Absolute regularity and ergodicity of Poisson count processes

Michael H. Neumann
Friedrich-Schiller-Universität Jena
Institut für Stochastik
Ernst-Abbe-Platz 2, 07743 Jena, Germany
E-mail: michael.neumann@uni-jena.de

Abstract

We consider a class of observation-driven Poisson count processes where the current value of the accompanying intensity process depends on previous values of both processes. We show under a contractive condition that the bivariate process has a unique stationary distribution and that the stationary version of the count process is absolutely regular. Moreover, since the intensities can be written as measurable functionals of the count variables we conclude that the bivariate process is ergodic. As an important application of these results, we show how a test method previously used in the case of independent Poisson data can be used in the case of Poisson count processes.

August 11, 2010

2010 Mathematics Subject Classification. Primary 60G10, 37A25; secondary 62M07.
Keywords and Phrases. Absolute regularity, ergodicity, integer-valued process, mixing, Poisson count process, test.
Running title. Ergodicity of Poisson count processes.

1. Introduction

The modeling and the analysis of count data has received increasing attention during the last decade. There are possible applications in various fields such as biometrics, econometrics and finance; see e.g. Davis, Dunsmuir and Wang (1999) and Davis and Wu (2009) for examples. A comprehensive account of models for time series of counts is given in Kedem and Fokianos (2002, Chapter 4). In the majority of cases the count variables are assumed to be Poisson distributed, conditioned on the past and perhaps some additional regressor variables. Models for count data consist of at least two processes, an observable process of counts and an accompanying intensity process that is usually not observed. Cox (1981) and later Davis, Dunsmuir and Wang (1999) classified these models into parameter-driven and observation-driven specifications. In the first case the accompanying intensity process evolves independently of the past history of the observation process while in the latter case the values on the intensity process do depend on past observations. The major aim of this paper is to derive important properties such as stationarity, mixing and ergodicity for a certain class of observation-driven processes. Davis, Dunsmuir and Wang (1999) mentioned that, in contrast to parameter-driven models where these properties are inherited by the observation process from the accompanying intensity process, there is much less theory available in the case of observation-driven models. Actually, ergodicity has been shown so far in a few special cases only; see e.g. **?**, Streett (2000), Davis, Dunsmuir and Streett (2003), Zheng and Basawa (2008) and Fokianos, Rahbek and Tjøstheim (2009). In these papers the authors could use classical Markov chain theory.

In the present paper we study a model where the observations N_t are Poisson distributed conditioned on the past, with an intensity λ_t depending on one lagged value of the count process and the intensity process, that is, $\lambda_t = f(\lambda_{t-1}, N_{t-1})$, for some function f. Models of this type have been considered before e.g. by Rydberg and Shephard (2000), Streett (2000), Davis, Dunsmuir and Streett (2003), Fokianos, Rahbek and Tjøstheim (2009), and Fokianos and Tjøstheim (2010). An important aspect is that such models allow for an AR feedback mechanism in the intensity process and it can be expected that this leads to a parsimonious parametrization. For clarity of exposition, we do not include additional regressor variables which are often also incorporated in specifications of the intensity. Under a contractive condition on f, we state in Section 2 that the bivariate process $((N_t, \lambda_t))_{t\in\mathbb{N}}$ has a unique stationary distribution. The proof of this result is based on a simple construction, where independently started versions of the process are coupled in such a way that they converge to each other. Section 3 contains the main results. For a stationary version of the process, we prove absolute regularity (β-mixing) of the univariate count process. Since the latter process is not Markovian we cannot rely on standard arguments from Markov chain theory; rather, we use again coupling arguments to derive this result. We also discuss an example that shows that the bivariate process $((N_t, \lambda_t))_{t\in\mathbb{Z}}$ is not absolutely regular in general. However, since the intensities can be written as measurable functionals of the count variables we conclude from the mixing property of the count process that the bivariate process is ergodic. In Section 4 we propose a test for a particular specification of the intensity process. We use a test statistic which has been applied before by several authors in connection with independent Poisson random variables. Using the ergodicity result from Section 3 we can show that the test statistic is asymptotically normal. All proofs are deferred to a final Section 5.

2. Stationarity of the bivariate process.

We assume that $(N_t)_{t\in\mathbb{N}}$ is a time series of counts, accompanied by an intensity process $(\lambda_t)_{t\in\mathbb{N}}$. Denote by $\mathcal{B}_t^{N,\lambda} = \sigma(\lambda_1, \ldots, \lambda_t, N_1, \ldots, N_t)$ the σ-field generated by the past and present values of the two processes at time t. We assume throughout that

$$N_t \mid \mathcal{B}_{t-1}^{N,\lambda} \sim \text{Poisson}(\lambda_t) \tag{2.1}$$

and

$$\lambda_t = f(\lambda_{t-1}, N_{t-1}), \tag{2.2}$$

for some function $f : [0,\infty)\times\mathbb{N}_0 \to [0,\infty)$ $(\mathbb{N}_0 = \mathbb{N}\cup\{0\})$. The starting value λ_1 may be random or nonrandom. It follows from the structure of the model that $\mathcal{B}_{t-1}^{N,\lambda} = \sigma(\lambda_1, N_1, \ldots, N_{t-1})$ and that the bivariate process $((N_t, \lambda_t))_{t\in\mathbb{N}}$ forms a homogeneous Markov chain. Throughout this paper we will assume that the function f satisfies the following contractive condition:

$$|f(\lambda, y) - f(\lambda', y')| \leq \kappa_1|\lambda - \lambda'| + \kappa_2|y - y'| \qquad \forall \lambda, \lambda' \geq 0, \forall y, y' \in \mathbb{N}_0, \tag{2.3}$$

where κ_1 and κ_2 are nonnegative constants with $\kappa := \kappa_1 + \kappa_2 < 1$. This includes as a special case a linear specification where $\lambda_t = \theta_0 + \theta_1\lambda_{t-1} + \theta_2 N_{t-1}$ and $\theta_0, \theta_1, \theta_2$ are nonnegative constants with $\theta_1 + \theta_2 < 1$. Rydberg and Shephard (2000) proposed such a model for describing the number of trades on the New York stock exchange in certain time intervals and called it BIN(1,1) model. Stationarity and other properties for this model where derived by Streett (2000), Ferland, Latour and Oraichi (2006) who referred to it as INGARCH(1,1) model, and Fokianos, Rahbek and Tjøstheim (2009). The generality of our condition (2.3) is chosen to include also nonlinear specifications such as the exponential AR model proposed in Fokianos, Rahbek and Tjøstheim (2009). In this case, the intensity function is specified as $f(\lambda, y) = (a + c\exp(-\gamma\lambda^2))\lambda + by$, where $a, b, c, \gamma > 0$. It follows from $\frac{\partial}{\partial y} f(\lambda, y) = b$ and $|\frac{\partial}{\partial \lambda} f(\lambda, y)| \leq a + c$ that (2.3) is fulfilled if $a + b + c < 1$.

Note that (2.3) implies that

$$f(\lambda, y) \leq f(0,0) + \kappa_1\lambda + \kappa_2 y. \tag{2.4}$$

It follows from (2.4) that $E(\lambda_t \mid \lambda_{t-1}) \leq f(0,0) + \kappa\lambda_{t-1}$ which leads to

$$E(N_t \mid \lambda_1) = E(\lambda_t \mid \lambda_1) \leq f(0,0)\frac{1-\kappa^{t-1}}{1-\kappa} + \kappa^{t-1}\lambda_1. \tag{2.5}$$

Hence, the bivariate chain $((N_t, \lambda_t))_{t\in\mathbb{N}}$ is bounded in probability on average. Moreover, it follows from (2.3) that, for any open set $O \in 2^{\mathbb{N}_0} \otimes \mathcal{B}$, the transition probabilities $P((N_t, \lambda_t) \in O \mid (N_{t-1}, \lambda_{t-1}) = \cdot)$ are a continuous, and therefore also a lower semicontinuous function. Hence, the Markov chain is a weak Feller chain and it follows from Theorem 12.1.2(ii) in Meyn and Tweedie (1993) that there exists at least one stationary distribution. Uniqueness of this stationary distribution, however, requires more than (2.4) only and will follow from the contractive condition (2.3). The following theorem summarizes this and a few other useful facts.

Theorem 2.1. *Suppose that the bivariate chain $((N_t, \lambda_t))_{t\in\mathbb{N}}$ obeys (2.1), (2.2) and (2.3). Then*

(i) *There exists a unique stationary distribution π.*
(ii) *If $(N_1, \lambda_1) \sim \pi$, then $E\lambda_1 < \infty$.*
(iii) *If $f(0,0) = 0$, then $\pi(\{0,0\}) = 1$. If $f(0,0) > 0$, then $\pi(\{y,\lambda\}) < 1$ for all $y \in \mathbb{N}_0$, $\lambda \in [0,\infty)$.*

Remark 1. Using the contractive property (2.3) we will show in the proof of Theorem 2.1 that the n-step transition probabilities $P((N_{n+1}, \lambda_{n+1}) \in A \mid (N_1, \lambda_1) = x)$ converge to a common limit $\pi(A)$ not depending on the starting value x, where π is a probability measure. This will imply that π is the unique stationary distribution.

There are alternative ways to prove Theorem 2.1. Introducing a sequence of independent "innovations" $(U_t)_{t\in\mathbb{N}}$ with $U_t \sim$ Uniform$[0,1]$ we could re-express the process values as

$$(N_{t+1}, \lambda_{t+1}) \;=\; G((N_t, \lambda_t), U_{t+1}) \;:=\; (F^{-1}_{f(\lambda_t, N_t)}(U_{t+1}), f(\lambda_t, N_t)),$$

where F_λ denotes the cumulative distribution function of a Poisson(λ) distribution. This gives us a representation of $((N_t, \lambda_t))_{t\in\mathbb{N}}$ as a randomly perturbed dynamical system with independent and identically distributed innovations. In such a context and under a contractive condition similar to our (2.3), Diaconis and Freedman (1999) also proved existence and uniqueness of a stationary distribution. To this end, these authors used backward iterations to identify a random variable which has the desired stationary distribution. The approach used here is more direct and uses elements of standard Markov chain theory as described e.g. in Meyn and Tweedie (1993). Finally, we would like to mention that Lasota and Mackey (1989) also proved the existence of a unique stationary distribution under conditions similar to our (2.3) and (2.4); see in particular equations (2) and (3) in their paper. Their proof contains similar ingredients as our proof, however, it is more analytic in nature while we established a coupling to represent the convergence facts in a simple stochastic language.

3. Absolute regularity of the count process and ergodicity

In this section we state the main results of our paper, absolute regularity of the count process and, as a consequence, ergodicity of the bivariate process $((N_t, \lambda_t))_t$. Actually, Grunwald, Hyndman, Tedesco and Tweedie (2000, Case II of Proposition 3), Streett (2000) and Davis, Dunsmuir and Streett (2003) proved ergodicity in a special case; however, they made heavy use of the particular form of their link function f and could show that Doeblin's condition is fulfilled. Hence, they could employ Markov chain technology to prove ergodicity. We cannot use this approach in the case considered here since Doeblin's condition will not be satisfied in general. Another commonly used approach to proving ergodicity, which is not restricted to the case of Markov chains, consists in proving first strong mixing as a sufficient condition for ergodicity.

It turns out, however, that the bivariate process $((N_t, \lambda_t))_t$ is not strongly mixing in general; a counterexample is given in Remark 3 below. The problem lies in the discreteness of the distribution of the “innovations” N_t while the λ_t take values on a continuous scale which make the commonly used coupling approach to proving mixing properties of Markov chains impossible. To give some idea why a discrete distribution of the innovations may cause problems, we recall the well-known example of a stationary AR(1)-process, $X_t = \theta X_{t-1} + \varepsilon_t$, where the innovations are independent with $P(\varepsilon_t = 1) = P(\varepsilon_t = -1) = 1/2$ and $0 < |\theta| \leq 1/2$. This process has a stationary distribution supported on $[-2, 2]$. It follows from the above model equation that X_t has with probability 1 the same sign as ε_t. Hence, we could perfectly recover $X_{t-1}, X_{t-2}, \ldots$ from X_t which clearly excludes any of the common mixing properties to hold. (Rosenblatt (1980) mentioned the fact that a process similar to $(X_t)_{t\in\mathbb{Z}}$ is purely deterministic going backwards in time. A rigorous proof that it is not strong mixing was given by Andrews (1984).) On the other hand, we can prove absolute regularity for the (univariate) count process $(N_t)_t$. For this purpose, the discrete nature of the distribution of the N_t does not harm. To see why, note that we have either $\pi(\{0,0\}) = 1$ or $P(\lambda_{t-1} > 0 \text{ or } \lambda_t > 0) = 1$; see the proof of part (iii) of Theorem 2.1. Therefore, the support of the conditional distribution of N_{t+2} given $N_t, N_{t+1}, \ldots$ is equal to the support of the stationary distribution of the N_t. Therefore, we can actually construct a successful coupling. Since absolute regularity implies strong mixing we obtain immediately ergodicity of the count process $(N_t)_t$. Moreover, as a by-product of our coupling, we see that the random intensities λ_t can be expressed as measurable functionals of past variables of the count process. Hence, we finally obtain the desired ergodicity of the bivariate process $((N_t, \lambda_t))_t$.

It was stated in Section 2 that the bivariate process $((N_t, \lambda_t))_t$ has a unique stationary distribution under the contractive condition (2.3). In this section, we will assume throughout that this process is in its stationary regime. Moreover, it proves to be quite convenient here to have a two-sided stationary version, with time domain $\mathbb{Z}$ rather than $\mathbb{N}$, which of course exists by Kolmogorov's extension theorem; see e.g. Durrett (1991, p. 293). Here is the main result of our paper.

Theorem 3.1. *Suppose that the bivariate chain $((N_t, \lambda_t))_{t\in\mathbb{Z}}$ is in its stationary regime and obeys (2.1), (2.2) and (2.3). Then*

(i) *The count process $(N_t)_{t\in\mathbb{Z}}$ is absolutely regular with coefficients satisfying*

$$\beta(n) \leq 2\, E\lambda_1\, \kappa^{n-1}/(1-\kappa_1).$$

(ii) *There exists a measurable function $g : \mathbb{N}_0^\infty = \{(n_1, n_2, \ldots) : n_i \in \mathbb{N}_0\} \longrightarrow [0, \infty)$ such that $\lambda_t = g(N_{t-1}, N_{t-2}, \ldots)$ holds almost surely.*

(iii) *The process $((N_t, \lambda_t))_{t\in\mathbb{Z}}$ is geometrically ergodic.*

(iv) $E\lambda_1^2 < \infty$.

Remark 2. In the case of a so-called INGARCH(1,1) process where λ_t is specified as $\lambda_t = \theta_0 + \theta_1\lambda_{t-1} + \theta_2 N_{t-1}$, Ferland, Latour and Oraichi (2006) proved the stronger result that all moments of λ_t and N_t are finite. Since it follows from (2.4) that $\lambda_t \leq f(0,0) + \kappa_1\lambda_{t-1} + \kappa_2 N_{t-1}$ we conjecture that their result can be generalized by simple majorization arguments to our more general framework. However, since higher than second moments are nowhere needed for the purposes in this paper we do not make the attempt to adapt their proof which was already quite involved in the special case of a linear specification of λ_t.

Remark 3. Theorem 3.1 states that the count process $(N_t)_{t\in\mathbb{Z}}$ is absolutely regular and therefore also strongly mixing under condition (2.3). This allowed us to conclude that the bivariate process $((N_t,\lambda_t))_{t\in\mathbb{Z}}$ is ergodic. However, the process $((N_t,\lambda_t))_{t\in\mathbb{Z}}$ and even the intensity process $(\lambda_t)_{t\in\mathbb{Z}}$ alone are not strongly mixing in general. To see this, consider the specification $f(\lambda,y)=g(\lambda)+y/2$, where g is strictly monotone and satisfies $0<c_1\leq g(\lambda)<0.5$ and $|g(\lambda)-g(\lambda')|\leq c_2|\lambda-\lambda'|$ for some $c_2<0.5$ and for all λ,λ'. Then f satisfies our contracte condition (2.3). Using the fact that $g(\lambda)\in[c_1,0.5)$ we obtain that $2g(\lambda_{t-1})=2\lambda_t-[2\lambda_t]$ which implies that we can perfectly recover λ_{t-1} once we know the value of λ_t. Iterating this argument we see that we can recover from λ_t the entire past of the intensity process. Taking into account that the above choice of f excludes the case that the intensity process is purely nonrandom we conclude that a stationary version of $(\lambda_t)_{t\in\mathbb{Z}}$ cannot be strongly mixing.

Remark 4. The primary intention of the author was to devise a method of proving ergodicity of certain count processes. This is done, mainly for clarity of presentation, for the simple case where the intensity depends only on one lagged value of the count process and the intensity process. In contrast to previous work in this area, the coupling approach used here does not require Markovianity of the process. The results of this paper, and in particular the ergodicity stated in Theorem 3.1, can be generalized to more complex models with more than one or even infinitely many lagged variables. Moreover, it seems to be possible to include covariates, at least if they are exogeneous. These generalizations are well beyond the scope of this paper and should be the subject of future research.

4. A specification test for the intensity function

There might be good reasons for assuming that the count variables are Poisson distributed, conditioned on the past. However, a particular specification for the intensity function seems to be more questionable and such a choice should be supported by a statistical test. Here we propose a test statistic which was originally designed for testing overdispersion in the context of i.i.d. observations, see e.g. Lee (1986) and Cameron and Trivedi (1986).

Assume that we have observations $N_1,\ldots,N_n$ from a stationary process $((N_t,\lambda_t))_{t\in\mathbb{Z}}$ obeying (2.1) and (2.2) and that we want to test the simple hypothesis

$$H_0: \quad f = f_0 \qquad \text{against} \qquad H_1: \quad f \neq f_0,$$

for some f_0 satisfying (2.3), or the composite hypothesis

$$H_0': \quad f \in \{f_\theta:\theta\in\Theta\} \qquad \text{against} \qquad H_1': \quad f \notin \{f_\theta:\theta\in\Theta\},$$

where $\Theta\subseteq\mathbb{R}^d$ and the f_θ satisfy (2.3).

To motivate a particular test statistic, pretend that we additionally observe the starting value λ_1 of the intensity process. Then we could take, for testing H_0 against H_1, the statistic

$$T_{n,0} \;=\; \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ (N_t - \lambda_t^0)^2 \;-\; N_t \right\},$$

where $\lambda_1^0 = \lambda_1$ and, for $t = 2, \ldots, n$, the λ_t^0 are recursively defined as $\lambda_t^0 = f_0(\lambda_{t-1}^0, N_{t-1})$. The idea behind this statistic is very simple. If the the intensity function f is correctly specified, then $E\left[(N_t - \lambda_t^0)^2 - N_t\right] = E\left[E\left((N_t - \lambda_t)^2 - N_t | \, \lambda_t\right)\right] = 0$ and, as stated in Proposition 4.1 below, $T_{n,0} \stackrel{d}{\longrightarrow} \mathcal{N}(0, 2E\lambda_1^2)$. On the other hand, if f is not correctly specified by f_0, then the random variables $(N_t - \lambda_t)^2 - N_t$ are not centered and we can expect consistency of the test.

In the more relevant case of unknown λ_1 we replace this by any arbitrarily chosen, random or nonrandom, starting value $\widetilde{\lambda}_1$, then define recursively $\widetilde{\lambda}_t = f_0(\widetilde{\lambda}_{t-1}, N_{t-1})$, for $t = 2, \ldots, n$, and take the test statistic

$$T_n \;=\; \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ (N_t - \widetilde{\lambda}_t)^2 \;-\; N_t \right\}.$$

In the case of testing H_0' against H_1' we estimate the parameter θ by some estimator $\widehat{\theta}_n$ first and take then the test statistic

$$\widehat{T}_n \;=\; \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ (N_t - \widehat{\lambda}_t)^2 \;-\; N_t \right\}.$$

Here $\widehat{\lambda}_1$ is again any starting value and $\widehat{\lambda}_t = f_{\widehat{\theta}_n}(\widehat{\lambda}_{t-1}, N_{t-1})$, for $t = 2, \ldots, n$.

Remark 5. In the context of independent observations, Lee (1986) and Cameron and Trivedi (1986) considered a test statistic similar to our's for testing the Poisson hypothesis against the alternative that the distribution belongs to the so-called Katz familiy of distributions. This family contains as special cases the Poisson, negative binomial and binomial distributions. While the variance equals the mean in the Poisson case, the latter two classes contain distributions for which the variance mean ratio is strictly greater or less than 1, respectively. Therefore, Lee (1986) and Cameron and Trivedi (1986) interpreted their tests as tests for over- or underdispersion. The same test statistic was also suggested in Cox (1983). It was also used by Brännäs and Johansson (1994) for testing for the existence of a latent process in the context of Poisson count models. Again in the case of independent data, Dean and Lawless (1989) and Dean (1992) came up with adjusted versions of Lee's and Cameron and Trivedi's test statistic that have the same limit distribution as the unadjusted statistic but are closer to this limit in small samples.

We will prove that the above statistics, $T_{n,0}$, T_n and $\widehat{T}_n$, are asymptotically normal with the same limit. This can be most easily done for $T_{n,0}$ since this statistic is a sum of martingale differences which allows us to apply an appropriate central limit theorem.

Proposition 4.1. *Suppose that the bivariate process is stationary and obeys (2.1) and (2.2). If H_0 is true and f_0 satisfies the contractive condition (2.3), then*

$$T_{n,0} \stackrel{d}{\longrightarrow} \mathcal{N}(0, 2E\lambda_1^2).$$

Next we will show that T_n and $\widehat{T}_n$ have the same limit distribution as $T_{n,0}$. To this end, we will simply show that the difference between the former statistics to $T_{n,0}$ converges to zero in probability. This is not surprising at all for T_n since it follows from (2.3) that $|\widetilde{\lambda}_t - \lambda_t| \leq \kappa_1^{t-1}|\widetilde{\lambda}_1 - \lambda_1|$. The following lemma shows that also $\widehat{\lambda}_t$ will be close to λ_t if $\widehat{\theta}_n$ is a $\sqrt{n}$-consistent estimator of θ and if $f_\theta(\lambda, y)$ is a smooth function in θ.

Lemma 4.1. *Suppose that the bivariate process obeys (2.1) and (2.2) with $f = f_{\theta_0}$. We assume that $\widehat{\theta}_n - \theta_0 = O_P(n^{-1/2})$. Furthermore, we assume that there exist $C < \infty$, $\kappa_1, \kappa_2 \geq 0$ with $\kappa := \kappa_1 + \kappa_2 < 1$ such that*

(i) $|f_{\theta'}(\lambda, y) - f_{\theta_0}(\lambda, y)| \leq C\, \|\theta' - \theta_0\|\, (\lambda + y + 1) \qquad \forall \lambda, y,$

(ii) $|f_{\theta'}(\lambda, y) - f_{\theta'}(\widetilde{\lambda}, \widetilde{y})| \leq \kappa_1|\lambda - \widetilde{\lambda}| + \kappa_2|y - \widetilde{y}|$

hold for all $\theta' \in \Theta$ with $\|\theta' - \theta_0\| \leq \delta$, for some $\delta > 0$.

Then

$$\sum_{t=1}^n (\lambda_t - \widehat{\lambda}_t)^2 = O_P(1).$$

We think that our assumption on the estimator $\widehat{\theta}_n$ is a realistic one in many cases. It is fulfilled, for example, by the conditional maximum likelihood estimator studied in Fokianos, Rahbek and Tjøstheim (2009).

Theorem 4.1. *Suppose that the assumptions of Lemma 4.1 are fulfilled.*

Then

$$\widehat{T}_n \stackrel{d}{\longrightarrow} \mathcal{N}(0, 2\, E\lambda_1^2).$$

Remark 6. The same assertion holds true for T_n instead of $\widehat{T}_n$ since this is obviously a special case of that considered in Theorem 4.1.

Note that the limit distribution of $\widehat{T}_n$ still contains the parameter $E\lambda_1^2$ that is usually not known in advance and has to be estimated. We obtain from Lemma 4.1 by the Minkowski inequality that $\left|\sqrt{n^{-1}\sum_{t=1}^n \widehat{\lambda}_t^2} - \sqrt{n^{-1}\sum_{t=1}^n \lambda_t^2}\right| \leq \sqrt{n^{-1}\sum_{t=1}^n(\lambda_t - \widehat{\lambda}_t)^2} = O_P(n^{-1/2})$, which leads in conjunction with ergodicity of $(\lambda_t)_{t\in\mathbb{Z}}$ to

$$\frac{1}{n}\sum_{t=1}^n \widehat{\lambda}_t^2 \stackrel{P}{\longrightarrow} E\lambda_1^2. \tag{4.1}$$

For a prescribed size $\alpha \in (0,1)$, we propose a test for H_0' against H_1' as

$$\varphi_n \,=\, I\left(\left(\frac{2}{n} \sum_{t=1}^n \widehat{\lambda}_t^2 \right)^{-1/2} \widehat{T}_n \,>\, u_\alpha \right),$$

where $u_\alpha = \Phi^{-1}(1-\alpha)$ denotes the $(1-\alpha)$-quantile of the standard normal distribution. From Theorem 4.1 and (4.1) we conclude that this test has asymptotically the correct size.

Theorem 4.2. *Suppose that the assumptions of Lemma 4.1 are fulfilled and that* $f_{\theta_0}(0,0) > 0$. *Then we have under* H_0'

$$\left((2/n) \sum_{t=1}^n \widehat{\lambda}_t^2 \right)^{-1/2} \widehat{T}_n \stackrel{d}{\longrightarrow} \mathcal{N}(0,1),$$

which implies that

$$P(\varphi_n \,=\, 1) \underset{n\to\infty}{\longrightarrow} \alpha.$$

5. Proofs

As already mentioned in the text, the main results of this paper, Theorem 2.1 and Theorem 3.1, are both proved by coupling arguments. Necessary technical prerequisites are summarized in the following lemma.

Lemma 5.1. *For arbitrary* $\lambda_1, \lambda_2 \geq 0$, *we can construct on an appropriate probability space* $X_1 \sim$ *Poisson*(λ_1) *and* $X_2 \sim$ *Poisson*(λ_2) *such that*

(i) $E|X_1 - X_2| \,=\, |\lambda_1 - \lambda_2|$,
(ii) $P(X_1 \neq X_2) \,\leq\, |\lambda_1 - \lambda_2|$.

Proof. Let, without loss of generality, $\lambda_1 \leq \lambda_2$. We take independent random variables $X_1 \sim$ Poisson(λ_1), $Z \sim$ Poisson$(\lambda_2 - \lambda_1)$, and define $X_2 = X_1 + Z$. Then $X_2 \sim$ Poisson(λ_2),

$$E|X_1 - X_2| \,=\, EZ \,=\, |\lambda_1 - \lambda_2|$$

and

$$P(X_1 \neq X_2) \,=\, 1 \,-\, P(Z=0) \,=\, 1 \,-\, e^{-(\lambda_2 - \lambda_1)} \,\leq\, |\lambda_1 - \lambda_2|.$$

□

Proof of Theorem 2.1. As mentioned above, we could use the fact that $((N_t, \lambda_t))_{t\in\mathbb{N}}$ is a weak Feller chain which is bounded in probability on average to conclude from Theorem 12.1.2(ii) in Meyn and Tweedie (1993) that it has at least one stationary distribution. Uniqueness could then eventually be derived from the contraction property (2.3). We think, however, that it is more instructive to the reader when a self-contained proof that uses arguments closely tied to the particular case at hand is presented.

Let P_λ^t be the conditional distribution of (N_t, λ_t) given $\lambda_1 = \lambda$, where $\lambda \in [0,\infty)$ is an arbitrarily chosen but fixed starting value. It follows from (2.5) that the sequence of distributions $(P_\lambda^t)_{t\in\mathbb{N}}$ is tight. Hence, there exists a subsequence $(n_k)_{k\in\mathbb{N}}$ of $\mathbb{N}$ such that $P_\lambda^{n_k}$ converges weakly to some probability measure π_λ, as $k \to \infty$. We will show that this limit does not depend on the starting value λ and that also the full sequence $(P_\lambda^n)_{n\in\mathbb{N}}$ converges. This will immediately imply that π is a stationary distribution which is unique.

The latter conclusions will follow after we have derived a few convergence properties of the process. To this end, we construct, on an appropriate probability space $(\Omega', \mathcal{A}', P')$, two Markov chains $((N_t', \lambda_t'))_{t\in\mathbb{N}}$ and $((N_t'', \lambda_t''))_{t\in\mathbb{N}}$ with transition laws according to (2.1) and (2.2) and with starting values λ_1' and λ_1'', respectively. We construct these chains iteratively. Given λ_1' and λ_1'', (i) of Lemma 5.1 allows us to construct N_1' and N_1'' in such a way that

$$E\left(|N_1' - N_1''| \,\middle|\, \lambda_1', \lambda_1''\right) \;=\; |\lambda_1' - \lambda_1''|.$$

The values of λ_2' and λ_2'' are then given by equation (2.2) and it follows from (2.3) that

$$\begin{aligned} E\left(|\lambda_2' - \lambda_2''| \,\middle|\, \lambda_1', \lambda_1''\right) \;&\leq\; \kappa_1|\lambda_1' - \lambda_1''| \,+\, \kappa_2 E\left(|N_1' - N_1''| \,\middle|\, \lambda_1', \lambda_1''\right) \\ &=\; \kappa|\lambda_1' - \lambda_1''|. \end{aligned}$$

In the next step we can construct N_2' and N_2'' such that $E(|N_2' - N_2''| \mid \lambda_1', \lambda_1'', \lambda_2', \lambda_2'') = |\lambda_2' - \lambda_2''|$ which also implies that

$$E\left(|N_2' - N_2''| \,\middle|\, \lambda_1', \lambda_1''\right) \;\leq\; \kappa|\lambda_1' - \lambda_1''|.$$

Now we can proceed in the same way and construct the pairs $(N_3', N_3''), (N_4', N_4''), \ldots$. With the above construction, we obtain that

$$E\left(|\lambda_t' - \lambda_t''| \,\middle|\, \lambda_1', \lambda_1''\right) \;\leq\; \kappa^{t-1}|\lambda_1' - \lambda_1''|. \tag{5.1}$$

and

$$E\left(|N_t' - N_t''| \,\middle|\, \lambda_1', \lambda_1''\right) \;\leq\; \kappa^{t-1}|\lambda_1' - \lambda_1''|. \tag{5.2}$$

Hence, it follows that $(P_{\lambda_1'}^{n_k})_{k\in\mathbb{N}}$ and $(P_{\lambda_1''}^{n_k})_{k\in\mathbb{N}}$ converge for any choice of λ_1' and λ_1'' to the same limit, which we denote by π in the following. Now we can translate this result to a convergence result for the conditional distributions of the Markov chain $((N_t, \lambda_t))_{t\in\mathbb{N}}$. Since the above convergence is uniform in λ_1' over compact sets and since f as a continuous function maps compact subsets of $[0,\infty) \times \mathbb{N}_0$ to compact subsets of $[0,\infty)$ we obtain that

$$\sup_{x\in K} |P^{n_k}(x, A) \,-\, \pi(A)| \underset{k\to\infty}{\longrightarrow} 0 \tag{5.3}$$

holds for every compact subset K of $\mathbb{N}_0 \times [0,\infty)$ and every π-continuity set A. Here $P^n(x, A) = P((N_{t+n}, \lambda_{t+n}) \in A \mid (N_t, \lambda_t) = x)$ denotes the n-step transition probability of the bivariate process. (5.3) will allow us to show convergence of the full sequence. For any $n \in \mathbb{N}$, let $k(n)$ be the largest integer such that $n_{k(n)} < n$. From tightness of $(P_\lambda^n)_{n\in\mathbb{N}}$ and (5.3) we conclude that

$$P_\lambda^n \;=\; \int P^{n_{k(n)}}(x, \cdot)\, P_\lambda^{n-n_{k(n)}}(dx) \;\Longrightarrow\; \pi \qquad \text{for all } \lambda \in [0,\infty). \tag{5.4}$$

It follows directly from this equation that π is a stationary distribution. To see this, observe that it follows from (5.4) that $Q_x^n := n^{-1}\sum_{t=1}^n P_\lambda^t$ converges weakly to π. Furthermore, it also follows that $\widetilde{Q}_\lambda^n := n^{-1}\sum_{t=1}^n P_\lambda^{t+1} \Longrightarrow \pi$, that is,

$$\widetilde{Q}_\lambda^n(A) \underset{k\to\infty}{\longrightarrow} \pi(A) \tag{5.5}$$

if A is a π-continuity set, i.e., $\pi(\partial A) = 0$. If A is an open set, then $x \mapsto P^1(x, A)$ is a continuous and bounded function. Therefore,

$$\widetilde{Q}_\lambda^n(A) \;=\; \int P^1(x, A)\, Q_\lambda^n(dx) \underset{n\to\infty}{\longrightarrow} \int P^1(x, A)\, \pi(dx). \tag{5.6}$$

From (5.5) and (5.6) we obtain that the probability measures π and $\int P^1(x,\cdot)\,\pi(dx)$ coincide for all open π-continuity sets A. Since these sets are stable under finite intersections and generate $2^{\mathbb{N}_0} \otimes \mathcal{B}$ we conclude that

$$\pi(A) \;=\; \int P^1(x, A)\, \pi(dx) \qquad \forall A \in 2^{\mathbb{N}_0} \otimes \mathcal{B},$$

that is, π is actually a stationary distribution. Let now π' be an arbitrary distribution. Then we obtain by majorized convergence, for any π-continuity set A,

$$\int P^n(x, A)\, \pi'(dx) \underset{n\to\infty}{\longrightarrow} \int \pi(A)\, \pi'(dx) \;=\; \pi(A).$$

If π' is a stationary distribution, then we also have that $\int P^n(x, A)\, \pi'(dx) \;=\; \pi'(A)$, which implies that $\pi = \pi'$. Hence, (i) is proved.

We obtain from (2.5) by Theorem 5.3 in Billingsley (1968) that

$$E_\pi \lambda_1 \;\le\; \liminf_{t\to\infty} E(\lambda_t \mid \lambda_1 = 0) \;\le\; f(0,0)/(1-\kappa),$$

which proves (ii).

To see (iii), note that $f(0, 0) = 0$ implies by (2.5) that $E(\lambda_t \mid \lambda_1 = 0) = 0$ holds for all t which in turn implies that $\pi(\{0, 0\}) = 1$. On the other hand, if $f(0, 0) > 0$, then we can conclude that $P(\lambda_{t-1} > 0 \text{ or } \lambda_t > 0) \;=\; P(\lambda_{t-1} > 0) + P(\lambda_{t-1} = 0, \lambda_t > 0) \;\ge\; P(\lambda_{t-1} > 0) + P(\lambda_{t-1} = 0) \;=\; 1$. This implies that $((N_t, \lambda_t))_{t\in\mathbb{Z}}$ cannot be nonrandom, as required. □

Proof of Theorem 3.1. Let, for $-\infty \le k \le l \le \infty$, $\mathcal{B}_{k,l}^N = \sigma(N_k, \ldots, N_l)$. Recall that the coefficients of absolute regularity of the count process $(N_t)_{t\in\mathbb{N}}$ are defined as

$$\beta(n) \;=\; E\left[\sup_{A\in\mathcal{B}_{n,\infty}^N} \left|P(A \mid \mathcal{B}_{-\infty,0}^N) \;-\; P(A)\right|\right].$$

Hence,

$$\beta(n) \;\le\; E\left[\sup_{A\in\mathcal{B}_{n,\infty}^N} \left|P(A \mid \sigma(\lambda_1, N_0, N_{-1}, \ldots)) \;-\; P(A)\right|\right].$$

Furthermore, it follows from $(N_n, N_{n+1}, \ldots) \mid \sigma(\lambda_1, N_0, N_{-1}, \ldots) \;=\; (N_n, N_{n+1}, \ldots) \mid \sigma(\lambda_1)$ that

$$\beta(n) \;\le\; E\left[\sup_{A\in\mathcal{B}_{n,\infty}^N} \left|P(A \mid \sigma(\lambda_1)) \;-\; P(A)\right|\right]. \tag{5.7}$$

Let $\mathcal{B}^\infty$ be the σ-field in $\mathbb{R}^\infty = \{(x_1, x_2, \dots) : x_i \in \mathbb{R}\}$ generated by the cylinder sets, that is,

$$\mathcal{B}^\infty \;=\; \sigma\left(\{B \times \mathbb{R}^\infty : B \in \mathcal{B}^k, k \in \mathbb{N}\}\right).$$

We can rewrite (5.7) in terms of the process variables as

$$\beta(n) \;\leq\; E\left[\sup_{A\in\mathcal{B}^\infty}\Big|P((N_n, N_{n+1}, \dots) \in A \mid \lambda_1) \;-\; P((N_n, N_{n+1}, \dots) \in A)\Big|\right]. \tag{5.8}$$

We will derive an upper estimate for the right-hand side of (5.8) via a coupling approach similar to that in the proof of Theorem 2.1. To this end, we will construct on an appropriate probability space $(\Omega', \mathcal{A}', P')$ two versions of the bivariate process, $((N_t', \lambda_t'))_{t\in\mathbb{N}}$ and $((N_t'', \lambda_t''))_{t\in\mathbb{N}}$, where the starting values λ_1' and λ_1'' are independent and distributed according to the stationary law π. Since, for any $A \in \mathcal{B}^\infty$,

$$P\left((N_n'', N_{n+1}'', \dots) \in A \mid \lambda_1'\right) \;=\; P\left((N_n, N_{n+1}, \dots) \in A\right)$$

it follows that

$$\begin{aligned}
&|P\left((N_n, N_{n+1}, \dots) \in A \mid \lambda_1 = u\right) \;-\; P\left((N_n, N_{n+1}, \dots) \in A\right)| \\
&\quad=\; \left|P\left((N_n', N_{n+1}', \dots) \in A \mid \lambda_1' = u\right) \;-\; P\left((N_n'', N_{n+1}'', \dots) \in A \mid \lambda_1' = u\right)\right| \\
&\quad\leq\; P\left((N_n', N_{n+1}', \dots) \neq (N_n'', N_{n+1}'', \dots) \mid \lambda_1' = u\right).
\end{aligned}$$

Therefore we obtain that

$$\beta(n) \;\leq\; P\left((N_n', N_{n+1}', \dots) \neq (N_n'', N_{n+1}'', \dots)\right). \tag{5.9}$$

Hence, to estimate $\beta(n)$, we will construct a coupling such that the processes $(N_t')_{t\in\mathbb{N}}$ and $(N_t'')_{t\in\mathbb{N}}$ coalesce after n steps with a high probability.

Using exactly the same construction as in the proof of Theorem 2.1 we can successively construct pairs $(N_1', N_1''), (N_2', N_2''), \dots$ such that

$$E\left(|\lambda_n' \;-\; \lambda_n''|\,\Big|\, \lambda_1', \lambda_1''\right) \;\leq\; \kappa^{n-1}\,|\lambda_1' \;-\; \lambda_1''|.$$

From here on we deviate from the approach in the proof of Theorem 2.1 where we constructed all pairs (N_t', N_t'') such that their mean distance was small. By (ii) of Lemma 5.1, we can construct N_n' and N_n'' such that

$$P\left(N_n' \neq N_n''\,\Big|\, \lambda_1', \lambda_1''\right) \;\leq\; \kappa^{n-1}\,|\lambda_1' \;-\; \lambda_1''|.$$

If the event $\{N_n' = N_n''\}$ occurs, then (2.3) reduces to

$$|\lambda_{n+1}' \;-\; \lambda_{n+1}''| \;\leq\; \kappa_1\,|\lambda_n' \;-\; \lambda_n''|,$$

which allows us to construct the next pair (N_{n+1}', N_{n+1}'') such that

$$P\left(N_n' = N_n'', N_{n+1}' \neq N_{n+1}''\,\Big|\, \lambda_1', \lambda_1''\right) \;\leq\; \kappa_1\,\kappa^{n-1}\,|\lambda_1' \;-\; \lambda_1''|.$$

Continuing in the same way we arrive at

$$P\left(N_n' = N_n'', \dots, N_{n+k-1}' = N_{n+k-1}'', N_{n+k}' \neq N_{n+k}''\,\Big|\, \lambda_1', \lambda_1''\right) \;\leq\; \kappa_1^k\,\kappa^{n-1}\,|\lambda_1' \;-\; \lambda_1''|.$$

Hence, we finally obtain that

$$\begin{aligned} & P\left((N_n', N_{n+1}', \dots) \neq (N_n'', N_{n+1}'', \dots)\right) \\ & = \; P(N_n' \neq N_n'') + \sum_{k=1}^{\infty} P\left(N_n' = N_n'', \dots, N_{n+k-1}' = N_{n+k-1}'', N_{n+k}' \neq N_{n+k}''\right) \\ & \leq \; C_0\, \kappa^{n-1}/(1-\kappa_1), \end{aligned} \tag{5.10}$$

where $C_0 := E|\lambda_1' - \lambda_1''| \leq 2E\lambda_1 < \infty$. This yields, in conjunction with (5.9), assertion (i).

To see (ii), define the functions $f_1 = f$ and, for $d \geq 2$, $f_d(\lambda; n_1, \dots, n_d) = f_{d-1}(f(\lambda, n_d); n_1, \dots, n_{d-1})$, where $n_1, \dots, n_d \in \mathbb{N}_0$ and $\lambda \geq 0$. It is clear from (2.2) that

$$\lambda_t \;=\; f_d(\lambda_{t-d}; N_{t-1}, \dots, N_{t-d}).$$

It follows from (2.3) that

$$E\,|\lambda_t \,-\, f_d(0; N_{t-1}, \dots, N_{t-d})| \;\leq\; \kappa_1^d\, E\lambda_{t-d}.$$

Hence, as $d \to \infty$, $f_d(0; N_{t-1}, \dots, N_{t-d})$ converges in L_1 to λ_t. By taking an appropriate subsequence we also get almost sure convergence. This means that there exists a measurable function $f_\infty : \mathbb{N}_0^\infty \longrightarrow [0, \infty)$ such that

$$\lambda_t \;=\; f_\infty(N_{t-1}, N_{t-2}, \dots) \qquad \text{almost surely.} \tag{5.11}$$

By stationarity, (5.11) holds for all $t \in \mathbb{Z}$, which proves (ii).

To show (iii), we first recall the well-known fact that absolute regularity implies strong mixing, that is, it follows from (i) that

$$\alpha(n) \;=\; \sup_{A \in \mathcal{B}^N_{-\infty,0},\, B \in \mathcal{B}^N_{n,\infty}} |P(A \cap B) \,-\, P(A)P(B)| \underset{n\to\infty}{\longrightarrow} 0; \tag{5.12}$$

see e.g. Doukhan (1994, p. 20). Furthermore, strong mixing implies ergodicity, see e.g. Remark 2.6 on page 50 in combination with Proposition 2.8 on page 51 in Bradley (2007). Finally, we conclude from the representation (5.11) by Proposition 2.10(ii) in Bradley (2007, p. 54) that also the bivariate process $((N_t, \lambda_t))_{t\in\mathbb{Z}}$ is ergodic.

To prove (iv), we study the asymptotics of the process $((\widetilde{N}_t, \widetilde{\lambda}_t))_{t\in\mathbb{N}}$ obeying (2.1), (2.2) and (2.3) which is started with $\widetilde{\lambda}_1 \equiv 0$. We obtain from (2.4) and $E(N_t^2 \mid \lambda_t) = \lambda_t^2 + \lambda_t$ that

$$\begin{aligned} E(\widetilde{\lambda}_t^2 \mid \widetilde{\lambda}_{t-1}) \;&\leq\; E\left((f(0,0) \,+\, \kappa_1 \widetilde{\lambda}_{t-1} \,+\, \kappa_2 \widetilde{N}_{t-1})^2 \mid \widetilde{\lambda}_{t-1}\right) \\ &=\; \left(f(0,0) \,+\, \kappa \widetilde{\lambda}_{t-1}\right)^2 \,+\, \kappa_2^2 \widetilde{\lambda}_{t-1} \\ &\leq\; K_0 \,+\, \bar{\kappa} \widetilde{\lambda}_{t-1}^2, \end{aligned}$$

for any $\bar{\kappa} > \kappa$ and appropriate $K_0 = K_0(\bar{\kappa})$. We choose $\bar{\kappa} < 1$. Then we obtain that

$$E(\widetilde{\lambda}_3^2 \mid \widetilde{\lambda}_1) \;\leq\; K_0 \,+\, \bar{\kappa}\, E(\widetilde{\lambda}_2^2 \mid \widetilde{\lambda}_1) \;\leq\; K_0 \,+\, \bar{\kappa}\, (K_0 \,+\, \bar{\kappa} \widetilde{\lambda}_1^2).$$

Continuing in the same way we arrive at the inequality

$$E(\widetilde{\lambda}_t^2 \mid \widetilde{\lambda}_1) \;\leq\; K_0\, (1 \,+\, \bar{\kappa} \,+\, \cdots \,+\, \bar{\kappa}^{t-2}).$$

Since $\widetilde{\lambda}_t \stackrel{d}{\longrightarrow} \lambda_1$ we conclude from Theorem 5.3 in Billingsley (1968) that

$$E\lambda_1^2 \,\leq\, \liminf_{t\to\infty} E\widetilde{\lambda}_t^2 \,\leq\, K_0/(1 \,-\, \bar{\kappa}),$$

which proves (iii). □

Proof of Proposition 4.1. We will use the central limit theorem for martingale-difference arrays given on page 171 in Pollard (1984). We define the filtration $(\mathcal{B}_t)_{t\in\mathbb{N}}$ with $\mathcal{B}_t = \sigma(\lambda_1, N_1, \ldots, N_t)$, for $t = 0, 1, \ldots$, and we set $Z_t = (N_t - \lambda_t)^2 - N_t$. Since $N_t \mid \mathcal{B}_{t-1} \sim \mathrm{Poisson}(\lambda_t)$ we obtain that

$$E(Z_t \mid \mathcal{B}_{t-1}) \,=\, 0$$

and

$$E(Z_t^2 \mid \mathcal{B}_{t-1}) \,=\, 2\lambda_t^2.$$

Hence, it follows from (iii) of Theorem 3.1 that

$$\frac{1}{n}\sum_{t=1}^n E(Z_t^2 \mid \mathcal{B}_{t-1}) \stackrel{P}{\longrightarrow} 2E\lambda_1^2.$$

It remains to verify the conditional Lindeberg condition,

$$n^{-1}\sum_{t=1}^n E(Z_t^2 I(|Z_t/\sqrt{n}| > \epsilon \mid \mathcal{B}_{t-1})) \stackrel{P}{\longrightarrow} 0 \qquad \forall\epsilon > 0.$$

We have $E\left[n^{-1}\sum_{t=1}^n E(Z_t^2 I(|Z_t/\sqrt{n}| > \epsilon \mid \mathcal{B}_{t-1}))\right] \,=\, E[Z_1^2 I(|Z_1/\sqrt{n}| > \epsilon)]$, which tends to zero as $n \to \infty$ since $E\lambda_1^2 < \infty$ implies that $EZ_1^2 < \infty$. Hence, the conditional Lindeberg condition is also satisfied and the assertion follows from the CLT mentioned above. □

Proof of Lemma 4.1. Assume for the time being that $\|\widehat{\theta}_n - \theta_0\| \leq \delta$, which allows us to conveniently exploit the smoothness assumptions on f_θ. Then we obtain that

$$\begin{aligned}
|\widehat{\lambda}_2 \,-\, \lambda_2| \;&\leq\; \left|f_{\widehat{\theta}_n}(\widehat{\lambda}_1, N_1) \,-\, f_{\widehat{\theta}_n}(\lambda_1, N_1)\right| \,+\, \left|f_{\widehat{\theta}_n}(\lambda_1, N_1) \,-\, f_{\theta_0}(\lambda_1, N_1)\right| \\
&\leq\; \kappa_1\, |\widehat{\lambda}_1 \,-\, \lambda_1| \,+\, C\, \|\widehat{\theta}_n - \theta_0\|\, (\lambda_1 + N_1 + 1)
\end{aligned}$$

and

$$\begin{aligned}
|\widehat{\lambda}_3 \,-\, \lambda_3| \;&\leq\; \kappa_1\, |\widehat{\lambda}_2 \,-\, \lambda_2| \,+\, C\, \|\widehat{\theta}_n - \theta_0\|\, (\lambda_2 + N_2 + 1) \\
&\leq\; C\, \|\widehat{\theta}_n - \theta_0\|\, \{(\lambda_2 + N_2 + 1) \,+\, \kappa_1\, (\lambda_1 + N_1 + 1)\} \,+\, \kappa_1^2\, |\widehat{\lambda}_1 \,-\, \lambda_1|.
\end{aligned}$$

Continuing in the same way we arrive at

$$\begin{aligned}
&|\widehat{\lambda}_t \,-\, \lambda_t| \\
&\quad\leq\; C\, \|\widehat{\theta}_n - \theta_0\|\, \Big\{(\lambda_{t-1} + N_{t-1} + 1) \,+\, \kappa_1(\lambda_{t-2} + N_{t-2} + 1) \,+\, \cdots \,+\, \kappa_1^{t-2}\, (\lambda_1 + N_1 + 1)\Big\} \\
&\qquad\quad +\, \kappa_1^{t-1}\, |\widehat{\lambda}_1 \,-\, \lambda_1|, \qquad (5.13)
\end{aligned}$$

which yields that

$$\begin{aligned}
&(\widehat{\lambda}_t \,-\, \lambda_t)^2 \\
&\quad\leq\; 2\, C^2\, \|\widehat{\theta}_n - \theta_0\|^2\, \Big\{(\lambda_{t-1} + N_{t-1} + 1) \,+\, \kappa_1(\lambda_{t-2} + N_{t-2} + 1) \,+\, \cdots \,+\, \kappa_1^{t-2}\, (\lambda_1 + N_1 + 1)\Big\}^2 \\
&\qquad\quad +\, 2\, \kappa_1^{2t-2}\, (\widehat{\lambda}_1 \,-\, \lambda_1)^2
\end{aligned}$$

holds for all $t \geq 2$. Hence, we obtain under $\|\widehat{\theta}_n - \theta_0\| \leq \delta$ that

$$\sum_{t=1}^{n}(\widehat{\lambda}_t - \lambda_t)^2 \leq \frac{2}{1-\kappa_1^2}\left\{(\widehat{\lambda}_1 - \lambda_1)^2 + C^2\,\|\widehat{\theta}_n - \theta_0\|^2\,\left(\sum_{t=1}^{n-1}(\lambda_t + N_t + 1)\right)^2\right\}.$$

The right-hand side is bounded in probability, which proves the assertion. □

Proof of Theorem 4.1. We show that the difference between the test statistic $\widehat{T}_n$ and $T_{n,0}$ tends to zero in probability. This will yield the assertion by Proposition 4.1. We have that

$$\widehat{T}_n - T_{n,0} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\widehat{\lambda}_t - \lambda_t)^2 + \frac{2}{\sqrt{n}}\sum_{t=1}^{n}(N_t - \lambda_t)(\lambda_t - \widehat{\lambda}_t). \tag{5.14}$$

According to Lemma 4.1, the first term on the right-hand side converges to zero in probability. The estimation of the second one, however, is more delicate since $\widehat{\lambda}_t$ depends via $\widehat{\theta}_n$ on the whole sample which means that this term is not a sum of martingale differences. To proceed, we take first any *nonrandom* θ' with $\|\theta' - \theta_0\| \leq \delta$ and consider the intensity process given by $\lambda_1' = \widehat{\lambda}_1$ and, for $t = 2, \ldots, n$, $\lambda_t' = f_{\theta'}(\lambda_{t-1}', N_{t-1})$. We obtain in complete analogy to (5.13) that

$$\begin{aligned}&|\lambda_t' - \lambda_t| \\ &\quad \leq C\,\|\theta' - \theta_0\|\,\Big\{(\lambda_{t-1} + N_{t-1} + 1) + \kappa_1(\lambda_{t-2} + N_{t-2} + 1) + \cdots + \kappa_1^{t-2}\,(\lambda_1 + N_1 + 1)\Big\} \\ &\qquad + \kappa_1^{t-1}\,|\widehat{\lambda}_1 - \lambda_1|.\end{aligned}$$

Therefore, we obtain that

$$E\left[\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(N_t - \lambda_t)(\lambda_t - \lambda_t')\right| I(|\widehat{\lambda}_1 - \lambda_1| \leq M)\right] = O\left(\|\theta' - \theta_0\| + n^{-1/2}\right). \tag{5.15}$$

Since $\|\widehat{\theta}_n - \theta_0\| = O_P(n^{-1/2})$ it suffices to establish (5.15) on a sequence of grids $\mathcal{G}_n$ on the set $\{\theta \in \Theta : \|\theta - \theta_0\| \leq \epsilon_n^{-1} n^{-1/2}\}$, where $\mathrm{mesh}(\mathcal{G}_n) \leq \epsilon_n n^{-1/2}$, $\#\mathcal{G}_n \leq \epsilon_n n^{1/2}$, for some null sequence $(\epsilon_n)_{n\in\mathbb{N}}$. It follows from (5.15) that

$$\sup_{\theta' \in \mathcal{G}_n}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(N_t - \lambda_t)(\lambda_t - \lambda_t')\right| = O_P(\epsilon_n). \tag{5.16}$$

Moreover, for any value of $\widehat{\theta}_n$ with $\|\widehat{\theta}_n - \theta_0\| \leq \epsilon_n^{-1} n^{-1/2}$ we will find some $\theta' \in \mathcal{G}_n$ with $\|\widehat{\theta}_n - \theta'\| \leq \epsilon_n n^{-1/2}$. Since

$$\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(N_t - \lambda_t)(\lambda' - \widehat{\lambda}_t)\right| \leq \sqrt{\frac{1}{n}\sum_{t=1}^{n}(N_t - \lambda_t)^2}\,\sqrt{\sum_{t=1}^{n}(\lambda_t' - \widehat{\lambda}_t)^2} = o_P(1)$$

we obtain, in conjunction with (5.16), that the second term on the right-hand side of (5.14) is $o_P(1)$. This completes the proof. □

Acknowledgment . This work was supported by the German Research Foundation DFG, project NE 606/2-1. I thank Richard A. Davis and an anonymous referee for helpful comments on an earlier version of this paper. I also thank Paul Doukhan and Konstantinos Fokianos for stimulating discussions.

References

Andrews, D. W. K. (1984). Non-strong mixing autoregressive processes. *J. Appl. Probab.* **21**, 930–934.

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: Wiley.

Brännäs, K. and Johansson, P. (1994). Time series count data regression. *Comm. Statist. - Theory and Methods* **23**, 2907–2925.

Bradley, R. C. (2007). *Introduction to Strong Mixing Conditions, Volume I.* Kendrick Press.

Cameron, A. C. and Trivedi, P. K. (1986). Econometric models based on count data: Comparison and applications of some estimators and tests. *Journal of Applied Econometrics* **1**, 29–53.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments (with discussion). *Scandinavian Journal of Statistics* **8**, 93–115.

Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika* **70**, 269–274.

Davis, R. A., Dunsmuir, W. T. M., and Streett, S. B. (2003). Observation-driven models for Poisson counts. *Biometrika* **90**, 777–790.

Davis, R. A., Dunsmuir, W. T. M., and Wang, Y. (1999). Modeling time series of count data. In *Asymptotics, Nonparametrics and Time Series*, Ed. S. Ghosh, pp. 63–114. New York: Dekker.

Davis, R. A. and Wu, R. (2009). A negative binomial model for time series of counts. *Biometrika* **96**, 735–749.

Dean, C. B. (1992). Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association* **87**, 451–457.

Dean, C. and Lawless, J. F. (1989). Tests for detecting overdispersion in Poisson regression models. *Journal of the American Statistical Association* **84**, 467–472.

Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Review* **41**, 45–76.

Doukhan, P. (1994). *Mixing: Properties and Examples.* Lecture Notes in Statistics **85**, Springer-Verlag.

Durrett, R. (1991). *Probability. Theory and Examples.* Wadsworth.

Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH processes. *Journal of Time Series Analysis* **27**, 923–942.

Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association* **104**, 1430–1439.

Fokianos, K. and Tjøstheim, D. (2010). Log-linear Poisson autoregression. Preprint.

Grunwald, G. K., Hyndman, R. J., Tedesco, L., and Tweedie, R. L. (2000). Non-Gaussian conditional linear AR(1) models. *Australian and New Zealand Journal of Statistics* **42** (4), 479-495.

Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis.* Wiley, Hoboken, New Jersey.

Lasota, A. and Mackey, M. C. (1989). Stochastic perturbation of dynamical systems: The weak convergence of measures. *Journal of Mathematical Analysis and Applications* **138**, 232–248.

Lee, L.-F. (1986). Specification test for Poisson regression models. *International Economic Review* **27**, 689–706.

Meyn, S. P. and Tweedie, R. L (1993). *Markov Chains and Stochastic Stability.* Springer, London.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

Rosenblatt, M. (1980). Linear processes and bispectra. *J. Appl. Probab.* **17**, 265–270.

Rydberg, T. H. and Shephard, N. (2000). BIN models for trade-by-trade data. Modelling the number of trades in a fixed interval of time. World Conference Econometric Society, 2000, Seattle, Contributed Paper 0740.

Streett, S. (2000). Some observation driven models for time series of counts. Ph.D. Thesis, Colorado State University, Department of Statistics.

Zheng, H. and Basawa, I. V. (2008). First-order observation-driven integer-valued autoregressive processes. *Statistics and Probability Letters* **78**, 1–9.