CrossMark

Goodness-of-fit testing of a count time series' marginal distribution

Christian H. Weiß[1]

Abstract

Popular goodness-of-fit tests like the famous Pearson test compare the estimated probability mass function with the corresponding hypothetical one. If the resulting divergence value is too large, then the null hypothesis is rejected. If applied to i. i. d. data, the required critical values can be computed according to well-known asymptotic approximations, e. g., according to an appropriate χ^2-distribution in case of the Pearson statistic. In this article, an approach is presented of how to derive an asymptotic approximation if being concerned with time series of autocorrelated counts. Solutions are presented for the case of a fully specified null model as well as for the case where parameters have to be estimated. The proposed approaches are exemplified for (among others) different types of CLAR(1) models, INAR(p) models, discrete ARMA models and Hidden-Markov models.

1 Introduction

In many applications, one is interested in testing a hypothesis with respect to the marginal distribution of a given count time series $X_1, \ldots, X_T$. This might be done by looking at a specific feature of the hypothetical count distribution, e. g., at the Poisson index of dispersion as in Schweer and Weiß (2014), or by deriving a test statistic which considers any kind of deviation from the null model. In Meintanis and Karlis (2014), such tests are developed based on probability generating functions (pgf), while here, we follow the textbook approach and consider goodness-of-fit statistics based on the

✉ Christian H. Weiß
 weissc@hsu-hh.de

1   Department of Mathematics and Statistics, Helmut Schmidt University, 22008 Hamburg, Germany

hypothetical and estimated probability mass function (pmf). As an important example within the power-divergence family (Cressie and Read 1984; Read and Cressie 1988), we shall concentrate on Pearson's goodness-of-fit statistic, but the presented approach could be adapted to other pmf-based statistics as well.

In Sect. 2, an approach is presented of how to explicitly compute the asymptotic distribution of Pearson's goodness-of-fit statistic (so no bootstrap implementation is required). This approach covers both scenarios, where the model parameters are either specified, or where they have to be estimated from the available data. The approach can be applied if the process satisfies certain mixing and moment conditions, and if the h-step-ahead conditional distributions (and corresponding moments) can be computed. A number of examples are presented, including types of CLAR(1) models, INAR(p) models, discrete ARMA models and Hidden-Markov models, also see the summaries in Appendix A. The goodness of the resulting asymptotic approximations to the actual Pearson statistic's distribution is investigated in Sect. 3, where results from a simulation study are presented. There, we also analyze the size and power of the test if applied in practice, and we briefly discuss two real-data examples. Finally, we conclude in Sect. 4.

2 Goodness-of-fit testing

Many common goodness-of-fit test statistics fall within the power-divergence family as discussed by Cressie and Read (1984) and Read and Cressie (1988). Since the asymptotic behavior of these statistics is the same as that of the famous Pearson's goodness-of-fit statistic G^2 to be defined below (Cressie and Read 1984; Read and Cressie 1988), we shall focus on the latter in the sequel.

If the given data $X_1, \ldots, X_T$ are independent and identically distributed (i.i.d.), and if the Pearson statistic G^2 is constructed by using k categories, it is known that its asymptotic distribution is a χ^2-distribution (Cressie and Read 1984; Read and Cressie 1988) [also note the results by Kißlinger and Stummer (2016) about statistics within the family of scaled Bregman divergences having such a limiting χ^2-distribution]. More precisely, if the hypothetical distribution is fully specified, then G^2 converges to the χ^2_{k-1}-distribution with $k-1$ degrees of freedom. If, in contrast, the hypothetical distribution has $r \geq 1$ unspecified parameters, which have to be estimated from the same data, the degrees of freedom have to be further reduced by r, i.e., the distribution of G^2 is approximated by the χ^2_{k-1-r}-distribution.

In the sequel, we shall consider both scenarios, i.e., a fully specified hypothetical distribution, or a distribution with estimated parameters, but in the case of *serially dependent* time series data $X_1, \ldots, X_T$. Such a scenario (although referring to continuous-valued data) was also considered by Moore (1982), where it was shown that the asymptotic distribution of the Pearson statistic (under rather general conditions) can be *formally* derived; however, the involved matrices appeared "intractible" such that explicit computations were presented only for a Gaussian AR(1) process (first-order autoregressive) with fully specified null distribution. In this work, an approach is presented for *explicitly* computing the limit distribution for a variety of count time

series models and under parameter estimation, such that the Pearson test can be applied in all these cases without the need for a bootstrap implementation.

2.1 Pearson's goodness-of-fit test

Let $(X_t)_{\mathbb{Z}}$ be a stationary count process. With a goodness-of-fit test based on $X_1, \ldots, X_T$, we want to test the hypothesis that the marginal distribution satisfies $P(X_t = i) = p_i$ for some pmf $(p_i)_i$. First, we define an appropriate set of categories, e. g., following one of the rules of thumb surveyed in Horn (1977). Like in Kim and Weiß (2015), we shall assume $b - a + 2$ categories of the form

$$\{0, \ldots, a\}, \{a+1\}, \ldots, \{b\}, \{b+1, \ldots\} \quad \text{with some } 0 \le a < b. \tag{1}$$

Then the hypothetical probabilities for X_t falling into one of these categories are computed from $\boldsymbol{p} = (p_0, \ldots, p_b)^\top$ as

$$\boldsymbol{\pi} := \begin{pmatrix} \pi_a \\ \pi_{a+1} \\ \vdots \\ \pi_b \\ \pi_{b+1} \end{pmatrix} := \begin{pmatrix} p_0 + \cdots + p_a \\ p_{a+1} \\ \vdots \\ p_b \\ 1 - p_0 - \ldots - p_b \end{pmatrix} = \mathbf{A}\,\boldsymbol{p} + \boldsymbol{e}_{b-a+2}$$

with $\boldsymbol{e}_{b-a+2} := (0, \ldots, 0, 1)^\top, \quad$ and $\tag{2}$

$$\mathbf{A} := \mathbf{A}(a, b) := \underbrace{\begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & & \vdots \\ \vdots & & \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 0 & & 1 \\ -1 & \cdots & -1 & -1 & \cdots & -1 \end{pmatrix}}_{\substack{a+1 \qquad b-a}} \left.\vphantom{\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix}}\right\} b - a + 2.$$

Defining $\hat{p}_i$ as the relative frequency of i within $X_1, \ldots, X_T$, i.e., $\hat{p}_i = \frac{1}{T} \sum_{t=1}^T \mathbb{1}_{\{X_t=i\}}$ with $\mathbb{1}$ denoting the indicator function, and setting $\hat{\boldsymbol{\pi}} := \mathbf{A}\,\hat{\boldsymbol{p}} + \boldsymbol{e}_{b-a+2}$ with $\hat{\boldsymbol{p}} = (\hat{p}_0, \ldots, \hat{p}_b)^\top$, Pearson's goodness-of-fit statistic is computed as

$$G^2 = T\,(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})^\top \operatorname{diag}(\boldsymbol{\pi})^{-1}\,(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}). \tag{3}$$

This can be rewritten

$$G^2 = T\,\boldsymbol{G}^\top \boldsymbol{G} \quad \text{with } \boldsymbol{G} := \operatorname{diag}\!\big(\pi_a^{-1/2}, \ldots, \pi_{b+1}^{-1/2}\big)\,\mathbf{A}(\hat{\boldsymbol{p}} - \boldsymbol{p}). \tag{4}$$

From now on, let us assume that the marginal pmf is determined by some parameter $\boldsymbol{\theta} \in \mathbb{R}^r$, i.e., $\boldsymbol{p} = \boldsymbol{p}(\boldsymbol{\theta})$ and $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}) = \mathbf{A}\,\boldsymbol{p}(\boldsymbol{\theta}) + \boldsymbol{e}_{b-a+2}$. We shall distinguish the

two scenarios, where either $\boldsymbol{\theta}$ is specified (so fully specified pmf), or $\boldsymbol{\theta}$ has to be estimated from $X_1, \ldots, X_T$ (goodness-of-fit with estimated parameters). In the latter case, we shall use simple moment estimators, i. e., we assume that $\hat{\boldsymbol{\theta}}$ is expressed as a differentiable function $\boldsymbol{h} : \mathbb{R}^r \to \mathbb{R}^r$ applied to the empirical marginal raw moments $\hat{\mu}_1 = \bar{X} = \frac{1}{T} \sum_{t=1}^T X_t, \ldots, \hat{\mu}_r = \frac{1}{T} \sum_{t=1}^T X_t^r$. The ultimate aim is to derive the asymptotic distributions of $\sqrt{T} \, \boldsymbol{G}(\hat{\boldsymbol{p}}, \boldsymbol{\theta})$ and $\sqrt{T} \, \boldsymbol{G}(\hat{\boldsymbol{p}}, \hat{\boldsymbol{\theta}})$, respectively, as well as of $G^2(\hat{\boldsymbol{p}}, \boldsymbol{\theta})$ and $G^2(\hat{\boldsymbol{p}}, \hat{\boldsymbol{\theta}})$, respectively.

2.2 A central limit theorem

To derive the asymptotic distributions of $\sqrt{T} \, \boldsymbol{G}(\hat{\boldsymbol{p}}, \boldsymbol{\theta})$ and $\sqrt{T} \, \boldsymbol{G}(\hat{\boldsymbol{p}}, \hat{\boldsymbol{\theta}})$, respectively, we start with a central limit theorem. For this purpose, we assume that $(X_t)_{\mathbb{Z}}$ satisfies appropriate mixing and moment conditions: in the examples considered below, it is sufficient to assume, e. g., that $(X_t)_{\mathbb{Z}}$ is α-mixing with geometrically decreasing weights, and that the $(2r + \delta)$-moments with some $\delta > 0$ exist. Then we can apply Theorem 1.7 in Ibragimov (1962) to obtain the required limit distribution.

Special attention is given to the case, where the lagged bivariate probabilities $P(X_t = i, X_{t-h} = j)$ are symmetric in i, j (property "SYM"), since then more simple closed-form expressions can be obtained. Note that SYM is implied by time reversibility. But our approach also works if the optional property SYM is violated, computations are just more complex in that case.

In view of (4), we define the $(b + 1 + r)$-dimensional process

$$
\boldsymbol{Z}_t = \begin{pmatrix} \mathbb{1}_{\{X_t=0\}} \\ \vdots \\ \mathbb{1}_{\{X_t=b\}} \\ X_t \\ \vdots \\ X_t^r \end{pmatrix} \quad \text{with} \quad \boldsymbol{\mu_Z} := E[\boldsymbol{Z}_t] = \begin{pmatrix} p_0 \\ \vdots \\ p_b \\ \mu_1 \\ \vdots \\ \mu_r \end{pmatrix}, \tag{5}
$$

where $\mu_k := E[X_t^k]$ with $\mu := \mu_1$ (marginal raw moments). The idea is to choose r sufficiently large such that all parameters in $\boldsymbol{\theta}$ can be estimated by the method-of-moments. Note that the mixing properties of $(X_t)_{\mathbb{Z}}$ carry over to $(\boldsymbol{Z}_t)_{\mathbb{Z}}$ such that Theorem 1.7 in Ibragimov (1962) is applicable.

Theorem 2.2.1 *Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary count process, which is α-mixing with geometrically decreasing weights and has existing $(2r + \delta)$-moments with some $\delta > 0$, and let $\boldsymbol{Z}_t$ be given by* (5). *Then*

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^T (\boldsymbol{Z}_t - \boldsymbol{\mu_Z}) \xrightarrow{D} N(\boldsymbol{0}, \boldsymbol{\Sigma}),
$$

where

$$\sigma_{ij} = E[Z_{0,i} \cdot Z_{0,j}] - \mu_{Z,i}\,\mu_{Z,j}$$
$$+ \sum_{h=1}^{\infty} \left(E[Z_{0,i} \cdot Z_{h,j}] + E[Z_{h,i} \cdot Z_{0,j}] - 2\,\mu_{Z,i}\,\mu_{Z,j} \right)$$
$$\overset{SYM}{=} E[Z_{0,i} \cdot Z_{0,j}] - \mu_{Z,i}\,\mu_{Z,j} + 2 \sum_{h=1}^{\infty} \left(E[Z_{h,i} \cdot Z_{0,j}] - \mu_{Z,i}\,\mu_{Z,j} \right).$$

Note that $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{Z}_t = \sqrt{T}\,(\hat{p}_0, \ldots, \hat{p}_b,\ \hat{\mu}_1, \ldots, \hat{\mu}_r)^{\top}$. Assuming some differentiable function $\boldsymbol{h} : \mathbb{R}^r \to \mathbb{R}^r$ such that $\boldsymbol{\theta} = \boldsymbol{h}(\mu_1, \ldots, \mu_r)$, the moment estimator of $\boldsymbol{\theta}$ follows as $\hat{\boldsymbol{\theta}} = \boldsymbol{h}(\hat{\mu}_1, \ldots, \hat{\mu}_r)$. Applying the Delta method, Theorem 2.2.1 leads to the asymptotic result

$$\sqrt{T}\left((\hat{\boldsymbol{p}}, \hat{\boldsymbol{\theta}}) - (\boldsymbol{p}, \boldsymbol{\theta})\right) \overset{D}{\to} \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Sigma}^*\right) \quad \text{with } \boldsymbol{\Sigma}^* := \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^{\top}, \tag{6}$$

with $\mathbf{D}$ denoting the Jacobian of $\left(z_0, \ldots, z_b, \boldsymbol{h}(z_{b+1}, \ldots, z_{b+r})\right)^{\top}$ evaluated at $(p_0, \ldots, \mu_r)^{\top}$. If, in contrast, $\boldsymbol{\theta}$ is specified, then it suffices to consider the asymptotic result

$$\sqrt{T}\,(\hat{\boldsymbol{p}} - \boldsymbol{p}) \overset{D}{\to} \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Sigma}^{(p)}\right) \quad \text{with } \boldsymbol{\Sigma}^{(p)} := (\sigma_{ij})_{i,j=0,\ldots,b}. \tag{7}$$

So formally, the asymptotic distributions (6) and (7) are easily derived, also see the analogous results by Moore (1982); in Sect. 2.4, we shall pick up again these asymptotic distributions for $\sqrt{T}\,(\hat{\boldsymbol{p}} - \boldsymbol{p})$ and $\sqrt{T}\left((\hat{\boldsymbol{p}}, \hat{\boldsymbol{\theta}}) - (\boldsymbol{p}, \boldsymbol{\theta})\right)$, respectively, and derive the resulting asymptotic distribution of the Pearson statistic with specified or estimated parameters. But before continuing in this direction, it is crucial to ask if the covariances occurring in Theorem 2.2.1 can be *computed* at all in practice, since otherwise we could not benefit from this result in applications. Therefore, in Sect. 2.3, it is shown that for many different types of count processes, explicit results are easily obtained.

2.3 Intermezzo: application and implementation of Theorem 2.2.1

The required moments $E[Z_{h,i} \cdot Z_{0,j}]$ with $h \geq 0$ in Theorem 2.2.1 are easy to compute in many practically relevant cases. As outlined in the sequel, the minimal requirement for computability is that the h-step-ahead conditional probabilities $p_{i|j}{}^{(h)} := P(X_t = i \mid X_{t-h} = j)$ are available.

For $i, j \in \{0, \ldots, b\}$, we have

$$E[Z_{h,i} \cdot Z_{0,j}] - \mu_{Z,i}\mu_{Z,j} = P(X_t = i, X_{t-h} = j) - p_i\,p_j$$
$$= \begin{cases} (\delta_{i,j} - p_i)\,p_j & \text{if } h = 0, \\ (p_{i|j}^{(h)} - p_i)\,p_j & \text{if } h > 0, \end{cases} \tag{8}$$

where $\delta_{i,j}$ denotes the Kronecker delta. So if being able to compute the $p_{i|j}{}^{(h)}$, then all σ_{ij} with $i, j = 0, \ldots, b$ [and hence $\boldsymbol{\Sigma}^{(p)}$ from (7)] are available.

Next, for $i, j \in \{1, \ldots, r\}$, we have $\mu_{Z,b+i} = \mu_i$ as well $\mu_{Z,b+j} = \mu_j$, and we obtain

$$E[Z_{h,b+i} \cdot Z_{0,b+j}] = E[X_t^i \cdot X_{t-h}^j] = \begin{cases} \mu_{i+j} & \text{if } h = 0, \\ E[X_t^i \cdot X_{t-h}^j] & \text{if } h > 0. \end{cases} \quad (9)$$

So joint moments of sufficiently high order need to be computed. For $r = 1$, it is indeed sufficient to compute the autocovariance function. If closed-form moment expressions are not available, a numerical computation based on the h-step-ahead bivariate distributions is possible, e.g., by truncating the summation after $M + 1$ summands with M sufficiently large.

Finally, for the remaining moments, the optional property SYM would be particularly useful, because then

$$E[Z_{h,i} \cdot Z_{0,b+j}] \overset{\text{SYM}}{=} E[Z_{h,b+j} \cdot Z_{0,i}]$$

holds for $i = 0, \ldots, b$ and $j = 1, \ldots, r$. For the latter moments, we obtain

$$E[Z_{h,b+j} \cdot Z_{0,i}] = E[X_t^j \mathbb{1}_{\{X_{t-h}=i\}}] = \begin{cases} i^j \, p_i & \text{if } h = 0, \\ E[X_t^j \mid X_{t-h} = i] \, p_i & \text{if } h > 0. \end{cases} \quad (10)$$

Note that such conditional moments (10) are often easy to compute in practice, as illustrated by the subsequent examples. If the "nice-to-have" property SYM does not hold, then $E[Z_{h,i} \cdot Z_{0,b+j}] = \sum_{x=0}^{\infty} x^j \, p_{i|x}^{(h)} \, p_x$ has to be calculated (numerically) from the h-step-ahead bivariate distribution.

Example 2.3.1 (i.i.d. Counts) Let $(X_t)_{\mathbb{Z}}$ be i.i.d. Then $p_{i|j}^{(h)} - p_i = 0$ for $h > 0$, so (8) simplifies and we obtain the well-known result that $\sigma_{ij} = (\delta_{i,j} - p_i) \, p_j$ for $i, j = 0, \ldots, b$. Similarly, (9) simplifies because of $E[X_t^i \cdot X_{t-h}^j] = \mu_i \, \mu_j$ for $h > 0$, so $\sigma_{b+i,b+j} = \mu_{i+j} - \mu_i \, \mu_j$ for $i, j = 1, \ldots, r$. Finally, the independence implies SYM, and $E[X_t^j \mid X_{t-h} = i] = \mu_j$ if $h > 0$. So $\sigma_{i,b+j} = (i^j - \mu_j) \, p_i$ because of (10).

Example 2.3.2 (CLAR(1) Model) Let $(X_t)_{\mathbb{Z}}$ follow a CLAR(1) model (Grunwald et al. 2000) as described in Appendix A.1. Applying (A.2) to (10), we obtain for $j = 1$ (referring to the estimation of the mean μ) that

$$E[Z_{h,b+1} \cdot Z_{0,i}] = \left(\alpha^h \cdot i + (1 - \alpha^h) \mu\right) p_i,$$

which also holds for $h = 0$. Hence, if SYM holds, $\sigma_{i,b+1}$ in Theorem 2.2.1 becomes

$$\sigma_{i,b+1} = E[Z_{0,b+1} \cdot Z_{0,i}] - \mu \, p_i + 2 \sum_{h=1}^{\infty} \left(E[Z_{h,b+1} \cdot Z_{0,i}] - \mu \, p_i\right)$$

$$= (i - \mu) \, p_i + 2 \sum_{h=1}^{\infty} \alpha^h \, (i - \mu) \, p_i = (i - \mu) \, p_i \, \frac{1 + \alpha}{1 - \alpha}.$$

Furthermore, (9) leads to $E[Z_{h,b+1} Z_{0,b+1}] - \mu^2 = Cov[X_t, X_{t-h}] = \sigma^2 \rho(h)$, where $\rho(h) = \alpha^h$ according to (A.2). So $\sigma_{b+1,b+1}$ in Theorem 2.2.1 becomes

$$\sigma_{b+1,b+1} = \sigma^2 + 2 \sum_{h=1}^{\infty} \sigma^2 \alpha^h = \sigma^2 \frac{1+\alpha}{1-\alpha}.$$

As we have seen in Example 2.3.2, the covariances $\sigma_{i,b+1}$ and $\sigma_{b+1,b+1}$ directly follow from the CLAR(1) property, without further distributional assumptions. To evaluate the remaining expressions, let us consider two particular instances within the CLAR(1) family, which date back to McKenzie (1985) and are often used in applications.

Example 2.3.3 (Poisson INAR(1) Model) Let $(X_t)_{\mathbb{Z}}$ follow the Poisson INAR(1) model from Appendix A.2. Then the marginal distribution $p_i = e^{-\mu} \mu^i / i!$ is determined by only one parameter, the mean parameter μ. So it suffices to consider only one estimator (thus $r = 1$), $\hat{\mu} = \bar{X}$. The results of Example 2.3.2 hold with an additional simplification: because of the Poisson's equidispersion property, we have $\sigma^2 = \mu$.

The remaining entries σ_{ij} in Theorem 2.2.1 for $i, j \in \{0, \ldots, b\}$ simplify with (8) to

$$\sigma_{ij} = (\delta_{i,j} - p_i) p_j + 2 p_j \sum_{h=1}^{\infty} (p_{i|j}{}^{(h)} - p_i)$$

which are computed by either using (A.3) for the Poisson INAR(1)'s h-step-ahead transition probabilities, or by utilizing that (X_t, X_{t-h}) are bivariately Poisson distributed according to $BPoi(\alpha^h \mu; (1 - \alpha^h) \mu, (1 - \alpha^h) \mu)$. Since a closed-form expression for σ_{ij} is rather complex, in practice, one approximates

$$\sigma_{ij} \approx (\delta_{i,j} - p_i) p_j + 2 p_j \sum_{h=1}^{M} (p_{i|j}{}^{(h)} - p_i) \qquad \text{with } M \text{ sufficiently large.}$$

Example 2.3.4 (Binomial AR(1) Model) Let $(X_t)_{\mathbb{Z}}$ follow the binomial AR(1) model from Appendix A.3, i.e., with binomial marginal distribution $p_i = \binom{n}{i} \pi^i (1 - \pi)^{n-i}$. The results of Example 2.3.2 hold with $\mu = n\pi, \sigma^2 = n\pi(1-\pi)$ as well as α replaced by ρ. The required moment estimator of π is defined as $\hat{\pi} := \bar{X}/n$. As a result, the Jacobian $\mathbf{D}$ used for (6) takes a very simple form, $\mathbf{D} = \text{diag}(1, \ldots, 1, 1/n)$.

The entries σ_{ij} in Theorem 2.2.1 for $i, j \in \{0, \ldots, b\}$ are computed as in Example 2.3.3, but using formula (A.7) for the h-step-ahead transition probabilities. Note that in this particular example, the asymptotic covariance matrix $\boldsymbol{\Sigma}^{(p)}$ from (7) can also be computed by using a result in Tavaré and Altham (1983) for finite Markov chains, see Kim and Weiß (2015) for details.

A CLAR(1) process *not* being time reversible is the geometric INAR(1) process.

Example 2.3.5 (Geometric INAR(1) Model) If $(X_t)_{\mathbb{Z}}$ follows the geometric INAR(1) model from Appendix A.2, then, in contrast to Example 2.3.3, the property SYM

does not hold. So numerical approximations are required for Theorem 2.2.1. These include the computation of the $p_{i|j}{}^{(h)}$, since a simple closed-form formula is not available: for N sufficiently large, define $\tilde{\mathbf{P}} := (p_{i|j})_{i,j=0,\ldots,N}$ with the transition probabilities (A.5); then the h-step-ahead transition probabilities $(p_{i|j}{}^{(h)})_{i,j=0,\ldots,N}$ are approximated by the matrix $\tilde{\mathbf{P}}^h$.

The geometric marginal distribution satisfies $\mu = (1-\pi)/\pi$ and $\sigma^2 = (1-\pi)/\pi^2$, and it requires one estimator (thus $r = 1$), $\hat{\pi} = 1/(1 + \bar{X})$. The Jacobian $\mathbf{D}$ used for (6) equals $\mathbf{D} = \mathrm{diag}\big(1, \ldots, 1, -\pi^2\big)$.

In an analogous way as in Example 2.3.5, one can also handle INAR(1) processes with other types of non-Poisson marginal distribution, e. g., a negative binomial distribution (McKenzie 1986). In the latter case, the main difference to Example 2.3.5 is the fact that the negative binomial distribution has two parameters, i. e., now $r = 2$ moments have to be considered for parameter estimation. The required formulae for higher-order joint moments can be found in Schweer and Weiß (2014).

The next examples demonstrate that the count processes to be considered are not limited to Markov chains. As an illustrative example for a higher-order AR-type model, the Poisson INAR(2) model by Alzaid and Al-Osh (1990) is presented.

Example 2.3.6 (Poisson INAR(2) Model) Let $(X_t)_{\mathbb{Z}}$ follow the Poisson INAR(2) model from Appendix A.2, especially Example A.2.2. Like any Poisson INAR(p) process in the sense of Alzaid and Al-Osh (1990), this process is time reversible such that the property SYM holds. All computations can be done in complete analogy to the INAR(1) case in Example 2.3.3, by using that $(X_t, X_{t-h}) \sim \mathrm{BPoi}\big(\rho(h)\,\mu;\ (1 - \rho(h))\,\mu,\ (1 - \rho(h))\,\mu\big)$ and $E[X_t \mid X_{t-h}] = \rho(h)\,X_{t-h} + (1 - \rho(h))\,\mu$, where the ACF satisfies $\rho(1) = \alpha_1$ and $\rho(h) = \alpha_1\,\rho(h-1) + \alpha_2\,\rho(h-2)$ for $h \geq 2$.

It should be pointed out that the argumentation in Example 2.3.6 is easily adapted to the family of Poisson INMA(q) processes (moving-average-type models), which are non-Markovian but q-dependent. The latter property implies that the infinite sums in Theorem 2.2.1 reduce to finite ones, with non-zero summands for $h \leq q$. For such Poisson INMA(q) processes, Weiß (2008) showed that again $(X_t, X_{t-h}) \sim$ BPoi$\big(\rho(h)\,\mu;\ (1 - \rho(h))\,\mu,\ (1 - \rho(h))\,\mu\big)$ holds, where the ACF has to be determined from the specific INMA(q) model. Note that the bivariate distributions of $(X_t, X_{t-1}), \ldots, (X_t, X_{t-q})$ require knowledge only about μ and $\rho(1), \ldots, \rho(q)$, which are easily estimated from given time series data.

A completely different approach than INARMA for obtaining ARMA-like count processes is given by the NDARMA models by Jacobs and Lewis (1983).

Example 2.3.7 (NDARMA Model) Let $(X_t)_{\mathbb{Z}}$ be an *NDARMA(p, q) process* as described in Appendix A.4. According to (A.8), these models satisfy the property SYM with $p_{i|j}{}^{(h)} - p_i = (\delta_{i,j} - p_i)\,\rho(h)$. Defining $c := 1 + 2\sum_{h=1}^{\infty}\rho(h)$, the entries σ_{ij} in Theorem 2.2.1 for $i, j \in \{0, \ldots, b\}$ become $\sigma_{ij} = c\,(\delta_{i,j} - p_i)\,p_j$. Note that this result coincides with the i. i. d. case from Example 2.3.1 except the additional factor c, also see formula (14) in Weiß (2013).

The conditional moments in (10) follow as

$$E[X_t^j | X_{t-h} = i] = \sum_{x=0}^{\infty} x^j \left(p_x + (\delta_{x,i} - p_x) \rho(h) \right) = \mu_j + \rho(h) (i^j - \mu_j).$$

So $\sigma_{i,b+j}$ in Theorem 2.2.1 becomes $\sigma_{i,b+j} = c\,(i^j - \mu_j)\,p_i$, compare again with the i.i.d. result from Example 2.3.1. Finally, the joint moments in (9) are computed as

$$\begin{aligned}
E[X_t^i X_{t-h}^j] &= \sum_{x,y=0}^{\infty} x^i y^j \left(p_x + (\delta_{x,y} - p_x)\rho(h) \right) p_y = \mu_i \mu_j \\
&\quad + \rho(h)(\mu_{i+j} - \mu_i \mu_j),
\end{aligned}$$

so $\sigma_{b+i,b+j}$ in Theorem 2.2.1 becomes $\sigma_{b+i,b+j} = c\,(\mu_{i+j} - \mu_i\,\mu_j)$. Hence, throughout, we obtain the i.i.d. results from Example 2.3.1 together with the additional factor c.

Another non-Markovian example, which could be relevant in practice, is a Hidden-Markov model for counts, see Zucchini and MacDonald (2009) for detailed information. If $\mathbf{A}$ denotes the transition matrix of the hidden Markov chain with $\boldsymbol{\pi}$ being the corresponding stationary marginal distribution, and if $\mathbf{P}(i) := \mathrm{diag}\big(p(i|\cdot)\big)$ denotes the diagonal matrix of all state-dependent probabilities leading to the count $i \in \mathbb{N}_0$, then the marginal and lagged bivariate probabilities are computed as

$$\begin{aligned}
P(X_t = i) &= \mathbf{1}^\top \mathbf{P}(i)\,\boldsymbol{\pi}, \\
P(X_t = i, X_{t-h} = j) &= \mathbf{1}^\top \mathbf{P}(i)\,\mathbf{A}^h\,\mathbf{P}(j)\,\boldsymbol{\pi},
\end{aligned} \tag{11}$$

where $\mathbf{1}$ denotes the vector of ones. So formulae (8)–(10) are again easily computed in practice.

Remark 2.3.8 A further model family commonly used for ARMA-like count time series are INGARCH models (integer-valued generalized autoregressive conditional heteroscedasticity), see Ferland et al. (2006) and Weiß (2018) for background information. Although these models typically satisfy the moment and mixing conditions of Theorem 2.2.1, see Ferland et al. (2006) and Neumann (2011), goodness-of-fit tests w.r.t. to the counts' marginal distribution are not applicable. INGARCH models are defined in terms of their *conditional* distribution (given the available past), e.g., by requiring a *conditional* Poisson distribution for Poisson INGARCH models. But analytic expressions concerning the resulting marginal distribution are not known, so marginal goodness-of-fit statistics cannot be computed for these models.

2.4 Asymptotic distribution of Pearson statistic

In Sect. 2.2, we ended up with the asymptotic normal distribution (6) for $\sqrt{T}\left((\hat{\boldsymbol{p}}, \hat{\boldsymbol{\theta}}) - (\boldsymbol{p}, \boldsymbol{\theta})\right)$. This is now used to derive the asymptotic distributions of $\sqrt{T}\,\boldsymbol{G}(\hat{\boldsymbol{p}}, \boldsymbol{\theta})$ and

$\sqrt{T}\,\boldsymbol{G}(\hat{\boldsymbol{p}},\hat{\boldsymbol{\theta}})$, respectively, remember (4). For this purpose, let $\boldsymbol{z} = (\boldsymbol{z}_1, \boldsymbol{z}_2)$ with $\boldsymbol{z}_1 \in \mathbb{R}^{b+1}$ and $\boldsymbol{z}_2 \in \mathbb{R}^r$, and define the function

$$\boldsymbol{g}(\boldsymbol{z}) := \boldsymbol{g}(\boldsymbol{z}_1, \boldsymbol{z}_2) := \operatorname{diag}\big(\pi_a(\boldsymbol{z}_2)^{-1/2}, \ldots, \pi_{b+1}(\boldsymbol{z}_2)^{-1/2}\big)\,\mathbf{A}\big(\boldsymbol{z}_1 - \boldsymbol{p}(\boldsymbol{z}_2)\big)$$

such that $\boldsymbol{G}(\hat{\boldsymbol{p}}, \boldsymbol{\theta}) = \boldsymbol{g}(\hat{\boldsymbol{p}}, \boldsymbol{\theta})$ and $\boldsymbol{G}(\hat{\boldsymbol{p}}, \hat{\boldsymbol{\theta}}) = \boldsymbol{g}(\hat{\boldsymbol{p}}, \hat{\boldsymbol{\theta}})$. To be able to apply the Delta method to (6), the Jacobian of $\boldsymbol{g}$ is required. Note that the kth component of $\boldsymbol{g}(\boldsymbol{z})$, $k = a, \ldots, b+1$, equals

$$g_k(\boldsymbol{z}) = \sum_{l=0}^{b} a_{kl}\, \pi_k(\boldsymbol{z}_2)^{-1/2}\, \big(z_{1,l} - p_l(\boldsymbol{z}_2)\big).$$

So we obtain the partial derivatives

$$\tfrac{\partial}{\partial z_{1,i}}\, g_k(\boldsymbol{z}) = a_{ki}\, \pi_k(\boldsymbol{z}_2)^{-1/2} \qquad \text{for } i = 0, \ldots, b,$$

$$\tfrac{\partial}{\partial z_{2,j}}\, g_k(\boldsymbol{z}) = \sum_{l=0}^{b} a_{kl} \left(\big(z_{1,l} - p_l(\boldsymbol{z}_2)\big) \tfrac{\partial}{\partial z_{2,j}} \pi_k(\boldsymbol{z}_2)^{-1/2} - \pi_k(\boldsymbol{z}_2)^{-1/2} \tfrac{\partial}{\partial z_{2,j}} p_l(\boldsymbol{z}_2) \right)$$

$$= -\sum_{l=0}^{b} \tfrac{a_{kl}}{2} \big(z_{1,l} - p_l(\boldsymbol{z}_2)\big) \pi_k(\boldsymbol{z}_2)^{-3/2} \sum_{m=0}^{b} a_{km} \tfrac{\partial}{\partial z_{2,j}} p_m(\boldsymbol{z}_2)$$

$$\qquad - \sum_{l=0}^{b} a_{kl}\, \pi_k(\boldsymbol{z}_2)^{-1/2} \tfrac{\partial}{\partial z_{2,j}} p_l(\boldsymbol{z}_2) \qquad \text{for } j = 1, \ldots, r.$$

In the specified-parameter case, we evaluate the reduced Jacobian

$$\mathbf{J}_{\boldsymbol{g}}(\boldsymbol{z}_1) = \big(\tfrac{\partial}{\partial z_{1,i}} g_k(\boldsymbol{z}_1, \boldsymbol{\theta})\big)_{k=a,\ldots,b+1,\ i=0,\ldots,b}$$

in $\boldsymbol{p}$. Together with (7), this leads to

$$\sqrt{T}\,\boldsymbol{G}(\hat{\boldsymbol{p}}, \boldsymbol{\theta}) \xrightarrow{D} \mathrm{N}\big(\boldsymbol{0}, \mathbf{D}_{\mathrm{kn}}\,\boldsymbol{\Sigma}^{(\boldsymbol{p})}\mathbf{D}_{\mathrm{kn}}^{\top}\big) \quad \text{with } \mathbf{D}_{\mathrm{kn}} = \operatorname{diag}\big(\boldsymbol{\pi}(\boldsymbol{\theta})^{-1/2}\big)\mathbf{A}. \qquad (12)$$

Here, $\boldsymbol{\pi}(\boldsymbol{\theta})^{-1/2}$ is computed by applying "$-1/2$" (i.e., $1/\sqrt{\cdot}$) separately in each component of $\boldsymbol{\pi}(\boldsymbol{\theta})$.

In the estimated-parameter case, we evaluate the full Jacobian

$$\mathbf{J}_{\boldsymbol{g}}(\boldsymbol{z}) = \big(\tfrac{\partial}{\partial z_l} g_k(\boldsymbol{z})\big)_{k=a,\ldots,b+1,\ l=0,\ldots,b+r}$$

in $(\boldsymbol{p}, \boldsymbol{\theta})$. Denoting the Jacobian of $\boldsymbol{p}(\boldsymbol{z}_2)$ by $\mathbf{J}_{\boldsymbol{p}}(\boldsymbol{z}_2)$, we obtain together with (6) that

$$\sqrt{T}\,\boldsymbol{G}(\hat{\boldsymbol{p}}, \hat{\boldsymbol{\theta}}) \xrightarrow{D} \mathrm{N}\big(\boldsymbol{0}, \mathbf{D}_{\mathrm{est}}\,\boldsymbol{\Sigma}^{*}\mathbf{D}_{\mathrm{est}}^{\top}\big) \quad \text{with } \mathbf{D}_{\mathrm{est}} = \big(\mathbf{D}_{\mathrm{kn}}, \ -\mathbf{D}_{\mathrm{kn}}\,\mathbf{J}_{\boldsymbol{p}}(\boldsymbol{\theta})\big) \qquad (13)$$

being a block matrix. Let us illustrate the computation of $\mathbf{J}_{\boldsymbol{p}}(\boldsymbol{\theta})$ and hence $\mathbf{D}_{\mathrm{est}}$ with some common types of hypothetical marginal distribution.

Example 2.4.1 (Poisson Marginal) If the hypothetical marginal distribution is a Poisson one, i.e., $X_t \sim \text{Poi}(\mu)$ with $p_i(\mu) = e^{-\mu} \mu^i / i!$, like for the Poisson INAR models in Examples 2.3.3, 2.3.6 and Appendix A.2, then the Jacobian $\mathbf{J}_p$ is easily computed:

$$\frac{\partial}{\partial \mu} p_i(\mu) = e^{-\mu} \frac{\mu^{i-1}}{(i-1)!} - e^{-\mu} \frac{\mu^i}{i!} = p_{i-1}(\mu) - p_i(\mu),$$

which also holds for $i = 0$ with the convention $p_j(\mu) = 0$ for $j < 0$.

Example 2.4.2 (Binomial Marginal) If the hypothetical marginal distribution is a binomial one, i.e., $X_t \sim \text{Bin}(n, \pi)$, i.e., $p_{n,i}(\pi) = \binom{n}{i} \pi^i (1-\pi)^{n-i}$, like for the binomial AR(1) model in Example 2.3.4 and Appendix A.3, then again the Jacobian $\mathbf{J}_p$ is easily computed:

$$\frac{\partial}{\partial \pi} p_{n,i}(\pi) = i \binom{n}{i} \pi^{i-1} (1-\pi)^{n-i} - (n-i) \binom{n}{i} \pi^i (1-\pi)^{n-i-1}$$

$$= n \left(p_{n-1,i-1}(\pi) - p_{n-1,i}(\pi) \right),$$

which also holds for $i = 0, n$ as well as $n = 1$ with the conventions $p_{0,0}(\pi) = 1$ and $p_{m,j}(\pi) = 0$ for $j < 0$ or $j > m$.

Example 2.4.3 (Geometric Marginal) If the hypothetical marginal distribution is the geometric distribution $\text{Geom}(\pi)$, i.e., $p_i(\pi) = \pi (1-\pi)^i$, like for the geometric INAR(1) model in Example 2.3.5 and Appendix A.2, then the Jacobian $\mathbf{J}_p$ is computed by using

$$\frac{\partial}{\partial \pi} p_i(\pi) = (1-\pi)^i - i \pi (1-\pi)^{i-1} = \frac{1}{\pi} p_i(\pi) - i p_{i-1}(\pi),$$

which also holds for $i = 0$ with the convention $p_j(\pi) = 0$ for $j < 0$.

As the final step, we apply Theorem 3.1 in Tan (1977) to (12) and (13) to obtain the asymptotic distribution of Pearson's goodness-of-fit test statistic (4) in both scenarios. This is a quadratic-form distribution, i.e., the distribution of an expression of the form $\sum_{i=1}^{u} \lambda_i Z_i^2$ with $Z_1, \ldots, Z_r$ being i.i.d. $N(0, 1)$-variates [also see Moore (1982)]. In the specified-parameter case, it holds that

$$G^2(\hat{p}, \theta) \xrightarrow{D} \sum_{i=1}^{u} \lambda_i Z_i^2, \tag{14}$$

where $\lambda_1, \ldots, \lambda_u$ are the non-zero eigenvalues of $\mathbf{D}_{kn} \mathbf{\Sigma}^{(p)} \mathbf{D}_{kn}^\top$ according to (12). Note that in the particular case of a binomial AR(1) process (Example 2.3.4 and Appendix A.3), the asymptotic distribution in (14) can also be computed according to the approach of Kim and Weiß (2015).

In the estimated-parameter case, it holds that

$$G^2(\hat{p}, \hat{\theta}) \xrightarrow{D} \sum_{j=1}^{v} \lambda_j^* Z_j^2, \tag{15}$$

where $\lambda_1^*, \ldots, \lambda_v^*$ are the non-zero eigenvalues of $\mathbf{D}_{est} \boldsymbol{\Sigma}^* \mathbf{D}_{est}^\top$ according to (13).

To evaluate such quadratic-form distributions, the R package `CompQuadForm` (Duchesne and Micheaux 2010) can be used, e. g., the functions `davies` or `imhof`. This is also done in the next Sect. 3, where we investigate the asymptotic distributions (14) and (15) from various viewpoints. Before doing this, we conclude with a brief remark concerning the NDARMA models from Example 2.3.7.

Example 2.4.4 (NDARMA Model) Since the involved covariance matrices of an NDARMA(p, q) process differ from those of the i. i. d. case only by the unique factor c as defined in Example 2.3.7, goodness-of-fit testing based on NDARMA processes, with or without estimated parameters, is done by computing the respective critical values for the i. i. d. case and by multiplying them by c, also see Weiß (2013).

3 Computations and simulations

In the sequel, some results from a simulation study are presented to demonstrate the goodness of the asymptotic approximations (14) and (15) to the Pearson statistic's distribution, and to illustrate the application of the Pearson test in practice.

3.1 Finite-sample performance of asymptotic approximations

Since the above approach concerning the distribution of Pearson's goodness-of-fit test statistic (4) relies on asymptotic results for sample size $T \to \infty$, the first question to be analyzed is the one about the performance of the resulting approximation for time series of finite length T. For this purpose, diverse types of count process were simulated (always with 10,000 replications), the Pearson statistics with specified or estimated parameters, respectively, were computed, and empirical properties of the resulting samples are now compared to the corresponding asymptotic properties according to (14) or (15), respectively.

Our first comparison refers to Kim and Weiß (2015), who derived the asymptotic distribution of the Pearson statistic for the special case of a binomial AR(1) process (Example 2.3.4 and Appendix A.3) with specified parameters. In the last four lines of their Table 1, however, Kim and Weiß (2015) gave simulated values of the quantiles $q_{0.25}, q_{0.50}, q_{0.75}, q_{0.95}, q_{0.99}$ also for the case of estimating the binomial parameter π by $\hat{\pi} := \bar{X}/n$. With the novel approach derived in this paper, see Example 2.3.4, we are able to also compute these quantiles asymptotically. This is shown in Table 1, where we see a rather good agreement between the asymptotic and the simulated quantiles, although the sample size $T = 70$ is rather small in a time series context.

Next, we check the finite-sample performance for a number of count models with an unbounded range: the Poisson INAR(1) model (Example 2.3.3 and Appendix A.2) as a first-order model satisfying the property "SYM", the geometric INAR(1) model (Example 2.3.5 and Appendix A.2) as a first-order model violating the property "SYM", and the Poisson INAR(2) model (Example A.2.2 and Appendix A.2) as a second-order model satisfying the property "SYM". In each case, the design parameters were chosen such that Cochran's rule is satisfied, i. e., the expected count per

Table 1 Quantiles of Pearson statistic (estimated parameters) for binomial AR(1) process: asymptotic versus simulated values, where the latter are taken from Table 1 in Kim and Weiß (2015)

n	π	ρ	T	a	b	$q_{0.25}$	$q_{0.50}$	$q_{0.75}$	$q_{0.95}$	$q_{0.99}$
5	0.4	0.00	70	0	4	1.923	1.914	3.357	3.268	5.385	5.246	9.49	9.28	13.28	13.66
5	0.4	0.50	70	0	4	2.449	2.351	4.292	3.962	6.921	6.293	12.33	11.73	17.44	18.19
22	0.4	0.00	70	5	11	3.483	3.554	5.377	5.418	7.870	7.903	12.62	12.59	16.84	16.65
22	0.4	0.50	70	5	11	3.845	3.911	5.951	6.000	8.744	8.621	14.14	13.78	19.03	18.80

Table 2 Mean, standard deviation and quantiles of Pearson statistic for Poisson INAR(1) process with $\mu = 3$: asymptotic versus simulated values

α	T	a	b	Mean		SD		q0.25		q0.50		q0.75		q0.95		q0.99	
Specified parameters																	
0.25	200	0	6	7.81	7.80	4.33	4.26	4.70	4.73	7.03	6.96	10.07	9.95	15.85	16.12	21.06	21.43
	500	0	7	8.82	8.82	4.58	4.49	5.54	5.52	8.06	8.00	11.26	11.15	17.27	17.60	22.61	23.26
	1000	0	8	9.86	9.83	4.79	4.72	6.39	6.44	9.07	9.16	12.44	12.33	18.65	18.66	24.10	25.03
0.50	200	0	6	9.88	9.87	5.92	5.85	5.70	5.68	8.66	8.66	12.68	12.68	20.97	21.15	29.22	29.70
	500	0	7	10.98	10.97	6.18	6.07	6.64	6.65	9.79	9.75	13.98	13.92	22.43	22.63	30.74	31.43
	1000	0	8	11.96	12.05	6.39	6.26	7.58	7.43	10.89	10.71	15.22	15.03	23.82	23.76	32.16	32.66
0.75	200	0	6	17.19	17.44	12.47	11.90	9.22	9.05	14.53	14.07	22.36	21.42	40.36	41.03	59.50	62.55
	500	0	7	18.83	18.97	12.73	12.18	10.53	10.39	16.11	15.73	24.15	23.49	42.35	43.26	61.56	64.42
	1000	0	8	20.46	20.40	13.06	12.39	11.79	11.65	17.57	17.36	25.79	25.59	44.10	44.89	63.33	68.57
Estimated parameters																	
0.25	200	0	6	6.13	6.15	3.50	3.55	3.54	3.59	5.48	5.45	8.03	7.94	12.90	12.78	17.23	17.34
	500	0	7	7.14	7.16	3.81	3.83	4.35	4.33	6.49	6.46	9.24	9.22	14.39	14.41	18.91	18.73
	1000	0	8	8.18	8.17	4.08	4.09	5.17	5.23	7.49	7.53	10.43	10.36	15.83	15.73	20.52	20.75
0.50	200	0	6	6.86	6.89	3.96	4.04	3.93	4.00	6.11	6.10	9.00	8.93	14.59	14.20	19.68	19.37
	500	0	7	7.94	7.98	4.33	4.34	4.81	4.81	7.20	7.18	10.30	10.26	16.19	15.91	21.48	21.17
	1000	0	8	8.95	9.05	4.64	4.61	5.69	5.59	8.27	8.17	11.56	11.36	17.71	17.64	23.16	23.45
0.75	200	0	6	10.16	10.49	6.35	6.68	5.72	5.74	9.04	8.84	13.65	13.12	23.29	21.88	32.82	31.66
	500	0	7	11.69	11.99	7.02	7.12	6.90	6.84	10.52	10.25	15.44	14.81	25.55	24.64	35.42	35.68
	1000	0	8	13.32	13.40	7.62	7.46	8.06	8.01	11.93	11.80	17.11	16.74	27.55	27.44	37.64	39.53

Table 3 Mean, standard deviation and quantiles of Pearson statistic for geometric INAR(1) process with $\mu = 3$: asymptotic versus simulated values

α	T	a	b	Mean		SD		$q_{0.25}$		$q_{0.50}$		$q_{0.75}$		$q_{0.95}$		$q_{0.99}$	
				asym	sim	asym	sim	asym	sim	asym	sim	asym	sim	asym	sim	asym	sim
Specified parameters																	
0.25	200	0	8	9.84	9.88	4.78	4.79	6.37	6.40	9.05	9.05	12.45	12.50	18.80	18.72	24.48	24.83
	500	0	11	12.84	12.79	5.38	5.44	8.95	8.87	12.07	12.00	15.89	15.83	22.78	22.78	28.76	29.26
	1000	0	13	14.84	14.95	5.74	5.96	10.70	10.67	14.08	14.14	18.16	18.30	25.38	25.90	31.55	32.67
0.50	200	0	8	12.21	12.12	6.65	6.61	7.52	7.55	10.89	10.88	15.40	15.05	24.75	24.55	34.22	34.50
	500	0	11	15.21	15.25	7.10	7.20	10.18	10.23	13.96	13.95	18.82	18.78	28.46	28.92	37.97	38.62
	1000	0	13	17.21	17.34	7.38	7.49	11.98	12.12	16.00	15.94	21.07	21.08	30.91	31.27	40.44	41.19
0.75	200	0	8	21.70	21.21	13.90	13.58	12.14	12.03	18.35	17.94	27.45	26.70	48.44	47.26	70.81	70.27
	500	0	11	24.98	24.96	14.25	14.51	15.16	15.06	21.73	21.69	31.07	30.85	52.22	52.16	74.63	75.92
	1000	0	13	27.06	27.26	14.44	14.90	17.12	17.01	23.88	23.84	33.36	33.79	54.57	55.95	76.98	79.05
Estimated parameters																	
0.25	200	0	8	8.25	8.32	4.11	4.14	5.24	5.30	7.57	7.63	10.52	10.58	15.97	16.05	20.73	20.67
	500	0	11	11.21	11.19	4.79	4.78	7.72	7.71	10.53	10.53	13.95	13.95	20.07	19.90	25.26	25.33
	1000	0	13	13.19	13.28	5.19	5.31	9.43	9.42	12.51	12.67	16.22	16.41	22.73	22.92	28.19	28.59
0.50	200	0	8	9.29	9.26	4.88	4.79	5.77	5.77	8.41	8.39	11.83	11.81	18.48	18.09	24.71	24.46
	500	0	11	12.25	12.35	5.48	5.54	8.29	8.38	11.39	11.45	15.25	15.32	22.44	22.84	28.94	29.48
	1000	0	13	14.23	14.34	5.85	5.85	10.02	10.17	13.39	13.51	17.51	17.57	25.02	25.21	31.69	31.70
0.75	200	0	8	14.86	14.67	8.97	8.70	8.59	8.52	12.87	12.87	18.87	18.72	31.96	31.51	45.58	43.51
	500	0	11	18.03	18.09	9.49	9.60	11.39	11.42	16.10	16.04	22.47	22.53	35.99	36.35	49.89	50.67
	1000	0	13	20.09	20.22	9.78	9.91	13.24	13.27	18.20	18.20	24.77	24.99	38.50	39.14	52.50	53.28

Table 4 Mean, standard deviation and quantiles of Pearson statistic for Poisson INAR(2) process with $\mu = 3$: asymptotic versus simulated values

| α_1	α_2	a	b	T	Mean		SD		$q_{0.25}$		$q_{0.50}$		$q_{0.75}$		$q_{0.95}$		$q_{0.99}$	
Specified parameters																		
0.25	0.25	0	6	200	9.42	9.37	5.85	5.64	5.42	5.37	8.24	8.13	12.08	11.94	20.11	20.03	29.22	28.28
		0	7	500	10.47	10.50	6.05	5.84	6.32	6.25	9.32	9.31	13.31	13.36	21.50	21.97	31.37	29.70
		0	8	1000	11.50	11.64	6.42	6.03	7.21	7.20	10.37	10.33	14.51	14.54	22.82	23.77	32.94	31.04
0.50	0.25	0	6	200	15.31	15.20	11.88	11.15	7.78	7.60	12.39	11.91	19.47	18.97	36.88	37.53	60.28	56.06
		0	7	500	16.59	16.51	11.95	11.36	8.90	8.64	13.72	13.25	20.97	20.53	38.49	39.00	61.21	57.72
		0	8	1000	17.78	17.70	12.07	11.51	9.97	9.91	14.97	14.51	22.33	21.98	39.91	40.70	63.30	59.15
0.75	0.10	0	6	200	25.84	25.20	20.82	19.71	12.48	12.05	20.49	19.31	33.15	31.26	64.16	64.21	104.81	97.95
		0	7	500	27.75	27.65	21.95	20.06	14.11	13.72	22.45	21.69	35.39	34.71	66.64	69.14	108.62	100.55
		0	8	1000	29.45	29.17	21.64	20.30	15.62	15.13	24.22	23.41	37.34	36.08	68.71	70.89	108.65	102.64
Estimated parameters																		
0.25	0.25	0	6	200	6.44	6.35	3.71	3.74	3.70	3.64	5.73	5.68	8.41	8.28	13.56	13.32	18.17	18.18
		0	7	500	7.48	7.43	4.04	4.03	4.53	4.47	6.76	6.71	9.65	9.58	15.09	15.06	19.79	19.89
		0	8	1000	8.50	8.49	4.34	4.28	5.37	5.33	7.79	7.68	10.86	10.84	16.55	16.42	22.12	21.52
0.50	0.25	0	6	200	8.36	8.11	5.00	5.25	4.60	4.58	7.24	7.10	10.86	10.46	18.37	17.61	24.98	25.80
		0	7	500	9.61	9.37	5.65	5.61	5.59	5.49	8.48	8.15	12.36	11.84	20.24	19.60	29.49	27.96
		0	8	1000	10.79	10.68	5.93	5.89	6.57	6.53	9.66	9.59	13.75	13.41	21.91	21.77	30.66	29.79
0.75	0.10	0	6	200	13.60	12.84	8.91	9.37	7.04	6.97	11.35	10.84	17.63	16.29	31.64	28.74	45.19	46.20
		0	7	500	15.44	14.86	10.01	9.96	8.45	8.24	13.15	12.62	19.85	18.69	34.56	32.74	50.32	49.67
		0	8	1000	17.12	16.65	10.70	10.39	9.80	9.47	14.82	14.29	21.85	21.00	37.01	36.05	53.78	52.43

category is ≥ 5. For each of these models, selected results are shown in Tables 2, 3 and 4 for illustration; further results can be obtained from the author upon request.

The results in Tables 2, 3 and 4 show a rather good agreement between the asymptotic approximation and the simulated values of mean, standard deviation as well as the quantiles $q_{0.25}, \ldots, q_{0.99}$ throughout, with slightly increasing discrepancy for increasing autocorrelation level. This agreement is better in the estimated-parameter case, which is more relevant for practice anyway.

3.2 Approximate critical values

After having demonstrated the goodness of the asymptotic approximation, let us next investigate the resulting (asymptotic) critical values in more detail, as they are required for applying the Pearson test. The graphs in Fig. 1 show the critical values (level 5%) for the Pearson statistic based on a time series of length $T = 200$, which stems from either a Poisson INAR(1), a geometric INAR(1), or a Poisson INAR(2) process. All processes have the marginal mean $\mu = 3$, and the INAR(2) process satisfies $\alpha_1 = \alpha, \alpha_2 = 0.2$. In view of Cochran's rule, we have two further categories for the geometric marginal distribution.

The respective solid graphs in Fig. 1 show the critical values for the case of specified parameters, while the dashed graphs refer to the estimated-parameter case. Like in the i. i. d. case, we have smaller critical values if the parameters have to be estimated; so ignoring the fact of estimated parameters (just plugging-in estimates instead of true parameter values) would lead to a (very) conservative test procedure. Furthermore, the critical values increase (without bound) with increasing dependence parameter α (note the analogous result in Moore (1982) for a fully-specified Gaussian AR(1) process). So if one would apply the i. i. d.-asymptotics to serially dependent data, the test performance would be severely deteriorated. Finally, comparing the graphs for the Poisson INAR models, we see that the additional dependence caused by $\alpha_2 = 0.2$ leads to a further increase of the critical values.
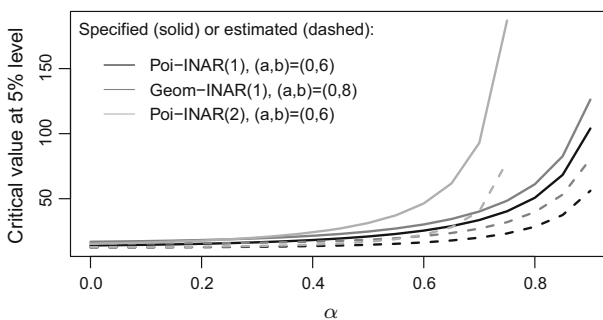
Fig. 1 Critical values (asymptotic approximation) with respect to Poisson or geometric marginal distribution having mean $\mu = 3$, sample size $T = 200$, see Sect. 3.2

3.3 Size and power in practice

Finally, we investigate the size and power of the Pearson test if applied *in practice*: since parameter values are usually not known, they have to be estimated from the available time series, and parameter estimates are also required for evaluating the asymptotic approximations ("plug-in approach"). The presented results rely on simulations (again with 10,000 replications), with μ being estimated by the arithmetic mean, and with α or α_1, α_2, respectively, being estimated from the sample autocorrelation function.

Table 5 shows the simulated sizes for diverse models if the Pearson tests are designed by assuming the level 5%. For the Poisson INAR(2) model as a data-generating process (DGP), also the robustness w.r.t. to a misspecification of the autoregressive order was investigated: the model order was chosen too small, i.e., the test design was done by falsely assuming a Poisson INAR(1) process. For the correctly specified model types and orders (but always with estimated parameters), the simulated sizes are very close to the nominal level of 5%. An exception is the scenario $\alpha = 0.75$ and $T = 200$ (strong autocorrelation), where somes sizes are slightly too large. More discrepancy is observed if the Poisson INAR(2) DGP is underfitted, so the model order should not be chosen too small in practice. Later in this section, we also analyze the effect of an overfitting as well as of misspecifying the model type.

Next, let us investigate the power of the Pearson test. If the test assumes Poisson INAR(1) but the data are generated by a geometric INAR(1) process (having the same mean and ACF), or vice versa, the power was equal to 1.000 in nearly any case (therefore, these values are not tabulated here). So such kind of alternative scenario is nearly always detected. More refined alternative scenarios are analyzed in Table 6. There, the Pearson test assumed a Poisson INAR(1) model, but if a value larger than 1 is given for the index of dispersion, $I = \sigma^2/\mu$, then the true DGP was a negative binomial (NB) INAR(1) process. More precisely, the innovations (ϵ_t) are NB-distributed in such a way that the observations exhibit the given values of μ, I and $\rho(1)$. It can be seen that the power increases with increasing I, and the increase is faster for larger sample sizes T, as expected. In applications, however, one has to be aware of the fact that the power of detection becomes worse with an increasing level of autocorrelation. Our conclusions slightly differ for the second power scenario considered in Table 6: there, the DGP was a Poisson INARCH(1) process (see Remark 2.3.8), which shows increasing overdispersion with increasing autocorrelation level (we have the relation $I = 1/(1 - \rho(1)^2)$). Thus, it is plausible that now the power improves with increasing autocorrelation.

Let us now return to the discussion of Table 5. There, we observed that an underfitting of the actual DGP is problematic and may lead to considerably increased sizes. So it is natural to ask for the effect of *over*fitting the model order instead. Table 7 provides results if an i.i.d. DGP ("model order 0") is overfitted by an INAR(1) model, and Table 8 if an INAR(1) DGP is overfitted by an INAR(2) model. It can be seen that the effect on size and power is very small, sometimes these values are slightly reduced if the model order is unnecessarily large. But the amount of reduction does not appear to be of practical relevance. So an overfitting of the DGP is substantially

Table 5 Simulated size of Pearson test (level 5%) for Poisson INAR(1), Poisson INAR(2) (also if underfitted by Poisson INAR(1)), and geometric INAR(1) process; all having mean $\mu = 3$

$\alpha = $ α_1	α_2	T	a	b	Poi-INAR(1)	Poi-INAR(2)	Underfitted as Poi-INAR(1)	a	b	Geom-INAR(1)
0.25	0.25	200	0	6	0.049	0.046	0.058	0	8	0.049
		500	0	7	0.047	0.049	0.062	0	11	0.052
		1000	0	8	0.048	0.047	0.059	0	13	0.051
0.50	0.25	200	0	6	0.047	0.055	0.104	0	8	0.055
		500	0	7	0.048	0.044	0.100	0	11	0.051
		1000	0	8	0.048	0.048	0.110	0	13	0.051
0.75	0.10	200	0	6	0.055	0.076	0.132	0	8	0.078
		500	0	7	0.045	0.057	0.125	0	11	0.058
		1000	0	8	0.052	0.052	0.127	0	13	0.052

Table 6 Simulated size of Pearson test (level 5%) for Poisson INAR(1) process ($I = 1$), and simulated power for NB-INAR(1) process ($I > 1$) or Poisson INARCH(1) process, respectively; all having mean $\mu = 3$

$\rho(1)$	T	a	b	Rejection rates; $I =$					INARCH(1)
				1	1.05	1.10	1.20	1.50	
0.25	200	0	6	0.049	0.062	0.093	0.220	0.802	0.068
	500	0	7	0.047	0.084	0.170	0.524	0.997	0.107
	1000	0	8	0.048	0.124	0.323	0.851	1.000	0.169
0.50	200	0	6	0.047	0.056	0.082	0.191	0.690	0.401
	500	0	7	0.048	0.073	0.157	0.447	0.984	0.834
	1000	0	8	0.048	0.110	0.285	0.775	1.000	0.991
0.75	200	0	6	0.055	0.054	0.068	0.121	0.452	0.980
	500	0	7	0.045	0.061	0.110	0.282	0.877	1.000
	1000	0	8	0.052	0.089	0.187	0.534	0.993	1.000

less problematic than an underfitting. Thus, we conclude that in practice, one should better choose a somewhat larger model order in case of doubt.

The previous robustness study allowed for a misspecification of the model order, but it assumed that the model family was chosen correctly. So finally, we also consider the case where the model type is misspecified, see Table 9. The considered types of misspecification are chosen such that in practice, there is a large risk of choosing the wrong model: the marginal distribution of the DGP is the same as that of the wrong model, and also the autocorrelation structure is very similar or even identical. In the left part of Table 9, the DGP is Poisson INMA(1), so it has a Poisson marginal distribution, and its autocorrelation structure might be confused with the one of a Poisson INAR(1) process. In the right part of Table 9, two types of DGP having a geometric marginal distribution and an AR(1)-like ACF are chosen (and confused with the geometric INAR(1) model): the geometric AR(1) process proposed by Ristić et al. (2009) is defined by an INAR(1)-like recursion but using negative-binomial thinning instead of binomial thinning (hence we abbreviate it as "NT-AR(1)" in Table 9), the one proposed by Al-Osh and Aly (1992) uses some kind of "iterated thinning" ("IT-AR(1)"), where first a binomial thinning and then a negative binomial thinning is applied at each time t. The sizes of the misspecified Poisson INAR(1) model in the left part of Table 9 are still very close to 5%, so erroneously treating the INMA(1) as INAR(1) data is not problematic. For the misspecified geometric INAR(1) models, the sizes become visibly smaller than 5% for large autocorrelation levels. So the Pearson test becomes conservative in these cases. On the other hand, for increasing autocorrelation, it is usually also more easy to distinguish between the different models, since then, e. g., sample paths or conditional variances are more pronounced.

3.4 Real-data examples

We conclude our investigations with two real-data examples. The count time series $x_1, \ldots, x_T$ shown in Fig. 2a is taken from the book by Weiß (2018). It consists of

Table 7 Simulated size and power of Pearson test (level 5%) for i.i.d. Poisson ($I = 1$) and i.i.d. NB counts ($I > 1$), respectively. Left part: correctly specified i.i.d. model; right part: overfitted INAR(1) model

μ	T	a	b	Rejection rates									
				For i.i.d.; $I =$					For INAR(1) overfit; $I =$				
				1	1.05	1.10	1.20	1.50	1	1.05	1.10	1.20	1.50
1.5	200	0	4	0.047	0.063	0.111	0.274	0.863	0.047	0.062	0.110	0.273	0.862
	500	0	5	0.050	0.104	0.232	0.633	0.999	0.050	0.104	0.232	0.633	0.999
	1000	0	5	0.051	0.134	0.375	0.901	1.000	0.051	0.134	0.374	0.901	1.000
3.0	200	0	6	0.050	0.062	0.104	0.231	0.834	0.049	0.062	0.103	0.231	0.833
	500	0	7	0.049	0.086	0.182	0.552	0.999	0.049	0.086	0.182	0.552	0.999
	1000	0	8	0.052	0.117	0.332	0.875	1.000	0.052	0.117	0.332	0.875	1.000
5.0	200	1	9	0.055	0.063	0.093	0.216	0.826	0.054	0.062	0.093	0.216	0.825
	500	1	10	0.050	0.083	0.170	0.523	0.998	0.050	0.083	0.169	0.523	0.998
	1000	0	11	0.049	0.119	0.300	0.842	1.000	0.049	0.119	0.300	0.842	1.000

Table 8 Simulated size and power of Pearson test (level 5%) for Poisson INAR(1) process ($I = 1$) and NB-INAR(1) process ($I > 1$), respectively; all having mean $\mu = 3$. Left part: correctly specified INAR(1) model; right part: overfitted INAR(2) model

α	T	a	b	Rejection rates									
				For INAR(1); $I =$					For INAR(2) overfit; $I =$				
				1	1.05	1.10	1.20	1.50	1	1.05	1.10	1.20	1.50
0.25	200	0	6	0.049	0.062	0.093	0.220	0.802	0.049	0.062	0.093	0.219	0.801
	500	0	7	0.047	0.084	0.170	0.524	0.997	0.047	0.083	0.169	0.524	0.997
	1000	0	8	0.048	0.124	0.323	0.851	1.000	0.048	0.124	0.322	0.851	1.000
0.50	200	0	6	0.047	0.056	0.082	0.191	0.690	0.045	0.054	0.081	0.188	0.687
	500	0	7	0.048	0.073	0.157	0.447	0.984	0.046	0.071	0.156	0.445	0.984
	1000	0	8	0.048	0.110	0.285	0.775	1.000	0.047	0.109	0.283	0.774	1.000
0.75	200	0	6	0.055	0.054	0.068	0.121	0.452	0.050	0.050	0.063	0.113	0.440
	500	0	7	0.045	0.061	0.110	0.282	0.877	0.041	0.056	0.104	0.274	0.872
	1000	0	8	0.052	0.089	0.187	0.534	0.993	0.049	0.086	0.180	0.528	0.993

Table 9 Simulated size of Pearson test (level 5%) under model misspecification: true DGP Poisson INMA(1), but test assumes Poisson INAR(1) (left table); true DGP geometric AR(1) using either negative-binomial thinning ("NT-AR(1)") or iterated thinning ("IT-AR(1)"), but test assumes geometric INAR(1) (right table); all having mean $\mu = 3$

Simulated size under Poi-INAR(1); DGP:				Poi-INMA(1)
$\rho(1)$	T	a	b	
0.25	200	0	6	0.048
	500	0	7	0.047
	1000	0	8	0.048
0.40	200	0	6	0.050
	500	0	7	0.045
	1000	0	8	0.051

Simulated size under Geom-INAR(1); DGP:				Geom-NT-AR(1)	Geom-IT-AR(1)
$\rho(1)$	T	a	b		
0.25	200	0	8	0.052	0.052
	500	0	11	0.049	0.042
	1000	0	13	0.050	0.048
0.50	200	0	8	0.048	0.034
	500	0	11	0.041	0.031
	1000	0	13	0.043	0.037
0.75	200	0	8	0.025	0.024
	500	0	11	0.014	0.015
	1000	0	13	0.011	0.013

the daily numbers of downloads of a TEX editor for the period Jun. 2006 to Feb. 2007 (hence length $T = 267$). As can be seen from the PACF plot, we are concerned with an AR(1)-like autocorrelation structure, with a rather low autocorrelation level: $\hat{\rho}(1) \approx 0.245$. The download counts have the mean ≈ 2.401 and the large dispersion index ≈ 3.127. In view of the autocorrelation structure, the INAR(1) family appears to be a plausible choice for the data, and in view of the strong degree of overdispersion, it appears plausible to test the null hypothesis of a geometric marginal distribution within this INAR(1) family (on level 5%).

The same null hypothesis is also to be tested for the second data example, plotted in Fig. 2b, which consists of $T = 800$ counts of so-called iceberg orders concerning the Deutsche Telekom shares traded in the XETRA system of Deutsche Börse, measured every 20 minutes for 32 consecutive trading days in the first quarter of 2004 (Jung and Tremayne 2011). According to the plotted PACF, we again have an AR(1)-like auto-correlation structure, but now of larger extend ($\hat{\rho}(1) \approx 0.635$). Mean and dispersion index equal ≈ 1.406 and ≈ 1.551, respectively.

For the download counts, we estimate the parameter of the hypothetical geometric distribution as $\hat{\pi} \approx 0.294$. The Pearson statistic is computed for the nine categories defined by $a = 0$ and $b = 7$, and it takes the value ≈ 1.179. The critical value obtained for the hypothetical geometric INAR(1) model equals ≈ 14.504 such that we cannot
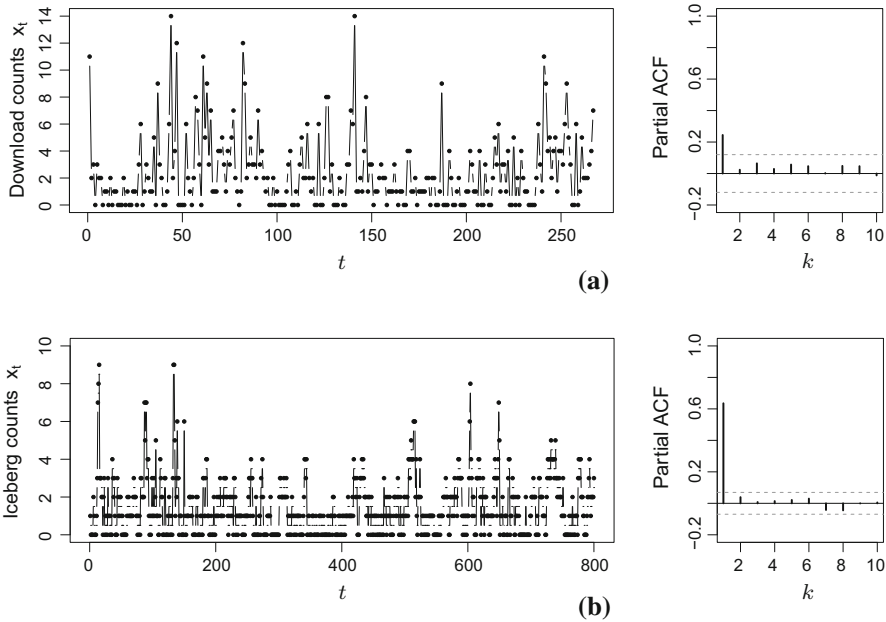
Fig. 2 Time series plot and PACF plot (against lag k) of data examples from Sect. 3.4: **a** download counts, **b** counts of iceberg orders
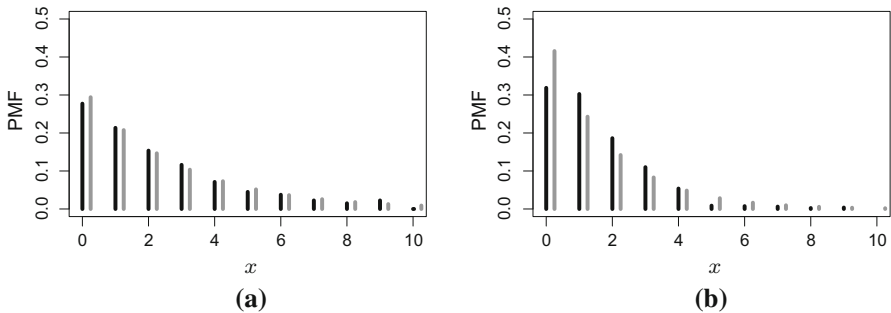
Fig. 3 Pmf plot (against count x) for data examples from Sect. 3.4: sample pmf (in black) and geometric pmf (in gray) for **a** download counts, **b** counts of iceberg orders

reject the null hypothesis. In fact, looking at Fig. 3a, where the sample pmf (black) is compared to the pmf of Geom($\hat{\pi}$) (gray), we see a rather good agreement. It should also be noted that the critical value under an i. i. d.-assumption would only be slightly smaller than the above critical value: ≈ 14.067 (95% quantile of χ^2_{9-1-1}-distribution). This small discrepancy is plausible in view of Fig. 1, where the critical values show a small slope for low values of α (here, we have $\hat{\alpha} \approx 0.245$).

Things differ for the second data example. The estimated geometric parameter equals $\hat{\pi} \approx 0.416$, so we choose the same categorization as before ($a = 0$, $b = 7$) and, thus, obtain the same critical value under an i. i. d.-assumption: ≈ 14.067. The critical value for the hypothetical geometric INAR(1) model is now much larger,

≈ 21.106, which is confirmed by Fig. 1 and the rather large estimate $\hat{\alpha} \approx 0.635$. The Pearson statistic, however, becomes even larger, ≈ 67.286, so this time, we have to reject the null of a geometric marginal distribution. Considering the pmf plots in Fig. 3b, this decision appears plausible as both pmfs deviate visibly from each other, especially for low counts $x \leq 2$.

4 Discussion

If Pearson's goodness-of-fit test statistic is applied to data stemming from a count process, its distribution can be asymptotically approximated with the help of a quadratic-form distribution. The specific distribution can be explicitly computed for a number of practically relevant count process models. The approach does not only cover the situation where the null model is fully specified, but also where parameters have to be estimated. The simulation study showed that the obtained asymptotic approximation works rather well for time series of finite length, and that the test can be successfully applied in practice to uncover model violations. Also the effect of different types of model misspecification was investigated. It turned out that an overfitting was clearly less problematic than an underfitting, so the model order should not be chosen too small. Also cases of misspecifying the model type have been analyzed, where the test's performance is affected especially for large autocorrelation levels. So careful model selection is recommended before applying the test to the given count time series.

For future research, it would be interesting to analyze how to best choose the categories (1) for the Pearson statistic such that the asymptotic approximation works well *and* we obtain an optimal power; in this work, we used the popular Cochran's rule for simplicity. Another research direction could be to investigate if and how the presented approach applies to the family of scaled Bregman divergences (Kißlinger and Stummer 2016). In view of Remark 2.3.8, an important question would be to analyze if a Pearson-like test could also be developed for the *conditional* distribution of a count process, since many important count time series models, like INGARCH and regression models, are defined by specifying the conditional distribution. Finally, returning to the possible problem of a model misspecification, it would be desirable for practice to have a nonparametric way of estimating the covariance matrices $\boldsymbol{\Sigma}^{(p)}$, $\boldsymbol{\Sigma}^*$ involved in computing the Pearson statistic's asymptotics. In this context, the Editor pointed out the work by Francq and Zakoïan (2013), where the estimation of the parameter vector (say, $\boldsymbol{\theta}$) for a time series' marginal distribution is considered. In addition, a nonparametric estimator for $\hat{\boldsymbol{\theta}}$'s covariance matrix is developed. For future research, it should be tried to develop an analogous approach for nonparametrically estimating $\boldsymbol{\Sigma}^{(p)}$, $\boldsymbol{\Sigma}^*$.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

A Specific models for count processes

The subsequent models are used in the main part of this article to illustrate the derivation of the asymptotic distributions of the considered goodness-of-fit tests.

A.1 CLAR(1) model

Many important models for count Markov chains belong to the class of *conditional linear autoregressive models* of order one (CLAR(1) models) as discussed by Grunwald et al. ([2000](#)). The homogeneous count Markov chain $(X_t)_{\mathbb{Z}}$ is said to have CLAR(1) structure if the conditional mean of X_t is linear in X_{t-1}, i.e., if

$$E[X_t \mid X_{t-1}] = \alpha \cdot X_{t-1} + \beta \qquad \text{for some } \alpha \in \mathbb{R} \text{ and } \beta > 0. \qquad \text{(A.1)}$$

The condition $|\alpha| < 1$ guarantees a finite stationary mean given by $\mu := E[X_t] = \beta/(1-\alpha)$. If also the stationary variance of X_t is finite, i.e., $\sigma^2 := V[X_t] < \infty$, then the autocorrelation function (ACF) is of AR(1)-type, i.e., it altogether holds that

$$\rho(h) := Corr[X_t, X_{t-h}] = \alpha^h, \qquad E[X_t \mid X_{t-h}] = \alpha^h \cdot X_{t-h} + (1-\alpha^h)\,\mu. \qquad \text{(A.2)}$$

A.2 INAR models

If X is a discrete random variable with range $\mathbb{N}_0$ and if $\alpha \in (0; 1)$, then the random variable $\alpha \circ X := \sum_{i=1}^{X} Z_i$ is said to arise from X by *binomial thinning* (Steutel and Harn [1979](#)). Here, the Z_i are i.i.d. binary random variables with $P(Z_i = 1) = \alpha$, which are also independent of X. Hence, $\alpha \circ X$ has a conditional binomial distribution given the value of X, i.e., $\alpha \circ X|X \sim \text{Bin}(X, \alpha)$. The boundary values $\alpha = 0$ and $\alpha = 1$ might be included into this definition by setting $0 \circ X := 0$ and $1 \circ X := X$.

Using the random operator "$\circ$", McKenzie ([1985](#)) defined the *INAR(1) model* in the following way.

Definition A.2.1 (*INAR(1) Model*) Let the innovations $(\epsilon_t)_{\mathbb{Z}}$ be an i.i.d. process with range $\mathbb{N}_0$, denote $E[\epsilon_t] = \mu_\epsilon$, $V[\epsilon_t] = \sigma_\epsilon^2$. Let $\alpha \in (0; 1)$. A process $(X_t)_{\mathbb{Z}}$ of observations, which follows the recursion

$$X_t = \alpha \circ X_{t-1} + \epsilon_t,$$

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 ϵ_t are independent of $(X_s)_{s<t}$.

The most popular instance of the INAR(1) family is the Poisson INAR(1) model (McKenzie 1985), which assumes the innovations $(\epsilon_t)_{\mathbb{Z}}$ to be i.i.d. according to the Poisson distribution Poi(λ). A Poisson INAR(1) process is an irreducible and aperiodic Markov chain with a unique stationary marginal distribution for $(X_t)_{\mathbb{Z}}$, the Poisson distribution Poi(μ) with $\mu = \frac{\lambda}{1-\alpha}$. It is also α-mixing with geometrically decreasing weights (Schweer and Weiß 2014). Furthermore, the Poisson INAR(1) model constitutes the only instance within the INAR(1) family, which is time reversible (McKenzie 1985; Schweer 2015).

The (Poisson) INAR(1) model belongs to the class of CLAR(1) models, so it satisfies (A.2). The h-step-ahead transition probabilities are given by (Freeland and McCabe 2004)

$$p_{i|j}^{(h)} = \sum_{m=0}^{\min\{i,j\}} \binom{j}{m} \alpha^{h\,m} (1-\alpha^h)^{j-m} \cdot e^{-\mu\,(1-\alpha^h)} \frac{\big(\mu\,(1-\alpha^h)\big)^{i-m}}{(i-m)!}, \quad \text{(A.3)}$$

and (X_t, X_{t-h}) are bivariately Poisson distributed (Johnson et al. 1997) according to BPoi$\big(\alpha^h\,\mu;\ (1-\alpha^h)\,\mu,\ (1-\alpha^h)\,\mu\big)$ (Alzaid and Al-Osh 1988). Here, BPoi$(\lambda_0;\ \lambda_1, \lambda_2)$ refers to the joint distribution of $(Y_0+Y_1, Y_0+Y_2)^\top$ with independent Poisson variates $Y_i \sim$ Poi(λ_i) for $i = 0, 1, 2$.

A simple example of an INAR(1) process *not* being time reversible is the geometric INAR(1) process (McKenzie 1985, 1986), which has marginal distribution Geom(π), i.e., $P(X_t = x) = \pi\,(1-\pi)^x$. Here, the innovations stem from a zero-inflated geometric distribution,

$$\epsilon_t := B_t\,G_t \quad \text{with independent } B_t \sim \text{Bin}(1, 1-\alpha),\ G_t \sim \text{Geom}(\pi), \quad \text{(A.4)}$$

with pmf $P(\epsilon = x) = \delta_{x,0}\,\alpha + (1-\alpha)\,\pi\,(1-\pi)^x$. Hence, the 1-step-ahead transition probabilities are computed as

$$p_{i|j} = \sum_{m=0}^{\min\{i,j\}} \binom{j}{m} \alpha^m (1-\alpha)^{j-m} \cdot P(\epsilon = i - m). \quad \text{(A.5)}$$

It is also possible to obtain any other member of the family of negative binomial distributions as a marginal distribution by choosing an appropritae innovations' distribution, see McKenzie (1986) for details.

It is also possible to extend the INAR(1) recursion in Definition A.2.1 to a pth-order autoregression of the form

$$X_t = \alpha_1 \circ X_{t-1} + \cdots + \alpha_p \circ X_{t-p} + \epsilon_t \quad \text{with } \alpha_\bullet := \sum_{j=1}^{p} \alpha_j < 1. \quad \text{(A.6)}$$

Due to the stochastic nature of the thinnings involved in (A.6), however, additional assumptions concerning the thinnings $(\alpha_1 \circ X_t, \ldots, \alpha_p \circ X_t)$ are required. While the INAR(p) model by Du and Li (1991) assumes the conditional independence of

$(\alpha_1 \circ X_t, \ldots, \alpha_p \circ X_t)$ given X_t, the one by Alzaid and Al-Osh (1990) supposes a conditional multinomial distribution. As shown by Schweer (2015), only the latter model continues the INAR(1)'s property that we have time reversibility exactly in the case of Poisson innovations (then also the observations are Poisson-distributed), while the INAR(p) model by Du and Li (1991) is neither time reversible nor does it have Poisson marginals. For this reason, we shall focus here on the time reversible *Poisson INAR(p) model* according to Alzaid and Al-Osh (1990), where the innovations are i.i.d. Poi(λ) and, hence, the observations have the stationary marginal distribution Poi(μ) with $\mu = \lambda/(1 - \alpha_\bullet)$.

Example A.2.2 (Poisson INAR(2) Model) Solving the Yule–Walker-type equations (3.6) and (3.8) in Alzaid and Al-Osh (1990), the ACF of the Poisson INAR(2) model becomes

$$\rho(1) = \alpha_1, \qquad \rho(h) = \alpha_1 \, \rho(h-1) + \alpha_2 \, \rho(h-2) \quad \text{for } h \geq 2.$$

As shown in Appendix B, the lagged observations X_t and X_{t-h} with $h \in \mathbb{N}$ are bivariately Poisson distributed,

$$(X_t, X_{t-h}) \sim \text{BPoi}\Big(\rho(h)\,\mu;\ \big(1 - \rho(h)\big)\,\mu,\ \big(1 - \rho(h)\big)\,\mu\Big),$$

with conditional mean $E[X_t \mid X_{t-h}] = \rho(h)\,X_{t-h} + \big(1 - \rho(h)\big)\,\mu$.

A.3 Binomial AR(1) model

In many applications, it is known that the observed count data cannot become arbitrarily large, but their range has a natural upper bound $n \in \mathbb{N}$ that can never be exceeded. For the case of such time series of counts supported on $\{0, \ldots, n\}$, McKenzie (1985) proposed the *binomial AR(1) model*.

Definition A.3.1 (*Binomial AR(1) Model*) Let $\rho \in \big(\max\{-\frac{\pi}{1-\pi}, -\frac{1-\pi}{\pi}\}\,;\ 1\big)$ and $\pi \in (0; 1)$. Define $\beta := \pi \cdot (1 - \rho)$ and $\alpha := \beta + \rho$. Fix $n \in \mathbb{N}$. The process $(X_t)_{\mathbb{Z}}$, defined by the recursion

$$X_t = \alpha \circ X_{t-1} + \beta \circ (n - X_{t-1}),$$

where all thinnings are performed independently of each other, and where the thinnings at time t are independent of $(X_s)_{s<t}$, is referred to as a *binomial AR(1) process*.

The condition on ρ guarantees that the derived parameters α, β satisfy $\alpha, \beta \in (0; 1)$, i.e., these parameters can indeed serve as thinning probabilities.

It is known that $(X_t)_{\mathbb{Z}}$ is a stationary, ergodic and ϕ-mixing finite Markov chain (again with geometrically decreasing weights), the marginal distribution of which is Bin(n, π) (McKenzie 1985; Kim and Weiß 2015). The binomial AR(1) model belongs

to the class of CLAR(1) models, so it satisfies (A.2), and it is time reversible (McKenzie 1985). The h-step-ahead transition probabilities are given by (Weiß and Pollett 2012)

$$p_{i|j}^{(h)} = \sum_{m=\max\{0,i+j-n\}}^{\min\{i,j\}} \binom{j}{m}\binom{n-j}{i-m} \alpha_h^m (1-\alpha_h)^{j-m} \beta_h^{i-m} (1-\beta_h)^{n-j+m-i}, \quad \text{(A.7)}$$

where $\beta_h := \pi \cdot (1 - \rho^h)$ and $\alpha_h := \beta_h + \rho^h$.

A.4 NDARMA model for counts

The "new" *discrete ARMA (NDARMA) models* have been proposed by Jacobs and Lewis (1983). They generate an ARMA-like dependence structure through some kind of random mixture.

Definition A.4.1 (*NDARMA Model for Counts*) Let the observations $(X_t)_{\mathbb{Z}}$ and the innovations $(\epsilon_t)_{\mathbb{Z}}$ be count processes, where $(\epsilon_t)_{\mathbb{Z}}$ is i. i. d. with $P(\epsilon_t = i) = p_i$, and where ϵ_t is independent of $(X_s)_{s<t}$. The random mixture is obtained through the i. i. d. multinomial random vectors

$$(\alpha_{t,1}, \ldots, \alpha_{t,p}, \beta_{t,0}, \ldots, \beta_{t,q}) \quad \sim \text{MULT}(1; \phi_1, \ldots, \phi_p, \varphi_0, \ldots, \varphi_q),$$

which are independent of $(\epsilon_t)_{\mathbb{Z}}$ and of $(X_s)_{s<t}$. Then $(X_t)_{\mathbb{Z}}$ is said to be an *NDARMA(p, q) process* if it follows the recursion

$$X_t = \alpha_{t,1} \cdot X_{t-1} + \cdots + \alpha_{t,p} \cdot X_{t-p} + \beta_{t,0} \cdot \epsilon_t + \cdots + \beta_{t,q} \cdot \epsilon_{t-q}.$$

The stationary marginal distribution of X_t is identical to that of ϵ_t, i. e., $P(X_t = i) = p_i = P(\epsilon_t = i)$, and we always have

$$p_{i|j}^{(h)} = p_i \cdot \left(1 - \rho(h)\right) + \delta_{i,j} \cdot \rho(h). \quad \text{(A.8)}$$

The autocorrelations are non-negative and can be determined from the Yule–Walker equations (Jacobs and Lewis 1983)

$$\rho(h) = \sum_{j=1}^{p} \phi_j \cdot \rho(|h-j|) + \sum_{i=0}^{q-h} \varphi_{i+h} \cdot r(i) \quad \text{for } h \geq 1,$$

where the $r(i)$ satisfy

$$r(i) = \sum_{j=\max\{0,i-p\}}^{i-1} \phi_{i-j} \cdot r(j) + \varphi_i \mathbb{1}(0 \leq i \leq q),$$

which implies $r(i) = 0$ for $i < 0$, and $r(0) = \varphi_0$. Mixing properties have been established by Weiß (2013).

B Bivariate distributions of INAR(2) model

We pick up the derivations of Alzaid and Al-Osh (1990) and extend them to obtain the bivariate distribution of X_t and X_{t-h} for $h \in \mathbb{N}$. Let $(X_t)_{\mathbb{Z}}$ follow an INAR(2) model, where we first do not further specify the innovations' distribution. Define the sequence of weights $(w_j)_{j \geq -1}$ by

$$w_{-1} = 0, \quad w_0 = 1, \quad w_j = \alpha_1 w_{j-1} + \alpha_2 w_{j-2} \quad \text{for } j = 1, 2, \ldots, \qquad \text{(B.1)}$$

which satisfy $\sum_{j=0}^{\infty} w_j = 1/(1 - \alpha_\bullet)$ (Alzaid and Al-Osh 1990 [p. 317]). Let us introduce a further sequence of coefficients $(\boldsymbol{a}_j)_{j \geq -1}$ with $\boldsymbol{a}_j = (a_{j,1}, a_{j,2})^\top$:

$$\boldsymbol{a}_{-1} = (0, 1)^\top, \quad \boldsymbol{a}_0 = (1, 0)^\top, \quad \boldsymbol{a}_j = \alpha_1 \boldsymbol{a}_{j-1} + \alpha_2 \boldsymbol{a}_{j-2} \quad \text{for } j = 1, 2, \ldots \text{ (B.2)}$$

Obviously, the first components are identical to the weights (B.1), $a_{j,1} = w_j$ for all $j \geq -1$, while the second components satisfy $a_{j,2} = \alpha_2 a_{j-1,1} = \alpha_2 w_{j-1}$ for $j \geq 0$.

Following Alzaid and Al-Osh (1990), p. 320, we define the bivariate process $(\boldsymbol{X}_t)_{\mathbb{Z}}$ by $\boldsymbol{X}_t = (X_t, \alpha_2 \circ X_{t-1})^\top$, which is a Markov chain satisfying

$$E\big[z_1^{X_{t,1}} z_2^{X_{t,2}} \mid \boldsymbol{X}_{t-1}, \ldots\big] = \text{pgf}_\epsilon(z_1) \big(1 + \alpha_1 (z_1 - 1) + \alpha_2 (z_2 - 1)\big)^{X_{t-1,1}} z_1^{X_{t-1,2}}. \tag{B.3}$$

In a first step, we extend this result to arbitrary time lags $h \in \mathbb{N}$. Using the coefficients (B.2), we rewrite (B.3) as

$$\begin{aligned}
E\Big[&\big(1 + a_{0,1} (z_1 - 1) + a_{0,2} (z_2 - 1)\big)^{X_{t,1}} \\
&\big(1 + a_{-1,1} (z_1 - 1) + a_{-1,2} (z_2 - 1)\big)^{X_{t,2}} \mid \boldsymbol{X}_{t-1}, \ldots\Big] \\
&= \text{pgf}_\epsilon \big(1 + a_{0,1} (z_1 - 1) + a_{0,2} (z_2 - 1)\big) \\
&\quad \cdot \big(1 + a_{1,1} (z_1 - 1) + a_{1,2} (z_2 - 1)\big)^{X_{t-1,1}} \\
&\quad \big(1 + a_{0,1} (z_1 - 1) + a_{0,2} (z_2 - 1)\big)^{X_{t-1,2}}.
\end{aligned} \tag{B.4}$$

Using the law of total expectation, we apply (B.4) and obtain

$$\begin{aligned}
E\big[z_1^{X_{t,1}} z_2^{X_{t,2}} \mid \boldsymbol{X}_{t-h}\big] &= \text{pgf}_\epsilon \ \big(1 + a_{0,1} (z_1 - 1) + a_{0,2} (z_2 - 1)\big) \\
&\cdots E\Big[\big(1 + a_{1,1} (z_1 - 1) + a_{1,2} (z_2 - 1)\big)^{X_{t-1,1}} \\
&\quad \big(1 + a_{0,1} (z_1 - 1) + a_{0,2} (z_2 - 1)\big)^{X_{t-1,2}} \mid \boldsymbol{X}_{t-h}\Big] \\
&= \cdots = \prod_{j=0}^{h-1} \text{pgf}_\epsilon \big(1 + a_{j,1} (z_1 - 1) + a_{j,2} (z_2 - 1)\big)
\end{aligned}$$

$$\cdot \left(1 + a_{h,1} \left(z_1 - 1\right) + a_{h,2} \left(z_2 - 1\right)\right)^{X_{t-h,1}}$$
$$\left(1 + a_{h-1,1} \left(z_1 - 1\right) + a_{h-1,2} \left(z_2 - 1\right)\right)^{X_{t-h,2}}. \tag{B.5}$$

On the one hand, this implies the marginal pgf as

$$\begin{aligned} \mathrm{pgf}_X(z) &= \textstyle\prod_{j=0}^{\infty} \mathrm{pgf}_\epsilon\left(1 + w_j \left(z_1 - 1\right) + \alpha_2 \, w_{j-1} \left(z_2 - 1\right)\right), \\ \mathrm{pgf}_X(z) &= \mathrm{pgf}_X(z, 1) = \textstyle\prod_{j=0}^{\infty} \mathrm{pgf}_\epsilon\left(1 + w_j \left(z - 1\right)\right), \end{aligned} \tag{B.6}$$

also see (4.4) and Theorem 2.1 in Alzaid and Al-Osh (1990). On the other hand, we compute from (B.5) the lagged bivariate pgf as

$$\mathrm{pgf}_{X_t, X_{t-h}}(z, y) =$$
$$\prod_{j=0}^{h-1} \mathrm{pgf}_\epsilon\left(1 + w_j \left(z_1 - 1\right) + \alpha_2 \, w_{j-1} \left(z_2 - 1\right)\right) \cdot E\left[y_1^{X_{t-h,1}} y_2^{X_{t-h,2}}\right.$$
$$\cdot \left(1 + w_h \left(z_1 - 1\right) + \alpha_2 \, w_{h-1} \left(z_2 - 1\right)\right)^{X_{t-h,1}}$$
$$\left(1 + w_{h-1} \left(z_1 - 1\right) + \alpha_2 \, w_{h-2} \left(z_2 - 1\right)\right)^{X_{t-h,2}}\right]$$
$$= \prod_{j=0}^{h-1} \mathrm{pgf}_\epsilon\left(1 + w_j \left(z_1 - 1\right) + \alpha_2 \, w_{j-1} \left(z_2 - 1\right)\right)$$
$$\cdot \mathrm{pgf}_X\left(y_1 \left(1 + w_h \left(z_1 - 1\right) + \alpha_2 \, w_{h-1} \left(z_2 - 1\right)\right),\right.$$
$$\left. y_2 \left(1 + w_{h-1} \left(z_1 - 1\right) + \alpha_2 \, w_{h-2} \left(z_2 - 1\right)\right)\right).$$

This implies that

$$\mathrm{pgf}_{X_t, X_{t-h}}(z, y) = \mathrm{pgf}_X\left(y \left(1 + w_h \left(z - 1\right)\right), \, 1 + w_{h-1} \left(z - 1\right)\right)$$
$$\prod_{j=0}^{h-1} \mathrm{pgf}_\epsilon\left(1 + w_j \left(z - 1\right)\right). \tag{B.7}$$

Now, we turn to the special case of the *Poisson* INAR(2) model (Example A.2.2), i.e., where the innovations $(\epsilon_t)_\mathbb{Z}$ satisfy $\mathrm{pgf}_\epsilon(z) = \exp\left(\lambda \left(z - 1\right)\right)$. Using that $\sum_{j=0}^{\infty} w_j = 1/(1 - \alpha_\bullet)$ and that $\mu = \lambda/(1 - \alpha_\bullet)$, (B.6) simplifies to

$$\mathrm{pgf}_X(z) = \exp\left(\mu \left((z_1 - 1) + \alpha_2 \left(z_2 - 1\right)\right)\right), \quad \mathrm{pgf}_X(z) = \exp\left(\mu \left(z - 1\right)\right), \tag{B.8}$$

also see (5.1) in Alzaid and Al-Osh (1990). In particular, the stationary marginal distribution is $\mathrm{Poi}(\mu)$. The bivariate pgf (B.7) becomes

$$\text{pgf}_{X_t, X_{t-h}}(z, y) = \exp\left(\mu\left(y - 1 + w_h\, y\,(z - 1) + \alpha_2\, w_{h-1}\,(z - 1)\right)\right)$$
$$\exp\left(\lambda\,(z - 1) \sum_{j=0}^{h-1} w_j\right).$$

This expression can be further simplified by considering (B.1). It follows that

$$(1 - \alpha_\bullet) \sum_{j=0}^{h-1} w_j = \sum_{j=0}^{h-1} w_j - \alpha_2\, w_{h-1} - \alpha_1\, w_0$$
$$- \sum_{j=2}^{h} (\alpha_1\, w_{j-1} + \alpha_2\, w_{j-2})$$
$$= 1 + \alpha_1 + \sum_{j=2}^{h-1} w_j - \alpha_2\, w_{h-1} - \alpha_1 - \sum_{j=2}^{h} w_j$$
$$= 1 - \alpha_2\, w_{h-1} - w_h,$$

so we continue

$$\text{pgf}_{X_t, X_{t-h}}(z, y) = \exp\left(\mu\left(y - 1 + z - 1 + w_h\,(y - 1)(z - 1)\right)\right)$$
$$= \exp\left(\mu\left((1 - w_h)\,(y - 1) + (1 - w_h)\,(z - 1) + w_h\,(yz - 1)\right)\right). \tag{B.9}$$

Obviously, this pgf is symmetric in z and y, confirming the time reversibility. For $h = 1$, it simplifies to (5.3) in Alzaid and Al-Osh (1990). In particular, (B.9) shows that (X_t, X_{t-h}) are bivariately Poisson distributed according to $\text{BPoi}\left(w_h\, \mu;\ (1 - w_h)\, \mu, (1 - w_h)\, \mu\right)$, see Johnson et al. (1997). This implies the conditional mean

$$E[X_t \mid X_{t-h}] = w_h\, X_{t-h} + (1 - w_h)\, \mu.$$

The proof of Example A.2.2 is completed by noting that the weights in (B.1) follow the same recursion as the ACF of the *Poisson* INAR(2) model, so $w_h = \rho(h)$ for $h \geq 0$ in this special case.

References

Al-Osh MA, Aly E-AAA (1992) First order autoregressive time series with negative binomial and geometric marginals. Commun Stat Theory Methods 21(9):2483–2492

Alzaid AA, Al-Osh MA (1988) First-order integer-valued autoregressive process: distributional and regression properties. Stat Neerl 42(1):53–61

Alzaid AA, Al-Osh MA (1990) An integer-valued pth-order autoregressive structure (INAR(p)) process. J Appl Probab 27(2):314–324

Cressie N, Read TRC (1984) Multinomial goodness-of-fit tests. J R Stat Soc Ser B 46(3):440–464

Du J-G, Li Y (1991) The integer-valued autoregressive (INAR(p)) model. J Time Ser Anal 12(2):129–142

Duchesne P, de Micheaux PL (2010) Computing the distribution of quadratic forms: further comparisons between the Liu–Tang–Zhang approximation and exact methods. Comput Stat Data Anal 54(4):858–862

Ferland R, Latour A, Oraichi D (2006) Integer-valued GARCH processes. J Time Ser Anal 27(6):923–942

Francq C, Zakoïan J-M (2013) Estimating the marginal law of a time series with applications to heavy-tailed distributions. J Bus Econ Stat 31(4):412–425

Freeland RK, McCabe BPM (2004) Forecasting discrete valued low count time series. Int J Forecast 20(3):427–434

Grunwald G, Hyndman RJ, Tedesco L, Tweedie RL (2000) Non-Gaussian conditional linear AR(1) models. Aust N Z J Stat 42(4):479–495

Horn SD (1977) Goodness-of-fit tests for discrete data: a review and an application to a health impairment scale. Biometrics 33(1):237–247

Ibragimov I (1962) Some limit theorems for stationary processes. Theory Probab Appl 7(4):349–382

Jacobs PA, Lewis PAW (1983) Stationary discrete autoregressive-moving average time series generated by mixtures. J Time Ser Anal 4(1):19–36

Johnson NL, Kotz S, Balakrishnan N (1997) Discrete multivariate distributions. Wiley, Hoboken

Jung RC, Tremayne AR (2011) Useful models for time series of counts or simply wrong ones? AStA Adv Stat Anal 95(1):59–91

Kim H-Y, Weiß CH (2015) Goodness-of-fit tests for binomial AR(1) processes. Statistics 49(2):291–315

Kißlinger A-L, Stummer W (2016) Robust statistical engineering by means of scaled Bregman distances. In: Agostinelli C et al (eds) Recent advances in robust statistics: theory and applications. Springer, New Delhi, pp 81–113

McKenzie E (1985) Some simple models for discrete variate time series. Water Resour Bull 21(4):645–650

McKenzie E (1986) Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. Adv Appl Probab 18(3):679–705

Meintanis S, Karlis D (2014) Validation tests for the innovation distribution in INAR time series models. Comput Stat 29(5):1221–1241

Moore DS (1982) The effect of dependence on chi squared tests of fit. Ann Stat 10(4):1163–1171

Neumann MH (2011) Absolute regularity and ergodicity of Poisson count processes. Bernoulli 17(4):1268–1284

Read TRC, Cressie N (1988) Goodness-of-fit statistics for discrete multivariate data. Springer, New York

Ristić MM, Bakouch HS, Nastić AS (2009) A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. J Stat Plan Inference 139(7):2218–2226

Schweer S (2015) On the time-reversibility of integer-valued autoregressive processes of general order. In: Steland A et al (eds) Stochastic models, statistics and their applications. Springer proceedings in mathematics & statistics, vol 122. Springer, Berlin, pp 169–177

Schweer S, Weiß CH (2014) Compound Poisson INAR(1) processes: stochastic properties and testing for overdispersion. Comput Stat Data Anal 77:267–284

Steutel FW, van Harn K (1979) Discrete analogues of self-decomposability and stability. Ann Probab 7(5):893–899

Tan WY (1977) On the distribution of quadratic forms in normal random variables. Can J Stat 5(2):241–250

Tavaré S, Altham PME (1983) Serial dependence of observations leading to contingency tables, and corrections to chi-squared statistics. Biometrika 70(1):139–144

Weiß CH (2008) Serial dependence and regression of Poisson INARMA models. J Stat Plan Inference 138(10):2975–2990

Weiß CH (2013) Serial dependence of NDARMA processes. Comput Stat Data Anal 68:213–238

Weiß CH (2018) An introduction to discrete-valued time series. Wiley, Chichester

Weiß CH, Pollett PK (2012) Chain binomial models and binomial autoregressive processes. Biometrics 68(3):815–824

Zucchini W, MacDonald IL (2009) Hidden Markov models for time series: an introduction using R. Chapman & Hall/CRC, London