival process $(A(t))_{t\in\mathbb{Z}}$. We assume that the available information about the behavior of this queue consists only of the counts of arrivals $(A(t))_{t\in\mathbb{Z}}$ and departures $(D(t))_{t\in\mathbb{Z}}$ from the queue in each time slot. More precisely, we consider a queue in which there is no possibility for the observer to distinguish between any of the customers, so that the matching of any departure to its respective arrival is impossible. Additionally, the number of customers present at the beginning of the observation is also unavailable. Our goal is the estimation of the entire service time distribution for which we assume no parametric form of any kind. Thus we are faced with a nonparametric estimation problem for $G$, for which we merely assume a finite mean and that its range is contained in $\mathbb{N}$.

The solution we present to this problem in this chapter starts out with the nonparametric estimation of a different distribution $H$ and uses the surprisingly simple relation between $G$ and $H$. The nonparametric nature of the estimates necessitates the application of functional approaches to the problem, which will be assumed to take place in the sequence space $c_0$. Since the resultant asymptotic expressions are rather involved and depend on unknown parameters, a bootstrapping procedure is suggested and is shown to be a viable option in our context under mild additional conditions. This chapter is an extended version of the article Schweer and Wichelhaus (2015a).

## 3.1 Introduction and Statement of the Problem

We begin by precisely defining the queueing model and stating the first results. The behavior of the queue is modeled as follows: denote the number of arrivals in the $t$-th time slot, the time slot between time $t$ and $t+1$, by $A(t)$ and the number of departures in this slot by $D(t)$. In each time slot $t \in \mathbb{Z}$, indistinguishable customers labeled $K_{t,1}, \ldots, K_{t,A(t)}$ arrive, where $A(t) = 0$ is interpreted as no customers arriving in the $t$-th time slot. We assume that the sequence $(A(t))_{t\in\mathbb{Z}}$ is i.i.d., has range $\mathbb{N}_0$ and that $\mathbb{E}[A(0)] < \infty$. Each customer $K_{k,j}$ receives upon arrival a sojourn time $S_{k,j}$ independently of all other customers arriving or present at the queue, where $S_{k,j}$ is distributed with cdf $G(\cdot)$, which has range $\mathbb{N}$ and a finite mean, i.e. $\sum_{i=1}^{\infty}(1 - G(i)) < \infty$. Denoting the probability masses of the distribution $G$ by $g_j$ for $j \in \mathbb{N}$, we thus have $\mathbb{P}(S_{k,j} = l) = g_l$ for any $k \in \mathbb{Z}$, $j, l \in \mathbb{N}$. Each customer $K_{k,j}$ then remains in service exactly the number of time steps that his service time $S_{k,j}$ demands and then leaves the queue. We point out that we make the assumption $G(0) = 0$ in order to ensure that each customer remains in the queue for at least one time step. We limit our knowledge about the considered system to the sequences $(A(t))_{t\in\mathbb{Z}}$ and $(D(t))_{t\in\mathbb{Z}}$ and we base our analysis of the behavior of this system solely on this information, i.e., we do not assume to have any possibility of matching the arrival of certain customers to their respective departures.

We define the "enlarged" process $(\xi(t))_{t\in\mathbb{Z}}$ by $\xi(t) := \{S_{t,1}\} \times \{S_{t,2}\} \times \cdots \times \{S_{t,A(t)}\}$, the collection of all information given for the process in the $t$-th time slot, i.e., $\xi(t)$

carries information about both the number of arrivals in the $t$-th time slot as well as the service time distribution for these arrivals. Notice that $\xi(t) \in \mathbb{N}^{\mathbb{N}} \cup \{0\}$ for each $t \in \mathbb{Z}$, where the state 0 represents the case of no arrivals in the $t$-th time slot. Since the sequences of random variables $(S_{k,\cdot})_{k\in\mathbb{Z}}$ and $(A(t))_{t\in\mathbb{Z}}$ are i.i.d., it follows that the process $(\xi(t))_{t\in\mathbb{Z}}$ is stationary and ergodic. Furthermore, recall the definition of the $\sigma$-algebras $\mathcal{F}_k(\xi) := \sigma\left(\xi(i)\ ; -\infty < i \leq k\right)$ for $k \in \mathbb{Z}$ as in Section 2.5. With this construction, the process $(\xi(t))_{t\in\mathbb{Z}}$ is an element of the space $\{\mathbb{N}^{\mathbb{N}} \cup \{0\}\}^{\mathbb{Z}}$. We consider this Baire space to be endowed with its product topology, and we denote the Borel-$\sigma$-algebra based on the open sets of this topology by $\mathcal{F}_\infty(\xi)$.

Under the assumption that the system has started in the infinite past, it follows from the construction of the process that we may express the queue length $Y(t)$, i.e., the number of customers in service during the $t$-th time slot by

$$Y(t) = \sum_{j=0}^{\infty} \sum_{l=1}^{A(t-j)} \mathbf{1}_{\{S_{t-j,l} > j\}}, \tag{3.1}$$

where customers who leave during the $t$-th time slot are not considered to be in service. From this representation and the assumption that the sequences $(S_{k,\cdot})_{k\in\mathbb{Z}}$ and $(A(t))_{t\in\mathbb{Z}}$ are i.i.d., it follows that the queue length process is stationary. Furthermore, the application of Wald's equation (see (3.2) below) implies that under the assumption of a finite mean of both service time distribution and arrival distribution, the stationary distribution of the queue length process has a finite mean. Finally, we remark that we may express the departure process as follows

$$D(t) = \sum_{j=1}^{\infty} \sum_{l=1}^{A(t-j)} \mathbf{1}_{\{S_{t-j,l} = j\}}.$$

Let us first record some important relations. First, a well-known result due to Wald as well as Blackwell and Girshick states that if $T, X_1, X_2, \dots$ are independent random variables with finite variance, and if $T$ has range $\mathbb{N}_0$ and the $X_1, X_2, \dots$ are identically distributed, then, with $S_T := \sum_{i=1}^{T} X_i$,

$$\mathbb{E}[S_T] = \mathbb{E}[T]\mathbb{E}[X_1] \quad \text{and} \quad \operatorname{Var}(S_T) = \mathbb{E}[X_1]^2 \operatorname{Var}(T) + \mathbb{E}[T] \operatorname{Var}(X_1). \tag{3.2}$$

The former relation is called Wald's equation. As immediate consequences of these relations, we find that, for all $t \in \mathbb{Z}$,

$$\mathbb{E}[D(t)] = \sum_{j=1}^{\infty} \mathbb{E}\left[\sum_{l=1}^{A(t-j)} \mathbf{1}_{\{S_{t-j,l} = j\}}\right] = \mathbb{E}[A(0)] \sum_{j=1}^{\infty} g_j = \mathbb{E}[A(0)].$$

We used that the sequence $(A(t))_{t\in\mathbb{Z}}$ is i.i.d. and the monotone convergence theorem. Furthermore, if $\mathbb{E}[A(0)^2] < \infty$,

$$\operatorname{Var}(D(0)) = \sum_{j=1}^{\infty} \operatorname{Var}\left(\sum_{l=1}^{A(-j)} \mathbf{1}_{\{S_{-j,l} = j\}}\right) = \sum_{j=1}^{\infty} \left[\mathbb{E}[A(0)] g_j (1 - g_j) + g_j^2 \operatorname{Var}(A(0))\right].$$

Notice that $\max\{\mathbb{E}[A(0)], \mathrm{Var}(A(0))\} \leq \max\left\{\mathbb{E}[A(0)], \mathbb{E}[A(0)^2]\right\}$, and since the random variable $A(0)$ is discrete-valued, $\max\left\{\mathbb{E}[A(0)], \mathbb{E}[A(0)^2]\right\} = \mathbb{E}[A(0)^2]$. We may thus conclude that $\mathrm{Var}(D(0)) \leq 2\mathbb{E}[A(0)^2]$ and $\mathrm{Var}\left(D(0)\mathbf{1}_{\{Z(0)\leq x\}}\right) \leq \mathbb{E}[D(0)^2] < \infty$.

Additionally, we note that in the popular case of Poisson distributed arrivals it holds that $\mathrm{Var}(A(0)) = \mathbb{E}[A(0)]$ and hence $\mathrm{Var}(D(0)) = \mathrm{Var}(A(0))$. Furthermore, in this case we easily calculate, for $s \neq 0$,

$$\begin{aligned}\mathrm{Cov}\left(D(0), D(s)\right) &= \sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty} \mathrm{Cov}\left(\sum_{l=1}^{A(-j_1)} \mathbf{1}_{\{S_{-j_1,l}=j_1\}}, \sum_{l=1}^{A(s-j_2)} \mathbf{1}_{\{S_{s-j_2,l}=j_2\}}\right) \\ &= \sum_{j_1=1}^{\infty} g_{j_1} g_{j_1+s}\left(\mathbb{E}[A(0)^2] - \mathbb{E}[A(0)] - \mathbb{E}[A(0)]^2\right) = 0, \qquad (3.3)\end{aligned}$$

a quite surprising result. In order to give an explanation for this curious result, we need to consider the time-reversibility of the process.

### 3.1.1 Time-Reversibility of the Queue Length Process

We now show that the special case of Poisson distributed arrivals $A(t)$ leads to a time-reversible queue length process $(Y(t))_{t\in\mathbb{Z}}$. Let us remark that this result was already mentioned in passing in Pickands and Stine (1997) but no formal proof was given.

**Lemma 3.1.1** (Cp. Pickands and Stine (1997), Sect. 2)**.** *Let* $A(t) \sim \mathrm{Poi}(\lambda)$ *for all* $t \in \mathbb{Z}$ *and some* $\lambda > 0$ *and let* $\sum_{i=1}^{\infty}(1 - G(i)) < \infty$*. Then* $(Y(t))_{t\in\mathbb{Z}}$ *is time-reversible.*

*Proof.* The finiteness of the expected service time immediately implies finiteness a.s. of $(Y(t))_{t\in\mathbb{Z}}$. Let $b(t_a; t_{b_1}, t_{b_2})$ denote the number of customers arriving at the system in time slot $t_a$ and departing in between the time slots $t_{b_1}$ and $t_{b_2}$ with $t_a \leq t_{b_1} \leq t_{b_2}$. We first notice that (setting $t_a = 0$ without loss of generality)

$$\mathbb{P}\left(\sum_{l=1}^{r} \mathbf{1}_{\{S_{0,l}\in\{t_{b_1},\ldots,t_{b_2}\}\}} = s\right) = \binom{r}{s}\left(\sum_{l=0}^{t_{b_2}-t_{b_1}} g_{t_{b_1}+l}\right)^{s}\left(1 - \sum_{l=0}^{t_{b_2}-t_{b_1}} g_{t_{b_1}+l}\right)^{r-s}.$$

Under the assumption of Poisson arrivals $A(t)$, it follows that

$$\begin{aligned}\mathbb{P}\left(b(0; t_{b_1}, t_{b_2}) = s\right) &= \sum_{r=s}^{\infty} \mathbb{P}\left(A(0) = r\right)\binom{r}{s}\left(\sum_{l=0}^{t_{b_2}-t_{b_1}} g_{t_{b_1}+l}\right)^{s}\left(1 - \sum_{l=0}^{t_{b_2}-t_{b_1}} g_{t_{b_1}+l}\right)^{r-s} \\ &= \exp\left(-\lambda \sum_{l=0}^{t_{b_2}-t_{b_1}} g_{t_{b_1}+l}\right)\frac{\left(\lambda\sum_{l=0}^{t_{b_2}-t_{b_1}} g_{t_{b_1}+l}\right)^{s}}{s!},\end{aligned}$$

hence $b(0; t_{b_1}, t_{b_2})$ is Poisson distributed. Let $B(t_{a_1}, t_{a_2}; t_{b_1}, t_{b_2}) = \sum_{i=t_{a_1}}^{t_{a_2}} b(i; t_{b_1}, t_{b_2})$, then this number is Poisson distributed again as a sum of independent Poisson random variables, this statement extends to cases where either $t_{a_1} = -\infty$ or $t_{a_2} = \infty$.

Now, let $n \in \mathbb{N}$, $t_1, \dots, t_n$ and $\tau \in \mathbb{Z}$ be arbitrary, without loss of generality we assume $t_1 < t_2 < \dots < t_n$. Additionally, we set $t_0 := -\infty$ and $t_{n+1} := \infty$. Then, it is easily seen that

$$(Y(t_i))_{i\in\{1,\dots,n\}} \overset{\mathcal{D}}{=} \left( \sum_{l=1}^{i} \sum_{m=i}^{n} B(t_{l-1}, t_l; t_m, t_{m+1}) \right)_{i\in\{1,\dots,n\}}.$$

Similarly, we find that

$$(Y(\tau - t_i))_{i\in\{1,\dots,n\}} \overset{\mathcal{D}}{=} \left( \sum_{l=1}^{i} \sum_{m=i}^{n} B(\tau - t_{m+1}, \tau - t_m; \tau - t_l, \tau - t_{l-1}) \right)_{i\in\{1,\dots,n\}}.$$

For any $l \le m$, $B(t_{l-1}, t_l; t_m, t_{m+1})$ is Poisson distributed with the parameter given by $\lambda \sum_{r=0}^{t_l - t_{l-1}} \sum_{s=0}^{t_{m+1}-t_m} g_{t_l - t_m + r + s}$, similarly for $B(\tau - t_{m+1}, \tau - t_m; \tau - t_l, \tau - t_{l-1})$. Since the summands appearing in the expressions of the distribution of $(Y(t_1), Y(t_2), \dots, Y(t_n))$ and $(Y(\tau - t_1), Y(\tau - t_2), \dots, Y(\tau - t_n))$ are mutually independent, this shows equality in distribution of these vectors and thus concludes the proof.

□

The result of Lemma 3.1.1 offers an explanation for the relation (3.3): the reversal of time for a queue length process implies the switching of the roles of arrival and departure process. Hence, if the queue length process is time-reversible, it necessarily follows that both departure process $(D(t))_{t\in\mathbb{Z}}$ and arrival process $(A(t))_{t\in\mathbb{Z}}$ share the same features. In particular, it follows that the departure process $(D(t))_{t\in\mathbb{Z}}$ is not only uncorrelated but independent.

From a probabilistic point of view, this is a rather nice result. Yet it complicates matters for possible statistical inferences, as it implies that observations of the departure process alone do not contain any information about the service time distribution. Estimators for the service time thus necessarily need to encompass both the arrival and the departure process. One particular instance of such an estimator is studied in the next section.

### 3.1.2 The Sequence of Differences

Let us first define the discrete-time *sequence of differences* $(Z(t))_{t\in\mathbb{Z}}$ as

$$Z(t) := t - \max\{n < t \mid A(n) > 0\} \quad \text{for } t \in \mathbb{Z},$$

which corresponds to the time elapsed since the most recent arrival for each time instant. As the next step, the following cdf $H(\cdot)$ and its estimator $\widehat{H}_n(\cdot)$ is defined for every $x \in \mathbb{N}$,

$$H(x) := \frac{\mathbb{E}\left[D(0)\mathbf{1}_{\{Z(0)\le x\}}\right]}{\mathbb{E}[D(0)]} \quad \text{and} \quad \widehat{H}_n(x) := \frac{\sum_{i=1}^{n} D(i)\mathbf{1}_{\{Z(i)\le x\}}}{\sum_{i=1}^{n} D(i)}. \tag{3.4}$$

Let us explain the rationale behind this estimated cdf by first considering the numerator: In each time slot, it estimates the cdf of the time elapsed since the last arrival, but only if there was at least one departure during in this slot. Put differently, it estimates the cdf of the service time distribution under the (obviously false) assumption that the nearest possible arrival is accountable for the departure of each customer. In particular, this distribution circumvents the underlying matchmaking problem by assuming the most simple, albeit false model, note that if we could match each departure to its arrival, the estimation of the service time would be trivial.

Given a realization of the arrival process $(A(t))_{t\in\{1,\ldots,n\}}$ and the departure process $(D(t))_{t\in\{1,\ldots,n\}}$, this distribution function can easily be established. Analogous to the continuous time model (cf. Lemma 2 in Brown (1970)), it can now be shown that there is a very simple relation linking $H$ to the sought after service time distribution $G$. For the discrete-time case, Edelmann and Wichelhaus (2014) showed that for every $x \in \mathbb{N}$,

$$H(x) = 1 - c^x(1 - G(x)), \tag{3.5}$$

where $c := \mathbb{P}(A(0) = 0)$. Defining the estimator $\hat{c}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{A(i)=0\}}$, we obtain the following estimator for the service time distribution,

$$\widehat{G}_n(x) := 1 - \hat{c}_n^{-x}\left(1 - \widehat{H}_n(x)\right).$$

### 3.1.3 Ergodicity of the Sequences

The following result deals with the measurability of the random variables $D(t)$ for $t \in \mathbb{Z}$, which is a crucial step towards the proof of ergodicity of this sequence.

**Lemma 3.1.2.** *Let $t \in \mathbb{Z}$ and let $\overline{\mathbb{N}}_0 = \mathbb{N}_0 \cup \{\infty\}$ and let $\sum_{i=1}^\infty (1 - G(i)) < \infty$. Then the functions*

$$\begin{aligned} D(t) \ : \ & \{\mathbb{N}^{\mathbb{N}} \cup \{0\}\}^{\mathbb{Z}} \to \overline{\mathbb{N}}_0, \\ & (\xi(i))_{i\in\mathbb{Z}} \mapsto \sum_{j=1}^{\infty} \sum_{k=1}^{A(t-j)} \mathbf{1}_{\{S_{t-j,l}=j\}} \end{aligned}$$

*and*

$$\begin{aligned} D(t)\cdot \mathbf{1}_{\{Z(t)\le x\}} \ : \ & \{\mathbb{N}^{\mathbb{N}} \cup \{0\}\}^{\mathbb{Z}} \to \overline{\mathbb{N}}_0, \\ & (\xi(i))_{i\in\mathbb{Z}} \mapsto \mathbf{1}_{\{Z(t)\le x\}} \sum_{j=1}^{\infty} \sum_{k=1}^{A(t-j)} \mathbf{1}_{\{S_{t-j,l}=j\}} \end{aligned}$$

*are $\mathcal{F}_\infty(\xi)$-$\sigma(\overline{\mathbb{N}}_0)$-measurable.*

*Proof.* Recall that we equipped the space $\{\mathbb{N}^{\mathbb{N}} \cup \{0\}\}^{\mathbb{Z}}$ with the product topology and that we defined the Borel-$\sigma$-algebra based on the open sets of this topology as $\mathcal{F}_\infty(\xi)$. Applying the usual construction, see, e.g., Ch. 3.14 in Aliprantis and Border (2006), we find a base of the topology in the collection of sets of the form (with a slight abuse of notation)

$$U_{N_1,N_2,t} = \cdots \times \{\mathbb{N}^{\mathbb{N}} \cup \{0\}\} \times V_{t-N_2} \times V_{t-N_2+1} \times \cdots \times V_{t+N_2} \times \{\mathbb{N}^{\mathbb{N}} \cup \{0\}\} \times \ldots$$

where

$$V_i = \{a_{i,1}\} \times \{a_{i,2}\} \times \dots \{a_{i,N_1}\} \times \mathbb{N} \times \mathbb{N} \times \dots$$

or $V_i = \{0\}$. Here, $a_{i,j} \in \mathbb{N}$ and $N_1, N_2 \in \mathbb{N}$ are arbitrary natural numbers. We now define the mappings

$$\begin{aligned} D(t)_{M_1,M_2} &: \{\mathbb{N}^{\mathbb{N}} \cup \{0\}\}^{\mathbb{Z}} \to \mathbb{N}_0, \\ \Big(A(i), S_{i,1}, \dots, S_{i,A(i)}\Big)_{i\in\mathbb{Z}} &\mapsto \sum_{j=1}^{M_2} \sum_{l=1}^{\max\{A(t-j),M_1\}} \mathbf{1}_{\{S_{t-j,l}=j\}}, \end{aligned}$$

for fixed natural numbers $M_1, M_2$. Now, let $k \in \mathbb{N}_0$ such that $k \leq M_1 M_2$. Then $(D(t)_{M_1,M_2})^{-1}(\{k\})$ is a (finite) union of sets $U_{M_1,M_2,t}$, where the entries $a_{ij}$ of the $V_i$'s satisfy the restrictions

$$|a_{t-1,l_1} = 1| = k_1, \dots, |a_{t-M_2,l_{M_2}} = M_2| = k_{M_2} \quad \text{for } 1 \leq l_i \leq M_1, \sum_{i=1}^{M_2} k_i = k,$$

if $V_{t-j} = \{0\}$, we consider the corresponding set empty. If $k > M_1 M_2$ then the preimage is the empty set, which is open. It follows that the preimage $(D(t)_{M_1,M_2})^{-1}(\{k\})$ of any $k \in \mathbb{N}$ is an element of $\mathcal{F}_\infty(\xi)$. Since the set of integers is a generator for $\sigma(\mathbb{N}_0)$, it follows that the map $D(t)_{M_1,M_2}$ is $\mathcal{F}_\infty(\xi)$-$\sigma(\mathbb{N}_0)$-measurable for all finite natural numbers $M_1, M_2$.

Now, for any fixed $M_2$, the sequence $(D(t)_{M_1,M_2})_{M_1\in\mathbb{N}}$ is a sequence of measurable functions, thus $\lim_{M_1\to\infty} D(t)_{M_1,M_2} := D(t)_{M_2}$ is $\mathcal{F}_\infty(\xi)$-$\sigma(\overline{\mathbb{N}}_0)$-measurable where $\infty$ is included in the image as the limit might be unbounded. As yet another sequence of measurable functions, the assertion follows for $\lim_{M_2\to\infty} D(t)_{M_2} := D(t)$. For the second assertion, a similar argument shows measurability of $\mathbf{1}_{\{Z(t)\leq x\}}$. The result follows since a product of measurable functions is measurable. □

The ergodicity of the sequence $\big(D(i)\mathbf{1}_{\{Z(i)\leq x\}}\big)_{i\in\mathbb{Z}}$, which is stationary due to the model assumptions made for the i.i.d. sequences $(S_{k,\cdot})_{k\in\mathbb{Z}}$ and $(A(t))_{t\in\mathbb{Z}}$ now follows immediately. We record it in the following Lemma, the result of which was shown with a different proof in Lemma 2 in Edelmann and Wichelhaus (2014). Their proof follows in the same vein as that of Lemma 1 in Brown (1970) and Proposition 3 in Blanghaps et al. (2013).

**Lemma 3.1.3** (Cp. Edelmann and Wichelhaus (2014), Lemma 2)**.** *Let $x \in \mathbb{N}$, then the sequences $\big(D(i)\mathbf{1}_{\{Z(i)\leq x\}}\big)_{i\in\mathbb{Z}}$ and $(D(i))_{i\in\mathbb{Z}}$ are stationary and ergodic.*

As a first consequence of Lemma 3.1.3, Birkhoff's ergodic theorem yields, for all $x \in \mathbb{N}$,

$$\frac{1}{n}\sum_{i=1}^{n} D(i)\mathbf{1}_{\{Z(i)\leq x\}} \to \mathbb{E}\left[D(0)\mathbf{1}_{\{Z(0)\leq x\}}\right] \quad \text{a.s.,}$$

so that $\widehat{H_n}(x) \to H(x)$ a.s. by the continuous mapping theorem. The relation between $G(x)$ and $H(x)$ as well as between $\widehat{G_n}(x)$ and $\widehat{H_n}(x)$ is a continuous one (see (3.5)). Additionally, the a.s. convergence $\hat{c}_n \to c$ is obvious since the sequence $(A(t))_{t\in\mathbb{Z}}$ is i.i.d. and we assumed $\mathrm{Var}(A(0)) < \infty$. We use the continuous mapping theorem to find

$$\widehat{G_n}(x) \to G(x) \quad \text{a.s.},$$

which holds pointwise for all $x \in \mathbb{N}$. This shows the consistency of the estimator $\widehat{G_n}(x)$.. In particular, it follows that $\widehat{G_n}(x) \to 1$ for $x \to \infty$.

### 3.1.4 Moment Relations and Bounds

Let us now record some rather technical results of moment relations and inequalities for various functions of the sequence of differences. These relations will be used in the proofs throughout this chapter.

**Lemma 3.1.4** (Schweer and Wichelhaus (2015a), Lemma 2.1)**.** *For $i \geq 1, k \geq 1$ and $j < i$ it holds that*

$$a) \quad \mathbb{E}\left[\mathbf{1}_{\{Z(i)>k\}}\middle|\mathcal{F}_0(\xi)\right] = \begin{cases} c^k, & i > k, \\ c^{i-1}\mathbf{1}_{\{Z(1)>k-i+1\}}, & i \leq k, \end{cases}$$

$$b) \quad \mathbb{E}\left[\mathbf{1}_{\{Z(i)>k\}}\sum_{l=1}^{A(i-j)}\mathbf{1}_{\{S_{i-j,l}=j\}}\middle|\mathcal{F}_0(\xi)\right] = \begin{cases} 0, & j \in \{1,\dots,k\}, \\ \mathbb{E}[D(0)]g_j c^k, & j \in \{k+1,\dots,i-1\}, \end{cases}$$

*where the set $\{k+1,\dots,i-1\}$ is considered empty if $i \leq k$.*

*Proof.* From the definition of the random variable $Z(i)$ it is clear that we can write

$$Z(i) = \mathbf{1}_{\{A(i-1)=0\}}\left(Z(i-1)+1\right) + \mathbf{1}_{\{A(i-1)>0\}} = \mathbf{1}_{\{A(i-1)=0\}}Z(i-1)+1.$$

This implies

$$\mathbf{1}_{\{Z(i)>k\}} = \mathbf{1}_{\{\mathbf{1}_{\{A(i-1)=0\}}Z(i-1)>k-1\}} = \mathbf{1}_{\{A(i-1)=0\}}\mathbf{1}_{\{Z(i-1)>k-1\}},$$

and with the tower rule for conditional expectations we find

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{\{Z(i)>k\}}\middle|\mathcal{F}_0(\xi)\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\{A(i-1)=0\}}\mathbf{1}_{\{Z(i-1)>k-1\}}\middle|\mathcal{F}_{i-2}(\xi)\right]\middle|\mathcal{F}_0(\xi)\right] \\ &= c\mathbb{E}\left[\mathbf{1}_{\{Z(i-1)>k-1\}}\middle|\mathcal{F}_0(\xi)\right]\end{aligned}$$

for $i \geq 2$, where the last equation used that $Z(i-1)$ is $\mathcal{F}_{i-2}(\xi)$-measurable and that $A(i-1)$ is independent of $\mathcal{F}_{i-2}(\xi)$. The case $i=2$ is established directly. By definition, $\mathbb{E}\left[\mathbf{1}_{\{Z(i)>0\}}\middle|\mathcal{F}_0(\xi)\right] = 1$ for all $i \in \mathbb{N}$, and due to the measurability of the random variables involved, $\mathbb{E}\left[\mathbf{1}_{\{Z(1)>k\}}\middle|\mathcal{F}_0(\xi)\right] = \mathbf{1}_{\{Z(1)>k\}}$ for all $k \in \mathbb{N}$. This allows us to prove relation a) for $i \geq 2$ recursively. For $i=1$ the statement is trivial.

To prove b), we first consider the case $j = 1$ and $i > 2$. This expectation equals, with the tower rule,

$$\mathbb{E}\left[\mathbb{E}\left[\sum_{l=1}^{A(i-1)} \mathbf{1}_{\{S_{i-1,l}=1\}}\mathbf{1}_{\{A(i-1)=0\}}\mathbf{1}_{\{Z(i-1)>k-1\}}\middle|\mathcal{F}_{i-2}(\xi)\right]\middle|\mathcal{F}_0(\xi)\right]$$
$$= \mathbb{E}\left[\sum_{l=1}^{A(i-1)} \mathbf{1}_{\{S_{i-1,l}=1\}}\mathbf{1}_{\{A(i-1)=0\}}\right]\mathbb{E}\left[\mathbf{1}_{\{Z(i-1)>k-1\}}\middle|\mathcal{F}_0(\xi)\right] = 0,$$

as $Z(i-1)$ is $\mathcal{F}_{i-2}(\xi)$-measurable and $A(i-1), S_{i-1,\cdot}$ are independent of $\mathcal{F}_{i-2}(\xi)$. Just as in the proof of the first assertion, this argumentation can be extended recursively for all $j \in \{1, \ldots, k\}$. Also as in the proof of the first assertion, the case $i = 2$ follows directly, without invoking the tower rule. Now, if $j \in \{k+1, \ldots, i-1\}$ then $i > k$. By the definition of $Z(i)$, the random variable $\mathbf{1}_{\{Z(i)>k\}}$ depends only on the arrivals $A(i-1), \ldots, A(i-k)$ and is independent of the random variables $A(i-k-1), \ldots, A(1)$. Thus, we have

$$\mathbb{E}\left[\mathbf{1}_{\{Z(i)>k\}}\sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}}\middle|\mathcal{F}_0(\xi)\right] = \mathbb{E}\left[\mathbf{1}_{\{Z(i)>k\}}\middle|\mathcal{F}_0(\xi)\right]\mathbb{E}\left[\sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}}\right].$$

As $\mathbb{E}[\sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}}] = \mathbb{E}[A(0)]g_j$ by Wald's equation, cf. (3.2), and $\mathbb{E}[D(0)] = \mathbb{E}[A(0)]$, application of a) for the case $i > k$ thus proves b). □

The next result provides an upper bound for an expression which appears several times during the course of this chapter.

**Lemma 3.1.5** (Schweer and Wichelhaus (2015a), Lemma 2.2)**.** *Assuming all expressions involved are finite, there exists a finite number $K$ such that for each $x \in \mathbb{N}$, $1 \le i \le x$ and $y \in \mathbb{N}_0$,*

$$\operatorname{Var}\left(\left(c^{i-1}\mathbf{1}_{\{Z(1)>x-i+1\}} - c^y(1-G(y))\right)\sum_{j=i}^{\infty}\sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}}\right)$$
$$\le \left[c^{2y}(1-G(y))^2 - 2c^{x+y}(1-G(y)) + c^{x+i-1}\right]\frac{\mathbb{E}[A(0)^2]K}{1-c}(1-G(i-1)).$$

*Proof.* We simplify the notation in this proof by setting $r_j := \sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}}$ and $R_i := \sum_{j=i}^{\infty} r_j$, recall that the second moment of $R_i$ is finite by (3.2). Let $z \in \mathbb{N}$, application of the law of total probability yields $\mathbb{E}[r_{i+z}^q | A(-z) > 0] = \frac{1}{1-c}\mathbb{E}[r_{i+z}^q]$ for $q \in \mathbb{N}$. For $q = 1$, $\mathbb{E}[r_{i+z}^q] = \mathbb{E}[A(0)]g_{i+z}$ by Wald's equation. For $q = 2$, the same equation and the inequalities established in the discussion following that expression for

$$\operatorname{Var}(r_{i+z}) = \mathbb{E}[A(0)](g_{i+z} - g_{i+z}^2) + g_{i+z}^2\operatorname{Var}(A(0)) \le \mathbb{E}[A(0)^2]g_{i+z}. \qquad (3.6)$$

We obtain

$$\mathbb{E}\left[r_{i+z}^2\right] \leq \mathbb{E}[A(0)^2]g_{i+z} + \mathbb{E}[A(0)]^2 g_{i+z}^2 \leq 2\mathbb{E}[A(0)^2]g_{i+z},$$

since Jensen's inequality implies $\mathbb{E}[A(0)]^2 \leq \mathbb{E}[A(0)^2]$. Together with the independence of $(A_t)_{t\in\mathbb{Z}}$, this implies $\mathbb{E}[R_{i+z+1}^2] \leq 2\mathbb{E}[A(0)^2](1-G(i+z))$. Now, the event $Z(1)=z$ entails that $A(0)=\cdots=A(-z+1)=0$ and $A(z)>0$. Combining all of these results with the linearity of the expectation, we find

$$\begin{aligned}\mathbb{E}\left[R_i^2|Z(1)=z\right] &= \mathbb{E}\left[R_{i+z+1}^2\right] + 2\mathbb{E}\left[R_{i+z+1}\right]\mathbb{E}[r_{i+z}|A(-z)>0] + \mathbb{E}[r_{i+z}^2|A(-z)>0] \\ &\leq 2\mathbb{E}[A(0)^2](1-G(i+z)) + \frac{2\mathbb{E}[A(0)]^2(1-G(i+z))g_{i+z}}{1-c} + \frac{2\mathbb{E}[A(0)^2]g_{i+z}}{1-c}.\end{aligned}$$

As $G(\cdot)$ is a cdf, it is monotonously increasing in its argument and there exists a positive constant $K$ such that this expression is bounded by $(K\mathbb{E}[A(0)^2](1-G(i-1))/(1-c)$. This bound holds for all $z \in \mathbb{N}$ and can thus be extended to conditions of the form $\{z \in B \subseteq \mathbb{N}\}$ by the law of total probability. Setting $A_0 = \{Z(1) > x-i+1\}$ and $A_1 = \{Z(1) \leq x-i+1\}$, we first have $\mathbb{P}(A_0) = c^{x-i+1}$. The assertion is now an easy consequence of the inequality $\mathrm{Var}(R_i) \leq \mathbb{E}[R_i^2]$ and the law of total probability, i.e. $\mathbb{E}[R_i^2] = \mathbb{P}(A_0)\mathbb{E}[R_i^2|A_0] + \mathbb{P}(A_1)\mathbb{E}[R_i^2|A_1]$. □

## 3.2 A Functional Central Limit Theorem for the Service Time Estimator

The estimator $\widehat{G}_n(x)$ was shown to be consistent in the previous section, thus ensuring that it eventually converges to the true value $G(x)$ for each $x \in \mathbb{N}$. For application purposes, such a result leaves something to be desired, since without knowledge of the speed of convergence the practitioner cannot gauge the quality of the estimation. A different problem is that each point $x \in \mathbb{N}$ is considered separately, yet one is usually more interested in the behavior of the entire distribution. Both of these concerns will be addressed by providing a functional central limit theorem for the service time estimator at the end of this section.

### 3.2.1 Finite Dimensional CLTs

We first prove finite dimensional CLTs for the estimator of the distribution function $H$, i.e., for a vector of the form $(\widehat{H}_n(x_1), \ldots, \widehat{H}_n(x_k))$. Due to the special structure of these estimators, see (3.4), it is necessary to first show CLTs for the numerator and the denominator and then combine these results with an appropriate expansion of the terms. The former results are shown in Theorem 3.2.1 and Corollary 3.2.2, respectively, the latter in Theorem 3.2.4.

In view of the two CLTs given in Section 2.5, we opt for employing Theorem 2.5.3 rather than trying to prove a mixing condition for the sequences involved. The reason for this is twofold: Firstly, the conditional expectation is more easily accessible as we have a lot of structure within our processes and thus (2.14) can be shown to hold for both sequences involved in the estimator (3.4). We refer to the proofs of the following theorems for details. Secondly, we remark that the establishment of classical mixing conditions in the framework of the continuous time queuing model proved elusive for Brown, as he states in his paper (cf. (Brown, 1970, p. 653)) that he "has been unable to verify the mixing conditions given by Billingsley [...]". We point out that he referred here to the first edition of Billingsley (1999), whereas the condition we apply here was published 30 years later, in the second edition of this book. We further point out that the mentioned mixing conditions are closely linked (yet not equal) to the concept of $\alpha$-mixing discussed in Section 2.5.

**Theorem 3.2.1** (Schweer and Wichelhaus (2015a), Theorem 3.1). *Let* $\mathbb{E}[A(0)^2] < \infty$ *and* $\sum_{n=1}^{\infty} \sqrt{1-G(n)} < \infty$. *Then, for each* $x \in \mathbb{N}$,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} D(i)\mathbf{1}_{\{Z(i)\leq x\}} - \mathbb{E}\left[D(0)\mathbf{1}_{\{Z(0)\leq x\}}\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma_x^2),$$

*where* $\sigma_x^2 = \operatorname{Var}\left(D(0)\mathbf{1}_{\{Z(0)\leq x\}}\right) + 2\sum_{j=1}^{\infty} \operatorname{Cov}\left(D(0)\mathbf{1}_{\{Z(0)\leq x\}}, D(j)\mathbf{1}_{\{Z(j)\leq x\}}\right)$.

*Proof.* Let $x \in \mathbb{N}$. By Lemma 3.1.3, the sequence $(D(i)\mathbf{1}_{\{Z(i)\leq x\}})_{i\in\mathbb{Z}}$ is stationary and ergodic. A direct implication of (3.2) is that the random variables $D(i)\mathbf{1}_{\{Z(i)\leq x\}}$ have finite second moments. It remains to be seen that the condition (2.14) is satisfied. We remark that $\sigma\left(D(i)\mathbf{1}_{\{Z(i)\leq x\}}; i\leq k\right) \subset \sigma\left(\xi(i); i\leq k\right) = \mathcal{F}_k(\xi)$, where the $\xi(i)$'s are the enlarged process defined in Section 3.1. Using (2.17) and the stationarity of $(D(i)\mathbf{1}_{\{Z(i)\leq x\}})_{i\in\mathbb{Z}}$, it suffices to show that

$$\sum_{i=1}^{\infty} \left\|\mathbb{E}\left[D(i)\mathbf{1}_{\{Z(i)\leq x\}} - \mathbb{E}\left[D(0)\mathbf{1}_{\{Z(0)\leq x\}}\right] \middle| \mathcal{F}_0(\xi)\right]\right\|_{L^2} < \infty. \tag{3.7}$$

First, let us consider the case $i > x$. In the first step, we separate the random variable given by the conditional expectation in the expression above into its probabilistic and deterministic parts. For instance, the arrivals occurring after the time slot 0 are independent of $\mathcal{F}_0(\xi)$ and, since $i > x$, so is the random variable $\mathbf{1}_{\{Z(i)\leq x\}}$. Similarly, since the random variable $\mathbf{1}_{\{Z(i)\leq x\}}$ is independent of the behavior of the process before time $i - x > 0$, we have

$$\begin{aligned}
&\mathbb{E}\left[D(i)\mathbf{1}_{\{Z(i)\leq x\}}\middle|\mathcal{F}_0(\xi)\right] - \mathbb{E}\left[D(i)\mathbf{1}_{\{Z(i)\leq x\}}\right] \\
&= \mathbb{E}\left[\mathbf{1}_{\{Z(i)\leq x\}}\right]\sum_{j=i}^{\infty}\sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}} - \mathbb{E}\left[\mathbf{1}_{\{Z(i)\leq x\}}\right]\mathbb{E}\left[\sum_{j=i}^{\infty}\sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}}\right].
\end{aligned}$$

Thus, as $\mathbb{E}[\mathbf{1}_{\{Z(i)\leq x\}}] = 1 - c^x$, and ince the random variables $S_{i-j_1,\cdot}$ and $S_{i-j_2,\cdot}$ are independent for $j_1 \neq j_2$, we find with (3.6)

$$\left\|\mathbb{E}\left[D(i)\mathbf{1}_{\{Z(i)\leq x\}} - \mathbb{E}\left[D(0)\mathbf{1}_{\{Z(0)\leq x\}}\right] \middle| \mathcal{F}_0(\xi)\right]\right\|_{L^2}^2 \leq (1-c^x)^2\, \mathbb{E}[A(0)^2](1-G(i-1)).$$

Now, let us consider the case $i \leq x$. We will use a similar approach as above, separating the conditional expectation in its probabilistic and deterministic parts. First,

$$\begin{aligned}
&\mathbb{E}\left[D(i)\mathbf{1}_{\{Z(i)\leq x\}}\middle|\mathcal{F}_0(\xi)\right] \\
&=\sum_{j=i}^{\infty}\sum_{l=1}^{A(i-j)}\mathbf{1}_{\{S_{i-j,l}=j\}}\mathbb{E}\left[\mathbf{1}_{\{Z(i)\leq x\}}\middle|\mathcal{F}_0(\xi)\right]+\mathbb{E}\left[\sum_{j=1}^{i-1}\sum_{l=1}^{A(i-j)}\mathbf{1}_{\{S_{i-j,l}=j\}}\mathbf{1}_{\{Z(i)\leq x\}}\middle|\mathcal{F}_0(\xi)\right] \\
&=\sum_{j=i}^{\infty}\sum_{l=1}^{A(i-j)}\mathbf{1}_{\{S_{i-j,l}=j\}}\left(1-c^{i-1}\mathbf{1}_{\{Z(1)>x-i+1\}}\right)+\mathbb{E}\left[\sum_{j=1}^{i-1}\sum_{l=1}^{A(i-j)}\mathbf{1}_{\{S_{i-j,l}=j\}}\right],
\end{aligned}$$

the second equality used Lemma 3.1.4 and the fact that the random variables $A(i), S_{i,l}$ are independent of $\mathcal{F}_0(\xi)$ for $i > 0$. Since the tower rule for conditional expectations implies that $\mathbb{E}[\mathbb{E}[D(i)\mathbf{1}_{\{Z(i)\leq x\}}|\mathcal{F}_0(\xi)]] = \mathbb{E}[D(i)\mathbf{1}_{\{Z(i)\leq x\}}]$, it follows that

$$\begin{aligned}
&\left\|\mathbb{E}\left[D(i)\mathbf{1}_{\{Z(i)\leq x\}}\middle|\mathcal{F}_0(\xi)\right]-\mathbb{E}\left[D(i)\mathbf{1}_{\{Z(i)\leq x\}}\right]\right\|_{L^2} \\
&=\operatorname{Var}^{\frac{1}{2}}\left(\sum_{j=i}^{\infty}\sum_{l=1}^{A(i-j)}\mathbf{1}_{\{S_{i-j,l}=j\}}\left(1-c^{i-1}\mathbf{1}_{\{Z(1)>x-i+1\}}\right)\right),
\end{aligned}$$

for which we find an upper bound using Lemma 3.1.5 and setting $y = 0$, notice that $G(0) = 0$. We are now able to combine the results for the cases $i > x$ and $i \leq x$. Changing the summation index for convenience, we obtain an upper bound for (3.7)

$$\sqrt{\mathbb{E}[A(0)^2]}\left[\sum_{i=0}^{x-1}\sqrt{\left(1-2c^x+c^{x+i}\right)\frac{K}{1-c}(1-G(i))}+(1-c^x)\sum_{i=x}^{\infty}\sqrt{1-G(i)}\right].$$

Since $1 - 2c^x + c^{x+i} \leq 1$ for all $0 \leq i \leq x - 1$ this expression has an upper bound $\sqrt{\frac{\mathbb{E}[A(0)^2]K}{1-c}}\sum_{i=0}^{\infty}\sqrt{1-G(i)}$ and since $\sum_{i=0}^{\infty}\sqrt{1-G(i)} < \infty$ by assumption, this concludes the proof. □

We point out that the inequality (3.6), which is used at a crucial part of the proof of Theorem 3.2.1, is not a very rough estimate. Apart from the upper bound on the moments, it can not be improved upon without a loss of generality. To see this, consider the very common assumption of Poisson distributed arrivals. In this case, the resultant variance in (3.6) actually equals $\lambda g_j$ rendering the upper bound found for (3.7) sharp. The following result is an easy consequence of Theorem 3.2.1, it shows that the denominator of the estimator (3.4) also obeys a CLT.

**Corollary 3.2.2** (Schweer and Wichelhaus (2015a), Corollary 3.2)**.** *Let* $\mathbb{E}[A(0)^2] < \infty$ *and* $\sum_{n=1}^{\infty}\sqrt{1-G(n)} < \infty$. *Then*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}D(i)-\mathbb{E}[D(0)]\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,\sigma^2),$$

*where* $\sigma^2 = \operatorname{Var}(D(0)) + 2\sum_{j=1}^{\infty}\operatorname{Cov}(D(0), D(j))$.

*Proof.* The sequence $(D(t))_{t\in\mathbb{Z}}$ is stationary and ergodic by Lemma 3.1.3, the second moment is finite due to the finiteness of $\mathbb{E}[A(0)^2]$ and (3.2). Now,

$$\mathbb{E}\left[D(i)|\mathcal{F}_0(\xi)\right] = \sum_{j=i}^{\infty}\sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}} + \mathbb{E}\left[\sum_{j=1}^{i-1}\sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}}\right],$$

so that similar to the proof of Theorem 3.2.1 we find that

$$\begin{aligned}\sum_{i=1}^{\infty}\left\|\mathbb{E}\left[D(i)-\mathbb{E}\left[D(0)\right]\Big|\mathcal{F}_0(\xi)\right]\right\| &= \sum_{i=1}^{\infty}\operatorname{Var}^{\frac{1}{2}}\left(\sum_{j=i}^{\infty}\sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}}\right)\\ &\overset{(3.6)}{\leq} \sqrt{\mathbb{E}[A(0)^2]}\sum_{i=0}^{\infty}\sqrt{1-G(i)},\end{aligned}$$

which is finite by assumption. □

Concerning the condition $\sum_{n=1}^{\infty}\sqrt{1-G(n)}<\infty$ employed in Theorem 3.2.1, it can be shown that a simple moment condition on the distribution $G$ implies this condition. Indeed, it is easily seen that $1-G(n)\leq\frac{1}{(n+1)^{2+\epsilon}}\sum_{j=1}^{\infty}g_j j^{2+\epsilon}$ for all $n\in\mathbb{N}$ and $\epsilon>0$, so that the following result immediately follows.

**Lemma 3.2.3** (Schweer and Wichelhaus (2015a), Lemma 3.3)**.** *Let $X$ be a random variable with distribution $G$. If $\mathbb{E}[|X|^{2+\epsilon}]<\infty$ for some $\epsilon>0$, then it follows that $\sum_{n=1}^{\infty}\sqrt{1-G(n)}<\infty$.*

In this section, we now piece together the asymptotic normality of $\widehat{H}_n(x_i)$ and $\hat{c}_n$. The first crucial step is given by the following result.

**Theorem 3.2.4** (Schweer and Wichelhaus (2015a), Theorem 3.4)**.** *Let $x_1,\ldots,x_l\in\mathbb{N}$, $l\in\mathbb{N}$, let $\mathbb{E}[A(0)^2]<\infty$ and $\sum_{n=1}^{\infty}\sqrt{1-G(n)}<\infty$. Then*

$$\sqrt{n}\begin{pmatrix}\hat{c}_n-c\\ \widehat{H}_n(x_1)-H(x_1)\\ \vdots\\ \widehat{H}_n(x_l)-H(x_l)\end{pmatrix}\overset{\mathcal{D}}{\to}\mathcal{N}(0,\mathbf{T}'),$$

*where the entries $\tau'_{i,j}$ of $\mathbf{T}'$ are given as follows: $\tau'_{1,1}=c(1-c)$,*

$$\tau'_{1,k+1}=\frac{1}{\mathbb{E}[D(0)]}\sum_{i=0}^{\infty}\mathbb{E}\left[D(i)\left(\mathbf{1}_{\{Z(i)\leq x_k\}}-H(x_k)\right)\mathbf{1}_{\{A(0)=0\}}\right]\quad\text{and}$$

$$\tau'_{m+1,k+1}=\frac{1}{\mathbb{E}[D(0)]^2}\sum_{i=-\infty}^{\infty}\mathbb{E}\Big[D(0)D(i)\left(H(x_m)-\mathbf{1}_{\{Z(0)\leq x_m\}}\right)\left(H(x_k)-\mathbf{1}_{\{Z(i)\leq x_k\}}\right)\Big]$$

*for $k,m\leq l$.*

*Proof.* First, let us define $\left(\widehat{H}_n(x_1), \ldots, \widehat{H}_n(x_l)\right)^T := \mathbf{H}_n$, $(H(x_1), \ldots, H(x_l))^T := \mathbf{H}$ and

$$\eta_i(k) := \frac{D(i)\left(\mathbf{1}_{\{Z(i)\leq k\}} - H(k)\right)}{\mathbb{E}[D(0)]}, \tag{3.8}$$

notice that $\mathbb{E}[\eta_i(k)] = 0$ for all $i \in \mathbb{N}_0$, $k \in \mathbb{N}$ by (3.4). We expand, for each $k \in \mathbb{N}$,

$$\widehat{H}_n(k) - H(k) = \frac{\frac{1}{n}\sum_{i=1}^n D(i)\mathbf{1}_{\{Z(i)\leq k\}} - H(k)\frac{1}{n}\sum_{i=1}^n D(i)}{\frac{1}{n}\sum_{i=1}^n D(i)},$$

which leads to

$$\sqrt{n}\left(\mathbf{H}_n - \mathbf{H}\right) = \sqrt{n}\frac{\mathbb{E}[D(0)]}{\frac{1}{n}\sum D(i)} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \eta_i(x_1) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n \eta_i(x_l) \end{pmatrix}. \tag{3.9}$$

Now, let $(t_0, t_1, \ldots, t_l) \in \mathbb{R}^{l+1}$ and consider the sequence of random variables

$$\vartheta_i := t_0\left(\mathbf{1}_{\{A(i)=0\}} - c\right) + \sum_{j=1}^{l} t_j\left(\eta_i(x_j)\right).$$

It is obviously stationary and ergodic and has finite second moments by (3.2), we now show that this sequence satisfies condition (2.14). Clearly,

$$\sigma\left(\vartheta_i; i \leq k\right) \subset \sigma\left(\xi(i); i \leq k\right) = \mathcal{F}_k(\xi).$$

By (2.17) and the triangle inequality of the $L^2$-norm we calculate an upper bound for (2.14):

$$\begin{aligned}
&\sum_{i=1}^{\infty} \left\|\mathbb{E}\left[\vartheta_i|\mathcal{F}_0(\xi)\right]\right\|_{L^2} \\
&\leq |t_0|\sum_{i=1}^{\infty}\left\|\mathbb{E}\left[\mathbf{1}_{\{A(i)=0\}} - c|\mathcal{F}_0(\xi)\right]\right\|_{L^2} + \sum_{i=1}^{\infty}\sum_{j=1}^{l}|t_j|\left\|\mathbb{E}\left[\eta_i(x_j)|\mathcal{F}_0(\xi)\right]\right\|_{L^2}.
\end{aligned}$$

Absolute convergence of the latter series is ensured by the proof of Theorem 3.2.1 and Corollary 3.2.2 and the triangle inequality in $L^2$. For the former series, notice that since $A(i)$ is independent of $\mathcal{F}_0(\xi)$ for all $i > 0$,

$$\left\|\mathbb{E}\left[\mathbf{1}_{\{A(i)=0\}} - c\middle|\mathcal{F}_0(\xi)\right]\right\|_{L^2} = 0$$

for $i > 0$. This implies finiteness of the entire expression, and in conclusion that $\frac{1}{\sqrt{n}}\sum_{i=1}^n \vartheta_i$ converges in distribution to a $\mathcal{N}(0,\xi)$-distribution. For the calculation of $\xi$ we use the fact that $\mathbb{E}[\eta_i(x_j)] = 0$ for any $i \in \mathbb{N}_0$, $j \in \mathbb{N}$ as well as the independence of $\mathbf{1}_{\{A(i)=0\}}$ and $\eta_0(x_j)$ for all $x_j$ and $i > 0$. Straightforward algebra yields

$$\xi = \sum_{j_1,j_2=1}^{k} t_{j_1} t_{j_2} \left[ \mathbb{E}\left[\eta_0(x_{j_1})\eta_0(x_{j_2})\right] + \sum_{i=1}^{\infty} \left( \mathbb{E}\left[\eta_0(x_{j_1})\eta_i(x_{j_2})\right] + \mathbb{E}\left[\eta_i(x_{j_1})\eta_0(x_{j_2})\right] \right) \right]$$
$$+ \sum_{j_1=1}^{k} t_{j_1} t_0 \left[ \mathbb{E}\left[\eta_0(x_{j_1})\mathbf{1}_{\{A(0)=0\}}\right] + \sum_{i=1}^{\infty} \mathbb{E}\left[\eta_i(x_{j_1})\mathbf{1}_{\{A(0)=0\}}\right] \right] + t_0^2 \operatorname{Var}\left(\mathbf{1}_{\{A(0)=0\}}\right),$$

which is easily seen to be the variance of the random variable $(t_0, t_1, \ldots, t_l) \cdot \mathbf{X}$, given by $\mathbf{X} \sim \mathcal{N}(0, \mathbf{T}')$. As we chose $(t_0, t_1, \ldots, t_l) \in \mathbb{R}^l$ arbitrarily, we may apply the Cramér-Wold device to show that $\sqrt{n}(\mathbf{H}_n - \mathbf{H})$ converges weakly to a multivariate normal distribution with zero mean and covariance matrix given by $\mathbf{T}'$. This follows from (3.9) and the application of Slutsky's Lemma, as the factor in this expression converges a.s., concluding the proof. □

The apparent differences in the expressions for the asymptotic covariances given in Theorem 3.2.1 and the proof of Theorem 3.2.4 are only a matter of notation, as the stationarity of the sequences involved ensures that, e.g.,

$$\mathbb{E}\left[\eta_0(k)\eta_0(m)\right] + \sum_{i=1}^{\infty} \left( \mathbb{E}\left[\eta_0(k)\eta_i(m)\right] + \mathbb{E}\left[\eta_i(k)\eta_0(m)\right] \right) = \sum_{i=-\infty}^{\infty} \mathbb{E}\left[\eta_0(k)\eta_i(m)\right].$$

### 3.2.2 Tightness

In this section, we provide an important technical result for the proof of tightness. Note that we do not have to make any further assumptions about our model for the assertion to hold. Thus, our result remains applicable to a wide range of models.

**Theorem 3.2.5** (Schweer and Wichelhaus (2015a), Theorem 4.1)**.** *Let* $\mathbb{E}[A(0)^2] < \infty$ *and* $\sum_{n=1}^{\infty} \sqrt{1 - G(n)} < \infty$. *Then* $\sum_{x=1}^{\infty} \sum_{i=-\infty}^{\infty} \mathbb{E}[\eta_0(x)\eta_i(x)]$ *converges absolutely.*

*Proof.* Let $\eta_i(x)$ be defined as in (3.8) for $i \in \mathbb{N}_0$, $x \in \mathbb{N}$. Clearly, it suffices to show that $\sum_{x=1}^{\infty} \sum_{i=0}^{\infty} \mathbb{E}[\eta_0(x)\eta_i(x)]$ converges absolutely. Furthermore, applying the Cauchy-Schwarz inequality as in (2.15) implies that

$$\left|\mathbb{E}\left[\eta_0(x)\eta_i(x)\right]\right| \leq \|\eta_0(x)\|_{L^2} \cdot \left\|\mathbb{E}\left[\eta_i(x)|\mathcal{F}_0(\xi)\right]\right\|_{L^2}.$$

Using stationarity again, $\|\eta_0(x)\|_{L^2} = \operatorname{Var}^{\frac{1}{2}}(\eta_1(x))$, as $\mathbb{E}[\eta_1(x)] = 0$. The variance corresponds exactly to the case $i = 1$ and $y = x$ in Lemma 3.1.5, so that

$$\mathbb{E}[D(0)]^2 \mathbb{E}[\eta_0(x)^2] = \mathbb{E}[D(0)]^2 \|\eta_0(x)\|_{L^2}^2 \leq \frac{\mathbb{E}[A(0)^2]K}{1-c} c^x, \tag{3.10}$$

as $c^{2x}(1 - G(x))^2 - 2c^{2x}(1 - G(x)) + c^{x+i-1} = c^{2x}(G(x)^2 - 1) + c^x \leq c^x$ for each $i, x \in \mathbb{N}$. By the Weierstraß M-test, $\sum_{x=1}^{\infty} \mathbb{E}[\eta_0(x)^2]$ converges absolutely, and furthermore the term $\|\eta_0(x)\|_{L^2}$ is uniformly bounded for each $x \in \mathbb{N}$. In order to complete the proof, we thus have to show absolute convergence of $\sum_{x=1}^{\infty} \sum_{i=1}^{\infty} \|\mathbb{E}[\eta_i(x)|\mathcal{F}_0(\xi)]\|_{L^2}$.

Just as in the proof of Theorem 3.2.1, we separate the probabilistic part of the conditional expectation from the deterministic part. Since we will be using the result of Lemma 3.1.4, we use $\mathbf{1}_{\{Z(i)\leq x\}} = 1 - \mathbf{1}_{\{Z(i)>x\}}$ and calculate, for each $i, x \in \mathbb{N}$,

$$\begin{aligned}
&\mathbb{E}\left[D(i)\left(H(x) - \mathbf{1}_{\{Z(i)\leq x\}}\right)|\mathcal{F}_0(\xi)\right] \\
&= \sum_{j=1}^{i-1} \mathbb{E}\left[\sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}} \mathbf{1}_{\{Z(i)>x\}} \middle| \mathcal{F}_0(\xi)\right] - (1 - H(x)) \sum_{j=1}^{i-1} \mathbb{E}\left[\sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}}\right] \\
&\qquad + \left(\mathbb{E}\left[\mathbf{1}_{\{Z(i)>x\}} \middle| \mathcal{F}_0(\xi)\right] - (1 - H(x))\right) \sum_{j=i}^{\infty} \sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}}, \qquad (3.11)
\end{aligned}$$

where we used independence or measurability of the random variables $(A(t)_{t\in\mathbb{Z}}$ and $(S_{j,\cdot})_{j\in\mathbb{Z}}$ with respect to $\mathcal{F}_0(\xi)$, as the case may be. Now, the second term in (3.11) is deterministic, and for the first term Lemma 3.1.4 implies that this conditional expectation is also deterministic. For the third term, we need to consider the cases $i > x$ and $i \leq x$ separately, beginning with the former case. Lemma 3.1.4 yields $\mathbb{E}\left[\mathbf{1}_{\{Z(i)>x\}} \middle| \mathcal{F}_0(\xi)\right] = c^x$, thus in this case (3.11) and the law of total expectation imply

$$\mathbb{E}[D(0)] \cdot \|\mathbb{E}[\eta_i(x)|\mathcal{F}_0(\xi)]\|_{L^2} = c^x G(x) \operatorname{Var}^{\frac{1}{2}}\left(\sum_{j=i}^{\infty} \sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}}\right),$$

noticing that $1 - H(x) = c^x(1 - G(x))$ by (3.5). Using the inequality (3.6) for this variance, we find the upper bound $c^x G(x)\sqrt{\mathbb{E}[A(0)^2]}\sqrt{1 - G(i-1)}$ for each $i > x$. For the case $i \leq x$, Lemma 3.1.4 implies that $\mathbb{E}\left[\mathbf{1}_{\{Z(i)>x\}} \middle| \mathcal{F}_0(\xi)\right] = c^{i-1}\mathbf{1}_{\{Z(1)>x-i+1\}}$. Using (3.11) as well as the law of total expectation again, we find

$$\begin{aligned}
&\mathbb{E}[D(0)] \cdot \|\mathbb{E}[\eta_i(x)|\mathcal{F}_0(\xi)]\|_{L^2} \\
&= \operatorname{Var}^{\frac{1}{2}}\left(\left(c^{i-1}\mathbf{1}_{\{Z(1)>x-i+1\}} - c^x(1 - G(x))\right) \sum_{j=i}^{\infty} \sum_{l=1}^{A(i-j)} \mathbf{1}_{\{S_{i-j,l}=j\}}\right).
\end{aligned}$$

We find an upper bound for this variance by applying Lemma 3.1.5 with $y = x$, we further notice that

$$c^{2x}(1 - G(x))^2 - 2c^{2x}(1 - G(x)) + c^{x+i-1} = c^{2x}G(x)^2 + c^{x+i-1} - c^{2x} \leq (c^x G(x) + c^{x/2})^2.$$

Combining the cases $i \leq x$ and $i > x$ and changing the summation index for convenience, we find

$$\mathbb{E}[D(0)] \sum_{i=1}^{\infty} \|\mathbb{E}[\eta_i(x)|\mathcal{F}_0(\xi)]\|_{L^2} \leq \sqrt{\frac{\mathbb{E}[A(0)^2]K}{1-c}} \left(\sum_{i=0}^{\infty} \sqrt{1 - G(i)}\right) \left[c^x G(x) + c^{\frac{x}{2}}\right],$$

and as $c \in (0, 1)$ by assumption, it follows that $\sum_{x=1}^{\infty} \sum_{i=1}^{\infty} \|\mathbb{E}[\eta_i(x)|\mathcal{F}_0(\xi)]\|_{L^2}$ converges. Absolute convergence of the expression is clear from the positivity of the variance, and the proof is concluded by applying the Weierstraß M-test. $\square$

### 3.2.3 Functional Central Limit Theorem

Let us introduce some convenient notation: We denote the sequences associated with the cdf of the distributions $G$ and $H$ by $\mathcal{G} := (G(k))_{k\in\mathbb{N}}$ and $\mathcal{H} := (H(k))_{k\in\mathbb{N}}$ and the respective estimators by $\mathcal{G}_n := (\widehat{G}_n(k))_{k\in\mathbb{N}}$ and $\mathcal{H}_n := (\widehat{H}_n(k))_{k\in\mathbb{N}}$. In this section, we are concerned with convergence in distribution in the separable Banach space $c_0$ of (2.1). We denote expressions of the type $\mathcal{H} - 1$ as $(H(k) - 1)_{k\in\mathbb{N}}$ while scalar multiplication of the form $a\mathcal{H}$ denotes the sequence $(aH(k))_{k\in\mathbb{N}}$. With this notation, the process we are interested in at first is an element of $\mathbb{R} \times c_0$ given by $\sqrt{n}[(\hat{c}_n,\ \mathcal{H}_n - 1) - (c,\ \mathcal{H} - 1)]$, we choose this representation in order to make the application of the functional delta method at the end of this section more transparent. For each $k \in \mathbb{N}$, $\sqrt{n}[\widehat{H}_n(k) - H(k)]$ is a random variable and hence Borel-measurable due to Lemma 3.1.2, implying that the process $\sqrt{n}[(\hat{c}_n,\ \mathcal{H}_n - 1) - (c,\ \mathcal{H} - 1)]$, is a random element of $\mathbb{R} \times c_0$. The following result shows that this process converges weakly to a limiting random element $\mathcal{W} \in \mathbb{R} \times c_0$.

**Theorem 3.2.6** (Schweer and Wichelhaus (2015a), Theorem 4.2)**.** *Let* $\mathbb{E}[A(0)^2] < \infty$ *and* $\sum_{n=1}^{\infty} \sqrt{1 - G(n)} < \infty$*. Then there exists a Gaussian element* $\mathcal{W} = \big(w, (W_k)_{k\in\mathbb{N}}\big)$ *in* $\mathbb{R} \times c_0$ *with zero mean such that* $\mathbb{E}\left[w^2\right] = c(1 - c)$ *as well as*

$$\mathbb{E}\left[wW_m\right] = \frac{1}{\mathbb{E}[D(0)]} \sum_{i=0}^{\infty} \mathbb{E}\left[D(i)\left(\mathbf{1}_{\{Z(i)\leq m\}} - H(m)\right)\mathbf{1}_{\{A(0)=0\}}\right] \quad \textit{and}$$

$$\mathbb{E}\left[W_k W_m\right] = \frac{1}{(\mathbb{E}[D(0)])^2} \sum_{i=-\infty}^{\infty} \mathbb{E}\Big[D(0)D(i)\left(H(m) - \mathbf{1}_{\{Z(0)\leq m\}}\right)\left(H(k) - \mathbf{1}_{\{Z(i)\leq k\}}\right)\Big]$$

*for* $k, m \in \mathbb{N}$*. Moreover, in* $c_0$*,*

$$\sqrt{n}[(\hat{c}_n,\ \mathcal{H}_n - 1) - (c,\ \mathcal{H} - 1)] \xrightarrow{\mathcal{D}} \mathcal{W}.$$

*Proof.* The convergence of the finite-dimensional distributions was shown in Theorem 3.2.4. The tightness of the sequence remains to be established. Since marginal tightness implies joint tightness, it suffices to show that the sequences $\sqrt{n}(\hat{c}_n - c)$ and $\sqrt{n}(\mathcal{H}_n - \mathcal{H})$ are tight. For the former this assertion is obvious, as the random variables $\mathbf{1}_{\{A(i)=0\}}$ are i.i.d. with bounded variance and the classical CLT yields tightness of $\sqrt{n}\,(\hat{c}_n - c)$.

For the latter sequence, we use Lemma 2.1.3. By Theorem 3.2.4 it follows that for each $l \in \mathbb{N}$, $\sqrt{n}(\widehat{H}_n(l) - H(l))$ converges weakly to a normal distribution with zero mean and bounded variance (cf. Theorem 3.2.5), thus condition (2.2) is established. The proof of the second condition is provided below, it is divided into three steps to make it more comprehensible.

**Step 1:**
We show that, with the notation of the proof of Theorem 3.2.4 and $S_n(k) := \sum_{i=1}^{n} \eta_i(k)$, it holds that $\sum_{k=1}^{\infty} \frac{1}{n}\mathbb{E}\left[S_n^2(k)\right]$ is finite. For this, let us recall that in the mentioned proof we showed that the stationary and ergodic sequence $(\eta_i(k))_{i\in\mathbb{Z}}$ satisfies the conditions of Theorem 2.5.3. It can further be shown that, under the assumptions of this Theorem,

for each $n$ (cf. (2.16)):

$$\left|\sum_{i=-\infty}^{\infty} \mathbb{E}[\eta_0(k)\eta_i(k)] - \frac{1}{n}\mathbb{E}\left[S_n^2(k)\right]\right| \leq 2\sum_{l=n}^{\infty} |\mathbb{E}\left[\eta_0(k)\eta_l(k)\right]| + \frac{2}{n}\sum_{i=1}^{n-1}\sum_{l=i}^{\infty} |\mathbb{E}\left[\eta_0(k)\eta_l(k)\right]|.$$

Furthermore,

$$|\mathbb{E}\left[\eta_0(k)\eta_l(k)\right]| \leq \|\eta_0(k)\|_{L^2} \left\|\mathbb{E}\left[\eta_l(k)|\mathcal{F}_0(\xi)\right]\right\|_{L^2},$$

cf. (2.15) and, by (3.10),

$$\mathbb{E}[D(0)] \|\eta_0(k)\|_{L^2} \leq \sqrt{(\mathbb{E}[A(0)^2]K)/(1-c)}$$

for all $k \in \mathbb{N}$. Both inequalities together with Theorem 3.2.5 yield

$$\begin{aligned}
&\left|\sum_{k=1}^{\infty}\sum_{i=-\infty}^{\infty} \mathbb{E}[\eta_0(x)\eta_i(x)] - \sum_{k=1}^{\infty}\frac{1}{n}\mathbb{E}\left[S_n^2(k)\right]\right| \\
&\leq \frac{2}{\mathbb{E}[D(0)]}\sqrt{\frac{\mathbb{E}[A(0)^2]K}{1-c}}\sum_{k=1}^{\infty}\left(\sum_{l=n}^{\infty}\left\|\mathbb{E}\left[\eta_l(k)|\mathcal{F}_0(\xi)\right]\right\|_{L^2} + \frac{1}{n}\sum_{i=1}^{n-1}\sum_{l=i}^{\infty}\left\|\mathbb{E}\left[\eta_l(k)|\mathcal{F}_0(\xi)\right]\right\|_{L^2}\right).
\end{aligned}$$

Since $\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\left\|\mathbb{E}\left[\eta_l(k)|\mathcal{F}_0(\xi)\right]\right\|_{L^2}$ converges absolutely as shown in the proof of Theorem 3.2.5, this expression is finite. As $\sum_{k=1}^{\infty}\sum_{i=-\infty}^{\infty}\mathbb{E}[\eta_0(x)\eta_i(x)]$ is finite by the same result, this implies finiteness of $\sum_{k=1}^{\infty}\frac{1}{n}\mathbb{E}\left[S_n^2(k)\right]$ and the convergence of the latter to the former, since the entire expression tends to 0 for $n \to \infty$. For the former series this is obvious, for the latter this follows from the dominated convergence theorem.

**Step 2:**
For any $\xi > 0$ we denote the event $\left\{\left|\frac{1}{n}\sum_{i=1}^{n} D(i) - \mathbb{E}[D(0)]\right| \leq \xi\right\}$ by $A(n,\xi)$ and the complementary event by $A^c(n,\xi)$. We observe that

$$\mathbb{E}\left[\left(\widehat{H}_n(k) - H(k)\right)^2 \middle| A^c(n,\xi)\right] \leq \mathbb{E}\left[\widehat{H}_n(k) - H(k)\middle| A^c(n,\xi)\right] \leq 1 - H(k),$$

we used that $\mathbb{E}[\widehat{H}_n(k)|A^c(n,\xi)] \leq 1$, as we have $\widehat{H}_n(k) \leq 1$ for each $k,n \in \mathbb{N}$ by definition (cf. (3.4)). Additionally, we used that $|\widehat{H}_n(k) - H(k)| \leq 1$ as both $0 \leq \widehat{H}_n(k) \leq 1$ and $0 \leq H(k) \leq 1$ hold for all $k,n \in \mathbb{N}$.

Now let us consider the behavior of $\mathbb{P}\left(A^c(n,\xi)\right)$ as $n$ increases, we follow the idea of (Kachurovskii, 1996, Theorem 11). It holds that $n\mathbb{P}\left(A^c(n,\xi)\right) \leq \mathrm{Var}\left(\sum_{i=1}^{n} D(i)\right)/(\xi^2 n)$ by Chebyshev's inequality. Recalling the results of Corollary 3.2.2 and the definition of $\sigma^2$ given therein, the ergodic and stationary sequence $(D(t))_{t\in\mathbb{Z}}$ satisfies the conditions of Theorem 2.5.3 and hence necessarily the inequality (2.16) with

$$\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D(i)\right) \leq \sigma^2 + \underbrace{2\sum_{l=n}^{\infty}\mathrm{Cov}\left(D(0),D(l)\right) + \frac{2}{n}\sum_{i=1}^{n-1}\sum_{l=i}^{\infty}\mathrm{Cov}\left(D(0),D(l)\right)}_{=:C(n)},$$

notice that $\lim_{n\to\infty} C(n) = 0$ can be shown as in Step 1. Combining these findings, we have

$$n\mathbb{P}\left(A^c(n,\xi)\right)\mathbb{E}\left[\left(\widehat{H}_n(k)-H(k)\right)^2\middle|A^c(n,\xi)\right] \leq \frac{\sigma^2+C(n)}{\xi^2}\left(1-H(k)\right). \tag{3.12}$$

Now, for each $k \in \mathbb{N}$, $\sqrt{n}(\widehat{H}_n(k) - H(k))$ is equal to (cf. (3.9)),

$$\frac{1}{\sqrt{n}}S_n(k)\left(1+\mathbb{E}[D(0)]\left(\frac{1}{\frac{1}{n}\sum_{i=1}^n D(i)}-\frac{1}{\mathbb{E}[D(0)]}\right)\right) = \frac{1}{\sqrt{n}}S_n(k)\left(\frac{\mathbb{E}[D(0)]}{\frac{1}{n}\sum_{i=1}^n D(i)}\right).$$

With this, for each $k \in \mathbb{N}$,

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{n}S_n^2(k)\left(\frac{\mathbb{E}[D(0)]}{\frac{1}{n}\sum_{i=1}^n D(i)}\right)^2\right] = \mathbb{E}\left[\frac{1}{n}S_n^2(k)\left(\frac{\mathbb{E}[D(0)]}{\frac{1}{n}\sum_{i=1}^n D(i)}\right)^2\left(\mathbf{1}_{\{A(n,\xi)\}}+\mathbf{1}_{\{A^c(n,\xi)\}}\right)\right]\\
&\leq \left(\frac{\mathbb{E}[D(0)]}{\mathbb{E}[D(0)]-\xi}\right)^2\mathbb{E}\left[\frac{1}{n}S_n^2(k)\right] + \mathbb{P}\left(A^c(n,\xi)\right)n\,\mathbb{E}\left[\left(\widehat{H}_n(k)-H(k)\right)^2\middle|A^c(n,\xi)\right]\\
&\overset{(3.12)}{=} \left(\frac{\mathbb{E}[D(0)]}{\mathbb{E}[D(0)]-\xi}\right)^2\mathbb{E}\left[\frac{1}{n}S_n^2(k)\right] + \frac{\sigma^2+C(n)}{\xi^2}\left(1-H(k)\right).
\end{aligned}$$

**Step 3:**
For the final step, let $\epsilon, \delta > 0$. Then, with Markov's inequality,

$$\begin{aligned}
\mathbb{P}\left(\sup_{k\geq l}\sqrt{n}\left|\widehat{H}_n(k)-H(k)\right|>\epsilon\right) &\leq \frac{1}{\epsilon^2}\sum_{k\geq l}\mathbb{E}\left[\frac{1}{n}S_n^2(k)\left(\frac{\mathbb{E}[D(0)]}{\frac{1}{n}\sum_{i=1}^n D(i)}\right)^2\right]\\
&\leq \frac{1}{\epsilon^2}\sum_{k\geq l}\left[\left(\frac{\mathbb{E}[D(0)]}{\mathbb{E}[D(0)]-\xi}\right)^2\mathbb{E}\left[\frac{1}{n}S_n^2(k)\right]+\frac{\sigma^2+C(n)}{\xi^2}\left(1-H(k)\right)\right],
\end{aligned}$$

we used Step 2 for the final inequality. By Step 1, the series $\sum_{k=1}^{\infty}\frac{1}{n}\mathbb{E}\left[S_n^2(k)\right]$ converges, for the series $\sum_{k=1}^{\infty}(1-H(k))$ convergence follows from (3.5). Now, let us define the integer $n_0 := \sup_{k\in\mathbb{N}} C(k)$, which is possible as $\lim_{n\to\infty} C(n) = 0$. As $\xi$ is arbitrary but fixed, it is clear that we are able to find $l_0 \in \mathbb{N}$ such that the expression above is less than $\delta$ for $l \geq l_0$ and $n \geq n_0$. Thus, (2.3) is satisfied as $\epsilon, \delta > 0$ were chosen arbitrarily. Since both conditions (2.2) and (2.3) combined yield tightness of the sequence $\sqrt{n}(\mathcal{H}_n - \mathcal{H})$, this concludes the proof. □

### 3.2.4 The Functional Delta Method

In the preceding sections, we established the asymptotic behavior of the estimator $\mathcal{H}_n$. However, our original goal was the estimation of the service time distribution $G$, by (3.5) this is done via $\widehat{G}_n(x) := 1 - \hat{c}_n^{-x}(1 - \widehat{H_n}(x))$. In the proof of the consistency of the estimator, the continuity of this mapping sufficed. In order to transfer the asymptotic normality, we have to consider the differentiability of the function. Let us define the mapping

$$\begin{aligned} \phi : \mathbb{R} \times c_0 &\to \mathbb{R}^{\mathbb{N}} \\ (a, (x_k)_{k\in\mathbb{N}}) &\mapsto \left(x_k a^{-k}\right)_{k\in\mathbb{N}.} \end{aligned}$$

It follows easily from (3.5) that $\sqrt{n}\phi\left(\hat{c}_n, (\mathcal{H}_n - 1)\right) = \sqrt{n}\left(\mathcal{G}_n - 1\right)$, and thus that

$$\sqrt{n}\ \left[\phi\left(\hat{c}_n, (\mathcal{H}_n - 1)\right) - \phi\left(c, (\mathcal{H} - 1)\right)\right] = \sqrt{n}\left(\mathcal{G}_n - \mathcal{G}\right),$$

our process of interest. Now, $\sqrt{n}\left(\mathcal{G}_n - \mathcal{G}\right) \in c_0$ (since both $\widehat{G}_n(\cdot)$ and $G(\cdot)$ are cdfs, thus both tend to 1). Therefore, the remaining step consists in applying a functional version of the Delta theorem.

**Theorem 3.2.7** (van der Vaart (2000), Theorem 20.8)**.** *Let $\mathbb{D}$ and $\mathbb{E}$ be normed linear spaces. Let $\phi : \mathbb{D}_\phi \subset \mathbb{D} \to \mathbb{E}$ be Hadamard differentiable at $\theta$ tangentially to $\mathbb{D}_0$, i.e., the derivative exists on $\mathbb{D}_0$. Let $T_n : \Omega_n \to \mathbb{D}_\phi$ be maps such that $r_n(T_n - \theta) \to T$ in distribution for some sequence $r_n \to \infty$ and a random element $T$ that takes its values in $\mathbb{D}_0$. Then $r_n(\phi(T_n) - \phi(\theta)) \to \phi'_\theta(T)$ in distribution, where $\phi'_\theta$ denotes the Hadamard derivative of $\phi$ at $\theta$.*

By this theorem, it suffices to show the existence of the Hadamard derivative at $\theta = (c, H(1) - 1, H(2) - 1, \dots) \in \mathbb{R} \times c_0$, denoted by $\phi'_\theta$, on a subset $\mathbb{D}_\phi \subset \mathbb{R} \times c_0$. Thus, we define $\mathbb{D}_\phi := \{(a, (x_k)_{k\in\mathbb{N}}) \in \mathbb{R} \times c_0 \,|\, \phi(a, (x_k)_{k\in\mathbb{N}}) \in c_0\}$, which is obviously is not an empty set, as for instance, $\theta \in \mathbb{D}_\phi$. This specificity is necessary because $\phi$ maps to $\mathbb{R}^{\mathbb{N}}$ and not to $c_0$, as for any sequence $(x_i)_{i\in\mathbb{N}} \in c_0$ it does not hold in general that $\phi(a, (x_k)_{k\in\mathbb{N}}) \in c_0$ for some $a \in \mathbb{R}$. We now show that $\phi$ is Hadamard differentiable.

**Lemma 3.2.8** (Schweer and Wichelhaus (2015a), Lemma 4.3)**.** *Let $\mathbb{E}[A(0)^2] < \infty$ and $\sum_{n=1}^{\infty} \sqrt{1 - G(n)} < \infty$. Then $\phi : \mathbb{D}_\phi \mapsto c_0$ is Hadamard differentiable at the parameter $\theta = (c, H(1) - 1, H(2) - 1, \dots)$ tangentially to $\mathbb{D}_\phi$ and*

$$\phi'_\theta\left(w, (x_k)_{k\in\mathbb{N}}\right) = \left(\frac{x_k}{c^k} - k\frac{1 - H(k)}{c^{k+1}}w\right)_{k\in\mathbb{N}}.$$

*Proof.* For any element $(w, (x_k)_{k\in\mathbb{N}}) \in \mathbb{D}_\phi$ we have that $\phi'_\theta(w, (x_k)_{k\in\mathbb{N}}) \in c_0$. This follows from $\lim_{k\to\infty} c^{-k} x_k = 0$ and $(k(1 - H(k))w)/c^{k+1} = (k(1 - G(k)w)/c$, where due to the inequality $k(1 - G(k)) \leq \mathbb{E}[G] - \sum_{i=1}^{k} i g_i$ the latter expression tends to 0. $\phi'_\theta$ as given is linear on $\mathbb{D}_\phi$, and as the projections $\pi_k$ of $\phi$ are continuous it follows that $\phi'_\theta$ is continuous on $\mathbb{D}_\phi$. Following van der Vaart (2000), it remains to be seen that $\|[\phi(\theta + th_t) - \phi(\theta)]/t - \phi'_\theta(h)\|_{c_0} \to 0$ as $t \to 0$, for every $h_t \to h$ such that $th_t$ is contained in the domain of $\phi$ for all small $t > 0$ and such that $h \in \mathbb{D}_\phi$. For convenience, we denote the first component of the vectors $h \in \mathbb{R} \times c_\phi$ by $h(0)$, the second by $h(1)$ and so on, similarly for $h_t$. We calculate

$$\begin{aligned}
&\left\| \frac{\phi(\theta + th_t) - \phi(\theta)}{t} - \phi'_\theta(h) \right\|_{c_0} \\
&= \sup_{k \in \mathbb{N}} \left| \frac{h_t(k)}{(c + th_t(0))^k} + (1 - H(k)) \frac{(c + th_t(0))^{-k} - c^{-k}}{t} - \frac{h(k)}{c^k} - k \frac{1 - H(k)}{c^{k+1}} h(0) \right| .
\end{aligned}$$

Now, for each $k$ it quite obviously holds that

$$\lim_{t \to 0} \frac{(c + th_t(0))^{-k} - c^{-k}}{t} = \lim_{t \to 0} \frac{c^k - (c + th_t(0))^k}{tc^k (c + th_t(0))^k} = \lim_{t \to 0} \frac{-kh_t(0)}{c^{k+1}}.$$

Under the assumptions for $h, h_t$, the expression above tends to 0 for $t \to 0, h_t \to h$, concluding the proof. □

Using Lemma 3.2.8 together with Theorem 3.2.6 and Theorem 3.2.7 yields that

$$\sqrt{n} \left[ \phi(\hat{c}_n, \mathcal{H}_n - 1) - \phi(c, \mathcal{H} - 1) \right] = \sqrt{n} \left( \mathcal{G}_n - \mathcal{G} \right) \xrightarrow{\mathcal{D}} \phi'_\theta(\mathcal{W}) = \mathcal{V}.$$

At this point, we have collected all necessary pieces of the puzzle and can now assemble them, resulting in the main theorem of this chapter.

**Theorem 3.2.9** (Schweer and Wichelhaus (2015a), Theorem 1.1). *Let* $\mathbb{E}[A(0)^2] < \infty$ *and* $\sum_{n=1}^{\infty} \sqrt{1 - G(n)} < \infty$. *Then there exists a Gaussian sequence* $\mathcal{V} = (V_k)_{k \in \mathbb{N}}$ *in* $c_0$ *such that* $\mathbb{E}[V_k] = 0$ *and*

$$\begin{aligned}
\mathbb{E}[V_k V_m] = &\frac{\tau_{k,m}}{c^{k+m}} + km(1 - H(k))(1 - H(m)) \frac{1 - c}{c^{k+m+1}} \\
&- k(1 - H(k)) \frac{\tau_{1,m}}{c^{k+m+1}} - m(1 - H(m)) \frac{\tau_{1,k}}{c^{k+m+1}}
\end{aligned}$$

*with*

$$\begin{aligned}
\tau_{1,m} &= \frac{1}{\mathbb{E}[D(0)]} \sum_{i=0}^{\infty} \mathbb{E}\left[ D(i) \left( \mathbf{1}_{\{Z(i) \le m\}} - H(m) \right) \mathbf{1}_{\{A(0)=0\}} \right] \quad \text{and} \\
\tau_{k,m} &= \frac{1}{(\mathbb{E}[D(0)])^2} \sum_{i=-\infty}^{\infty} \mathbb{E}\Big[ D(0) D(i) \left( H(m) - \mathbf{1}_{\{Z(0) \le m\}} \right) \left( H(k) - \mathbf{1}_{\{Z(i) \le k\}} \right) \Big]
\end{aligned}$$

*for* $k, m \in \mathbb{N}$. *Moreover, in* $c_0$,

$$\sqrt{n} \left( \mathcal{G}_n - \mathcal{G} \right) \xrightarrow{\mathcal{D}} \mathcal{V}.$$

A short discussion of the assumptions of this theorem: First, we remark that the condition $\sum_{n=1}^{\infty} \sqrt{1 - G(n)} < \infty$ is a condition on the tail behavior of the service time distribution $G$. As proven in Lemma 3.2.3, it can be replaced by a moment condition on the distribution $G$, i.e., by assuming that $G$ has finite moments of order at least $2 + \epsilon$ for some $\epsilon > 0$. Thus, the moment conditions under which this result holds are very mild. Additional to the condition on the tail behavior of $G$, we only assume finiteness of the second moment of the arrival distribution. Due to this characteristic, it applies to a wide range of discrete-time queues with an infinite buffer size. For instance, the popular integer-valued auto-regressive models of the first order (cf. Chapters 4 and 5) can be interpreted as $GI/G/\infty$-models satisfying the condition $\sum_{n=1}^{\infty} \sqrt{1 - G(n)} < \infty$, as the service time distribution is geometrical in this case.

A different application is given by considering a discretized version of the continuous time $M/G/\infty$ estimation problem discussed in Brown (1970). Suppose that we are given arrival and departure points of a continuous time $M/G/\infty$ process. We denote the sequence of arrival points by $\{A_{\text{cont.}}(t)\}_{t \in \mathbb{Z}}$, governed by a Poisson process of intensity $\lambda$, and the departure points by $\{D_{\text{cont.}}(t)\}_{t \in \mathbb{Z}}$. Notice that $A_{\text{cont.}}(t), D_{\text{cont.}}(t) \in \mathbb{R}$ in this case. We now discretize the time domain with a certain step size $h > 0$ and define a discrete version of this process by setting $A_{\text{discr.}}(i) := \# \{A_{\text{cont.}}(t) \in [h(i-1), hi) \,|t \in \mathbb{Z}\}$ and similarly for $D_{\text{discr.}}(t)$. Elementary properties of Poisson processes imply that the $A_{\text{discr.}}(i)$'s are i.i.d. according to a Poi$(\lambda h)$ distribution. The assumption that the general (continuous) service time distribution $G$ satisfies the conditions of Theorem 3.2.9, or alternatively the moment condition imposed by Lemma 3.2.3, implies that the same assumption holds for the discretized version $G_{\text{discr.}}$. Hence, we can apply our main result in this situation and obtain, for example, confidence bounds on the estimation of $G$. Notice that we may choose the parameter $h > 0$ arbitrarily small, thus ensuring that we can approximate the continuous $G$ arbitrarily well. Furthermore, notice that the smaller $h > 0$, the larger $c = \mathbb{P}(A(0) = 0) = \exp(-\lambda h)$ becomes, thus improving the asymptotic variance of the estimator given in Theorem 3.2.9.

## 3.3 Moving Blocks Bootstrap

The results of the previous sections are difficult to apply in practice as the covariance kernel established in Theorem 3.2.9 is very involved and depends on the a priori unknown distributions $H$ or $G$ in a complicated manner. This problem was already pointed out in the original article introducing this estimation technique, where it is stated that (cf. (Brown, 1970, p. 653)) "[w]ere one to verify the mixing conditions there would still remain the difficulty of computing the covariance kernel of the limiting process". We address this problem by showing that bootstrapping techniques are available for the estimation of the kernel of the limiting distribution and lead to correct results.

In our situation, we assume to be given data sets of the form $(A(t))_{t \in \{1,\dots,n\}}$ and $(D(t))_{t \in \{1,\dots,n\}}$. The covariance kernel established in Theorem 3.2.9 involves the parameter $c$. The estimator $\hat{c}_n$ is based on i.i.d. observations of the arrival process and can be established independently from the estimators $\widehat{H}_n(\cdot)$. We suggest using $\hat{c}_n$ as plug-in estimate in the covariance kernel, we will not focus on this here.

Let us consider a bootstrap technique for the estimators $\widehat{H}_n(\cdot)$. Since we are not dealing with i.i.d. random variables but rather with a stationary and ergodic sequence, we will apply the moving blocks bootstrap resampling procedure as first introduced by Künsch (1989). In order to do so, we fix a block size $b$ for a set of data of size $n$ and denote $k = \frac{n}{b}$. We introduce the random variables $\{I_j\}_{j\in\{1,\dots,k\}}$ which are i.i.d. uniformly distributed on $\{1,\dots,n-b+1\}$. The bootstrap sample is then given as $\{(X_{I_d+1}, X_{I_d+2}, \dots, X_{I_d+b})\}_{d\in\{1,\dots,k\}}$. In order to keep the notation concise, we write $\mathbb{P}^*$, $\mathbb{E}^*$ and $\mathrm{Var}^*$ for the conditional probability, expectation and variance with respect to $\{I_j\}_{j\in\{1,\dots,k\}}$, respectively, and we abbreviate the bootstrap sample by $\{X_i^*\}_{i\in\{1,\dots,n\}}$. Expressions like $\frac{1}{n}\sum_{i=1}^n X_i^*$ are thus shorthand for $\frac{1}{k}\sum_{d=1}^k \frac{1}{b}\sum_{i=1}^b X_{I_d+i}$ and so forth. First, we prove a more general result than needed for our specific purposes, as this general result is of interest on its own.

**Theorem 3.3.1** (Schweer and Wichelhaus (2015a), Theorem 5.1)**.** *Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary and ergodic sequence of random variables satisfying* (2.14)*, $\mathbb{E}[X_0] = 0$ and $\mathbb{E}[X_0^4] < \infty$. Let $b = b(n)$ be a sequence of real numbers such that $b = o(n^\alpha)$ with $\alpha \in (0, 2/5)$. Then*

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left[\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \le x\right] - \mathbb{P}^*\left[\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i^* - \mathbb{E}^*[X_i^*]\right) \le x\right]\right| \to 0$$

*in probability.*

*Proof.* First note that under the assumptions on the sequence $b$ there exists a sequence $M$ such that $M \to \infty$, $b^2M^2/n \to 0$ and $b/M^4 \to 0$ as $n \to \infty$. Our goal is the application of Theorem 3 in Radulović (2012). The first condition is satisfied, as the assumptions of Theorem 2.5.3 are satisfied, thus $\mathrm{Var}\left(1/\sqrt{n}\sum_{i=1}^n X_i\right) \to \rho^2 > 0$. Markov's inequality implies

$$b\mathbb{E}\left[X_1^2 \mathbf{1}_{\{|X_1|>M\}}\right] \le b\mathbb{E}\left[X_1^4 \mathbf{1}_{\{|X_1|>M\}}\right] \le \frac{b}{M^4}\mathbb{E}\left[X_1^4\right],$$

which converges to 0 by assumption, proving the second condition. For the third condition we show that with the notation $Y_{i,b} := (1/\sqrt{b})\sum_{j=i}^{b+i-1} X_j$, the expression $(1/n-b+1)\sum_{i=1}^{n-b+1} Y_{i,b}^2$ converges in probability. Since the $X_i$ (and thus the $Y_{i,b}$) are stationary, $\mathbb{E}[1/(n-b+1)\sum_{i=1}^{n-b+1} Y_{i,b}^2] = 1/b\ \mathbb{E}[(\sum_{j=1}^b X_j)^2] \to \rho^2$, as $b \to \infty$ by (2.16). Now, since $\mathbb{E}[X_i^4] < \infty$ it follows that $\mathbb{E}[Y_{i,b}^4] < \infty$ and due to the stationarity of the $Y_{i,b}$ we have

$$\mathrm{Var}\left[\sum_{i=1}^{n-b+1} Y_{i,b}^2\right] \le (2b+1)(n-b+1)\,\mathrm{Var}\left[Y_{1,b}^2\right] + (n-b+1)\sum_{i=b}^{n-b+1} \mathrm{Cov}\left(Y_{1,b}^2, Y_{i,b}^2\right),$$

where we used the Cauchy-Schwarz inequality. Now, for any $i \in \mathbb{N}$ (including the case $i = 1$), the multi-linearity of the covariance, the application of equation (13) in

Bohrnstedt and Goldberger (1969) (notice that $\mathbb{E}[X_i] = 0$ for all $i$) and the stationarity of the $X_i$'s yield

$$\begin{aligned}
&\text{Cov}(Y_{1,b}^2, Y_{i,b}^2) \\
&= 2\left(\text{Cov}\left(X_1, X_i\right) + \sum_{j=1}^{b-1} \frac{b-j}{b} \text{Cov}\left(X_1, X_{i-j}\right) + \sum_{j=1}^{b-1} \frac{b-j}{b} \text{Cov}\left(X_1, X_{i+j}\right)\right)^2 . \quad (3.13)
\end{aligned}$$

Now, since $\mathbb{E}[X_i] = 0$, $\text{Cov}(X_1, X_i) = \mathbb{E}[X_1 X_i] \leq \|X_1\|_{L^2} \cdot \|\mathbb{E}\left[X_i|\mathcal{F}_1(X)\right]\|_{L^2}$ (cf. (2.15)). Thus, we first have

$$\text{Var}\left[Y_{1,b}^2\right] \leq 2\|X_1\|_{L^2}^2 \left(\|X_1\|_{L^2} + 2\sum_{j=1}^{b-1} \|\mathbb{E}\left[X_j|\mathcal{F}_1(X)\right]\|_{L^2}\right)^2 ,$$

as $b \to \infty$ this expression is bounded by the assumption of this theorem. For the covariances we use the same argument, and discuss each summand in (3.13) separately. Let $i = b + a$ for some $a \in \{0, 1, \ldots, n - 2b + 1\}$, then clearly $\|\mathbb{E}\left[X_{b+a}|\mathcal{F}_1(X)\right]\|_{L^2}$ tends to 0. By dominated convergence,

$$\lim_{b\to\infty} \sum_{j=1}^{b-1} \frac{b-j}{b} \|\mathbb{E}\left[X_{b+a-j}|\mathcal{F}_1(X)\right]\|_{L^2} = \lim_{b\to\infty} \sum_{j=a+1}^{b+a-1} \frac{j-a}{b} \|\mathbb{E}\left[X_j|\mathcal{F}_1(X)\right]\|_{L^2} = 0,$$

for each value of $a$, as the series $\sum_{j=1}^{\infty} \|\mathbb{E}\left[X_j|\mathcal{F}_1(X)\right]\|_{L^2}$ converges. Analogous argumentation applies to the third summand in (3.13). Combining all of these results we find that

$$\text{Var}\left[\frac{Y_{1,b}^2}{n-b+1}\right] \leq \frac{2b+1}{n-b+1} \text{Var}\left[Y_{1,b}^2\right] + \frac{n-2b+1}{n-b+1} \sup_{i\in\{b,\ldots,n-b+1\}} \text{Cov}\left(Y_{1,b}^2, Y_{i,b}^2\right) .$$

As $n \to \infty$ (and thus $b \to \infty$), the latter expression tends to 0. For the former expression, notice that $b^2/n \to 0$, thus $(2b + 1)(n - b + 1) \to 0$. In conclusion, we showed that $\mathbb{E}[(1/n - b + 1) \sum_{i=1}^{n-b+1} Y_{i,b}^2] \to \rho^2$ and $\text{Var}[(1/n - b + 1)Y_{1,b}^2] \to 0$, implying convergence in probability. Thus, as argued in the proof of Theorem 2 in Radulović (2012), it follows that $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^* - \mathbb{E}^*[X_i^*]\right)$ converges to an asymptotically normal distribution $\mathcal{N}(0, \rho^2)$ in probability. The assertion follows due to the continuity of the normal distribution. □

A general problem for the application of moving block bootstrap results is that there is no canonical choice for the block length $b$. The theoretical optimal block length in a different albeit similar situation to the one discussed here was calculated to the order $o(n^{1/3})$, we refer to (Lahiri, 2003, Corollary 7.1.). The conditions of Theorem 3.3.1 allow for such a choice of $b$ and we suggest to use this block length in practical applications. However, we point out to the authors' best knowledge there exists no result on the theoretical optimal block length for the specific situation discussed in this paper.

Returning to the problem at hand, we first construct the bootstrap samples of the processes $\{D(t)\mathbf{1}_{\{Z(t)\leq x\}}\}_{t\in\mathbb{Z}}$ and $\{D(t)\}_{t\in\mathbb{Z}}$ using the scheme described above, we denote these samples by $\{(D(t)\mathbf{1}_{\{Z(t)\leq x\}})^*\}_{t\in\{1,\dots,n\}}$ and $\{(D(t))^*\}_{t\in\{1,\dots,n\}}$. We now construct the bootstrap estimator as

$$\widehat{H}_n^*(x) := \frac{\sum_{i=1}^n \left(D(i)\mathbf{1}_{\{Z(i)\leq x\}}\right)^*}{\sum_{i=1}^n (D(i))^*} \text{ and } H^*(x) := \frac{\mathbb{E}^*\left[\left(D(0)\mathbf{1}_{\{Z(0)\leq x\}}\right)^*\right]}{\mathbb{E}^*[(D(0))^*]}.$$

We denote $(\widehat{H}_n^*(x_1),\dots,\widehat{H}_n^*(x_l))^T := \mathbf{H}_n^*$ and $(H^*(x_1),\dots,H^*(x_l))^T := \mathbf{H}^*$ in complete analogy with the notation $\mathbf{H}_n$ and $\mathbf{H}$ in the proof of Theorem 3.2.4.

**Corollary 3.3.2** (Schweer and Wichelhaus (2015a), Corollary 5.2)**.** *Let $x_1,\dots,x_l \in \mathbb{N}$, $l \in \mathbb{N}$. Let $b = b(n), M = M(n)$ be sequences of real numbers such that $b, M \to \infty$, $\frac{b^2M^2}{n} \to 0$ and $\frac{b}{M^4} \to 0$ as $n \to \infty$. Let the conditions of Theorem 3.2.9 be satisfied and let $\mathbb{E}\left[A(0)^4\right] < \infty$. Then*

$$\sup_{x\in\mathbb{R}^l} \left|\mathbb{P}\left(\sqrt{n}\left(\mathbf{H}_n - \mathbf{H}\right) \leq x\right) - \mathbb{P}^*\left(\sqrt{n}\left(\mathbf{H}_n^* - \mathbf{H}^*\right) \leq x\right)\right| \to 0$$

*in probability.*

*Proof.* Let $(t_1,\dots,t_l) \in \mathbb{R}^l$. In the proof of Theorem 3.2.4 we showed that the stationary and ergodic sequence $(X_i)_{i\in\mathbb{Z}}$ with $X_i := \sum_{j=1}^l t_j(\eta_i(x_j))$ satisfies the assumptions of Theorem 2.5.3. All other conditions of Theorem 3.3.1 are satisfied, it remains to be seen that $\mathbb{E}[X_1^4] < \infty$. It is clear that we can find a constant $C$ such that $\mathbb{E}[X_1^4] \leq C\mathbb{E}[D(1)^4]$. Recall the notation of the $r$-th order cumulant introduced in Section 2.2, then with Lemma 2.2.1 (iii),

$$\kappa_r(D(1)) = \sum_{j=1}^{\infty} \kappa_r\left(\sum_{l=1}^{A(1-j)} \mathbf{1}_{\{S_{1-j,l}=j\}}\right).$$

Let $r = 4$ then $\kappa_4 = \mu_4 - 3\mu_2^2$ in terms of central moments. For any $j \in \mathbb{N}$, the variance of $\sum_{l=1}^{A(1-j)} \mathbf{1}_{\{S_{1-j,l}=j\}}$ is finite by (3.6), for the fourth moment we calculate

$$\mathbb{E}\left[\left(\sum_{l=1}^{A(1-j)} \mathbf{1}_{\{S_{1-j,l}=j\}}\right)^4\right] = \sum_{m=1}^{\infty} \mathbb{P}\left(A(0)=m\right)\mathbb{E}\left[\left(\sum_{l=1}^{m} \mathbf{1}_{\{S_{1-j,l}=j\}}\right)^4\right] \leq g_j\mathbb{E}\left[A(1)^4\right].$$

Thus, $\kappa_4(D(1)) < \infty$, the finiteness of the other cumulants follows analogously. As $\mathbb{E}[D(1)^4]$ can be written as a polynomial in the first four cumulants of $D(1)$, this shows that $\mathbb{E}[X_1^4] < \infty$. Application of the Cramér-Wold device as in the proof of Theorem 3.2.4 concludes the proof. □

# 4 INAR(1) Processes - Stochastic Properties

Let us begin this chapter by deriving first elementary characteristics of INAR(1) processes according to Definition 1.1.1. Quite clearly, it is a homogeneous Markov chain with 1-step transition probabilities

$$\mathbb{P}(Y_t = k \mid Y_{t-1} = l) \;=\; \sum_{j=0}^{\min\{k,l\}} \binom{l}{j} \alpha^j (1-\alpha)^{l-j} \mathbb{P}(\epsilon_t = k - j). \tag{4.1}$$

Applying the recursion involved in Definition 1.1.1 iteratively leads to the representation (cp. Alzaid and Al-Osh (1988), eq. (2.2))

$$Y_t \stackrel{\mathcal{D}}{=} \sum_{i=0}^{\infty} \alpha^j \circ \epsilon_{t-j}, \tag{4.2}$$

where we used the notion of the $i$-time iteration of the $\circ$ operator, i.e., the application of $i$ independent thinning operations on a random variable $X$. This is defined as

$$\alpha^i \circ X := \underbrace{\alpha \circ \alpha \circ \cdots \circ \alpha \circ X}_{i \text{ times}} = \sum_{j=1}^{X} \prod_{k=1}^{i} \xi_{j,k}, \tag{4.3}$$

where $\xi_{a,b}$ are mutually independent identically distributed Bernoulli random variables, independent of $X$, with probability of success $\alpha$ for $(a,b) \in \mathbb{N} \times \{1,\ldots,i\}$, where we set $\alpha^0 \circ X := X$. Note that in the cited reference, the random variable $\alpha^i \circ X$ does not represent iterated thinning as we defined it above, but one single thinning with parameter $\alpha^i$. As the resultant random variables are equal in distribution, however, the result holds. A further useful relation for the joint distribution of the process is

$$(Y_k, Y_l) \stackrel{\mathcal{D}}{=} \left( Y_k, \alpha^{l-k} \circ Y_k + \sum_{s=0}^{l-k-1} \alpha^s \circ \epsilon_{l-s} \right) \tag{4.4}$$

for any $l, k \in \mathbb{Z}$, $l \geq k$. This relation can be extended to higher order joint distributions in an obvious manner.

From (4.4), it immediately follows that the autocovariance function $\gamma(\cdot)$ of an INAR(1) process $(Y_t)_{t\in\mathbb{Z}}$ decays exponentially, i.e., $\gamma(k) = \alpha^k \gamma(0)$, cf. eq. (3.3) in Al-Osh and Alzaid (1987). This implies for the ACF that $\rho(k) = \alpha^k$. Similarly, it can be seen that the PACF satisfies $\rho_{part}(1) = \alpha$ and $\rho_{part}(k) \equiv 0$ everywhere else, which corresponds to the empirical behavior seen in the data of Figure 1.2. Hence, the INAR(1) model provides the practitioner with a simple variation of the popular continuous AR(1) process, which allows for the consideration of count data instead of real-valued observations.

The relation between the INAR(1) model and the queuing model of Chapter 3 can also nicely be described. In the notation of that chapter, set $g_k = (1-\alpha)\alpha^{k-1}$ and thus $G(k) = 1 - \alpha^k$, that is, assume a geometric service time distribution for the queue with parameter $\alpha$. Comparing (3.1) with (4.2), it is quite obvious that we have equality in distribution and hence that the INAR(1) process is a special case of a discrete-time queueing process. Indeed, for each time step $t$, the summand $\alpha \circ Y_{t-1}$ can be seen as the portion of customers who were present at time $t-1$ and who remain at the queue until time $t$. The $\epsilon_t$ random variables thus correspond to the arrivals $A(t)$.

In this chapter, we focus on stochastic properties of the INAR(1) model. In the first part, we derive closed expressions for joint moments and cumulants for these processes, which will facilitate the calculation of asymptotic expressions of a number of statistics in subsequent chapters. Furthermore, the question of conditions for the time-reversibility of INAR(1) processes is shown to have a quite satisfying answer.

In the second part of this chapter, the special case of Compound Poisson INAR(1) models is introduced and discussed in detail. We show that in this semi-parametric setting, the derivation of important characteristics such as the stationary distribution or mixing properties of the processes is possible. Additionally, we endeavor to highlight the intricate way in which the notion of Compound Poisson distribution, infinite divisibility and the class of possible marginal distributions of INAR(1) models are linked together. Large parts of the following chapter consist of rearranged material from these articles: Schweer and Weiß (2014), Schweer and Weiß (2015) and Schweer (2015b).

## 4.1 Stochastic Properties of INAR(1) Processes

In this first part, we make no assumption on the distribution of the arrivals $(\epsilon_t)_{t\in\mathbb{Z}}$ of Definition 1.1.1 other than the existence of appropriate moments. Let us first discuss the question of stationarity of INAR(1) processes. Due to their Markovian structure, we first consider the irreducibility of these processes on the state space $\mathbb{N}_0$.

**Lemma 4.1.1.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a INAR(1) process. Then $(Y_t)_{t\in\mathbb{Z}}$ is irreducible if and only if $\mathbb{P}(\epsilon_0 = 0) \in (0,1)$.*

*Proof.* First, let $\mathbb{P}(\epsilon_0 = 0) \in (0,1)$, thus there exists a $k \in \mathbb{N}$ with $\mathbb{P}(\epsilon_0 = k) \in (0,1)$. Let $k_0 \in \mathbb{N}$ be arbitrary, define $l := \lceil k_0/k \rceil$, i.e., the smallest integer larger than $k_0/k$. Let us assume the following behavior of the process: it is zero at some time point $t$, and there are $k$ arrivals at each time instant $t+1, t+2, \ldots, t+l$ and no departures, and there are 0 arrivals at time $t+l+1$ and $l \cdot k - k_0$ departures. The transition probability is, with (4.1), bounded from below by the expression

$$\alpha^{\frac{k}{2}l(l+1)+k_0}(1-\alpha)^{l\cdot k-k_0}\binom{l\cdot k-k_0}{k_0}\mathbb{P}(\epsilon_0 = 0)\mathbb{P}(\epsilon_0 = k)^l > 0.$$

Therefore, for any state $l_0 \in \mathbb{N}$, we have $p_Y(0|l_0) = (1-\alpha)^{l_0}\mathbb{P}(\epsilon_0 = 0) > 0$. Hence, the process can move with a positive probability from any state $l_0$ to any other state $k_0$, proving irreducibility of the Markov chain. For the converse assertion, assume that $\mathbb{P}(\epsilon_0 = 0) \notin (0,1)$, so either $\mathbb{P}(\epsilon_0 = 0) = 0$ or $\mathbb{P}(\epsilon_0 = 0) = 1$. In the former case, the state 0 can never be reached, in the latter, the state 0 is absorbing. Thus, $(Y_t)_{t\in\mathbb{Z}}$ is not irreducible which concludes the proof. □

As we pointed out in the introduction, INAR(1) processes are closely linked to a special type of discrete-time branching process with immigration. This connection is exploited in the next result, which holds for a slightly larger class than that of Definition 1.1.1, as only finiteness of the first moment is assumed.

**Lemma 4.1.2.** *Let* $(Y_t)_{t\in\mathbb{Z}}$ *be a process satisfying the recursion* (1.1) *with* $\alpha < 1$*, let* $\mathbb{E}[\epsilon_t] < \infty$ *and let* $\mathbb{P}(\epsilon_0 = 0) \in (0,1)$*. Then* $(Y_t)_{t\in\mathbb{Z}}$ *is a stationary Markov chain.*

*Proof.* By Lemma 4.1.1 (which does not necessitate a finite second moment), $(Y_t)_{t\in\mathbb{Z}}$ is irreducible, it is aperiodic because $p_Y(k|k) \geq \alpha^k \cdot \mathbb{P}(\epsilon_0 = 0) > 0$, see formula (4.1). Since it is also a branching process with immigration, where the offspring distribution has mean $\alpha < 1$, we conclude from Heathcote (1966) that the condition $\mathbb{E}[\epsilon_t] < \infty$ is sufficient for a nontrivial stationary distribution to exist. Since the recursion is assumed to be satisfied for all $t \in \mathbb{Z}$, this concludes the proof. □

This result allows for a very simple condition of stationarity of INAR(1) processes of Definition 1.1.1: They only have to satisfy $\mathbb{P}(\epsilon_0 = 0) \in (0,1)$. Clearly, this is not a sufficient condition for stationarity, yet since it is much easier verified (and holds for most cases), we opt to employ this condition throughout this thesis.

### 4.1.1 Moments and Cumulants of INAR(1) Processes

In this section, we collect a number of results concerned with joint moments and joint cumulants of INAR(1) processes. As a first consequence of (4.2), we obtain, if $\mu_\epsilon, \sigma_\epsilon < \infty$,

$$\mu_Y = \frac{\mu_\epsilon}{1-\alpha} \quad \text{and} \quad \sigma_Y^2 = \frac{\sigma_\epsilon^2 + \alpha\mu_\epsilon}{1-\alpha^2}, \quad \text{i.e.,} \quad I_Y = \frac{I_\epsilon + \alpha}{1+\alpha}, \tag{4.5}$$

recalling the definition of the index of dispersion $I_Y$ in (2.5). Like in Weiß (2012), we introduce the notation

$$\mu(s_1,\ldots,s_{r-1}) := \mathbb{E}[Y_t Y_{t+s_1} Y_{t+s_{r-1}}] \quad \text{with } 0 \leq s_1 \leq \ldots \leq s_{r-1}, r \in \mathbb{N}.$$

In particular, the case $r = 1$ corresponds to the marginal mean $\mu_Y = \mathbb{E}[Y_t]$.

**Theorem 4.1.3** (Schweer and Weiß (2014), Theorem 3.3.1)**.** *Let* $(Y_t)_{t\in\mathbb{Z}}$ *be an INAR(1) process with* $\mathbb{P}(\epsilon_0 = 0) \in (0,1)$*, where the innovations* $(\epsilon_t)_{t\in\mathbb{Z}}$ *have existing moments* $\mu_{\epsilon,r}$ *for* $r \leq 4$*. Then*

$$\begin{aligned}
\mu(k) &= \sigma_Y^2\alpha^k + \mu_Y^2,\\
\mu(k,l) &= (\bar{\mu}_{Y,3} - \sigma_Y^2)\alpha^{l+k} + (1+\mu_Y)\sigma_Y^2\alpha^l + \mu_Y\sigma_Y^2\left(\alpha^{l-k} + \alpha^k\right) + \mu_Y^3,\\
\mu(k,l,m) &= \alpha^{m+l+k}\left(\bar{\mu}_{Y,4} - 3\bar{\mu}_{Y,3} + \sigma_Y^2(2 - 3\sigma_Y^2)\right) + \mu_Y^4\\
&\quad + (\bar{\mu}_{Y,3} - \sigma_Y^2)\left((2+\mu_Y)\alpha^{m+l} + (1+\mu_Y)\alpha^{m+k} + \mu_Y\left(\alpha^{m+l-2k} + \alpha^{l+k}\right)\right)\\
&\quad + (1+\mu_Y)^2\sigma_Y^2\alpha^m + \mu_Y\sigma_Y^2(1+\mu_Y)\left(\alpha^{m-k} + \alpha^l\right)\\
&\quad + \mu_Y^2\sigma_Y^2\left(\alpha^{m-l} + \alpha^{l-k} + \alpha^k\right) + \sigma_Y^4\left(\alpha^{m-l+k} + 2\alpha^{m+l-k}\right),
\end{aligned}$$

*holds for any* $0 \leq k \leq l \leq m$.

*Proof.* $(Y_t)_{t\in\mathbb{Z}}$ is stationary by Lemma 4.1.2, and the mixed moment $\mu(s_1,\ldots,s_{r-1})$ for the case $r=2$ is easily derived using that the autocorrelation function equals $\rho_Y(k)=\alpha^k$:

$$\mu(k)=\mathrm{Cov}[Y_t,Y_{t+k}]+\mu_Y^2=\sigma_Y^2\alpha^k+\mu_Y^2.$$

To obtain expressions for $r>2$, we first have to consider the conditional moments, recalling the notations of the raw moments $\mu_{\epsilon,k}$ and central moments $\bar{\mu}_{\epsilon,k}$ of Section 2.3,

$$\mathbb{E}[Y_t^k|Y_{t-1},\ldots]=\mathbb{E}\left[\sum_{j=0}^{k}\binom{k}{j}\epsilon_t^{k-j}(\alpha\circ Y_{t-1})^j|Y_{t-1}\right]=\sum_{j=0}^{k}\binom{k}{j}\mu_{\epsilon,k-j}\mathbb{E}\left[(\alpha\circ Y_{t-1})^j|Y_{t-1}\right]$$

where the last factor is just $j$th moment of the binomial distribution $\mathrm{Bin}(Y_{t-1},\alpha)$. Using the formulae given on p. 110 in Johnson et al. (2005), we obtain

$$\begin{aligned}
\mathbb{E}[Y_t|Y_{t-1},\ldots]&=\alpha Y_{t-1}+\mu_\epsilon,\\
\mathbb{E}[Y_t^2|Y_{t-1},\ldots]&=\mu_{\epsilon,2}+2\mu_\epsilon\alpha Y_{t-1}+\alpha Y_{t-1}+\alpha^2Y_{t-1}(Y_{t-1}-1)\\
&=\alpha^2Y_{t-1}^2+\alpha(1-\alpha)(1+2\mu_Y)Y_{t-1}+\mu_{\epsilon,2}\quad\text{as well as}\\
\mathbb{E}[Y_t^3|Y_{t-1},\ldots]&=\mu_{\epsilon,3}+3\mu_{\epsilon,2}\alpha Y_{t-1}+3\mu_\epsilon\left(\alpha Y_{t-1}+\alpha^2Y_{t-1}(Y_{t-1}-1)\right)\\
&\quad+\alpha Y_{t-1}+3\alpha^2Y_{t-1}(Y_{t-1}-1)+\alpha^3Y_{t-1}(Y_{t-1}-1)(Y_{t-1}-2)\\
&=\mu_{\epsilon,3}+(1+3\mu_\epsilon+3\mu_{\epsilon,2})\alpha Y_{t-1}-3(1+\mu_\epsilon)\alpha^2Y_{t-1}+2\alpha^3Y_{t-1}\\
&\quad+3(1+\mu_\epsilon)\alpha^2Y_{t-1}^2-3\alpha^3Y_{t-1}^2+\alpha^3Y_{t-1}^3\\
&=\alpha^3Y_{t-1}^3+3\alpha^2(1-\alpha)(1+\mu_Y)Y_{t-1}^2+\mu_{\epsilon,3}\\
&\quad+\alpha(1-\alpha)\left(1-2\alpha+3\mu_\epsilon+3\frac{\mu_{\epsilon,2}}{1-\alpha}\right)Y_{t-1},\quad\text{and}\\
\mathbb{E}[Y_t^4|Y_{t-1},\ldots]&=\mu_{\epsilon,4}+4\mu_{\epsilon,3}\mathbb{E}\left[\alpha\circ Y_{t-1}|Y_{t-1}\right]+6\mu_{\epsilon,2}\mathbb{E}\left[(\alpha\circ Y_{t-1})^2|Y_{t-1}\right]\\
&\quad+4\mu_\epsilon\mathbb{E}\left[(\alpha\circ Y_{t-1})^3|Y_{t-1}\right]+\mathbb{E}\left[(\alpha\circ Y_{t-1})^4|Y_{t-1}\right]\\
&=\mu_{\epsilon,4}+4\mu_{\epsilon,3}\alpha Y_{t-1}+6\mu_{\epsilon,2}\left(\alpha Y_{t-1}+\alpha^2Y_{t-1}(Y_{t-1}-1)\right)\\
&\quad+4\mu_\epsilon\left(\alpha Y_{t-1}+3\alpha^2Y_{t-1}(Y_{t-1}-1)+\alpha^3Y_{t-1}(Y_{t-1}-1)(Y_{t-1}-2)\right)\\
&\quad+\alpha Y_{t-1}+7\alpha^2Y_{t-1}(Y_{t-1}-1)+6\alpha^3Y_{t-1}(Y_{t-1}-1)(Y_{t-1}-2)\\
&\quad+\alpha^4Y_{t-1}(Y_{t-1}-1)(Y_{t-1}-2)(Y_{t-1}-3)\\
&=(1+4\mu_\epsilon+6\mu_{\epsilon,2}+4\mu_{\epsilon,3})\,\alpha Y_{t-1}+(6\mu_{\epsilon,2}+12\mu_\epsilon+7)\,\alpha^2(Y_{t-1}^2-1)\\
&\quad+\alpha^3(Y_{t-1}^2-Y_{t-1})[(4\mu_\epsilon+6)\,(Y_{t-1}-2)+\alpha(Y_{t-1}^2-5Y_{t-1}+6)]+\mu_{\epsilon,4}\\
&=\left(1-6\alpha(1-\alpha)+4\mu_\epsilon(1-2\alpha)+6\mu_{\epsilon,2}+4\frac{\mu_{\epsilon,3}}{1-\alpha}\right)(1-\alpha)\alpha Y_{t-1}\\
&\quad+\left(7-11\alpha+12\mu_\epsilon+6\frac{\mu_{\epsilon,2}}{1-\alpha}\right)(1-\alpha)\alpha^2Y_{t-1}^2\\
&\quad+2\,(2\mu_Y+3)\,(1-\alpha)\alpha^3Y_{t-1}^3+\alpha^4Y_{t-1}^4+\mu_{\epsilon,4}.
\end{aligned}$$

As a consequence, we obtain for the raw moments that

$$
\begin{aligned}
(1-\alpha^2)\mu_{Y,2} &= \mu_{\epsilon,2} + \alpha\mu_\epsilon(1+2\mu_Y),\\
(1-\alpha^3)\mu_{Y,3} &= 3\alpha^2(1-\alpha)(1+\mu_Y)\mu_{Y,2} + \mu_{\epsilon,3} + \alpha\mu_\epsilon\left(1-2\alpha+3\mu_\epsilon+3\frac{\mu_{\epsilon,2}}{1-\alpha}\right), \quad (4.6)\\
(1-\alpha^4)\mu_{Y,4} &= \left(1-6\alpha(1-\alpha)+4\mu_\epsilon(1-2\alpha)+6\mu_{\epsilon,2}+4\frac{\mu_{\epsilon,3}}{1-\alpha}\right)\alpha\mu_\epsilon + \mu_{\epsilon,4}\\
&\quad + \left(7-11\alpha+12\mu_\epsilon+6\frac{\mu_{\epsilon,2}}{1-\alpha}\right)(1-\alpha)\alpha^2\mu_{Y,2} + 2\left(2\mu_Y+3\right)(1-\alpha)\alpha^3\mu_{Y,3}.
\end{aligned}
$$

Finally, we shall use the well-known relations (see, e.g., p. 450 in Douglas (1980))

$$
\begin{aligned}
\mu_{Y,2} &= \bar{\mu}_{Y,2} + \mu_Y^2 = \sigma_Y^2 + \mu_Y^2,\\
\mu_{Y,3} &= \bar{\mu}_{Y,3} + 3\mu_Y\sigma_Y^2 + \mu_Y^3, \qquad (4.7)\\
\mu_{Y,4} &= \bar{\mu}_{Y,4} + 4\mu_Y\bar{\mu}_{Y,3} + 6\mu_Y^2\sigma_Y^2 + \mu_Y^4.
\end{aligned}
$$

**Third-Order Moments:**
To derive an explicit expression for $\mu(k,l)$ with $0 \leq k \leq l$, we distinguish between the following cases:

(i) $l > k$. Here, we have

$$
\begin{aligned}
\mu(k,l) &= \mathbb{E}\left[Y_tY_{t+k}\mathbb{E}[Y_{t+l}|Y_{t+l-1},\ldots]\right] = \alpha\mu(k,l-1) + \mu_\epsilon\mu(k) = \ldots\\
&= \alpha^{l-k}\mu(k,k) + (1-\alpha)\mu_Y\mu(k)\sum_{j=0}^{l-k-1}\alpha^j = \alpha^{l-k}\left(\mu(k,k)-\mu_Y\mu(k)\right) + \mu_Y\mu(k).
\end{aligned}
$$

(ii) $l = k > 0$. Using the relations (4.6) and (4.7), we have

$$
\begin{aligned}
\mu(k,k) &= \mathbb{E}\left[Y_t\mathbb{E}[Y_{t+k}^2|Y_{t+k-1},\ldots]\right]\\
&= \alpha^2\mu(k-1,k-1) + \alpha(1-\alpha)(1+2\mu_Y)\mu(k-1) + \mu_{\epsilon,2}\mu_Y\\
&= \alpha^2\mu(k-1,k-1) + (1-\alpha)(1+2\mu_Y)\sigma_Y^2\alpha^k + (1-\alpha^2)\mu_Y\mu_{Y,2}\\
&= \ldots = \alpha^{2k}\mu(0,0) + (1-\alpha)(1+2\mu_Y)\sigma_Y^2\sum_{j=0}^{k-1}\alpha^{k+j} + (1-\alpha^2)\mu_Y\mu_{Y,2}\sum_{j=0}^{k-1}\alpha^{2j}\\
&= \alpha^{2k}\left(\mu_{Y,3} - (1+2\mu_Y)\sigma_Y^2 - \mu_Y\mu_{Y,2}\right) + (1+2\mu_Y)\sigma_Y^2\alpha^k + \mu_Y\mu_{Y,2}\\
&= \alpha^{2k}(\bar{\mu}_{Y,3} - \sigma_Y^2) + (1+2\mu_Y)\left(\mu(k) - \mu_Y^2\right) + \mu_Y\mu_{Y,2},
\end{aligned}
$$

which also holds for $k = 0$. So it follows that

$$
\mu(k,l) = \alpha^{l-k}\left(\alpha^{2k}(\bar{\mu}_{Y,3} - \sigma_Y^2) + (1+\mu_Y)\left(\mu(k) - \mu_Y^2\right) + \mu_Y\sigma_Y^2\right) + \mu_Y\mu(k)
$$

holds for any $0 \leq k \leq l$.

**Fourth-Order Moments:**
To derive an explicit expression for $\mu(k,l,m)$ with $0 \leq k \leq l \leq m$, we distinguish between the following cases:

(i) $m > l$. Then, similar to above, we have

$$\begin{aligned}\mu(k,l,m) &= \mathbb{E}\left[Y_t Y_{t+k} Y_{t+l} \mathbb{E}[Y_{t+m}|Y_{t+m-1},\ldots]\right] \alpha\mu(k,l,m-1) + \mu_\epsilon \mu(k,l) = \ldots \\ &= \alpha^{m-l}\left(\mu(k,l,l) - \mu_Y \mu(k,l)\right) + \mu_Y \mu(k,l).\end{aligned}$$

(ii) $m = l > k$. Then we have (using the above relation (4.6))

$$\begin{aligned}\mu(k,l,l) &= \mathbb{E}\left[Y_t Y_{t+k} \mathbb{E}[Y_{t+l}^2|Y_{t+l-1},\ldots]\right] \\ &= \alpha^2\mu(k,l-1,l-1) + \alpha(1-\alpha)(1+2\mu_Y)\mu(k,l-1) + \mu_{\epsilon,2}\mu(k) \\ &= \alpha^2\mu(k,l-1,l-1) + \left(\alpha(1-\alpha)\mu_Y(1+2\mu_Y) + \mu_{\epsilon,2}\right)\mu(k) \\ &\quad + (1-\alpha)(1+2\mu_Y)\left(\mu(k,k) - \mu_Y\mu(k)\right)\alpha^{l-k} \\ &= \ldots = \alpha^{2(l-k)}\mu(k,k,k) + (1-\alpha^2)\mu_{Y,2}\mu(k)\sum_{j=0}^{l-k-1}\alpha^{2j} \\ &\quad + (1-\alpha)(1+2\mu_Y)\left(\mu(k,k) - \mu_Y\mu(k)\right)\sum_{j=0}^{l-k-1}\alpha^{l-k+j} \\ &= \alpha^{2(l-k)}\left(\mu(k,k,k) - \mu_{Y,2}\mu(k) - (1+2\mu_Y)\left(\mu(k,k) - \mu_Y\mu(k)\right)\right) \\ &\quad + (1+2\mu_Y)\left(\mu(k,k) - \mu_Y\mu(k)\right)\alpha^{l-k} + \mu_{Y,2}\mu(k).\end{aligned}$$

(iii) $m = l = k > 0$. Then we have (using the relations (4.6) and (4.7))

$$\begin{aligned}\mu(k,k,k) &= \mathbb{E}\left[Y_t \mathbb{E}[Y_{t+k}^3|Y_{t+k-1},\ldots]\right] \\ &= \alpha^3\mu(k-1,k-1,k-1) + 3\alpha^2(1-\alpha)(1+\mu_Y)\mu(k-1,k-1) \\ &\quad + \alpha(1-\alpha)\left(1 - 2\alpha + 3\mu_\epsilon + 3\frac{\mu_{\epsilon,2}}{1-\alpha}\right)\mu(k-1) + \mu_{\epsilon,3}\mu_Y \\ &= \alpha^3\mu(k-1,k-1,k-1) + 3(1-\alpha)(1+\mu_Y)(\bar{\mu}_{Y,3} - \sigma_Y^2)\alpha^{2k} \\ &\quad + (1-\alpha^2)\sigma_Y^2(1+3\mu_Y+3\mu_{Y,2})\alpha^k + (1-\alpha^3)\mu_Y\mu_{Y,3} \\ &= \ldots = \alpha^{3k}\mu(0,0,0) + 3(1-\alpha)(1+\mu_Y)(\bar{\mu}_{Y,3} - \sigma_Y^2)\sum_{j=0}^{k-1}\alpha^{2k+j} \\ &\quad + (1-\alpha^2)\sigma_Y^2(1+3\mu_Y+3\mu_{Y,2})\sum_{j=0}^{k-1}\alpha^{k+2j} + (1-\alpha^3)\mu_Y\mu_{Y,3}\sum_{j=0}^{k-1}\alpha^{3j} \\ &= \alpha^{3k}\left(\bar{\mu}_{Y,4} - 3\bar{\mu}_{Y,3} + \sigma_Y^2(2 - 3\sigma_Y^2)\right) + 3(1+\mu_Y)(\bar{\mu}_{Y,3} - \sigma_Y^2)\alpha^{2k} \\ &\quad + \sigma_Y^2(1+3\mu_Y+3\mu_{Y,2})\alpha^k + \mu_Y\mu_{Y,3},\end{aligned}$$

which also holds for $k=0$. So it follows that

$$
\begin{aligned}
\mu(k,l,m) &= \alpha^{m+l-2k}\left(\mu(k,k,k)-\mu_{Y,2}\mu(k)-(1+2\mu_Y)\left(\mu(k,k)-\mu_Y\mu(k)\right)\right)\\
&\quad+\alpha^{m-l}(1+\mu_Y)\left(\mu(k,l)-\mu_Y\mu(k)\right)+\alpha^{m-l}\sigma_Y^2\mu(k)+\mu_Y\mu(k,l)\\
&=\alpha^{m+l-2k}\Big(\mu(k,k,k)-(\sigma_Y^2+\mu_Y^2)\sigma_Y^2\alpha^k-(\sigma_Y^2+\mu_Y^2)\mu_Y^2\\
&\quad-(1+2\mu_Y)\Big(\alpha^{2k}(\bar{\mu}_{Y,3}-\sigma_Y^2)+(1+\mu_Y)\sigma_Y^2\alpha^k+\mu_Y\sigma_Y^2\Big)\Big)\\
&\quad+\alpha^{m-k}(1+\mu_Y)\Big(\alpha^{2k}(\bar{\mu}_{Y,3}-\sigma_Y^2)+(1+\mu_Y)\sigma_Y^2\alpha^k+\mu_Y\sigma_Y^2\Big)\\
&\quad+\alpha^{m-l}\sigma_Y^2\Big(\sigma_Y^2\alpha^k+\mu_Y^2\Big)\\
&\quad+\mu_Y(\bar{\mu}_{Y,3}-\sigma_Y^2)\alpha^{l+k}+\mu_Y(1+\mu_Y)\sigma_Y^2\alpha^l+\mu_Y^2\sigma_Y^2(\alpha^{l-k}+\alpha^k)+\mu_Y^4\\
&=\alpha^{m+l-2k}\Big(\mu(k,k,k)-(1+2\mu_Y)(\bar{\mu}_{Y,3}-\sigma_Y^2)\alpha^{2k}\\
&\quad-(1+3\mu_Y+3\mu_Y^2+\sigma_Y^2)\sigma_Y^2\alpha^k-\mu_Y^4-\mu_Y(1+3\mu_Y)\sigma_Y^2\Big)\\
&\quad+(1+\mu_Y)(\bar{\mu}_{Y,3}-\sigma_Y^2)\alpha^{m+k}+(1+\mu_Y)^2\sigma_Y^2\alpha^m+\mu_Y(1+\mu_Y)\sigma_Y^2\alpha^{m-k}\\
&\quad+\sigma_Y^4\alpha^{m-l+k}+\mu_Y^2\sigma_Y^2\alpha^{m-l}\\
&\quad+\mu_Y(\bar{\mu}_{Y,3}-\sigma_Y^2)\alpha^{l+k}+\mu_Y(1+\mu_Y)\sigma_Y^2\alpha^l+\mu_Y^2\sigma_Y^2(\alpha^{l-k}+\alpha^k)+\mu_Y^4\\
&=\alpha^{m+l-2k}\Big(\alpha^{3k}\left(\bar{\mu}_{Y,4}-3\bar{\mu}_{Y,3}+\sigma_Y^2(2-3\sigma_Y^2)\right)\\
&\quad+(2+\mu_Y)(\bar{\mu}_{Y,3}-\sigma_Y^2)\alpha^{2k}+2\sigma_Y^4\alpha^k+\mu_Y(\bar{\mu}_{Y,3}-\sigma_Y^2)\Big)\\
&\quad+(1+\mu_Y)(\bar{\mu}_{Y,3}-\sigma_Y^2)\alpha^{m+k}+(1+\mu_Y)^2\sigma_Y^2\alpha^m\\
&\quad+\mu_Y(1+\mu_Y)\sigma_Y^2(\alpha^{m-k}+\alpha^l)+\sigma_Y^4\alpha^{m-l+k}\\
&\quad+\mu_Y(\bar{\mu}_{Y,3}-\sigma_Y^2)\alpha^{l+k}+\mu_Y^2\sigma_Y^2(\alpha^{m-l}+\alpha^{l-k}+\alpha^k)+\mu_Y^4\\
&=\alpha^{m+l+k}\left(\bar{\mu}_{Y,4}-3\bar{\mu}_{Y,3}+\sigma_Y^2(2-3\sigma_Y^2)\right)\ +\sigma_Y^4(\alpha^{m-l+k}+2\alpha^{m+l-k})\\
&\quad+(\bar{\mu}_{Y,3}-\sigma_Y^2)\Big((2+\mu_Y)\alpha^{m+l}+(1+\mu_Y)\alpha^{m+k}+\mu_Y(\alpha^{m+l-2k}+\alpha^{l+k})\Big)\\
&\quad+(1+\mu_Y)^2\sigma_Y^2\alpha^m+\mu_Y(1+\mu_Y)\sigma_Y^2(\alpha^{m-k}+\alpha^l)\\
&\quad+\mu_Y^2\sigma_Y^2(\alpha^{m-l}+\alpha^{l-k}+\alpha^k)+\mu_Y^4
\end{aligned}
$$

holds for any $0\le k\le l\le m$.

□

We now consider a special case, the so-called *Poisson INAR(1) model*. In this model, the arrivals $(\epsilon_t)_{t\in\mathbb{Z}}$ are assumed to be Poisson distributed, this is the most popular instance of the INAR(1) family. Let us first show that in this case, the stationary marginal distribution is also a Poisson distribution, $\mathrm{Poi}(\frac{\lambda}{1-\alpha})$. The following assertion is well-known, yet the proof via a discrete-time queueing model seems to be a novel idea.

**Lemma 4.1.4.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process, where $\epsilon_t\sim\mathrm{Poi}(\lambda)$ for all $t\in\mathbb{Z}$. Then $Y_0\sim\mathrm{Poi}(\frac{\lambda}{1-\alpha})$.*

*Proof.* As pointed out in the introduction of this chapter, an INAR(1) process is equal in distribution to (3.1) when setting $G(k) = 1 - \alpha^k$. A closer look at the proof of Lemma 3.1.1 reveals that $Y_0 = \sum_{i=-\infty}^{0} b(i; 0, \infty) = \sum_{i=0}^{\infty} b(0; i, \infty)$, hence $Y_0$ is Poisson distributed with parameter $\lambda \sum_{i=0}^{\infty} \sum_{l=i}^{\infty} g_l = \lambda \sum_{i=0}^{\infty}(1-G(i))$, concluding the proof. □

It is easily seen that in this case, we have $\bar{\mu}_{Y,3} - \sigma_Y^2 = 0$ and, for higher order moments, $\bar{\mu}_{Y,4} - 3\bar{\mu}_{Y,3} + \sigma_Y^2(2 - 3\sigma_Y^2) = 0$, leading to a simplification of the results of Theorem 4.1.3.

**Corollary 4.1.5** (Weiß (2012), Proposition 1)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process, where $\epsilon_t \sim \mathrm{Poi}(\lambda)$ for all $t \in \mathbb{Z}$. Then*

$$
\begin{aligned}
\mu(k) &= \mu_Y(\alpha^k + \mu_Y),\\
\mu(k,l) &= \mu_Y \alpha^l + \mu_Y^2(\alpha^{l-k} + \alpha^k + \alpha^l) + \mu_Y^3,\\
\mu(k,l,m) &= \mu_Y^2 \left(\alpha^{m-l+k} + 2\alpha^{m+l-k}\right) + \mu_Y(1+\mu_Y)^2\alpha^m + \mu_Y^4\\
&\quad + \mu_Y^2(1+\mu_Y)\left(\alpha^{m-k} + \alpha^l\right) + \mu_Y^3\left(\alpha^{m-l} + \alpha^{l-k} + \alpha^k\right)
\end{aligned}
$$

*holds for any $0 \le k \le l \le m$.*

The next result presents a surprisingly simple relation for the joint cumulants of Poisson INAR(1) processes. This relation was suggested and used in Pickands and Stine (1997), however, these authors only gave a sketch of a proof. A more formal proof is presented here.

**Theorem 4.1.6** (Schweer and Weiß (2015), Theorem 2.1.1)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process, where $\epsilon_t \sim \mathrm{Poi}(\lambda)$ for all $t \in \mathbb{Z}$. Let $r \in \mathbb{N}$ and $i_1 \le i_2 \le \cdots \le i_r$ with $i_j \in \mathbb{Z}$. Then*

$$
\mathrm{cum}\left(Y_{i_1}, \ldots, Y_{i_r}\right) = \frac{\lambda}{1-\alpha}\alpha^{i_r - i_1} = \mu_Y \alpha^{i_r - i_1}.
$$

*Proof.* The joint cumulant is well-defined as the marginal distribution of a Poisson INAR(1) process is Poisson distributed (cf. Lemma 4.1.4) and thus all moments are finite. Let $j_0 := 0 \le j_1 \le j_2 \le \ldots j_r$ and $j_i \in \mathbb{N}_0$. Then we calculate, with (4.3),

$$
\begin{aligned}
&\mathrm{cum}\left(\alpha^{j_1} \circ \epsilon_0, \alpha^{j_2} \circ \epsilon_0, \ldots, \alpha^{j_r} \circ \epsilon_0 \mid \epsilon_0\right)\\
&= \mathrm{cum}\left(\sum_{m=1}^{\epsilon_0}\prod_{k=1}^{j_1}\xi_{m,k}, \ldots, \sum_{m=1}^{\epsilon_0}\prod_{k=1}^{j_r}\xi_{m,k} \,\middle|\, \epsilon_0\right) = \mathrm{cum}\left(\sum_{m=1}^{\epsilon_0}\prod_{k=1}^{j_1}\xi_{m,k}, \ldots, \prod_{k=1}^{j_r}\xi_{m,k} \,\middle|\, \epsilon_0\right)\\
&= \epsilon_0\,\mathrm{cum}\left(\prod_{k=1}^{j_1}\xi_{1,k}, \ldots, \prod_{k=1}^{j_r}\xi_{1,k}\right),
\end{aligned}
$$

where the penultimate equation follows from the mutual independence of the random variables $\xi_{a,b}$ and Lemma 2.2.1 (ii). The last equation follows from relation (2.4). With

Theorem 2.2.2 and the notation used there Lemma 2.2.1 (iv) shows that the following calculations hold:

$$
\begin{aligned}
&\operatorname{cum}\left(\alpha^{j_1}\circ\epsilon_0,\alpha^{j_2}\circ\epsilon_0,\dots,\alpha^{j_r}\circ\epsilon_0\right)\\
&=\sum_{\pi\in\Pi_r}\prod_{i=1}^{|\pi|}\left[\operatorname{cum}\left(\prod_{k=1}^{j_{b_i(\pi,1)}}\xi_{1,k},\dots,\prod_{k=1}^{j_{b_i(\pi,p_i)}}\xi_{1,k}\right)\cdot\kappa_{|\pi|}\left(\epsilon_0\right)\right]\\
&=\kappa_1\left(\epsilon_0\right)\sum_{\pi\in\Pi_r}\prod_{i=1}^{|\pi|}\operatorname{cum}\left(\prod_{k=1}^{j_{b_i(\pi,1)}}\xi_{1,k},\dots,\prod_{k=1}^{j_{b_i(\pi,p_i)}}\xi_{1,k}\right)=\kappa_1\left(\epsilon_0\right)\mathbb{E}\left[\prod_{l=1}^{r}\prod_{k=1}^{j_l}\xi_{1,k}\right].
\end{aligned}
\tag{4.8}
$$

The penultimate equation exploited the fact that $\kappa_1(X)=\kappa_2(X)=\dots=\kappa_r(X)$ for any $r\in\mathbb{N}$ for Poisson distributed $X$, and the last equation uses Lemma 2.2.1 (vi). Due to the mutual independence of the random variables $\xi$, and since $\xi_{m,k}^u=\xi_{m,k}$ for any $u\in\mathbb{N}$ for Bernoulli random variables, we conclude

$$
\operatorname{cum}\left(\alpha^{j_1}\circ\epsilon_0,\alpha^{j_2}\circ\epsilon_0,\dots,\alpha^{j_r}\circ\epsilon_0\right)=\kappa(\epsilon_0)\prod_{l=1}^{r}\mathbb{E}\left[\prod_{k=j_{l-1}}^{j_l}\xi_{1,k}\right]=\alpha^{j_r}\kappa(\epsilon_0). \tag{4.9}
$$

Now, since equality in distribution implies equality of the corresponding cumulants, we apply (4.2) and Lemma 2.2.1 (vii) repeatedly (note the mutual independence of the $\epsilon$'s and the independence of $\epsilon_t$ of $Y_s$ for $s<t$). We obtain

$$
\begin{aligned}
\operatorname{cum}\left(Y_{j_1},\dots,Y_{j_r}\right)&=\operatorname{cum}\left(Y_{j_1},\alpha^{j_2-j_1}\circ Y_{j_1},\dots,\alpha^{j_r-j_1}\circ Y_{i_1}\right)\\
&=\operatorname{cum}\left(\sum_{s=0}^{\infty}\alpha^s\circ\epsilon_{j_1-s},\sum_{s=0}^{\infty}\alpha^{j_2-j_1+s}\circ\epsilon_{j_1-s},\dots,\sum_{s=0}^{\infty}\alpha^{j_r-j_1+s}\circ\epsilon_{j_1-s}\right)\\
&=\sum_{s=0}^{\infty}\operatorname{cum}\left(\alpha^s\circ\epsilon_{j_1-s},\alpha^{j_2-j_1+s}\circ\epsilon_{j_1-s},\dots,\alpha^{j_r-j_1+s}\circ\epsilon_{j_1-s}\right),
\end{aligned}
$$

and conclude the proof with an appeal to (4.9). □

Note that the result of Theorem 4.1.6 can be used to extend the existing results for the joint moments of Poisson INAR(1) processes as given in Weiß (2012), cp. Corollary 4.1.5, by applying Lemma 2.2.1 (vi). For general INAR(1) processes we find the following corollary.

**Corollary 4.1.7.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be an INAR(1) process with $\mathbb{P}(\epsilon_0=0)\in(0,1)$ and let $r\in\mathbb{N}$ with $\mathbb{E}[Y_t^r]<\infty$. Let $i_1\le i_2\le\dots\le i_r$ with $i_j\in\mathbb{Z}$. Then*

$$
\operatorname{cum}\left(Y_{i_1},\dots,Y_{i_r}\right)=\sum_{s=0}^{\infty}\sum_{\pi\in\Pi_r}\kappa_{|\pi|}\left(\epsilon_0\right)\prod_{i=1}^{|\pi|}\operatorname{cum}\left(\prod_{k=1}^{j_{b_i(\pi,1)}(s)}\xi_{1,k},\dots,\prod_{k=1}^{j_{b_i(\pi,p_i)}(s)}\xi_{1,k}\right),
$$

*where we denote the sets of indices $\{s,i_2-i_1+s,i_3-i_1+s,\dots,i_r-i_1+s\}:=\{j_1(s),\dots,j_r(s)\}$.*

In comparison with the general result of Corollary 4.1.7, the simple relation of Theorem 4.1.6 becomes even more striking. On the other hand we would like to remark that the direct calculation of joint cumulants via Corollary 4.1.7 might seem daunting, yet turns out to be quite simple. We demonstrate this by giving a short example.

**Example 4** (Joint Moments of INAR(1) Processes)**.** *First, the second order joint cumulant corresponds to the covariance of two random variables. We calculate*

$$\begin{aligned} \operatorname{cum}(Y_0, Y_h) &= \sum_{s=0}^{\infty} \left( \operatorname{cum}\left( \prod_{k=1}^{s} \xi_{1,k}, \prod_{k=1}^{s+h} \xi_{1,k} \right) \kappa_1(\epsilon_0) + \kappa_2(\epsilon_0)\alpha^s \alpha^{h+s} \right) \\ &= \sum_{s=0}^{\infty} \left( \kappa_1(\epsilon_0) \left( \alpha^{h+s} - \alpha^s \alpha^{h+s} \right) + \kappa_2(\epsilon_0)\alpha^s \alpha^{h+s} \right) \\ &= \alpha^h \frac{\kappa_2(\epsilon_0) + \alpha\kappa_1(\epsilon_0)}{1-\alpha^2}. \end{aligned}$$

*Now, from* (4.6) *together with* (4.7) *and recalling that* $\operatorname{Var}(X) = \kappa_2(X)$ *and* $\mathbb{E}[X] = \kappa(X)$ *for any random variable with existing second moment, we have that*

$$\begin{aligned} \operatorname{Var}(Y_0) &= \frac{\kappa_2(\epsilon_0) + \alpha\kappa_1(\epsilon_0)}{1-\alpha^2} + \frac{\kappa_1(\epsilon_0)^2 + 2\alpha\kappa_1(\epsilon_0)\kappa_1(Y_0)}{1-\alpha^2} - \kappa_1(Y_0)^2 \\ &= \frac{\kappa_2(\epsilon_0) + \alpha\kappa_1(\epsilon_0)}{1-\alpha^2} + \kappa_1(\epsilon_0)^2 \frac{1+\alpha}{1-\alpha} \frac{1}{1-\alpha^2} - \frac{\kappa_1(\epsilon_0)^2}{(1-\alpha)^2} = \frac{\kappa_2(\epsilon_0) + \alpha\kappa_1(\epsilon_0)}{1-\alpha^2}, \end{aligned}$$

*so that* $\operatorname{cum}(Y_0, Y_h) = \alpha^h \operatorname{Var}(Y_0)$*, as already shown above. We used* $\mathbb{E}[Y_0] = \mathbb{E}[\epsilon_0]/(1-\alpha)$ *repeatedly in this calculation, this follows by taking expectations in* (4.2)*.*

### 4.1.2 Time-Reversibility

Let us discuss the time-reversibility of INAR(1) processes. Following the discussion in the introduction of this chapter, an INAR(1) process is a special case of a $GI/G/\infty$-queueing system and thus Lemma 3.1.1 is applicable. This implies that if we assume that the innovations $(\epsilon_t)_{t\in\mathbb{Z}}$ are i.i.d. according to the Poisson distribution $\operatorname{Poi}(\lambda)$ for some $\lambda > 0$, the resulting INAR(1) process is time-reversible.

Lemma 4.1.4 allows for a very short proof of time-reversibility of a stationary Poisson INAR(1) process. As an INAR(1) process $(Y_t)_{t\in\mathbb{Z}}$ is Markovian, it suffices to prove the relation $\pi_Y(k)p_Y(l|k) = \pi_Y(l)p_Y(k|l)$ for all $l, k \in \mathbb{N}_0$, where $\pi_Y(k) := \mathbb{P}(Y_0 = k)$ denotes the probability mass function (pmf) of the stationary distribution, and $p_Y(l|k) := \mathbb{P}(Y_1 = l|Y_0 = k)$ denotes the transition probabilities. In fact, for any $l, k \in \mathbb{N}_0$ it follows with (4.1) and Lemma 4.1.4 that

$$\begin{aligned} \pi_Y(k)p_Y(l|k) &= \exp(-2\lambda)\frac{\lambda^k}{(1-\alpha)^k k!} \sum_{j=0}^{\min\{k,l\}} \binom{k}{j} \alpha^j (1-\alpha)^{k-j} \frac{\lambda^{l-j}}{(l-j)!} \\ &= \exp(-2\lambda)\frac{\lambda^l}{(1-\alpha)^l l!} \sum_{j=0}^{\min\{k,l\}} \binom{l}{j} \alpha^j (1-\alpha)^{l-j} \frac{\lambda^{k-j}}{(k-j)!} = \pi_Y(l)p_Y(k|l). \end{aligned}$$

The time-reversibility of Poisson INAR(1) processes was first shown in Walrand (1983). The converse of this statement is also true.

**Theorem 4.1.8** (Schweer (2015b), Theorem 19.1)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be an INAR(1) process with $\mathbb{P}(\epsilon_0 = 0) \in (0,1)$. Then $(Y_t)_{t\in\mathbb{Z}}$ is time-reversible if and only if there exists a $\lambda > 0$ such that $\epsilon_0 \sim \mathrm{Poi}(\lambda)$.*

*Proof.* The sufficiency of the Poisson assumption is shown above. Now, let $(Y_t)_{t\in\mathbb{Z}}$ be time-reversible and let $i \in \mathbb{N}$. By Theorem 2.4.2 applied to the sequence $j_1 = 0, j_2 = 1, j_3 = i$, it necessarily holds that $p_Y(1|0)p_Y(i|1)p_Y(0|i) = p_Y(i|0)p_Y(1|i)p_Y(0|1)$. With (4.1), this implies

$$\begin{aligned}&\mathbb{P}\left(\epsilon_0 = 1\right)\left[(1-\alpha)\mathbb{P}\left(\epsilon_0 = i\right) + \alpha\mathbb{P}\left(\epsilon_0 = i-1\right)\right](1-\alpha)^i\mathbb{P}\left(\epsilon_0 = 0\right)\\ &= \mathbb{P}\left(\epsilon_0 = i\right)\left[(1-\alpha)^i\mathbb{P}\left(\epsilon_0 = 1\right) + \binom{i}{i-1}\alpha(1-\alpha)^{i-1}\mathbb{P}\left(\epsilon_0 = 0\right)\right](1-\alpha)\mathbb{P}\left(\epsilon_0 = 0\right),\end{aligned}$$

which, due to $1 - \alpha > 0$, is equivalent to

$$\mathbb{P}\left(\epsilon_0 = 1\right)\mathbb{P}\left(\epsilon_0 = i-1\right) = i\mathbb{P}\left(\epsilon_0 = i\right)\mathbb{P}\left(\epsilon_0 = 0\right). \tag{4.10}$$

Summation over $i$ on both sides leads to $\mathbb{P}\left(\epsilon_0 = 1\right) = \mathbb{E}[\epsilon_0]\mathbb{P}\left(\epsilon_0 = 0\right)$. We assumed that $\mathbb{P}(\epsilon_0 = 0) \in (0,1)$, hence it holds that $\mathbb{E}[\epsilon_0] > 0$ and therefore that $\mathbb{P}(\epsilon_0 = 1) > 0$. Applying (4.10) recursively,

$$\mathbb{P}\left(\epsilon_0 = i\right) = \frac{1}{i}\mathbb{E}[\epsilon_0]\mathbb{P}(\epsilon_0 = i-1) = \cdots = \frac{1}{i!}\mathbb{E}[\epsilon_0]^i\mathbb{P}(\epsilon_0 = 0).$$

Normalization yields $\mathbb{P}\left(\epsilon_0 = 0\right)^{-1} = \sum_{l=0}^{\infty}\frac{1}{l!}\mathbb{E}[\epsilon_0]^l$, concluding the proof. □

Theorem 4.1.8 requires that $\mathbb{P}(\epsilon_0 = 0) \in (0,1)$. We could also require that $(Y_t)_{t\in\mathbb{Z}}$ is irreducible on $\mathbb{N}_0$, the equivalence of these conditions was shown in Lemma 4.1.1.

## 4.2 Compound Poisson INAR(1) Processes

Let us begin by demonstrating an alternative proof for Lemma 4.1.4 via pgfs. For this we first calculate for any integer-valued random variable $X$ and $\alpha \in (0,1)$,

$$\mathrm{pgf}_{\alpha\circ X}(z) = \sum_{k=0}^{\infty}\mathbb{E}\left[z^{\alpha\circ X}\middle| X = k\right]\mathbb{P}(X = k) = \sum_{k=0}^{\infty}\mathbb{E}[z_1^{\xi}]^k\mathbb{P}(X = k) = \mathrm{pgf}_X(1 - \alpha + \alpha z),$$

where $\xi_1$ denotes a thinning random variable, i.e., a Bernoulli random variable with probability of success $\alpha$. This argument may be extended to iterated applications of the thinning operation, as long as the thinnings involved are executed independently. In this case, the relation above shows that

$$\mathrm{pgf}_{\alpha\circ(\beta\circ X)}(z) = \mathrm{pgf}_{\beta\circ X}(1 - \alpha + \alpha z) = \mathrm{pgf}_X(1 - \alpha\beta + \alpha\beta z),$$

implying that

$$\alpha \circ (\beta \circ X) \stackrel{\mathcal{D}}{=} (\alpha \cdot \beta) \circ X \tag{4.11}$$

for independent thinning operations. Now, by elementary properties of power series, it follows that the pgf of the marginal distribution of an INAR(1) process $\text{pgf}_Y(z)$ is uniquely determined by the relation

$$\text{pgf}_Y(z) = \text{pgf}_Y(1 - \alpha + \alpha z)\, \text{pgf}_\epsilon(z) \quad \text{for} \quad z \in [0, 1]. \tag{4.12}$$

Using this relation, a second proof of Lemma 4.1.4 can be given.

*Proof.* Since $\epsilon \sim \text{Poi}(\lambda)$, it follows that $\text{pgf}_\epsilon(z) = \exp(\lambda(z - 1))$, cf. Definition 2.3.1. It is clear that setting $P_Y(z) = \exp(\frac{\lambda}{1-\alpha}(z - 1))$ is a solution to (4.12) and thus the unique solution, concluding the proof. □

We have just seen that for the Poisson INAR(1) model, the marginal distribution of the observations and that of the innovations belong to the same distribution family, the Poisson distribution. It should be noted, however, that the more general family of Compound Poisson distributions has analogous invariance properties.

**Lemma 4.2.1** (Schweer and Weiß (2014), Lemma B.4)**.** *Let $Y_1, Y_2$ be independent with $Y_i \sim \text{ComPoi}_\nu(\lambda_i, H_i)$ (including the case $\nu = \infty$) and let $\alpha \in (0, 1)$. Then*

*(i) $Y_1 + Y_2 \sim \text{ComPoi}_\nu(\lambda, H)$ with*

$$\lambda = \lambda_1 + \lambda_2, \quad \lambda H(z) = \sum_{x=1}^{\nu} (\lambda_1 h_{1;x} + \lambda_2 h_{2;x}) z^x,$$

*(ii) $\alpha \circ Y \sim \text{ComPoi}_\nu(\mu, G)$, where*

$$\mu = \lambda\left(1 - H(1 - \alpha)\right), \quad \mu G(z) = \lambda \sum_{j=1}^{\nu} \alpha^j \left( \sum_{i=j}^{\nu} h_i \binom{i}{j} (1 - \alpha)^{i-j} \right) z^j.$$

*Proof.* The additivity property stated in part (i) follows from

$$\begin{aligned} \text{pgf}_{Y_1+Y_2}(z) &= \text{pgf}_{Y_1}(z)\, \text{pgf}_{Y_2}(z) = \exp\left(\lambda_1 H_1(z) + \lambda_2 H_2(z) - (\lambda_1 + \lambda_2)\right) \\ &= \exp\left( (\lambda_1 + \lambda_2) \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} H_1(z) + \frac{\lambda_2}{\lambda_1 + \lambda_2} H_2(z) - 1 \right) \right). \end{aligned}$$

To prove the invariance with regard to binomial thinning as stated in part (ii), we consider (recall that $h_0 := 0$ may be assumed as shown in Section 2.3)

$$H(1 - \alpha + \alpha z) = \sum_{j=0}^{\nu} z^j \sum_{i=j}^{\nu} h_i \binom{i}{j} (1 - \alpha)^{i-j} \alpha^j =: \sum_{j=0}^{\nu} \tilde{h}_j z^j,$$

where $\tilde{h}_0 = \sum_{i=1}^{\nu} h_i(1-\alpha)^i = H(1-\alpha)$. Hence,

$$\begin{aligned}\mathrm{pgf}_{\alpha\circ Y}(z) &= \mathrm{pgf}_Y(1-\alpha+\alpha z) = \exp\left(\lambda\left(H(1-\alpha+\alpha z)-1\right)\right)\\ &= \exp\left(\lambda(1-\tilde{h}_0)\left(\sum_{j=1}^{\nu}\frac{\tilde{h}_j}{1-\tilde{h}_0}z^j - 1\right)\right),\end{aligned}$$

which completes the proof. □

These invariance principles suggest that the consideration of Compound Poisson distribution in the context of INAR(1) processes yields nice mathematical properties. A further connection between these two concepts can be established via discrete self-decomposability, as can be gleaned from the similarity of the relations (2.7) and (4.12). This motivates the following definition of the Compound Poisson integer-valued autoregressive process of first order (CPINAR(1)).

**Definition 4.2.2** (CPINAR(1) Process)**.** *An INAR(1) process $(Y_t)_{t\in\mathbb{Z}}$ according to Definition 1.1.1 is referred to as a* CPINAR(1) process *if there exists a $\lambda > 0$ and a pgf $H(z)$ with* $\deg(H(z)) := \nu \in \mathbb{N}\cup\{\infty\}$ *such that* $\epsilon_t \sim \mathrm{ComPoi}_\nu(\lambda, H)$ *for all $t \in \mathbb{Z}$.*

Note that any CPINAR(1) process is stationary by Lemma 4.1.2, as Proposition 2.3.6 (i) implies that $\mathbb{P}(\epsilon_0 = 0) = \exp(-\lambda) \in (0,1)$ for any $\lambda > 0$, independent of the choice of compounding distribution $H(z)$. The general approach of Definition 4.2.2 comprises a number of specialized INAR(1) models within one model. The popular Poisson INAR(1) model described above just corresponds to the case $\nu = 1$. Further models which are included in the above definition were considered in Jung et al. (2005), Pedeli and Karlis (2011) (negative binomial innovations) and Schweer and Wichelhaus (2015b) (finite compounding structure).

### 4.2.1 Forecasting

As a first consequence of the semi-parametric choice of Definition 4.2.2, the $k$-step-ahead conditional distribution of a CPINAR(1) process for arbitrary $k \in \mathbb{N}$ is derived.

**Theorem 4.2.3** (Schweer and Weiß (2014), Theorem 3.1.1)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a CPINAR(1) process according to Definition 4.2.2. Then the conditional pgf of $Y_{t+k}$ given $Y_t$ satisfies*

$$\mathrm{pgf}_{Y_{t+k}|Y_t}(z) = \left(1-\alpha^k+\alpha^k z\right)^{Y_t}\cdot \mathrm{pgf}_{\epsilon^{(k)}}(z), \tag{4.13}$$

*where $\epsilon^{(k)}$ is a* $\mathrm{ComPoi}_\nu(\lambda^{(k)}, H^{(k)})$*-distributed random variable with*

$$\begin{aligned}\lambda^{(k)} &= \lambda\sum_{i=1}^{k}\left(1-H(1-\alpha^{i-1})\right),\\ \lambda^{(k)}\left(H^{(k)}(z)-1\right) &= \lambda\sum_{i=1}^{k}\left(H(1-\alpha^{i-1}+\alpha^{i-1}z)-1\right).\end{aligned} \tag{4.14}$$

Note that the first equation in (4.14) is included in the second one for $z = 0$ (recalling that $H^{(k)}(0) = 0$ according to Definition 2.3.1).

*Proof.* We abbreviate $C_k := \lambda^{(k)}/\lambda$ and we prove the theorem by induction. The 1-step-ahead conditional pgf is given by

$$\mathrm{pgf}_{Y_{t+1}|Y_t}(z) = (1-\alpha+\alpha z)^{Y_t}\,\mathrm{pgf}_\epsilon(z), \tag{4.15}$$

i.e., (4.13) holds for $k = 1$ with $C_1 = 1$ and $H^{(1)}(z) = H(z)$. Now, suppose that equations (4.13), (4.14) hold for some $k \geq 1$. For the $(k+1)$-step-ahead conditional pgf, we calculate

$$\begin{aligned}
\mathrm{pgf}_{Y_{t+k+1}|Y_t}(z) &= \mathbb{E}\left[z^{Y_{t+k+1}}\middle|Y_t\right] = \mathbb{E}\left[\mathbb{E}\left[z^{Y_{t+k+1}}\middle|Y_{t+1}\right]\middle|Y_t\right] \\
&= \mathbb{E}\left[\mathrm{pgf}_{Y_{t+k+1}|Y_{t+1}}(z)\middle|Y_t\right] \overset{(4.13)}{=} \mathbb{E}\left[(1-\alpha^k+\alpha^k z)^{Y_{t+1}}\,\mathrm{pgf}_{\epsilon^{(k)}}(z)\middle|Y_t\right] \\
&= \mathbb{E}\left[y^{Y_{t+1}}\middle|Y_t\right]\mathrm{pgf}_{\epsilon^{(k)}}(z) \overset{(4.15)}{=} (1-\alpha+\alpha y)^{Y_t}\,\mathrm{pgf}_\epsilon(y)\,\mathrm{pgf}_{\epsilon^{(k)}}(z),
\end{aligned}$$

where $y := 1-\alpha^k+\alpha^k z$. Re-substitution yields

$$\begin{aligned}
\mathrm{pgf}_{Y_{t+k+1}|Y_t}(z) &= \Big(1-\alpha+\alpha(1-\alpha^k+\alpha^k z)\Big)^{Y_t}\,\mathrm{pgf}_\epsilon\Big(1-\alpha^k+\alpha^k z\Big)\,\mathrm{pgf}_{\epsilon^{(k)}}(z) \\
&= (1-\alpha^{k+1}+\alpha^{k+1}z)^{Y_t}\,\mathrm{pgf}_\epsilon\Big(1-\alpha^k+\alpha^k z\Big)\,\mathrm{pgf}_{\epsilon^{(k)}}(z).
\end{aligned} \tag{4.16}$$

Now, the last product on the RHS of (4.16) is the pgf of a sum of two independent random variables $\epsilon^*$ and $\epsilon^{(k)}$, where (see Lemma 4.2.1 (ii))

$$\begin{aligned}
&\epsilon^{(k)} \sim \mathrm{ComPoi}_\nu(\lambda^{(k)}, H^{(k)}), \quad \epsilon^* \sim \mathrm{ComPoi}_\nu(\mu^*, G^*) \quad \text{with} \\
&\mu^* = \lambda\Big(1-H(1-\alpha^k)\Big), \quad \mu^*\left(G^*(z)-1\right) = \lambda\Big(H(1-\alpha^k+\alpha^k z)-1\Big).
\end{aligned}$$

Lemma 4.2.1 (i) yields that $\epsilon^{(k+1)} := \epsilon^* + \epsilon^{(k)}$ is $\mathrm{ComPoi}_\nu(\lambda^{(k+1)}, H^{(k+1)})$-distributed with

$$\begin{aligned}
C_{k+1} &= 1 - H(1-\alpha^k) + C_k, \\
C_{k+1}\Big(H^{(k+1)}(z)-1\Big) &= \Big(H(1-\alpha^k+\alpha^k z)-1\Big) + C_k\Big(H^{(k)}(z)-1\Big).
\end{aligned} \tag{4.17}$$

To conclude the proof, we apply (4.17) inductively. □

As the CPINAR(1) model comprises a number of specialized INAR(1) models, this forecasting result also includes some known results as a special case. For a Poisson INAR(1) model ($\nu = 1$), Theorem 4.2.3 implies that

$$\epsilon^{(k)} \sim \mathrm{Poi}\left(\lambda\frac{1-\alpha^k}{1-\alpha}\right),$$

this result is known from Theorem 1 in Freeland (2010). If $\nu = \infty$ with $\epsilon_t \sim \text{NegBin}(n, \pi)$, then formula (4.14) and Example 1 imply that

$$\text{pgf}_{\epsilon^{(k)}}(z) = \prod_{i=0}^{k-1} \left(1 + \frac{1-\pi}{\pi}\alpha^i(1-z)\right)^{-n}.$$

This particular result was also shown in formula (5.4) and Corollary 2 in Pedeli and Karlis (2011) (setting $s_2 = 1$ and $\beta = n^{-1}$, $\lambda_1 = n(1-\pi)/\pi$ in their notation).

### 4.2.2 Stationarity

The CPINAR(1) process $(Y_t)_{t\in\mathbb{Z}}$ of Definition 4.2.2 is a homogeneous Markov chain, where the 1-step transition probabilities are given by formula (4.1). We assume that $H'(1) < \infty$ such that $\mu_\epsilon = \mathbb{E}[\epsilon_t] = \lambda H'(1) < \infty$ holds, see part (ii) of Proposition 2.3.6. Note that $H'(1) < \infty$ is always satisfied if $\nu < \infty$, but also if we are concerned, e.g., with a negative binomial distribution, see Example 1. First, however, we need to prove two rather technical upper bounds related to Lemma 1' of Heathcote (1966), where it was shown that whenever $H'(1) < \infty$ holds, we also have $\sum_{i=0}^{\infty} \left(1 - H(1-\alpha^i)\right) < \infty$. This property, in turn, is of importance for Theorem 4.2.5.

**Lemma 4.2.4** (Schweer and Weiß (2014), Lemma A.2.1)**.** *Let the pgf $H(z)$ of the* ComPoi$(\lambda, H)$-*distribution satisfy $\sum_{i=0}^{\infty} \left(1 - H(1-\alpha^i)\right) < \infty$, and let $\alpha \in (0,1)$. Then*

*(i)* $\sum_{i=0}^{\infty} \left(1 - H(1-\alpha^i+\alpha^i z)\right) < \infty$ *for all* $z \in [0,1]$,

*(ii)* $\prod_{i=0}^{\infty} \exp\left(\lambda\left(H(1-\alpha^i+\alpha^i z) - 1\right)\right) < \infty$ *for all* $z \in [0,1]$.

*Proof.* As $H(z)$ is monotonically increasing in $z$ on the interval $z \in [0,1]$, we have $1 - H(1-\alpha^i+\alpha^i z) \in \left[0, 1 - H(1-\alpha^i)\right]$. Hence, by the Weierstrass M-test, the expression $\sum_{i=0}^{\infty} \left(H(1-\alpha^i+\alpha^i z) - 1\right)$ converges if the term $\sum_{i=0}^{\infty} \left(1 - H(1-\alpha^i)\right)$ converges. So part (i) follows from the assumption.

For the convergence of the infinite product in (ii), it suffices to show that the series

$$\sum_{i=0}^{\infty} \left|\exp\left(\lambda\left(H(1-\alpha^i+\alpha^i z) - 1\right)\right) - 1\right| \tag{4.18}$$

converges. As $H(1-\alpha^i+\alpha^i z) \le 1$ for all $\alpha \in (0,1)$ and $z \in [0,1]$, we find that $0 < \exp\left(\lambda\left(H(1-\alpha^i+\alpha^i z) - 1\right)\right) \le 1$. Using the inequality $\exp(x) \ge 1 + x$, which holds for all $x \in \mathbb{R}$, we have

$$1 - \exp\left(\lambda\left(H(1-\alpha^i+\alpha^i z) - 1\right)\right) \le \lambda\left(1 - H(1-\alpha^i+\alpha^i z)\right).$$

By the Weierstrass M-test, the convergence of (4.18) is thus ensured by the convergence of $\sum_{i=0}^{\infty} \left(1 - H(1-\alpha^i+\alpha^i z)\right)$, which, in turn, was shown in part (i). □

We now present the main result of this section. As shown in Lemma 4.1.2, the condition $H'(1) < \infty$ guarantees that the CPINAR(1) process is an ergodic Markov chain, and that it possesses a unique stationary marginal distribution. Properties of this stationary marginal distribution, refining assertion (2) in Pakes (1971), are also studied.

**Theorem 4.2.5** (Schweer and Weiß (2014), Theorem 3.2.1)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a CPINAR(1) process according to Definition 4.2.2. If $H'(1) < \infty$, then $(Y_t)_{t\in\mathbb{Z}}$ is an ergodic Markov chain and it holds that $Y_0 \sim \mathrm{ComPoi}_\nu(\mu, G)$, where*

$$\mu = \lambda \sum_{i=0}^{\infty} \left(1 - H(1-\alpha^i)\right),$$
$$\mu\left(G(z) - 1\right) = \lambda \sum_{i=0}^{\infty} \left(H(1-\alpha^i+\alpha^i z) - 1\right) \text{ for all } z \in [0,1].$$

In the case $\nu = 1$, we obtain the well-known expressions $\mu = \frac{\lambda}{1-\alpha}$ and $G(z) = H(z) = z$ for the stationary marginal distribution of a Poisson INAR(1) process, providing yet another proof for Lemma 4.1.4.

*Proof.* With Proposition 2.3.6 (i), $\mathbb{P}(\epsilon_0 = 0) = \exp(-\lambda) \in (0,1)$, so Lemma 4.1.2 applies, showing that a unique stationary distribution $\pi_Y(k) := \mathbb{P}(Y_t = k)$ exists. It follows that (cf. Theorem 7.1, Ch. XV in Feller (1968))

$$\mathbb{P}\left(Y_{t+k} = y | Y_t\right) \overset{k\to\infty}{\longrightarrow} \pi_Y(y) \quad \text{for all } y \in \mathbb{N}_0.$$

This is equivalent to the convergence of the respective pgfs, which by Theorem 4.2.3 results in

$$\left(1-\alpha^k+\alpha^k z\right)^{Y_t} \exp\left(\lambda^{(k)}\left(H^{(k)}(z)-1\right)\right) \overset{k\to\infty}{\longrightarrow} \mathrm{pgf}_Y(z) \quad \text{for } z \in [0,1].$$

This, in turn, implies

$$\mathrm{pgf}_Y(z) = \lim_{k\to\infty} \exp\left(\lambda^{(k)}\left(H^{(k)}(z)-1\right)\right) \quad \text{for all } z \in [0,1],$$

as $\lim_{k\to\infty}(1-\alpha^k+\alpha^k z)^{Y_t} = 1$. Using the result of Theorem 4.2.3, we find

$$\begin{aligned}\mathrm{pgf}_Y(z) &= \prod_{i=1}^{\infty} \exp\left(\lambda\left(H(1-\alpha^{i-1}+\alpha^{i-1}z)-1\right)\right)\\ &= \exp\left(\lambda \sum_{i=1}^{\infty}\left(H(1-\alpha^{i-1}+\alpha^{i-1}z)-1\right)\right), \end{aligned} \tag{4.19}$$

where the convergence of this expression is guaranteed by Lemma 4.2.4. By part (ii) of Lemma 4.2.1, we know that

$$\exp\left(\lambda\left(H(1-\alpha^i+\alpha^i z)-1\right)\right) = \exp\left(\mu_i\left(G_i(z)-1\right)\right),$$

where $\mu_i = \lambda\left(1 - H(1-\alpha^i)\right)$, and where $G_i(z)$ satisfies

$$\mu_i G_i(z) = \lambda \sum_{j=1}^{\nu} \left( \alpha^{ij} \sum_{r=j}^{\nu} h_r \binom{r}{j} (1-\alpha^i)^{r-j} \right) z^j.$$

Thus, we can rewrite $\mathrm{pgf}_Y(z)$ from (4.19) in the form

$$\mathrm{pgf}_Y(z) = \exp\left( \sum_{i=0}^{\infty} \mu_i \left(G_i(z) - 1\right) \right) = \exp\left(\mu\left(G(z)-1\right)\right),$$

where $\mu = \sum_{i=0}^{\infty} \mu_i = \lambda \sum_{i=0}^{\infty} \left(1 - H(1-\alpha^i)\right)$, and the expression $\mu(G(z)-1)$ satisfies the assertion. □

If $\nu < \infty$ then $H'(1) < \infty$, and Theorem 4.2.5 yields for the stationary marginal distribution of $(Y_t)_{t\in\mathbb{Z}}$ that $\mathrm{pgf}_Y(z) = \exp\left(\mu\left(G(z)-1\right)\right)$ with $G(z) = \sum_{i=1}^{\nu} g_i z^i$. The parameters $\mu$ and $g_1, \ldots, g_\nu$ can be computed explicitly from $\lambda$ and $h_1, \ldots, h_\nu$ (the ones from the innovations' distribution) by solving the following linear system of equations, see Theorem 2.2 in Schweer and Wichelhaus (2015b):

$$\begin{aligned} &g_1 + \ldots + g_\nu = 1, \ \frac{\lambda}{\mu} - (1-\alpha)g_1 - \ldots - (1-\alpha)^\nu g_\nu = 0, \\ &\quad h_k \frac{\lambda}{\mu} - (1-\alpha^k)\cdot g_k + \alpha^k \sum_{i=k+1}^{\nu} \binom{i}{k} (1-\alpha)^{i-k} g_i = 0 \text{ for } k = \nu, \ldots, 2. \end{aligned} \tag{4.20}$$

Let us illustrate the application of Scheme (4.20) in the following example.

**Example 5** (Finite Compounding Structure)**.** *Let us consider a CPINAR(1) process* $(Y_t)_{t\in\mathbb{Z}}$*, where* $H(z) = (1-h)z + hz^2$ *with parameter* $h \in [0,1]$ *and* $\lambda > 0$ *is arbitrary. Obviously* $H(0) = 0$ *is satisfied, and we may apply the scheme* (4.20) *to explicitly calculate the stationary distribution* ComPoi$(\mu, G)$ *of* $(Y_t)_{t\in\mathbb{Z}}$*. We already know that* $G(z) = g_1 z + g_2 z^2$*, and the third relation yields, for* $k = 2$*,* $g_2 = \frac{\lambda}{\mu}\frac{h}{1-\alpha^2}$*. With the second relation we find that* $g_1 = \frac{\lambda}{\mu}\frac{1+\alpha-h(1-\alpha)}{1-\alpha^2}$*. Thus* $g_1 + g_2 = 1$ *necessitates*

$$1 = \frac{\lambda}{\mu}\frac{1+\alpha+h\alpha}{1-\alpha^2} \quad \text{or} \quad \frac{\mu}{\lambda} = \frac{1+\alpha+h\alpha}{1-\alpha^2}.$$

*Combining these results we find explicit expressions for* $\lambda, g_1$ *and* $g_2$*:*

$$\mu = \lambda\frac{1+\alpha+\alpha h}{1-\alpha^2}, \quad g_1 = 1 - \frac{h}{1+\alpha+\alpha h}, \quad g_2 = \frac{h}{1+\alpha+\alpha h}.$$

### 4.2.3 Mixing Properties

It is well known that each stationary, aperiodic and ergodic Markov chain with a countable range is necessarily $\alpha$-mixing (see Definition 2.5.1) with certain weights $\alpha_Y(n) \to 0$, see, e.g., Theorem 3.2 in Bradley (2005). But the speed of convergence of $\alpha_Y(n)$ is not clear in advance. This speed, however, is an essential prerequisite for further results, such as the central limit theorem of Theorem 2.5.2.

**Theorem 4.2.6** (Schweer and Weiß (2014), Theorem 3.4.1)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a CPINAR(1) process according to Definition 4.2.2 and let $H'(1) < \infty$. Then $(Y_t)_{t\in\mathbb{Z}}$ is $\alpha$-mixing with exponentially decreasing weights $\alpha_Y(n)$.*

*Proof.* Denote the stationary distribution of $(Y_t)_{t\in\mathbb{Z}}$ by $\pi_Y(\cdot)$. A CPINAR(1) process $(Y_t)_{t\in\mathbb{Z}}$ satisfies all conditions required in Pakes (1971)[1] for the subcritical case (Section 2), as $H'(1) < \infty$ implies $\mathbb{E}[\epsilon_t] < \infty$. Thus, it is geometrically ergodic, see Theorem 1 in Pakes (1971), i.e., there exist finite constants $M_{ij}$ such that

$$|\mathbb{P}\left(Y_{t+k} = i \;|Y_t = j\right) - \pi_Y(i)| \leq M_{ij} \cdot \alpha^k$$

for each $(i, j) \in \mathbb{N}_0^2$. For such a geometrically ergodic, irreducible and aperiodic Markov chain (see the proof of Theorem 4.2.5), Theorem 1 of Nummelin and Tweedie (1978) implies the existence of finite constants $M_j$ for $j \in \mathbb{N}_0$ such that

$$|\mathbb{P}\left(Y_{t+k} = i \;|Y_t = j\right) - \pi_Y(i)| \leq M_j \cdot \alpha^k. \tag{4.21}$$

Since $(Y_t)_{t\in\mathbb{Z}}$ is a strictly stationary Markov chain satisfying (4.21), it is also $\beta$-mixing (and thus $\alpha$-mixing) at least exponentially fast as $k \to \infty$, see Theorem 3.7 in Bradley (2005). □

Theorem 4.2.6 now allows us to apply an appropriate type of central limit theorem to the considered CPINAR(1) process. Since our CPINAR(1) process as in Theorem 4.2.6 has exponentially decreasing weights $\alpha_Y(n)$, the condition (2.13) is satisfied for any $\delta > 0$. A concrete application of Theorem 2.5.2. is presented in Section 5.1.3 below.

### 4.2.4 Infinite Divisibility of the Marginal Distribution

As pointed out before, there is a striking similarity between the pgf of the marginal distribution of an INAR(1) process (4.12) and the defining relation for discrete self-decomposable distributions (2.7). In fact, it is easily seen that for any discrete self-decomposable distribution $H$ there exists a stationary INAR(1) process with marginal distribution $H$. By Theorem 2.3.5, this statement can be extended to infinitely divisible distributions under a certain restriction on the canonical measure. The question remains whether this restriction is necessary, i.e., whether *any* infinitely divisible distribution is also a marginal distribution of a stationary INAR(1) process. The following example shows that the answer to this question is negative.

[1]Note that the requirement "$a_0 + a_1 < 1$" in Condition A of Pakes (1971) is not necessary for the proof of his Theorem 1.

**Example 6** (Infinitely Divisible but not Marginal of INAR(1))**.** *Let us assume there exists a stationary INAR(1) process* $(Y_t)_{t\in\mathbb{Z}}$ *with marginal distribution* $\mathrm{ComPoi}(\mu, G)$ *with* $G(z) = z^2$*, this distribution is infinitely divisible by Theorem 2.3.3. By* (4.12)*, the function* $F(z) = \exp(\mu(G(z) - G(1 - \alpha + \alpha z)))$ *emerges as the unique candidate for the pgf of the arrival distribution. We calculate*

$$\begin{aligned} F(z) &= \exp\left(\mu(z-1)(1-\alpha)[(1+\alpha)z + (1-\alpha)]\right) \\ &= \sum_{l=0}^{\infty} \frac{\mu^l (1-\alpha)^l}{l!} \left( \sum_{k_1=0}^{l} \binom{l}{k_1} (-1)^{l-k_1} z^{k_1} \right) \left( \sum_{k_2=0}^{l} \binom{l}{k_2} (1-\alpha)^{l-k_2} ((1+\alpha)z)^{k_2} \right). \end{aligned}$$

*It is easily seen that this is an absolutely convergent sum for* $z \le 1$ *and hence that* $F(z) = \sum_{i=0}^{\infty} a_i z^i$*. For the coefficient of* $z^0$*, we find that* $a_0 = \exp(-\mu(1-\alpha)^2)$*. Moreover, for the coefficient of* $z$*, simple calculation shows that*

$$\begin{aligned} a_1 &= \sum_{l=1}^{\infty} \frac{\mu^l (1-\alpha)^l}{l!} \left( (-1)^{l-1} \binom{l}{1} (1-\alpha)^l + (-1)^l (1-\alpha)^{l-1} (1+\alpha) \binom{l}{1} \right) \\ &= \mu(1-\alpha) \sum_{l=1}^{\infty} \frac{\mu^{l-1} (-1)^{l-1} \left((1-\alpha)^2\right)^{l-1}}{(l-1)!} ((1-\alpha) + (-1)(1+\alpha)) \\ &= -2\mu\alpha(1-\alpha) \exp\left(-\mu(1-\alpha)^2\right). \end{aligned}$$

*Thus,* $a_1 < 0$ *for any choice of* $\mu > 0$ *and* $\alpha \in (0,1)$*, showing that* $F(z)$ *is not a pgf. We may conclude that excluding trivial cases (i.e., excluding the possibility* $\alpha = 0$*) there exists no stationary INAR(1) process with a marginal distribution* $\mathrm{ComPoi}(\mu, G)$ *of the assumed form.*

In order to complete the picture we now ask whether the class of marginal distributions of stationary INAR(1) processes is a proper subset of the class of infinitely divisible distributions. In the following extensive example, we show that the answer to this question is also negative.

**Example 7** (Marginal Distribution but not Infinitely Divisible)**.** *Let* $(Y_t)_{t\in\mathbb{Z}}$ *be a stationary INAR(1) process where* $\epsilon_t \sim \mathrm{Bin}(1,p)$*, i.e., Bernoulli random variables with a probability of success* $p \in (0,1)$*. Using the relation* (4.2) *it is clear that the pgf of the marginal distribution of* $(Y_t)_{t\in\mathbb{Z}}$ *can be written as*

$$\mathrm{pgf}_Y(z) = \prod_{j=0}^{\infty} \left(1 - p\alpha^j + p\alpha^j z\right),$$

*the convergence of this infinite product can be shown as in Lemma 4.2.4 (ii). Now, we check the infinite divisibility of the marginal distribution by invoking Theorem 2.3.3. We calculate, noting that* $p > 0$*,*

$$\log \frac{\prod_{j=0}^{\infty} \left(1 - p\alpha^j + p\alpha^j z\right)}{\prod_{j=0}^{\infty} (1 - p\alpha^j)} = \sum_{j=0}^{\infty} \log\left(1 + \frac{p\alpha^j z}{1 - p\alpha^j}\right) =: H(z).$$

*Choosing* $p \leq 0.5$ *ensures* $\frac{p\alpha^j z}{1-p\alpha^j} \leq 1$ *so that the series expansion of the logarithm* $\log(1+x) = \sum_n (-1)^{n+1} x^n/n$ *is applicable. Thus* $H(z) = \sum_i a_i z^i$ *for some coefficients* $a_i$ *with radius of convergence at least* 1*, note that* $\log(1+x) \leq x$ *holds for* $x > -1$*. The strict positivity of the coefficients* $a_i$ *remains to be checked. It is clear that* $2a_2 = (\mathrm{d}^2/\mathrm{d}^2 z)H(0)$*, where the derivative is evaluated at* 0*. Denoting* $f_\alpha(z) = 1 + \frac{p\alpha^j z}{1-p\alpha^j}$*, we calculate*

$$\frac{\mathrm{d}^2}{\mathrm{d}^2 z} \sum_{j=0}^{\infty} \log\left(f_\alpha(z)\right) = \sum_{j=0}^{\infty} \frac{f_\alpha(z)\left(\frac{\mathrm{d}^2}{\mathrm{d}^2 z} f_\alpha(z)\right) - \left(\frac{\mathrm{d}}{\mathrm{d}z} f_\alpha(z)\right)^2}{f_\alpha^2(z)}.$$

*Evaluating this derivative at zero yields*

$$2a_2 = -\sum_{j=0}^{\infty} \left(\frac{p\alpha^j}{1-p\alpha^j}\right)^2 \in \left[-\frac{1}{(1-p)^2}\frac{p^2}{1-\alpha^2}, -\frac{p^2}{1-\alpha^2}\right].$$

*Hence,* $a_2 < 0$ *and, by Theorem 2.3.3, the marginal distribution of* $(Y_t)_{t\in\mathbb{Z}}$ *is* not *infinitely divisible.*

Similarly, we could arrive at the same result by invoking Proposition 2.3.6 (ii), which shows in particular that the cumulants of Compound Poisson distributions are nondecreasing. Hence, the ratio of the variance to the mean, $\kappa_{2,Y}/\kappa_{1,Y}$ of a Compound Poisson distributed random variable $Y$ is greater or equal to 1, i.e., Compound Poisson distributions are never underdispersed. This concept will be of great interest in Section 5.1. For now, it suffices to refer the reader to an example of an underdispersed INAR(1) process discussed in Weiß (2013) which gives another example for the assertion we just discussed.

With these examples, the relation between the class of infinitely divisible distributions and the class of marginal distributions of stationary INAR(1) processes is now better understood. Yet the question remains whether we can find a concise characterization of the intersection of these sets. In other words, what characterizes a distribution that is both infinitely divisible and also arises as the marginal distribution of an INAR(1) process? Are these distributions necessarily discrete self-decomposable (that this characteristic is sufficient is clear)? Once again, we find a counter-example to this conjecture.

**Example 8** (Marginal of INAR(1) and Infinitely Divisible but not DSD)**.** *Let* $(Y_t)_{t\in\mathbb{Z}}$ *be the stationary INAR(1) process of Example 5. It is clear that the stationary distribution of* $(Y_t)_{t\in\mathbb{Z}}$ *is infinitely divisible. For this distribution to be DSD, Theorem 2.3.5 necessitates that the sequence* $(n \cdot g_n)_{n\in\mathbb{N}}$ *be nondecreasing. However, this is only the case if* $h < (1+\alpha)/(3-\alpha)$ *holds, hence we can choose values for* $h, \alpha$ *such that the stationary distribution of* $(Y_t)_{t\in\mathbb{Z}}$ *is no longer DSD.*

# 5 INAR(1) Processes - Statistical Inference

Whereas Chapter 4 focuses on probabilistic results for INAR(1) processes, we now shift our attention towards the statistical inference for these processes. More precisely, in this chapter we assume to be given a data set similar to that of Figure 1.1 and we present statistical tools designed to deal with such data. First, let us consider the parameter estimation.

In contrast to the situation in Chapter 3, the dependency of an INAR(1) process on the past is entirely described by the single parameter $\alpha \in (0,1)$. Under the assumption that the arrival distribution is parametrized by a functional of its first few moments, it is clear from relation such as (4.6) that we can recover these moments from the marginal moments and an estimation of $\alpha$. Since the underlying structure of an INAR(1) process resembles that of an AR(1) process closely it is hardly surprising that classical estimation techniques from time series analysis (Conditional Least Squares, Yule-Walker etc.) may be applied with satisfying results. We will not discuss details here, an extensive discussion of several estimation approaches can be found in Al-Osh and Alzaid (1987).

The main question to address is the following one: Is the model we have chosen an appropriate one? How should we assess a question such as this? For instance, given the data $(y_1, \ldots, y_{96})$ as shown in Figure 1.1, Freeland (1998) suggested that the Poisson INAR(1) model is a good fit for this data. In view of the integer-valued data points and the behavior of the ACF and PACF, an INAR(1) structure seems appropriate. Yet the assumption of Poisson distributed innovations $\epsilon_t$ should be verified. In terms of the classical hypothesis testing theory, we are faced with the following scenario: the null hypothesis is given by

$$H_0 : (y_1, \ldots, y_{96}) \text{ stem from a Poisson INAR(1) process with } \alpha \in (0,1) \tag{5.1}$$

and the alternative hypothesis must needs be specified under additional considerations. Obviously, there are other possible hypotheses for testing in the framework of INAR(1) processes, a more general approach is presented in Chapter 6. The first section is based on the article Schweer and Weiß (2014), whereas the second part follows Schweer and Weiß (2015). The third section of this chapter contains unpublished material and has a slightly different focus, it considers the asymptotic distribution of the dependency functions introduced in Section 2.6. In this rather classical setting it is only fitting, yet still surprising, that the classical findings of Quenouille (1949) are actually reproducible in the case of the Poisson INAR(1) process.

## 5.1 Testing the Index of Dispersion

By Lemma 4.1.4, the marginal distribution of a stationary Poisson INAR(1) process $(Y_t)_{t\in\mathbb{Z}}$ is Poisson distributed. Amongst many other things, this implies for the first and second central moment of $Y_t$ that $\mathbb{E}[Y_t] := \mu_Y$ is equal to $\mathrm{Var}(Y_t) := \sigma_Y$ for all $t \in \mathbb{Z}$, in other words, the marginal distribution is equidispersed. Let us define the *empirical index of dispersion* similarly to its theoretical counterpart of (2.5) as follows:

$$I_Y = \frac{\sigma_Y^2}{\mu_Y}, \quad \text{thus} \quad \hat{I}_Y := \frac{S_Y^2}{\overline{Y}}, \tag{5.2}$$

where $S_Y^2 = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \overline{Y})^2 = \left(\frac{1}{T}\sum_{t=1}^{T} Y_t^2\right) - \overline{Y}^2$, the empirical variance. For a Poisson INAR(1) process $(Y_t)_{t\in\mathbb{Z}}$ it thus holds that $I_Y = 1$. It turns out, however, that many if not most real data sets consisting of count data show a large amount of empirical overdispersion, i.e., $\hat{I}_Y > 1$.

In the literature, several theories have been put forward in order to account for this behavior. One explanation that is often used is the presence of positive correlation between the monitored events, this would correspond to an additional dependency in time of the $\epsilon_t$ in Definition 1.1.1. Other explanations focus on the variation in the probability $\alpha$, i.e., suggest replacing the fixed parameter $\alpha$ with a time-varying parameter $\alpha_t$. In this thesis (and in the article Schweer and Weiß (2014)) a variation of the first explanation is put forward, we refer the reader to Section 5.1.3.

Now, let us see whether Example 1.1 also fits the pattern just described. We calculate $\overline{y} \approx 8.604$ and $S_y^2 \approx 11.24$, so that $\hat{I}_y \approx 1.306$ and the data $y_1, \ldots, y_{96}$ are indeed overdispersed. We have about 31% of empirical overdispersion and the question arises whether this is indicative of a violation of the equidispersion property of the Poisson distribution. In terms of hypothesis testing, this means that we would like to test the null hypothesis (5.1) against the alternative hypothesis

$$H_1 : (y_1, \ldots, y_{96}) \text{ stem from an overdispersed INAR(1) process with } \alpha \in (0,1). \quad (5.3)$$

The approach we employ is that of developing (asymptotic) distributional theory for the estimator $\hat{I}_y$ and using these results for statistical inference on the data given Figure 1.1. We remark that in the case of i.i.d. counts, the index of dispersion has been analyzed in detail in Rao and Chakravarti (1956) as well as in Böhning (1994).

### 5.1.1 Asymptotic Distribution of Index of Dispersion

The main result of this section, Theorem 5.1.2, makes use of the moment formulae derived in Theorem 4.1.3, which hold for any stationary INAR(1) process (with existing moments). Therefore, also Theorem 5.1.2 itself is formulated in a rather general way. However, the essential mixing condition (2.13), which is required for Theorem 5.1.2 to hold, is satisfied particularly for a CPINAR(1) process as in Theorem 4.2.6. It should be pointed out that the much more general Theorem 6.1.2 provides a sufficient condition for all INAR(1) processes with a finite mean and $\mathbb{P}(\epsilon_0 = 0) \in (0,1)$. In the following preliminary result, we consider moment expressions which will be important for our asymptotic results.

**Lemma 5.1.1** (Schweer and Weiß (2014), Lemma A.5.1)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be an INAR(1) process with $\mathbb{P}(\epsilon_0 = 0) \in (0,1)$, $\mu_{\epsilon,4} > \infty$ and let $\mathbf{X}_t := \left(Y_t - \mu_Y, Y_t^2 - \mu_Y^2 - \sigma_Y^2\right)$. Then the series*

$$\sigma_{ij} = \mathbb{E}\left[X_{0,i}X_{0,j}\right] + \sum_{k=1}^{\infty}\left(\mathbb{E}[X_{0,i}X_{k,j}] + \mathbb{E}[X_{k,i}X_{0,j}]\right)$$

*converge absolutely for any $1 \le i, j \le 2$ where $X_{k,i}$ denotes the $i$-th entry of $\mathbf{X}_k$.*

*Furthermore, the following expressions hold:*

$$\sigma_{11} = \sigma_Y^2 \frac{1+\alpha}{1-\alpha},$$
$$\sigma_{22} = \left(\bar{\mu}_{Y,4} - (1-2\mu_Y)\bar{\mu}_{Y,3} - 2\mu_Y\sigma_Y^2 - \sigma_Y^4\right)\frac{1+\alpha^2}{1-\alpha^2} + (1+2\mu_Y)(\bar{\mu}_{Y,3} + 2\mu_Y\sigma_Y^2)\frac{1+\alpha}{1-\alpha},$$
$$\sigma_{12} = \sigma_{21} = \frac{1}{2}(\bar{\mu}_{Y,3} - \sigma_Y^2)\frac{1+\alpha^2}{1-\alpha^2} + \frac{1}{2}\left(\bar{\mu}_{Y,3} + \sigma_Y^2(1+4\mu_Y)\right)\frac{1+\alpha}{1-\alpha}.$$

*Proof.* Using that $\gamma(k) = \sigma_Y^2\alpha^k$, we obtain that

$$\sigma_{11} = \mathbb{E}\left[X_{0,1}^2\right] + 2\sum_{k=1}^{\infty}\mathbb{E}\left[X_{0,1}X_{k,1}\right] = \sigma_Y^2 + 2\sum_{k=1}^{\infty}\gamma(k) \;= \sigma_Y^2\frac{1+\alpha}{1-\alpha}.$$

Next, we have using results and notation of Theorem 4.1.3

$$\begin{aligned}
\mathbb{E}\left[X_{0,2}X_{k,2}\right] &= \mathbb{E}\left[(Y_0^2 - \mu_Y^2 - \sigma_Y^2)(Y_k^2 - \mu_Y^2 - \sigma_Y^2)\right] = \mu(0,k,k) - (\mu_Y^2 + \sigma_Y^2)^2 \\
&= \alpha^{2k}\left(\bar{\mu}_{Y,4} - 3\bar{\mu}_{Y,3} + \sigma_Y^2(2 - 3\sigma_Y^2)\right) + \mu_Y^4 + (1+\mu_Y)^2\sigma_Y^2\alpha^k \\
&\quad + (\bar{\mu}_{Y,3} - \sigma_Y^2)\Big((2+\mu_Y)\alpha^{2k} + (1+\mu_Y)\alpha^k + \mu_Y(\alpha^{2k} + \alpha^k)\Big) \\
&\quad + \mu_Y^2\sigma_Y^2(\alpha^k + 2) + \sigma_Y^4(1 + 2\alpha^{2k}) - (\mu_Y^2 + \sigma_Y^2)^2 + 2\mu_Y(1+\mu_Y)\sigma_Y^2\alpha^k \\
&= \alpha^{2k}\Big(\bar{\mu}_{Y,4} - 3\bar{\mu}_{Y,3} + \sigma_Y^2(2 - 3\sigma_Y^2) + 2(1+\mu_Y)(\bar{\mu}_{Y,3} - \sigma_Y^2) + 2\sigma_Y^4\Big) \\
&\quad + \alpha^k\left((\bar{\mu}_{Y,3} - \sigma_Y^2)(1+2\mu_Y) + (1+\mu_Y)^2\sigma_Y^2 + 2\mu_Y(1+\mu_Y)\sigma_Y^2 + \mu_Y^2\sigma_Y^2\right) \\
&\quad + \sigma_Y^4 + 2\sigma_Y^2\mu_Y^2 + \mu_Y^4 - (\mu_Y^2 + \sigma_Y^2)^2 \\
&= \alpha^{2k}\left(\bar{\mu}_{Y,4} - (1-2\mu_Y)\bar{\mu}_{Y,3} - 2\mu_Y\sigma_Y^2 - \sigma_Y^4\right) + \alpha^k(1+2\mu_Y)(\bar{\mu}_{Y,3} + 2\mu_Y\sigma_Y^2),
\end{aligned}$$

which implies the expression for $\sigma_{22} = \mathbb{E}[X_{0,2}^2] + 2\sum_{k=1}^{\infty}\mathbb{E}[X_{0,2}X_{k,2}]$. We further calculate

$$\begin{aligned}
\mathbb{E}\left[X_{0,1}X_{k,2}\right] &= \mathbb{E}\left[(Y_0 - \mu_Y)(Y_k^2 - \mu_Y^2 - \sigma_Y^2)\right] = \mu(k,k) - \mu_Y(\mu_Y^2 + \sigma_Y^2) \\
&= (\bar{\mu}_{Y,3} - \sigma_Y^2)\alpha^{2k} + (1+\mu_Y)\sigma_Y^2\alpha^k \; + \mu_Y\sigma_Y^2(1+\alpha^k) + \mu_Y^3 - \mu_Y(\mu_Y^2 + \sigma_Y^2) \\
&= \alpha^{2k}(\bar{\mu}_{Y,3} - \sigma_Y^2) + \alpha^k\sigma_Y^2(1+2\mu_Y),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left[X_{0,2}X_{k,1}\right] &= \mathbb{E}\left[(Y_0^2 - \mu_Y^2 - \sigma_Y^2)(Y_k - \mu_Y)\right] = \mu(0,k) - \mu_Y(\mu_Y^2 + \sigma_Y^2) \\
&= (\bar{\mu}_{Y,3} - \sigma_Y^2)\alpha^k + (1+\mu_Y)\sigma_Y^2\alpha^k \; + \mu_Y\sigma_Y^2(\alpha^k + 1) + \mu_Y^3 \; - \; \mu_Y(\mu_Y^2 + \sigma_Y^2) \\
&= \alpha^k(\bar{\mu}_{Y,3} + 2\mu_Y\sigma_Y^2).
\end{aligned}$$

With this, we find

$$\sigma_{12} = \bar{\mu}_{Y,3} + 2\mu_Y\sigma_Y^2 \; + \sum_{k=1}^{\infty}\Big(\alpha^{2k}\left(\bar{\mu}_{Y,3} - \sigma_Y^2\right) + \alpha^k\left(\bar{\mu}_{Y,3} + \sigma_Y^2(1+4\mu_Y)\right)\Big),$$

which completes the proof. □

With the results of the previous Lemma, we are now able to prove the main result of this section.

**Theorem 5.1.2** (Schweer and Weiß (2014), Theorem 4.1.1). *Let* $(Y_t)_{t\in\mathbb{Z}}$ *be a INAR(1) process, which is also* $\alpha$*-mixing with weights* $\alpha_Y(n)$. *Let* $\mathbb{E}[Y_t^{2(2+\delta)}] < \infty$ *for some* $\delta > 0$, *and let the mixing condition* (2.13) *hold. Then*

$$\sqrt{T}(\hat{I}_Y - I_Y) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2\right) \quad \text{as } T \longrightarrow \infty,$$

*where*

$$\sigma^2 = \frac{1+\alpha}{1-\alpha}(\mu_Y - \sigma_Y^2)\left(\frac{\bar{\mu}_{Y,3}}{\mu_Y^3} - \frac{\sigma_Y^4}{\mu_Y^4}\right) + \frac{1+\alpha^2}{1-\alpha^2}\left(\frac{\bar{\mu}_{Y,4}}{\mu_Y^2} - \frac{\bar{\mu}_{Y,3}}{\mu_Y^3}(\mu_Y + \sigma_Y^2) + \frac{\sigma_Y^4}{\mu_Y^3}(1-\mu_Y)\right).$$

*Proof.* As in the proof of Lemma 5.1.1, we consider the vector-valued process defined by $\mathbf{X}_t := \left(Y_t - \mu_Y, Y_t^2 - \mu_Y^2 - \sigma_Y^2\right)$ which obviously satisfies $\mathbb{E}[\mathbf{X}_t] = \mathbf{0}$. The conditions required by Theorem 5.1.2 are chosen such that Theorem 2.5.2 is applicable, see the discussion at the end of Section 4.2.3 (note that $\mathbb{E}\big[Y_t^{2(2+\delta)}\big] < \infty$ implies finiteness of $\mathbb{E}\big[|Y_t^2 - \mu_Y^2 - \sigma_Y^2|^{2+\delta}\big]$). Using this result together with the Cramér-Wold device, we conclude that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{X}_t \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right) \quad \text{with } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix},$$

where the $\sigma_{ij}$ are given as in Lemma 5.1.1. In a final step, we can apply the Delta theorem to derive the asymptotic behavior of $\hat{I}_Y$ from formula (5.2). For this purpose, we introduce the function

$$g : \mathbb{R}^2 \to \mathbb{R}, \quad g(y_1, y_2) := \frac{y_2}{y_1} - y_1.$$

For this function, we find $g\left(\bar{Y}, \frac{1}{T}\sum_{t=1}^{T} Y_t^2\right) = \hat{I}_Y$ as well as $g(\mu_Y, \mu_Y^2 + \sigma_Y^2) = I_Y$. Furthermore,

$$\mathbf{D} := \operatorname{grad} g(\mu_Y, \mu_Y^2 + \sigma_Y^2) = \left(-\frac{\sigma_Y^2}{\mu_Y^2} - 2, \ \frac{1}{\mu_Y}\right).$$

We calculate for $\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^\top$

$$\begin{aligned}
&= \frac{1+\alpha}{1-\alpha}\left[\frac{(1+2\mu_Y)(\bar{\mu}_{Y,3} + 2\mu_Y\sigma_Y^2)}{\mu_Y^2} + \sigma_Y^2\left(\frac{\sigma_Y^2}{\mu_Y^2} + 2\right)^2 \right.\\
&\qquad\qquad \left. - \frac{\bar{\mu}_{Y,3} + \sigma_Y^2(1+4\mu_Y)}{\mu_Y}\left(\frac{\sigma_Y^2}{\mu_Y^2} + 2\right)\right]\\
&\quad + \frac{1+\alpha^2}{1-\alpha^2}\left(\frac{1}{\mu_Y^2}\left(\bar{\mu}_{Y,4} - (1-2\mu_Y)\bar{\mu}_{Y,3} - 2\mu_Y\sigma_Y^2 - \sigma_Y^4\right) - \frac{\bar{\mu}_{Y,3} - \sigma_Y^2}{\mu_Y}\left(\frac{\sigma_Y^2}{\mu_Y^2} + 2\right)\right)\\
&= \frac{1+\alpha}{1-\alpha}(\mu_Y - \sigma_Y^2)\left(\frac{\bar{\mu}_{Y,3}}{\mu_Y^3} - \frac{\sigma_Y^4}{\mu_Y^4}\right) + \frac{1+\alpha^2}{1-\alpha^2}\left(\frac{\bar{\mu}_{Y,4}}{\mu_Y^2} - \frac{\bar{\mu}_{Y,3}}{\mu_Y^3}(\mu_Y + \sigma_Y^2) + \frac{\sigma_Y^4}{\mu_Y^3}(1-\mu_Y)\right).
\end{aligned}$$

Application of the Delta theorem completes the proof. □

Since a Poisson INAR(1) process has existing moments up to any order, and since it satisfies the mixing condition (2.13), see Section 4.2.3, we can apply Theorem 5.1.2 in this case. Remembering that for $Y \sim \text{Poi}(\lambda)$ we have $\lambda = \sigma_Y^2 = \bar{\mu}_{Y,3} = \bar{\mu}_{Y,4} - 3\sigma_Y^4$, the following Corollary 5.1.3 is an immediate consequence.

**Corollary 5.1.3** (Schweer and Weiß (2014), Corollary 4.1.2)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process with $\epsilon_t \sim \text{Poi}(\lambda)$ for all $t \in \mathbb{Z}$. Then*

$$\sqrt{T}(\hat{I}_Y - 1) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ 2\frac{1+\alpha^2}{1-\alpha^2}\right) \quad \text{as } T \longrightarrow \infty.$$

Theorem 5.1.2 shows that $\hat{I}_Y$ is an asymptotically unbiased estimator of $I_Y$. If computed from a time series of finite length $T$, however, we expect this estimator to be negatively biased. Our conjecture relies on the well-known fact that $S_Y^2$ is generally negatively biased David (1985): $\mathbb{E}[S_Y^2] = \sigma_Y^2 - \text{Var}[\bar{Y}]$. For AR(1)-like models, we compute

$$\text{Var}[\bar{Y}] = \frac{\sigma_Y^2}{T} + \frac{2\sigma_Y^2}{T}\frac{\alpha}{1-\alpha}\left(1 - \frac{1}{T}\frac{1-\alpha^T}{1-\alpha}\right).$$

This increases in $\alpha$ if $T > 1$, so we expect $\hat{I}_Y$ to be visibly biased for large $\alpha$ and small $T$. In fact, we can give a precise result on the bias of $\hat{I}_Y$:

**Proposition 5.1.4** (Weiß and Schweer (2015), Remark A.3.1)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process with $\epsilon_t \sim \text{Poi}(\lambda)$ for all $t \in \mathbb{Z}$. Then*

$$\mathbb{E}\left[\hat{I}_Y\right] \approx 1 - \frac{1}{T}\frac{1+\alpha}{1-\alpha}.$$

*Proof.* The bias correction is obtained based on the second-order Taylor expansion of the function $g$ from the proof of Theorem 5.1.2. We obtain the derivatives

$$\frac{\partial^2}{\partial x_1^2}g(x_1,x_2) = \frac{2x_2}{x_1^3}, \quad \frac{\partial^2}{\partial x_1 \partial x_2}g(x_1,x_2) = -\frac{1}{x_1^2}, \quad \frac{\partial^2}{\partial x_2^2}g(x_1,x_2) = 0.$$

So the Hessian of $g$, evaluated in $(\mu_Y, \mu_Y^2 + \sigma_Y^2) = (\mu_Y, \mu_Y + \mu_Y^2)$, is given by

$$\mathbf{H}_g(\mu_Y, \mu_Y + \mu_Y^2) = \begin{pmatrix} \frac{2(1+\mu_Y)}{\mu_Y^2} & -\frac{1}{\mu_Y^2} \\ -\frac{1}{\mu_Y^2} & 0 \end{pmatrix}.$$

Hence, the function $g$ is twice differentiable and Taylor's theorem may be applied. Taking expectations, this implies

$$\begin{aligned} T\mathbb{E}\left[\hat{I}_Y - 1\right] &\approx \mathbb{E}\left[\frac{1}{2}\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\mathbf{X}_t\right)\mathbf{H}_g(\mu_Y, \mu_Y + \mu_Y^2)\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\mathbf{X}_t\right)^\top\right] \\ &\approx \frac{1}{2}\left(\sigma_{11}h_{11} + 2h_{12}\sigma_{12} + h_{22}\sigma_{22}\right) = \frac{1+\alpha}{1-\alpha}\left[\frac{1+\mu_Y}{\mu_Y} - \frac{1+2\mu_Y}{\mu_Y}\right] = -\frac{1+\alpha}{1-\alpha}, \end{aligned}$$

concluding the proof. □

The result of Corollary 5.1.3 is important if we want to test the null hypothesis (5.1) against the alternative (5.3) (as it would be the case for a CPINAR(1) model with $\nu \geq 2$). Based on the asymptotic result in Corollary 5.1.3, we will reject $H_0$ on significance level $\beta$ if the observed value of the index of dispersion, $\hat{I}_y$, exceeds the critical value

$$1 + z_{1-\beta}\sqrt{\frac{2}{T}\frac{1+\alpha^2}{1-\alpha^2}}, \tag{5.4}$$

where $z_{1-\beta}$ denotes the $(1-\beta)$-quantile of the $\mathcal{N}(0,1)$-distribution. Alternatively, we can check if the p-value,

$$1 - \Phi\left(\sqrt{\frac{T}{2}\frac{1-\alpha^2}{1+\alpha^2}}(\hat{I}_y - 1)\right), \tag{5.5}$$

falls below $\beta$, where $\Phi$ denotes the distribution function of the $\mathcal{N}(0,1)$-distribution. If a hypothetical value for the dependence parameter $\alpha$ is not available, we recommend to use a plug-in approach, i.e., to replace $\alpha$ by $\widehat{\rho}_y(1)$.

### 5.1.2 Power Analysis

Let us now exemplify how Theorem 5.1.2 can be applied to analyze the power of the test for overdispersion described above with regard to the alternative of a certain type of CPINAR(1) model with $\nu \geq 2$. As the alternative model, we consider a CPINAR(1) model with negative binomial innovations (Example 1). This three-parameter model is particularly well-suited for a theoretical power analysis, since 1) it allows us to control the marginal mean $\mu_Y$, the true index of dispersion $I_Y$ and the autocorrelation level $\alpha$ separately from each other (remember the properties in (4.5)); 2) it allows for the full range of $(\mu_Y, I_Y, \alpha)$; and 3) it includes the null model at least as a boundary case. For the case of $\text{Poi}_\nu$-distributed innovations with $\nu \geq 2$ (Example 1), in contrast, the range of $(\mu_Y, I_Y, \alpha)$ is restricted by the relation between $I_Y$ and $\alpha$ given in (4.5) as well as $I_\epsilon = (2\nu+1)/3$. Therefore, a further analysis of the $\text{Poi}_\nu$-INAR(1) model is postponed to the application presented in Section 5.1.3.

To be able to evaluate the expressions in Theorem 5.1.2, we use formula (2.9) for the central moments of the Negative Binomial innovations (and later formula (2.8) for the case of the $\text{Poi}_\nu$-innovations), formula (4.7) to switch between central and raw moments, and formula (4.6) to obtain the observations' moments from the innovations' moments. The graphs in Figure 5.1 show the asymptotic approximations of selected power functions computed in this way, the critical value (5.4) assumes the significance level $\beta = 0.05$.

Figure 5.1 (a) shows, as expected, that the power of the dispersion test becomes better if $T$ is increased. It is, however, interesting to see that also in the Negative Binomial case, the actual mean $\mu_Y$ is nearly without effect on the power of the test (Figure 5.1 (b), for the Poisson case, an analogous property is known from Corollary 5.1.3). In contrast, see Figure 5.1 (c), increasing the autocorrelation level $\alpha$ of the underlying INAR(1) process leads to a clear deterioration of the power, which can only be compensated by increasing the length $T$ of the time series. As an example, see Figure 5.1 (d), if $\alpha = 0.8$, we have

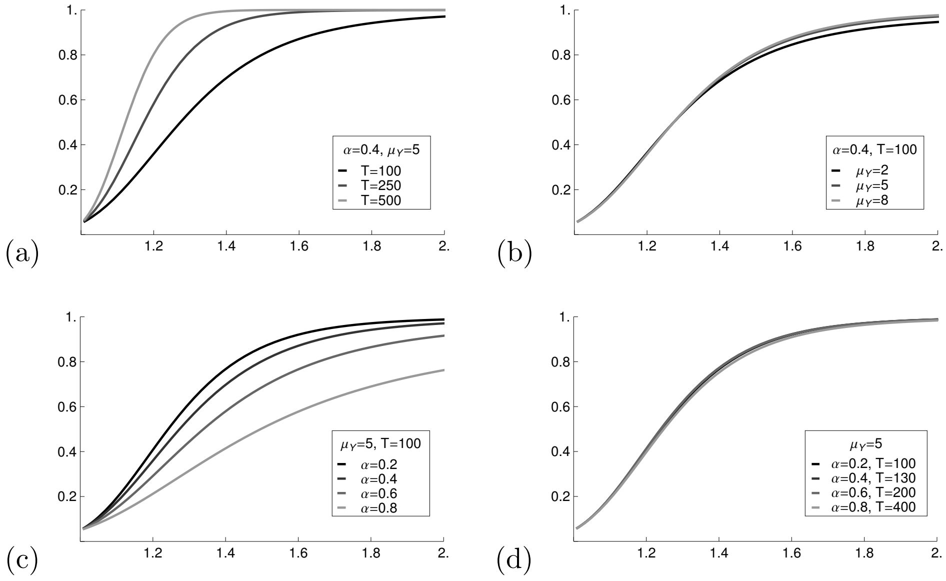


Figure 5.1: Asymptotic power against $I_Y$ for NegBin-INAR(1) models.

to choose $T = 400$ to reach the same power as in the case $(\alpha, T) = (0.2, 100)$, i.e., the sample size has to be quadrupled. This deterioration of the power as $\alpha$ increases appears plausible if we look at the asymptotic variance in Corollary 5.1.3: Under the null, we have more and more noise if $\alpha$ increases, which, in turn, makes our test less sensitive towards overdispersion.

### 5.1.3 Application: A Time Series of Claims Counts

In this section, we consider the concrete application of the test (5.4) to the real data set of Figure 1.1, and also the finite-sample performance of the test is considered in this context. Let us begin by applying the test for overdispersion, recalling that the empirical mean and variance are given by $\bar{y} \approx 8.604$ and $s_y^2 \approx 11.24$, respectively, so the data are empirically overdispersed with $\hat{I}_y \approx 1.306$. About 31 % of empirical overdispersion seems quite high, but is this already a *significant* violation of the equidispersion property of the Poisson distribution?

We apply the test for overdispersion described in Section 5.1.1 with significance level $\beta = 0.05$. Plugging-in $\widehat{\rho}_y(1) \approx 0.452$ instead of $\alpha$, we compute the critical value according to formula (5.4) as 1.292, i.e., the observed value $\hat{I}_y \approx 1.306$ leads to a rejection of the null hypothesis of a Poisson INAR(1) model. However, in view of having observed about 31 % of empirical overdispersion, it might be surprising that this was a quite narrow decision, also the p-value of about 0.0424 according to formula (5.5) is only slightly below 0.05. But remembering that the asymptotic variance in Corollary 5.1.3 is a strictly increasing function in $\alpha$, also see the discussion of Figures 5.1 (c,d) above, it is clear that the sensitivity of the test becomes worse for increasing $\alpha$. So dispersion ratios computed

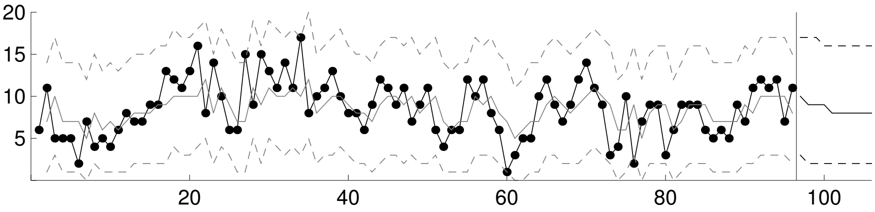


Figure 5.2: Plot of claim counts from Figure 1.1, with 1-step-ahead forecasts (median and 95 % interval; left part) and $k$-step-ahead forecasts ($k = 1, \dots, 10$; median and 95 % interval; right part), respectively.

from *dependent* data $y_1, \dots, y_T$ have to be interpreted with caution, especially if $T$ is not particularly large.

Weiß (2009) fitted the so-called INARCH(1) model to the data, which is an AR(1)-like model with overdispersion. In addition, we now also consider the CPINAR(1) model with the innovations being NegBin$(n, \pi)$-distributed (i.e., infinite compounding structure) as well as being $\text{Poi}_\nu(\lambda)$-distributed (i. e., finite compounding structure with uniform compounding distribution), see Example 1. All models are fitted with a conditional maximum likelihood (CML) approach to the data $y_T, \dots, y_2$ given $y_1$. Estimates and approximate standard errors are computed by using R's `optim`, which is initialized with the moment estimates for the respective parameters: $\hat{\mu}_\epsilon$ and $\hat{\sigma}_\epsilon^2$ are obtained from $\bar{y}, s_y^2$ by using relation (4.5), then the moment formulae (2.9) and (2.8) are applied. The probabilities required for the $\text{Poi}_\nu(\lambda)$-innovations (which, in turn, are necessary for the transition probabilities (4.1)) are computed by using the recursive scheme in Proposition 2.3.6 (i). Results are summarized in Table 5.1.

| Model | Par. 1 | Par. 2 | Par. 3 | AIC | BIC |
|---|---|---|---|---|---|
| Poi-INAR(1) | 0.396 | 5.232 | | 485.4 | 490.5 |
| $(\alpha, \lambda)$ | (0.068) | (0.612) | | | |
| INARCH(1) | 0.483 | 4.486 | | 482.0 | 487.2 |
| $(\alpha, \beta)$ | (0.090) | (0.777) | | | |
| NegBin-INAR(1) | 0.426 | 14.785 | 0.748 | 485.3 | 493.0 |
| $(\alpha, n, \pi)$ | (0.075) | (13.160) | (0.158) | | |
| $\text{Poi}_2$-INAR(1) | 0.466 | 3.092 | | 483.2 | 488.3 |
| $(\alpha, \lambda)$ | (0.069) | (0.428) | | | |
| $\text{Poi}_3$-INAR(1) | 0.524 | 2.078 | | 487.7 | 492.8 |
| $(\alpha, \lambda)$ | (0.062) | (0.305) | | | |

Table 5.1: Claim counts from Figure 1.1: CML estimates for diverse models.

Table 5.1 shows that the NegBin-INAR(1) model only leads to a slight improvement in the AIC compared to the Poi-INAR(1) model, but its BIC is clearly worse and, in particular, the standard error for the parameter $n$ is rather large. So we conclude that this model is not appropriate for the data. The $\text{Poi}_2$-INAR(1) model, in contrast, leads to a clear improvement in both AIC and BIC, while orders $\nu \geq 3$ are not appropriate for the data. So among the considered CPINAR(1) models, the $\text{Poi}_2$-INAR(1) model is the best choice. It has to be mentioned that the INARCH(1) model as proposed by Weiß (2009) leads to further improved values of AIC and BIC, but this model has the practical disadvantages that neither expressions for the stationary marginal distribution nor for the $k$-step-ahead conditional distributions with $k \geq 2$ are known. For the $\text{Poi}_\nu$-INAR(1) models with their finite compounding structure, in contrast, these distributions are easily computed according to Example 5 and Theorem 4.2.3, respectively (see the next paragraph for further details). In addition, its model parameters have an intuitive interpretation. The estimate for $\alpha$ can be understood as the rate of claimants in a month $t$ that continue collecting benefits also in month $t+1$. The fitted $\text{Poi}_\nu$-model for the innovations might be interpreted in view of Definition 2.3.1 as follows: $N$ describes the number of accidents (having the estimated mean $\hat{\lambda}$), and the distribution of the number of persons being injured per accident is approximated by a uniform distribution on $\{1, \ldots, \nu\}$.

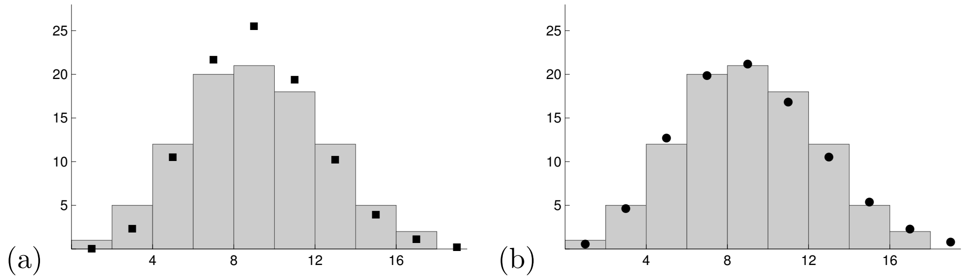


Figure 5.3: Claim counts from Figure 1.1: Histogram with marginal distribution from fitted Poi- and $\text{Poi}_2$-INAR(1) model, respectively.

Let us illustrate the above-mentioned practical advantages of the $\text{Poi}_2$-INAR(1) model. Figure 5.3 shows the marginal distributions of the CML-fitted Poi- and $\text{Poi}_2$-INAR(1) model, respectively (the latter being computed according to the Scheme of Example 5), and compares them to a histogram of the data. Obviously, the marginal distribution of the $\text{Poi}_2$-INAR(1) model leads to a much better fit to the overdispersed data. We also used the CML-fitted $\text{Poi}_2$-INAR(1) model for forecasting, by evaluating the respective conditional pgf from Theorem 4.2.3 via numerical series expansion. From these forecast distributions, we computed the median (solid lines in Figure 5.2) as well as the limits of a 95 % prediction interval (dashed lines) at each point in time, thus guaranteeing coherent forecasts, cf. Section 2 in Freeland (2010). In the left part of Figure 5.2, the 1-step-ahead forecasts at each time $t$ are shown, being obtained by conditioning on $y_1, \ldots, y_{T-1}$. In most cases, the actual observations are rather close to the predicted median value, and none of the observations is beyond the interval limits, which again indicates the goodness of our CML-fitted $\text{Poi}_2$-INAR(1) model. In the right part of

Figure 5.2, each prediction is conditioned on the last observation in our time series, $y_T$, while we increased the prediction horizon from $k = 1$ to 10. As expected due to the geometrical ergodicity, cf. (4.21), the $k$-step-ahead forecast distribution converges rather quickly to the stationary marginal distribution (as plotted in Figure 5.3 (b)) such that the forecasts remain constant for $k \geq 5$.

| $T$ | mean | s.d. | s.d.$_\mathrm{a}$ | skew. | $\hat{q}_{0.05}$ | $q_{0.05,\mathrm{a}}$ | $\hat{q}_{0.95}$ | $q_{0.95,\mathrm{a}}$ | r.r.$_\alpha$ | r.r.$_{\hat{\alpha}}$ | r.r.$_\mathrm{a}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 0.977 | 0.162 | 0.166 | 0.421 | 0.731 | 0.727 | 1.262 | 1.273 | 0.044 | 0.038 | 0.050 |
| 250 | 0.991 | 0.104 | 0.105 | 0.265 | 0.828 | 0.827 | 1.169 | 1.173 | 0.047 | 0.042 | 0.050 |
| 500 | 0.995 | 0.074 | 0.074 | 0.191 | 0.878 | 0.878 | 1.120 | 1.122 | 0.047 | 0.043 | 0.050 |
| 1000 | 0.998 | 0.052 | 0.053 | 0.143 | 0.914 | 0.914 | 1.086 | 1.086 | 0.049 | 0.047 | 0.050 |

Table 5.2: Simulated Poi-INAR(1) with $(\alpha, \lambda) = (0.40, 5.1)$, i.e., $\mu_Y = 8.5$ and $I_Y = 1$, and significance level $\beta = 0.05$ (100,000 repl.): empirical and asymptotic properties of $\hat{I}_Y$; empirical rejection rates.

Finally, we present some results from a simulation study to investigate the goodness of our asymptotic approximations according to Theorem 5.1.2 and Corollary 5.1.3. The choice of the shown Poi- and Poi$_2$-INAR(1) models is motivated by the fitted models according to Table 5.1, and the model parameters are chosen such that all models have the same marginal mean, $\mu_Y = 8.5$. Columns 2–9 of Tables 5.2 and 5.3 consider stochastic properties of $\hat{I}_Y$ and show both the empirically observed results as well as the values obtained from the asymptotic approximation. Columns $\geq 10$ show the rates of rejecting the null hypothesis of a Poi-INAR(1) with parameters $(\alpha, \lambda)$, where the critical value according to (5.4) was computed either with the true $\alpha$ or by plugging-in $\hat{\alpha} := \widehat{\rho}_y(1)$. The rejection rate being expected from the respective asymptotic approximation ("r.r.$_\mathrm{a}$") equals the significance level 0.05 for the Poi-INAR(1) model in Table 5.2, while it equals the asymptotic power in the case of Table 5.3, also see the power graphs in Figure 5.4.

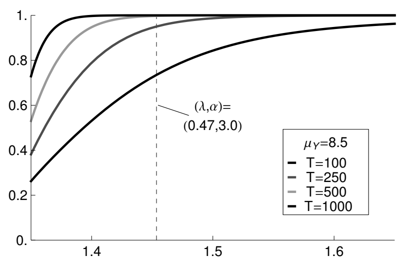


Figure 5.4: Asymptotic power against $I_Y$ for Poi$_2$-INAR(1) model (dashed line: model from Table 5.3).

The stochastic properties considered in columns 2–9 of Tables 5.2 and 5.3 show that the asymptotic normal approximation gives values being close to the empirically observed ones, at least for $T \geq 250$. In accordance with the result of Propostition 5.1.4, we notice a (moderate) negative bias especially for $T = 100$. The goodness of the normal approximation is also confirmed by the respective quantile plots (not shown).

| $T$ | mean | s.d. | s.d.$_{\text{a}}$ | skew. | $\hat{q}_{0.05}$ | $q_{0.05,\text{a}}$ | $\hat{q}_{0.95}$ | $q_{0.95,\text{a}}$ | r.r.$_{\alpha}$ | r.r.$_{\hat{\alpha}}$ | r.r.$_{\text{a}}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 1.413 | 0.251 | 0.258 | 0.532 | 1.041 | 1.029 | 1.855 | 1.879 | 0.665 | 0.683 | 0.735 |
| 250 | 1.436 | 0.161 | 0.163 | 0.327 | 1.187 | 1.185 | 1.714 | 1.722 | 0.952 | 0.959 | 0.950 |
| 500 | 1.445 | 0.115 | 0.116 | 0.237 | 1.264 | 1.263 | 1.643 | 1.644 | 0.999 | 0.999 | 0.997 |
| 1000 | 1.449 | 0.082 | 0.082 | 0.170 | 1.318 | 1.319 | 1.588 | 1.588 | 1.000 | 1.000 | 1.000 |

Table 5.3: Simulated $\text{Poi}_2$-INAR(1) with $(\alpha, \lambda) = (0.47, 3.0)$, i.e., $\mu_Y = 8.5$ and $I_Y = 1.454$, and signif. level $\beta = 0.05$ (100,000 repl.): emp. and asympt. properties of $\hat{I}_Y$; emp. and asympt. rejection rates.

The empirical rejection rates in Table 5.2 (*false* rejections) are even below the chosen significance level of $\beta = 0.05$. The empirical rejection rates in Table 5.3 express the power of the dispersion test, and these values are close to the ones expected from the asymptotic approximation. The latter is displayed in Figure 5.4, where increasing values for $I_Y$ are obtained through decreasing the autocorrelation parameter $\alpha$ according to formula (4.5) (remember the discussion in Section 5.1.2). In a nutshell, our asymptotic results lead to reasonable approximations also for finite $T$.

## 5.2 Testing Time-Reversibility in INAR(1) processes Via Moments

In the previous section, we developed a statistical test for the hypothesis (5.1) based on the characteristic that a Poisson INAR(1) process has a Poisson marginal distribution. This characteristic is not the only unique feature of a Poisson INAR(1) process, since by Theorem 4.1.8 these processes are the only time-reversible INAR(1) processes. In Section 2.4 we presented one possibility to derive a test to check a given time series for time-reversibility. This approach is discussed in detail here, including the explicit calculation of the asymptotic distributions and the consideration of bias correction. We will also present a second approach, which evaluates the marginal skewness of the observations $y_1, \ldots, y_T$, motivated in part by the same idea as the test based on (5.2) in Section 5.1.

### 5.2.1 Testing via Generalized Autocovariances

Let us now describe the first approach in detail. Using the moment formulae from Theorem 4.1.3, we can explicitly calculate the values of the generalized autocovariance function $\beta(\cdot)$ of (2.10) in terms of the moments of $(Y_t)_{t\in\mathbb{Z}}$. In particular,

$$\beta(k) = \alpha^k(1-\alpha^k)\left(\bar{\mu}_{Y,3} - \sigma_Y^2\right) \quad \text{for } k \in \mathbb{N}_0. \tag{5.6}$$

Note that for the *Poisson* INAR(1) model, we have $\bar{\mu}_{Y,3} = \sigma_Y^2$ so that $\beta(k) \equiv 0$, corresponding to Theorem 4.1.8. Since $\alpha^k \to 0$ for increasing $k$, it seems advisable to check whether $\beta(1) = 0$ or not in the proposed test.

Asymptotic characteristics of the estimator $\hat{\beta}_T(k)$ as defined in (2.11) can be inferred from two results. Firstly, all moments of $(Y_t)_{t\in\mathbb{Z}}$ exist because the marginal distribution is Poisson-distributed. Secondly, a Poisson INAR(1) process is geometrically ergodic and $\alpha$-mixing with exponentially decreasing weights by Theorem 4.2.6. With Birkhoff's ergodic theorem, it follows that $\hat{\beta}_T(k)$ is a consistent estimator of $\beta(k)$, and the classical CLT of Theorem 2.5.2 leads to the following result.

**Theorem 5.2.1** (Schweer and Weiß (2015), Theorem 3.1.1)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process, with $\epsilon_t \sim \mathrm{Poi}(\lambda)$ for all $t \in \mathbb{Z}$, let $\mu_Y = \lambda/(1-\alpha)$ and let $k \in \mathbb{N}$. Then*

$$\sqrt{T-k} \cdot \hat{\beta}_T(k) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_k^2\right),$$

*where*

$$\sigma_k^2 = \mathbb{E}\left[\left(Y_0^2 Y_{-k} - Y_0 Y_{-k}^2\right)^2\right] + 2\sum_{t=1}^{\infty} \mathbb{E}\left[\left(Y_0^2 Y_{-k} - Y_0 Y_{-k}^2\right)\left(Y_t^2 Y_{t-k} - Y_t Y_{t-k}^2\right)\right].$$

For the case $k = 1$, which is of special interest according to the discussion above, we explicitly calculate the asymptotic variance.

**Lemma 5.2.2** (Schweer and Weiß (2015), Lemma 3.1.2)**.** *Let the conditions and notations of Theorem 5.2.1 hold. Then*

$$\sigma_1^2 = 4\mu_Y^2(1-\alpha)^2\left[2\frac{\alpha}{1+\alpha} + \mu_Y \frac{(1-\alpha)(1+\alpha)^2}{1+\alpha+\alpha^2}\right].$$

*Proof.* Note that the expressions in $\sigma_1^2$ consist of terms of the form $\mathbb{E}[Y_0 Y_a Y_1 Y_t Y_b Y_{t+1}]$ with $a \in \{0, 1\}$ and $b \in \{t, t+1\}$ due to stationarity and $t \in \mathbb{N}_0$. For each term, we apply Lemma 2.2.1 (vi) to convert the expectation into an expression of the cumulants, finding

$$\mathbb{E}[Y_0 Y_a Y_1 Y_t Y_b Y_{t+1}] = \sum_{\pi\in\Pi_6} \prod_{B\in\pi} \mathrm{cum}(Y_i \ ; \ i \in B). \tag{5.7}$$

Let $t > 0$, and define the summands of the latter expression as $\nu_\pi(a, b)$ for each $\pi \in \Pi_6$. We calculate

$$\nu_\pi(0, t) + \nu_\pi(1, t+1) - \nu_\pi(0, t+1) - \nu_\pi(1, t) =: \nu_\pi$$

by using Theorem 4.1.6, i.e., we evaluate this expression for different forms of partitions $\pi$ separately. We denote each partition in an obvious manner. For instance, the notation (123)(456) stands for the partition consisting of two blocks, where the three random variables with the three lowest indices, represented by 1, 2 and 3, respectively, are in one block, and the other random variables are contained in the second block. So, for $a = 0$ and $b = t$, this corresponds to the summand $\nu_{(123)(456)}(0, t) = \mathrm{cum}(Y_0, Y_0, Y_1) \cdot \mathrm{cum}(Y_t, Y_t, Y_{t+1})$.

Note that $a$ and $b$ correspond to the second and fifth index, all other indices are equal. Hence, for any partition $\pi \in \Pi_6$, for which none of either the highest or the lowest indices of any block is either 2 or 5, it holds that $\nu_\pi = 0$, for example, $\nu_{(123456)} = 0$. Furthermore, if only one of the highest or lowest index of any block of a given partition $\pi \in \Pi_6$ is either 2 or 5, it also holds that $\nu_\pi = 0$. For example,

$$\nu_{(14)(2356)} = \mu_Y^2 \alpha^t(\alpha^{t+1} + \alpha^t - \alpha^{t+1} - \alpha^t) = 0.$$

Thus, it suffices to consider only those partitions $\pi \in \Pi_6$, in which both 2 and 5 appear as either the highest or the lowest index of any block, given that these blocks contain at least two elements. In what follows, these partitions are referred to as the "remaining partitions". We list the calculation for all possible partitions separately in ascending order with respect to the number of blocks in the partitions. When using letters instead of numbers, we want to express that a certain relation holds for all possible permutations of a given partition, e.g., $(a)(bcdef)$ stands for the partitions (1)(23456), (2)(13456), (3)(12456) and so forth. Note that we have already shown that $\nu_{(123456)} = 0$, it similarly follows that $\sum_{\pi=(a)(bcdef)} \nu_\pi = 0$.

| Remaining partitions | $\nu_\pi(a, b)$ |
|---|---|
| (15)(2346),(16)(2345), (25)(1346), (26)(1345) | $\mu_Y^2\alpha^{t+1+b-a}$ |
| (135)(246),(136)(245), (145)(236), (146)(235) | $\mu_Y^2\alpha^{t+1+b-a}$ |
| (1)(2345)(6) | $\mu_Y^3\alpha^{b-a}$ |
| (12)(34)(56) | $\mu_Y^3\alpha^{t+1-b+t-1+a}$ |
| (12)(35)(46), (12)(36)(45) | $\mu_Y^3\alpha^{t+1+b-t-1+a}$ |
| (13)(24)(56), (14)(23)(56) | $\mu_Y^3\alpha^{t+1-b+t+1-a}$ |
| (13)(25)(46), (13)(26)(45), (15)(23)(46), (16)(23)(45) | $\mu_Y^3\alpha^{t+1+b-t+1-a}$ |
| (14)(25)(36), (14)(26)(35), (15)(24)(36), | |
| (15)(26)(34), (16)(24)(35), (16)(25)(34) | $\mu_Y^3\alpha^{t+1+b+t-1-a}$ |
| (2)(56)(134) | $\mu_Y^3\alpha^{2t-b-a+1}$ |
| (6)(12)(345) | $\mu_Y^3\alpha^{b+a-1}$ |
| (6)(13)(245), (6)(23)(145), (1)(45)(236), (1)(46)(235) | $\mu_Y^3\alpha^{b-a+1}$ |
| (6)(14)(235), (6)(15)(234), (6)(24)(135), (6)(25)(134) | |
| (6)(25)(346), (1)(26)(345), (1)(35)(246), (1)(36)(245) | $\mu_Y^3\alpha^{b-a+t}$ |
| (6)(15)(236), (6)(15)(246), (4)(16)(235), (5)(16)(245) | |
| (1)(25)(136), (1)(25)(146), (4)(26)(135), (4)(26)(145) | $\mu_Y^3\alpha^{t+1+b-a}$ |

In the case $(ab)(cd)(e)(f)$, we first take a closer look at the remaining partitions. Here, some partitions cancel each other out, i.e., $\nu_{(12)(45)(3)(6)} + \nu_{(1)(23)(4)(56)} + \nu_{(12)(3)(4)(56)} + \nu_{(1)(23)(45)(6)} = 0$. We continue with the list:

| Remaining partitions | $\nu_\pi(a,b)$ |
|---|---|
| (5)(6)(12)(34) | $\mu_Y^4\alpha^{b+a-1}$ |
| (1)(3)(24)(56) | $\mu_Y^4\alpha^{-b-a+2t+1}$ |
| (4)(6)(13)(25), (4)(6)(15)(23), (1)(3)(25)(46), (1)(3)(26)(45) | $\mu_Y^4\alpha^{b-a+1}$ |
| (3)(6)(14)(25), (3)(6)(15)(24), (1)(4)(25)(36), (1)(4)(26)(35) | $\mu_Y^4\alpha^{b-a+t}$ |
| (3)(4)(16)(25), (3)(4)(15)(26) | $\mu_Y^4\alpha^{b-a+t+1}$ |
| (1)(6)(24)(35), (1)(6)(25)(34) | $\mu_Y^4\alpha^{b-a+t-1}$ |
| (1)(235)(4)(6), (1)(245)(3)(6) | $\mu_Y^4\alpha^{b-a}$ |
| (1)(25)(3)(4)(6) | $\mu_Y^5\alpha^{b-a}$. |

Now, we can conclude from (5.7) that, for $t > 0$,

$$\begin{aligned}
&\mathbb{E}[(Y_0^2Y_{-1} - Y_0Y_{-1}^2)(Y_t^2Y_{t-1} - Y_tY_{t-1}^2)] \\
&= -\mu_Y^2(1-\alpha)^2\Big(8\alpha^{2t} + \mu_Y\left[4\alpha^{t+1} + 8\alpha^{2t-1} + 8\alpha^{2t} + 6\alpha^{3t-1}\right] \\
&\quad + \mu_Y^2\left(4\alpha^t + 2\alpha^{2t-2}(1+\alpha)^2\right) + \mu_Y^3\alpha^{t-1}\Big). \qquad (5.8)
\end{aligned}$$

Next, let us consider the case $t = 0$. It is quite clear that the same approach is viable, since the calculations are almost entirely analogous, we refrain from presenting the considerations in a s much detail as above. Note that in this case, the ordering is altered in the sense that we consider the expectation $\mathbb{E}[Y_0Y_0Y_aY_bY_1Y_1]$ with $a, b \in \{0, 1\}$. We denote $\nu_\pi(a,b)$ and $\nu_\pi$ in analogy to the case $t > 0$. All except the third and fourth indices are equal. Hence, for any partition $\pi \in \Pi_6$, for which none of either the highest or the lowest indices of any block is either 3 or 4, it holds that $\nu_\pi = 0$. For example, $\nu_{(123456)} = 0$. Furthermore, certain pairs of partitions cancel each other out. This happens if there is a partition in which 3 is either the highest or lowest index in one block and 4 is not, and if there is the corresponding partition in which 4 is the respective highest or lowest index and 3 is not. For example, $\nu_{(13)(2456)} + \nu_{(14)(2356)} = 0$.

The fact that $\nu_{(123456)} = 0$ and $\nu_{(a)(bcdef)} = 0$ follows analogously as before. The remaining partitions in the case $(ab)(cdef)$ are (12)(3456), (34)(1256) and (56)(1234). For partitions of the form $(abc)(def)$, we find $\nu_{(123)(456)} = \mu_Y^2(2\alpha - 2)$, $\nu_{(124)(356)} = \mu_Y^2(2\alpha - 2\alpha^2)$ and $\nu_{(125)(346)} = \nu_{(126)(345)} = \nu_{(134)(256)} = \nu_{(156)(234)} = \mu_Y^2(\alpha - \alpha^2)$.

The case $(a)(b)(cdef)$ yields the summands $\nu_{(1)(2)(3456)} + \nu_{(1234)(5)(6)} = 2\mu_Y^3(1-\alpha)$. For the partitions $(ab)(cd)(ef)$, $(a)(bc)(def)$ and $(ab)(cd)(e)(f)$, we provide the following list, where $a \times b$ means that the term $b$ corresponds to $\nu_\pi(a,b)$ for $a$ different partitions.

| Form of partition | $\nu_\pi(a,b)$ |
|---|---|
| $(ab)(cd)(ef)$ | $1\times\mu_Y^3\alpha^{b-a}$, $2\times\mu_Y^3\alpha^{2-b-a}$, $2\times\mu_Y^3\alpha^{b+a}$ |
| | $4\times\mu_Y^3\alpha^{2-b+a}$, $6\times\mu_Y^3\alpha^{2+b-a}$ |
| $(a)(bc)(def)$ | $2\times\mu_Y^3\alpha^{-a+1}$, $4\times\mu_Y^3\alpha^{-a+2}$,$4\times\mu_Y^3\alpha^{b+1}$, |
| | $2\times\mu_Y^3\alpha^{b}$, $2\times\mu_Y^3\alpha^{b}$, $8\times\mu_Y^3\alpha^{b-a+1}$ |
| $(ab)(cd)(e)(f)$ | $2\times\mu_Y^4\alpha^{b+a}$, $2\times\mu_Y^4\alpha^{2-b-a}$, $2\times\mu_Y^4\alpha^{b-a}$, |
| | $4\times\mu_Y^4\alpha^{-b+a+1}$, $8\times\mu_Y^4\alpha^{b-a+1}$. |

For the penultimate case, $(a)(b)(c)(def)$, the remaining partitions are (1)(2)(5)(346), (1)(2)(4)(356), (2)(5)(6)(134) and (1)(5)(6)(234) leading to the expression $4\mu_Y^4(1-\alpha)$. For the final case $(a)(b)(c)(d)(ef)$, the remaining partition is (1)(34)(2)(5)(6) and we obtain $\nu_{(1)(34)(2)(5)(6)} = 2\mu_Y^5(1-\alpha)$. In total, we have

$$\sum_{\pi\in\Pi_6} \nu_\pi = 2\mu_Y^2(1-\alpha)\left(4\alpha + \mu_Y(2+6\alpha+6\alpha^2) + \mu_Y^2(2+6\alpha) + \mu_Y^3\right).$$

Combining all of these results yields

$$\begin{aligned}
\sigma_1^2 &= \mu_Y^2(1-\alpha)\Bigg[8\alpha + \mu_Y(4+12\alpha+12\alpha^2) + \mu_Y^2(4+12\alpha) \\
&\qquad + 2\mu_Y^3 + 2\sum_{t=1}^{\infty} \mathbb{E}[(Y_0^2 Y_{-1} - Y_0 Y_{-1}^2)(Y_t^2 Y_{t-1} - Y_t Y_{t-1}^2)]\Bigg] \\
&= 4\mu_Y^2(1-\alpha)\left[2\alpha + \mu_Y(1+3\alpha+3\alpha^2) - 4\frac{\alpha^2}{1+\alpha} - \mu_Y\left(2\alpha^2 + 4\alpha + \frac{3\alpha^2}{1+\alpha+\alpha^2}\right)\right] \\
&= 8\mu_Y^2\frac{\alpha(1-\alpha)^2}{1+\alpha} + 4\mu_Y^3\frac{(1-\alpha)^3(1+\alpha)^2}{1+\alpha+\alpha^2},
\end{aligned}$$

concluding the proof. □

Theorem 5.2.1 together with Lemma 5.2.2 gives an explicit expression for the asymptotic distribution of $\sqrt{T-1}\cdot\hat{\beta}_T(1)$ under the null hypothesis of a Poisson INAR(1) model. Using this result, we could design a corresponding test procedure. However, an initial simulation study showed that the true variance of $\sqrt{T-1}\cdot\hat{\beta}_T(1)$ for finite $T$ is larger than the asymptotic one, and that this true variance converges only slowly for increasing $T$. Therefore, to make our test also applicable to short time series, we continue with deriving the exact variance of $\hat{\beta}_T(1)$.

**Corollary 5.2.3** (Schweer and Weiß (2015), Corollary 3.1.3)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process with $\epsilon_t \sim \mathrm{Poi}(\lambda)$ for all $t\in\mathbb{Z}$, let $\mu_Y = \lambda/(1-\alpha)$ and let $T\in\mathbb{N}$ with $T\geq 3$. Then*

$$\begin{aligned}
&\mathrm{Var}\left(\sqrt{T}\hat{\beta}_{T+1}(1)\right) = \sigma_1^2 + \\
&\frac{2\mu_Y^2}{T}\left[\mu_Y(\mu_Y+2\alpha)^2(1-\alpha^T) + 2[\mu_Y(1+\alpha)+2\alpha]^2\frac{1-\alpha^{2T}}{(1+\alpha)^2} + 6\alpha^2\mu_Y\frac{1-\alpha^{3T}}{(1+\alpha+\alpha^2)^2}\right].
\end{aligned}$$

*Proof.* We begin by noticing that for any $p\in[0,1)$ and any $N\in\mathbb{N}$ with $N\geq 2$,

$$\sum_{t=1}^{N-1}\left(1-\frac{t}{N}\right)p^{t-1} = \frac{1}{1-p} - \frac{1}{N}\frac{1-p^N}{(1-p)^2} = \sum_{t=1}^{\infty}p^{t-1} - \frac{1}{N}\frac{1-p^N}{(1-p)^2}. \tag{5.9}$$

Now, the process $(Y_t)_{t\in\mathbb{Z}}$ is stationary and it is easily seen that $\mathbb{E}[\hat{\beta}_T(1)] = 0$. Thus, the variance equals

$$\mathbb{E}\left[(Y_0^2 Y_{-1} - Y_0 Y_{-1}^2)^2\right] + 2\sum_{t=1}^{T-1}\frac{T-t}{T}\mathbb{E}\left[(Y_0^2 Y_{-1} - Y_0 Y_{-1}^2)(Y_t^2 Y_{t-1} - Y_t Y_{t-1}^2)\right].$$

By (5.8), $\mathbb{E}\left[(Y_0^2 Y_{-1} - Y_0 Y_{-1}^2)(Y_t^2 Y_{t-1} - Y_t Y_{t-1}^2)\right]$ is an expression consisting of terms with coefficients $\alpha^{t-1}$, $(\alpha^2)^{t-1}$ and so forth, thus (5.9) is applicable. We find

$$\begin{aligned}
&= \mathbb{E}\left[(Y_0^2 Y_{-1} - Y_0 Y_{-1}^2)^2\right] + 2\sum_{t=1}^{\infty}\mathbb{E}\left[(Y_0^2 Y_{-1} - Y_0 Y_{-1}^2)(Y_t^2 Y_{t-1} - Y_t Y_{t-1}^2)\right] \\
&+ \frac{1}{T}\left[16\mu_Y^2\alpha^2\frac{1-\alpha^{2T}}{(1+\alpha)^2} + 4\mu_Y^3\alpha\left(2\alpha(1-\alpha^T) + 4(1+\alpha)\frac{1-\alpha^{2T}}{(1+\alpha)^2} + 3\alpha\frac{1-\alpha^{3T}}{(1+\alpha+\alpha^2)^2}\right)\right] \\
&+ \frac{1}{T}\left[8\mu_Y^4\alpha(1-\alpha^T) + 4(1-\alpha^{2T}) + 2\mu_Y^5(1-\alpha^T)\right],
\end{aligned}$$

concluding the proof. □

The finite-sample performance of the presented asymptotic approximation is analyzed in Section 5.2.3 below.

### 5.2.2 Testing Skewness in INAR(1) Processes

Looking at relation (5.6), we find another possibility to check for Poisson-distributed innovations. Although originally designed as a time-reversibility test, $\hat{\beta}_T(k)$ also compares the marginal skewness and variance of the process. Therefore, in analogy to the empirical dispersion index in (5.2), we define

$$\theta_Y := \frac{\bar{\mu}_{Y,3}}{\sigma_Y^2} \quad \text{and} \quad \hat{\theta}_Y := \frac{m_{Y,3}}{S_Y^2},$$

where $S_Y^2 = \frac{1}{T}\sum_{t=1}^T (Y_t - \bar{Y})^2$ and $m_{Y,3} = \frac{1}{T}\sum_{t=1}^T (Y_t - \bar{Y})^3$. Using the result of Lemma 4.1.4, if $(Y_t)_{t\in\mathbb{Z}}$ is a Poisson INAR(1) process, it follows that $Y_0$ is Poisson distributed, and thus that $\theta_Y = 1$. In order to show the asymptotic normality of $\hat{\theta}_Y$, we first consider the vector-valued process $\mathbf{X}_t := (Y_t - \mu_Y, Y_t^2 - \mu_Y - \mu_Y^2, Y_t^3 - \mu_Y - 3\mu_Y^2 - \mu_Y^3)$ with $\mathbb{E}[\mathbf{X}_t] = \mathbf{0}$. Since $(Y_t)_{t\in\mathbb{Z}}$ is $\alpha$-mixing with exponentially decreasing weights $\alpha_Y(n)$ by Theorem 4.2.6, so is $(\mathbf{X}_t)_{t\in\mathbb{Z}}$, and we conclude with Theorem 2.5.2 that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbf{X}_t \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0},\ \mathbf{T}).$$

Here, $\mathbf{T}$ has the entries $\tau_{i,j}$ with

$$\tau_{i,j} = \mathbb{E}[X_{0,i}X_{0,j}] + \sum_{k=1}^{\infty}\left(\mathbb{E}[X_{0,1}X_{k,j}] + \mathbb{E}[X_{k,i}X_{0,j}]\right),$$

where $X_{k,j}$ denotes the $j$-th entry of $\mathbf{X}_k$. We calculate the expressions for the entries $\tau_{i,j}$ explicitly, noting that $\tau_{1,1}, \tau_{1,2}$ and $\tau_{2,2}$ were already calculated in Lemma 5.1.1.

**Lemma 5.2.4** (Schweer and Weiß (2015), Lemma 3.2.1). *Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process with $\epsilon_t \sim \text{Poi}(\lambda)$ for all $t \in \mathbb{Z}$ and $\mu_Y = \lambda/(1-\alpha)$. Then*

$$\tau_{1,3} = \mu_Y \left(1 + 6\mu_Y + 3\mu_Y^2\right) \frac{1+\alpha}{1-\alpha},$$
$$\tau_{2,3} = \mu_Y \left(1 + 6\mu_Y + 3\mu_Y^2\right) (1 + 2\mu_Y) \frac{1+\alpha}{1-\alpha} + 6\mu_Y^2 (1+\mu_Y) \frac{1+\alpha^2}{1-\alpha^2},$$
$$\tau_{3,3} = \mu_Y \left(1 + 6\mu_Y + 3\mu_Y^2\right)^2 \frac{1+\alpha}{1-\alpha} \; + \; 18\mu_Y^2 (1+\mu_Y)^2 \frac{1+\alpha^2}{1-\alpha^2} + 6\mu_Y^3 \frac{1+\alpha^3}{1-\alpha^3}.$$

*Proof.* For $\tau_{1,3}$, we consider $\mathbb{E}[Y_0 Y_k^3] = \sum_{\pi\in\Pi_4} \prod_{i=1}^{|\pi|} \text{cum}(Y_{B_i(\pi)})$, where $\Pi_4$ and $B_i(\pi)$ are defined as in (2.4). Denote the summands of the latter expression as $\rho_\pi$ for each $\pi \in \Pi_4$ and recall approach of the second part of the proof of Lemma 5.2.2. With Theorem 4.1.6, we find the following list:

| Form of partition | $\rho_\pi$ |
|---|---|
| $(abcd)$ | $1 \times \mu_Y \alpha^k$ |
| $(a)(bcd)$ | $1 \times \mu_Y^2$, $3 \times \mu_Y^2 \alpha^k$ |
| $(ab)(cd)$ | $3 \times \mu_Y^2 \alpha^k$ |
| $(a)(b)(cd)$ | $3 \times \mu_Y^3$, $3 \times \mu_Y^3 \alpha^k$ |
| $(a)(b)(c)(d)$ | $1 \times \mu_Y^4$. |

Due to the symmetry of the expressions, the same calculations hold for $\mathbb{E}[Y_0^3 Y_k]$, and they also hold for $k = 0$. We thus conclude that

$$\tau_{1,3} = \mathbb{E}[Y_0^4] - \mu_Y^2 - 3\mu_Y^3 - \mu_Y^4 + \sum_{k=1}^{\infty} \left(\mathbb{E}[Y_0^3 Y_k] + \mathbb{E}[Y_0 Y_k^3] - 2\mu_Y^2 - 6\mu_Y^3 - 2\mu_Y^4\right)$$
$$= \mu_Y \left(1 + 6\mu_Y + 3\mu_Y^2\right) (1 + 2\sum_{k=1}^{\infty} \alpha^k) = \mu_Y \left(1 + 6\mu_Y + 3\mu_Y^2\right) \frac{1+\alpha}{1-\alpha}.$$

For the calculation of $\tau_{2,3}$, we proceed analogously:

| Form of partition | $\rho_\pi$ |
|---|---|
| $(abcde)$ | $1 \times \mu_Y \alpha^k$ |
| $(a)(bcde)$ | $5 \times \mu_Y^2 \alpha^k$ |
| $(ab)(cde)$ | $1 \times \mu_Y^2$, $3 \times \mu_Y^2 \alpha^k$, $6 \times \mu_Y^2 \alpha^{2k}$ |
| $(a)(b)(cde)$ | $1 \times \mu_Y^3$, $9 \times \mu_Y^3 \alpha^k$ |
| $(a)(bc)(de)$ | $3 \times \mu_Y^3$, $6 \times \mu_Y^3 \alpha^k$, $6 \times \mu_Y^3 \alpha^{2k}$ |
| $(a)(b)(c)(de)$ | $4 \times \mu_Y^4$, $6 \times \mu_Y^4 \alpha^k$ |
| $(a)(b)(c)(d)(e)$ | $1 \times \mu_Y^5$. |

Due to the symmetry of the expressions, the same calculations hold for the expectation $\mathbb{E}[Y_0^2 Y_k^3]$, and they also hold for $k = 0$. The assertion for this entry now follows from easy calculations, analogous to the case above. Finally, we calculate $\tau_{3,3}$:

| Form of partition | $\rho_\pi$ |
|---|---|
| $(abcdef)$ | $1 \times \mu_Y \alpha^k$ |
| $(a)(bcdef)$ | $6 \times \mu_Y^2 \alpha^k$ |
| $(ab)(cdef)$ | $6 \times \mu_Y^2 \alpha^k$, $9 \times \mu_Y^2 \alpha^{2k}$ |
| $(abc)(def)$ | $1 \times \mu_Y^2$, $9 \times \mu_Y^2 \alpha^{2k}$ |
| $(a)(b)(cdef)$ | $15 \times \mu_Y^3 \alpha^k$ |
| $(a)(bc)(def)$ | $6 \times \mu_Y^3$, $18 \times \mu_Y^3 \alpha^k$, $36 \times \mu_Y^3 \alpha^{2k}$ |
| $(ab)(cd)(ef)$ | $9 \times \mu_Y^3 \alpha^k$, $6 \times \mu_Y^3 \alpha^{2k}$ |
| $(a)(b)(c)(def)$ | $2 \times \mu_Y^4$, $18 \times \mu_Y^4 \alpha^k$ |
| $(a)(b)(cd)(ef)$ | $9 \times \mu_Y^4$, $18 \times \mu_Y^4 \alpha^k$, $18 \times \mu_Y^4 \alpha^{2k}$ |
| $(a)(b)(c)(d)(ef)$ | $6 \times \mu_Y^5$, $9 \times \mu_Y^5 \alpha^k$, |
| $(a)(b)(c)(d)(e)(f)$ | $1 \times \mu_Y^6$. |

This list extends to the case $k = 0$, and easy calculations conclude the proof. □

Lemma 5.2.4 is now applied to derive a closed-form expression for the asymptotic distribution of $\hat{\theta}_Y$. In addition, we also develop an expression to correct the bias for small $T$ in analogy to Remark A.3.1 in Weiß and Schweer (2015).

**Theorem 5.2.5** (Schweer and Weiß (2015), Theorem 3.2.2)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process with $\epsilon_t \sim \text{Poi}(\lambda)$ for all $t \in \mathbb{Z}$ and $\mu_Y = \lambda/(1-\alpha)$. Then*

$$\sqrt{T}(\hat{\theta}_Y - 1) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ 8\frac{1+\alpha^2}{1-\alpha^2} + 6\mu_Y \frac{1+\alpha^3}{1-\alpha^3}\right) \quad \text{for } T \to \infty.$$

*An improved approximation for the mean of $\hat{\theta}_Y$ is given by*

$$\mathbb{E}\left[\hat{\theta}_Y\right] \approx 1 - \frac{2}{T}\frac{3+2\alpha+3\alpha^2}{1-\alpha^2}.$$

*Proof.* We introduce the function

$$f : \mathbb{R}^3 \to \mathbb{R}, \quad f(y_1, y_2, y_3) := \frac{y_3 - 3y_1y_2 + 2y_1^3}{y_2 - y_1^2} = \frac{y_3 - y_1^3}{y_2 - y_1^2} - 3y_1.$$

It is easily seen that, on the one hand $f(\mu_Y, \mu_Y^2 + \mu_Y, \mu_Y^3 + 3\mu_Y^2 + \mu_Y) = 1$, on the other hand $f(\frac{1}{T}\sum_{t=1}^T Y_t, \frac{1}{T}\sum_{t=1}^T Y_t^2, \frac{1}{T}\sum_{t=1}^T Y_t^3) = \hat{\theta}_Y$. We readily find

$$\frac{\partial}{\partial y_1} f(y_1, y_2, y_3) = \frac{2y_1(y_3 - y_1^3)}{(y_2 - y_1^2)^2} - \frac{3y_1^2}{y_2 - y_1^2} - 3,$$
$$\frac{\partial}{\partial y_2} f(y_1, y_2, y_3) = \frac{y_1^3 - y_3}{(y_2 - y_1^2)^2} \quad \text{and} \quad \frac{\partial}{\partial y_3} f(y_1, y_2, y_3) = \frac{1}{y_2 - y_1^2},$$

and thus that

$$\mathbf{D} := \text{grad}\, f(\mu_Y, \mu_Y^2 + \mu_Y, \mu_Y^3 + 3\mu_Y^2 + \mu_Y) = \frac{1}{\mu_Y}\left(3\mu_Y^2 - \mu_Y, -1 - 3\mu_Y, 1\right).$$

In order to simplify notation, we abbreviate some polynomials appearing in the following calculation by $A := 1+6\mu_Y+3\mu_Y^2$, $B := 1+2\mu_Y$, $C := 1+3\mu_Y$ and $D := \mu_Y(3\mu_Y-1)$. The dependence on the parameter $\mu_Y$ is omitted in a slight abuse of notation. We calculate

$$\begin{aligned}\mu_Y^2 \mathbf{D}\mathbf{T}\mathbf{D}^T &= \tau_{1,1}D^2 + C^2\tau_{2,2} + \tau_{3,3} + 2D\tau_{1,3} - 2DC\tau_{1,2} - 2C\tau_{2,3} \\ &= \frac{1+\alpha}{1-\alpha}\mu_Y\left(D^2 + B^2C^2 + A^2 + 2AD - 2BCD - 2ABC\right) \\ &+ 2\frac{1+\alpha^2}{1-\alpha^2}\mu_Y^2\left(C^2 - 6C(1+\mu_Y) + 9(1+\mu_Y)^2\right) + 6\mu_Y^3\frac{1+\alpha^3}{1-\alpha^3}.\end{aligned}$$

Simple manipulations yield $C^2 - 6C(1+\mu_Y) + 9(1+\mu_Y)^2 = (C - 3(1+\mu_Y))^2 = 4$. Furthermore, notice that, in our notation, $D = BC - A$. Applying this relation yields $D^2 + B^2C^2 + A^2 + 2AD - 2BCD - 2ABC = 0$, and thus concludes the proof of the asymptotic result. The bias correction is derived in the same way as in Proposition 5.1.4. First, the Hessian of $f(y_1, y_2, y_3)$ is computed and $(\mu_Y, \mu_Y^2+\mu_Y, \mu_Y^3+3\mu_Y^2+\mu_Y)$ is inserted into this Hessian, leading to

$$\mathbf{H}_f = \begin{pmatrix} 8+\frac{2}{\mu_Y}+12\mu_Y & -9-\frac{4}{\mu_Y} & \frac{2}{\mu_Y} \\ -9-\frac{4}{\mu_Y} & \frac{2}{\mu_Y^2}+\frac{6}{\mu_Y} & -\frac{1}{\mu_Y^2} \\ \frac{2}{\mu_Y} & -\frac{1}{\mu_Y^2} & 0 \end{pmatrix}$$

Then, with $\mathbf{Z}_T := \frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbf{X}_t$, we find, using the formulae for $\tau_{i,j}$ from Lemma 5.2.4,

$$\mathbb{E}\left[\frac{1}{2}\mathbf{Z}_T^\top \mathbf{H}_f \mathbf{Z}_T\right] = -2\frac{3+2\alpha+3\alpha^2}{1-\alpha^2}.$$

□

Let us now analyze the finite-sample performance of the presented asymptotic approximation. In the following sections, we summarize the results of a simulation study designed to demonstrate the behavior of the estimators and tests developed above. The first two parts of this section deal with the finite-sample performance of the asymptotic results 5.2.1 and 5.2.5. The third part provides a power analysis for the tests developed here, and compares their ability to detect violations from the Poisson assumption with the test based on the empirical dispersion index (5.2) of Section 5.1.

### 5.2.3 Finite-Sample Performance of the generalized autocovariance

To analyze the finite-sample performance of the asymptotic result of Theorem 5.2.1 and Corollary 5.2.3, a simulation study was done with 10,000 replications per scenario. We only show some illustrative results here. The empirically observed means and standard deviations nearly perfectly coincide with their theoretical values (given by 0 and Corollary 5.2.3, respectively). Approximate normality becomes visible not only by checking the empirical quantiles, but also from the rejection rates being summarized in Table 5.4.

| $\alpha$ | $T \setminus \lambda$ | Size (specified parameters) 0.5 | 1 | 2 | 4 | Size (estimated parameters) 0.5 | 1 | 2 | 4 |
|---|---|---|---|---|---|---|---|---|---|
| 0.3 | 100 | 0.050 | 0.055 | 0.060 | 0.057 | 0.048 | 0.054 | 0.052 | 0.052 |
| | 250 | 0.048 | 0.056 | 0.054 | 0.053 | 0.051 | 0.051 | 0.051 | 0.052 |
| | 500 | 0.055 | 0.051 | 0.053 | 0.049 | 0.052 | 0.049 | 0.053 | 0.048 |
| | 1000 | 0.051 | 0.051 | 0.049 | 0.052 | 0.053 | 0.049 | 0.049 | 0.050 |
| 0.5 | 100 | 0.062 | 0.055 | 0.054 | 0.059 | 0.045 | 0.045 | 0.044 | 0.055 |
| | 250 | 0.055 | 0.048 | 0.054 | 0.052 | 0.051 | 0.044 | 0.050 | 0.050 |
| | 500 | 0.052 | 0.050 | 0.049 | 0.053 | 0.053 | 0.050 | 0.050 | 0.053 |
| | 1000 | 0.056 | 0.049 | 0.051 | 0.052 | 0.054 | 0.049 | 0.050 | 0.050 |
| 0.7 | 100 | 0.049 | 0.050 | 0.058 | 0.061 | 0.040 | 0.043 | 0.054 | 0.059 |
| | 250 | 0.052 | 0.052 | 0.058 | 0.059 | 0.045 | 0.047 | 0.055 | 0.057 |
| | 500 | 0.050 | 0.052 | 0.054 | 0.057 | 0.047 | 0.047 | 0.053 | 0.057 |
| | 1000 | 0.052 | 0.051 | 0.051 | 0.056 | 0.049 | 0.049 | 0.050 | 0.055 |

Table 5.4: Simulated sizes if testing null of Poisson INAR(1) process (5 % level, two-sided test) via $\hat{\beta}_T(1)$.

Columns 3–6 show the sizes if testing the null of a Poisson INAR(1) process (5 % level, two-sided test) with specified parameters, and columns 7–10 the sizes if the parameter values for $(\lambda, \alpha)$ are estimated from the same data. In any case, the observed sizes are very close to the nominal level 0.05, confirming the good performance of our asymptotic approximations.

| | $\lambda$ | 0.5 | | 1 | | 2 | | 4 | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $T$ | sim. | asymp. | sim. | asymp. | sim. | asymp. | sim. | asymp. |
| 0.3 | 100 | 0.921 | 0.915 | 0.923 | 0.915 | 0.924 | 0.915 | 0.914 | 0.915 |
| | 250 | 0.968 | 0.966 | 0.971 | 0.966 | 0.966 | 0.966 | 0.966 | 0.966 |
| | 500 | 0.984 | 0.983 | 0.980 | 0.983 | 0.986 | 0.983 | 0.988 | 0.983 |
| | 1000 | 0.992 | 0.991 | 0.992 | 0.991 | 0.992 | 0.991 | 0.989 | 0.991 |
| 0.5 | 100 | 0.887 | 0.873 | 0.887 | 0.873 | 0.886 | 0.873 | 0.879 | 0.873 |
| | 250 | 0.950 | 0.949 | 0.948 | 0.949 | 0.951 | 0.949 | 0.953 | 0.949 |
| | 500 | 0.977 | 0.975 | 0.974 | 0.975 | 0.980 | 0.975 | 0.978 | 0.975 |
| | 1000 | 0.988 | 0.987 | 0.989 | 0.987 | 0.987 | 0.987 | 0.991 | 0.987 |
| 0.7 | 100 | 0.821 | 0.770 | 0.811 | 0.770 | 0.813 | 0.770 | 0.806 | 0.770 |
| | 250 | 0.912 | 0.908 | 0.915 | 0.908 | 0.917 | 0.908 | 0.915 | 0.908 |
| | 500 | 0.952 | 0.954 | 0.962 | 0.954 | 0.947 | 0.954 | 0.970 | 0.954 |
| | 1000 | 0.981 | 0.977 | 0.977 | 0.977 | 0.976 | 0.977 | 0.982 | 0.977 |

Table 5.5: Simulated mean of $\hat{\theta}_Y$ vs. asymptotic approximations for Poisson INAR(1) process.

### 5.2.4 Finite-Sample Performance of $\hat{\theta}_Y$

To analyze the finite-sample performance of the asymptotic result according to Theorem 5.2.5, we continued the simulation study from Section 5.2.3. Let us first compare the empirically observed means and standard deviations with their asymptotic approx-

| | λ | 0.5 | | 1 | | 2 | | 4 | |
|---|---|---|---|---|---|---|---|---|---|
| α | T | sim. | asymp. | sim. | asymp. | sim. | asymp. | sim. | asymp. |
| 0.3 | 100 | 0.316 | 0.376 | 0.380 | 0.432 | 0.475 | 0.526 | 0.626 | 0.677 |
| | 250 | 0.221 | 0.238 | 0.258 | 0.273 | 0.314 | 0.333 | 0.412 | 0.428 |
| | 500 | 0.161 | 0.168 | 0.187 | 0.193 | 0.230 | 0.235 | 0.296 | 0.303 |
| | 1000 | 0.118 | 0.119 | 0.134 | 0.136 | 0.162 | 0.166 | 0.213 | 0.214 |
| 0.5 | 100 | 0.359 | 0.459 | 0.456 | 0.536 | 0.565 | 0.665 | 0.762 | 0.866 |
| | 250 | 0.265 | 0.290 | 0.312 | 0.339 | 0.394 | 0.420 | 0.523 | 0.548 |
| | 500 | 0.194 | 0.205 | 0.228 | 0.240 | 0.291 | 0.297 | 0.379 | 0.387 |
| | 1000 | 0.142 | 0.145 | 0.167 | 0.170 | 0.205 | 0.210 | 0.272 | 0.274 |
| 0.7 | 100 | 0.477 | 0.662 | 0.597 | 0.802 | 0.805 | 1.025 | 1.120 | 1.367 |
| | 250 | 0.355 | 0.419 | 0.445 | 0.507 | 0.579 | 0.649 | 0.782 | 0.865 |
| | 500 | 0.268 | 0.296 | 0.334 | 0.358 | 0.425 | 0.459 | 0.579 | 0.611 |
| | 1000 | 0.203 | 0.209 | 0.244 | 0.253 | 0.315 | 0.324 | 0.428 | 0.432 |

Table 5.6: Simulated standard deviation of $\hat{\theta}_Y$ vs. asymptotic approximations for Poisson INAR(1) process.

imations, see Tables 5.5 and 5.6. The approximations for the mean tend to give lower values than empirically observed in the simulation study (especially if the autocorrelation parameter $\alpha$ is large), but the difference rapidly decreases with increasing sample size. The asymptotic standard deviations are, in contrast, larger than the simulated ones (again stronger if $\alpha$ is large), so we expect a test being designed on these approximations to be conservative. In fact, looking at the empirical sizes in Table 5.7 (columns 3–6), these values are always below the nominal level of 5 %, i.e., the rate of false rejections is even better than required by design. Again, the effect of using estimated parameters for the null model (columns 7–10) is small.

| | | Size (specified parameters) | | | | Size (estimated parameters) | | | |
|---|---|---|---|---|---|---|---|---|---|
| α | T \ λ | 0.5 | 1 | 2 | 4 | 0.5 | 1 | 2 | 4 |
| 0.3 | 100 | 0.026 | 0.030 | 0.032 | 0.033 | 0.025 | 0.028 | 0.031 | 0.033 |
| | 250 | 0.035 | 0.039 | 0.039 | 0.042 | 0.033 | 0.039 | 0.038 | 0.042 |
| | 500 | 0.040 | 0.042 | 0.043 | 0.044 | 0.040 | 0.042 | 0.043 | 0.044 |
| | 1000 | 0.047 | 0.044 | 0.045 | 0.048 | 0.046 | 0.043 | 0.044 | 0.048 |
| 0.5 | 100 | 0.023 | 0.029 | 0.026 | 0.029 | 0.019 | 0.028 | 0.025 | 0.029 |
| | 250 | 0.032 | 0.035 | 0.036 | 0.040 | 0.033 | 0.035 | 0.036 | 0.040 |
| | 500 | 0.037 | 0.037 | 0.047 | 0.044 | 0.038 | 0.037 | 0.047 | 0.044 |
| | 1000 | 0.043 | 0.044 | 0.045 | 0.050 | 0.044 | 0.044 | 0.044 | 0.049 |
| 0.7 | 100 | 0.020 | 0.020 | 0.021 | 0.027 | 0.017 | 0.017 | 0.019 | 0.025 |
| | 250 | 0.027 | 0.029 | 0.031 | 0.034 | 0.024 | 0.028 | 0.030 | 0.035 |
| | 500 | 0.032 | 0.038 | 0.037 | 0.038 | 0.032 | 0.037 | 0.038 | 0.039 |
| | 1000 | 0.042 | 0.040 | 0.042 | 0.050 | 0.042 | 0.041 | 0.041 | 0.050 |

Table 5.7: Simulated sizes if testing null of Poisson INAR(1) process (5 % level, two-sided test) via $\hat{\theta}_Y$.

### 5.2.5 Power Analysis

In Proposition 5.1.4, it is shown that if $Y_1, \ldots, Y_T$ stem from a Poisson INAR(1) process, then the distribution of $\hat{I}_Y$ can be approximated by a normal distribution with

$$\mathbb{E}\left[\hat{I}_Y\right] \approx 1 - \frac{1}{T}\frac{1+\alpha}{1-\alpha} \quad \text{and} \quad \operatorname{Var}\left(\hat{I}_Y\right) \approx \frac{2}{T}\frac{1+\alpha^2}{1-\alpha^2}.$$

These relations can be utilized to test the null of a Poisson INAR(1) model against the alternative of an INAR(1) model with differently distributed innovations. In addition to this dispersion test, we developed two further tests for this situation: the reversibility test $\hat{\beta}_T(1)$ in Section 5.2.1, and the skewness test $\hat{\theta}_Y$ in Section 5.2.2. For all these tests, the finite-sample properties under the null hypothesis have been shown in simulations to fit reasonably well to those properties being expected from asymptotic results, see the previous sections. In this section, we will compare the power of all three tests under diverse alternative scenarios.

| | $n$ | 1 | | | 2 | | | 8 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $T$ | $\hat{I}$ | $\hat{\beta}_T(1)$ | $\hat{\theta}_Y$ | $\hat{I}$ | $\hat{\beta}_T(1)$ | $\hat{\theta}_Y$ | $\hat{I}$ | $\hat{\beta}_T(1)$ | $\hat{\theta}_Y$ |
| 0.3 | 100 | 0.556 | 0.370 | 0.394 | 0.576 | 0.324 | 0.351 | 0.611 | 0.227 | 0.225 |
| | 250 | 0.869 | 0.542 | 0.723 | 0.894 | 0.488 | 0.666 | 0.920 | 0.310 | 0.427 |
| | 500 | 0.987 | 0.738 | 0.938 | 0.992 | 0.639 | 0.909 | 0.996 | 0.412 | 0.673 |
| | 1000 | 1.000 | 0.890 | 0.999 | 1.000 | 0.832 | 0.998 | 1.000 | 0.560 | 0.907 |
| 0.5 | 100 | 0.385 | 0.414 | 0.257 | 0.408 | 0.350 | 0.229 | 0.407 | 0.155 | 0.132 |
| | 250 | 0.698 | 0.641 | 0.524 | 0.709 | 0.569 | 0.446 | 0.736 | 0.267 | 0.251 |
| | 500 | 0.921 | 0.847 | 0.792 | 0.937 | 0.781 | 0.709 | 0.948 | 0.440 | 0.421 |
| | 1000 | 0.996 | 0.971 | 0.968 | 0.998 | 0.943 | 0.935 | 0.998 | 0.685 | 0.662 |
| 0.7 | 100 | 0.226 | 0.328 | 0.126 | 0.232 | 0.217 | 0.109 | 0.226 | 0.105 | 0.065 |
| | 250 | 0.435 | 0.646 | 0.276 | 0.437 | 0.489 | 0.218 | 0.450 | 0.124 | 0.124 |
| | 500 | 0.677 | 0.884 | 0.460 | 0.693 | 0.782 | 0.370 | 0.693 | 0.201 | 0.193 |
| | 1000 | 0.922 | 0.985 | 0.736 | 0.921 | 0.963 | 0.607 | 0.933 | 0.461 | 0.323 |

Table 5.8: Simulated power for alternative $\epsilon_t \sim \text{NegBin}(n, 2/3)$, i.e., with $I_\epsilon = 1.5$.

As the first alternative, we assume the innovations to be negative binomially distributed, i.e., $\epsilon_t \sim \text{NegBin}(n, \pi)$, such that the innovations' mean equals $n(1-\pi)/\pi$, their dispersion index $1/\pi$. This assumption implies that the means of $\hat{I}$ and $\hat{\theta}_Y$ are increased compared to the null situation, while the mean of $\hat{\beta}_T(1)$ is decreased towards negative values. Since the most obvious difference between a Poisson distribution and a negative binomial one is overdispersion, we expect the test based on $\hat{I}$ to perform particularly well. Looking at Tables 5.8 and 5.9, it becomes clear, however, that this statement does not universally hold. While the $\hat{I}$-test is always superior if $\alpha = 0.3$, the $\hat{\beta}_T(1)$-test is often superior if $\alpha \geq 0.5$ (and if the innovations' mean is small, say $\leq 1$). In fact, the degradation of the $\hat{I}$-test's power for increasing $\alpha$ was one of the disadvantages pointed out by Weiß and Schweer (2015).

In view of our results here, the application of the $\hat{\beta}_T(1)$-test can be recommended as a remedy in situations with a low mean and a high $\alpha$. It is also worth pointing out that the $\hat{\theta}_Y$-test, which considers the skewness of the counts, is also rather sensitive

| | $n$ | 0.5 | | | 1 | | | 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $T$ | $\hat{I}$ | $\hat{\beta}_T(1)$ | $\hat{\theta}_Y$ | $\hat{I}$ | $\hat{\beta}_T(1)$ | $\hat{\theta}_Y$ | $\hat{I}$ | $\hat{\beta}_T(1)$ | $\hat{\theta}_Y$ |
| 0.3 | 100 | 0.879 | 0.615 | 0.739 | 0.922 | 0.598 | 0.729 | 0.952 | 0.434 | 0.523 |
| | 250 | 0.998 | 0.845 | 0.978 | 0.999 | 0.812 | 0.974 | 1.000 | 0.621 | 0.851 |
| | 500 | 1.000 | 0.957 | 1.000 | 1.000 | 0.942 | 1.000 | 1.000 | 0.779 | 0.985 |
| | 1000 | 1.000 | 0.996 | 1.000 | 1.000 | 0.991 | 1.000 | 1.000 | 0.922 | 1.000 |
| 0.5 | 100 | 0.736 | 0.700 | 0.565 | 0.787 | 0.659 | 0.535 | 0.830 | 0.352 | 0.332 |
| | 250 | 0.972 | 0.920 | 0.905 | 0.984 | 0.894 | 0.865 | 0.992 | 0.635 | 0.625 |
| | 500 | 1.000 | 0.992 | 0.993 | 1.000 | 0.987 | 0.989 | 1.000 | 0.862 | 0.876 |
| | 1000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.983 | 0.990 |
| 0.7 | 100 | 0.497 | 0.634 | 0.323 | 0.527 | 0.517 | 0.271 | 0.539 | 0.162 | 0.141 |
| | 250 | 0.836 | 0.939 | 0.637 | 0.857 | 0.891 | 0.556 | 0.881 | 0.301 | 0.301 |
| | 500 | 0.979 | 0.996 | 0.890 | 0.988 | 0.989 | 0.816 | 0.994 | 0.639 | 0.488 |
| | 1000 | 1.000 | 1.000 | 0.992 | 1.000 | 1.000 | 0.979 | 1.000 | 0.941 | 0.752 |

Table 5.9: Simulated power for alternative $\epsilon_t \sim \text{NegBin}(n, 1/2)$, i.e., with $I_\epsilon = 2$.

towards the Negative Binomial alternative provided that the innovations' mean is small. As a final remark concerning Tables 5.8 and 5.9, we point out that the reported values were calculated with respect to a fully specified model. We also considered the case with estimated null parameters, but since the use of estimated parameters had nearly no influence on the power values, these results are not reported here.

As another type of alternative, we consider the Good INAR(1) model discussed in Weiß (2013) with parameters $(q', \nu)$ chosen such that the Good-distributed innovations exhibit equidispersion, even though they are not Poisson distributed. So the innovations share mean and variance with Poisson-distributed innovations, but higher-order moments differ, as can be seen in Table 5.10.

| | | Good$(q', \nu)$ | | | | Poi$(\lambda)$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $q'$ | $\nu$ | Mean | Var. | Skew. | Exc. | $\lambda$ | Mean | Var. | Skew. | Exc. |
| –2.51 | –2.65 | 0.500 | 0.500 | 1.455 | 2.335 | 0.5 | 0.500 | 0.500 | 1.414 | 2.000 |
| –1.96 | –2.91 | 1.000 | 1.000 | 1.103 | 1.571 | 1 | 1.000 | 1.000 | 1.000 | 1.000 |
| –1.51 | –3.55 | 2.000 | 2.000 | 0.926 | 1.289 | 2 | 2.000 | 2.000 | 0.707 | 0.500 |
| –1.25 | –5.25 | 4.000 | 4.000 | 0.800 | 0.961 | 4 | 4.000 | 4.000 | 0.500 | 0.250 |

Table 5.10: Comparison of higher-order moments of Poisson and Good distributions.

For this reason, we expect the dispersion test to perform badly, but the skewness test to be sensitive to this type of alternative situation, especially with an increasing mean. The resulting simulated power values are summarized in Table 5.11.

First of all, the power values are much lower than in the previous alternative situations, but this is reasonable since the considered Good model deviates only little from the Poisson model. Now looking at the power values, we indeed realize that the $\hat{I}$-test completely fails, while both the $\hat{\beta}_T(1)$-test and the $\hat{\theta}_Y$-test are sensitive to this type of alternative. In most cases, the $\hat{\theta}_Y$-test performs better, with the best performance for small $\alpha$ and large innovations' mean. However, as in the case of negative binomially

| $\alpha$ | $(q',\nu)$ $T$ | (–2.5119, –2.6470) $\hat{I}$ | $\hat{\beta}_T(1)$ | $\hat{\theta}_Y$ | (–1.9560, –2.9135) $\hat{I}$ | $\hat{\beta}_T(1)$ | $\hat{\theta}_Y$ | (–1.2498, –5.2489) $\hat{I}$ | $\hat{\beta}_T(1)$ | $\hat{\theta}_Y$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.3 | 100 | 0.040 | 0.058 | 0.037 | 0.054 | 0.069 | 0.049 | 0.058 | 0.076 | 0.108 |
| | 250 | 0.048 | 0.061 | 0.048 | 0.054 | 0.072 | 0.064 | 0.060 | 0.091 | 0.174 |
| | 500 | 0.055 | 0.067 | 0.054 | 0.058 | 0.074 | 0.084 | 0.061 | 0.107 | 0.290 |
| | 1000 | 0.056 | 0.065 | 0.065 | 0.060 | 0.081 | 0.107 | 0.061 | 0.166 | 0.486 |
| 0.5 | 100 | 0.039 | 0.069 | 0.029 | 0.043 | 0.063 | 0.036 | 0.045 | 0.070 | 0.061 |
| | 250 | 0.042 | 0.056 | 0.040 | 0.051 | 0.068 | 0.054 | 0.053 | 0.081 | 0.104 |
| | 500 | 0.049 | 0.067 | 0.051 | 0.053 | 0.076 | 0.066 | 0.055 | 0.118 | 0.148 |
| | 1000 | 0.053 | 0.067 | 0.055 | 0.053 | 0.085 | 0.074 | 0.058 | 0.192 | 0.247 |
| 0.7 | 100 | 0.034 | 0.052 | 0.019 | 0.038 | 0.053 | 0.026 | 0.037 | 0.063 | 0.036 |
| | 250 | 0.042 | 0.060 | 0.027 | 0.047 | 0.063 | 0.039 | 0.046 | 0.062 | 0.057 |
| | 500 | 0.046 | 0.064 | 0.035 | 0.046 | 0.067 | 0.043 | 0.049 | 0.071 | 0.075 |
| | 1000 | 0.049 | 0.061 | 0.045 | 0.052 | 0.080 | 0.055 | 0.051 | 0.096 | 0.110 |

Table 5.11: Simulated power for alternative $\epsilon_t \sim \text{Good}(q',\nu)$.

distributed innovations, the $\hat{\beta}_T(1)$-test fares best in situations with a low mean and a high $\alpha$.

## 5.3 Asymptotics for ACF and PACF of Poisson INAR(1) processes

We now return to more classical theory and consider the asymptotic distribution of the functions introduced in Section 2.6. In particular, we show in this section that the results of Theorems 2.6.2 and 2.6.3 hold in the case of INAR(1) processes. The $\alpha$-mixing property shown in 4.2.6 together with a bound on the fourth moment of the arrival turns out to be sufficient for this assertion.

In what follows, we restrict ourselves to the special case of Poisson INAR(1) processes for two connected reasons. First of all, this restriction allows for an explicit calculation of the resultant terms due ot the surprisingly simple result of Theorem 4.1.6. Secondly, we can show that in this case a very classical result dating back to Quenouille (1949) holds.

**Theorem 5.3.1.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process, where $\epsilon_t \sim \text{Poi}(\lambda)$ for all $t \in \mathbb{Z}$. Then*

$$\sqrt{T}\left(\widehat{\gamma}(0)-\gamma(0),\ldots,\widehat{\gamma}(K)-\gamma(K)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0},\mathbf{T}+\mathbf{U}\right),$$

*where the entries $\tau_{i,j}$ of the matrix $\mathbf{T}$ are given by*

$$\tau_{i+1,j+1} = \gamma(0)^2\left[\left(\frac{1+\alpha^2}{1-\alpha^2}+(j-i)\right)\alpha^{j-i}+\left(\frac{1+\alpha^2}{1-\alpha^2}+(j+i)\right)\alpha^{i+j}\right].$$

*and the entries $u_{i,j}$ of the matrix $\mathbf{U}$ are given by*

$$u_{i+1,j+1} = \gamma(0)\alpha^j\left(\frac{1+\alpha}{1-\alpha}+j-i\right).$$

*Proof.* As shown in Theorem 4.2.6, INAR(1) processes with Poisson innovations are $\alpha$-mixing with exponentially decreasing weights $\alpha_Y(n)$ and all moments exist, thus Theorem 2.5.2 is applicable, implying joint convergence of the estimators for the autocovariance function, where the asymptotic variance matches that given in Theorem 2.6.2, cf. Theorem 3.1 in Romano and Thombs (1996). For $\tau_{i,j}$ with $j \geq i$, we find

$$\frac{1}{\gamma(0)^2} \sum_{d=-\infty}^{\infty} [\gamma(d)\gamma(d+j-i) + \gamma(d+j)\gamma(d-i)] = \sum_{d=-\infty}^{\infty} \left[\alpha^{|d|}\alpha^{|d+j-i|} + \alpha^{|d+j|}\alpha^{|d-i|}\right]$$
$$= \sum_{d=1}^{\infty} \alpha^{2d+j-i} + \sum_{d=i-j}^{-1} \alpha^{j-i} + \sum_{d=0}^{\infty} \alpha^{2d+j-i} + \sum_{d=1}^{\infty} \alpha^{2d+i+j} + \sum_{d=-j}^{i-1} \alpha^{i+j} + \sum_{d=0}^{\infty} \alpha^{2d+i+j}$$
$$= \left(\frac{1+\alpha^2}{1-\alpha^2} + (j-i)\right) \alpha^{j-i} + \left(\frac{1+\alpha^2}{1-\alpha^2} + (j+i)\right) \alpha^{i+j}.$$

Similarly, with Theorem 4.1.6, we find

$$u_{i+1,j+1} = \sum_{d=-\infty}^{\infty} \mathrm{cum}(Y_0, Y_i, Y_d, Y_{d+j}) = \gamma(0) \sum_{d=-\infty}^{i-j} \alpha^{i-d} + \gamma(0) \sum_{d=1}^{\infty} \alpha^{d+j} + \gamma(0) \sum_{d=0}^{j-i-1} \alpha^j$$
$$= \gamma(0)\alpha^j \left(\frac{1+\alpha}{1-\alpha} + j - i\right).$$

□

This result allows the application of Theorems 2.6.2 and 2.6.3, implying that the estimators of the ACF and PACF are jointly asymptotically normal as well. For the covariance matrices of the asymptotic distributions, those of the empirical ACF corresponding to Theorem 2.6.2 are easily derived. For the estimator of the PACF, we now show that the covariance matrix has a surprisingly simple structure. Indeed, we show that the sample partial autocorrelations $\widehat{\rho}_{\text{part}}(k), \widehat{\rho}_{\text{part}}(k+1), \ldots$ are asymptotically independent for $k \geq 2$ for an underlying Poisson INAR(1) process.

**Theorem 5.3.2** (Cp. Mills and Seneta (1991), Sect. 3)**.** *Let* $(Y_t)_{t\in\mathbb{Z}}$ *be a Poisson INAR(1) process where* $\epsilon_t \sim \mathrm{Poi}(\lambda)$ *for all* $t \in \mathbb{Z}$ *and let* $K \geq 2$*. Then*

$$\sqrt{T}\left(\widehat{\rho}_{part}(2), \ldots, \widehat{\rho}_{part}(K)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \Sigma\right),$$

*where* $\Sigma$ *is the matrix with entries* $\sigma_{i,j}$ *for* $i, j \geq 2$ *given by*

$$\sigma_{i,j} = \begin{cases} 1 + \frac{\alpha^i}{\gamma(0)(1+\alpha)^2} & \text{for } i = j, \\ 0 & \text{else.} \end{cases}$$

*Proof.* For the partial autocorrelation function of a Poisson INAR(1) process we have $\rho_{part}(1) = \rho(1) = \alpha$, see Definition 2.6.1, and $\gamma(0) = \frac{\lambda}{1-\alpha}$. The conditions of Theorem

2.6.3 are clearly satisfied, and recalling that $\alpha_0 := -1$ and $\alpha_1 = \alpha$, we calculate with Theorem 5.3.1

$$\begin{aligned}
\sigma_{1,1}\gamma(0)(1-\alpha^2)^2 &= \sum_{u=0}^{2}\sum_{m+n=u}\alpha_m\alpha_n\sum_{r=0}^{2}\sum_{s+t=r}\alpha_s\alpha_t\alpha^{\max\{2-u,2-r\}}\left(\frac{1+\alpha}{1-\alpha}+|u-r|\right)\\
&= \frac{1+\alpha}{1-\alpha}\left[\alpha^2-2\alpha^3+\alpha^4-2\alpha^3+4\alpha^3-2\alpha^4+\alpha^4-2\alpha^4+\alpha^4\right]\\
&\quad -2\alpha^3+2\alpha^4-2\alpha^3-2\alpha^4+2\alpha^4-2\alpha^4\\
&= \frac{1+\alpha}{1-\alpha}\left(\alpha^2-\alpha^4\right)-4\alpha^3 = \alpha^2(1-\alpha)^2 \quad \text{and, similarly,}\\
\sigma_{1,2}\gamma(0)(1-\alpha^2)^2 &= \frac{1+\alpha}{1-\alpha}\left[\alpha^3-2\alpha^4+\alpha^5-2\alpha^3+4\alpha^4-2\alpha^5+\alpha^4-2\alpha^4+\alpha^5\right]\\
&\quad +\alpha^3-4\alpha^4+3\alpha^5+4\alpha^4-4\alpha^5+\alpha^4+\alpha^5\\
&= \frac{1+\alpha}{1-\alpha}\left(\alpha^4-\alpha^3\right)+\alpha^3+\alpha^4 = 0.
\end{aligned}$$

Finally, if $i > 2$ and recalling the function $h(\cdot)$ of (2.20),

$$\begin{aligned}
\sigma_{1,i}\gamma(0)(1-\alpha^2)^2 &= \frac{1+\alpha}{1-\alpha}h(i+1)+\alpha^{i+1}[i-2(i-1)+(i-2)]\\
&\quad +\alpha^{i+2}(-2(i+1)+4i-2(i-1))+\alpha^{i+3}(i+2-2(i+1)+i) = 0,
\end{aligned}$$

recalling Property 3 (a) in Choi (1990), that $h(k) = 0$ for $k = p, p+1, \ldots$. The remaining entries of the matrix $\Sigma$ can be established from these calculations, as on the one hand, the expression is symmetric in $i$ and $j$, so that the above result hold for the entries $(i, 1)$ as well. On the other hand, letting $(a, b)$ be an arbitrary entry of the matrix with $b \geq a$, it is clear that

$$\begin{aligned}
&\sum_{d=0}^{\infty}\alpha^{\max\{a+1-u,b+1-r+d\}}+\sum_{d=1}^{\infty}\alpha^{\max\{b+1-r,a+1-u+d\}}\\
&= \alpha^a\left(\sum_{d=0}^{\infty}\alpha^{\max\{1-u,b-a+1-r+d\}}+\sum_{d=1}^{\infty}\alpha^{\max\{b-a+1-r,1-u+d\}}\right).
\end{aligned}$$

This implies that $\Sigma$ is a diagonal matrix with the asserted structure. □

We remark that this result coincides with the findings of Mills and Seneta (1991), even though their result in equation (6) looks a little different. This difference is due to a mistake in the calculation there, as can easily be seen by calculating their expression (6) from (4). We provided a theoretical foundation for the celebrated Quenouille's test, cf. Quenouille (1949). By our calculations, for large $T$ it holds approximately that

$$T\sum_{k=2}^{r}\frac{\widehat{\rho}_{\text{part}}(k)}{\widehat{\sigma}_{k,k}} \sim \chi_r^2,$$

where $\chi_r^2$ denotes the chi-squared distribution with $r$ degrees of freedom and where $\widehat{\sigma}_{k,k}$ denotes a consistent estimate of the matrix entry $\sigma_{k,k}$ as defined in Theorem 5.3.2.

# 6 Goodness-of-Fit Testing in Markovian Models

In Chapters 4 and 5 three specification tests for INAR(1) processes were introduced, which were designed to detect the deviation of a given data sample from the assumption of a Poisson INAR(1) model. In this chapter, a similar problem is addressed in a more general framework. We suppose that we are presented with a count data time series which exhibits an AR(1)-like structure in terms of ACF and PACF such as the motivating example in the introduction. However, we now refrain from assuming the structure of the underlying model to adhere to the recursion (1.1), and allow a general Markovian structure. For instance, we now include the INARCH(1) model, which assumes $Y_t$ to be conditionally Poisson distributed with

$$Y_t \mid Y_{t-1}, Y_{t-2}, \ldots \quad \sim \text{Poi}\left(\beta + \alpha \cdot Y_{t-1}\right) \text{ for all } t \in \mathbb{Z}. \tag{6.1}$$

The naming of these models evolved from Ferland et al. (2006), where a more general class of INGARCH(p,q) models was considered, one of the first instances of the name INARCH(1) process can be found in Weiß (2010). In many variations both in name and formulation, the INARCH(1) model of (6.1) has been previously considered, leading to many modifications. One such instance can be found in Zeger and Qaqish (1988), another one in Xu et al. (2012).

Other candidates for models describing count data and exhibiting an AR(1)-like structure of the ACF and PACF abound, for instance the Random Coefficient INAR(1) model of Zheng et al. (2007) or variations of INAR(1) models with different thinning operations as those of Weiß (2008). Yet, asides from these differences, all of the models just discussed satisfy

$$\mathbb{P}_\theta\left(Y_t = l \mid Y_{t-1} = k, Y_{t-2} = j, \ldots\right) = \mathbb{P}_\theta\left(Y_t = l \mid Y_{t-1} = k\right) := p_\theta(l|k), \tag{6.2}$$

where $\theta = (\theta_1, \ldots, \theta_d) \in \Theta \subseteq \mathbb{R}^d$, the parameter space and where the probability measure $\mathbb{P}_\theta$ is a function of the underlying parameter $\theta$. Additionally, the stationary distribution, which may be assumed to exist, will be denoted by $\mathbb{P}_\theta(Y_0 = k) := \pi_\theta(k)$ to emphasize the dependence on the parameter $\theta$. We are now interested in setting up a statistical test for the hypothesis

$$\mathbf{H}_0^{(s)} \ : f = p_{\theta_0} \quad \text{against} \quad \mathbf{H}_1^{(s)} \ : f \neq p_{\theta_0}, \tag{6.3}$$

where $\theta_0$ is the true parameter. Moreover, it would be even better to have at hand a test for the composite hypothesis

$$\mathbf{H}_0 \ : f \in \{p_\theta | \theta \in \Theta\} \quad \text{against} \quad \mathbf{H}_1 \ : f \notin \{p_\theta | \theta \in \Theta\}. \tag{6.4}$$

In two seminal contributions, Neumann (2011) and Fokianos and Neumann (2013), tests are established for both scenarios (6.3) and (6.4) in the context of Poisson count processes. The first article derives the asymptotic normality of certain test statistics, providing easily implemented goodness-of-fit tests. However, these test statistics are not suitable to detect local alternatives in general. In the second article, an alternative test statistic based on Pearson residuals is proposed which can be shown to have nontrivial power against local alternatives. Both works consider a nonlinear model allowing for a wide range of models, cf. eq. (1.1) in Fokianos and Neumann (2013). It should also be noted that both approaches can be extended to higher and even infinite order dependencies, see for instance Remark 4 in Neumann (2011).

Relaxing the distributional assumption of a Poisson count process is necessary to account for more general models. For instance, consider the NegBin-INARCH(1) model of Xu et al. (2012), where $Y_t$ follows the recursion (6.1) with a conditional Negative Binomial distribution. Another set of models exceeding this scope, amongst many others, is the Compound Poisson INAR(1) model studied in Schweer and Weiß (2014).

In Meintanis and Karlis (2014), a goodness-of-fit test based on the joint probability generating function (jpgf) for the INAR(1) process is suggested, note that Baringhaus and Henze (1992) considered a similar test in the case of i.i.d. count data. For a stationary process $(Y_t)_{t\in\mathbb{Z}}$ the jpgf is given by $\psi(u,v;\theta) := \mathbb{E}_\theta\left[u^{Y_0}v^{Y_1}\right]$ for all $u,v\in[0,1]$. For $T+1$ consecutive counts $\{y_1,\dots,y_{T+1}\}$, its empirical counterpart is based on the $T$ pairs of consecutive observations at hand, and is defined as

$$\widehat{\psi}_T(u,v) := \frac{1}{T}\sum_{i=1}^{T} u^{y_i}v^{y_{i+1}} \text{ for } u,v\in[0,1], \tag{6.5}$$

the so-called empirical joint probability generating function (ejpgf). One possible test statistic is then given by

$$W_{T,a}(y_1,\dots,y_{T+1};\theta) := T\int_0^1\int_0^1 \left(\widehat{\psi}_T(u,v)-\psi(u,v;\theta)\right)^2 u^a v^a\,du\,dv, \tag{6.6}$$

where $a\geq 0$ is a weighting factor. The null hypothesis (6.3) is rejected for large values of the statistic (6.6) with $\theta=\theta_0$, for the null hypothesis (6.4) with $\theta=\hat{\theta}_T$.

In this chapter, we derive the asymptotic properties of this statistic for a general class of Markovian models satisfying a drift condition, resulting in a goodness-of-fit test. Additionally, we show that there exists a surprising connection between this test and those of previous chapters. An application of the resulting test in the second part reveals the usefulness of this approach. This chapter is largely based on the article Schweer (2015a).

## 6.1 Introduction and Main Results

The following definition provides the framework for the processes under consideration in this chapter.

**Condition 1** (Drift condition).
*$(Y_t)_{t\in\mathbb{Z}}$ is a time-homogeneous, irreducible, aperiodic first order Markov chain on the sample space $\mathbb{N}_0$ with finite mean and satisfies (6.2) for some $\theta=(\theta_1,\dots,\theta_d)\in\Theta\subseteq\mathbb{R}^d$. Furthermore, there exists a function $V:\mathbb{N}_0\to\mathbb{R}$, a finite set $A\subset\mathbb{N}_0$, $b<\infty$ and $\delta\in(0,1)$ such that $V(x)\geq 1$ for all $x\in\mathbb{N}_0$ and*

$$\mathbb{E}_\theta\left[V(Y_t)|Y_{t-1}=y\right]\leq V(y)(1-\delta)+b\mathbf{1}_{\{y\in A\}} \textit{ for all } y\in\mathbb{N}_0.$$

A large class of processes satisfying the drift condition can be defined as follows, based on an idea by Grunwald et al. (2000): A process $(Y_t)_{t\in\mathbb{Z}}$ is called an integer-valued conditional linear autoregressive process of first order (INCLAR(1)) if it is a time-homogeneous, irreducible, aperiodic first-order Markov chain on $\mathbb{N}_0$ with finite mean satisfying

$$\mathbb{E}_\theta[Y_t|Y_{t-1}] = \alpha Y_{t-1} + \lambda \tag{6.7}$$

for some $\alpha \in (0,1)$, $\lambda > 0$. We now show that each INCLAR(1) process satisfies Condition 1.

**Lemma 6.1.1** (Proposition 3, Grunwald et al. (2000))**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be an INCLAR(1) process satisfying* (6.7)*. Then Condition 1 holds for $(Y_t)_{t\in\mathbb{Z}}$.*

*Proof.* Define the function $V(x) := x + 1$ and further choose $\delta > 0$ such that the set $A := \{0, 1, \dots, \lfloor(\lambda + \delta)/(1 - \delta - \alpha)\rfloor\}$ is not empty, where $\lfloor\cdot\rfloor$ denotes the floor function which maps $r$ to the highest integer value smaller than $r$. Now, it is easily seen that $y \in \mathbb{N}_0\backslash A$ implies $\lambda+1+\alpha y \leq (1-\delta)(y+1)$ and thus that $\mathbb{E}[V(Y_t)|Y_{t-1} = y] \leq V(y)(1-\delta)$ holds for $y \in \mathbb{N}_0 \setminus A$. It is clear that $\mathbb{E}[V(Y_t)|Y_{t-1} = y]$ is uniformly bounded for $y \in A$, concluding the proof. □

### 6.1.1 Asymptotics for the Empirical Joint Probability Generating Function

We begin this section by showing that Condition 1 implies the $\alpha$-mixing of the process involved. This result has as a consequence, that Theorem 4.2.6 can be extended to a more general class of processes.

**Theorem 6.1.2** (Schweer (2015a), Theorem 1)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a process satisfying Condition 1. Then $(Y_t)_{t\in\mathbb{Z}}$ is $\alpha$-mixing with exponentially decreasing weights $\alpha_Y(n)$.*

*Proof.* By Theorem 2 of Popov (1977), each process satisfying Condition 1 is geometrically ergodic. From this, the assertion follows with the proof of Theorem 4.2.6. □

One implication among many others of the preceding assertion is that (Poisson) INARCH(1) processes are $\alpha$-mixing with exponentially decreasing weights, this was previously shown to hold in Neumann (2011). The mixing property of Theorem 6.1.2 allows for the establishment of the asymptotic behavior of the ejpgf. In Theorem 6.1.6, both cases pertaining to the null hypotheses (6.3) and (6.4) are considered separately, this structure is reflected in the presentation of the necessary assumptions.

**Assumption (A1).**

*(i) The true data generating process satisfies Condition 1.*

*(ii) It holds that $\mathbb{E}_{\theta_0}[|Y_0Y_1|^{2+\xi}]$ for some $\xi > 0$.*

If the hypothesis is of a composite nature, it is necessary to estimate the parameter $\theta_0$. The following assumption ensures an appropriate behavior of both the estimator and the test statistic as the number of observations increases.

**Assumption (A2).**

*(i) The sequence of estimators* $\hat{\theta}_T$ *of* $\theta \in \Theta \subseteq \mathbb{R}^d$ *satisfies the expansion*

$$\hat{\theta}_T - \theta_0 = \frac{1}{T} \sum_{i=1}^{T} l(Y_i, Y_{i+1}; \theta_0) + \mathbf{r}_T,$$

*where* $\mathbf{r}_T = o_{\mathbb{P}}(T^{-\frac{1}{2}})$ *and where* $\mathbf{l}(\theta; i) = (l(\theta; i)_1, \dots, l(\theta; i)_d)$ *is a measurable function of* $(Y_{i+1}, Y_i, \dots)$ *and* $\theta$ *such that, for all* $k \in \{1, \dots, d\}$, $\mathbb{E}_{\theta_0}[l(\theta_0; i)_k] = 0$, $\mathbb{E}_{\theta_0}[l(\theta_0; i)_k^{2+\delta}] < \infty$ *for some* $\delta > 0$ *and, if* $(Y_t)_{t\in\mathbb{Z}}$ *is* $\alpha$*-mixing with exponentially decreasing weights, the same holds for* $(l(\theta_0; t)_k))_{t\in\mathbb{Z}}$.

*(ii) The function* $\psi(u, v; \theta)$ *as a function of* $\theta$ *is twice continuously differentiable for all* $u, v \in [0, 1]^2$ *at the point* $\theta_0$.

*(iii) The series* $\sum_{k,l=0}^{\infty} kl \frac{\partial}{\partial\theta} \mathbb{P}_{\theta_0}(Y_0 = k, Y_1 = l)$ *and* $\sum_{k,l=0}^{\infty} \frac{\partial^2}{\partial^2\theta} \mathbb{P}_{\theta'}(Y_0 = k, Y_1 = l)$ *converge, where* $\theta' \in \Theta$.

Assumption (i) can easily seen to be satisfied if $\mathbf{l}(\theta_0; i)$ is a function of a finite number of $Y_i$'s only. The usual estimation techniques for count data times series, i.e., conditional maximum likelihood, conditional least squares, Yule-Walker estimators and the moment estimators employed in Chapter 5 satisfy the latter assumption for many useful models. Part (ii) and (iii) constitute higher order assumptions on the process $(Y_t)_{t\in\mathbb{Z}}$ and its dependency on the parameter $\theta \in \Theta$. Given the structure of the proof, assumptions of this kind seem inevitable, thus the following two lemmata show they may be verified for specific models.

**Lemma 6.1.3** (Schweer (2015a), Example 1)**.** *Let* $(Y_t)_{t\in\mathbb{Z}}$ *be a Compound Poisson INAR(1) process as in Definition 4.2.2, where* $\epsilon_t \sim \text{ComPoi}(\lambda, H_\theta)$ *for all* $t \in \mathbb{Z}$*, such that* $H_\theta$ *is twice differentiable in the parameter* $\theta$ *and* $\frac{\partial^2}{\partial^2\theta} H_\theta(1 + \epsilon)$ *is finite for some* $\epsilon > 0$. *Then* $(Y_t)_{t\in\mathbb{Z}}$ *satisfies Assumptions (A2) (ii) and (iii).*

*Proof.* First of all, we calculate from (4.15)

$$\psi(u, v; \theta) = \mathbb{E}_\theta \left[u^{Y_0} v^{Y_1}\right] = \mathbb{E}_\theta \left[u^{Y_0} \mathbb{E}\left[v^{Y_1} \middle| Y_0\right]\right] = \text{pgf}_Y \left(u(1 - \alpha + \alpha v)\right) \text{pgf}_\epsilon(v),$$

which, by Theorem 4.2.5 may be rewritten as

$$\psi(u, v; \theta) = \exp\left(\lambda \sum_{i=0}^{\infty} [H_\theta(1 - \alpha^i + \alpha^i u(1 - \alpha + \alpha v)) - 1]\right) \exp(\mu(H_\theta(v) - 1)).$$

This immediately implies Assumptions (A2) (ii). For part (iii), a sufficient condition is the existence of an $\epsilon' > 0$ such that $\frac{\partial}{\partial\theta}\psi(1+\epsilon', 1+\epsilon'; \theta_0) < \infty$ and $\frac{\partial^2}{\partial^2\theta}\psi(1+\epsilon', 1+\epsilon'; \theta') < \infty$ for $\theta' \in \Theta$. Again, this follows from the equation above, as we may simply choose $\epsilon'$ so that $\epsilon'(1 + 2\alpha) < \epsilon$. $\square$

**Lemma 6.1.4** (Schweer (2015a), Example 1)**.** *Let* $(Y_t)_{t\in\mathbb{Z}}$ *be a Poisson INARCH(1) process as in* (6.1)*. Then* $(Y_t)_{t\in\mathbb{Z}}$ *satisfies Assumptions (A2) (ii) and (iii).*

*Proof.* All moments of $(Y_t)_{t\in\mathbb{Z}}$ are finite (cf. Ferland et al. (2006)), hence we may consider the joint mgf

$$\begin{aligned}\mathbb{E}_\theta\left[\exp\left(sY_0+tY_1\right)\right] &= \mathbb{E}_\theta\left[\exp\left(sY_0\right)\mathbb{E}\left[\exp\left(tY_1\right)\middle|Y_0\right]\right] \\ &= \exp\left(\beta\left(\exp(t)-1\right)\right)\mathbb{E}_\theta\left[\exp\left(Y_0(s+\alpha(\exp(t)-1))\right)\right], \qquad (6.8)\end{aligned}$$

where the second equality follows from (6.1) since $Y_1$ given $Y_0$ is Poisson distributed, and the mgf of $X\sim\text{Poi}(\lambda)$ satisfies $\text{mgf}_X(t)=\text{pgf}_X(e^t)=\exp(\lambda(e^t-1))$. The last term of (6.8) may be expressed in terms of the moments of $Y_0$ as $\sum_{i=0}^{\infty}\mathbb{E}_\theta[Y_0^i]x^i/i!$, where $x:=s+\alpha(\exp(t)-1)$. From Example 2 in Weiß (2009) it can easily be seen that for each $i\in\mathbb{N}$, $\mathbb{E}_\theta[Y_0^i]$ is a twice continuously differentiable function in $\theta=(\beta,\alpha)$. Now, substituting $s:=\log(u)$ and $t:=\log(v)$ in (6.8) proves that Assumption (A2) (ii) holds for fixed $u,v>0$, the special cases $u=0$ and $v=0$ can be dealt with by calculating $\psi(0,v;\theta)$ and $\psi(u,0;\theta)$ directly, using (6.1). Similarly, $\frac{\partial^2}{\partial^2\theta}\mathbb{E}_\theta[\exp(\epsilon(Y_0+Y_1))]$ can be seen to be finite for some $\epsilon>0$. As in Lemma 6.1.3, this suffices for (A2) (iii). ☐

Using Assumptions (A1) and (A2), the first result of this chapter can be shown, which provides a functional central limit theorem for the integrands of (6.6). This proof deals with convergence in distribution in the Banach space $C[0,1]^2$ of all continuous real functions $g$ on the square $[0,1]^2$, equipped with the uniform norm $\|g\|_\infty=\sup_{0\le u,v\le 1}|g(u,v)|$. As in the more commonly used space $C[0,1]$, convergence in distribution is shown by proving convergence of the finite-dimensional distributions and proving tightness of the sequences involved. Tightness criteria in $C[0,1]^k$ are hard to come by. The classical reference for such criteria is Bickel and Wichura (1971). However, their results are not applicable in the context of Theorem 6.1.6, as the condition of vanishing along the lower boundary does not hold for the processes under consideration. As a remedy, a generalization of Theorem 12.3. in Billingsley (1968) is given in the following lemma.

**Lemma 6.1.5** (Schweer (2015a), Lemma A.1)**.** *Let* $(X_n(u,v))_{n\in\mathbb{N}}$ *with* $u,v\in[0,1]$ *be a sequence of random elements of* $C[0,1]^2$*. Let there exist constants* $\gamma_0,\gamma_1,\gamma_2,\gamma_3>0$ *and* $\alpha_1,\alpha_2,\alpha_3>1$*, continuous increasing functions* $F_1,F_2$ *on* $[0,1]$ *and a finite nonnegative measure* $F_3$ *on* $[0,1]^2$ *with continuous marginals such that*

*(i)* $\mathbb{E}[|X_n(0,0)|^{\gamma_0}]<\infty$,

*(ii)* $\mathbb{E}[|X_n(u_2,0)-X_n(u_1,0)|^{\gamma_1}]\le|F_1(u_2)-F_1(u_1)|^{\alpha_1}$ *for all* $u_1,u_2\in[0,1]$,

*(iii)* $\mathbb{E}[|X_n(0,v_2)-X_n(0,v_1)|^{\gamma_2}]\le|F_2(v_2)-F_2(v_1)|^{\alpha_2}$ *for all* $v_1,v_2\in[0,1]$ *and*

*(iv)* $\mathbb{E}[|X_n(u_2,v_2)-X_n(u_2,v_1)-X_n(u_1,v_2)+X_n(u_1,v_1)|^{\gamma_3}]\le F_3([u_1,u_2]\times[v_1,v_2])^{\alpha_3}$ *for all* $u_1,u_2,v_1,v_2\in[0,1]$ *with* $u_1\le u_2$ *and* $v_1\le v_2$.

*Then the sequence* $(X_n(u,v))_{n\in\mathbb{N}}$ *is tight in* $C[0,1]^2$.

The proof is omitted, it follows immediately from Theorem 1 in Lachout (1988) together with the restatement in terms of moments, cf. eq. (12.51) in Billingsley (1968). Note that the first condition is in place to ensure that the probability of the sequence exceeding a given bound is low enough. For brevity, the presentation of this result is restricted to the 2-dimensional case. To simplify notation, we introduce the functions $a_k(u,v) := u^{Y_k}v^{Y_{k+1}} - \psi(u,v;\theta_0)$ and $b_k(u,v) := \mathbf{l}(\theta_0;k)\cdot(\frac{\partial}{\partial\theta}\psi(u,v;\theta_0))^\top$ for each $u,v\in[0,1]$, $k\in\mathbb{Z}$, where $\mathbf{l}(\cdot)$ is given as in Assumption (A2).

**Theorem 6.1.6** (Schweer (2015a), Theorem 2)**.**

*(i) Let Assumption (A1) be satisfied. Then for each* $(u_1,v_1),(u_2,v_2)\in[0,1]^2$*, the series*

$$\kappa_1(u_1,v_1;u_2,v_2) = \sum_{k\in\mathbb{Z}} \mathbb{E}_{\theta_0}\left[a_0(u_1,v_1)a_k(u_2,v_2)\right]$$

*converges absolutely. Furthermore,*

$$\sqrt{T}\left(\widehat{\psi}_T(u,v) - \psi(u,v;\theta_0)\right) \xrightarrow{\mathcal{D}} \Psi_1,$$

*a zero mean Gaussian element in* $C[0,1]^2$*, with covariance function* $\kappa_1(\cdot)$*.*

*(ii) Let Assumptions (A1) and (A2) be satisfied. Then for* $(u_1,v_1),(u_2,v_2)\in[0,1]^2$*, the series*

$$\kappa_2(u_1,v_1;u_2,v_2) = \sum_{k\in\mathbb{Z}} \mathbb{E}_{\theta_0}\left[\left(a_0(u_1,v_1)-b_0(u_1,v_1)\right)\left(a_k(u_2,v_2)-b_k(u_2,v_2)\right)\right]$$

*converges absolutely . Furthermore,*

$$\sqrt{T}\left(\widehat{\psi}_T(u,v) - \psi(u,v;\hat{\theta}_T)\right) \xrightarrow{\mathcal{D}} \Psi_2,$$

*a zero mean Gaussian element in* $C[0,1]^2$ *with covariance function* $\kappa_2(\cdot)$*.*

*Proof.* Let us begin with (i). Define

$$A_T(u,v) := \sqrt{T}(\widehat{\psi}_T(u,v) - \psi(u,v;\theta_0)) = \frac{1}{\sqrt{T}}\sum_{i=1}^{T} a_i(u,v),$$

by (6.5) and the stationarity of $(Y_t)_{t\in\mathbb{Z}}$,

$$\mathbb{E}_{\theta_0}\left[\left|A_T(0,0)\right|^2\right] = \mathrm{Var}_{\theta_0}\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\mathbf{1}_{\{Y_i=0,Y_{i+1}=0\}}\right). \tag{6.9}$$

Now, let $u_1,u_2\in[0,1]$ with $u_1<u_2$. The function $\sum_{k=0}^{\infty} k u^{k-1}\mathbb{P}_{\theta_0}(Y_0=k,Y_1=0)$ is a power series in $u$ which converges at $u=1$, as $\mathbb{E}[Y_0\mathbf{1}_{\{Y_{i+1}=0\}}]$ is finite. Thus, Abel's uniform convergence test ensures that summation and derivation may be interchanged,

yielding $\frac{\partial}{\partial u}\psi(u,0;\theta_0) = \mathbb{E}_{\theta_0}[Y_0 u^{Y_0-1}\mathbf{1}_{\{Y_{i+1}=0\}}]$. For the function $\widehat{\psi}_T(u,0)$ similar arguments apply. It follows that $A_T(u,0)$ is continuously differentiable in $u$ on the interval $[0,1]$, and the mean value theorem yields the existence of a $c \in (u_1,u_2)$ such that $(A_T(u_2,0) - A_T(u_1,0))/(u_2-u_1) = \frac{\partial}{\partial u}A_T(c,0)$. It follows that

$$\mathbb{E}_{\theta_0}\left[|A_T(u_2,0) - A_T(u_1,0)|^2\right] \leq |u_2-u_1|^2 \sup_{c_1\in[0,1]} \operatorname{Var}_{\theta_0}\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{T} Y_i c_1^{Y_i-1}\mathbf{1}_{\{Y_{i+1}=0\}}\right). \tag{6.10}$$

An analogous expression can be obtained for $\mathbb{E}_{\theta_0}[|A_T(0,v_2) - A_T(0,v_1)|^2]$. Now, let $u_1,u_2,v_1,v_2 \in [0,1]$, with $u_1 < u_2, v_1 < v_2$. Since $\mathbb{E}_{\theta_0}[Y_0Y_1]$ is bounded by assumption, Abel's uniform convergence test implies $\frac{\partial^2}{\partial u \partial v}\psi(u,v;\theta_0) = \mathbb{E}_{\theta_0}[\frac{\partial^2}{\partial u\partial v}u^{Y_0}v^{Y_1}]$, similar arguments apply to $\widehat{\psi}_T(u,v)$. Hence, using the mean value theorem twice (first in the first, then in the second argument of the function),

$$\begin{aligned}&\mathbb{E}_{\theta_0}\left[|A_T(u_2,v_2) - A_T(u_2,v_1) - A_T(u_1,v_2) + A_T(u_1,v_1)|^2\right]\\ &\leq |v_2-v_1|^2\,|u_2-u_1|^2 \sup_{c_2,c_3\in[0,1]} \operatorname{Var}_{\theta_0}\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{T} Y_iY_{i+1}c_2^{Y_i-1}c_3^{Y_{i+1}-1}\right).\end{aligned} \tag{6.11}$$

For any fixed $c_1,c_2,c_3 \in [0,1]$, the stationary sequences appearing in the variances above, $(\mathbf{1}_{\{Y_t=0,Y_{t+1}=0\}})_{t\in\mathbb{Z}}$ in (6.9), $(Y_t c^{Y_t-1}\mathbf{1}_{\{Y_{t+1}=0\}})_{t\in\mathbb{Z}}$ in (6.10) and the third sequence, $(Y_tY_{t+1}c_2^{Y_t-1}c_3^{Y_{t+1}-1})_{t\in\mathbb{Z}}$ in (6.11) are all functions of $(Y_t)_{t\in\mathbb{Z}}$ and thus $\alpha$-mixing with exponentially decreasing weights, cf. Lemma 6.1.2. Now, each element of these sequences has finite moments of order at least $2+\xi$ for some $\xi > 0$. For the first sequence, this is obvious; for the second and third use $\mathbb{E}_{\theta_0}[|Y_tY_{t+1}c_2^{Y_t-1}c_3^{Y_{t+1}-1}|^{2+\xi}] \leq \mathbb{E}_{\theta_0}[|Y_tY_{t+1}|^{2+\xi}]$ for any $c_2,c_3 \in [0,1]$ and employ the assumption. Using Theorem 2.5.2, it follows that all of these variances are uniformly bounded for all $T \in \mathbb{N}$ and all $c_2,c_3 \in [0,1]$, and with Lemma 6.1.5 it follows that the sequence $A_T$ is tight on $C[0,1]^2$.

For the finite-dimensional distributions, let $l \in \mathbb{N}$ and $(u_1,v_1),\ldots,(u_l,v_l) \in [0,1]^2$ and let $r_1,\ldots,r_l \in \mathbb{R}$. Then the sequence $\sum_{j=1}^{l} r_j a_i(u_j,v_j)$ is $\alpha$-mixing with exponentially decreasing weights and it has finite moments of order at least $2+\xi$ by assumption. Theorem 2.5.2 yields asymptotic normality of the random variable $\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\sum_{j=1}^{l} r_j a_i(u_j,v_j)$, and application of the Cramér-Wold device shows the convergence of the finite-dimensional distributions. The asserted covariance structure follows immediately, cp. the proof of Theorem 3.2.4.

For assertion (ii), first define

$$B_T(u,v) := \frac{1}{\sqrt{T}}\sum_{i=1}^{T} b_i(u,v),$$

then with the Cauchy-Schwarz inequality,

$$\mathbb{E}_{\theta_0}[|B_T(u,v)|^2] \leq \left\|\frac{\partial}{\partial\theta}\psi(u,v;\theta_0)\right\|^2 \sum_{k=1}^{d}\mathbb{E}_{\theta_0}\left[\frac{1}{T}\left(\sum_{i=1}^{T} l(\theta_0;i)_k\right)^2\right],$$

where $\|\cdot\|$ denotes the Euclidean norm. By Theorem 6.1.2, $(Y_t)_{t\in\mathbb{Z}}$ is $\alpha$-mixing with exponentially decreasing weights and due to the assumptions on $\mathbf{l}(\cdot)$, Theorem 1.7 in Ibragimov (1962) is applicable. This shows that the summands converge for $T\to\infty$ and are uniformly bounded in $T$ by $C_1 := \sum_{k=1}^{d}\sum_{t\in\mathbb{Z}}\mathbb{E}_{\theta_0}[l(\theta_0;0)_k l(\theta_0;t)_k]$, which is finite. Hence, the finite-dimensional distributions can be shown to converge in an analogous fashion as in part (a) of this theorem.

With Abel's uniform convergence criterion, the assumption implies uniform convergence of $\sum_{k,l=0}^{\infty} ku^{k-1}lv^{l-1}\frac{\partial}{\partial\theta}\mathbb{P}_{\theta_0}(Y_0=k,Y_1=l)$ on $[0,1]^2$. Thus, summation and differentiation may be exchanged, yielding that the function $\frac{\partial^3}{\partial u\partial v\partial\theta}\psi(u,v;\theta_0)$ is continuous in $u,v$ on $[0,1]^2$, similarly for $\frac{\partial^2}{\partial u\partial\theta}\psi(u,v;\theta_0)$ and $\frac{\partial^2}{\partial v\partial\theta}\psi(u,v;\theta_0)$. As a first consequence, $\mathbb{E}_{\theta_0}[|B_T(0,0)|^2]\le C_1\|\frac{\partial}{\partial\theta}\psi(0,0;\theta_0)\|^2<\infty$. Application of the mean value theorem as in the proof of Theorem 6.1.6 (a) implies

$$\frac{\mathbb{E}_{\theta_0}[|B_T(u_2,0)-B_T(u_1,0)|^2]}{C_1|u_2-u_1|^2}\le \sup_{c_4\in[0,1]}\left\|\frac{\partial^2}{\partial u\partial\theta}\psi(c_4,0;\theta_0)\right\|^2.$$

The supremum is finite, as it is taken over a compact set. It is obvious that an analogous result can be shown for $\mathbb{E}_{\theta_0}[|B_n(0,v_2)-B(0,v_1)|^2]$. Now, for the last condition of Lemma 6.1.5, applying the mean value theorem twice yields

$$\begin{aligned}&\mathbb{E}_{\theta_0}\left[|B_T(u_2,v_2)-B_T(u_2,v_1)-B_T(u_1,v_2)+B_T(u_1,v_1)|^2\right]\\ &\quad\le C_1\,|u_2-u_1|^2\,|v_2-v_1|^2\sup_{c_5,c_6\in[0,1]}\left\|\frac{\partial^3}{\partial u\partial v\partial\theta}\psi(c_5,c_6;\theta_0)\right\|^2,\end{aligned}$$

implying tightness of the sequence. Taylor's theorem, applicable to the function $\psi(u,v;\theta)$ w.r.t. the parameter $\theta$ by assumption, yields the existence of a random vector $\tilde\theta\in\Theta$ on the line segment between $\theta_0$ and $\hat\theta_T$ such that

$$\begin{aligned}&\psi(u,v;\hat\theta_T)\\ &=\psi(u,v;\theta_0)+(\hat\theta_T-\theta_0)\left(\frac{\partial}{\partial\theta}\psi(u,v;\theta_0)\right)^\top+\frac{1}{2}(\hat\theta_T-\theta_0)\frac{\partial^2}{\partial^2\theta}\psi(u,v;\tilde\theta)(\hat\theta_T-\theta_0)^\top.\end{aligned}$$

Abel's uniform convergence criterion shows that both $\frac{\partial}{\partial\theta}\psi(u,v;\theta_0)$ and $\frac{\partial^2}{\partial^2\theta}\psi(u,v;\theta')$ are bounded on $[0,1]^2$. Since the estimator $\hat\theta_T$ is $\sqrt{T}$-consistent, Slutsky's lemma implies

$$\mathbb{P}_{\theta_0}\left(\sup_{(u,v)\in[0,1]^2}\left|\mathbf{r}_T\left(\frac{\partial}{\partial\theta}\psi(u,v;\theta_0)\right)^\top+\frac{\sqrt{T}}{2}(\hat\theta_T-\theta)\frac{\partial^2}{\partial^2\theta}\psi(u,v;\tilde\theta)(\hat\theta_T-\theta)^\top\right|\ge\epsilon\right)\to 0.$$

for $T\to\infty$, recall the definition of $\mathbf{r}_T$ in Assumption (A2) (i). Both of these results together show that $\sqrt{T}(\widehat{\psi}_T(u,v)-\psi(u,v;\hat\theta_T))=A_T(u,v)-B_T(u,v)+o_{\mathbb{P}}(1)$ holds. Since the sum of two tight sequences is tight again, this concludes the proof. □

It should be noted that in the special cases of INAR(1) and Poisson INARCH(1) processes, a similar result as that of Corollary 6.1.7 is shown in Hudecová et al. (2015), Theorem 4.1 and Theorem 4.4, respectively, with two differences: First, the authors use the particular structure of these models to obtain semi-parametric estimates of the marginal pgf under the null hypothesis, whereas here the structure is nonparametric, rendering the result viable in a much more general setting. Second, the authors consider the one-dimensional empirical probability generating function (epgf) instead of the (two-dimensional) ejpgf. For further comparisons of these two approaches within a simulation study, the reader is referred to Section 6.4.3.

In view of establishing asymptotic distributional results, Theorem 6.1.6 immediately implies the convergence in distribution of statistics (6.6) to $\int_0^1\int_0^1 \Psi_1(u,v)^2 u^a v^a dudv$ and $\int_0^1\int_0^1 \Psi_2(u,v)^2 u^a v^a dudv$, respectively. It is well known (cf. for instance the paragraph preceding Theorem 2.1 in Baringhaus and Henze (1992)) that the distribution of the latter integral coincides with that of $\sum_{j\geq 1}\lambda_{1;j}Z_{1;j}^2$, where the $Z_{1;j}$'s are independent standard normal random variables. The $\lambda_{1;j}$'s are the eigenvalues of the integral operator associated with the covariance function $\kappa_1(u_1,v_1;u_2,v_2)(u_1v_1)^a(u_2v_2)^a$, i.e.

$$\int_0^1\int_0^1 \kappa_1(u_1,v_1;u,v)(u_1v_1)^a(uv)^a g_{1;j}(u,v)dudv = \lambda_{1;j}g_{1;j}(u_1,v_1), \text{ for } u,v\in[0,1]. \tag{6.12}$$

Here, $g_{1;j}(u,v)$ denotes a continuous eigenfunction for each $j\in\mathbb{N}$, defined on $[0,1]^2$. This leads to the following result.

**Corollary 6.1.7** (Schweer (2015a), Corollary 3)**.**

*(i) Let Assumption (A1) be satisfied. Then there exist eigenvalues, denoted by $\lambda_{1;j}$, satisfying* (6.12) *for the covariance function $\kappa_1(u_1,v_1;u_2,v_2)(u_1v_1)^a(u_2v_2)^a$ such that*

$$W_{T,a}(Y_1,\dots,Y_{T+1};\theta_0) \xrightarrow{\mathcal{D}} \sum_{j\geq 1}\lambda_{1;j}Z_{1;j}^2,$$

*where the $Z_{1;j}$'s denote independent standard normal random variables and where the series converges in the $L^2$-sense.*

*(ii) Let Assumptions (A1) and (A2) be satisfied. Then there exist eigenvalues, denoted by $\lambda_{2;j}$, satisfying* (6.12) *for the covariance function $\kappa_2(u_1,v_1;u_2,v_2)(u_1v_1)^a(u_2v_2)^a$ such that*

$$W_{T,a}(Y_1,\dots,Y_{T+1};\hat{\theta}_T) \xrightarrow{\mathcal{D}} \sum_{j\geq 1}\lambda_{2;j}Z_{2;j}^2$$

*where the $Z_{2;j}$'s denote independent standard normal random variables and where the series converges in the $L^2$-sense.*

The eigenvalues present in the formulation of Corollary 6.1.7 depend on the particular transition probabilities given by the model (6.2) as well as on the unknown parameter $\theta_0$. We strongly believe that the explicit calculation of these eigenvalues is either extremely difficult or entirely impossible. Hence, the use of a resampling technique is advised in order to establish quantiles of the asymptotic distributions of the statistics given in Corollary 6.1.7. Since the underlying structure of the processes considered is a parametric one, a parametric bootstrap approach is employed, see Section 6.3.

### 6.1.2 Asymptotics for the Empirical CDF of the Stationary Distribution

Turning towards the possible application of these results, it is clear that one obstacle is given by the necessity to calculate $\mathbb{E}_\theta[u^{Y_0}v^{Y_1}] = \sum_{k,l=0}^{\infty} u^k v^l \pi_\theta(k) p_\theta(l|k)$. For models satisfying (6.2), the transition probabilities $p_\theta(k,l)$ are readily available. Unfortunately, the same does not hold true for the stationary distribution $\pi_\theta(k)$ or its cdf $\Pi_\theta(x) = \sum_{k=0}^{x} \pi_\theta(k)$. There are notable exceptions to this rule, see for instance Theorem 4.2.5. However, there are no general results available for the stationary distribution of processes satisfying Condition 1. One possibility to circumvent this problem is to estimate the stationary distribution nonparametrically. The necessary distributional theory is provided by the following Theorem 6.1.8 in the form of a functional central limit theorem. As this theorem is dealing with discrete distributions, we consider the space $c_0$ of (2.1), cf. Section 2.1.

**Theorem 6.1.8** (Schweer (2015a), Theorem 3)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a process satisfying Condition 1 with parameter $\theta \in \Theta$ and assume that $\mathbb{E}_\theta[Y_0^{2+\xi}] < \infty$, where $\xi > 0$. Then for all $x, y \in \mathbb{N}_0$, the sequences $\sigma_{xy}^2 = \sum_{i\in\mathbb{Z}} [\mathbb{P}_\theta(Y_0 \le x, Y_i \le y) - \Pi_\theta(x)\Pi_\theta(y)]$ converge and there exists a zero mean Gaussian element $\Xi$ of $c_0$ such that $\mathbb{E}_\theta[\Xi(x)\Xi(y)] = \sigma_{xy}^2$. Furthermore,*

$$\sqrt{T}\left(\frac{1}{T}\sum_{i=1}^{T} \mathbf{1}_{\{Y_i \le x\}} - \Pi_\theta(x)\right) \xrightarrow{\mathcal{D}} \Xi.$$

*Proof.* $(Y_t)_{t\in\mathbb{Z}}$ is $\alpha$-mixing with exponentially decreasing weights by Theorem 6.1.2, implying the same characteristic for the sequence $(\mathbf{1}_{\{Y_t\le x\}})_{t\in\mathbb{Z}}$ for any $x \in \mathbb{N}_0$. The moments of this sequence are trivially bounded, so that the CLT of Theorem 2.5.2 is applicable. Similar to the proof of Theorem 6.1.6, this result is extended to the multivariate case, proving the convergence of the finite-dimensional distributions.

Lemma 2.1.3 provides two sufficient and necessary conditions for tightness in $c_0$. The first condition is satisfied since $\sqrt{T}\left(\frac{1}{T}\sum_{i=1}^{T} \mathbf{1}_{\{Y_i\le x\}} - \Pi_\theta(x)\right)$ converges in distribution to a normal distribution with zero mean and bounded variance for each $x \in \mathbb{N}_0$. The second condition necessitates that for each positive numbers $\delta, \epsilon$ there exist integers $T_0$ and $l_0$ such that

$$\mathbb{P}_\theta\left(\sup_{k\ge l_0}\sqrt{T}\left|\frac{1}{T}\sum_{i=1}^{T}\mathbf{1}_{\{Y_i\le k\}} - \Pi_\theta(k)\right| > \epsilon\right) \le \delta \quad \text{for all } T \ge T_0. \tag{6.13}$$

Since $(\mathbf{1}_{\{Y_t\le k\}})_{t\in\mathbb{Z}}$ is $\alpha$-mixing, Davydov's inequality (cf. eq. (2.2) in Davydov (1968)) implies the upper bound

$$\left|\mathrm{Cov}_\theta\left(\mathbf{1}_{\{Y_0\le k\}}, \mathbf{1}_{\{Y_i\le k\}}\right)\right| \le 12\mathbb{E}_\theta\left[\left|\mathbf{1}_{\{Y_0\le k\}} - \Pi_\theta(k)\right|^4\right]^{\frac{1}{2}} \alpha(i)^{\frac{1}{2}} \le 12(1-\Pi_\theta(k))^{\frac{1}{2}}\alpha(i)^{\frac{1}{2}}.$$

Now, let $\delta, \epsilon > 0$. With Markov's inequality,

$$\mathbb{P}_\theta\left(\sup_{k\geq l_0}\sqrt{T}\left|\frac{1}{T}\sum_{i=1}^{T}\mathbf{1}_{\{Y_i\leq k\}}-\Pi_\theta(k)\right|>\epsilon\right)\leq\sum_{k\geq l_0}^{\infty}\mathbb{P}_\theta\left(\sqrt{T}\left|\frac{1}{T}\sum_{i=1}^{T}\mathbf{1}_{\{Y_i\leq k\}}-\Pi_\theta(k)\right|>\epsilon\right)$$
$$\leq\frac{1}{\epsilon^2}\sum_{k\geq l_0}^{\infty}\mathrm{Var}_\theta\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\mathbf{1}_{\{Y_i\leq k\}}\right)\leq\frac{1}{\epsilon^2}\left(1+24\sum_{i=1}^{T-1}\frac{T-i}{T}\sqrt{\alpha(i)}\right)\sum_{k\geq l_0}^{\infty}\sqrt{1-\Pi_\theta(k)},$$

the last inequality follows by Davydov's inequality and the stationarity of $(\mathbf{1}_{\{Y_t\leq k\}})_{t\in\mathbb{Z}}$. The former series converges as the weights $\alpha(i)$ are exponentially decreasing by Theorem 6.1.2, the latter series $\sum_{k\geq 1}^{\infty}\sqrt{1-\Pi_\theta(k)}$ converges under the assumption of existing moments up to order $2+\xi$ for the distribution $\Pi_\theta$ (cf. Lemma 3.2.3). Thus it is possible to chose a $l_0$ large enough so that the RHS of the expression above is less than $\delta$, proving (6.13) and rendering Lemma 2.1.3 applicable. □

## 6.2 Auxiliary Results

In this section, several results extending the scope of Corollary 6.1.7 and Theorem 6.3.2 are recorded. In Section 3 of Meintanis and Karlis (2014), the authors point out that letting $a$ grow large in (6.6) means putting more weight on higher values of $u, v$. Since the moments of a random variable can be recovered via the one-sided derivative of the pgf evaluated at points tending to one, it is expected that the limiting behavior of the statistics should be connected to deviations of the corresponding moment estimators. In the following two results, two different types of such connections are proven.

As the integrands in the expression (6.6) are two-dimensional functions, the following generalization of Proposition 1.1. of Baringhaus et al. (2000) will be employed. The proof is essentially the same as that of Corollary 1a, Ch. V of Widder (1946).

**Proposition 6.2.1** (Cp. Baringhaus et al. (2000), Proposition 1.1)**.** *Let the function $g : [0,\infty)\times[0,\infty)\to\mathbb{R}$ be measurable and integrable over compact subsets and let $\int_0^\infty\int_0^\infty g(s,t)e^{-as}e^{-at}dsdt$ be finite for each $a > 0$. Denote the Gamma function by $\Gamma(\cdot)$. If for some $\gamma_1, \gamma_2 \geq 0$ and some real constant $A$*

$$\lim_{s,t\to 0} g(s,t)\Gamma(\gamma_1+1)\Gamma(\gamma_2+1)s^{-\gamma_1}t^{-\gamma_2} = A,$$

*then*

$$\lim_{a\to\infty} a^{\gamma_1+\gamma_2+2}\int_0^\infty\int_0^\infty g(s,t)e^{-as}e^{-at}dsdt = A.$$

*Proof.* First, we note that by the definition of the Gamma function and integration by substitution,

$$\Gamma(\gamma+1)=\int_0^\infty e^{-t}t^\gamma dt = a^{\gamma+1}\int_0^\infty e^{-at}t^\gamma dt \quad \text{for } a \geq 0. \tag{6.14}$$

This immediately shows that

$$
\begin{aligned}
&a^{\gamma_1+\gamma_2+2}\int_0^\infty\int_0^\infty g(s,t)e^{-as}e^{-at}dsdt - A \\
&= a^{\gamma_1+\gamma_2+2}\int_0^\infty\int_0^\infty e^{-as}e^{-at}\left(g(s,t) - \frac{As^{\gamma_1}t^{\gamma_2}}{\Gamma(\gamma_1+1)\Gamma(\gamma_2+1)}\right)dsdt. \qquad (6.15)
\end{aligned}
$$

Let us separate the latter integral in several parts. With $T > 0$, the integral over $[0,T]\times[0,T]$ is bounded by

$$
\sup_{0\le t,s\le T}\left(g(s,t)\Gamma(\gamma_1+1)\Gamma(\gamma_2+1)s^{-\gamma_1}t^{-\gamma_2} - A\right)\cdot\int_0^T\int_0^T\frac{e^{-as}(as)^{\gamma_1}e^{-at}(at)^{\gamma_2}}{\Gamma(\gamma_1+1)\Gamma(\gamma_2+1)}dsdt,
$$

the latter integral is bounded by 1 due to (6.14). Concerning the remaining integrals of 6.14, the assumption implies that $g(s,t) = o(e^{\epsilon_1 s+\epsilon_2 t})$ for $\epsilon_1,\epsilon_2 > 0$ by Theorem 2.2(a) in Widder (1946). This shows that all of these integrals can be bounded by expressions of the form

$$
M_1\frac{a^{\gamma_1+\gamma_2+1}}{a-\epsilon_1}e^{(\epsilon_1-a)T},
$$

which converge to 0 as $a\to\infty$. In summary, the expression (6.15) is bounded the supremum $\sup_{0\le t,s\le T}\left(g(s,t)\Gamma(\gamma_1+1)\Gamma(\gamma_2+1)s^{-\gamma_1}t^{-\gamma_2} - A\right)$ independently of $T$, so letting $T\to 0$ concludes the proof. □

In order to see why the preceding result is of use when dealing with statistics of the form (6.6), perform the substitution $s := -\log(u)$ and $t := -\log(v)$ which yields

$$
\begin{aligned}
&W_{T,a}(y_1,\ldots,y_{T+1};\theta) \\
&= T\int_0^\infty\int_0^\infty\left(\widehat{\psi}_T\left(e^{-s},e^{-t}\right) - \psi\left(e^{-s},e^{-t};\theta\right)\right)^2 e^{-(a+1)s}e^{-(a+1)t}dsdt.
\end{aligned}
$$

### 6.2.1 Connection to the Index of Dispersion

Under the assumption of $(Y_t)_{t\in\mathbb{Z}}$ being a Poisson INAR(1) process (see (1.1)), two parameters are of interest, the mean of the observations $\lambda$ and the thinning operator $\alpha$. Since the marginal distribution of a Poisson INAR(1) process is Poisson distributed by Lemma 4.1.4, it holds that $\lambda\alpha = \mathrm{Cov}_\theta(Y_0,Y_1) =: \gamma(1)$, so that a different choice for the governing parameter is possible, $\theta = (\lambda,\gamma(1))$. Now, recalling that the data is implicitly assumed to be in pairs of the form $(Y_1,Y_2),(Y_2,Y_3)$ and so forth, there are actually two estimators for $\lambda$, given by $\hat\lambda_{T,1} := \frac{1}{T}\sum_{i=1}^T Y_i$ and $\hat\lambda_{T,2} := \frac{1}{T}\sum_{i=1}^T Y_{i+1}$, corresponding to the respective position of the data in the pair of observations. For $\gamma(1)$, the Yule-Walker estimator is given by $\hat\gamma_T(1) := (\sum Y_iY_{i+1})/T - \hat\lambda_{T,1}\hat\lambda_{T,2}$. This meticulousness in the choice of estimators is necessary, because the result of Proposition 6.2.2 does not deal with asymptotic results but holds for finite $T$. Other choices of estimators for this situation only lead to comparable results when letting $T$ grow at at a rate proportional to $a$, cp. the statement and assertion of Proposition 6.2.3. Throughout this section, let $R_a(s,t)$ denote a generic, infinite collection of terms, each of which is in $O(s^xt^y)$ for some $x,y$ satisfying $x+y>a$.

**Proposition 6.2.2** (Schweer (2015a), Proposition 4.1). *Let $\hat{\theta}_T = (\hat{\lambda}_{T,1}, \hat{\lambda}_{T,2}, \hat{\gamma}_T(1))$ be given as above and let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process. Then*

$$\lim_{a\to\infty} a^6 W_{T,a}\left(Y_1,\dots,Y_{T+1};\hat{\theta}_T\right) = \frac{14}{T}\left[\sum_{i=1}^{T} Y_i^2 - \sum_{i=1}^{T} Y_i - \frac{1}{T}\left(\sum_{i=1}^{T} Y_i\right)^2\right]^2 + O\left(\frac{1}{T}\right).$$

*Proof.* Taking hints from the proof of Theorem 4.1. in Baringhaus et al. (2000), first write

$$\widehat{\psi}_T(e^{-s},e^{-t}) = \frac{1}{T}\sum_{i=1}^{T} e^{-sY_i}e^{-tY_{i+1}} = \sum_{\vartheta_1,\vartheta_2=0}^{\infty} \frac{(-1)^{\vartheta_1+\vartheta_2}}{\vartheta_1!\vartheta_2!}\left(\frac{1}{T}\sum_{i=1}^{T} Y_i^{\vartheta_1}Y_{i+1}^{\vartheta_2}\right)s^{\vartheta_1}t^{\vartheta_2},$$

similarly for $\psi(e^{-s},e^{-t};\hat{\theta}_T)$. In the resultant expression for $\widehat{\psi}_T(e^{-s},e^{-t})-\psi(e^{-s},e^{-t};\hat{\theta}_T)$, it is clear that the coefficient of the term $s^0t^0$ is zero. The coefficients of the terms $s$, $t$ and $st$ are zero by the definition of $\hat{\lambda}_{T,1}$, $\hat{\lambda}_{T,2}$ and $\hat{\gamma}_T(1)$, respectively. For $\vartheta_1 = 2, \vartheta_2 = 0$, define $S_{Y,T;1} := \frac{1}{T}\sum_{i=1}^{T} Y_i^2 - \mathbb{E}_{\hat{\theta}_T}[Y_0^2]$. Since $\mathbb{E}_{\hat{\theta}_T}[Y_0^2] = \hat{\lambda}_{T,1}^2 + \hat{\lambda}_{T,1}$ it holds that $S_{Y,T;1} \neq 0$ with positive probability, similarly comments apply for the case $\vartheta_1 = 0, \vartheta_2 = 2$. Hence,

$$\left(\widehat{\psi}_T(e^{-s},e^{-t}) - \psi(e^{-s},e^{-t};\hat{\theta}_T)\right)^2 e^{-s}e^{-t} = \left[\frac{1}{2}S_{Y,T;1}s^2 + \frac{1}{2}S_{Y,T;2}t^2 + R_2(s,t)\right]^2 e^{-s-t}$$
$$= \frac{1}{4}S_{Y,T;1}^2 s^4 + \frac{1}{4}S_{Y,T;2}^2 t^4 + \frac{1}{2}S_{Y,T;1}S_{Y,T;2}s^2t^2 + R_4(s,t).$$

It follows with Proposition 6.2.1, that

$$\lim_{a\to\infty} Ta^6 \int_0^\infty\int_0^\infty \left(\widehat{\psi}_T(e^{-s},e^{-t}) - \psi(e^{-s},e^{-t};\hat{\theta}_T)\right)^2 e^{-s}e^{-t}e^{-as-at}dsdt$$
$$= T\frac{\Gamma(5)}{4}\left(S_{Y,T;1}^2\right)^2 + T\frac{\Gamma(5)}{4}\left(S_{Y,T;2}^2\right)^2 + T\frac{\Gamma(3)^2}{2}\left(S_{Y,T;1}^2\right)\left(S_{Y,T;2}^2\right).$$

Since $S_{Y,T;1}^2 = S_{Y,T;2}^2 + O(1/T)$, this concludes the proof. □

The connection between this limiting expression and the index of dispersion of $(Y_t)_{t\in\mathbb{Z}}$ and its empirical counterpart of (5.2) can now be evaluated directly. From Proposition 6.2.2 and some algebra,

$$\lim_{a\to\infty} a^6 W_{T,a}\left(Y_1,\dots,Y_{T+1};\hat{\theta}_T\right) = 14\left(\frac{1}{T}\sum_{i=1}^{T} Y_i\right)^2 \cdot \left(\sqrt{T}(\hat{I}_Y - 1)\right)^2 + O\left(\frac{1}{T}\right).$$

The latter expression converges almost surely to a constant by the ergodicity of $(Y_t)_{t\in\mathbb{Z}}$, hence this expression showcases the connection between the goodness-of-fit test in this special case and the test for Poissonity established in Corollary 5.1.3.

### 6.2.2 Testing for Time-Reversibility

In general, the characteristic of a stochastic process being time-reversible is not common, yet most of the popular models share it. For instance, in continuous time it is well known that an AR(p) process is time-reversible if and only if the innovation distribution is Gaussian. We show a similar result to hold for AAINAR(p) processes in 7.1.2 with the Poisson distribution replacing the Gaussian distribution. Using this, a test for time-reversibility could be used as a validation test for a Poisson INAR(1) process.

If a stochastic process $(Y_t)_{t\in\mathbb{Z}}$ is time-reversible, it follows immediately that the jpgf is symmetric (cf. Section 2.4), i.e., $\psi(u,v;\theta_0) = \psi(v,u;\theta_0)$. Hence, it would be sensible to reject the null hypothesis of $(Y_t)_{t\in\mathbb{Z}}$ being a time-reversible process if the (nonparametric) statistic

$$V_{T,a}(Y_1,\ldots,Y_{T+1}) := T\int_0^1\int_0^1 \left(\widehat{\psi}_T(u,v) - \widehat{\psi}_T(v,u)\right)^2 u^a v^a du dv,$$

exceeds a certain critical value. It should be noted that under the null hypothesis,

$$\widehat{\psi}_T(u,v) - \widehat{\psi}_T(v,u) = (\widehat{\psi}_T(u,v) - \psi(u,v;\theta_0)) - (\widehat{\psi}_T(v,u) - \psi(u,v;\theta_0)), \tag{6.16}$$

so that the results of Section 6.1.1 and Section 6.3 are (almost) directly applicable. Furthermore, the following result highlights the connection between the statistic and the moment-based criterion of (5.6), cf. (2.10).

**Proposition 6.2.3** (Schweer (2015a), Proposition 4.2)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a time-reversible process satisfying Condition 1 and let all moments of $Y_0$ exist. Let $\hat{\beta}_T(1)$ denote the estimator defined in* (2.11) *and let $\frac{a^4}{T} \to 0$ as $a \to \infty$. Then*

$$\lim_{a,T\to\infty} a^8 V_{T,a}(Y_1,\ldots,Y_{T+1}) = \lim_{T\to\infty} 6\left(\sqrt{T}\cdot\hat{\beta}_T(1)\right)^2.$$

*Proof.* Similar to the proof of Proposition 6.2.2, note that

$$\widehat{\psi}_T\left(e^{-s},e^{-t}\right) - \widehat{\psi}_T\left(e^{-t},e^{-s}\right) = \sum_{\vartheta_1,\vartheta_2=0}^{\infty} \frac{(-1)^{\vartheta_1+\vartheta_2}}{\vartheta_1!\vartheta_2!}\left[\frac{1}{T}\sum_{i=1}^{T}\left(Y_i^{\vartheta_1}Y_{i+1}^{\vartheta_2} - Y_i^{\vartheta_2}Y_{i+1}^{\vartheta_1}\right)\right] s^{\vartheta_1}t^{\vartheta_2},$$

so that the square of this expression equals

$$\sum_{x,y}\frac{(s^x-t^x)(s^y-t^y)\left(Y_1^x - Y_{T+1}^x\right)\left(Y_1^y - Y_{T+1}^y\right)}{x!y!T^2} + \sum_x \frac{(s^x-t^x)\left(Y_1^x - Y_{T+1}^x\right)}{x!T^2}R_3(s,t)$$
$$+\sum_x \frac{(s^x-t^x)\left(Y_1^x - Y_{T+1}^x\right)(s-t)st}{x!T^2}\sum_{i=1}^{T}\left(Y_i^2Y_{i+1} - Y_iY_{i+1}^2\right)$$
$$+ s^2t^2\frac{(s-t)^2}{4T^2}\left(\sum_{i=1}^{T}\left(Y_i^2Y_{i+1} - Y_iY_{i+1}^2\right)\right)^2 + R_6(s,t),$$

where the indices $x, y$ range between 1 and 3. In view of Proposition 6.2.1 it is clear that the terms of interest are those with the lowest exponents of $s, t$. For the first line in the expression above, this term is given by $((s-t)(Y_1 - Y_{T+1})/T)^2$ with the limiting behavior

$$\lim_{a,T\to\infty} Ta^8 \int_0^1 \int_0^1 \left(\frac{(s-t)(Y_1 - Y_{T+1})}{T}\right)^2 e^{-a(s+t)} ds dt = \lim_{a,T\to\infty} 2a^4 \frac{(Y_1 - Y_{T+1})^2}{T} = 0,$$

as $a^4/T \to 0$ by assumption. For terms with higher order exponents a similar argument applies, so that all the terms of the first sum converge to zero as $a, T \to \infty$. For the second line, consider the term with the lowest exponent for $s, t$:

$$\lim_{a,T\to\infty} Ta^8 \int_0^1 \int_0^1 2 \frac{\left(-t^3 s + 2t^2 s^2 - s^3 t\right)(Y_1 - Y_{T+1})}{T^2} \sum_{i=1}^{T} \left(Y_i^2 Y_{i+1} - Y_i Y_{i+1}^2\right) e^{-a(s+t)} ds dt$$

$$= \lim_{a,T\to\infty} -4\left(Y_1 - Y_{T+1}\right) \frac{a^2}{\sqrt{T}} \left[\frac{1}{\sqrt{T}} \sum_{i=1}^{T} \left(Y_i^2 Y_{i+1} - Y_i Y_{i+1}^2\right)\right].$$

Since $(Y_t)_{t\in\mathbb{Z}}$ is $\alpha$-mixing and all moments exist, the latter factor is asymptotically normal by Theorem 2.5.2. Slutsky's lemma together with $a^2/\sqrt{T} \to 0$ implies the convergence of this expression to 0, similar comments apply to higher order exponents. It is easily seen that the limiting behavior of the remaining term

$$\lim_{a,T\to\infty} Ta^8 \int_0^1 \int_0^1 s^2 t^2 \frac{(s-t)^2}{4T^2} \left(\sum_{i=1}^{T} \left(Y_i^2 Y_{i+1} - Y_i Y_{i+1}^2\right)\right)^2 e^{-a(s+t)} ds dt$$

coincides with the assertion and that the higher order terms all converge to zero by analogous arguments as used before in this proof. □

Similar to the above, this result links the goodness-of-fit test based on the statistic $V_{T,a}(Y_1, \ldots, Y_{T+1})$ (in the special case of a Poisson INAR(1) process, which satisfies the conditions by Lemma 4.1.4 and Theorem 4.1.8) to that established by Theorem 5.2.1. It is noteworthy that in this special case, the resulting covariance kernel of the integrand of the statistic $V_{T,a}$ may be explicitly calculated. We state the result here, mainly in order to show the complexity of the resulting structures.

**Corollary 6.2.4.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process and let $f(x) := \exp(\lambda x)$. Then for each $(u_1, v_1), (u_2, v_2) \in [0,1]^2$, the series*

$$\sigma(u_1, v_1; u_2, v_2) =$$
$$A(u_1, u_2; v_1, v_2) + 2B(u_1, u_2; v_1, v_2) \sum_{k=0}^{\infty} C_k(u_1, u_2; v_1, v_2) D_k(u_1, u_2; v_1, v_2)$$

*converges absolutely, where*

$$A(u_1, u_2; v_1, v_2) = f\left(\frac{\alpha u_1 u_2 v_1 v_2 - 2 + \alpha}{1-\alpha}\right)\left[2f\left(u_1 u_2 + v_1 v_2\right) - 2f\left(v_1 u_2 + u_1 v_2\right)\right],$$

$$
\begin{aligned}
B(u_1,u_2;v_1,v_2) &= f\left(u_1+u_2+v_1+v_2-3+\frac{\alpha(u_1v_1+u_2v_2)-1-\alpha}{1-\alpha}\right),\\
C_k(u_1,u_2;v_1,v_2) &= f\left(\alpha^k\frac{(1-\alpha u_1v_1)(1-\alpha u_2v_2)}{1-\alpha}\right) \quad \textit{and}\\
D_k(u_1,u_2;v_1,v_2) &= \Big[f\left(\alpha^k(-v_1-u_2+\alpha(u_1v_1u_2+v_1u_2v_2)+(1-\alpha)v_1u_2)\right)\\
&\quad - f\left(\alpha^k(-u_1-u_2+\alpha(u_1v_1u_2+u_1u_2v_2)+(1-\alpha)u_1u_2)\right)\\
&\quad - f\left(\alpha^k(-v_1-v_2+\alpha(u_1v_1v_2+v_1u_2v_2)+(1-\alpha)v_1v_2)\right)\\
&\quad + f\left(\alpha^k(-u_1-v_2+\alpha(u_1v_1v_2+u_1u_2v_2)+(1-\alpha)u_1v_2)\right)\Big].
\end{aligned}
$$

*Furthermore,*

$$
\sqrt{T}\left(\widehat{\psi}_T(u,v)-\widehat{\psi}_T(v,u)\right)\xrightarrow{\mathcal{D}}\Psi_3,
$$

*a zero mean Gaussian element in* $C[0,1]^2$*, with covariance function* $\sigma(u_1,v_1;u_2,v_2)$*.*

*Proof.* Using (6.16), Theorem 6.1.6 (i) may be applied. The sum and difference of jointly normal distributions is normally distributed again, and the sum of two tight sequences is tight again, this proves the convergence in distribution on $C[0,1]^2$. Only the covariance kernel remains to be calculated, yielding

$$
\begin{aligned}
&\sigma(u_1,v_1;u_2,v_2)\\
&= \kappa_1(u_1,v_1;u_2,v_2)-\kappa_1(v_1,u_1;u_2,v_2)-\kappa_1(u_1,v_1;v_2,u_2)+\kappa_1(v_1,u_1;v_2,u_2),
\end{aligned}
$$

and each of these summands satisfies (barring a permutation of the argument)

$$
\kappa_1(u_1,v_1;u_2,v_2)=\sum_{k\in\mathbb{Z}}\mathbb{E}_{\theta_0}\left[u_1^{Y_0}v_1^{Y_1}u_2^{Y_k}v_2^{Y_{k+1}}\right].
$$

Since $(Y_t)_{t\in\mathbb{Z}}$ is an INAR(1) process, successive conditioning of the expectations leads to

$$
\begin{aligned}
\mathbb{E}\left[z_0^{Y_0}z_1^{Y_1}\cdots z_{k+1}^{Y_{k+1}}\right] &= \mathrm{pgf}_Y\left((1-\alpha)\sum_{j=0}^{k}\alpha^j\prod_{i=0}^{j}z_i+\alpha^{k+1}\prod_{i=0}^{k+1}z_i\right)\mathrm{pgf}_\epsilon(z_{k+1})\\
&\quad \prod_{r=1}^{k}\mathrm{pgf}_\epsilon\left((1-\alpha)\sum_{j=0}^{r-1}\alpha^j\prod_{i=r-j}^{r}z_{k+1-i}+\alpha^r\prod_{i=k+1-r}^{k+1}z_i\right). \qquad (6.17)
\end{aligned}
$$

Setting $z_2=\cdots=z_{k-1}=1$, this relation simplifies to

$$
\begin{aligned}
&\mathbb{E}\left[z_0^{Y_0}z_1^{Y_1}z_k^{Y_k}z_{k+1}^{Y_{k+1}}\right]\\
&= \mathrm{pgf}_Y\left((1-\alpha)\left[z_0+z_0z_1\sum_{j=1}^{k-1}\alpha^j+\alpha^kz_0z_1z_k\right]+\alpha^{k+1}z_0z_1z_kz_{k+1}\right)\cdot\mathrm{pgf}_\epsilon(z_{k+1})\cdot
\end{aligned}
$$

$$\text{pgf}_\epsilon\left((1-\alpha)z_k + \alpha z_k z_{k+1}\right) \cdots \text{pgf}_\epsilon\left((1-\alpha)\sum_{j=0}^{k-3}\alpha^j + (1-\alpha)\alpha^{k-2}z_k + \alpha^{k-1}z_k z_{k+1}\right) \cdot$$
$$\text{pgf}_\epsilon\left((1-\alpha)\sum_{j=0}^{k-2}\alpha^j z_1 + (1-\alpha)\alpha^{k-1}z_1 z_k + \alpha^k z_1 z_k z_{k+1}\right).$$

In our case, it holds that $\text{pgf}_Y(z) = \exp(\frac{\lambda}{1-\alpha}(z-1))$ and $\text{pgf}_\epsilon(z) = \exp(\lambda(z-1))$. Thus, we obtain

$$= (z_0 + z_1 + z_k + z_{k+1} - 3) + \frac{1}{1-\alpha}\left(\alpha(z_0 z_1 + z_k z_{k+1}) - 1 - \alpha\right)$$
$$+ \alpha^{k-1}\left(\alpha(z_0 z_1 z_k + z_1 z_k z_{k+1}) - (z_1 + z_k) + (1-\alpha)z_1 z_k + \frac{(1-\alpha z_0 z_1)(1-\alpha z_k z_{k+1})}{1-\alpha}\right).$$

Now, let us consider special cases. First, (6.17) yields

$$\mathbb{E}\left[z_0^{Y_0} z_1^{Y_1} z_2^{Y_2}\right] = \exp\left[\lambda\left(z_0 + z_1 + z_2 - 2 - \frac{1}{1-\alpha} + \alpha(z_0 z_1 + z_1 z_2 - z_1) + \frac{\alpha^2 z_0 z_1 z_2}{1-\alpha}\right)\right],$$

furthermore,

$$\begin{aligned}\mathbb{E}\left[z_0^{Y_0} z_1^{Y_1}\right] &= \text{pgf}_Y\left((1-\alpha)z_0 + \alpha z_0 z_1\right)\text{pgf}_\epsilon(z_1)\\ &= \exp\left(\lambda\left[z_0 + z_1 - 1 - \frac{1}{1-\alpha} + \frac{\alpha z_0 z_1}{1-\alpha}\right]\right).\end{aligned}$$

In total, these results imply the assertion. □

## 6.3 Parametric Bootstrap

Looking at the resultant covariance structure in Theorem 6.1.6 and especially in Corollary 6.2.4, it is reasonable to assume that this structure is not available to us explicitly. Hence, in order to apply these results in practice, we establish an appropriate form of resampling technique, i.e., we employ a bootstrapping procedure similar to that of Section 3.3.

For the implementation of such a (parametric) bootstrap procedure, first note that the statistic (6.6) can be explicitly calculated with $\int_0^1 u^x du = \frac{1}{1+x}$. For an underlying process $(Y_t)_{t\in\mathbb{Z}}$ and a realization $(y_1, \ldots, y_{T+1})$,

$$\begin{aligned}W_{T,a}(y_1,\ldots,y_{T+1};\theta) &= \frac{1}{T}\sum_{i,j=1}^{T}\frac{1}{(y_i+y_j+a')(y_{i+1}+y_{j+1}+a')}\\ &- 2\sum_{i=1}^{T}\sum_{k,l=0}^{\infty}\frac{\pi_\theta(k)p_\theta(l|k)}{(y_i+k+a')(y_{i+1}+l+a')} + T\sum_{k_1,k_2,l_1,l_2=0}^{\infty}\frac{\pi_\theta(k_1)p_\theta(l_1|k_1)\pi_\theta(k_2)p_\theta(l_2|k_2)}{(k_1+k_2+a')(l_1+l_2+a')},\end{aligned} \tag{6.18}$$

recalling the notation of (6.2) and setting $a' := a + 1$. Algorithm 1 is formulated for testing the (composite) null hypothesis (6.4), it can easily be modified for the situation (6.3).

**Algorithm 1** (Bootstrap)**.**

*(i)* *Estimate the parameter* $\theta$ *by* $\hat{\theta}_T$ *based on the realization* $(y_1, \ldots, y_{T+1})$*, where* $\hat{\theta}_T$ *satisfies Assumption (A2) (i).*

*(ii)* *Compute the test statistic* $W_0 := W_{T,a}(y_1, \ldots, y_{T+1}; \hat{\theta}_T)$ *given in* (6.18)*.*

*(iii)* *Generate stationary bootstrap data* $(Y_1^*, \ldots, Y_{T+1}^*)$ *by using* (6.2) *with* $f_{\hat{\theta}_T}$*.*

*(iv)* *Estimate the parameter* $\hat{\theta}_T$ *by* $\hat{\theta}_T^*$ *based on the realization* $(Y_1^*, \ldots, Y_{T+1}^*)$*, where* $\hat{\theta}_T^*$ *satisfies Assumption (A2*) (i) below.*

*(v)* *Compute the test statistic* $W^* := W_{T,a}^*(Y_1^*, \ldots, Y_{T+1}^*; \hat{\theta}_T^*)$ *given in* (6.18)*.*

*(vi)* *Repeat steps (v)-(vii) B times to obtain the sequence of statistics* $W_1^*, \ldots W_B^*$*.*

Denoting the corresponding order statistics by $W_{(1)}^* \leq W_{(2)}^* \leq \cdots \leq W_{(B)}^*$, the null hypothesis is then rejected at a significance level $\beta$ if $W_0 > W_{(B\cdot(1-\beta))}^*$. In the algorithm, the notation $W_{T,a}^*(\cdot)$ denotes the expression (6.18), where the measure $\mathbb{P}_\theta$ is replaced with the measure $\mathbb{P}_\theta^*$ conditional on the realization $(y_1, \ldots, y_{T+1})$. Similar comments apply to $\mathbb{E}^*$, $\mathrm{Var}^*$, $\mathrm{Cov}^*$ as well as $\psi^*(u, v; \theta)$. The notation $R_T^* = o_{\mathbb{P}^*}(a_T)$ is used if $\mathbb{P}^*(\|R_T^*\|/|a_T| > \epsilon) \to 0$ in probability for all $\epsilon > 0$. Quite obviously, the bootstrap process also needs to adhere to a certain set of assumptions in order for theoretical results to be available.

**Assumption (A1$^*$).**

*(i)* *The (stationary) bootstrap process satisfies Condition 1 for* $\theta = \hat{\theta}_T$*.*

*(ii)* *The function* $\psi(u, v; \theta)$ *is continuous as a function of* $\theta$ *for all* $u, v \in [0, 1]^2$*.*

*(iii)* *It holds that* $\hat{\theta}_T \to \theta_0$ *almost surely.*

*(iv)* *It holds that* $\mathbb{E}_{\hat{\theta}_T}^*[|Y_0^* Y_1^*|^{2+\xi}]$ *for some* $\xi > 0$*.*

It should be pointed out that the assumption of stationarity is not a very strong one as the structure imposed by the drift condition implies geometric ergodicity, cf. Theorem 6.1.2. Thus, starting at some point and simulating an appropriate pre-run to later discard should be enough to fulfil this assumption in practice, see Section 6.4.

Furthermore, we record the following result, which considers the mixing behavior of the bootstrapped process resulting from Algorithm 1. It turns out that this process maintains the mixing behavior of the true data generating process.

**Lemma 6.3.1** (Schweer (2015a), Lemma B.1). *Let Assumption (A1$^*$) hold. Then the conditionally Markov chain $(Y_t^*)_{t\in\mathbb{Z}}$ is conditionally $\alpha$-mixing with exponentially decreasing weights $\alpha^*(n)$. Furthermore, $\alpha^*(n) \to \alpha(n)$ almost surely, where $\alpha(n)$ are the weights defined in Theorem 6.1.2.*

*Proof.* By Assumption (A1$^*$), $(Y_t^*)_{t\in\mathbb{Z}}$ satisfies Condition 1 for $\theta = \hat{\theta}_T$, hence Theorem 6.1.2 applies, proving the mixing condition for the bootstrap process for finite $T$, for large $T$ the assumption on the convergence $\hat{\theta}_T \to \theta$ implies (a.s.) the mixing condition. The continuity of the jpgf in $\theta$ implies the a.s. convergence $\psi(u, v; \hat{\theta}_T) \to \psi(u, v; \theta_0)$. Since the transition and stationary probabilities of $(Y_t^*)_{t\in\mathbb{Z}}$ can be continuously recovered from the jpgf, this implies the a.s. convergence of these probabilities. As both processes are (conditionally) Markovian, these considerations imply the assertion. □

As in Section 6.1.1, the more difficult situation of assessing the statistics (6.6) with an estimated parameter requires stronger assumptions concerning the behavior of the considered estimator ans well as the underlying (bootstrap) process.

**Assumption (A2$^*$).**

*(i) The sequence of estimators $\hat{\theta}_T^*$ satisfies the expansion*

$$\hat{\theta}_T^* - \hat{\theta}_T = \frac{1}{T}\sum_{i=1}^{T} \mathbf{l}^*(\hat{\theta}_T; i) + \mathbf{r}_T^*,$$

*where $\mathbf{r}_T^* = o_{\mathbb{P}^*}(T^{-\frac{1}{2}})$ and where $\mathbf{l}^*(\theta; i)$ is a measurable function of $(Y_{i+1}^*, Y_i^*, \dots)$ and $\theta$ such that $\mathbb{E}^*_{\hat{\theta}_T}[l^*(\theta_0; i)_k] = 0$, $\mathbb{E}_{\hat{\theta}_T}[l^*(\theta_0; i)_k^{2+\delta}] < \infty$ for some $\delta > 0$ and, if $(Y_t^*)_{t\in\mathbb{Z}}$ is conditionally $\alpha$-mixing with exponentially decreasing weights, the same holds for $(l^*(\theta_0; t)_k))_{t\in\mathbb{Z}}$ for all $k \in \{1, \dots, d\}$.*

*(ii) The series $\sum_{k,l=0}^{\infty} kl \frac{\partial}{\partial\theta}\mathbb{P}^*_{\hat{\theta}_T}(Y_0 = k, Y_1 = l)$ converges.*

The main bootstrap result can now be formulated, providing the analogon of Corollary 6.1.7 for the bootstrapped statistics. The idea of the proof stems from (Rajarshi, 1990, Theorem 2.2.).

**Theorem 6.3.2** (Schweer (2015a), Theorem 3.1). *Let Assumptions (A1) and (A2) be satisfied and let the notation of Corollary 6.1.7 hold.*

*(i) If Assumption (A1$^*$) is satisfied, then, almost surely,*

$$W_{T,a}^*(Y_1^*, \dots, Y_{T+1}^*; \hat{\theta}_T) \xrightarrow{\mathcal{D}} \sum_{j\geq 1} \lambda_{1;j} Z_{1;j}^2,$$

*where the series converges in the $L^2$-sense.*

*(ii) If Assumption (A1*) and Assumption (A2*) are satisfied, then, almost surely,*

$$W^*_{T,a}(Y^*_1, \dots, Y^*_{T+1}; \hat{\theta}^*_T) \xrightarrow{\mathcal{D}} \sum_{j \geq 1} \lambda_{2;j} Z^2_{2;j},$$

*where the series converges in the $L^2$-sense.*

*Proof.* A closer look at the proof of Theorem 6.1.6 (a) reveals that tightness of the process follows analogously under Assumption (A1*), when replacing $\mathbb{P}$ with the conditional $\mathbb{P}^*$ and so forth. Thus, let $(u, v) \in [0, 1]^2$, by continuous mapping theorem it follows that $\psi^*(u, v; \hat{\theta}_T) \to \psi(u, v; \theta_0)$ almost surely. By Assumption (A1*), there exists a set $E$ in the underlying $\sigma$-algebra of the process, such that $\theta_T(\omega) \to \theta$ holds surely for each $\omega \in E$, with $\mathbb{P}_{\theta_0}(E) = 1$.

Let $\omega \in E$ be arbitrary but fixed and let $\epsilon > 0$. By Theorem 6.1.2, the weights $\alpha^*(j)$ decrease exponentially, so that $\sum_{j=1}^{\infty} \alpha^*(j) < \infty$. Since $\alpha^*(j) \to \alpha(j)$ for $T \to \infty$, an integer $t_1(\omega)$ may be chosen such that $2 \sum_{j=t_1(\omega)}^{\infty} \alpha^*(j) < \epsilon$ and $2 \sum_{j=t_1(\omega)}^{\infty} \alpha(j) < \epsilon$ for all $T \geq t_1(\omega)$. By Lemma 6.3.1, the bootstrap process is $\alpha$-mixing for $\omega \in E$, thus Lemma 1.2. in Ibragimov (1962) is applicable (note that $\psi^*(u, v; \hat{\theta}_T) \leq 1$), yielding $\mathrm{Cov}^*_{\hat{\theta}_T}(u^{Y^*_0} v^{Y^*_1}, u^{Y^*_j} v^{Y^*_{j+1}}) \leq C\alpha^*(j)$. As the next step, choose $t_2(\omega)$ so that $T \geq t_2(\omega)$ implies

$$\left| \mathrm{Cov}^*_{\hat{\theta}_T} \left( u^{Y^*_0} v^{Y^*_1}, u^{Y^*_j} v^{Y^*_{j+1}} \right) - \mathrm{Cov}_{\theta_0} \left( u^{Y_0} v^{Y_1}, u^{Y_j} v^{Y_{j+1}} \right) \right| < \frac{\epsilon}{2t_1(\omega) - 1}$$

for all $j = 0, 1, 2, \dots, t_1(\omega)$. Now, the variance of $\sqrt{T}(\widehat{\psi}^*_T(u, v) - \psi^*(u, v; \hat{\theta}_T))$ calculates due to stationarity to $\mathrm{Var}^*_{\hat{\theta}_T}(u^{Y^*_0} v^{Y^*_1}) + 2 \sum_{j=1}^{T-1} (T - j) \mathrm{Cov}^*_{\hat{\theta}_T}(u^{Y^*_0} v^{Y^*_1}, u^{Y^*_j} v^{Y^*_{j+1}})/T$. By the exponential decrease of $\alpha^*(j)$ the series $\sum_{j=1}^{T-1} j |\mathrm{Cov}^*_{\hat{\theta}_T}(u^{Y^*_0} v^{Y^*_1}, u^{Y^*_j} v^{Y^*_{j+1}})|$ converges for $T \to \infty$ and there is a $t_3(\omega)$ such that this expression is less than $\epsilon T$ for all $T \geq t_3(\omega)$. Choosing a $T \geq \max\{t_1(\omega), t_2(\omega), t_3(\omega)\}$ and recalling the notation of Theorem 6.1.6, it follows that

$$\begin{aligned}
&\left| \kappa_1(u_1, v_1; u_2, v_2) - \sum_{j=-T+1}^{T-1} \frac{T - j}{T} \mathrm{Cov}^*_{\hat{\theta}_T} \left( u^{Y^*_0} v^{Y^*_1}, u^{Y^*_j} v^{Y^*_{j+1}} \right) \right| \\
&\leq \sum_{j=-t_1(\omega)+1}^{t_1(\omega)-1} \left| \mathrm{Cov}_{\theta_0} \left( u^{Y_0} v^{Y_1}, u^{Y_j} v^{Y_{j+1}} \right) - \mathrm{Cov}^*_{\hat{\theta}_T} \left( u^{Y^*_0} v^{Y^*_1}, u^{Y^*_j} v^{Y^*_{j+1}} \right) \right| \\
&+ 2 \sum_{j=t_1(\omega)}^{\infty} \left| \mathrm{Cov}_{\theta_0} \left( u^{Y_0} v^{Y_1}, u^{Y_j} v^{Y_{j+1}} \right) \right| + \frac{1}{T} \sum_{j=1}^{T-1} j \left| \mathrm{Cov}^*_{\hat{\theta}_T} \left( u^{Y^*_0} v^{Y^*_1}, u^{Y^*_j} v^{Y^*_{j+1}} \right) \right| \\
&+ 2 \sum_{j=t_1(\omega)}^{T-1} \left| \mathrm{Cov}^*_{\hat{\theta}_T} \left( u^{Y^*_0} v^{Y^*_1}, u^{Y^*_j} v^{Y^*_{j+1}} \right) \right| < C\epsilon,
\end{aligned}$$

showing that the asymptotic variances of the bootstrapped ejpgf coincide with that of the true ejpgf. Since the moment condition of Assumption (A1$^*$) is in place, Theorem 1 in Ekström (2014) is applicable, hence it follows that

$$\sqrt{T}\left(\widehat{\psi}_T^*(u,v) - \psi^*(u,v;\hat{\theta}_T)\right) \xrightarrow{\mathcal{D}} \Psi_1$$

in $C[0,1]^2$ almost surely. Quite analogously, we can show that under Assumption (A1$^*$) and Assumption (A2$^*$) it holds that

$$\sqrt{T}\left(\widehat{\psi}_T^*(u,v) - \psi^*(u,v;\hat{\theta}_T^*)\right) \xrightarrow{\mathcal{D}} \Psi_2$$

in $C[0,1]^2$ almost surely. Using the reasoning preceding Corollary 6.1.7 concludes the proof. □

### 6.3.1 Effect of Simulating the Stationary Distribution

In this section, the particular case is discussed where the calculation of the statistics $W_{T,a}(y_1,\dots,y_{T+1};\theta')$ and $W^*_{T,a}(Y_1^*,\dots,Y_{T+1}^*;\theta')$ is complicated because the stationary distribution of the process is not analytically available, cp. the discussion in Section 6.1.2. A remedy is suggested below. Since the result of Theorem 6.1.8 is used, the following additional assumption besides those listed above needs to be in place for its applicability:

**Assumption (A3$^*$).** *For any fixed $\theta' \in \Theta$, it holds that $\mathbb{E}_{\theta'}[Y_0^{2+\xi}] < \infty$, where $\xi > 0$.*

Under this assumption, the following steps can be added to Algorithm 1 between steps (i) and (ii) and steps (iv) and (v), respectively, replacing $\theta'$ and $Y_i$ with $\hat{\theta}_T$ and $y_i$ or $\hat{\theta}_T^*$ and $Y_i^*$ where appropriate.

**Algorithm 2** (Simulating the Stationary Distribution)**.**

- *(i)$^*$ Generate stationary bootstrap data $(Y_1^{**},\dots,Y_S^{**})$ by using (6.2) with $p_{\theta'}$.*
- *(iv)$^*$ Calculate the statistic $\widehat{W}_{S,T,a}(Y_1,\dots,Y_{T+1};\theta')$ by replacing $\pi_{\theta'}(k)$ with the estimator $\hat{\pi}_{\theta'}(k) := \frac{1}{S}\sum_{i=1}^S \mathbf{1}_{\{Y_i^{**}=k\}}$ for all $k \in \mathbb{N}_0$.*

Certainly, the implementation of these additional steps leads to a higher volatility of the calculated test statistics on the one hand, and a higher computational demand on the other hand. The following Lemma addresses the former concern in the more intricate case of testing the hypothesis (6.4), the easier case (6.3) then follows directly. Note that the proof draws heavily from Theorem 6.1.8.

**Lemma 6.3.3** (Schweer (2015a), Lemma 3.2)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ satisfy Condition 1, with $\mathbb{E}_\theta[Y_0^{4+\xi}] < \infty$, where $\xi > 0$. Let $\theta \in \Theta$. Then for large $T$ it holds almost surely that*

$$\mathbb{E}\left[\left(W_{T,a}(Y_1,\dots Y_{T+1};\theta) - \widehat{W}_{S,T,a}(Y_1,\dots,Y_{T+1};\theta)\right)^2\right] = O\left(\frac{T^2}{S}\right).$$

*Proof.* Recall the denotation of (6.2) and of $\hat{\pi}_\theta(k)$. First, it follows from (6.18) that

$$\begin{aligned} &W_{T,a}(Y_1,\dots,Y_{T+1};\hat{\theta}_T) - \widehat{W}_{S,T,a}(Y_1,\dots,Y_{T+1};\hat{\theta}_T) \\ &= T\sum \frac{p_\theta(l_1|k_1)p_\theta(l_2|k_2)\,(\pi_\theta(k_1)\pi_\theta(k_2) - \hat{\pi}_\theta(k_1)\hat{\pi}_\theta(k_2))}{(k_1+k_2+a')(l_1+l_2+a')} \\ &\quad - 2\sum \frac{p_\theta(l|k)\,(\pi_\theta(k) - \hat{\pi}_\theta(k))}{(Y_i+k+a')(Y_{i+1}+l+a')}. \end{aligned}$$

Employing the triangle inequality, these two series can be dealt with separately, beginning with the second series. Here, the nonnegativity of the random variables $Y_i$ together with the triangle inequality yields the following upper bounds:

$$\mathbb{E}_\theta\left[\left|2\sum_{i=1}^{T}\sum_{k,l=0}^{\infty}\frac{p_\theta(l|k)\,(\pi_\theta(k)-\hat{\pi}_\theta(k))}{(Y_i+k+a')(Y_{i+1}+l+a')}\right|^2\right] \le \frac{CT^2}{S}\sum_{k=0}^{\infty}\operatorname{Var}_\theta\left(\frac{1}{\sqrt{S}}\sum_{i=1}^{S}\mathbf{1}_{\{Y_i^{**}\le k\}}\right),$$

the latter series is bounded by the proof of Theorem 6.1.8. With an analogous argumentation, there exists an upper bound for the first series of the form

$$T^2 C\sum_{k_1,k_2=0}^{\infty}\operatorname{Var}_\theta[\hat{\pi}_\theta(k_1)\hat{\pi}_\theta(k_2)].$$

The exact expression for the variance of a product of random variables is provided in eq. (5) in Bohrnstedt and Goldberger (1969) and $\mathbb{E}_\theta[(\hat{\pi}_\theta(k_1)-f_\theta(k_1))^2(\hat{\pi}_\theta(k_2)-f_\theta(k_2))^2] \le |\operatorname{Cov}_\theta(\hat{\pi}_\theta(k_1),\hat{\pi}_\theta(k_2))|$ (since $|\hat{\pi}_\theta(k_1)-f_\theta(k_1)|\le 1$ holds deterministically). Thus, this expression is bounded from above by

$$\frac{CT^2}{S}\sum_{k_1,k_2=0}^{\infty}\left|\operatorname{Cov}_\theta\left(\frac{1}{\sqrt{S}}\sum_{i=1}^{S}\mathbf{1}_{\{Y_i^{**}\le k_1\}},\frac{1}{\sqrt{S}}\sum_{i=1}^{S}\mathbf{1}_{\{Y_i^{**}\le k_2\}}\right)\right|.$$

An argument very similar to that proving (6.13) shows that this is bounded by

$$\frac{CT^2}{S}\left(1+24\sum_{i=1}^{S-1}\frac{S-i}{S}\sqrt{\alpha(i)}\right)\sum_{k_1,k_2=0}^{\infty}(1-\Pi_\theta(k_1))^{\frac{1}{4}}(1-\Pi_\theta(k_2))^{\frac{1}{4}}.$$

The latter series converges under the assumption of existing moments of order $4+\xi$ for $G$ (see Lemma 3.2.3) and the assertion follows by Chebyshev's inequality. □

It should be noted that this result merely serves as an orientation for finding appropriate values for $S$ in Algorithm 2. In practice, the choice of $S$ will most often be guided by the computational aspects mentioned above. This is discussed in detail for the special case of the Poisson INARCH(1) model below, including an approach for reducing the computational times. Furthermore, an empirical simulation study providing more insight into this topic is also given there.

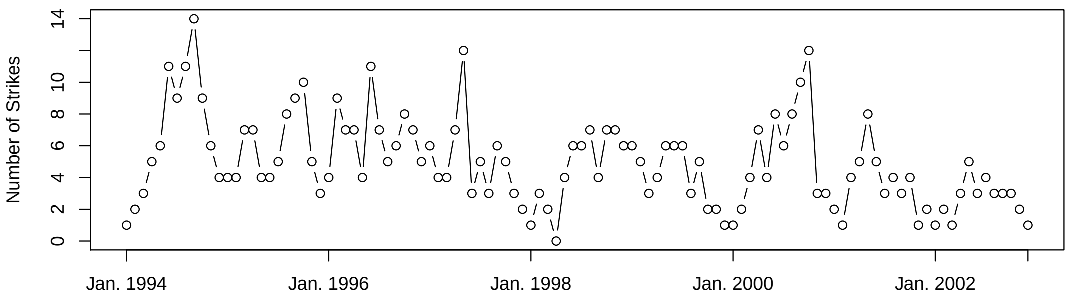


Figure 6.1: Plot of monthly numbers of work stoppages (strikes) leading to 1000 or more workers being idle in this period, reported by the US Bureau of Labor Statistics.

## 6.4 An Application to Real Data

In order to assess the performance of the proposed tests, an application to a real data set is given in this section. The data example $(y_1, \dots, y_{108})$ is given by monthly strikes counts published by the U.S. Bureau of Labor Statistics for the period January 1994 to December 2002, thus totaling 108 observations. The data is plotted in Figure 6.1, each count represents a work stoppage in the United States which lead to 1000 or more workers being idle in this period. The corresponding ACF and PACF are reported in Figure 6.2. This time series was already studied in Jung et al. (2005), where the authors came to the conclusion that an INCLAR(1) model seems an appropriate fit for the date in view of the exponential decay of the autocorrelation function. Their findings suggested that the Poisson INAR(1) model is not a good fit for the data and employed a Negative Binomial INAR(1) model and a Random Coefficient INAR(1) model as alternatives.

In Table 6.1, the results of fitting different INCLAR(1) models to the strike count data are reported. Three scenarios are considered, the (Poisson) INARCH(1) model of (6.1), the Poisson INAR(1) process, i.e., (1.1) with Poisson innovations, and the INAR(1) process with a Poisson distribution of order 2 of Example 1. All models are fitted with a conditional maximum likelihood (CML) approach to the data $y_{108}, \dots, y_2$ given $y_1$. Estimates and approximate standard errors are computed by using R's `optim`, which is initialized with the moment estimates for the respective parameters. The probabilities required for the $\text{Poi}_2(\lambda)$-innovations are computed by using Proposition 2.3.6.

Comparing the different fitted models in terms of AIC and BIC it becomes clear that the INARCH(1) model fares best amongst the models under consideration, yet the results for the $\text{Poi}_2$-INAR(1) model show that this model is also an adequate fit. The classical Poisson INAR(1) model receives the worst AIC and BIC values, thus suggesting that this model should be rejected.

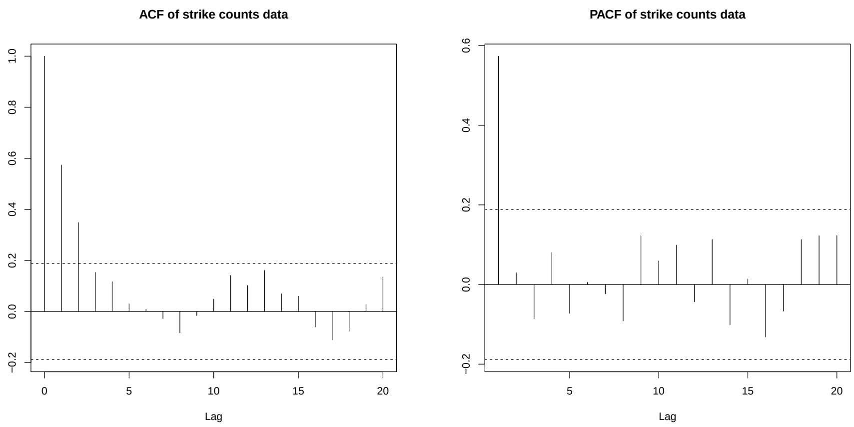


Figure 6.2: Strike counts from Figure 6.1: Plots of ACF and PACF.

| Model | Par. 1 | Par. 2 | AIC | BIC |
|---|---|---|---|---|
| INARCH(1) | 1.8114 | 0.6364 | 464.3 | 469.7 |
| $(\beta, \alpha)$ | (0.386) | (0.081) | | |
| Poi($\lambda$)-INAR(1) | 2.4603 | 0.5061 | 473.1 | 478.5 |
| $(\lambda, \alpha)$ | (0.299) | (0.056) | | |
| Poi$_2$($\lambda$)-INAR(1) | 1.4296 | 0.5696 | 467.6 | 473.0 |
| $(\lambda, \alpha)$ | (0.199) | (0.051) | | |

Table 6.1: Strike counts from Figure 6.1: CML estimates for diverse models.

### 6.4.1 Simulation Study

In the INARCH(1) case, the application of the test presented in this chapter involves simulating the stationary distribution via the Algorithm 2, hence an empirical simulation study discussing the effect of this procedure is presented. First, a Poisson INARCH(1) process of length $T = 100$ as given in (6.1) is simulated with the parameter $\theta = (\beta, \alpha)$ chosen to be $\beta = 1.8114$ and $\alpha = 0.6364$ in order to resemble the (possibly) true parameters of the data example given above (cf. Table 6.1). Concerning stationarity, the bootstrap time series is started in $Y_0 \sim \text{Poi}(\beta \cdot (1-\alpha))$ for want of a better approximation of the true stationary distribution and a total number of $T+500$ realizations is simulated, where the first 500 are discarded. In view of the geometrical ergodicity of $(Y_t)_{t\in\mathbb{Z}}$ (cf. Theorem 6.1.2) it may be assumed that the influence of the initialization is negligible. Then, the statistics $\widehat{W}_{S,T,a}(Y_1, \ldots, Y_{T+1}; \theta)$ are calculated $B = 1000$ times according to Algorithm 2 provided for varying values of $a$ and $S$. In Table 6.2, the sample variance as well as the difference between the maximum and the minimum of the statistics for the respective parameter values are reported.

From Table 6.2 it can be easily seen that, not surprisingly, a higher value of $S$ yields

|  | $S = 10^4$ |  | $S = 10^6$ |  |
|---|---|---|---|---|
| $a$ | max − min | sample variance | max − min | sample variance |
| 0 | 0.0273 | $1.936 \cdot 10^{-5}$ | 0.0047 | $4.186 \cdot 10^{-7}$ |
| 2 | 0.0072 | $1.559 \cdot 10^{-6}$ | 0.0012 | $3.883 \cdot 10^{-8}$ |
| 4 | 0.0032 | $2.510 \cdot 10^{-7}$ | 0.0005 | $8.435 \cdot 10^{-9}$ |
| 6 | 0.0016 | $6.778 \cdot 10^{-8}$ | 0.0003 | $2.667 \cdot 10^{-9}$ |
| 8 | 0.0009 | $2.352 \cdot 10^{-8}$ | 0.0002 | $9.355 \cdot 10^{-10}$ |
| 10 | 0.0005 | $9.053 \cdot 10^{-9}$ | 0.0001 | $3.951 \cdot 10^{-10}$ |

Table 6.2: Simulated statistics $\widehat{W}_{S,T,a}(Y_1, \ldots, Y_{T+1}; \theta)$ for INARCH(1) process with varying $a, S$.

better results both for the variances as well as for the maximal differences for all studied values of $a$, in accordance with Lemma 6.3.3. The variances improve at a faster rate with increasing $S$ than the maximal differences, again for all values of $a$. For each choice of $S$ and with increasing $a$, the variances and the maximal differences decline. Such a behavior is expected when looking at the expression (6.18). In summation, the result of Table 6.2 suggest that the usage of Algorithm 2 achieves reasonable results for appropriate choices of $S$.

### 6.4.2 Goodness-of-fit Test for Strike Count Data

The goal of this section consists in applying the test based on (6.18) to test the goodness-of-fit of the INARCH(1) model for the strike count data. The parameters $\beta, \alpha$ are estimated by the CML estimators $\hat{\beta}_T, \hat{\alpha}_T$, using R's `optim` which is initialized with the respective moment estimators of the parameters. Using these estimates, an INARCH(1) process with parameters $\hat{\beta}_T, \hat{\alpha}_T$ and length $S = 10^6$ is simulated and used to estimate the stationary distribution, and the statistic $\widehat{W}_{10^6,107,a}(y_1, \ldots, y_{108}; \hat{\theta}_{107})$ is calculated. In the next step, the same procedure is applied to the bootstrap variables $(Y_1^*, \ldots, Y_{T+1}^*)$ which are generated as described in Algorithm 1. Concerning stationarity, the comments of the previous section apply *mutatis mutandis*. Estimating $\hat{\theta}$ by the CML estimator $\hat{\theta}_T^*$, the statistic $\widehat{W}^*_{10^6,107,a}(Y_1^*, \ldots, Y_{108}^*; \hat{\theta}_{107}^*)$ is calculated. For $B = 1000$ replications, the p-value is reported in Table 6.3 for five different values of $a$.

The choice of $S = 10^6$ represents a good compromise between accuracy and computational speed and is supported by the results of the previous section, cp. Table 6.2. Note that the simulation of the INARCH(1) process is achieved with a small `C++` script, which employs a suitable modification of the Box-Muller method to simulate Poisson distributed random variables. This script decreases computing times, one single simulation of $10^6$ realizations takes roughly 0.6 second, whereas for $10^7$ realizations more than 6 seconds elapse.

For the second scenario, an underlying Poisson INAR(1) process is assumed, governed by unknown parameters $\lambda_0, \alpha_0$. The hypothesis (6.4) is tested using Algorithm 1. As before, the parameters are estimated using conditional maximum likelihood estimators

$\hat{\lambda}_T, \hat{\alpha}_T$. For $B = 1000$, the bootstrapped parameters $\hat{\alpha}_T^*$ and $\hat{\lambda}_T^*$ are again estimated with the conditional likelihood method, the results of this procedure are reported in Table 6.3. The third scenario is handled analogously to the second one with one difference: For the calculation of the stationary distribution as well as the transition probabilities, the results of Proposition 2.3.6 and Theorem 4.2.5 are used. It should be noted that all of the models and the estimation method chosen satisfy the Assumptions (A1) through (A2*), cf. Lemmata 6.1.3 and 6.1.4.

| Model | $a$ | statistic | critical value | p-value |
|---|---|---|---|---|
| | 0 | 0.00492 | 0.03347 | 0.512 |
| | 2 | 0.00117 | 0.00463 | 0.321 |
| INARCH(1) | 6 | 0.00028 | 0.00076 | 0.191 |
| | 8 | 0.00002 | 0.00037 | 0.689 |
| | 10 | 0.00015 | 0.00020 | 0.084 |
| | 0 | 0.01886 | 0.01535 | 0.033 |
| Poi($\lambda$)-INAR(1) | 2 | 0.00501 | 0.00282 | 0.011 |
| | 10 | 0.00021 | 0.00011 | 0.009 |
| | 0 | 0.80792 | 1.02489 | 0.376 |
| $\text{Poi}_2(\lambda)$-INAR(1) | 2 | 0.53182 | 0.60359 | 0.344 |
| | 10 | 0.18372 | 0.18831 | 0.243 |

Table 6.3: Strike counts: $p$-values and critical values of statistic $W_{107,a}(y_1, \ldots, y_{108}; \hat{\theta}_{107})$ for diverse models.

For the first scenario discussed, Table 6.3 shows that all values of $a$ lead to the same result: That the null hypothesis of $(y_1, \ldots, y_{T+1})$ stemming from an INARCH(1) process is not rejected. The $p$-value behaves rather erratically with respect to the weight parameter $a$, however, such a behavior is expected when taking into account the findings of previous sections. For instance, a closer look at the values for $a = 10$, Table 6.2 suggests that variations of $\widehat{W}_{10^6,107,10}(y_1, \ldots, y_{108}; \hat{\theta}_{107})$ of order 0.0001 are unlikely but possible. Comparing this with the discrepancy between the calculated statistic and the critical value in Table 6.3 shows that such a variation would be large enough to let the statistic exceed the critical value, thus producing a rejection of the null hypothesis. On the other hand, the comparison of the respective values for $a = 0$, e.g., shows that here the variations of $\widehat{W}_{10^6,107,0}(y_1, \ldots, y_{108}; \hat{\theta}_{107})$ are too small for such occurrences. Thus, these findings should make it clear that in the case of high weight parameters $a$, the procedures introduced in this chapter should only be applied with caution.

For the second scenario, the estimation of the stationary distribution is not necessary. The p-values of the statistics recommend rejection of the null hypothesis of a Poisson INAR(1) process across the board of $a$ values. Even though this assessment matches that of Jung et al. (2005), it is nevertheless surprising how strongly the goodness-of-fit test of (6.6) rejects this null hypothesis, given that the time series consists of merely 107 pairs of observations.

In the third case, where the null hypothesis is given by a $\text{Poi}_2$-INAR(1) process, the result is yet again easily interpreted, the tests suggests not to reject the null hypothesis for all values of $a$. In contrast to the behavior of the p-value of the statistic for the INARCH(1) null hypothesis, the p-value decreases only slowly with increasing $a$. The higher smoothness is to be expected since in this case, the stationary distribution is explicitly available and does not need to be simulated. The decreasing p-value could, in connection with Proposition 6.2.2 be seen as an indicator that there is a discrepancy between the asymptotic moments of the data and of those of the $\text{Poi}_2$-INAR(1) process. On the other hand, since $T = 107$ is a very short realization, such interpretations should always be taken with a grain of salt.

### 6.4.3 Comparison with Hudecová et al. (2015)

In Hudecová et al. (2015), a goodness-of-fit test very similar to the one discussed above was considered. As an application, the authors analyzed several time series previously considered in Freeland (1998). For five count data time series, the appropriateness of a stationary Poisson INAR(1) model was tested and the p-value reported. In order to assess differences in performance, the procedure of the previous section was carried out for three selected time series for the same weights $a$ as well as the same number of bootstrap samples, the results are reported in Table 6.4. The bold font values are those calculated from Algorithm 1, the others from Table 3 of Hudecová et al. (2015).

| | values of $a$ | | | | |
|---|---|---|---|---|---|
| series | 0 | 1 | 2 | 5 | 10 |
| # 1 | **0.575** | **0.602** | **0.598** | **0.639** | **0.655** |
| | 0.653 | 0.652 | 0.698 | 0.771 | 0.862 |
| # 3 | **0.005** | **0.002** | **0.000** | **0.000** | **0.000** |
| | 0.020 | 0.002 | 0.001 | 0.000 | 0.000 |
| # 4 | **0.233** | **0.216** | **0.256** | **0.316** | **0.376** |
| | 0.270 | 0.270 | 0.295 | 0.386 | 0.506 |

Table 6.4: Claim counts data: Comparison of p-values of statistic (6.6) to the statistic of Hudecová et al. (2015).

The results in Table 6.4 corroborate those of Freeland (1998) in that for the series #1 and #4, the null hypothesis of a Poisson INAR(1) model is not rejected. For series # 3, this model seems to be inappropriate. Comparing the p-values, it can clearly be seen that even though there is not a large difference, the p-values calculated from Algorithm 1 are lower across the entire range of series and weights $a$. It should be noted that after comparing their simulation results for various different models and varying weight parameters $a$, the authors in Hudecová et al. (2015) conclude that it "seems that it is not possible to generally recommend one particular value of $a$". As Table 6.4 indicates that both of the goodness-of-fit tests behave similarly, it seems likely that the same assertion could hold true in the case of the statistics (6.6).

# 7 Higher Order Integer-Valued Autoregressive Processes

The INAR(1) model as discussed in Chapters 4 and 5 can be seen as an integer-valued version of the continuous AR(1) model, which is quite easily extended to a higher order AR(p) model, cf. (2.18). It is therefore not surprising that the modeling of higher order autoregressive structures has been considered in the literature. Indeed, the first contributions considering INAR(p) processes (with $p > 1$) were published shortly after the introduction of the INAR(1) model in McKenzie (1985) and Al-Osh and Alzaid (1987). However, the two most prominent attempts, given by Alzaid and Al-Osh (1990) and Du and Li (1991), differ substantially so that there exists no canonical extension of the INAR(1) process. In terms of popularity, there does not seem to be a large difference between the two formulations when comparing the number of times these articles were cited. As of this writing (April 2015), the web page `scholar.google.com` reports 161 citations for the article Alzaid and Al-Osh (1990) and 166 for Du and Li (1991).

In this chapter, we will present both formulation and discuss various merits and disadvantages of both models. In order to avoid confusion, we use the initials of their respective creators as prefixes to mark the respective model, i.e., we discuss the integer-valued autoregressive process of $p$-th order in the formulation of Al-Osh and Alzaid (AAINAR(p)) and the integer-valued autoregressive process of $p$-th order in the formulation of Du and Li (DLINAR(p)). We begin with a short analysis of the properties of the former process, focusing especially on its behavior with respect to time-reversibility. This characteristic is a recurring motif in this thesis, and for INAR(p) processes, it leads to an interesting result. The first part shows that AAINAR(p) processes behave in analogy to the INAR(1) process, i.e., they are time-reversible if and only if the arrival distribution is Poisson distributed, cf. Theorem 4.1.8.

The DLINAR(p) process on the other hand behaves very differently in relation to time-reversibility, as shown in Theorem 7.2.10, it is only time-reversible in trivial cases. Additionally, we consider several other aspects of DLINAR(p) process, such as the stationary distribution and joint cumulants. Since the ACF of this process satisfies the Yule-Walker equations, some parts of the theory of previous chapters can easily be generalized to this setting. Finally, let us point out that several parts of this chapter are directly lifted from the contribution Schweer (2015b).

## 7.1 The INAR(p) Model of Al-Osh and Alzaid

**Definition 7.1.1** (AAINAR(p) Process). *Let $(\epsilon_t)_{t\in\mathbb{Z}}$ be an i.i.d. process with range $\mathbb{N}_0$, let $\sigma_\epsilon^2 < \infty$. Let $p \in \mathbb{N}$, $\alpha_1, \dots, \alpha_p \in [0,1)$ with $\sum_{i=1}^p \alpha_i < 1$. A process $(Y_t)_{t\in\mathbb{Z}}$, which follows the recursion*

$$Y_t = \sum_{i=1}^{p} \alpha_i \circ Y_{t-i} + \epsilon_t \quad \text{for all } t \in \mathbb{Z} \tag{7.1}$$

*is said to be an AAINAR(p) process, if the conditional distribution of the vector $(\alpha_1 \circ Y_t, \alpha_2 \circ Y_t, \dots, \alpha_p \circ Y_t)$ given $Y_t = y_t$ is multinomial with parameters $(\alpha_1, \alpha_2, \dots, \alpha_p, y_t)$ and if, given $Y_t = y_t$, the random variables $a_i \circ Y_t$ and $\epsilon_t$ are independent of $Y_{t-k}$ and its survivals (and the thinning operations) $a_j \circ Y_{t-k}$ for $i, j = 1, 2, \dots, p$ and $k > 0$.*

### 7.1.1 First Properties

In order to keep the expositions of this chapter and that of Chapter 4 parallel, let us record some properties of the AAINAR(p) process shown in Alzaid and Al-Osh (1990). It is quite clear that the defining recursion (7.1) directly generalizes the recursion (1.1). Yet an AAINAR(p) process is not Markovian for $p \geq 1$ as its present realization depends on the thinning operations in the $p$ time steps before. Concerning the stationary distribution of an AAINAR(p) process $(Y_t)_{t\in\mathbb{Z}}$ with parameters $\alpha_j$, this can be written down by first defining the weights

$$w_0 := 1 \quad \text{and} \quad w_j := \sum_{i=1}^{\min(j,p)} \alpha_i w_{i-j},$$

which allows for the following representation of the process (cf. Theorem 2.1 in Alzaid and Al-Osh (1990))

$$Y_t \stackrel{\mathcal{D}}{=} \sum_{j=0}^{\infty} w_j \circ \epsilon_{t-j}. \tag{7.2}$$

Concerning the joint distribution of the process, it can be shown (cf. eq. (3.8) of Alzaid and Al-Osh (1990)) that the covariance structure resembles that of a Gaussian ARMA$(p, p-1)$ process. In this sense, this model is not a true generalization of the AR(p) process (2.18).

Finally, let us discuss the relationship between this model and the classical branching processes. The original article presents the following heuristic model: Consider a biological population of a species in which we count only the number of members able to reproduce. Let us assume that these members can give birth to at most one offspring which is able to reproduce, and let us further assume that the reproductive span of each member is split into $p$ periods without overlap. Then, if the probability of each member reproducing within the $i$th period of the reproductive span is $\alpha_i$, and if there is an immigration process $\epsilon_t$ from the outside, the resulting total of members follows an AAINAR(p) process of Definition 7.1.1.

### 7.1.2 Concerning Time-Reversibility

In their Section 5.2., Alzaid and Al-Osh (1990) show the time-reversibility of AAINAR(2) processes with Poisson innovations and indicate how this result may be established in higher-order autoregressive structures. In the next theorem it is shown that the converse assertion is also true, the proof generalizes the approach of Theorem 4.1.8 AAINAR(p) processes of a general order.

**Theorem 7.1.2** (Schweer (2015b), Theorem 19.2)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a time-reversible AAINAR(p) process with $p > 1$ and let $\mathbb{P}(\epsilon_0 = 0) \in (0,1)$. Then $(Y_t)_{t\in\mathbb{Z}}$ is time-reversible if and only if $\epsilon_0 \sim \mathrm{Poi}(\lambda)$ for some $\lambda > 0$.*

*Proof.* Sufficiency of the assertion remains to be shown. First, let $p = 2$, and define the vector-valued process $\mathbf{Y}_t := (Y_t, \alpha_2 \circ Y_{t-1})$. and denote $\mathbb{P}(\epsilon_0 = k) := p_\epsilon(k)$. As shown in Section 4 of Alzaid and Al-Osh (1990), this process is Markovian. Denote the transition probabilities of this process by $\mathbb{P}(\mathbf{Y}_t = (a_1, a_2)|\mathbf{Y}_{t-1} = (b_1, b_2)) := p_{\mathbf{Y}}((a_1, a_2)|(b_1, b_2))$ for $a_1, a_2, b_1, b_2 \in \mathbb{N}_0$. Now, the event $\{Y_t = 0\}$ implies that $\{\alpha_j \circ Y_t = 0\}$ for $j = 1, 2$ and any $t \in \mathbb{Z}$ by the definition of the thinning operation. Now, $\{Y_t = 0\}$ implies that $\{\alpha_2 \circ Y_{t-2} = 0\}$ for any $t \in \mathbb{Z}$, since by Definition 7.1.1, $Y_t = \alpha_1 \circ Y_{t-1} + \alpha_2 \circ Y_{t-2} + \epsilon_t$ and all random variables are assumed to be nonnegative. Therefore, for any $i \in \mathbb{N}_0$,

$$\begin{aligned}&\mathbb{P}\left(\{Y_{-1} = 0, Y_0 = 0, Y_1 = 1, Y_2 = i, Y_3 = 0, Y_4 = 0, \alpha_2 \circ Y_t = 0; t = -2, -1, \ldots, 3\}\right)\\ &= \mathbb{P}\left(Y_{-1} = 0, Y_0 = 0, Y_1 = 1, Y_2 = i, Y_3 = 0, Y_4 = 0\right)\end{aligned}$$

Using the time-reversibility of $(Y_t)_{t\in\mathbb{Z}}$ and the Markovian structure of $(\mathbf{Y}_t)_{t\in\mathbb{Z}}$, the argumentation above implies for every $i \in \mathbb{N}_0$ that

$$\begin{aligned}&\mathbb{P}(\mathbf{Y}_{-1} = (0,0), \mathbf{Y}_0 = (0,0), \mathbf{Y}_1 = (1,0), \mathbf{Y}_2 = (i,0), \mathbf{Y}_3 = (0,0), \mathbf{Y}_4 = (0,0)) =\\ &= \mathbb{P}(\mathbf{Y}_{-1} = (0,0), \mathbf{Y}_0 = (0,0), \mathbf{Y}_1 = (i,0), \mathbf{Y}_2 = (1,0), \mathbf{Y}_3 = (0,0), \mathbf{Y}_4 = (0,0)). \quad (7.3)\end{aligned}$$

Furthermore, by Definition 7.1.1, it holds that

$$p_{\mathbf{Y}}((a,0)|(b,0)) = \sum_{l=0}^{\min(a,b)} p_\epsilon(a-l)\frac{b!}{l!(b-l)!}(1-\alpha_1-\alpha_2)^{b-l}\alpha_1^l \quad \text{for } a, b \in \mathbb{N}_0,$$

implying $p_{\mathbf{Y}}((0,0)|(0,0)) > 0$. Following Alzaid and Al-Osh (1990), the process $(\mathbf{Y}_t)_{t\in\mathbb{Z}}$ is a stationary Markov process on the state space $\mathbb{N}_0^2$, under the assumption $p_\epsilon(0) > 0$ it follows that it is both irreducible and aperiodic, cp. Lemma 4.1.1. Hence, each state $(a, b) \in \mathbb{N}_0$ is positive recurrent and, in particular, $p_{\mathbf{Y}}((0,0)) > 0$. This implies that (7.3) is equivalent to

$$\begin{aligned}&p_\epsilon(1)\left[p_\epsilon(i)(1-\alpha_1-\alpha_2) + p_\epsilon(i-1)\alpha_1\right]p_\epsilon(0)(1-\alpha_1-\alpha_2)^i\\ &= p_\epsilon(i)\left[p_\epsilon(1)(1-\alpha_1-\alpha_2)^i + p_\epsilon(0)\binom{i}{i-1}(1-\alpha_1-\alpha_2)^{i-1}\alpha_1\right](1-\alpha_1-\alpha_2)p_\epsilon(0).\end{aligned}$$

This last expression is easily seen to be equivalent to (4.10), as $1 - \alpha_1 - \alpha_2 > 0$. Thus, the proof of Theorem 4.1.8 applies *mutatis mutandis*, proving the assertion for $p = 2$.

Now, let $p > 2$. Replace the process $(\mathbf{Y}_t)_{t\in\mathbb{Z}}$ with the process (see Sect. 4 in Alzaid and Al-Osh (1990))

$$\mathbf{Y}_t^* = \left(Y_t, \sum_{i=2}^{p}\alpha_i Y_{t+1-i}, \sum_{i=3}^{p} Y_{t+2-i}, \ldots, \alpha_p \circ Y_{t-1}\right).$$

For the sequence of events $\mathbf{Y}^*_{-p+1} = (0, 0, \ldots, 0) = \mathbf{Y}^*_{-p+2} = \cdots = \mathbf{Y}^*_0$, together with $\mathbf{Y}^*_{-p+1} = \mathbf{Y}^*_3 = \cdots = \mathbf{Y}^*_{p+2}$ and $\mathbf{Y}^*_1 = (1, 0, \ldots, 0)$ as well as $\mathbf{Y}^*_2 = (i, 0, \ldots, 0)$, a short moment of reflection should convince the reader that an extension of (7.3) holds. Again, this relation can be shown to be equivalent to (4.10), concluding the proof. □

Hence we find that while certain characteristics of the AAINAR(p) model do not generalize the behavior of the INAR(1) model as anticipated, it still mimics the time-reversibility of the INAR(1) process rather well. In the remainder of this chapter we will discuss several ways in which the formulation of Du and Li (1991) differs from this behavior.

## 7.2 The INAR(p) Model of Du and Li

In this section a competing formulation for an INAR(p) process given in Du and Li (1991) is considered.

**Definition 7.2.1** (DLINAR(p) Process)**.** *Let $(\epsilon_t)_{t\in\mathbb{Z}}$ be an i.i.d. process with range $\mathbb{N}_0$, let $\sigma_\epsilon^2 < \infty$. Let $p \in \mathbb{N}$, $\alpha_1, \ldots, \alpha_p \in [0,1)$ with $\sum_{i=1}^{p} \alpha_i < 1$. A process $(Y_t)_{t\in\mathbb{Z}}$, which follows the recursion (7.1) for all $t \in \mathbb{Z}$ is said to be an DLINAR(p) process, if all thinning operations are mutually independent and if the $(\epsilon_t)_{t\in\mathbb{Z}}$ are independent of all thinning operations and the random variable $\epsilon_t$ is independent of $Y_{t-k}$ and its survivals (and the thinning operations) $a_j \circ Y_{t-k}$ for $j = 1, 2, \ldots, p$ and $k > 0$.*

In contrast to the AAINAR(p) model, this process is a $p$-th order Markov chain. For the transition probabilities, the definition implies (cf. Bu and McCabe (2008), (2))

$$\begin{aligned} &p_Y(a|b_1, \ldots b_p) \\ &= \sum_{l_k \le b_k, \sum l_k \le a} \mathbb{P}\left(\epsilon_0 = a - \sum l_i\right) \binom{b_1}{l_1} \alpha_1^{l_1} (1-\alpha_1)^{b_1 - l_1} \cdots \binom{b_p}{l_p} \alpha_p^{l_p} (1-\alpha_p)^{b_p - l_p} \end{aligned} \quad (7.4)$$

for $a, b_i \in \mathbb{N}_0$. For a representation of this process similar to that of (7.2) we refer to the next section. The ACF of a DLINAR(p) process can be shown to coincide with that of a Gaussian AR(p) process as it satisfies

$$\rho(k) = \sum_{i=1}^{p} \alpha_i \rho(k-i) \quad \text{for } k \ge 1 \quad (7.5)$$

(cf. eq. (3.5) in Du and Li (1991)), demonstrating another crucial difference between these models.

### 7.2.1 Stationary Distribution

In this section, we concern ourselves with the stationary distribution of DLINAR(p) processes. It will turn out that these may be represented in a fashion similar to the quasi-linear representations of (4.2) and (7.2) but with certain differences. Let us begin by defining the following functions $f_i(\epsilon, \boldsymbol{\alpha})$, with arguments $\epsilon$, a random variable and $\boldsymbol{\alpha}$, the shorthand notation for the parameter vector $(\alpha_1, \ldots, \alpha_p)$, where

$$f_0(\epsilon, \boldsymbol{\alpha}) = \epsilon \quad \text{and} \quad f_i(\epsilon, \boldsymbol{\alpha}) := \sum_{\substack{i_1, \ldots, i_k \ge 1 \\ \sum i_l = i}} \left[ (\alpha_{i_1} \cdots \alpha_{i_k}) \circ \epsilon \right]. \quad (7.6)$$

With the help of these functions, we can state the first result on the stationary distribution.

**Lemma 7.2.2.** *For a stationary DLINAR(p) Process it holds that*

$$Y_t \overset{\mathcal{D}}{=} \sum_{i=0}^{\infty} f_i\left(\epsilon_{t-i}, \boldsymbol{\alpha}\right), \tag{7.7}$$

*where the thinning operations within each $f_i(\epsilon_{t-i}, \boldsymbol{\alpha})$ are mutually independent.*

*Proof.* We start this proof with an induction argument for an altered process. Let $\epsilon'_p \equiv 0$ for all $p > 0$ and let $(Y'_{-p}, \ldots, Y'_{-1}) = (0, 0, \ldots, 0)$. Let the process $(Y'_t)_{\mathbb{Z}}$ satisfy Definition 7.2.1 for $t \geq 0$, then $Y'_0 \sim \epsilon_0 = f_0(\epsilon_0, \boldsymbol{\alpha})$. Now, we assume that $Y_t \overset{\mathcal{D}}{=} \sum_{i=0}^{t} f_i(\epsilon_{t-i}, \boldsymbol{\alpha})$ for some $t \in \mathbb{N}$, where the thinning operations in each $f_i(\epsilon_{t-i}, \boldsymbol{\alpha})$ are mutually independent. Denoting the thinning operations involved by $\circ^t$ if the thinning operator is applied at time $t$, we find for all $i = 0, 1, 2, \ldots$

$$\begin{aligned}
Y'_{t+1} &\overset{\mathcal{D}}{=} \sum_{j=1}^{p} \alpha_j \circ^{t+1} \sum_{\substack{i_1, \ldots, i_k \geq 1 \\ \sum i_l = i}} \left[(\alpha_{i_1} \cdots \alpha_{i_k}) \circ^{t+1-i_1-j} \epsilon'_{t+1-i-j}\right] \\
&\overset{\mathcal{D}}{=} \sum_{j=1}^{p} \sum_{\substack{i'_1, \ldots, i'_k \geq 1; i'_1 = j \\ \sum i'_l = i+j}} \left[\left(\alpha_{i'_1} \cdots \alpha_{i'_k}\right) \circ^{t+1} \epsilon'_{t+1-i-j}\right],
\end{aligned}$$

for the latter equality notice that the thinning operations at different times (notice that $j > 0$) and apply (4.11). Rearranging yields

$$\begin{aligned}
Y'_{t+1} &\overset{\mathcal{D}}{=} \sum_{j=1}^{p} \left( \sum_{m=1}^{t+1} \sum_{\substack{i'_1, \ldots, i'_k \geq 1; i'_1 = j \\ \sum i'_l = m}} \left[\left(\alpha_{i'_1} \cdots \alpha_{i'_k}\right) \circ \epsilon'_{t+1-m}\right] \right) + \epsilon'_{t+1} \\
&\overset{\mathcal{D}}{=} \sum_{m=1}^{t+1} \left( \sum_{\substack{m_1, \ldots, m_k \geq 1 \\ \sum m_l = m}} \left[(\alpha_{m_1} \cdots \alpha_{m_k}) \circ \epsilon'_{t+1-m}\right] \right) + \epsilon'_{t+1} \overset{\mathcal{D}}{=} \sum_{m=0}^{t+1} f_m\left(\epsilon'_{t+1-m}, \boldsymbol{\alpha}\right),
\end{aligned} \tag{7.8}$$

where the second equality in distribution is due to the fact that the thinning operations in the resulting sum are executed mutually independently, for any given fixed $j \in \{1, \ldots, p\}$ by induction assumption, for varying $j$ due to the independence of the $\epsilon_t$ for all $t \in \mathbb{Z}$. The independence of the thinning operations at different time instants given by Definition 7.2.1 shows the induction step. As a result, both $(Y'_t)_{t\in\mathbb{N}}$ and $(Y_t)_{t\in\mathbb{Z}}$ are $p$-th order irreducible and aperiodic Markov chains, i.e., the vectors $(Y'_{t+1}, \ldots, Y'_{t+p})_{t\in\mathbb{N}}$ and

$(Y_{t+1},\ldots,Y_{t+p})_{t\in\mathbb{Z}}$ are irreducible and aperiodic Markov chains on $\mathbb{N}^p$. Since the respective transition probabilities are equal, both Markov chains have the same stationary distribution if it exists. Thus, it follows that the stationary distribution of $(Y_t')_{t\in\mathbb{N}}$ and $(Y_t)_{t\in\mathbb{Z}}$ are equal.

For any process $(Y_t)_{t\in\mathbb{Z}}$ to be stationary, we necessarily have to initialize it at some point $t_0$ with the stationary distribution (if it exists). For the DLINAR(p) process $(Y_t)_{t\in\mathbb{Z}}$ we have just shown that we may use the stationary distribution of $(Y_t')_{t\in\mathbb{N}}$ for this purpose. Thus, $Y_0 \sim \sum_{i=0}^{\infty} f_i(\epsilon_{t-i},\boldsymbol{\alpha})$ with independent thinning operations. It is easily seen that the considerations of (7.8) and following it hold for $t\to\infty$ and in the limit as well (if it exists). This concludes the proof. □

The similarities between the result of Lemma 7.2.2 and the representations (4.2) and (7.2) are obvious, yet the following results will highlights essential differences. Let us first calculate the pgf of the functions $f_i(\cdot)$ of (7.6).

**Lemma 7.2.3.** *For all $i\geq 0$ and all $t\in\mathbb{Z}$ we have*

$$\mathrm{pgf}_{f_i(\epsilon_{t-i},\boldsymbol{\alpha})}(z)=\mathrm{pgf}_{\epsilon_t}\left(\prod_{\substack{i_1,\ldots,i_k\geq 1\\ \sum i_j=i}}\left(1-\prod_{j=1}^{k}\alpha_{i_j}+\prod_{j=1}^{k}\alpha_{i_j}z\right)\right).$$

*Proof.* First, we notice that by (4.11), for independent thinning operations,

$$\mathrm{pgf}_{\alpha_1\circ\alpha_2\circ\cdots\circ\alpha_k\circ X}(z)=\mathrm{pgf}_X\left(1-\prod_{i=1}^{k}\alpha_i+\prod_{i=1}^{k}\alpha_i z\right).$$

Moreover, we have for any $\alpha,\beta\in[0,1]$, $X$ as above and $\xi_i,\zeta_i$ mutually independent Bernoulli random variables with $\mathbb{E}[\xi_i]=\alpha$, $\mathbb{E}[\zeta_i]=\beta$ that

$$\begin{aligned}\mathrm{pgf}_{\alpha\circ X+\beta\circ X}(z)= &\sum_{l=0}^{\infty}\mathbb{E}\left[z^{\alpha\circ X+\beta\circ X}|X=l\right]\mathbb{P}(X=l)=\sum_{l=0}^{\infty}\mathbb{E}\left[z^{\sum_{i=1}^{l}\xi_i+\sum_{i=1}^{l}\zeta_i}\right]\mathbb{P}(X=l)\\ =&\sum_{l=0}^{\infty}\left(\mathbb{E}\left[z^{\xi_1}\right]\mathbb{E}\left[z^{\zeta_1}\right]\right)^{l}\mathbb{P}(X=l)=\mathrm{pgf}_X\left((1-\alpha+\alpha z)(1-\beta+\beta z)\right),\end{aligned}$$

where the penultimate equation holds due to the mutual independence of the $\xi_i,\zeta_i$ as well as their independence of $X$. We now calculate for each $i\geq 0$,

$$\mathrm{pgf}_{f_i(\epsilon_{t-i},\boldsymbol{\alpha})}(z)=\mathrm{pgf}_{\sum[(\alpha_{i_1}\cdots\alpha_{i_k})\circ\epsilon_{t-i}]}(z)=\mathrm{pgf}_{\epsilon_t}\left(\prod_{\substack{i_1,\ldots,i_k\geq 1\\ \sum i_j=i}}\left(1-\prod_{j=1}^{k}\alpha_{i_j}+\prod_{j=1}^{k}\alpha_{i_j}z\right)\right).$$

Since the $\epsilon_t$ were assumed to be i.i.d., this proves the assertion for all $t\in\mathbb{Z}$. □

With these results we are able to state the main result of this section. It shows that the assumptions made on DLINAR(p) processes in Definition 7.2.1 suffice to ensure the existence (and the form) of the stationary distribution as given by Lemma 7.2.2.

**Theorem 7.2.4.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a DLINAR(p) process as given in Definition 7.2.1. Then the stationary distribution of $(Y_t)_{t\in\mathbb{Z}}$ exists and is given by* (7.7).

*Proof.* Any stationary process satisfying Definition 7.2.1 necessarily satisfies (7.7) as well, and thus, by Lemma 7.2.3 we formally have

$$\operatorname{pgf}_{Y_t}(z) = \prod_{i=0}^{\infty} \operatorname{pgf}_{\epsilon_t}\left( \prod_{\substack{i_1,\ldots,i_k\geq 1\\ \sum i_j = i}} \left(1 - \prod_{j=1}^{k} \alpha_{i_j} + \prod_{j=1}^{k} \alpha_{i_j} z\right)\right). \tag{7.9}$$

At this point, it is unclear whether the above expression converges for any $z \in [0,1]$ or not. At the same time, if the above expression can be shown to converge for all $z \in [0,1]$, it is immediately clear that the random variable generated from this pgf is that of the stationary distribution (as the $\epsilon_t$ are assumed to be i.i.d.).

Quite obviously it holds that $\operatorname{pgf}_{Y_t}(1) = 1$. The argument of $\operatorname{pgf}_{\epsilon_t}$ in (7.9) is a monotonically increasing function in $z$, and each $\operatorname{pgf}_{\epsilon_t}(z)$ is monotonically increasing in $z$. Thus by the Weierstrass M-test, (7.9) converges or diverges with $\operatorname{pgf}_{Y_t}(0)$. This infinite product converges absolutely or diverges with the infinite series

$$\sum_{i=0}^{\infty} \left[1 - \operatorname{pgf}_{\epsilon_{t-i}}\left( \prod_{\substack{i_1,\ldots,i_k\geq 1\\ \sum i_j = i}} \left(1 - \prod_{j=1}^{k} \alpha_{i_j}\right)\right)\right] \tag{7.10}$$

as each term of the product in (7.9) is positive and less than unity, also see the proof of Lemma 4.2.4. Notice that for each $M \in \mathbb{N}$ and probabilities $p_1, \ldots, p_M$ we can show the following inequality by induction:

$$\prod_{i=1}^{M} (1 - p_i) \geq 1 - \sum_{i=1}^{M} p_i.$$

Now, referring to Lemma 1 in Heathcote (1965)[1] we find that under the assumption $\mathbb{E}[\epsilon_t] < \infty$, a sufficient condition for (7.10) to converge is the convergence of

$$\sum_{i=0}^{\infty} \left[1 - \prod_{\substack{i_1,\ldots,i_k\geq 1\\ \sum i_j = i}} \left(1 - \prod_{j=1}^{k} \alpha_{i_j}\right)\right] \leq \sum_{i=0}^{\infty} \sum_{\substack{i_k\\ \sum i_k = i}} \prod_{j=1}^{k} \alpha_{i_j} < \infty. \tag{7.11}$$

[1] We acknowledge that a later paper Heathcote (1966) corrected a mistake made in the cited Lemma 1. However, this mistake concerns only the necessary condition, not the sufficient condition, which is all that is needed for this proof.

For the summands of the RHS of this expression, we have

$$\sum_{\substack{i_1,\ldots,i_k\geq 1\\ \sum i_j=i}} \prod_{j=1}^{k} \alpha_{i_j} = \sum_{\substack{i_1,\ldots,i_k\geq 1\\ \sum i_j=i}} \prod_{1\leq l\leq p} \alpha_l^{l_j} = \sum_{\sum j\cdot l_j=i} \binom{\sum_{j=1}^{p} l_j}{l_1,\ldots,l_p} \prod_{1\leq l\leq p} \alpha_l^{l_j}.$$

It is easily seen that

$$\bigcup_{n\in\mathbb{N}} \left\{(l_1,\ldots,l_p)\in\mathbb{N}^p \middle| \sum_{i=1}^{p} l_i = n\right\} = \bigcup_{n\in\mathbb{N}} \left\{(l_1,\ldots,l_p)\in\mathbb{N}^p \middle| \sum_{i=1}^{p} i\cdot l_i = n\right\}.$$

With the multinomial theorem, we find the following upper bound for (7.11):

$$\sum_{i=0}^{\infty}\left[1-\prod_{\substack{i_1,\ldots,i_k\geq 1\\ \sum i_j=i}}\left(1-\prod_{j=1}^{k}\alpha_{i_j}\right)\right] \leq \sum_{i=0}^{\infty}\left(\sum_{l=1}^{p}\alpha_l\right)^i,$$

where convergence is ensured due to $\sum_{l=1}^{p}\alpha_l < 1$. This concludes the proof.

□

### 7.2.2 Connections to Branching Process Theory

In connection with an DLINAR(p) process $(Y_t)_{t\in\mathbb{Z}}$, several contributions such as Dion et al. (1995), Silva and Silva (2006) as well as the original contribution Du and Li (1991) have employed the following matrix:

$$\mathbf{M} = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_p \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \vdots & \vdots \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

Defining $\mathbf{Z}_t := (Y_t, Y_{t-1},\ldots,Y_{t-p})$ and $\mathbf{I}_t = (\epsilon_t, 0,\ldots,0)$, it is easily seen that, in distribution,

$$\begin{pmatrix} Y_{t+1} \\ Y_t \\ \vdots \\ Y_{t-p+1} \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_p \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix} \circ \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{7.12}$$

Here, the "$\circ$" operator is applied to matrices in the same fashion that multiplication is applied in the usual matrix multiplication, and, for iterations of this procedure we will use the exponent "$\circ t$" instead of $t$ to mark this difference. Let us record some consequences of this representation of the process.

**Lemma 7.2.5** (Cp. Dion et al. (1995))**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be an DLINAR(p) process as given by Definition 7.2.1. Then the following holds:*

*(i) The process $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ is a multitype Galton-Watson process with immigration vectors $\mathbf{I}_t = (\epsilon_t, 0, \ldots, 0)$ and matrix of offspring $\mathbf{M}$.*

*(ii) Let $\mathbf{Z}_0 = (i_1, \ldots, i_p)$ be any vector of nonnegative integers, then*

$$\mathbf{Z}_t \overset{\mathcal{D}}{=} \sum_{\nu=0}^{t-1} \mathbf{Y}_{t,\nu} + \mathbf{X}_t \overset{\mathcal{D}}{=} \sum_{s=0}^{t-1} \mathbf{M}^{\circ s} \circ \mathbf{I}_{t-s} + \mathbf{M}^{\circ t} \circ \mathbf{Z}_0, \tag{7.13}$$

*where $\mathbf{Y}_{t,\nu}$ is a $\nu$-th generation multitype Galton-Watson process (without immigration) with random initial vector distributed as $\mathbf{I}_{t-\nu}$. $\mathbf{X}_t$ is a $t$-th generation multitype Galton-Watson process with fixed initial vector $(i_1, \ldots, i_p)$. The random vectors on the right of* (7.13) *are independent.*

*(iii) For any $t \in \mathbb{Z}$,*

$$\mathbf{Z}_t \overset{\mathcal{D}}{=} \sum_{\nu=0}^{\infty} \mathbf{M}^{\circ \nu} \circ \mathbf{I}_{t-\nu}.$$

*(iv) Let $m_{i,j}(\nu)$ denote the $(i,j)$-th entry of the matrix $\mathbf{M}^{\circ\nu}$. Then, for $1 \le i \le p$ and $\nu \in \mathbb{N}$,*

$$m_{i,1}(\nu) \circ \epsilon_t \overset{\mathcal{D}}{=} f_{\nu-i+1}(\epsilon_t, \boldsymbol{\alpha}),$$

*where we define $f_{-i}(\epsilon_t, \boldsymbol{\alpha}) := 0$ for $i \in \mathbb{N}$.*

*Proof.* The assertion (i) is almost precisely Proposition A Dion et al. (1995), with two small differences. First, we replaced the name Bienamé-Galton-Watson-process with immigration with the name multitype Galton-Watson, as this is the name employed by Mode (1971) to which the cited paper refers. Second, the matrix of offspring means becomes the matrix of offspring, as the DLINAR(p) model discussed here is restricted to Bernoulli offspring distributions. We refer to the comment in Dion et al. (1995) after Corollary 1. For (ii), we use (i), the first equation follows with eq. (2.7.2) in Mode (1971) or eq. (7) in Dion et al. (1995). The second equation follows by iteratively applying (7.12). The condition $\sum_{i=1}^{p} \alpha_i < 1$ implies that the spectral radius of the matrix $\mathbf{M}$ is less than unity (cf. Proposition B in Dion et al. (1995)). The resultant expression of (iii) can be found in eq. (9) in Dion et al. (1995) as well as eq. (4a) in Silva and Silva (2006). Now, we use induction over $\nu$ to show (iv). For $\nu = 1$, the statement clearly holds, as $f_0(\epsilon_t, \boldsymbol{\alpha}) = \epsilon_t$ and $f_1(\epsilon_t, \boldsymbol{\alpha}) = \alpha_1 \circ \epsilon_t$. Assume the statement holds for some $\nu \in \mathbb{N}$. Simple matrix multiplication of $\mathbf{M} \circ \mathbf{M}^{\circ\nu}$ yields the equations

$$m_{1,1}(\nu+1) \overset{\mathcal{D}}{=} \sum_{i=1}^{p} \alpha_i \circ m_{i,1}(\nu) \quad \text{and} \quad m_{i,1}(\nu+1) = m_{i-1,1}(\nu) \quad \text{for } i = 2, \ldots, p.$$

The latter equations shows that the statement holds for $i = 2, \ldots, p$. For the former we find, similarly to the proof of Lemma 7.2.2,

$$m_{1,1}(\nu+1) \circ \epsilon_t \stackrel{\mathcal{D}}{=} \sum_{i=1}^{p} \alpha_i \circ m_{i,1}(\nu) \circ \epsilon_t \stackrel{\mathcal{D}}{=} \sum_{i=1}^{p} \alpha_i \circ f_{\nu-i+1}(\epsilon_t, \boldsymbol{\alpha}) \stackrel{\mathcal{D}}{=} f_{\nu+1}(\epsilon_t, \boldsymbol{\alpha}).$$

The justifications for these equalities are the same as the ones given in the proof of the referenced Lemma.

□

We conclude this section with the following remark: On p. 131 in Dion et al. (1995), it is stated that

$$Y_t \stackrel{\mathcal{D}}{=} \mathrm{Poi}\left(\frac{\lambda}{1-\sum_{i=1}^{p}\alpha_i}\right).$$

However, it is clear from Theorem 7.2.4 that this result cannot hold, as the resultant pgf is that of a Compound Poisson distribution, not a Poisson distribution (excepting the well known results in case $p = 1$). The mistake made in the cited paper seems to be in the derivation of the stationary distribution from their equation (10). Their derivation only holds if, for Poisson random variables $\epsilon_{n-\nu}$, the sum $\sum_{i=1}^{d} m_{i1}(\nu) \circ \epsilon_{n-\nu}$ is Poisson distributed again. From Lemma 7.2.3 we can see that this is only satisfied in degenerate cases, or if $d = 1$.

### 7.2.3 Joint Cumulants of DLINAR(p) Processes

In the following sections we will derive asymptotic results similar to those of Section 5.3 for DLINAR(p) processes, and it is quite clear that we need explicit expressions for the joint cumulants of DLINAR(p) processes. The first relation is an immediate consequence of the argumentation leading up to (4.8).

**Lemma 7.2.6.** *Let $t \in \mathbb{Z}$, let $\epsilon_t \sim \mathrm{Poi}(\lambda)$ and let $0 \le i \le j \le k \le l$. Then*

$$\mathrm{cum}\left(f_i(\epsilon_t,\boldsymbol{\alpha}), f_j(\epsilon_t,\boldsymbol{\alpha}), f_k(\epsilon_t,\boldsymbol{\alpha}), f_l(\epsilon_t,\boldsymbol{\alpha})\right) = \lambda \mathbb{E}\left[f_i(\mathbf{1},\boldsymbol{\alpha}) f_j(\mathbf{1},\boldsymbol{\alpha}) f_k(\mathbf{1},\boldsymbol{\alpha}) f_l(\mathbf{1},\boldsymbol{\alpha})\right],$$

*where $\mathbf{1}$ denotes a constant random variable with value* 1.

We point out that the calculation of the expression $\mathbb{E}[f_i(\mathbf{1},\boldsymbol{\alpha}) f_j(\mathbf{1},\boldsymbol{\alpha}) f_k(\mathbf{1},\boldsymbol{\alpha}) f_l(\mathbf{1},\boldsymbol{\alpha})]$ is very complex for $p \ge 2$. In the case $p = 1$, a closed expression is calculated in (4.9). The result of Lemma 7.2.6 can now be employed to calculate the fourth joint cumulant of DLINAR(p) processes.

**Theorem 7.2.7.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson DLINAR(p) Process, where $\epsilon_t \sim \mathrm{Poi}(\lambda)$ for all $t \in \mathbb{Z}$. Then, for $i, j, k, l \in \mathbb{Z}$ with $i \le j \le k \le l$, the fourth joint cumulant of the process calculates to*

$$\mathrm{cum}\left(Y_i, Y_j, Y_k, Y_l\right) = \lambda \sum_{\nu=0}^{\infty} \mathbb{E}\left[f_\nu(\mathbf{1},\boldsymbol{\alpha}) f_{\nu+j-i}(\mathbf{1},\boldsymbol{\alpha}) f_{\nu+k-i}(\mathbf{1},\boldsymbol{\alpha}) f_{\nu+l-i}(\mathbf{1},\boldsymbol{\alpha})\right].$$

*Proof.* Notice that $Y_t = \mathbf{e}_1 \cdot \mathbf{Z}_t$, where $e_1 = (1, 0, \ldots, 0)$. Using Lemma 7.2.5 (ii), we have

$$(Y_i, Y_j, Y_k, Y_l) = (\mathbf{e}_1 \circ \mathbf{Z}_i, \mathbf{e}_1 \circ \mathbf{Z}_j, \mathbf{e}_1 \circ \mathbf{Z}_k, \mathbf{e}_1 \cdot \mathbf{Z}_l) \stackrel{\mathcal{D}}{=}$$
$$\Big(\mathbf{e}_1 \cdot \mathbf{Z}_i,\ \mathbf{e}_1 \cdot \sum_{s=0}^{j-i-1} \mathbf{M}^{\circ s} \circ \mathbf{I}_{j-s} + \mathbf{e}_1 \cdot \mathbf{M}^{\circ j-i} \circ \mathbf{Z}_i,$$
$$\mathbf{e}_1 \cdot \sum_{s=0}^{k-i-1} \mathbf{M}^{\circ s} \circ \mathbf{I}_{k-s} + \mathbf{e}_1 \cdot \mathbf{M}^{\circ k-i} \circ \mathbf{Z}_i,\ \mathbf{e}_1 \cdot \sum_{s=0}^{l-i-1} \mathbf{M}^{\circ s} \circ \mathbf{I}_{l-s} + \mathbf{e}_1 \cdot \mathbf{M}^{\circ l-i} \circ \mathbf{Z}_i\Big).$$

Using the multilinearity of the cumulant function (see Lemma 2.2.1) together with the mutual independence of the random vectors in (7.13), we find

$$\operatorname{cum}(Y_i, Y_j, Y_k, Y_l) = \operatorname{cum}\Big(\mathbf{e}_1 \cdot \mathbf{Z}_i,\ \mathbf{e}_1 \cdot \mathbf{M}^{\circ j-i} \circ \mathbf{Z}_i,\ \mathbf{e}_1 \cdot \mathbf{M}^{\circ k-i} \circ \mathbf{Z}_i,\ \mathbf{e}_1 \cdot \mathbf{M}^{\circ l-i} \circ \mathbf{Z}_i\Big).$$

With Lemma 7.2.5 (iii), it follows that

$$\Big(\mathbf{e}_1 \cdot \mathbf{Z}_i,\ \mathbf{e}_1 \cdot \mathbf{M}^{\circ j-i} \circ \mathbf{Z}_i,\ \mathbf{e}_1 \cdot \mathbf{M}^{\circ k-i} \circ \mathbf{Z}_i,\ \mathbf{e}_1 \cdot \mathbf{M}^{\circ l-i} \circ \mathbf{Z}_i\Big)$$
$$\stackrel{\mathcal{D}}{=} \Bigg(\mathbf{e}_1 \cdot \sum_{\nu=0}^{\infty} \mathbf{M}^{\circ \nu} \circ \mathbf{I}_{i-\nu}, \mathbf{e}_1 \cdot \mathbf{M}^{\circ j-i} \circ \sum_{\nu=0}^{\infty} \mathbf{M}^{\circ \nu} \circ \mathbf{I}_{i-\nu},$$
$$\mathbf{e}_1 \cdot \mathbf{M}^{\circ k-i} \circ \sum_{\nu=0}^{\infty} \mathbf{M}^{\circ \nu} \circ \mathbf{I}_{i-\nu}, \mathbf{e}_1 \cdot \mathbf{M}^{\circ l-i} \circ \sum_{\nu=0}^{\infty} \mathbf{M}^{\circ \nu} \circ \mathbf{I}_{i-\nu}\Bigg),$$

which, together with

$$\mathbf{e}_1 \cdot \mathbf{M}^{\circ a} \circ \mathbf{I}_b = (m_{1,1}(a), m_{2,1}(a), \ldots, m_{p,1}(a)) \circ \mathbf{I}_b = f_a(\epsilon_b, \boldsymbol{\alpha})$$

implies

$$\operatorname{cum}(Y_i, Y_j, Y_k, Y_l)$$
$$= \sum_{\nu=0}^{\infty} \operatorname{cum}\left(f_\nu(\epsilon_{i-\nu}, \boldsymbol{\alpha}), f_{\nu+j-i}(\epsilon_{i-\nu}, \boldsymbol{\alpha}), f_{\nu+k-i}(\epsilon_{i-\nu}, \boldsymbol{\alpha}), f_{\nu+l-i}(\epsilon_{i-\nu}, \boldsymbol{\alpha})\right).$$

We used Lemma 7.2.5 (iv) and Lemma 2.2.1 (v) repeatedly. Both results are applicable, as $\mathbb{E}[Y_i^4] < \infty$. Application of Theorem 7.2.6 yields the result. □

In order to generalize the findings of Section 5.3 to DLINAR(p) processes, it suffices to show that the conditions of both Theorem 2.6.2 and Theorem 2.6.3 are satisfied. Note that the assumption $\sum_{i=1}^{p} \alpha_i < 1$ was shown in Proposition B of Dion et al. (1995) to be equivalent to the statement that the spectral radius of the matrix $\mathbf{A}$ is less than unity. The latter statement implies convergence of the series of $\sum_{i=0}^{\infty} \mathbf{A}^i$, hence Theorem 1 ofSilva and Silva (2006) is applicable. Together with (7.5), we have proven the following assertion:

**Theorem 7.2.8.** *Let $(Y_t)_{t \in \mathbb{Z}}$ be a Poisson DLINAR(p) process with $\epsilon_t \sim \operatorname{Poi}(\lambda)$ for all $t \in \mathbb{Z}$. Let $q \geq 1$ and denote $K := p + q$. Then $(Y_t)_{t \in \mathbb{Z}}$ satisfies the conditions of both Theorem 2.6.2 and Theorem 2.6.3.*

### 7.2.4 Concerning Time-Reversibility

Let us now show the following analogue of Theorem 7.1.2. In this section, we abbreviate the sequence of states $Y_0 = a_1, Y_1 = a_2, \dots Y_l = a_l$ by writing $\overline{a_1, a_2, \dots, a_l}$.

**Lemma 7.2.9** (Schweer (2015b), Lemma 19.1)**.** *For $p > 1$, let $(Y_t)_{t\in\mathbb{Z}}$ be a time-reversible DLINAR(p) process and let $\mathbb{P}(\epsilon_0 = 0) \in (0, 1)$. Then there exists a $\lambda > 0$ such that $\epsilon_0 \sim \mathrm{Poi}(\lambda)$ .*

*Proof.* First, let $p = 2$. By the time-reversibility of the process $(Y_t)_{t\in\mathbb{Z}}$, it follows that for any $i \in \mathbb{N}_0$, $\mathbb{P}\left(\overline{0, 0, 1, i, 0, 0}\right) = \mathbb{P}\left(\overline{0, 0, i, 1, 0, 0}\right)$. Using the Markovian structure of $(Y_t)_{t\in\mathbb{Z}}$, the notation of (7.4) and the fact that $p_Y(0) > 0$ as well as $p_Y(0|0) > 0$ (which is deduced analogously as in Theorem 7.1.2), this is equivalent to

$$p_Y(0|0, i)p_Y(0|i, 1)p_Y(i|1, 0)p_Y(1|0, 0) = p_Y(0|0, 1)p_Y(0|1, i)p_Y(1|i, 0)p_Y(i|0, 0),$$

reminiscent of Kolmogorov's criterion of Theorem 2.4.2. With simple manipulations, this is equivalent to (4.10) as $(1 - \alpha_2) > 0$ by Definition 7.2.1. For $p > 2$, consideration of the sequence $\overline{0, 0, \dots, 0, 1, i, 0, \dots, 0}$ and its inverse where the dots represent $p$ zeroes, yields an equivalent relation as above, as $(1 - \alpha_k) > 0$ holds for all $k = 3, \dots, p$. The assertion follows analogously to Theorem 4.1.8. □

At first glance, since the result of Lemma 7.2.9 is exactly analogous to that of Theorem 7.1.2 there doesn't seem to be a difference in the characteristics of DLINAR(p) processes and AAINAR(p) processes with respect to time-reversibility. However, the following result shows that a DLINAR(p) process is time-reversible only if the parameters take on degenerate values, i.e., if it is in fact an INAR(1) process.

**Theorem 7.2.10** (Schweer (2015b), Theorem 19.3)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a time-reversible DLINAR(p) process with $p > 1$ and $\alpha_1 > 0$ and let $0 < \mathbb{P}(\epsilon = 0) < 1$. Then $\alpha_j = 0$ for $j = 2, \dots, p$.*

*Proof.* Let $p = 2$ and let $(1 - \alpha_1) := \overline{\alpha_1}, (1 - \alpha_2) := \overline{\alpha_2}$. Since $(Y_t)_{t\in\mathbb{Z}}$ is time-reversible, the transition probabilities necessarily satisfy $\mathbb{P}\left(\overline{0, 0, 1, 3, 2, 0, 0}\right) = \mathbb{P}\left(\overline{0, 0, 2, 3, 1, 0, 0}\right)$. By (7.4) and the fact that $p_Y(0|0), p_Y(0) > 0$, this is equivalent to

$$\overline{\alpha_2}^5\overline{\alpha_1}^2\lambda \left[ \sum_{\substack{l_1\le 3, l_2\le 1\\ l_1+l_2\le 2}} \frac{\lambda^{2-l_1-l_2}}{(2-l_1-l_2)!}\binom{3}{l_1}\alpha_1^{l_1}\overline{\alpha_1}^{3-l_1}\alpha_2^{l_2}\overline{\alpha_2}^{1-l_2}\right]\left(\frac{\lambda^3}{6}\overline{\alpha_1} + \frac{\lambda^2}{2}\alpha_1\right) =$$

$$\overline{\alpha_2}^4\overline{\alpha_1}\frac{\lambda^2}{2}\left[\lambda\overline{\alpha_1}^3\overline{\alpha_2}^2 + 2\overline{\alpha_1}^3\overline{\alpha_2}\alpha_2 + 3\overline{\alpha_1}^2\alpha_1\overline{\alpha_2}^2\right]\left(\frac{\lambda^3}{6}\overline{\alpha_1}^2 + \lambda^2\overline{\alpha_1}\alpha_1 + \lambda\alpha_1^2\right). \tag{7.14}$$

This, in turn (recall that $\lambda > 0$ and $\overline{\alpha_1}, \overline{\alpha_2} > 0$ by Definition 7.2.1, is equivalent to $\frac{1}{2}\overline{\alpha_1}\alpha_1^2\alpha_2 = 0$. Since $\overline{\alpha_1} > 0$, this implies the assertion for $p = 2$.

Now, let $p > 2$ be arbitrary but fixed. It remains to be seen that $\alpha_j = 0$ for $j = 2, \dots, p$ which is done by first showing that $\alpha_2 = 0$ and then proceeding inductively. An appeal to (7.4) reveals that the relations considered in the first part of this proof hold similarly for the sequence $\overline{0,0,\ldots,0,1,3,2,0,\ldots,0}$ and its inverse, here the first and last $p$ entries are 0. Since there are at most three consecutive nonzero states in these sequences, the relation of the transition probabilities is equivalent to that of (7.14) save for the factors of the form $e^{-k\lambda}$ for some $k$, $p_Y(0|0,0), p_Y(0|0,0,0)$ and so forth and $\prod_{i=3}^{p}\overline{\alpha_i}^6$. By (7.4), $p_Y(0|0,0), p_Y(0|0,0,0)$ etc. are all larger than zero, and $\overline{\alpha_i}, e^{-\lambda} > 0$ by definition. Thus, time-reversibility of the process and $\alpha_1 > 0$ implies $\alpha_2 = 0$.

Now, let $2 < k \leq p$, and let $\alpha_i = 0$ for $1 < i < k$, it is shown that this implies $\alpha_k = 0$. Consider the sequence of states

$$\overline{\underbrace{0,0,\ldots,0}_{p \text{ times}},1,3,\underbrace{0,\ldots,0}_{k-2 \text{ times}},2,\underbrace{0,\ldots,0,0}_{p \text{ times}}}$$

and its inverse. For a time-reversible process, the transition probabilities for this sequence has to equal the transition probability of its inverse. With (7.4) and recalling that $\overline{\alpha_i}, e^{-\lambda}, p_Y(0|0,0), p_Y(0|0,0,0)$ etc. are all larger than zero, this relation is equivalent to

$$\begin{aligned}&\overline{\alpha_1}^3\overline{\alpha_k}^4\left(\lambda\overline{\alpha_1}^3\overline{\alpha_k}^2 + 3\alpha_1\overline{\alpha_1}^2\overline{\alpha_k}^2 + 2\overline{\alpha_1}^3\alpha_k\overline{\alpha_k}\right)\frac{\lambda^5}{12}\\ =&\overline{\alpha_1}^5\overline{\alpha_k}^5\left(\frac{\lambda^2}{2}\overline{\alpha_k} + \lambda\alpha_k\right)\left(\frac{\lambda^3}{6}\overline{\alpha_1} + \frac{\lambda^2}{2}\alpha_1\right)\lambda.\end{aligned}$$

This relation can be simplified to yield $0 = \frac{1}{2}\alpha_1\alpha_k$. The assertion thus follows by induction over $k$. □

It may be pointed out that the result of Theorem 7.2.10 is only partial in the sense that $\alpha_1 > 0$ was assumed. In the author's opinion, this is a quite natural assumption for DLINAR(p) processes, and the investigation is stopped at this point. However, given the previous result, the author conjectures the following assertion to be true:

**Conjecture 1** (Schweer (2015b), Conjecture 19.1)**.** *Let $(Y_t)_{t\in\mathbb{Z}}$ be a DLINAR(p) process with $p > 1$ and $\alpha_j > 0$ for some $j \in \{1,\ldots,p\}$ and let $0 < \mathbb{P}(\epsilon = 0) < 1$. Then $(Y_t)_{t\in\mathbb{Z}}$ is time-reversible if and only if $\alpha_l = 0$ for $l \in \{1,\ldots,p\}, l \neq j$.*

To illustrate why the conjecture contains the reverse implication as well, consider the result of Lemma 7.2.9 for the case $p = 2$. It shows that for time-reversibility of the process, either $\alpha_1 = 0$ or $\alpha_2 = 0$ has to hold. The case $\alpha_1 = 0$ is clearly a degenerate case, and it can indeed be shown quite easily that it has the same stationary distribution as that of the corresponding INAR(1) process (i.e., with $\alpha = \alpha_2$) and that it is time-reversible. Similar arguments apply to higher order autoregressive structures.

# List of Figures

# List of Tables

# List of Acronyms

| | |
|---|---|
| AAINAR(p) | integer-valued autoregressive process of $p$-th order in the formulation of Al-Osh and Alzaid. 130 |
| ACF | autocorrelation function. 2 |
| AR(1) | (continuous) autoregressive process of first order. 3 |
| cdf | cumulative distribution function. 26 |
| cgf | cumulant generating function. 17 |
| CLT | central limit theorem. 19 |
| CPINAR(1) | Compound Poisson integer-valued autoregressive process of first order. 64 |
| DLINAR(p) | integer-valued autoregressive process of $p$-th order in the formulation of Du and Li. 130 |
| DSD | discrete self-decomposable. 16 |
| ejpgf | empirical joint probability generating function. 103 |
| epgf | empirical probability generating function. 110 |
| i.i.d. | independent and identically distributed. 3 |
| INAR(1) | integer-valued autoregressive process of first order. 3 |
| INCLAR(1) | integer-valued conditional linear autoregressive process of first order. 104 |
| jpgf | joint probability generating function. 103 |
| mgf | moment generating function. 17 |
| PACF | partial autocorrelation function. 2 |
| pgf | probability generating function. 15 |
| pmf | probability mass function. 61 |

# List of Symbols

| | |
|---|---|
| $A^-$ | Closure of a set $A$. 11 |
| $\alpha_Y(n)$ | Strong mixing coefficient of a stationary process $(Y_t)_{t\in\mathbb{Z}}$. 20 |
| $\boldsymbol{\alpha}$ | Parameter of autoregression parameters in AAINAR(p) and DLINAR(p) processes, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$. 133 |
| $c_0$ | Space of sequences of real numbers $x = (x_k)_{k\in\mathbb{N}}$ converging to zero. 10 |
| $\mathrm{ComPoi}_\nu(\lambda, H)$ | Compound Poisson distribution with compounding pgf $H$ with $\deg(H(z)) = \nu$ and parameter $\lambda$. 15 |
| $\mathrm{cum}(X_1, \ldots, X_r)$ | $r$-th order joint cumulant of the random variables $X_1, \ldots, X_r$. 13 |
| $\mathcal{F}_k(Y)$ | $\sigma$-algebra $\sigma\left(Y(i)\ ; -\infty < i \leq k\right)$ of the past behavior of a process $(Y(t))_{t\in\mathbb{Z}}$. 20 |
| $\gamma(\cdot)$ | Autocovariance function of a stationary process $(Y_t)_{t\in\mathbb{Z}}$. 22 |
| $I_X$ | Index of Dispersion of a random variable $X$. 15 |
| $\kappa_r(X)$ | $r$-th order cumulant of a random variable $X$. 13 |
| $\overline{X}$ | Empirical mean of an observation of a random variable $X$. 22 |
| $\bar{\mu}_{X,k}$ | $k$-th central moment of a random variable $X$. 17 |
| $\mu(s_1, \ldots, s_{r-1})$ | Joint moment of a stationary process $(Y_t)_{t\in\mathbb{Z}}$. 54 |
| $\mu_{X,k}$ | $k$-th moment about the origin of a random variable $X$. 17 |
| $\lVert\cdot\rVert_{c_0}$ | Norm on the space $c_0$, given by $\lVert x\rVert_{c_0} = \sup_{k\in\mathbb{N}} \lvert x_k\rvert$. 10 |
| $\lVert\cdot\rVert_{L^2}$ | $L^2$-norm on a measure space, given by $\lVert X\rVert_{L^2} = (\mathbb{E}[\lvert X\rvert^2])^{\frac{1}{2}}$. 21 |

$\|\cdot\|$ — Euclidean norm on $\mathbb{R}^d$ for $d \in \mathbb{N}$. 109

$\|\cdot\|_\infty$ — Uniform norm on $C[0,1]^2$, given by $\|g\|_\infty = \sup_{0\leq u,v\leq 1} |g(u,v)|$. 106

$p_Y(l|k)$ — Transition probabilities of a process $(Y_t)_{t\in\mathbb{Z}}$, also written as $p_\theta(k,l)$ to emphasize dependency on underlying parameter $\theta \in \mathbb{R}^d$. 19, 61, 102

$\pi_Y(k)$ — Stationary distribution of a process $(Y_t)_{t\in\mathbb{Z}}$, also written as $\pi_\theta(k)$ to emphasize dependency on underlying parameter $\theta \in \mathbb{R}^d$. 61, 102

$\mathrm{pgf}_X(\cdot)$ — Probability generating function of a random variable $X$, $\mathrm{pgf}_X = \mathbb{E}[z^X]$. 15

$\mathrm{Poi}(\lambda)$ — Poisson distribution with parameter $\lambda$. 15

$\mathrm{Poi}_\nu(\lambda)$ — Poisson distribution of order $\nu$ with parameter $\lambda$. 16

$\rho_{part}(\cdot)$ — Partial autocorrelation function of a stationary process $(Y_t)_{t\in\mathbb{Z}}$. 22

$\rho(\cdot)$ — Autocorrelation function of a stationary process $(Y_t)_{t\in\mathbb{Z}}$. 22

$S_X^2$ — Empirical variance of an observation of a random variable $X$. 75

$\alpha \circ X$ — Binomial thinning of a random variable $X$ with parameter $\alpha$. 3

# Index

# Bibliography


Aki, S. (1985). Discrete distributions of order $k$ on a binary sequence. *Annals of the Institute of Statistical Mathematics*, 37(1):205–224.

Aki, S., Kuboki, H., and Hirano, K. (1984). On discrete distributions of order $k$. *Annals of the Institute of Statistical Mathematics*, 36(1):431–440.

Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive INAR(1) processes. *Journal of Time Series Analysis*, 8(3):261–275.

Aliprantis, C. and Border, K. (2006). *Infinite Dimensional Analysis.* Springer, Berlin, 3rd edition.

Alzaid, A. and Al-Osh, M. (1990). An integer-valued $p$th-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, 27(2):314–324.

Alzaid, A. A. and Al-Osh, M. A. (1988). First-order integer-valued autoregressive (INAR(1)) process: Distributional and regression properties. *Statistica Neerlandica*, 42(1):53–61.

Baringhaus, L., Gürtler, N., and Henze, N. (2000). Theory & Methods: Weighted integral test statistics and components of smooth tests of fit. *Australian & New Zealand Journal of Statistics*, 42(2):179–192.

Baringhaus, L. and Henze, N. (1992). A goodness of fit test for the Poisson distribution based on the empirical generating function. *Statistics & Probability Letters*, 13(4):269–274.

Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Mathematical Statistics*, 42(5):1656–1670.

Billingsley, P. (1968). *Convergence of probability measures.* John Wiley & Sons Inc., New York, NY, 1st edition.

Billingsley, P. (1999). *Convergence of probability measures.* John Wiley & Sons Inc., New York, NY, 2nd edition.

Bingham, N., Goldie, C., and Teugels, J. (1989). *Regular Variation.* Cambridge University Press, Cambridge, 1st edition.

Bingham, N. and Pitts, S. (1999). Non-parametric estimation for the M/G/? queue. *Annals of the Institute of Statistical Mathematics*, 51:71–97.

Blanghaps, N., Nov, Y., and Weiss, G. (2013). Sojourn time estimation in an $M/G/\infty$ queue with partial information. *Journal of Applied Probability*, 50(4):1044–1056.

Böhning, D. (1994). A note on a test for Poisson overdispersion. *Biometrika*, 81(2):418–419.

Bohrnstedt, G. W. and Goldberger, A. S. (1969). On the exact covariance of products of random variables. *Journal of the American Statistical Association*, 64(328):1439–1442.

Bradley, R. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2:107–144.

Brillinger, D. (1969). The calculation of cumulants via conditioning. *Annals of the Institute of Statistical Mathematics*, 21(1):215–218.

Brillinger, D. (1974). Cross-spectral analysis of processes with stationary increments including the stationary $G/G/\infty$ queue. *The Annals of Probability*, 2(5):815–827.

Brillinger, D. (1981). *Time series: Data analysis and theory.* McGraw-Hill, New York, NY, expanded edition.

Brown, M. (1970). An M/G/$\infty$ estimation problem. *The Annals of Mathematical Statistics*, 41(2):651–654.

Bu, R. and McCabe, B. (2008). Model selection, estimation and forecasting in INAR(p) models: A likelihood-based markov chain approach. *International Journal of Forecasting*, 24(1):151–162.

Choi, B. (1990). A derivation of the asymptotic distribution of the partial autocorrelation function of an autoregressive process. *Communications in Statistics - Theory and Methods*, 19(2):547–553.

David, H. (1985). Bias of $S^2$ under dependence. *The American Statistician*, 39(3):201.

Davydov, Y. A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory of Probability & Its Applications*, 13(4):691–696.

Dion, J.-P., Gauthier, G., and Latour, A. (1995). Branching processes with immigration and integer-valued time series. *Serdica Mathematical Journal*, 21(2):123–136.

Douglas, J. (1980). *Analysis with standard contagious distributions.* International Co-operative Publishing House, Fairland, MD.

Du, J.-G. and Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, 12(2):129–142.

Edelmann, D. and Wichelhaus, C. (2014). Nonparametric inference for queueing networks of $\mathrm{GEOM}^X/\mathrm{G}/\infty$ queues in discrete time. *Advances in Applied Probability*, 46(3):790–811.

Ekström, M. (2014). A general central limit theorem for strong mixing sequences. *Statistics & Probability Letters*, 94:236–238.

Erlang, A. (1909). The theory of probabilities and telephone conversations. *Nyt Tidsskrift for Matematik B*, 20(16):33–39.

Feller, W. (1968). *An introduction to probability theory and its applications*. John Wiley & Sons, Inc., New York, NY, 3rd edition.

Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, 27(6):923–942.

Fokianos, K. and Neumann, M. H. (2013). A goodness-of-fit test for Poisson count processes. *Electronic Journal of Statistics*, 7:793–819.

Freeland, R. and McCabe, B. (2004). Forecasting discrete valued low count time series. *International Journal of Forecasting*, 20(3):427–434.

Freeland, R. K. (1998). *Statistical analysis of discrete time series with application to the analysis of workers' compensation claims data*. PhD thesis, University of British Columbia, Canada.

Freeland, R. K. (2010). True integer value time series. *AStA Advances in Statistical Analysis*, 94(3):217–229.

Grunwald, G. K., Hyndman, R. J., Tedesco, L., and Tweedie, R. L. (2000). Theory & Methods: Non-gaussian conditional linear AR(1) models. *Australian & New Zealand Journal of Statistics*, 42(4):479–495.

Hall, P. and Park, J. (2004). Nonparametric inference about service time distribution from indirect measurements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(4):861–875.

Hanche-Olsen, H. and Holden, H. (2010). The Kolmogorov-Riesz compactness theorem. *Expositiones Mathematicae*, 28(4):385–394.

Heathcote, C. (1965). A branching process allowing immigration. *Journal of the Royal Statistical Society. Series B (Methodological)*, 27(1):138–143.

Heathcote, C. (1966). Corrections and comments on the paper "A branching process allowing immigration". *Journal of the Royal Statistical Society. Series B (Methodological)*, 28(1):213–217.

Henze, N. (1996). Empirical-distribution-function goodness-of-fit tests for discrete models. *Canadian Journal of Statistics*, 24(1):81–93.

Hudecová, Š., Hušková, M., and Meintanis, S. G. (2015). Tests for time series of counts based on the probability-generating function. *Statistics*, 49(2):316–337.

Ibragimov, I. (1962). Some limit theorems for stationary processes. *Theory of Probability & Its Applications*, 7(4):349–382.

Johnson, N., Kotz, S., and Kemp, A. (2005). *Univariate Discrete Distributions.* John Wiley & Sons, Inc., Hoboken, NJ, 3rd edition.

Jung, R. C., Ronning, G., and Tremayne, A. R. (2005). Estimation in conditional first order autoregression with discrete support. *Statistical Papers*, 46(2):195–224.

Kachurovskii, A. (1996). The rate of convergence in ergodic theorems. *Russian Mathematical Surveys*, 51(4):653–703.

Kelly, F. (1979). *Reversibility and stochastic networks.* John Wiley & Sons, Inc., Chichester, 1st edition.

Kemp, C. (1967). "Stuttering-poisson" distributions. *Journal of the Statistical and Social Inquiry Society of Ireland*, 21(5):151–157.

Ku, S. and Seneta, E. (1996). Quenouille-type theorem on autocorrelations. *Annals of the Institute of Statistical Mathematics*, 48(4):621–630.

Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241.

Lachout, P. (1988). Billingsley-type tightness criteria for multiparameter stochastic processes. *Kybernetika*, 24(5):363–371.

Lahiri, S. (2003). *Resampling methods for dependent data.* Springer, New York, NY, 1st edition.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, 21(4):645–650.

McKenzie, E. (1988). Some ARMA models for dependent sequences of poisson counts. *Advances in Applied Probability*, 20(4):822–835.

Meintanis, S. G. and Karlis, D. (2014). Validation tests for the innovation distribution in INAR time series models. *Computational Statistics*, 29(5):1221–1241.

Mills, T. and Seneta, E. (1991). Independence of partial autocorrelations for a classical immigration branching process. *Stochastic Processes and their Applications*, 37(2):275–279.

Mode, C. J. (1971). *Multitype branching processes.* American Elsevier, New York, NY, 1st edition.

Neumann, M. H. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli*, 17(4):1268–1284.

Nummelin, E. and Tweedie, R. (1978). Geometric ergodicity and R-positivity for general Markov chains. *The Annals of Probability*, 6(3):404–420.

Pakes, A. (1971). Branching processes with immigration. *Journal of Applied Probability*, 8(1):32–42.

Pedeli, X. and Karlis, D. (2011). A bivariate INAR(1) process with application. *Statistical modelling*, 11(4):325–349.

Pickands, J. and Stine, R. (1997). Estimation for an M/G/$\infty$ queue with incomplete information. *Biometrika*, 84(2):295–308.

Popov, N. (1977). Conditions for geometric ergodicity of countable Markov chains. *Soviet Mathematics Doklady*, 18(3):676–679.

Quenouille, M. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(1):68–84.

Radulović, D. (2012). Necessary and sufficient conditions for the moving blocks bootstrap central limit theorem of the mean. *Journal of Nonparametric Statistics*, 24(2):343–357.

Rajarshi, M. (1990). Bootstrap in Markov-sequences based on estimates of transition density. *Annals of the Institute of Statistical Mathematics*, 42(2):253–268.

Ramsey, J. B. and Rothman, P. (1996). Time irreversibility and business cycle asymmetry. *Journal of Money, Credit and Banking*, 28(1):1–21.

Rao, C. and Chakravarti, I. (1956). Some small sample tests of significance for a poisson distribution. *Biometrics*, 12(3):264–282.

Romano, J. and Thombs, L. (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association*, 91(434):590–600.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1):43–47.

Schweer, S. (2015a). A goodness-of-fit test for integer-valued autoregressive processes. *Journal of Time Series Analysis*. To appear.

Schweer, S. (2015b). On the time-reversibility of integer-valued autoregressive processes of general order. In Steland, A., Rafajłowicz, E., and Szajowski, K., editors, " Stochastic Models, Statistics and Their Applications", *Springer Proceedings in Mathematics & Statistics*, (122):169–177. Springer, Heidelberg.

Schweer, S. and Weiß, C. H. (2014). Compound Poisson INAR (1) processes: Stochastic properties and testing for overdispersion. *Computational Statistics and Data Analysis*, 77(9):267–284.

Schweer, S. and Weiß, C. H. (2015). Testing for Poisson arrivals in INAR(1) processes via time-reversibility. Submitted.

Schweer, S. and Wichelhaus, C. (2015a). Nonparametric estimation of the service time distribution in the discrete-time GI/G/$\infty$ queue with partial information. *Stochastic Processes and their Applications*, 125(1):233–253.

Schweer, S. and Wichelhaus, C. (2015b). Queueing systems of INAR(1) processes with Compound Poisson arrival distributions. Submitted.

Silva, I. and Silva, M. (2006). Asymptotic distribution of the Yule-Walker estimator for INAR processes. *Statistics & Probability Letters*, 76(15):1655–1663.

Steutel, F. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability*, 7(5):893–899.

van der Vaart, A. (2000). *Asymptotic statistics.* Cambridge University Press, Cambridge, $1^{st}$ edition.

Walrand, J. (1983). A discrete-time queueing network. *Journal of Applied Probability*, 20(4):903–909.

Weiß, C. H. (2008). Thinning operations for modeling time series of counts - a survey. *AStA Advances in Statistical Analysis*, 92(3):319–341.

Weiß, C. H. (2009). Modelling time series of counts with overdispersion. *Statistical Methods and Applications*, 18(4):507–519.

Weiß, C. H. (2010). INARCH(1) processes: Higher-order moments and jumps. *Statistics & Probability Letters*, 80(23-24):1771–1780.

Weiß, C. H. (2012). Process capability analysis for serially dependent processes of Poisson counts. *Journal of Statistical Computation and Simulation*, 82(3):383–404.

Weiß, C. H. (2013). Integer-valued autoregressive models for counts showing underdispersion. *Journal of Applied Statistics*, 40(9):1931–1948.

Weiß, C. H. and Schweer, S. (2015). Detecting overdispersion in INARCH(1) processes. *Statistica Neerlandica.* To appear.

Weiss, G. (1975). Time-reversibility of linear stochastic processes. *J. Appl. Prob*, 12(831-836):264.

Wichelhaus, C. and Langrock, R. (2012). Nonparametric inference for stochastic feedforward networks based on cross-spectral analysis of point processes. *Electronic Journal of Statistics*, 6:1670–1714.

Widder, D. V. (1946). *The Laplace transform.* Princeton University Press, Princeton, $2^{nd}$ edition.

Xu, H.-Y., Xie, M., Goh, T. N., and Fu, X. (2012). A model for integer-valued time series with conditional overdispersion. *Computational Statistics & Data Analysis*, 56(12):4229–4242.

Zeger, S. and Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, 44(4):1019–1031.

Zheng, H., Basawa, I. V., and Datta, S. (2007). First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference*, 137(1):212–229.

Zhu, R. and Joe, H. (2003). A new type of discrete self-decomposability and its application to continuous-time Markov processes for modeling count data time series. *Stochastic Models*, 19(2):235–254.