Changepoint tests for INARCH time series

Jürgen Franke † * Claudia Kirch ‡
Joseph Tadjuidje Kamgaing †

† University of Kaiserslautern, ‡ Karlsruhe Institute of Technology, Germany

August 28, 2011

Abstract

In this paper, we discuss the problem of testing for a changepoint in the structure of an integer-valued time series. In particular, we consider a test statistic of cumulative sum (CUSUM) type for general Poisson autoregressions of order 1. We investigate the asymptotic behaviour of conditional least-squares estimates of the parameters in the presence of a changepoint. Then, we derive the asymptotic distribution of the test statistic under the hypothesis of no change, allowing for the calculation of critical values. We prove consistency of the test, i.e. asymptotic power 1, and consistency of the corresponding changepoint estimate. As an application, we have a look at changepoint detection in daily epileptic seizure counts from a clinical study.

Keywords: Integer-valued time series, Poisson autoregression, INGARCH, changepoint test, CUSUM statistic, conditional least-squares

This preprint is an earlier version of the extended paper *Changepoints in Times Series of Counts* submitted for publication. It contains more detailed versions of the proofs.

*Corresponding Author: Jürgen Franke, University of Kaiserslautern, Department of Mathematics, Erwin-Schroedinger-Str., 67663 Kaiserslautern, Germany. E-mail: franke@mathematik.uni-kl.de

1 Introduction

Time series of counts assume values in $\mathbb{N}_0$, the natural numbers including 0. They have a wide range of applications, e.g. in life sciences where they show up as counts of events, e.g. epileptic seizures (compare [1], [13]), in consecutive time intervals or as counts of individuals or objects over time, e.g. of animals in a given habitat (compare [15]) or of infections during an epidemic [8]. Sequences of counts which exhibit temporal dependence appear also in many other fields of applications like statistical quality control [30] or finance [27], [9], [19] and insurance [34]. For a review of the field we refer to the recent survey by Fokianos [12].

As in standard time series analysis, deviations from stationarity, in particular sudden changes in the data-generating mechanism, are of interest. There is some demand for tests which allow for deciding if the structure of the data is homogeneous in time, and, if this is not the case, to estimate the changepoint in time. For linear autoregressive processes with a continuous range several authors have derived related tests. Kulperger [24] as well as Horváth [16] use test statistics based on partial sums of residuals. Similary, Bai [2] makes use of partial sums of residuals for the change analysis in ARMA-models. Davis et al. [4] obtain a Gaussian-type likelihood ratio statistic which is related to the approach using weighted sums of residuals by Hušková , Prášková and Steinebach [18], both of which have power against a larger class of residuals than statistics based on partial sums of residuals.

Recently, Kirch and Tadjuidje-Kamgaing [23] have developped various CUSUM-type tests for changepoints in nonlinear, possibly misspecified autoregressions corresponding to neural network based autoregressive functions. For integer-valued time series, Weiß and Testik [31] have investigated CUSUM charts for detecting changes in integer-valued autoregressions (INAR) from a practical point of view. We consider another class of integer-valued time series, the Poisson autoregressions or INARCH-processes, for which Fokianos and Fried [10] have already discussed some kind of changepoint problems. Their approach, however, is targeted at another type of changes in the structure of the time series (compare the discussion preceding subsection 2.1).

Time series of counts have many features in common with time series with a continuous range, but in some regards, they show a different behaviour from a technical point of view due to the countability of the possible values. In contrast to preliminary expectations, the discreteness of the range makes it more difficult to prove some properties like, e.g., mixing, which are basic for dealing with the asymptotics of estimates and tests. Recently, there has been a breakthrough in that respect. Fokianos et al [9], Neumann [22] and Franke [14] showed various kinds of weak dependence properties for Poisson autoregressions. Similar results have been derived by Fokianos and Tjostheim [11] for log-linear Poisson autoregressions resp. by Triebsch [29] for integer-valued autoregressions.

In this paper, we use that theoretical basis to develop a CUSUM-type test for a changepoint and an estimate of the changepoint as well as the corresponding asymptotic theory. As a paradigm we consider a popular specific model for time series of counts, the Poisson autoregression. In the simplest case of order 1, it has the form

$$\mathcal{L}\{X_t|\mathcal{F}_{t-1}\} = \text{Poisson}(\lambda_t), \qquad \lambda_t = \omega + \alpha X_{t-1}, \tag{1.1}$$

i.e. the conditional distribution of the observation at time t given the past data is simply Poisson with a random parameter depending on the previous observation only. This process is also known in the literature as integer-valued ARCH process of order 1 or INARCH(1), as the conditional variance

$$\text{var}\{X_t|\mathcal{F}_{t-1}\} = \lambda_t = \omega + \alpha X_{t-1}$$

shares a similar dependence on the past with the classical ARCH(1) process of Engle [7]. The INARCH(1) model is a special case of the integer-valued GARCH or INGARCH process studied by Ferland et al. [8].

For a review of the properties of the Poisson autoregression model and related ones, we refer to Chapter 3 of Fokianos [12]. Here, we need that a choice of parameters satisfying $\omega > 0, 0 \leq \alpha < 1$ guarantees the existence of a stationary solution of (1.1). In that case, the time series of counts has the same autocorrelation structure as a linear Gaussian autoregression of order 1 with autoregressive parameter α, i.e. $\text{corr}(X_s, X_{t+s}) = \alpha^t, t \geq 0$. The process may be written in the form

$$X_t = \omega + \alpha X_{t-1} + \varepsilon_t, \qquad \text{with } \varepsilon_t = X_t - \lambda_t,$$

where the residuals ε_t are zero-mean white noise, i.e. a sequence of pairwise uncorrelated random variables, with var $\varepsilon_t = \text{E}X_t$. For $\alpha > 0$, the process in its stationary state exhibits overdispersion (compare Weiß [33]), i.e.

$$\text{E}X_t = \frac{\omega}{1-\alpha} < \frac{\omega}{(1-\alpha)(1-\alpha^2)} = \text{var}X_t. \tag{1.2}$$

This is a desirable feature as many time series of counts in practice show an overdispersion effect empirically.

In the following, we consider the more general class of nonlinear parametric Poisson autoregressions of order 1

$$\mathcal{L}\{X_t|\mathcal{F}_{t-1}\} = \text{Poisson}(\lambda_t), \qquad \lambda_t = g_\theta(X_{t-1}) \tag{1.3}$$

where g_θ is some known positive function on $\mathbb{N}_0$ depending on the parameter $\theta \in \Theta \subset \mathbb{R}^d$. Similarly, higher-order autoregressions could be dealt with, but we restrict our considerations to order 1 to keep the notation as simple as possible. As for the INARCH(1)-model one may write this process in the form $X_t = g_\theta(X_{t-1}) + \varepsilon_t$, where the residuals

$$\varepsilon_t = X_t - g_\theta(X_{t-1}) \tag{1.4}$$

are a zero-mean white-noise process. This is an important observation since the changepoint tests developed in this paper are based on estimated residuals.

To avoid a distinction of cases, we always assume that the autoregressive function g_θ involves a constant term as the first parameter, i.e.

$$g_\theta(x) = \theta_1 + g_{\theta_2,\ldots,\theta_d}(x). \tag{1.5}$$

This is a question of parametrization and does not restrict the scope of the model. Later on, we consider conditional least-squares (CLS) estimates for the model parameter. There, we need the following technical condition, which is typical for the asymptotics of M-estimates:

(A1) The parameter set Θ is compact.

In case of model (1.1), we just have to constrain the two parameters by assuming $\delta \le \omega \le \Delta, 0 \le \alpha \le 1-\delta$ for some arbitrary small $\delta > 0$ and some arbitrary large $\Delta < \infty$.

In the next chapter, we formally introduce the changepoint problem. As a first step towards a corresponding test, we derive the asymptotics, i.e. consistency and asymptotic normality, for the CLS estimate of the model parameter θ. Then, we derive the asymptotic distribution of the changepoint test statistic under the hypothesis, which allows to construct the test, followed by some asymptotics under the alternative which shows that the test is consistent in the sense that its asymptotic power is 1. As a by-product, we get consistency of an estimate of the changepoint.

In chapter 3, we apply the changepoint procedure to some artificially generated data as well as to a real data example from pharmaceutical research where the observations are daily epileptical seizure counts. All the proofs are postponed to the appendix.

2 Testing for a changepoint

We are interested in a change of the model parameters during the period where data are collected, and we consider the following formulation of this changepoint detection problem: Given two independent stationary time series $\{Y_t\}, \{Y_t^*\}$ both following model (1.3), but with different parameters $\theta = \theta_0$ and $\theta = \theta_0^*$ resp. We observe $X_0, \ldots, X_N$, where

$$X_t = \begin{cases} Y_t & t \le k^* \\ Y_t^* & t > k^* \end{cases} \tag{2.1}$$

for some $1 \le k^* \le N$. If $k^* = N$, we have $X_t = Y_t, t = 0, \ldots, N$, and there is no change in the observed time series of counts X_t. For $k^* < N$, we have a change of parameters at time k^*. We look for a test solving the following decision problem:

$$\mathrm{H}_0 : k^* = N \quad \text{against} \quad \mathrm{H}_1 : k^* < N \tag{2.2}$$

For asymptotic considerations, we need a growing number of observations before and after the changepoint if there is one. Therefore, we assume

(A2) $k^* = [\gamma N]$ for some $0 < \gamma < 1$, if H_1 holds.

We consider a test based on a statistic of cumulative sum (CUSUM) type. First, we have to estimate the parameter θ to get the sample residuals $\widehat{\varepsilon}_t$. As a parameter estimate we consider the conditional least-squares (CLS) estimate $\widehat{\theta}_N$ which minimizes

$$Q_N(\theta) = \sum_{t=1}^{N} \big(X_t - g_\theta(X_{t-1})\big)^2 = \sum_{t=1}^{N} q_t(\theta)$$

over $\theta \in \Theta$. For the linear Poisson autoregression (1.1) we have by a straightforward calculation (compare also [32]),

$$\widehat{\theta}_{N,2} = \widehat{\alpha}_N = \widehat{\rho}_N, \quad \widehat{\theta}_{N,1} = \widehat{\omega}_N = (1 - \widehat{\alpha}_N)\overline{X}_N \tag{2.3}$$

where $\overline{X}_N$, $\widehat{\rho}_N$ denote the sample mean and the lag 1 sample autocorrelation of the data. From the parameter estimates, we get the sample residuals and their cumulative sums

$$\widehat{\varepsilon}_t = X_t - g_{\widehat{\theta}_N}(X_{t-1}), \;\; t = 1, \ldots, N, \qquad \widehat{S}_N(k) = \sum_{t=1}^{k} \widehat{\varepsilon}_t, \;\; k = 1, \ldots, N. \tag{2.4}$$

As test statistic we use the maximum of the normalized cumulative sums

$$T_N = \max_{1 \le k < N} \sqrt{\frac{N}{k(N-k)}}\, \big|\widehat{S}_N(k)\big|.$$

Such statistics are common tools for changepoint detection. Originally, they were derived as the maximum likelihood statistic in a sequence of independent normal random variables with constant variance and a mean that changes from μ_0 to μ_0^* at some unknown time. This statistic converges almost surely to infinity (as can be seen by the law of iterated logarithm) but using appropriate normalizing factors one obtains a distributional convergence to the Gumbel extreme value distribution (cf. Theorem 2.3). Because the convergence is slow and many practitioners are not used to this kind of convergence in hypothesis testing, the following related statistics are often considered:

$$T_{N,2} = \max_{\tau N \le k \le N - \tau N} \sqrt{\frac{N}{k(N-k)}}\, \big|\widehat{S}_N(k)\big|,$$

$$T_{N,3} = \max_{1 \le k < N} \left(\frac{N^2}{k(N-k)}\right)^{\beta} \frac{1}{\sqrt{N}}\, \big|\widehat{S}_N(k)\big|$$

for $\tau > 0$ and $0 \leq \beta < 1/2$. These two statistics converge in distribution to $\sup_{\tau<t<1-\tau} |B(t)|$ resp. $\sup_{0<t<1} |B(t)|/(t(1-t))^{\beta}$. Analogous results for these statistics can be obtained along the lines of the proofs for T_N. For details and further references we refer to the monograph by Csörgő and Horváth [3].

Fokianos and Fried [10] recently have also considered changepoint detection for linear INGARCH processes. Beyond the type of processes considered, i.e. nonlinear INARCH(1) compared to linear INGARCH(p,q), their approach differs from ours as they are interested in deterministic changes of the conditional means. In terms of model (1.1), they consider the following situation:

$$\mathrm{E}\{X_t|\mathcal{F}_{t-1}\} = \lambda_t = g_\theta(X_{t-1}) + \nu z_t$$

where the deterministic series z_t may represent various types of intervention effects and the hypothesis of no change corresponds to $\nu = 0$. In case of a level shift where $z_t = 0, t \leq k^*$, and $z_t = 1, t > k^*$, this intervention model and our approach (2.1) with a change solely in the constant parameter θ_1 of (1.5) coincide up to asymptotically negligible differences. However, the model of Fokianos and Fried allows for a larger variety of dependencies of the level on time, e.g. transient level shifts and additive outliers, whereas our model allows for changes in the way how λ_t depends on past data. The latter is of interest in certain practical problems. Franke and Seligmann [13] discuss, e.g., the effectiveness of a drug on epileptic seizure counts. Epileptic seizures themselves stress the brain and make subsequent seizures more likely. Therefore, a positive correlation between the counts X_t, X_{t-1} is plausible and can be verified with the data. The drug was, among other effects, meant to improve recovery of the brain and, therefore, to reduce this kind of dependence. In the simple model (1.1) this corresponds to a change in the autoregressive parameter α.

2.1 Asymptotics of the CLS estimate

Weiß [32] studied the asymptotics of the CLS estimates for the linear Poisson autoregression (1.1), proving, e.g., consistency and asymptotic normality. We generalize those results to the nonlinear model (1.3) and extend them to the situation where a changepoint is present in the sample. Observe, that under the alternative H_1, the estimate $\widehat{\theta}_N$ will converge neither to θ_0 nor to θ_0^* but to some $\tilde{\theta}_0$ which is defined in the following manner. First, we modify the summands $q_t(\theta) = (X_t - g_\theta(X_{t-1}))^2$ of $Q_N(\theta)$ slightly by introducing

$$\tilde{q}_t(\theta) = \begin{cases} \big(Y_t - g_\theta(Y_{t-1})\big)^2 & t \leq [\gamma N] \\ \big(Y_t^* - g_\theta(Y_{t-1}^*)\big)^2 & t > [\gamma N], \end{cases}$$

and we set $\widetilde{Q}_N(\theta) = \sum_{t=1}^N \tilde{q}_t(\theta)$. We have $q_t(\theta) = \tilde{q}_t(\theta)$ for all t under the hypothesis H_0 as, then, $X_t = Y_t$ for all t. Otherwise, $q_t(\theta), \tilde{q}_t(\theta)$ differ only at the changepoint such that the difference between $Q_N(\theta)$ and $\widetilde{Q}_N(\theta)$ becomes negligible for increasing N. Let

$$e(\theta) = \gamma \, \mathrm{E}\tilde{q}_1(\theta) + (1-\gamma)\mathrm{E}\tilde{q}_N(\theta).$$

We have to assume existence and uniqueness of the minimizer $\tilde{\theta}_0$ of $e(\theta)$:

(A3) There is a unique $\tilde{\theta}_0 \in \Theta$ such that $e(\tilde{\theta}_0) = \min_{\theta\in\Theta} e(\theta)$, and $\tilde{\theta}_0$ lies in the interior of Θ.

To show convergence of $\widehat{\theta}_N$ to $\tilde{\theta}_0$, we need a contraction property of the autoregressive function $g_\theta(x)$. Among other features, it guarantees the existence of a stationary and ergodic solution of (1.3) by Theorem 3.1 of Neumann [22].

(A4) For all $\theta \in \Theta$, $g_\theta(x)$ is uniformly Lipschitz in x with constant $L_\theta < 1$:

$$\big|g_\theta(x) - g_\theta(x')\big| \le L_\theta|x - x'| \qquad \text{for all } x, x' \in \mathbb{N}_0.$$

For the simple model (1.1), (A4) is equivalent to assuming $\alpha < 1$. Additionally, we need some regularity of $g_\theta(x)$ as a function of the parameter. ∇ and ∇^2 denote the gradient and the Hessian w.r.t. to θ.

(A5) $g_\theta(x)$ is twice continuously differentiable w.r.t. θ for all $x \in \mathbb{N}_0$.

(A6) If $\{Y_t\}$ is a stationary solution of (1.3) with $\theta = \theta_0$ or $\theta = \theta_0^*$,

$$\mathrm{E}\sup_{\theta\in\Theta} g_\theta^2(Y_t) < \infty, \quad \mathrm{E}\sup_{\theta\in\Theta} ||\nabla g_\theta(Y_t)\nabla^T g_\theta(Y_t)|| < \infty,$$
$$\mathrm{E}\big(Y_t \sup_{\theta\in\Theta} ||\nabla^2 g_\theta(Y_{t-1})||\big) < \infty.$$

For linear Poisson autoregressions (1.1), (A5) is obvious, and (A6) is automatically satisfied if $\alpha < 1$.

Theorem 2.1. *Assume that $X_0, \ldots, X_N$ are generated by model (2.1) where $\{Y_t\}, \{Y_t^*\}$ are independent stationary Poisson autoregressions (1.3). Moreover, let (A1)-(A6) be satisfied. Then, the conditional least-squares estimate $\widehat{\theta}_N$ is strongly consistent for $\tilde{\theta}_0$:*

$$\widehat{\theta}_N \to \tilde{\theta}_0 \qquad a.s. \text{ for } N \to \infty.$$

If we additionally assume

(A7) The Hessian $\nabla^2 e(\tilde{\theta}_0)$ is positive definite,

(A8) $\mathrm{E}\big(Y_t - g_{\tilde{\theta}_0}(Y_{t-1})\big)^{4+\nu} < \infty,\ \mathrm{E}\big(Y_t^* - g_{\tilde{\theta}_0}(Y_{t-1}^*)\big)^{4+\nu} < \infty$ for some $\nu > 0$, we also have asymptotic normality of $\widehat{\theta}_N$:

Theorem 2.2. *If, additionally to the assumptions of Theorem 2.1, (A7), (A8) are satisfied, we have*

$$\sqrt{N}(\widehat{\theta}_N - \tilde{\theta}_0) \underset{d}{\longrightarrow} \mathcal{N}(0, A^{-1}BA^{-1}) \qquad \text{for } N \to \infty$$

with $A = \nabla^2 e(\tilde{\theta}_0)$ *and, with* $\nabla \widetilde{Q}_N(\theta)$ *denoting the gradient of* $\widetilde{Q}_N(\theta)$ *w.r.t.* θ*,*

$$B = \lim_{N\to\infty} \frac{1}{N} \mathrm{E}\big(\nabla \widetilde{Q}_N(\tilde{\theta}_0) \nabla^T \widetilde{Q}_N(\tilde{\theta}_0)\big).$$

Mark that condition (A7) is automatically satisfied for the linear Poisson autoregression (1.1) as, then,

$$A = 2\gamma \begin{pmatrix} 1 & \mathrm{E}Y_0 \\ \mathrm{E}Y_0 & \mathrm{E}(Y_0)^2 \end{pmatrix} + 2(1-\gamma) \begin{pmatrix} 1 & \mathrm{E}Y^*_{N-1} \\ \mathrm{E}Y^*_{N-1} & \mathrm{E}(Y^*_{N-1})^2 \end{pmatrix},$$

by a straightforward calculation, and $\det(A) = 2\gamma \mathrm{var} Y_0 + 2(1-\gamma)\mathrm{var} Y^*_{N-1} > 0$. Moreover, (A8) is also satisfied provided the autoregressive parameters are less than 1 as, then, all moments of Y_t, Y_t^* exist by Proposition 6 of Ferland et al. [8].

In case of a linear Poisson autoregression (1.1) with parameter vector $\theta_0 = (\omega, \alpha)$ and no change, one can easily check that $\tilde{\theta}_0 = \theta_0$ and hence Theorems 2.1 and 2.2 reprove the consistency and asymptotic normality of the CLS estimates already obtained in Weiß [32].

2.2 The test statistics under the hypothesis

In this section, we derive the asymptotic distribution of the test statistic T_N under the hypothesis. This allows to derive critical values for a test of the null hypothesis of no change which asymptotically has the prescribed level. We need a further condition:

(A9) The $(2+\nu)$th central moment of $\nabla g_{\theta_0}(Y_t)$ exists for some $\nu > 0$.

This condition is satisfied for the model (1.1) with $\alpha < 1$ as $\nabla g_{\theta_0}(y) = (1, y)^T$ and all moments of Y_t exist by Proposition 6 of Ferland et al. [8].

Using the abbreviations

$$a(u) = \sqrt{2 \log u}, \qquad b(u) = 2\log u + \frac{1}{2}\log\log u - \frac{1}{2}\log \pi,$$

we have

Theorem 2.3. *Let the assumptions of Theorem 2.2 and (A9) be satisfied. If the hypothesis* H_0 *holds, i.e.* $X_t = Y_t, t = 0, \ldots, N$, *we have*

$$\mathrm{pr}\Big(\; a(\log N) \, \frac{T_N}{\tau} - b(\log N) \leq x \; \Big) \to \exp(-2e^{-x}) \quad \textit{for} \;\; N \to \infty$$

This result involves the unkown variance τ^2 of the residuals ε_t. The following corollary shows that we may replace it by the sample variance of the sample residuals. Observe that, by assumption (1.5) and by the definition of the CLS estimates, the sum of $\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_N$ and therefore their sample mean is 0. For the details, compare the proof of Lemma 4.1.

Corollary 2.1. *Under the assumptions of Theorem 2.3*

$$\hat{\tau}^2 = \frac{1}{N-d} \sum_{t=1}^{N} \widehat{\varepsilon}_t^{\,2} = \tau^2 + \mathrm{O}_p\Big(\frac{1}{\sqrt{N}}\Big).$$

2.3 The test statistics under the alternative

In this section, we have a look at the asymptotic behaviour of the test statistic under the alternative which implies consistency of the test, i.e. asymptotic power 1. We need

(A10) $\qquad |\mathrm{E}Y_t - \mathrm{E}g_{\tilde{\theta}_0}(Y_{t-1})| = |\mathrm{E}g_{\theta_0}(Y_{t-1}) - \mathrm{E}g_{\tilde{\theta}_0}(Y_{t-1})| = C > 0.$

From (A3), we have $\nabla e(\tilde{\theta}_0) = 0$ for large enough N, and, looking only at the derivative w.r.t. θ_1, we get in view of (1.5)

$$|\mathrm{E}Y_t^* - \mathrm{E}g_{\tilde{\theta}_0}(Y_{t-1}^*)| = \frac{\gamma}{1-\gamma} |\mathrm{E}Y_t - \mathrm{E}g_{\tilde{\theta}_0}(Y_{t-1})|. \tag{2.5}$$

Therefore, (A10) can be equivalently expressed as $|\mathrm{E}Y_t^* - \mathrm{E}g_{\tilde{\theta}_0}(Y_{t-1}^*)| > 0$. (A10) is kind of an identifiability condition, excluding $\tilde{\theta}_0 = \tilde{\theta}_0^*$ and some other degenerate situations.

In case of a linear Poisson autoregression (1.1) with parameter $\theta_0 = (\omega, \alpha)$ before the change and $\theta_0^* = (\omega^*, \alpha^*)$ after the change, (A10) simplifies considerably. Using $g_{\tilde{\theta}_0}(y) = \tilde{\omega}_0 + \tilde{\alpha}_0 y$ and (2.5), a straightforward calculation shows that (A10) is equivalent to

$$\mathrm{E}Y_1 = \frac{\omega}{1-\alpha} \neq \frac{\omega^*}{1-\alpha^*} = \mathrm{E}Y_1^*,$$

i.e. the unconditional mean changes at k^*. This is a typical condition needed if test statistics are based on estimated residuals and even occur in a simple linear regression situation with continuous errors (cf. Hušková and Koubková [17]).

Theorem 2.4. *Let the assumptions of Theorem 2.3 and (A10) be satisfied. If the alternative* H_1 *holds, we have for all* $c > 0$

$$\mathrm{pr}\Big(\; a(\log N)\, \frac{T_N}{\tau} - b(\log N) \geq c \; \Big) \to 1 \quad \textit{for} \quad N \to \infty$$

As a by-product, we get a consistent estimate of the changepoint if there is one, i.e. if the alternative holds. It is the point in time where the maximum of the absolute cumulative sums is attained.

Corollary 2.2. *If the assumptions of Theorem 2.4 hold, we have under the alternative* H_1

$$\frac{\hat{k}^*}{N} \underset{p}{\longrightarrow} \gamma \qquad \textit{where} \quad \hat{k}^* = \arg \max_{1 \leq k < N} |\widehat{S}_N(k)|, \quad \textit{i.e.} \;\; |\widehat{S}_N(\hat{k}^*)| = \max_{1 \leq k < N} |\widehat{S}_N(k)|.$$

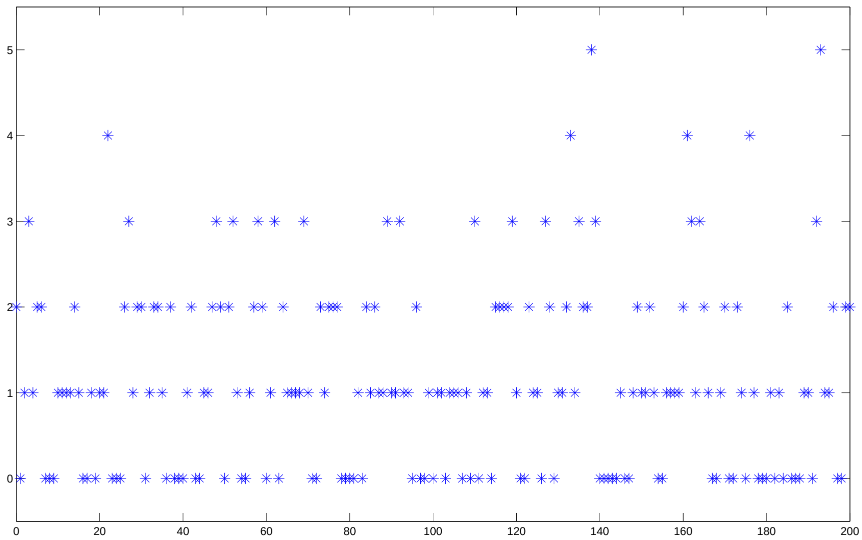

Figure 1: Poisson autoregression with parameters $\omega = 1, \alpha = 0.15$.

3 Numerical examples

In this section, we look at the practical performance of the changepoint test for some artificial and real data. The parameters of the simulated data are chosen such that they roughly match the two subsequent real data examples regarding visual appearance of the time series and sample size. In all cases, the changepoint test procedure based on Theorem 2.3 works quite well even though the sample size was around 200 only and not excessively large. Using Corollary 2.1, we replaced the standard deviation τ of the residuals, showing up in the asymptotic distribution of the test, by the sample standard deviation $\hat{\tau}$ of the sample residuals $\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_N$.

3.1 Artificially generated data

First, we generate some data $X_0, \ldots, X_N, N = 200$, from the linear Poisson autoregression (1.1) with parameters $\omega = 1, \alpha = 0.15$. The data are shown in Figure 1. Then, we calculate the CLS estimate of the parameters from (2.3) resulting in $\hat{\omega} = 0.945, \hat{\alpha} = 0.178$. Next, we calculate the residuals $\widehat{\varepsilon}_t$ and the standardized cumulative sums $T_N(k) = \sqrt{\frac{N}{k(N-k)}}|\widehat{S}_N(k)|, 1 \leq k < N$. Finally, for some level $0 < \phi < 1$, we calculate

$$C = \hat{\tau}\,\frac{\log 2 - \log(\log \frac{1}{1-\phi}) + b(\log N)}{a(\log N)},$$

such that, by Theorem 2.3, the changepoint test rejects the hypothesis H_0 asymptotically on the level ϕ iff $T_N = \max_{1 \leq k < N} T_N(k) < C$.

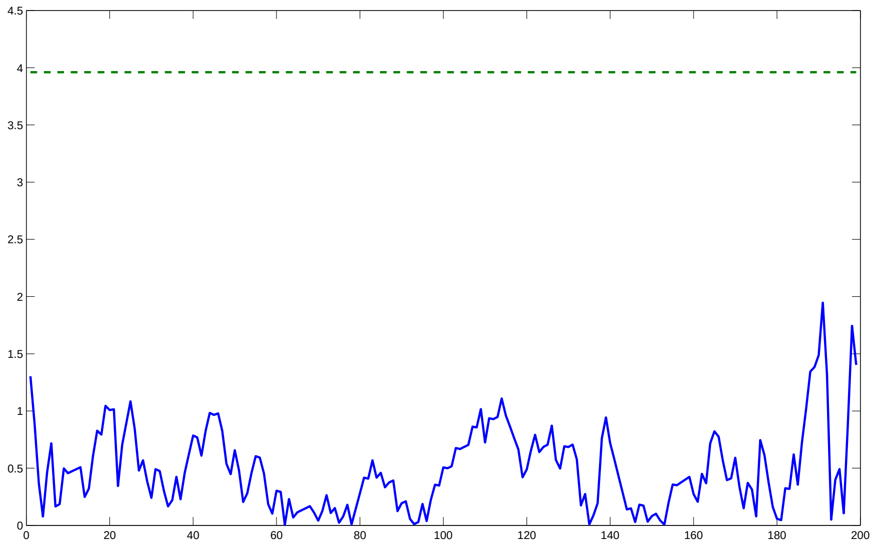

Figure 2: Standardized cumulative sums (hypothesis)

Figure 2 shows the standardized cumulative sums $T_N(k)$ as the solid line and the bound C of significance (for $\phi = 0.05$) as the dashed horizontal line. As $T_N(k)$ stays well below the bound C, the hypothesis is accepted.

The second data set $X_0, \ldots, X_N, N = 200$, shown in Figure 3, has a changepoint at time $k^* = 100$. It consists of two linear Poisson autoregressions with parameters $\omega = 1, \alpha = 0.5$ before and $\omega = 0.3, \alpha = 0.15$ after the changepoint. Figure 4 shows the corresponding standardized cumulative sums with the vertical line marking the changepoint. A changepoint is detected by the level $\phi = 0.05$ test, and the estimated changepoint is $\hat{k}^* = 99$, compare Corollary 2.2.

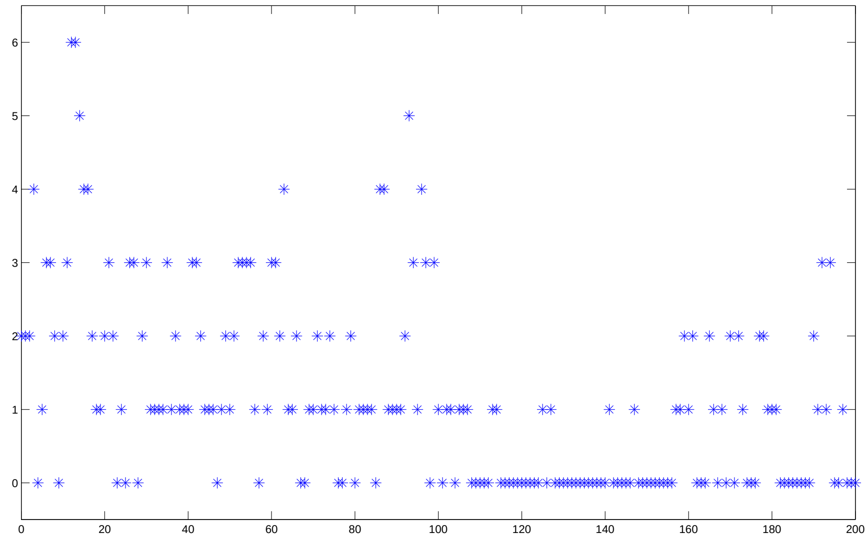

Figure 3: Poisson autoregression with change of parameters at $t = 100$.

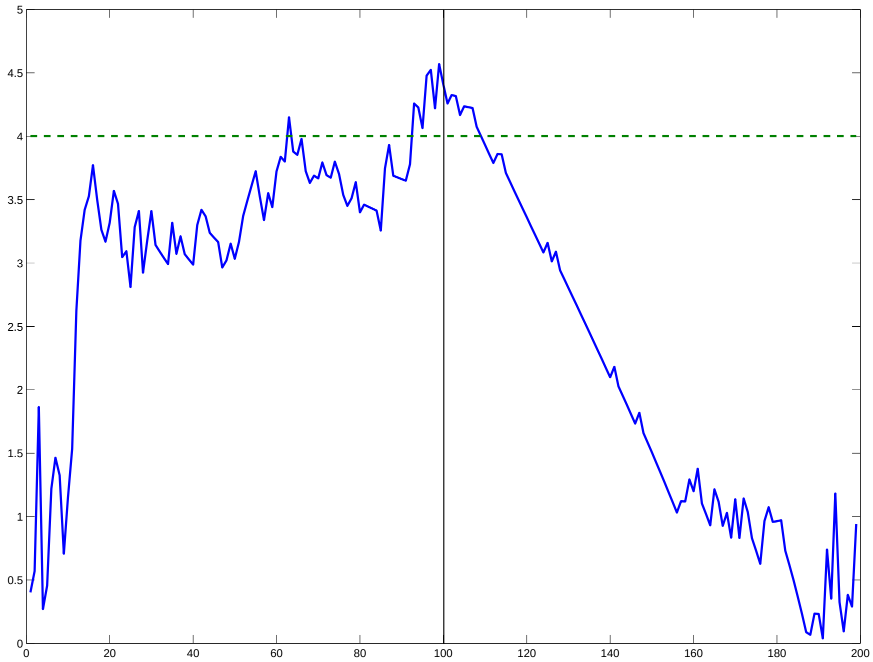

Figure 4: Standardized cumulative sums (alternative)

3.2 Epileptic seizure counts

Daily counts of epileptic seizures are an important source of information in evaluating the effectiveness of various treatments and, in particular, of the administration of drugs against epilepsy. Such data $\{X_t\}$ show correlation in time as well as a significant amount of overdispersion, i.e. $\mathrm{E}X_t < \mathrm{var}X_t$. To model those effects, Albert [1] considered a hidden Markov chain model controlled by a Markov chain with two states and data, which are independent Poisson distributed conditional on the hidden state variable. Franke and Seligmann [13] started from integer-valued autoregressions of order 1 and introduced overdispersion by considering innovations with a Poisson mixture distribution. Here, we use the quite parsimonous linear Poisson autoregression of order 1 which also shows the desired overdispersion by (1.2). Mark that this model requires only two parameters, whereas the model by Albert has 4 parameters (2 for the hidden Markov chain and 2 Poisson parameters) and the model by Franke and Seligmann has 3 parameters (2 Poisson parameters for the mixture distribution and the autoregressive parameter).

The data are part of a larger clinical study where the effect of a drug on partial seizures has been investigated. More details about the background of the data are given by Franke and Seligmann [13]. Here, we are only interested in the applicability of the above changepoint detection method to such data. During the study, the number of epileptic seizures was first recorded for some time under the usual medication. Then, from a certain day onwards, the new drug was additionally administered. In a second control group, a placebo was handed out to the patients instead. The test has to decide if there is a change in the structure of the time series of counts. Mark that this is not a classical straightforward changepoint detection problem as we expect the change, if at all, to happen shortly after the change of medication, but we do not exploit this information for decision making. The following study only serves as an illustration for the performance of the changepoint test in a realistic setting.

Figures 5 and 7 show the daily seizure counts for a patient from the control group resp. the test group. Figures 6 and 8 display the corresponding standardized cumulative sums together with the bound of significance. The vertical lines mark here the date where the change of medication took place. For the data from the placebo group, the test statistic, i.e. the maximum of the curve in Figure 6, is too small to raise some doubt and much less to reject the hypothesis that the administration of placebo has no effect on the time series of counts. Figure 8, on the other hand, shows a highly significant changepoint at $\hat{k}^* = 81$, i.e. at the day after the first administration of the drug. Mark, that not all time series of seizure counts from the test group showed a significant changepoint; for some patients, epileptic seizures were very rare from the beginning, so it was not possible to show a further reduction due to the drug. For other patients in the test group, the estimated changepoint was significant, but not close to the change of medication; the structure of the time series switched to the new state considerably later, perhaps an indication for a certain retardation of the drug effect at some patients.

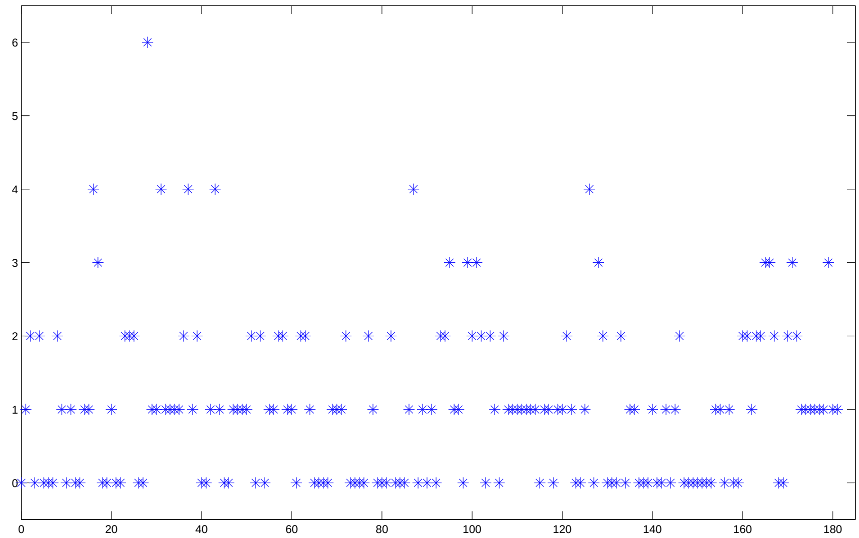

Figure 5: Epileptic seizure counts (placebo group), change of medication at $t = 83$.

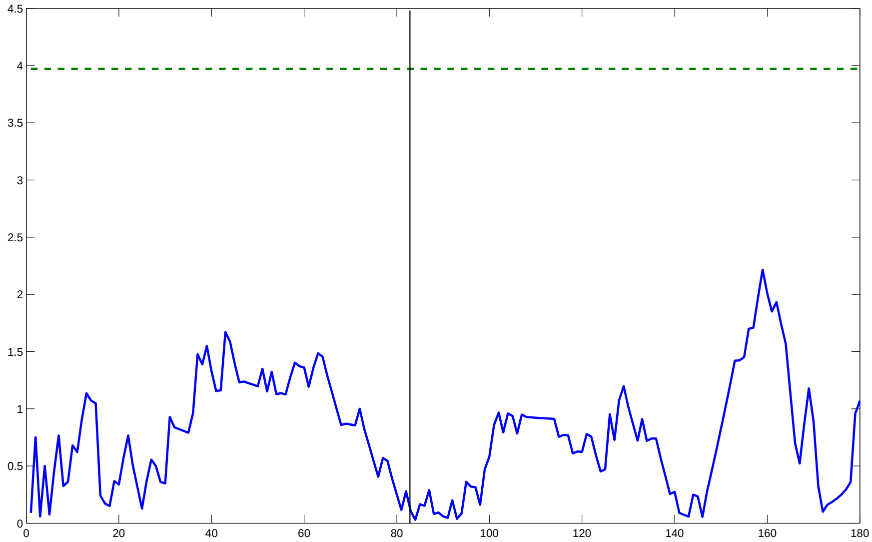

Figure 6: Standardized cumulative sums (placebo group)

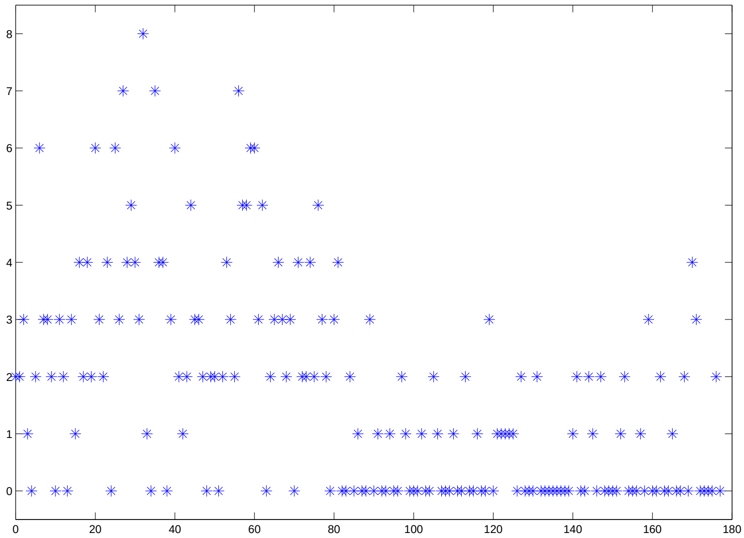

Figure 7: Epileptic seizure counts (test group), change of medication at $t = 80$.

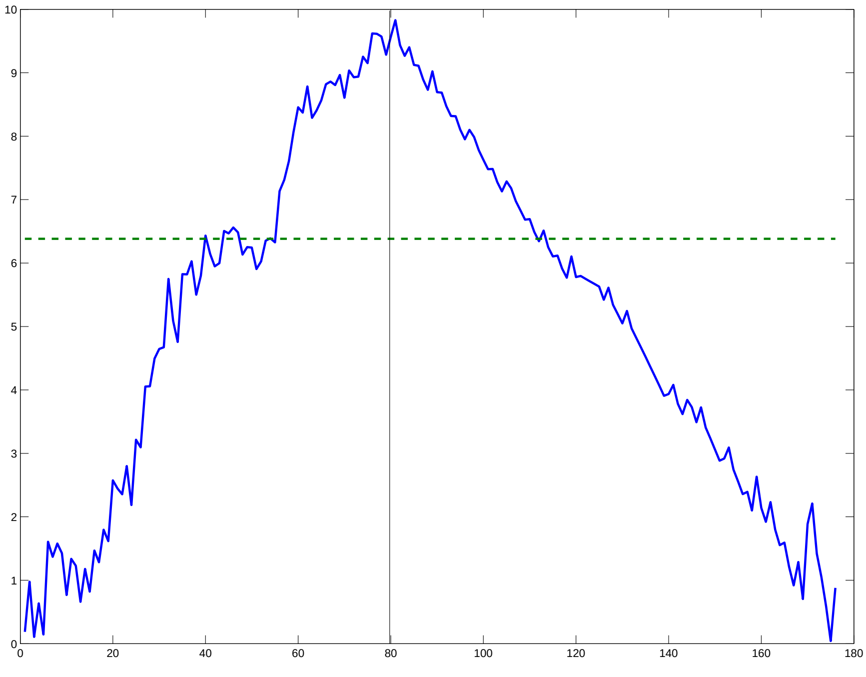

Figure 8: Standardized cumulative sums (test group)

4 Appendix

Before starting with the proofs, let us summarize some properties of nonlinear Poisson autoregressions (1.3) in the following corollary which follows immediately from Theorems 2.1 and 3.1 of Neumann [22].

Corollary 4.1. *Assume (A4). Then, there exists a stationary ergodic time series $\{X_t\}$ of counts satisfy (1.3) which is β-mixing (also called absolutely regular, compare Doukhan [6]) with exponential rate. Moreover,* $\mathrm{E}\lambda_t^2 = \mathrm{E}g_\theta^2(X_{t-1}) < \infty$.

From the last statement, we also have

$$\mathrm{E}X_t^2 = \mathrm{E}\mathrm{E}\{X_t^2|\mathcal{F}_{t-1}\} = \mathrm{E}(\lambda_t + \lambda_t^2) < \infty \text{ and } \mathrm{E}\ \varepsilon_t^2 = \mathrm{E}(X_t - \lambda_t)^2 < \infty. \tag{4.1}$$

4.1 Proofs of Section 2.1

As usual for consistency proofs, we first show uniform convergence of the empirical loss.

Proposition 4.1. *Let (A1), (A2), (A4)-(A6) be satisfied. Then, for $N \to \infty$,*

$$a)\ \ \sup_{\theta\in\Theta} \Big|\frac{1}{N}\widetilde{Q}_N(\theta) - e(\theta)\Big| \to 0 \quad a.s.$$

$$b)\ \ \sup_{\theta\in\Theta} \Big|\Big|\frac{1}{N}\nabla^2\widetilde{Q}_N(\theta) - \nabla^2 e(\theta)\Big|\Big| \to 0 \quad a.s.$$

where ∇^2 denotes the Hessian w.r.t. θ.

Proof: Let us first consider the situation where the hypothesis holds, i.e. where $X_t = Y_t,\ 0 \le t \le N$. By Corollary 4.1, $\{Y_t\}$ is stationary and ergodic, and, hence, $\tilde{q}_t(\theta) = (Y_t - g_\theta(Y_{t-1}))^2, \nabla^2\tilde{q}_t(\theta)$ are (pointwise for each $\theta \in \Theta$) stationary and ergodic. From assumption (A6), we get, using that $Y_t \in \mathbb{N}_0$, i.e. $Y_t = 0$ or $Y_t \ge 1$,

$$\mathrm{E}\sup_{\theta\in\Theta}\tilde{q}_t(\theta) < \infty\,, \quad \mathrm{E}\sup_{\theta\in\Theta}\|\nabla^2\tilde{q}_t(\theta)\| < \infty.$$

Now, the assertion follows from the uniform law of large numbers of Ranga Rao (Theorem 6 of [26]) for stationary and ergodic processes.
Under the alternative, that there is a changepoint $k^* = [\gamma N]$, we have to apply the same kind of argument before and after the changepoint. E.g., we have

$$\frac{1}{N}\widetilde{Q}_N(\theta) = \frac{[\gamma N]}{N}\frac{1}{k^*}\sum_{t=1}^{k^*}(Y_t - g_\theta(Y_{t-1}))^2 + \frac{N-[\gamma N]}{N}\frac{1}{N-k^*}\sum_{t=k^*+1}^{N}(Y_t^* - g_\theta(Y_{t-1}^*))^2$$

$$\to \gamma\mathrm{E}(Y_t - g_\theta(Y_{t-1}))^2 + (1-\gamma)\ \mathrm{E}(Y_t^* - g_\theta(Y_{t-1}^*))^2 = e(\theta)$$

uniformly in $\theta \in \Theta$. ■

Proof of Theorem 2.1:
The result follows immediately from Proposition 4.1 a) and the identifiability condition (A3) by applying Lemma 3.1 of Pötscher and Prucha [25]. ■

Proof of Theorem 2.2:
i) First, we consider the case where the hypothesis $X_t = Y_t,\ 0 \le t \le N$, holds. By the definition of $\tilde\theta_0$ and assumptions (A3)-(A6), we have

$$0 = \nabla e(\tilde\theta_0) = \nabla \mathrm{E}\tilde q_t(\tilde\theta_0) = \mathrm{E}\nabla\tilde q_t(\tilde\theta_0)$$

using dominated convergence and $\mathrm{E}\sup_{\theta\in\Theta}\|\nabla\tilde q_t(\theta)\| < \infty$ which follows from (A6), (4.1) and the Cauchy-Schwarz inequality.

From Corollary 4.1 and Proposition 1.1.1 of Doukhan [6], $\{Y_t\}$ is also α-mixing with exponential rate. As

$$\nabla\tilde q_t(\tilde\theta_0) = -2(Y_t - g_{\tilde\theta_0}(Y_{t-1}))\ \nabla g_{\tilde\theta_0}(Y_{t-1}) = Z_t$$

is a measurable function of (Y_t, Y_{t-1}), it has the same kind of mixing condition. Therefore, together with (A8), we may apply the strong invariance principle for α-mixing time series of Kuelbs and Philipp (Theorem 4 of [20]), compare also Dehling and Philipp [5], to get that

$$B_{ij} = \sum_{k=-\infty}^{\infty} \mathrm{cov}(Z_{ti}, Z_{t+k,j})$$

converges absolutely for all $1 \le i,j \le d$, and, with W(s) denoting a d-variate Brownian motion with covariance matrix $B = (B_{ij})_{1\le i,j\le d}$ on an appropriate probability space

$$\frac{1}{\sqrt{s}}\left\{\sum_{t\le s} Z_t - W(s)\right\} = O\left(\frac{1}{s^r}\right) \quad \text{a.s. for some } r > 0. \tag{4.2}$$

In particular we have

$$\frac{1}{\sqrt{N}}\nabla\widetilde{Q}_N(\tilde\theta_0) = \frac{1}{\sqrt{N}}\sum_{t=1}^{N} Z_t \underset{d}{\longrightarrow} \mathcal{N}(0,B) \tag{4.3}$$

and

$$B = \lim_{N\to\infty}\frac{1}{N}\sum_{s,t=1}^{N}\mathrm{E}Z_tZ_s^T = \lim_{N\to\infty}\frac{1}{N}\mathrm{E}\,\nabla\widetilde{Q}_N(\tilde\theta_0)\nabla^T\widetilde{Q}_N(\tilde\theta_0).$$

ii) If the alternative holds, i.e. $X_t = Y_t,\ t \le k^*$ and $X_t = Y_t^*, t > k^*$, we have to apply the same kind of argument separately before and after the changepoint. We still have $\nabla \tilde{q}_t(\tilde{\theta}_0) = Z_t,\ t \le k^*$, but

$$\nabla \tilde{q}_t(\tilde{\theta}_0) = -2(Y_t^* - g_{\tilde{\theta}_0}(Y_{t-1}^*))\ \nabla g_{\tilde{\theta}_0}(Y_{t-1}^*) = Z_t^*,\ t > k^*,$$

such that

$$\begin{aligned}
\frac{1}{\sqrt{N}} \nabla \widetilde{Q}_N(\tilde{\theta}_0) &= \sqrt{\frac{k^*}{N}} \frac{1}{\sqrt{k^*}} \sum_{t=1}^{k^*} Z_t + \sqrt{\frac{N-k^*}{N}} \frac{1}{\sqrt{N-k^*}} \sum_{t=k^*+1}^{N} Z_t^* \\
&= \sqrt{\frac{k^*}{N}} \frac{1}{\sqrt{k^*}} \sum_{t=1}^{k^*} (Z_t - \mathrm{E} Z_t) + \sqrt{\frac{N-k^*}{N}} \frac{1}{\sqrt{N-k^*}} \sum_{t=k^*+1}^{N} (Z_t^* - \mathrm{E} Z_t^*) + R_N
\end{aligned}$$

where

$$\begin{aligned}
R_N &= \sqrt{N} \left(\frac{k^*}{N} E Z_t + \frac{N-k^*}{N} E Z_t^* \right) = \sqrt{N} \left(\gamma E Z_t + (1-\gamma) E Z_t^* + O\Big(\frac{1}{N}\Big) \right) \\
&= \sqrt{N} \left(\nabla e(\tilde{\theta}_0) + O\Big(\frac{1}{N}\Big) \right) = O\Big(\frac{1}{\sqrt{N}}\Big)
\end{aligned}$$

as $\nabla e(\tilde{\theta}_0) = 0$. Using independence of $\{Y_t\}, \{Y_t^*\}$ and, hence of $\{Z_t\}, \{Z_t^*\}$, we have, using i), that $\frac{1}{\sqrt{N}} \nabla \widetilde{Q}_N(\tilde{\theta}_0)$ consists of two independent asymptotically normal components with mean 0, and, hence, (4.3) also holds under the alternative.

iii) We now get asymptotic normality of $\tilde{\theta}_N$ by the delta method as usual. For some $\theta_N^\dagger \in \Theta$ with $||\theta_N^\dagger - \tilde{\theta}_0|| \le ||\widehat{\theta}_N - \tilde{\theta}_0||$, we have from the mean-value theorem

$$0 = \nabla \widetilde{Q}_N(\hat{\theta}_N) = \nabla \widetilde{Q}_N(\tilde{\theta}_0) + \nabla^2 \widetilde{Q}_N(\theta_N^\dagger)(\widehat{\theta}_N - \tilde{\theta}_0)$$

and therefore

$$-\frac{1}{\sqrt{N}} \nabla \widetilde{Q}_N(\tilde{\theta}_0) = \frac{1}{N} \nabla^2 \widetilde{Q}_N(\theta_N^\dagger) \sqrt{N} (\tilde{\theta}_N - \tilde{\theta}_0).$$

The left-hand side is asymptotically normal with mean 0 and covariance matrix B by i) and ii). As, by Theorem 2.1, $\widehat{\theta}_N \to \tilde{\theta}_0$ a.s. and, hence, $\widehat{\theta}_N^\dagger \to \tilde{\theta}_0$ a.s. too, we get by applying Proposition 4.1 b) that the right-hand side is asymptotically equivalent to $A\sqrt{N}(\widehat{\theta}_N - \tilde{\theta}_0)$. The assertion of the theorem immediately follows from Slutsky's Lemma. ■

4.2 Proofs of Section 2.2

The following lemma allows to replace the sample residuals $\hat{\varepsilon}_t$ by the centered residuals

$$\varepsilon_t^0 = \varepsilon_t - \frac{1}{N} \sum_{s=1}^{N} \varepsilon_s\,, \quad t = 1, \ldots, N,$$

in the test statistic T_N without changing the asymptotics.

Lemma 4.1. *Under the assumptions of Theorem 2.3,*

$$\max_{1\le k<N} \sqrt{\frac{N}{k(N-k)}}\Big|\sum_{t=1}^{k}(\widehat{\varepsilon}_t-\varepsilon_t^0)\Big| = \mathrm{O}_p\Big(\sqrt{\frac{\log\log N}{N}}\Big)$$

Proof: By (A5), (A6), and using that $Y_{t-1}=0$ or $Y_{t-1}\ge 1$ and that Θ is compact, we have $E(\sup_{\theta\in\Theta}\|\nabla^2 g_\theta(Y_{t-1})\|)<\infty$. Again, an application of the uniform law of large numbers of Ranga Rao (Theorem 6.4 of [26]) implies, similar to the argument in the proof of Proposition 4.1,

$$\sup_{1\le k<N}\frac{1}{k}\sum_{t=1}^{k}\|\nabla^2 g_\theta(Y_{t-1})\| = \mathrm{O}_p(1)$$

uniformly in $\theta\in\Theta$, and, from a Taylor expansion, we get, uniformly in $k\le N$,

$$\frac{1}{k}\sum_{t=1}^{k}\Big(g_{\widehat{\theta}_N}(Y_{t-1})-g_{\theta_0}(Y_{t-1})\Big) = \frac{1}{k}\sum_{t=1}^{k}\nabla g_{\theta_0}(Y_{t-1})(\widehat{\theta}_N-\theta_0)+\mathrm{O}_p(||\widehat{\theta}_N-\theta_0||^2). \quad (4.4)$$

Setting

$$Z_t=\nabla g_{\theta_0}(Y_{t-1})-\mathrm{E}\,\nabla g_{\theta_0}(Y_{t-1})$$

we get, using (A9) and the same kind of argument as in the proof of Theorem 2.2, that $\{Z_t\}$ satisfies a strong invariance principle (4.2) which implies a bounded law of the iterated logarithm, compare Theorem 6 of Dehling and Philipp [5],

$$\sum_{t=1}^{k} Z_t = \mathrm{O}(\sqrt{k\log\log k}) \quad \text{a.s.} \quad (4.5)$$

From (1.5)

$$\frac{\partial}{\partial\theta_1}Q_N(\widehat{\theta}_N) = -2\sum_{t=1}^{N}(Y_t-g_{\widehat{\theta}_N}(Y_{t-1})) = -2\sum_{t=1}^{N}\hat{\varepsilon}_t.$$

By Theorem 2.1, $\widehat{\theta}_N\to\theta_0$ a.s. and, therefore, $\widehat{\theta}_N$ will be in the interior of Θ for large enough N, implying $\frac{\partial}{\partial\theta_1}Q_N(\widehat{\theta}_N)=0$ by definition of $\widehat{\theta}_N$. Hence, with $U_t=\hat{\varepsilon}_t-\varepsilon_t=g_{\theta_0}(Y_{t-1})-g_{\widehat{\theta}_N}(Y_{t-1})$

$$\sum_{t=1}^{k}(\hat{\varepsilon}_t-\varepsilon_t^0)=\sum_{t=1}^{k}(U_t-\overline{U}_N), \qquad \overline{U}_N=\frac{1}{N}\sum_{s=1}^{N}U_s$$

and, with $V_t=\nabla g_{\theta_0}(Y_{t-1})$, we have from (4.4)

$$\frac{1}{k}\sum_{t=1}^{k}(\hat{\varepsilon}_t-\varepsilon_t^0) = -\frac{1}{k}\sum_{t=1}^{k}(V_t-\overline{V}_N)(\widehat{\theta}_N-\theta_0)+\mathrm{O}_p(\|\widehat{\theta}_N-\theta_0\|^2).$$

By the same kind of argument as in part i) of the proof of Theorem 2.2, we have that $\overline{V}_N - \mathrm{E}\,V_t = \mathrm{O}_p(\frac{1}{\sqrt{N}})$. As, by Theorem 2.2, $\widehat{\theta}_N - \theta_0 = \mathrm{O}_p(\frac{1}{\sqrt{N}})$, too, and as $Z_t = V_t - \mathrm{E}\,V_t$, we have

$$\frac{1}{k}\sum_{t=1}^{k}(\hat{\varepsilon}_t - \varepsilon_t^0) = -\frac{1}{k}\sum_{t=1}^{k} Z_t\ (\widehat{\theta}_N - \theta_0) + \mathrm{O}_p\left(\frac{1}{N}\right). \tag{4.6}$$

The same assertion holds with $N-k$ replacing k and summation from $t = k+1, \ldots, N$. By an elementary calculation, we have

$$\begin{aligned}
\sqrt{\frac{N}{k(N-k)}}\left|\sum_{1}^{k}(\hat{\varepsilon}_t - \varepsilon_t^0)\right| &= \sqrt{\frac{N}{k(N-k)}}\left|\sum_{1}^{k}(U_t - \overline{U}_N)\right| \\
&= \sqrt{\frac{k(N-k)}{N}}\left|\frac{1}{k}\sum_{t=1}^{k}(U_t - \overline{U}_N) - \frac{1}{N-k}\sum_{t=k+1}^{N}(U_t - \overline{U}_N)\right| \\
&= \sqrt{\frac{k(N-k)}{N}}\left|\frac{1}{k}\sum_{t=1}^{k}(\hat{\varepsilon}_t - \varepsilon_t^0) - \frac{1}{N-k}\sum_{t=k+1}^{N}(\hat{\varepsilon}_t - \varepsilon_t^0)\right|
\end{aligned}$$

Combining (4.5) and (4.6) and using again Theorem 2.2, we finally get the assertion of the lemma. ∎

Proof of Theorem 2.3:
By Lemma 4.1, we may replace $\hat{\varepsilon}_t$ in the test statistic T_N by ε_t^0 if the hypothesis holds. By the same argument as in the proof of Theorem 2.2, we have the strong invariance principle (4.2) also for $Z_t = \varepsilon_t = Y_t - g_{\theta_0}(Y_{t-1})$. Now, we can replace ε_t by i.i.d. normal variables ξ_t with mean 0 and variance $\tau^2 = \mathrm{var}\ \varepsilon_t$ using standard arguments from changepoint analysis which we sketch now. For more details and references, we refer to the book by Csörgö and Horváth [3].

Recall $\mathrm{E}\varepsilon_t = 0$. The invariance principle (4.2) implies a law of the iterated logarithm from which we get

$$\max_{1\le k\le \log N}\sqrt{\frac{N}{k(N-k)}}\left|\sum_{t=1}^{k}\varepsilon_t^0\right| = \mathrm{o}_p\left(\frac{b(\log N)}{a(\log N)}\right)$$

and an analogous expression for the maximum over $N - \log N \le k < N$. The same kind of assertion also holds for $\xi_t^0 = \xi_t - \frac{1}{N}\sum_{s=1}^{N}\xi_s$ instead of ε_t^0. This implies

$$\begin{aligned}
&\mathrm{pr}\left(a(\log N)\max_{1\le k<N}\sqrt{\frac{N}{k(N-k)}}\left|\sum_{t=1}^{k}\varepsilon_t^0\right| - b(\log N) \le y\right) \\
= \ &\mathrm{pr}\left(a(\log N)\max_{\log N\le k\le N-\log N}\sqrt{\frac{N}{k(N-k)}}\left|\sum_{t=1}^{k}\varepsilon_t^0\right| - b(\log N) \le y\right) + \mathrm{o}(1)
\end{aligned}$$

and, again from the invariance principle (4.2)

$$\max_{\log N \le k \le N - \log N} \sqrt{\frac{N}{k(N-k)}} \left| \sum_{t=1}^{k} \varepsilon_t^0 - \sum_{t=1}^{k} \xi_t^0 \right| = \mathrm{o}_p(1)$$

which finishes the proof. ∎

Proof of Corollary 2.1:
By the definition of $\hat{\varepsilon}_t, \varepsilon_t$:

$$\sum_{t=1}^{N} \hat{\varepsilon}_t^2 = \sum_{t=1}^{N} \varepsilon_t^2 - 2 \sum_{t=1}^{N} \varepsilon_t \left(g_{\widehat{\theta}_N}(Y_{t-1}) - g_{\theta_0}(Y_{t-1}) \right) + \sum_{t=1}^{N} \left(g_{\widehat{\theta}_N}(Y_{t-1}) - g_{\theta_0}(Y_{t-1}) \right)^2 .$$

From the proof of Theorem 2.3, ε_t satisfies a strong invariance principle. As, from (A8), $\mathrm{E}\varepsilon_t^4 < \infty$, and as $\mathrm{E}\varepsilon_t^2 = \mathrm{var}\ \varepsilon_t = \tau^2$, we get from the central limit theorem

$$\frac{1}{N-d} \sum_{t=1}^{N} \varepsilon_t^2 = \tau^2 + O_p\left(\frac{1}{\sqrt{N}} \right).$$

As in deriving (4.4), but with a Taylor expansion to first order only,

$$\frac{1}{N-d} \sum_{t=1}^{N} \left(g_{\widehat{\theta}_N}(Y_{t-1}) - g_{\theta_0}(Y_{t-1}) \right)^2 = O_p\left(\|\widehat{\theta}_N - \theta_0\|^2 \right) = O_p\left(\frac{1}{N} \right)$$

by Theorem 2.2. Bounding the mixed term with the Cauchy-Schwarz inequality finishes the proof. ∎

4.3 Proofs of Section 2.3

Proof of Theorem 2.4:
As in the proof of Lemma 4.1, we have for large enough N that $\sum_{t=1}^{N} \hat{\varepsilon}_t = 0$. Therefore,

$$\widehat{S}_N(k^*) = \sum_{t=1}^{k^*} \left(X_t - g_{\widehat{\theta}_N}(X_{t-1}) \right) = - \sum_{t=k^*+1}^{N} (X_t - g_{\widehat{\theta}_N}(X_{t-1})). \tag{4.7}$$

Under the alternative, we have $X_t = Y_t,\ 1 \le t \le k^*,\ X_t = Y_t^*,\ k^*+1 \le t \le N$. Therefore, with $B(\theta), B^*(\theta)$ as in assumption (A10),

$$\begin{aligned}
\widehat{S}_N(k^*) &= \sum_{t=1}^{k^*} (Y_t - g_{\widehat{\theta}_N}(Y_{t-1})) \\
&= \sum_{t=1}^{k^*} (Y_t - \mathrm{E}Y_1) + k^* B(\widehat{\theta}_N) + O_p\left(\sup_{\theta \in \Theta} \left| \sum_{t=1}^{k^*} \left(g_\theta(Y_{t-1}) - \mathrm{E}g_\theta(Y_0) \right) \right| \right) \\
&= [\gamma N]\ B(\widehat{\theta}_N) + \mathrm{o}_p(N)
\end{aligned}$$

as we have the law of large numbers for $\{Y_t\}$ and, similar to the proof of Proposition 4.1, a uniform law of large numbers for $g_\theta(Y_{t-1})$. Mark that the condition $\mathrm{E}\sup_{\theta\in\Theta}|g_\theta(Y_{t-1})| < \infty$ needed for applying Theorem 6.4 of Ranga Rao [26] is satisfied by assumption (A6). Using our regularity assumptions regarding the dependence of $g_\theta(x)$ on θ, in particular (A6), we have for some $\theta_N^\dagger$ between $\widehat{\theta}_N$ and $\tilde{\theta}_0$

$$\left|B(\widehat{\theta}_N) - B(\tilde{\theta}_0)\right| = \left|\mathrm{E}\nabla^T g_{\theta_N^\dagger}(Y_0)(\widehat{\theta}_N - \tilde{\theta}_0)\right| \leq \beta\|\widehat{\theta}_N - \theta_0\| = \mathrm{O}_p\left(\frac{1}{\sqrt{N}}\right)$$

for some $\beta > 0$, applying Theorem 2.2. Hence, we have

$$\widehat{S}_N(k^*) = [\gamma N]\ B(\tilde{\theta}_0) + \mathrm{o}_p(N).$$

Analogously, we get by (4.7)

$$\widehat{S}_N(k^*) = -(N - [\gamma N])B^*(\tilde{\theta}_0) + \mathrm{o}_p(N).$$

By assumption (A10)

$$|\widehat{S}_N(k^*)| \geq N\ C\min(\gamma, 1-\gamma) + \mathrm{o}_p(N)$$

from which we get, using (A2)

$$\begin{aligned} T_N &\geq \sqrt{\frac{N}{k^*(N-k^*)}}\Big[N\ C\min(\gamma, 1-\gamma) + \mathrm{o}_p(N)\Big] \\ &= \sqrt{N}C_\gamma + \mathrm{o}_p(\sqrt{N}) \end{aligned}$$

for a constant $C_\gamma > 0$ depending only on C and γ. The assertion of the theorem follows immediately. ■

Proof of Corollary 2.2:
For $0 \leq u \leq 1$, we define

$$R(u) = \begin{cases} u|B(\tilde{\theta}_0)| & u \leq \gamma \\ (1-u)|B^*(\tilde{\theta}_0)| & u > \gamma \end{cases},$$

where $B(\theta) = \mathrm{E}Y_t - \mathrm{E}g_\theta(Y_{t-1})$, $B^*(\theta) = \mathrm{E}Y_t^* - \mathrm{E}g_\theta(Y_{t-1}^*)$. By (A3), we have $\gamma|B(\tilde{\theta}_0)| = (1-\gamma)|B^*(\tilde{\theta}_0)|$, such that R is a continuous function with a unique maximum in $u = \gamma$ by (A10).

From the proof of Theorem 2.4, we know

$$\sup_{0\leq u\leq 1}\left|\frac{1}{N}|\widehat{S}_N([uN])| - R(u)\right| = \mathrm{o}_p(1).$$

Hence, by standard arguments

$$\frac{\hat{k}^*}{N} = \arg\max_{0\le u\le 1} \frac{1}{N}|\widehat{S}_N([uN])| \to \arg\max_{0\le u\le 1} R(u) = \gamma.$$

■

Acknowledgements

The work was supported by the *Center for Mathematical and Computational Modelling* $(CM)^2$ funded by the state of Rhineland-Palatinate as well as DFG grant KI 1443/2-1. The position of Claudia Kirch was financed by the Stifterverband für die Deutsche Wissenschaft by funds of the Claussen-Simon trust.

References

[1] Albert, P. (1991): A two-state Markov mixture model for a time series of epileptic seizure counts. Biometrics 47, 1371-1381.

[2] Bai, J. (1993): On the partial sums of residuals in autoregressive and moving average models. J. Time Ser. Anal. 14, 247-260.

[3] Csörgő, M. and Horváth, L. (1997): Limit Theorems in Changepoint Analysis. Wiley, New York.

[4] Davis, R.A., Huang, D., and Yao, Y.-C. (1995): Testing for a change in the parameter values and order of an autoregressive model. Ann. Statist. 23, 282-304.

[5] Dehling, H., and Philipp, W. (1982): Almost sure invariance principles for weakly dependent vector-valued random variables. Ann. Probab. 10, 689-701.

[6] Doukhan, P. (1985): Mixing. Lecture Notes in Statistics 85. Springer-Verlag, Berlin-Heidelberg-New York.

[7] Engle, R.F. (1982): Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflations. Econometrica 50, 987-1007.

[8] Ferland, R., Latour, A., and Oraichi, D. (2006): Integer-valued GARCH process. J. Time Ser. Anal. 27, 923-942.

[9] Fokianos, K., Rahbek, A., and Tjostheim, D. (2009): Poisson autoregression. J. Amer. Statist. Assoc. 104, 1430-1439.

[10] Fokianos, K., and Fried, R. (2010): Interventions in INGARCH processes. J. Time Series Anal. 31, 210-225.

[11] Fokianos, K., and Tjostheim, D. (2011): Log-linear Poisson autoregression. J. Multivariate Anal. 102, 563-578.

[12] Fokianos, K. (2011): Count Time Series Models. To appear in Handbook of Statistics Vol 30: Time Series - Methods and Applications, eds. C.R. Rao and T. Subba Rao, Elsevier B.V., Amsterdam.

[13] Franke, J., and Seligmann, T. (1993): Conditional maximum likelihood estimates for INAR(1)-models and their application to modelling epileptic seizure counts. In: Developments in Time Series Analysis, ed. T. Subba Rao, Chapman and Hall, Boca Raton, FL.

[14] Franke, J. (2010): Weak Dependence of Functional INGARCH Processes. Report in Wirtschaftsmathematik 126, University of Kaiserslautern, 2010.

[15] Gandini, D. (1998): Time series models for deer populations using density dependent reproduction rates. Diploma Thesis, Dept. of Mathematics, University of Kaiserslautern.

[16] Horváth, L. (1993): Change in autoregressive processes. Stoch. Process. Appl. 44, 221-242.

[17] Hušková , M., and Koubková , A. (2005): Monitoring jump changes in linear models. J. Statist. Res. 39, 59-78.

[18] Hušková , M., Prášková, Z., and Steinebach, J. (2007): On the detection of changes in autoregressive time series, I. Asymptotics. J. Statist. Plann. Infer. 137, 1243-1259.

[19] Jung, R.C., and Tremayne, D. (2011): Convolution-closed models for count time series with applications. J. Time Series Anal. 32, 268-280.

[20] Kuelbs, J., and Philipp, W. (1980): Almost sure invariance principles for partial sums of mixing B-valued random variables. Ann. Probab. 8, 1003-1036.

[21] MacDonald, I.L., and Zucchini, W. (1997): Hidden Markov and other models for discrete-valued time series. Chapman & Hall. New York.

[22] Neumann, M.H. (2011): Poisson count processes: Ergodicity and goodness-of-fit. Bernoulli, to appear.

[23] Kirch, C., and Tadjuidje Kamgaing, J. (2011): Testing for parameter stability in nonlinear autoregressive models. Report in Wirtschaftsmathematik 137, University of Kaiserslautern, 2011.

[24] Kulperger, R.J. (1985): On the residuals of autoregressive processes and polynomial regression. Stoch. Process. Appl. 21, 107-118.

[25] Pötscher, B.M., and Prucha, I. (1997): Dynamic Nonlinear Econometric Models - Asymptotic Theory. Springer-Verlag, Berlin-Heidelberg-New York.

[26] Ranga Rao, R. (1962): Relation between weak and uniform convergence of measures with applications. Ann. Math. Statist. 33, 659-680.

[27] Rydberg, T.H., and Shephard, N. (2000): BIN models for trade-by-trade data. Modelling the number of trades in fixed interval of time. Paper 0740 (www.nuff.ox.ac.uk/users/rydberg/publications.htm), Econometric Society World Congress

[28] Streett, S. (2008): Some observation driven models for time series of counts. PhD Thesis, Colorado State University, Depart. of Statistics

[29] Triebsch, L.K. (2008): New integer-valued autoregressive and regression models with state-dependent parameters. PhD Thesis, University of Kaiserslautern, Depart. of Mathematics, Verlag Dr. Hut, München, 2008

[30] Weiß, C.H. (2007): Controlling correlated processes of Poisson counts. Quality and Reliability Engineering Intern. 23, 741-754.

[31] Weiß, C.H., and Testik, M.C. (2009): CUSUM monitoring of first-order integer-valued autoregressive processes of Poisson counts. J. Quality Techn. 41, 389-400.

[32] Weiß, C.H. (2010): The INARCH(1) Model for Overdispersed Time Series of Counts. Comm. Statist. - Simulation and Computation 39, 1269-1291.

[33] Weiß, C.H. (2010): INARCH(1) processes: Higher-order moments and jumps. Statistics and Probab. Letters 803, 1771-1780.

[34] Zhu, R., and Joe, H. (2006): Modelling count data time series with Markov processes based on binomial thinning. J. Time Series Anal. 27, 725-738.