Clemson University
TigerPrints

All Dissertations | Dissertations

8-2009

Integer-valued time series and renewal processes

Yunwei Cui
Clemson University, ycui@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

Part of the Statistics and Probability Commons

Recommended Citation
Cui, Yunwei, "Integer-valued time series and renewal processes" (2009). *All Dissertations*. 401.
https://tigerprints.clemson.edu/all_dissertations/401

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

Integer-Valued Time Series and Renewal Processes

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematical Sciences

by
Yunwei Cui
August 2009

Accepted by:
Dr. Robert B Lund, Committee Chair
Dr. Colin M Gallagher
Dr. Peter C Kiessler
Dr. Karunarathna B Kulasekera

Abstract

This research proposes a new but simple model for stationary time series of integer counts. Previous work in the area has focused on mixture and thinning methods and links to classical time series autoregressive moving-average difference equations; in contrast, our methods use a renewal process to generate a correlated sequence of Bernoulli trials. By superpositioning independent copies of such processes, stationary series with binomial, Poisson, geometric, or any other discrete marginal distribution can be readily constructed. The model class proposed is parsimonious, non-Markov, and readily generates series with either short or long memory autocovariances. The model can be fitted with linear prediction techniques for stationary series. Estimation of process parameters based on conditional least squares methods is considered. Asymptotic properties of the estimators are derived. The models sometimes have an autoregressive moving-average structure and we consider the AR(1) count process case in detail. Unlike previous methods based on mixture and thinning tactics, series with negative autocorrelations can be produced.

Dedication

To my family.

Acknowledgments

I am grateful to my dissertation committee for providing tremendous help and guidance during the process of completing this dissertation. My deep gratitude goes to my dissertation committee chair, Dr. Robert Lund. Your knowledge and insights in the area of time series modeling and analysis helped me choose this fruitful topic. Your suggestions and encouragements guided me through the most difficult time of this research when there seemed to be no progress ever made and the oral exam was impending. Your office door is always open to me. No matter how busy you are, you always can find time to meet me as long as you are on campus. Your directions and insightful suggestions have significantly improved the quality of this work. Moreover, Dr. Lund, you patiently helped me in writing and mentored me how to be a better teacher and improve my teaching evaluation.

To Dr. Gallagher whose valuable comments have improved the overall quality of the research. Your teaching style and method greatly impressed me and will definitely help me improve my class delivery on those subject which could be better explained through examples.

To Dr. Kiessler whose classes first exposed me to Markov Chain and probability theory. I was inspired to take several advanced courses which contributed to my decision to choose time series analysis and related field as the focus of my Ph.D program. You are always willing to help me solve difficult probability problems. You also kindly pointed out the deficiencies in my presentation and encouraged to improve my presentation skills.

To Dr. K.B. whose mathematical statistics courses are among the most important courses I ever took at Clemson. You keep high standards of your teaching. Your advice is valuable to anyone who wants to be successful in his Ph.D. program.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Integer-valued time series arise in many practical settings. Counts of objects or occurrences of events taken sequentially in time have such structure; for examples, consider the annual counts of hurricanes, the number of rainy days in successive weeks, the number of patients treated each day in an emergency department, the number of U.S. soldiers injured in Iraq in each month, or the daily counts of new swine flu cases in Mexico. McKenzie [2003] and Fokianos and Kedem [2003] provide recent overviews.

Count series are non-negative integers and are usually correlated over time. They cannot be well approximated by continuous variables, especially when the counts are relatively small (Brockwell and Davis, 2002, Chapter 8). Modeling and analyzing count series remains one of the most challenging and undeveloped areas of time series analysis. For example, developing a non-negative integer-valued time series with the autocorrelation function $\rho(h) = \phi^h$ with $0 < \phi < 1$ is not easy.

Since the late 1970s, many authors have investigated ways to model count series with a preset marginal distributions. Many of the results are based on classical autoregressive moving-average (ARMA) methods. While mathematically innovative, these models have some unresolved shortcomings. New models continue to emerge as the area is still in its infancy (McKenzie, 2003). To acquaint the reader with this area, we next review classical count models.

1.1 DARMA Models

The discrete autoregressive moving-average (DARMA) model (Jacobs and Lewis, 1978a, 1978b, 1983) represents the first attempt to define a general stationary series of counts. The simplest form of the model, the first order discrete autoregession (DAR(1)) process (Jacobs and Lewis, 1978a, 1978b), is based on mixing copies of random variables having the prescribed marginal distribution. The difference equation governing a DAR(1) model is

$$X_t = V_t X_{t-1} + (1 - V_t)A_t, \qquad t \geq 1, \tag{1.1}$$

where $\{V_t\}$ is a sequence of independent and identically distributed Bernoulli random variables with $P(V_t = 1) = \zeta \in (0, 1)$, and $\{A_t\}$ is a sequence of independent and identically distributed integer-valued random variables having the prescribed marginal distribution (call this π). If X_0 has distribution π, then it is easy to see that $\{X_t\}$ is stationary and has marginal distribution π. Moreover, the lag h autocorrelation of $\{X_t\}$ is $\rho(h) = \zeta^h$ for $h \geq 0$. In fact, $P(A_t = A_{t-1}) = \zeta$. Note that DAR(1) autocorrelations must be non-negative since $\zeta \in (0, 1)$.

By introducing another random sequence $\{Z_t\}$ into (1.1), one can define the DAR(p) model:

$$X_t = V_t X_{t-Z_t} + (1 - V_t)A_t, \qquad t \geq 1,$$

where $\{A_t\}$ and $\{V_t\}$ are defined as before, and $\{Z_t\}$ are independent and identically distributed random variables taking values in $\{1, 2, \ldots, p\}$. Let $\phi_k = P(Z_t = k)$. If $\{X_0, \ldots, X_{-p+1}\}$ are independent and identically distributed with distribution π, then $\{X_t\}$ is stationary and also has marginal distribution π. Its autocorrelation function satisfies the AR(p)-type Yule-Walker equation

$$\rho_X(k) = \zeta \sum_{i=1}^{p} \phi_i \rho_X(k-i),$$

where $\rho_X(k) = \text{Corr}(X_t, X_{t+k})$ for $k \geq 0$.

The DARMA(p, q) process is built with two component models:

$$\begin{aligned} X_t &= V_t Y_{t-q} + (1 - V_t) A_{t-S_t}, \quad t \geq 1; \\ Y_t &= U_t Y_{t-Z_t} + (1 - U_t) A_t, \quad t \geq -q+1, \end{aligned}$$

where $\{A_t\}$ and $\{Z_t\}$ are as above, $\{U_t\}$ is a sequence of independent and identically distributed Bernoulli random variables, and $\{S_t\}$ are independent and identically distributed random variables taking values in $\{1, 2, \ldots, q\}$. If $\{Y_{-q}, \ldots, Y_{-q-p+1}\}$ are independent and identically distributed with distribution π, then $\{X_t\}$ for $t > 1$ is stationary with marginal distribution π and has an ARMA(p, q) type autocorrelation structure (Jacobs and Lewis, 1983) .

While the DAMA model generates integer-valued time series with any preset distribution π, it is handicapped by the fact that for high correlations, the model tends to generate series that have runs of constant values. This is not realistic in practice since time series of counts usually exhibit variability over time. As a result, DARMA models are rarely used in practice (MacDonald and Zucchini, 1997).

1.2 INARMA Models

Integer-valued autoregressive moving-average (INARMA) models were first proposed independently by Mckenzie [1986] and Al-Osh and Alzaid [1987]. INARMA models are based on the binomial thinning operator $\circ$ (Steutel and Van Harn, 1979), which combines a non-negative integer valued random variable N and a probability p via

$$p \circ N = \sum_{j=1}^{N} B_j.$$

Here, $\{B_j\}$ is a sequence of independent Bernoulli random variables with success probability p. Hence, $p \circ N$ is the sum of N independent Bernoulli random variables, each of which is unity with probability p.

The simplest model in the INARMA class is the integer-valued first order autoregressive (INAR(1)) process $\{X_t\}$ (McKenzie, 1985, 1988; Al-Osh and Alzaid, 1987). This model obeys an AR(1) difference equation with a thinning operator:

$$X_t = \phi \circ X_{t-1} + Z_t, \qquad 0 < \phi < 1, \tag{1.2}$$

where $\{Z_t\}$ is a sequence of independent and identically distributed count random variables. It can be shown that the autocorrelation function of $\{X_t\}$ at lag h is $\rho(h) = \phi^h$. But since ϕ must be positive in the above, only positively correlated count series can be produced.

Different marginal distributions of Z_t give rise to different marginal distributions of $\{X_t\}$. McKenzie [1986] [1988] used (1.2) to generate integer-valued count series with Poisson, negative binomial, and geometric marginal distributions. To generate a stationary sequence $\{X_t\}$ with a certain prescribed marginal distribution, one can first derive the generating function of Z_t through (1.2) and then invert this generating function to obtain the distribution of Z_t. For example, to generate a series $\{X_t\}$ with Poisson marginals, $\{Z_t\}$ should be chosen to have a Poisson distribution (McKenzie, 1986). It can be shown that if $\{Z_t\}$ is a sequence of independent and identically distributed Poisson random variables with mean $\theta(1-\phi)$ $(\theta > 0)$ and X_0 is Poisson, mean θ, then $\{X_t\}$ generated through (1.2) is stationary with Poisson marginals, with mean θ and the $AR(1)$ type autocorrelation structure $\rho(h) = \phi^h$.

Not all marginal distributions can be constructed with (1.2). In fact, only random variables with the so-called discrete self-decomposable distributions of Steutel and Van Harn [1979] can be solutions to equation (1.2). One implication of this is that stationary count series with binomial marginal distributions cannot be generated, because the binomial distribution is not self-decomposable.

In (1.2), the thinning operation $\circ$ replaces scalar multiplication in the classical AR(1) process. Extensions of the method to other ARMA models are possible. For example, the Poisson INMA(q) process by McKenzie [1988] satisfies

$$X_t = Z_t + \theta_1 \circ Z_{t-1} + \cdots + \theta_q \circ Z_{t-q}, \qquad 0 < \theta_i < 1, \qquad i = 1, \ldots, q,$$

where $\{Z_t\}$ is a sequence of independent and identically distributed Poisson random variables. Let $\theta = 1 + \sum_{i=1}^{q} \theta_i$. If Z_t has mean λ/θ and all thinning operations are independently performed, then $\{X_t\}$ generated by the above model has Poisson marginals with mean λ and the classical moving-average (MA(q)) type autocorrelation structure

$$\begin{aligned} \rho(0) &= \sum_{i=0}^{q-h} \theta_i \theta_{i+h}/\theta, \qquad h = 1, \ldots, q; \\ \rho(h) &= 0, \quad h > q. \end{aligned}$$

Al-Osh and Alzaid [1990] and Du and Li [1991] proposed the INAR(p) process

$$X_t = \phi_1 \circ X_{t-1} + \cdots + \phi_p \circ X_{t-p} + Z_t, \quad 0 < \phi_i < 1, \quad i = 1, \ldots, p,$$

where $\{Z_t\}$ is a sequence of independent and identically distributed count random variables. The marginal distributions of $\{X_t\}$ are difficult to identify in the above model. By Al-Osh and Alzaid's assumption that given X_t, the thinned vector $\{\phi_1 \circ X_t, \ldots, \phi_p \circ X_t\}$ has a multinomial distribution with parameters $(\phi_1, \ldots, \phi_p, X_t)$, $\{X_t\}$ is stationary with an ARMA(p, p-1) autocorrelation structure. However, by the assumption of Du and Li [1991] that the thinning operations $\{\phi_1 \circ X_t, \ldots, \phi_p \circ X_t\}$ are performed independently, $\{X_t\}$ is stationary with an AR(p) autocorrelation structure. Most authors use Du and Li's assumption [1991].

Gauthier and Latour [1994] extended the thinning operator $\circ$ to the generalized thinning operation $\bullet$ defined

$$a \bullet N = \sum_{j=1}^{N} X_j, \quad a > 0,$$

where N is a non-negative integer-valued random variable and $\{X_j\}$ is a sequence of independent and identically distributed non-negative integer-valued random variables with mean a and finite variance. Gauthier and Latour [1994] and Latour [1997] [1998] extended INAR(p) processes to the so-called generalized integer-valued autoregressive (GINAR(p)) process by substituting $\bullet$ for $\circ$ in the INAR(p) model of Du and Li [1991].

Estimation methods for INARMA models have also been studied. Al-Osh and Alzaid [1987] and Ronning and Jung [1992] investigate maximum likelihood methods in the Poisson INAR(1) model. Al-Osh and Alzaid [1987] and Du and Li [1991] studied conditional least squares estimation methods (Klimko and Nelson, 1978) in the Poisson INAR(1) and INAR(p) models. Brännäs [1994] proposed generalized method of moments estimators for the Poisson INAR(1) model.

Applications of INARMA models are frequent; for example, the study of epileptic seizure counts (Franke and Seligmann, 1993) and applications to economics (Brännäs and Hellström, 2001; Brännäs and Shahiduzzaman, 2004; Böckenholt, 1999b; Böckenholt, 2003; Rudholm, 2001; Freeland and McCabe, 2004).

1.3 Regression Models

Zeger [1988] proposed an important Poisson-based model to incorporate covariates (regressors) into the analysis of time series of counts. Given values for a stationary series $\{\epsilon_t\}$, $\{Y_t\}$ is assumed to be an independent sequence of counts with Poisson distributions having the conditional moments

$$\mathrm{E}(Y_t|\epsilon_t) = \exp(D_t'\beta)\epsilon_t \quad \text{and} \quad \mathrm{Var}(Y_t|\epsilon_t) = \exp(D_t'\beta)\epsilon_t,$$

where $\{D_t\}$ is a $p \times 1$ vector of covariates and β is a $p \times 1$ vector of unknown coefficients to be estimated. If $\{\epsilon_t\}$ has mean $\mathrm{E}(\epsilon_t) = 1$ and autocovariance $\mathrm{Cov}(\epsilon_t, \epsilon_{t+h}) = \sigma^2 \rho_\epsilon(h)$, then $\{Y_t\}$ has

$$\begin{aligned} \mu_t &= \mathrm{E}(Y_t) = \exp(D_t'\beta); \\ \upsilon_t &= \mathrm{Var}(Y_t) = \mu_t + \sigma^2 \mu_t^2; \\ \rho_Y(t,h) &= \mathrm{Corr}(Y_t, Y_{t+h}) = \frac{\rho_\epsilon(h)}{[1 + (\sigma^2\mu_t)^{-1}]^{\frac{1}{2}}[1 + (\sigma^2\mu_{t+h})^{-1}]^{\frac{1}{2}}}. \end{aligned}$$

The correlations of $\{Y_t\}$ are determined by the latent process $\{\epsilon_t\}$. In general, regression models for Poisson time series of counts are not stationary. A quasilikelihood method was used to estimate parameters in this model.

Campbell [1994] extended Zeger's work to higher orders of dependence and studied the occurrences of sudden infant death syndrome. Brännäs and Johansson [1994] also studied the model, developing Poisson pseudomaximum likelihood estimation methods.

Zeger and Qaqish [1988] also designed a model based on the assumption that the past history of $\{Y_t\}$ affects Y_t only through the values of $Y_{t-1}, \ldots, Y_{t-q}$. In the literature, this model is also called a Markov regression model. Let $\{D_t\}$ be as before, H_t be the history of the covariates and $\{Y_t\}$ through time t—say

$$H_t = \{D_t, D_{t-1}, \ldots, D_0; Y_{t-1}, Y_{t-2}, \ldots, Y_0\}.$$

Define $\mu_t = E(Y_t|H_t)$ and $V_t = \mathrm{Var}(Y_t|H_t)$. Conditional on H_t, the model assumes

$$\begin{aligned} \log(\mu_t) &= D_t'\beta + \sum_{i=1}^{q} \theta_i[\ln(Y_{t-i}^*) - D_{t-i}'\beta]; \\ V_t &= \phi\mu_t, \end{aligned} \tag{1.3}$$

where $Y_{t-i}^* = \max(Y_{t-i}, c)$, $0 < c < 1$. An alternative model replaces Y_{t-i} by $Y_{t-i}^* = Y_{t-i} + c$, $c > 0$. Some transformation is necessary since $\log(Y_{t-i})$ is undefined when $Y_{t-i} = 0$, a possible value in the Poisson support set.

Quasilikelihood methods were originally used to estimate β. If additional assumptions about the conditional distribution of Y_t are available, likelihood estimation can be conducted. Fahrmeir and Tutz [1994] applied the model to analyze the monthly number of polio cases in the United States; Cameron and Leon [1993] used the model to investigate the monthly number of strikes in the United States. But since the transformation of Y_{t-i} to Y_{t-i}^* is ad hoc and its impact on μ_t is hard to evaluate, the model is not widely used today (MacDonald and Zucchini, 1997).

More generally, (1.3) has the form

$$\begin{aligned} h(\mu_t) &= D_t'\beta + \sum_{i=1}^{q} \theta_i f_i(H_t); \\ V_t &= g(\mu_t)\phi, \end{aligned} \tag{1.4}$$

where $h(\mu_t)$ is called a 'link' function and the $f_i(H_t)'s$ are functions based on the process history. The model in (1.4) is very flexible. An autoregressive model of order q (AR(q)) is obtained when $\{Y_t\}$ is Gaussian, $h(\mu_t) = \mu_t$, $g(\mu_t) = 1$, and $f_i(H_t) = Y_{t-i} - D_{t-i}'\beta$. Also, trend and seasonality can be added to the model by incorporating a trend function of t and *sine* and *cosine* terms into $h(\mu_t)$ and $f_i(H_t)$.

Li [1991] proposed two methods of assessing the adequacy of the above model. Li [1994] also extended Markov regression models by adding autoregressive and moving-average terms. Albert [1994] used the model for magnetic resonance imaging. It is important to note that Poisson marginal distributions are assumed in the above models. It may not be possible to construct explicit covariance structures and have any marginal distributional type desired. In fact, this point is key in what follows.

1.4 Hidden Markov Models

In a hidden Markov model, a count series $\{Y_t\}$ is affected by the value of an unobservable series $\{C_t\}$. It is assumed that C_t takes values in the set $\{1, 2, \ldots, m\}$ and evolves according to a strictly positive Markov chain with transitional and stationary distributions

$$\gamma_{ij} = P(C_t = j | C_{t-1} = i) \quad \text{and} \quad \delta_j = P(C_t = j), \quad j = 1, 2, \ldots, m.$$

In the simplest case, a Poisson Markov model, it is assumed that conditional on $C_t = j$ and D_t (a $p \times 1$ vector of covariates), that Y_t has Poisson distribution with mean

$$\mu_{tj} = \exp(D_t'\beta_j),$$

where β_j is a $p \times 1$ vector of unknown coefficients to be estimated. Then Y_t has the conditional moments

$$\begin{aligned} E(Y_t|D_t) &= \sum_{j=1}^{m} \delta_j \mu_{tj}; \\ E(Y_t^2|D_t) &= \sum_{j=1}^{m} \delta_j (\mu_{tj} + \mu_{tj}^2); \\ E(Y_t Y_{t+k}|D_t) &= \sum_{i=1}^{m} \sum_{j=1}^{m} \delta_i \gamma_{ij}(k) \mu_{ti} \mu_{t+k,j}, \end{aligned}$$

where $\gamma_{ij}(k) = P(C_{t+k} = j|C_t = i)$. The parameters of the model, β_j and γ_{ij}, can be estimated by maximum likelihood. MacDonald and Zucchini [1997] applied this model to describe the daily number of epileptic seizures of a particular patient. More details about this class of models can be found in MacDonald and Zucchini [1997].

1.5 State-Space Models

In the state-space model for a count series $\{Y_t\}$, Y_t is specified to have a certain distribution based on a state variable X_t. The state variable X_t evolves stochastically according to a particular distribution determined by X_t and Y_{t-1}. Simple state-space models are discussed in Brockwell and Davis [2002, Chapter 8]. For this model, Y_t is assumed to be Poisson with mean $\exp(X_t)$ with

$$P(Y_t = k|X_t) = \frac{e^{-e^{X_t}} e^{kX_t}}{k!}.$$

Moreover, X_t is assumed to obey a regression model with Gaussian noise

$$X_t = \beta' \mu_t + W_t,$$

where μ_t is a $p \times 1$ vector of covariates and β is a $p \times 1$ vector of unknown coefficients to be estimated. The noise term W_t is defined to be $AR(1)$:

$$W_t = \phi W_{t-1} + Z_t, \qquad \{Z_t\} \sim \text{IID N}(0, \sigma^2).$$

Then, conditional on X_t, X_{t+1} has a normal distribution with mean $\mu'_{t+1}\beta + \phi(X_t - \beta'\mu_t)$ and variance σ^2. Estimation of $\theta = (\beta', \phi, \sigma^2)$ can be conducted by maximum likelihood techniques. Brockwell and Davis [2002] fit the model to the polio data of Zeger [1988].

Several generalizations of basic state space models can be made. One approach builds the model in a Bayesian framework. The most mathematically tractable model is given by Harvey [1989]. In Harvey's model, Y_t, conditional on μ_t, has a Poisson distribution with

$$P(Y_t = k | \mu_t) = \frac{e^{-\mu_t} \mu_t^k}{k!},$$

and μ_t (conditional on Y_{t-1}) is taken to have a gamma distribution with density

$$f(\mu_t | Y_{t-1}) = \frac{b_t^{a_t} x^{a_t - 1} e^{-b_t \mu_t}}{\Gamma(a_t)},$$

where a_t and b_t is given by

$$a_t = \omega a_{t-1}, \qquad b_t = \omega b_{t-1}.$$

Then given the observation of Y_t, the posterior distribution for μ_t is gamma with density

$$f(\mu_t|Y_t) = \frac{b^a x^{a-1} e^{-b\mu_t}}{\Gamma(a)}, \quad a = a_t + Y_t, \quad b = b_t + 1.$$

The conditional distribution of Y_t given Y_{t-1} is negative binomial with parameters a_t and b_t. Estimation of ω can be conducted with likelihood techniques. Harvey and Fernandes [1989] used this model to analyze count series data on scores in soccer games, fatalities of van drivers, and purse snatchings in Chicago. Johansson [1996] applied this model to Swedish traffic accident fatalities.

So far, the five major model classes for count serries have been briefly introduced. DARMA and INARMA are the two major types of models for stationary count series. The remaining three model classes are generally not stationary as the observation at time t, Y_t, is assumed to be affected by some latent processes that evolves stochastically. The nuances of the later three model types lie with the assumptions about the mechanisms that govern the latent processes and the conditional distributions of Y_t. The simplest cases of these models are usually based on the assumption of a linear relationship between the conditional mean of Y_t and the latent process.

Complete overviews of models for integer-valued time series can be found in McKenzie [2003], Chapter 7 of MacDonald and Zucchini [1997], and Chapter 1 of Cameron and Trivedi [1998]. The literature for stationary time series models with non-Gaussian marginals is by now vast. Besides count series, stationary series with exponential marginals (Lawrance and Lewis, 1977a, 1977b), gamma marginals (Al-Osh and Alzaid, 1993), multinomial marginals (Böckenholt, 1999a), binomial marginals (Weiβ, 2009), and conditional exponential family marginals (Benjamin et al., 2003) exist.

1.6 Outline of the Thesis

The rest of this document proceeds as follows. In Chapter 2, a completely new model for stationary count series will be introduced that is based upon renewal processes. Here, we first introduce notation and review simple renewal processes. Section 2.2 establishes some general time series properties of the model. Sections 2.3 considers how to construct binomial, Poisson, and geometric marginal distributions, respectively. The last section contains proofs and technical derivations.

Chapter 3 studies renewal AR(1) count series. Section 3.1 establishes general results about the AR(1) count process. Section 3.2 considers estimation issues, and Section 3.3 gives examples and simulation results. Section 3.4 concludes with proofs and technical details.

Chapter 2

A New Look at Time Series of Counts

This chapter introduces a new model class to describe stationary integer count time series. The model class is simple, parsimonious, non-Markov, and easily generates all classical discrete marginal distributions for counts such as binomial, Poisson, and geometric. The model can readily produce either long or short memory series.

In Chapter 1, two types of general models for stationary count series were presented. The discrete autoregressive moving-average (DARMA) model can generate a stationary integer-valued time series with any prescribed marginal distribution; however, as noted by McKenzie [1985] [2003], these series tend to have sample paths that are constant for long runs, a trait generally not seen in data. The integer-valued autoregressive moving-average (INARMA) model uses a thinning operation in a difference equation scheme that mimics autoregressive moving-average methods to generate stationary integer-valued series that have negative binomial, geometric, and Poisson marginal distributions. Unlike discrete autoregressive moving-average methods, thinning techniques cannot produce an arbitrary marginal count distribution.

However, the sample paths of thinned models seem to be more data-realistic than those of discrete autoregressive moving-average models.

Here, we take a completely different approach to the problem. We regard the marginal distribution as known and study methods that generate a stationary series with the known marginal distribution. Our methods use a simple on/off renewal process to generate a correlated sequence of Bernoulli trials. By superpositioning independent copies of such processes, we will be able to construct series with any of the classical count marginal distributions in an efficient manner. In fact, since a draw from any discrete distribution can be constructed from a sequence of independent coin tosses, the methods can generate discrete series with any specified marginal distribution. By selecting the renewal lifetimes to have an infinite second moment, long memory count series are obtained. Long memory count series cannot be produced with non-unit root autoregressive moving-average difference equations and thinning methods. By linking time series and renewal processes, several short proofs of classical renewal results are obtained; these are pointed out in §2.2.

The only other paper linking count series to renewal processes seems to be Blight [1989], who focuses on autoregressive moving-average structures of renewal processes. It is also noted that Yule's original formulation of autoregressive models involved pea-shooters and point processes. An autoregressive moving-average slant is not pursued here for two reasons. First, because our model's parameters are limited to those governing the renewal interarrival distribution, the class is naturally parsimonious and the parsimonizing effects of an autoregressive moving-average structure are not needed. Second, as we show in §2.4 , our model can be fitted via general linear prediction techniques for stationary series. Fokianos and Kedem [2003] discuss general inference methods for time series of counts; again, an autoregressive moving-average structure is not needed. This said, we state that some, but not all, of the series constructed below indeed obey an autoregressive moving-average difference equation. We prove that our renewal series are not Markov in general and demonstrate through examples that their autocovariance structures can be intricate. For feel, sample paths of several count series and their sample autocorrelations and partial autocorrelations are provided.

2.1 The Renewal Process Building Block

This section establishes notation and the simple renewal process that we use to build our model. Let $f_n = P(L = n)$ be the distribution of a random variable L taking values in $\{1, 2, \ldots\}$ with $f_1 < 1$. Later, L will also be called a lifetime. Let L_0, L_1, $L_2, \ldots$ be independent nonnegative integer-valued random variables with L_i distributed as L for all $i \geq 1$; notice that we allow L_0 to have a different distribution than the rest of the L_i's. If $L_0 + L_1 + \cdots + L_k = n$ for some $k \geq 0$, then a renewal is said to have taken place at time n. If $L_0 \equiv 0$, the process is called non-delayed; otherwise, it is called delayed.

In the non-delayed situation, let u_n be the probability that a renewal occurs at time n. Then u_n satisfies $u_n = \sum_{k=1}^{n} u_{n-k} f_k$ for $n \geq 1$ with $u_0 = 1$. In the delayed case, let w_n be the probability of a renewal at time n. Conditioning on L_0 gives $w_0 = b_0$ and $w_n = \sum_{k=0}^{n} b_k u_{n-k}$ for $n \geq 1$, where $b_n = P(L_0 = n)$ is the distribution of the first lifetime. When L is non-lattice and L has finite mean, which we henceforth assume, $w_n \to E[L]^{-1} = \mu^{-1}$ as $n \to \infty$ (Feller, 1968, chapter XIII). If $b_n = \mu^{-1} P(L > n)$ for $n \geq 0$, the so-called first derived distribution of L, then the delayed process is stationary in that $w_n \equiv \mu^{-1}$ (Feller, 1968).

In a stationary renewal process, define $X_t = 1$ if a renewal occurs at time t;

otherwise, take $X_t = 0$. Then

$$\begin{aligned}
P(X_{1,t} = 1, X_{1,t+h} = 0) &= P(X_{1,t} = 0, X_{1,t+h} = 1) = \mu^{-1}(1 - u_h);\\
P(X_{1,t} = 0, X_{1,t+h} = 0) &= 1 - 2\mu^{-1} + u_h\mu^{-1};\\
P(X_{1,t} = 1, X_{1,t+h} = 1) &= \mu^{-1}u_h, \qquad (2.1)
\end{aligned}$$

where u_h is the non-delayed renewal probability at time h. Since $E(X_t) = E(X_{t+h}) = \mu^{-1}$, we see that $\{X_t\}$ is second-order stationary with

$$\gamma(h) = \text{cov}(X_t, X_{t+h}) = \mu^{-1}(u_h - \mu^{-1}). \qquad (2.2)$$

2.2 General Results

The fact that $\gamma(\cdot)$ is a stationary autocovariance function has immediate implications. For example, time series theory provides the following result.

THEOREM 1. *If $E(L) < \infty$, then the $n \times n$ renewal matrix U_n with (i, j)th entry $(U_n)_{i,j} = u_{|i-j|}$ is invertible for every $n \geq 1$.*

The proof of this and all subsequent results are in the §2.6.

A time series with autocovariance function $\gamma(\cdot)$ is said to have long memory if $\sum_{h=0}^{\infty} |\gamma(h)| = \infty$ and short memory if the sum is finite. It is easy to generate long memory series with this model. In fact, we offer the following result.

THEOREM 2. *If $E(L) < \infty$, then $\{X_t\}$ has long memory if and only if $E(L^2) = \infty$.*

More can be said about the convergence speed of $\gamma(h)$ to zero as $h \to \infty$. By (2.2), the convergence rate of $\gamma(h)$ to zero is the same as the convergence rate of u_h to

μ^{-1}. The latter problem has been extensively studied (Pitman, 1974; Lindvall, 1992; Hansen and Frenk, 1991; and Berenhaut and Lund 2001). For example, if $E(L^r) < \infty$ for some $r \geq 2$, then it is known that $n^{r-1}(u_n - \mu^{-1}) \to 0$ as $n \to \infty$. If L has a finite generating function in that $E(r^L) < \infty$ for some $r > 1$, then $\gamma(h)$ decays to zero geometrically in that $|\gamma(h)| \leq \kappa s^{-h}$ for some $\kappa < \infty$ and some $s > 1$, implying a short memory autocovariance (Kendall, 1959).

In general, $\{X_t\}$ will not be a Markov chain. The following result gives necessary and sufficient conditions for $\{X_t\}$ to be Markov.

THEOREM 3. *The series* $\{X_t\}$ *is Markov if and only if* L *has a constant hazard rate after lag* 1; *that is,* $h_k = P(L = k \mid L \geq k)$ *is constant over* $k \geq 2$.

From independent copies of the above Bernoulli processes, we will easily be able to construct time series with the classical count distributions. For a nonnegative integer $M \geq 1$, consider $\{Y_t\}$ defined by

$$Y_t = \sum_{i=1}^{M} X_{i,t}, \qquad n \geq 0, \tag{2.3}$$

where $\{X_{i,t}\}_{i=1}^{M}$ are independent copies of $\{X_t\}$. Then $\{Y_t\}$ is stationary and Theorem 3 can be seen to apply without modification.

THEOREM 4. *The series* $\{Y_t\}$ *is Markov if and only if* L *has a constant hazard rate after lag* 1.

For notation, let $\alpha = 1 - 2\mu^{-1} + \mu^{-1}u_h$, $\beta = \mu^{-1}(1 - u_h)$, and $\nu = \mu^{-1}u_h$. Use (2.1) to get that Y_t and Y_{t+h} have the joint generating function

$$E[s_1^{Y_t} s_2^{Y_{t+h}}] = \{\alpha + \beta(s_1 + s_2) + \nu s_1 s_2\}^M, \tag{2.4}$$

and probability distribution

$$P(Y_t = i, Y_{t+h} = j) = \sum_{\ell=\max(0,i+j-M)}^{\min(i,j)} \frac{M!\alpha^{M+\ell-i-j}\beta^{i+j-2\ell}\nu^{\ell}}{\ell!(i-\ell)!(j-\ell)!(M+\ell-i-j)!}. \tag{2.5}$$

This joint distribution has been called the bivariate binomial distribution (Kocherlakota and Kocherlakota, 1992). The conditional distribution of Y_{t+h} given $Y_t = i$ is

$$P(Y_{t+h} = j | Y_t = i) = \sum_{\ell=\max(0,i+j-M)}^{\min(i,j)} \begin{pmatrix} i \\ \ell \end{pmatrix} \begin{pmatrix} M-i \\ j-\ell \end{pmatrix} \alpha^{M+\ell-i-j}\beta^{i+j-2\ell}\nu^{\ell}\mu^{i}(1-1/\mu)^{i-M}. \tag{2.6}$$

In general, it is not possible to write our model with the classical thinning operator $\circ$ as defined in McKenzie [1986]. To see this on a superficial level, note that a thinning based model cannot produce series with negative autocorrelations since all thinning probabilities must be nonnegative. However, it is easy to have a negative lag one autocorrelation in our model: simply choose a renewal lifetime where $u_1 < \mu^{-1}$. To explore the issue more deeply, consider the simple case where $M = 1$ and the renewal process is non-delayed. Then Y_t is either zero or one for each t. For a fixed time t, let $S_{N(t-1)}$ be the time of the most recent renewal prior to time t. Then a renewal happens at time t if and only if the item put in use at time $S_{N(t-1)}$ lasts exactly $t - S_{N(t-1)}$ time units; this statement is conditional upon the event that this item lasts more than $t - 1 - S_{N(t-1)}$ time units. Since $Y_{S_{N(t-1)}} = 1$, we have

$$Y_t = h_{t-S_{N(t-1)}} \circ 1 = h_{t-S_{N(t-1)}} \circ Y_{S_{N(t-1)}},$$

where h_k is the hazard rate of L at index k. Because $\{t - S_{N(t-1)}\}$ is random with time-varying dynamics, one cannot work the above equation into a difference scheme of finite order with non-random coefficients. Introduction of $M > 1$ does not simplify the issue.

This said, the model can be connected to the thinning operator in special cases. For example, if L has a constant hazard rate after lag 1, then the process is Markov by Theorem 4, and summing over all components gives

$$Y_t = c_1 \circ Y_{t-1} + c_2 \circ (M - Y_{t-1}),$$

where $c_1 = P(X_{i,n} = 1 \mid X_{i,n-1} = 1)$ and $c_2 = P(X_{i,n} = 1 \mid X_{i,n-1} = 0)$. Explicit expressions of c_1 and c_2 are derived in §2.6. This yields the stationary binomial series considered in McKenzie [1985]. Other generalities are possible when h_k is constant for k larger than some prescribed constant.

Likewise, few of our models obey autoregressive moving-average recursions. For a case where a first order autoregressive structure does arise, suppose that L has a constant hazard rate after lag 1. Then the lifetime probabilities can be expressed as $f_1 = 1 - f_2/(1-r)$ and $f_n = f_2 r^{n-2}$ for $n \geq 2$, where $r < 1$, and $f_1 < 1$. From Theorem 4, we know that $\{Y_t\}$ is Markov. Hence, $E(Y_n \mid Y_{t-1}, \ldots, Y_0) = E(Y_t \mid Y_{t-1})$. Since the joint distribution of (Y_t, Y_{t-1}) is bivariate binomial, Kocherlakota and Kocherlakota [1992, page 62] gives

$$E(Y_t|Y_{t-1}) = M\mu^{-1}(1 - \phi) + \phi Y_{t-1} \tag{2.7}$$

where $\phi = (f_1 - \mu^{-1})/(1 - \mu^{-1})$. Notice that $\phi \in (-1, 1)$ and note that ϕ can be negative. Let $W_t = Y_t - E(Y_t \mid Y_{t-1}, \ldots, Y_0)$, and use the above to get

$$Y_t - \frac{M}{\mu} = \phi \left(Y_{t-1} - \frac{M}{\mu} \right) + W_t.$$

Since $\{W_t\}$ is a martingale difference with respect to $\{Y_t\}$, $\{W_t\}$ is white noise.

The general result, which we do not prove here, is that if L_1 has a constant hazard rate after lag k, then the model satisfies an autoregressive moving-average difference equation with autoregressive order k and moving-average order $k-1$.

The joint probability mass function of any n-tuple $Y_1, \ldots, Y_n$, useful for likelihood estimation, can be produced, but the complexity of the problem increases as a function of 2^n. This distribution is given explicitly in (2.5) when $n = 2$. The results can be extended to higher orders inductively. For example, when $n = 3$, partition the outcomes of the 2-dimensional components $(X_{i,1}, X_{i,2})$, $i = 1, \ldots, M$, into four categories, which we denote by $(1, 1)$, $(1, 0)$, $(0, 1)$, and $(0, 0)$. The probability of each outcome is easily computed from the renewal probabilities. For example, $P(X_{i,1} = 1, X_{i,2} = 1) = f_1/\mu$. We can also easily compute $P(X_{i,3} = 1 | X_{i,1} = i_1, X_{i,2} = i_2)$ for any i_1, $i_2 \in \{0, 1\}$. Now use these conditional probabilities along with the multinomial distribution, akin to (2.5), to obtain the joint distribution of (Y_1, Y_2, Y_3).

Hence, theoretically, one can calculate the joint distribution of any order; practically, the computations become unwieldy for moderate n since the computation at "level n" involves 2^n categories to sum over. Later, this issue will lead us to estimate parameters via general linear prediction methods for stationary series.

It will sometimes be advantageous to take M in (2.3) as random, independent of $\{X_{i,t}\}$ for all i. In this case, the covariance function simply becomes $\mathrm{cov}(Y_t, Y_{t+h}) = E[M]\mu^{-1}(u_h - \mu^{-1})$.

Next, we show how to use our model to generate stationary series with binomial, Poisson, and geometric marginal distributions. We comment that it is easy to simulate all renewal processes involved from independent and identically distributed copies of L and its first derived lifetime L_0. The simulation of discrete random variables having specified probability distributions is accomplished as in Ross [2006].

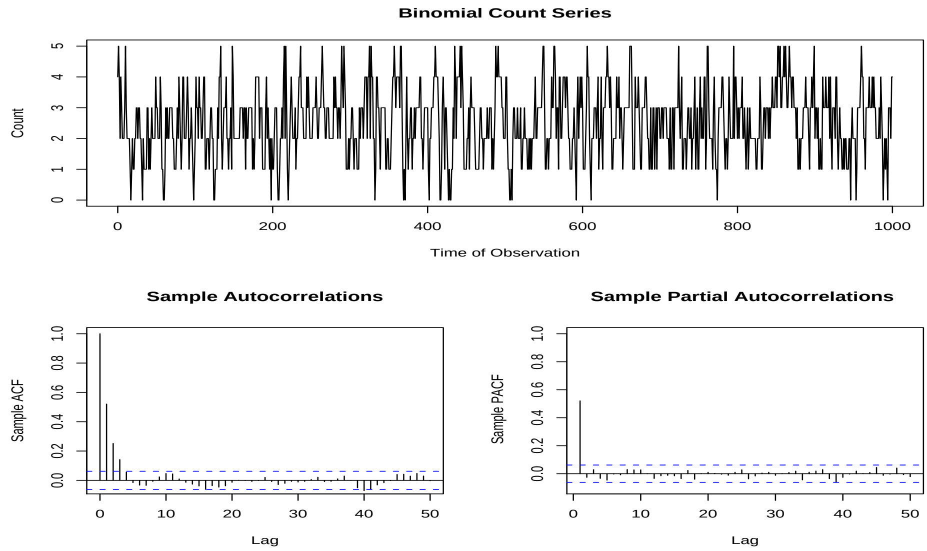

Figure 2.1: A Realization of a Stationary Series Having Binomial Marginals along with Sample Autocorrelations and Partial Autocorrelations.

2.3 The Classical Count Marginals

Stationary series with the classical count marginal distribution structures are easily produced. When M is a unit point mass at $M = m$, Y_t has a binomial distribution for each fixed n with m trials and success probability μ^{-1}. Figure 2.1 shows a sample path of 1000 observations sampled from a stationary series with a binomial marginal distribution with $m = 5$ and success probability 1/2. The sample autocorrelations and partial autocorrelations are also shown. The dashed lines are 95% confidence bounds for white noise (pointwise). The renewal lifetime used here had $f_1 = 3/4$ and $f_n = (16)^{-1}(3/4)^{n-2}$ for $n \geq 2$. This lifetime has a constant hazard rate past lag 1; hence, by Theorem 4, the series is Markov and, as discussed in the last section, satisfies a first order autoregressive difference equation. Discrete autoregressive moving-average and other methods can generate the above series; we offer it mainly as a baseline example.

When M is Poisson with mean λ, a simple calculation will verify that Y_t has a Poisson distribution for each n with mean λ/μ. The joint probability generating function of Y_t and Y_{t+h} is

$$E[s_1^{Y_t} s_2^{Y_{t+h}}] = e^{-\lambda[1-\{\alpha+\beta(s_1+s_2)+\nu s_1 s_2\}]}, \tag{2.8}$$

where α, β, and ν are as in the last section. From (2.8), it follows that (Y_t, Y_{t+h}) has the bivariate Poisson distribution discussed in Holgate [1964] and Loukas et al. [1986]:

$$P(Y_t = i, Y_{t+h} = j) = \sum_{\ell=0}^{\min(i,j)} e^{-(2\lambda_1-\lambda_2)} \frac{(\lambda_1 - \lambda_2)^{i+j-2\ell} \lambda_2^{\ell}}{\ell!(i-\ell)!(j-\ell)!},$$

where $\lambda_1 = \lambda(\beta + \nu)$, and $\lambda_2 = \lambda\nu$.

Figure 2.2 shows a sample path of 1000 points from a series with Poisson marginal distributions with $\lambda = 20$. The renewal lifetime used here was Pareto: $f_n = C/n^{2.5}$ for $n \geq 1$, where C is a constant making the distribution's probabilities sum to unity. This lifetime has a finite mean but infinite second moment. By Theorem 2, the series has long-memory, a trait that can be seen in the sample autocorrelations. We do not know of other methods that can generate a long-memory Poisson count series with such ease, although an unpublished technical report by A. M. M. Q. Quoreshi entitled "A long memory count data time series model for financial application" attempts to do this via fractional differencing.

Stationary series with geometric marginals can also be constructed. Given a collection $\{X_{i,t}\}$ of independent and identically distributed copies of the renewal processes in §2.1 , set $Y_t = \inf\{m \geq 1 : X_{m,t} = 1\}$. Since $P(X_{i,t} = 1) = \mu^{-1}$ for all i and t, Y_t is the time of first success in independent Bernoulli trials having success

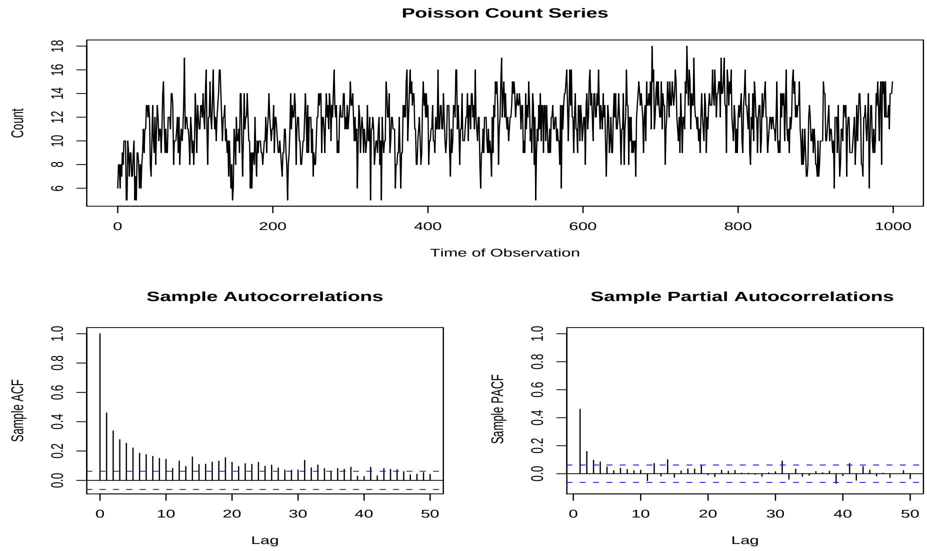

Figure 2.2: A Realization of a Long Memory Stationary Series Having Poisson Marginals along with Sample Autocorrelations and Partial Autocorrelations.

probability μ^{-1}. Hence, Y_t has a geometric distribution with success probability μ^{-1} for each t.

The autocovariance function of $\{Y_t\}$ is derived as follows. First, we will show that $E(Y_tY_{t+h}) = (2-u_h)^{-1}(2\mu^2-\mu)$. To do this, observe that the event $\{Y_t = k, Y_{t+h} = j\}$ with $k > j$ happens precisely when $X_{i,t} = X_{i,t+h} = 0$ for each i satisfying $1 \leq i < j$, $X_{j,t} = 0$ and $X_{j,t+h} = 1$, and $X_{i,t} = 0$ for $j+1 \leq i \leq k-1$ and $X_{k,t} = 1$. Using these and the independence of $\{X_{i,t}\}$ in i gives $P(Y_t = k, Y_{t+h} = j) = \alpha^{j-1}\beta(\mu-1)^{k-j-1}/\mu^{k-j}$ for $k > j$, $P(Y_t = k, Y_{t+h} = j) = \alpha^{k-1}\beta(\mu-1)^{j-k-1}/\mu^{j-k}$ when $k < j$ and $P(Y_t = k, Y_{t+h} = j) = \alpha^{k-1}\nu$ when $k = j$.

The above three probabilities now give

$$E[Y_tY_{t+h}] = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} kjP(Y_t = k, Y_{t+h} = j) = \frac{2(2-u_h)}{(1-\alpha)^2} - \frac{1}{1-\alpha}. \qquad (2.9)$$

Using $1-\alpha = (2-u_h)\mu^{-1}$ in (2.9) gives the aforementioned form of $E[Y_nY_{n+h}]$. Now

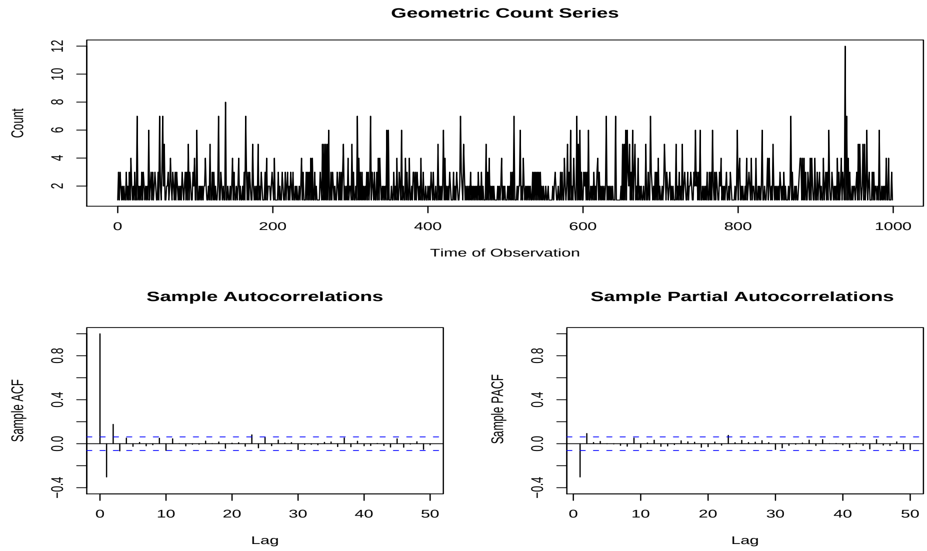

Figure 2.3: A Realization of a Stationary Series Having Geometric Marginals along with Sample Autocorrelations and Partial Autocorrelations.

apply $E(Y_t) = E(Y_{t+h}) = \mu$ to get $\mathrm{Cov}(Y_t, Y_{t+h}) = (\mu^2 u_h - \mu)/(2 - u_h)$ for $h \geq 0$.

The joint probability generating function of Y_t and Y_{t+h} can be found as in Marshall and Olkin [1985]:

$$E[s_1^{Y_t} s_2^{Y_{t+h}}] = \frac{\nu - (s_1 + s_2)\tau - s_1 s_2(\beta^2 - \alpha\tau)}{\{1 - (1 - 1/\mu)s_1\}\{1 - (1 - 1/\mu)s_2\}(1 - \alpha s_1 s_2)} s_1 s_2,$$

where $\tau = \mu^{-1}(u_h - \mu^{-1})$, which is also called a bivariate geometric distribution.

Figure 2.3 shows a sample path of 1000 points from a stationary series with geometric marginal distributions. Here, the renewal lifetime used has $f_1 = 1/4$ and $f_n = (9/16)(1/4)^{n-2}$ for $n \geq 2$. Observe that the lag one sample autocorrelations for this process are negative; correlation structures with negative dependence have been difficult to produce for some count model classes (McKenzie, 2003).

One can construct a stationary series from the above methods with any discrete marginal distribution desired. To do this, let $\{X_{i,t}\}_{t=0}^{\infty}$ be independent and identically distributed stationary Bernoulli sequences for each $i \geq 1$ and let f be a function such that $f(X_{1,t}, X_{2,t}, \ldots)$ has the discrete marginal distribution in question. Such a function exists since every discrete distribution can be generated from independent and identically distributed coin tosses. In many cases, f requires only a finite number of processes. Then $Y_t = f(X_{1,t}, X_{2,t}, \ldots)$ is the desired stationary series. Of course, one needs to derive the autocovariance function $\{Y_t\}$ for each different f.

2.4 Fitting the Renewal Model

This section shows how to fit the renewal model to data. Figure 2.4 shows the number of days in which non-zero rainfall was recorded at Key West, Florida in the $n = 210$ week period spanning from January 2, 2005 — January 3, 2009, inclusive. The sample autocorrelations of this data are displayed in the bottom graphic of the figure and show that the data are correlated. It is natural to fit a model that has binomial marginal distributions with $M = 7$ trials. To proceed further, one must specify the renewal lifetime. We will work with the three-parameter mixture geometric lifetime

$$P(L = k) = \xi p_1(1 - p_1)^{k-1} + (1 - \xi)p_2(1 - p_2)^{k-1}, \quad k \geq 1$$

where ξ, p_1, and p_2 lie in [0,1]. When $p_1 = p_2$ (or $\xi = 0$ or $\xi = 1$) the model reduces to a simple geometric lifetime, which has $\gamma(h) = 0$ for all $h \geq 1$ and represents uncorrelated data. One could try other discrete lifetime distributions or geometric mixtures of three or more components, but the two component geometric mixture is flexible and illustrates the general techniques. Also, Key West, a tropical locality, was chosen for study because its rainfall is largely non-seasonal.

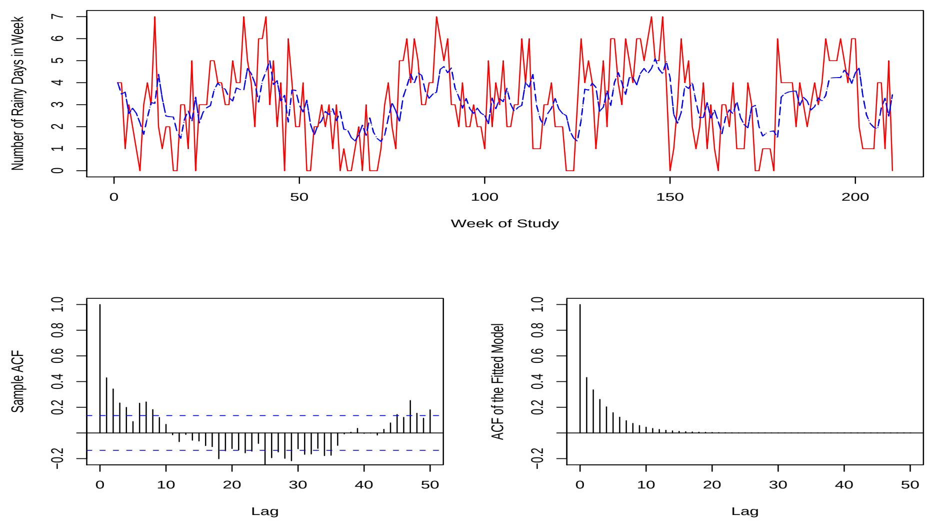

Figure 2.4: Number of Rainy Days in 210 Consecutive Weeks with Fitted Model One-Step-Ahead Linear Predictions. Also Shown Are Sample Autocorrelations and Partial Autocorrelations of the Series.

Estimates of ξ, p_1, and p_2 are chosen as those that minimize the sum of squares

$$S(\xi, p_1, p_2) = \sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2,$$

where $\hat{Y}_t$ is the best linear prediction of Y_t from a constant and the history $Y_1, \ldots, Y_{t-1}$: $\hat{Y}_t = P(Y_t \mid Y_1, \ldots, Y_{t-1}, 1)$. Linear prediction was used in lieu of conditional expectation as the latter seems untractable in this setting. Linear prediction is easily accomplished as follows. For given values of ξ, p_1, and p_2, first compute the renewal probabilities of the model and the autocovariance function in (2.2). From this autocovariance function and the mean of L, use the Durbin–Levinson algorithm (Brockwell and Davis, 1991, chapter 5) to compute the coefficients of the linear predictions recursively. From these predicting coefficients, it is a simple matter to compute the one-step-ahead predictions $\{\hat{Y}_t\}$ and $S(\xi, p_1, p_2)$. No explicit form for the parameter estimates exist; however, values of ξ, p_1 and p_2 that minimize $S(\xi, p_1, p_2)$ were found with a gradient step and search algorithm.

Because $S(\xi, p_1, p_2) = S(1-\xi, p_2, p_1)$, the restriction $\xi \in [0, 1/2]$ is imposed for parameter identifiability. Following the sum of squares quasilikelihood theory in Klimko and Nelson (1978) and the best linear prediction methods for stationary series in Sørensen [2000], minimizers of the sum of squares can be shown to be asymptotically normal. Moreover, estimates of the information matrix can be obtained from the second derivative matrix of S evaluated at the estimated parameters; i.e., the inverse of the so-called Hessian.

The estimated parameters of the model (the error margins listed are one standard error) are $\hat{\xi} = 0.1496 \pm 0.0316$, $\hat{p}_1 = 0.1241 \pm 0.0203$, and $\hat{p}_2 = 0.7764 \pm 0.0288$. The one-step-ahead predictions are plotted in Figure 2.4 and seem to track the data well. The sample autocorrelations of the fitted model are also plotted and reasonably match those of the data for smaller lags, although some midrange dependence remains unmodeled. From these estimators, a z-score for the hypothesis test that $p_1 = p_2$ is $z = 25.1856$, which has a p-value of nearly zero; the same conclusion is obtained if one tests a null hypothesis of $\xi = 0$ or $\xi = 1$. The mean of the fitted model is 3.0425, while the sample mean is 3.0857. Overall, the model seems to fit the data roughly. Future work will formalize the asymptotics of the fitting methods and incorporate the possibility of periodic dynamics.

2.5 Comments

Nonstationary processes can also be produced with the renewal model. Such a setup could also accommodate covariates. To do this, one can simply allow the parameters in the model to vary with time or the covariates. For example, periodically stationary series could be constructed with a periodic renewal process. A process with binomial marginals that depend on a time-varying deterministic covariate c_t at time t could be obtained by allowing M and the renewal probabilities at time t to depend on c_t.

2.6 Proofs

Proof of Theorem 1. Let Γ_n be the $n \times n$ matrix with (i, j)th entry $\gamma(|i-j|)$ for $1 \leq i, j \leq n$. Since $u_h \to \mu^{-1}$ as $h \to \infty$, $\gamma(h) \to 0$ as $h \to \infty$. Applying Proposition 5.1.1 in , Brockwell and Davis [1991] we infer that Γ_n is non-singular for every n. Now let $\vec{a}$ be any nonzero $n \times 1$ vector. Invertibility of Γ_n gives $\vec{a}'\Gamma_n\vec{a} > 0$, or equivalently, from (2.2), $\vec{a}'U_n\vec{a} - \mu^{-1}\vec{a}'J_n\vec{a} > 0$, where J_n denotes an $n \times n$ matrix whose entries are all unity. Hence, $\vec{a}'U_n\vec{a} > \mu^{-1}\vec{a}'J_n\vec{a} \geq 0$ and the result is proven. □

Proof of Theorem 2. Equation (2.2) gives

$$\sum_{h=0}^{\infty} |\gamma(h)| = \frac{1}{\mu} \sum_{t=0}^{\infty} |P_0(X_t = 1) - P_S(X_t = 1)|, \tag{2.10}$$

where $P_0(X_t = 1)$ denotes the probability of a renewal at time n in a non-delayed renewal process and $P_S(X_t = 0) \equiv \mu^{-1}$ is the stationary version of this probability; i.e., L_0 has the first derived distribution as discussed in §2.1. Given that $E[L] < \infty$, Pitman [1974] shows that the sum in right hand side of (2.10) is finite if and only if $E[L_0] < \infty$. But since $E[L_0] = E[L^2]/(2E[L])$, the theorem follows. □

Proof of Theorem 3. Recall that X_t is either zero or one for each n. If $\{X_t\}$ is Markov, then

$$P(X_t = 1 \mid X_{t-1} = 0, X_{t-2} = 0, \ldots, X_{t-k+1} = 0, X_{t-k} = 1) = P(X_t = 1 \mid X_{t-1} = 0) \tag{2.11}$$

for all $k \geq 2$. But the left hand side of (2.11) is

$$\frac{P(X_t = 1, X_{t-1} = 0, \ldots, X_{t-k+1} = 0, X_{t-k} = 1)}{P(X_{t-1} = 0, \ldots, X_{t-k+1} = 0, X_{t-k} = 1)} = \frac{\mu^{-1} P(L = k)}{\mu^{-1} P(L > k-1)} = h_k,$$

where h_k is the hazard rate of L at index k. The right hand side of (2.11) is seen to be constant in t:

$$P(X_t = 1 \mid X_{t-1} = 0) = \frac{P(X_t = 1) - P(X_{t-1} = 1, X_t = 1)}{P(X_{t-1} = 0)} = \frac{\mu^{-1} P(L > 1)}{1 - \mu^{-1}}. \tag{2.12}$$

Therefore, L must have constant hazard rates past lag 1 for $\{X_t\}$ to be Markov.

Now suppose that L has a constant hazard rate past lag 1 and write $f_n = f_2 r^{n-2}$ for $n \geq 2$ and some $r < 1$. The hazard rate for L at lag $k \geq 2$ is $h_k = f_k / (\Sigma_{\ell=k}^{\infty} f_\ell) = 1 - r$. To verify the Markov property, we need to show that

$$P(X_t = i_t \mid X_{t-1} = i_{t-1}, X_{t-2} = i_{t-2}, \cdots, X_0 = i_0) = P(X_t = i_t \mid X_{t-1} = i_{t-1})$$

for every $i_k \in \{0, 1\}$ and all $n \geq 2$. Because X_t is either zero or one for all t, we need only consider the case where $i_t = 1$ (the case where $i_t = 0$ then follows by complementation).

Our work proceeds by cases. If $X_{t-1} = 1$ then

$$P(X_t = 1 \mid X_{t-1} = 1, X_{t-2} = i_{t-2}, \cdots, X_0 = i_0) = P(X_t = 1 \mid X_{t-1} = 1) = f_1.$$

If $X_{t-1} = 0$, there are two subcases. First, suppose that $i_k = 0$ for $0 \le k \le t-1$. Then the lifetime L_0 is in use at time $t-1$ and

$$P(X_t = 1 \mid X_i = 0, 0 \le i \le n-1) = \frac{P(L_0 = t)}{P(L_0 \ge t)} = 1 - r,$$

where the fact that $P(L_0 = t) = \mu^{-1} P(L > t)$ for $t \ge 0$ and $f_t = f_2 r^{t-2}$ for $t \ge 2$ have been applied.

From (2.12), we have $P(X_t = 1 \mid X_{t-1} = 0) = (1 - f_1)/(\mu - 1)$. Using $E[L] = 1 + f_2(1-r)^{-2}$ and $1 - f_1 = f_2(1-r)^{-1}$ gives $P(X_t = 1 \mid X_{t-1} = 0) = 1 - r$, which verifies the Markov property for this case.

Our second subcase entails the situation where $i_k = 1$ for some k with $0 \le k \le t-2$. Let $\ell = \max\{k : i_k = 1 \text{ and } 0 \le k \le t-1\}$ be the maximum such index and argue as above to finish our work:

$$P(X_t = 1 \mid X_{t-1} = 0, \ldots, X_\ell = 1, X_{\ell-1} = i_{\ell-1}, \ldots, X_0 = i_0) = \frac{P(L = t - \ell)}{P(L \ge t - \ell)} = 1 - r.$$

$\square$

Proof of Theorem 4. Suppose that $\{Y_t\}$ is Markov. Then

$$P(Y_t = M \mid Y_{t-1} = 0) = P(Y_t = M \mid Y_{t-1} = 0, \ldots, Y_0 = M)$$

for all $t \ge 2$. Applying the independence of the M component processes gives

$$\Pi_{i=1}^{M} P(X_{i,t} = 1 \mid X_{i,t-1} = 0) = \Pi_{i=1}^{M} P(X_{i,t} = 1 \mid X_{i,t-1} = 0, \ldots, X_{i,0} = 1). \quad (2.13)$$

We have selected special values for Y_i to take on over $0 \le i \le t$, extreme in that they are either M or 0. Since the terms in the products in (2.13) are constant in i, we infer that

$$P(X_{1,t} = 1 \mid X_{1,t-1} = 0) = P(X_{1,t} = 1 \mid X_{1,t-1} = 0, \ldots, X_{1,0} = 1).$$

Now argue as in the proof of Theorem 3 to infer that L must have constant hazard rates past lag 1.

Now suppose that L has constant hazard rates past lag 1. We need to show that

$$P(Y_t = j \mid Y_{t-1} = i, Y_{t-2}, \cdots, Y_0) = P(Y_t = j \mid Y_{t-1} = i). \tag{2.14}$$

Conditional on the event $Y_{t-1} = i$ and any values of $Y_{t-2}, \ldots Y_0$, the ways in which $Y_t = j$ can happen are enumerated as follows. If ℓ of the i component processes $\{X_{i,t}\}$ which are unity at time $t-1$ are still unity at time n, then $j - \ell$ of the component processes which were zero at time $t-1$ must have transitioned unity to make $Y_t = j$. Summing over all possible ℓ and applying the Markov property of the component processes provides

$$P(Y_t = j \mid Y_{t-1} = i, Y_{n-2}, \cdots, Y_0) = \sum_{\ell=\max(0,i+j-M)}^{\min(i,j)} \binom{i}{\ell} \binom{M-i}{j-\ell} J_\ell, \tag{2.15}$$

where $J_\ell = f_1^\ell (1-f_1)^{i-\ell}(1-r)^{j-\ell} r^{M+\ell-j-i}$. The expression for J_ℓ is a simple multinomial probability. In fact, the proof of Theorem 3 showed that

$$\begin{aligned}
P(X_{\ell,t} = 1 \mid X_{\ell,t-1} = 1, X_{\ell,t-2} = i_{t-2}, \ldots, X_{\ell,0} = i_0) &= f_1 \\
P(X_{\ell,t} = 0 \mid X_{\ell,t-1} = 1, X_{\ell,t-2} = i_{t-2}, \ldots, X_{\ell,0} = i_0) &= 1 - f_1 \\
P(X_{\ell,t} = 1 \mid X_{\ell,t-1} = 0, X_{\ell,t-2} = i_{t-2}, \ldots, X_{\ell,0} = i_0) &= 1 - r \\
P(X_{\ell,t} = 0 \mid X_{\ell,t-1} = 0, X_{\ell,t-2} = i_{t-2}, \cdots, X_{\ell,0} = i_0) &= r.
\end{aligned}$$

for each of the component processes ($1 \le \ell \le M$).

When $h = 1$, the Section §2.2 parameters can be evaluated as $\alpha = r(1 - \mu^{-1})$, $\beta = (1 - r)(1 - \mu^{-1})$, and $\nu = \mu^{-1} f_1$. For the right side of (2.14), apply the above results to the conditional distribution derived in (2.6) to obtain

$$P(Y_t = j \mid Y_{t-1} = i) = \sum_{\ell=\max(0,i+j-M)}^{\min(i,j)} \binom{i}{\ell} \binom{M-i}{j-\ell} J_\ell,$$

which completes our work. $\square$

Chapter 3

Renewal AR(1) Count Series

In time series, the classical first order autoregressive process (AR(1)) is widely used because of its simple structure and vast applicability. Count versions of AR(1) models have been proposed before. In particular, the first order discrete autoregession (DAR(1)) and the first order integer autoregression (INAR(1)) are two commonly used count series with AR(1) structure. By an AR(1) structure, we refer to a stationary series with lag h autocorrelation ρ^h for $h \geq 0$.

DAR(1) series (Jacobs and Lewis, 1978a, 1978b) are based on mixing copies of random variables having the prescribed marginal distribution. They can have any prescribed marginal distributions. However, DAR(1) series have the drawback that sample paths of the process contain many runs of a constant value, which is frequently unrealistic. As noted in Chapter 1, DAR(1) autocorrelations must be non-negative.

The INAR(1) process $\{X_t\}$ (McKenzie, 1985, 1988; Al-Osh and Alzaid 1987) is based on thinning operation. It can have Poisson, geometric, and negative binomial distributions, but cannot have a binomial marginal distributions (only discrete self-decomposable distributions in the sense of Steutel and Van Harn [1979] can be achieved). Latour [1998] extends thinning techniques to general non-negative random variables in introducing generalized integer autoregressive models, but again, such processes cannot generate count series with negative correlations (or binomial marginals).

In this chapter, we present how to use renewal series to generate stationary integer-valued time series with any AR(1) structure ($\phi \in (-1, 1)$). The method is based on superpositioning method of the last chapter. The superpositioning method is simple and versatile; it is also parsimonious in that the correlation structure of the model is determined by the lifetime distribution of the underlying renewal processes. The model does not tend to generate series with constant runs such as DAR(1) series.

3.1 AR(1) Count Series

In Chapter 2, it is shown that by aggregating independent copies of $\{X_t\}$ (the stationary Bernoulli trials in §2.1) in various ways, one obtains a rich class of stationary processes. When $M \geq 0$ is a fixed integer, $\{Y_t\}$ defined by (2.3) is a stationary series with binomial marginal distributions. The autocovariance of $\{Y_t\}$ is

$$\gamma_Y(h) = \text{cov}(Y_t, Y_{t+h}) = M\mu^{-1}(u_h - \mu^{-1}).$$

Stationary series with Poisson marginals can also be constructed: simply take M in (2.3) as a Poisson random variate with mean λ that is independent of each $\{X_{i,t}\}$. The value of M is generated up front and is needed to define Y_t at each t. The autocovariance of $\{Y_t\}$ is

$$\gamma_Y(h) = E[E\{(Y_t - E(Y_t))(Y_{t+h} - E(Y_{t+h}))|M\}] = \lambda\mu^{-1}(u_h - \mu^{-1}).$$

We now relate the renewal model to AR(1) models. The classical stationary and causal AR(1) process $\{X_t\}$ with mean c satisfies the difference equation

$$X_t - c = \phi(X_{t-1} - c) + \epsilon_t, \tag{3.1}$$

where $|\phi| < 1$ and $\{\epsilon_t\}$ is zero mean white noise with variance σ^2. The autocovariance and autocorrelation functions of the AR(1) model are

$$\mathrm{Cov}(X_t, X_{t+h}) = \frac{\sigma^2\phi^h}{1-\phi^2} \quad \text{and} \quad \mathrm{Corr}(X_t, X_{t+h}) = \phi^h, \tag{3.2}$$

for $h \geq 0$. While DAR(1) and INAR(1) models are capable of generating count series with an AR(1) autocorrelation function with any $\phi > 0$, they cannot generate AR(1) count structures with $\phi < 0$. Below, we show that our renewal class can easily accommodate negative correlations.

Let $h_k = P(L = k | L \geq k)$ be the hazard rate of the lifetime L at index k. If a lifetime distribution has constant hazard rate after lag 1, then its probability mass function has the form $f_2 = (1 - f_1)(1 - r)$ and $f_n = f_2 r^{n-2}$ for $n \geq 2$ for some $f_1 = P(L = 1)$ and $r \in (0, 1)$. Since $h_1 = P(L = 1|L \geq 1) = f_1$ and $h_k = P(L = k|L \geq k) = f_k/(\Sigma_{\ell=k}^{\infty} f_\ell) = 1 - r$ the distribution of L is simply $f_1 = h_1$, $f_2 = (1 - h_1)h_2$, and $f_n = f_2(1 - h_2)^{n-2}$ for $n \geq 2$. The mean of L is

$$\mu = 1 + \frac{f_2}{(1-r)^2} = \frac{1 + h_2 - h_1}{h_2}. \tag{3.3}$$

Our first result is the following. The proofs of all results in this section are delegated to §3.5.

Theorem 5. $\{Y_t\}$ *satisfies the* AR(1) difference equation if and only if h_k *is* *constant over over* $k \geq 2$. *When* h_k *is constant over* $k \geq 2$, *the* AR(1) *model can be written as*

$$Y_t - \frac{M}{\mu} = \phi\left(Y_{t-1} - \frac{M}{\mu}\right) + W_t,$$

where $\mu = E[L]$, $\phi = (f_1 - \mu^{-1})/(1 - \mu^{-1}) = h_1 - h_2$, *and* $\{W_t\}$ *is zero mean white noise with variance* $M\mu^{-1}(1 - \mu^{-1})(1 - \phi^2)$. *This model is causal in that* $|\phi| < 1$.

Observe that negative correlation in the renewal model can arise: simply select a renewal lifetime L with $h_1 < h_2$. If L is geometric, ($f_n = pq^{n-1},\ n \geq 1,\ E(L) = 1/p$), then Theorem 5 shows that $\phi = 0$ and $\{Y_t\}$ is a white noise sequence of counts with mean Mp variance $Mp(1-p)$.

Our next result shows that the renewal class can generate an AR(1) count series with any autocorrelation parameter ϕ in the range (-1,1).

THEOREM 6. *Given a* $\phi \in (-1, 1)$, *there exists a renewal count* AR(1) *process with the autocorrelation function in* (3.2). *This process can be chosen to have either binomial or Poisson marginal distributions.*

While Theorem 6 states that our renewal model can produce integer-valued time series with any AR(1) autocorrelation function whatsoever ($-1 < \phi < 1$), it cannot produce any AR(1) autocovarince function whatsoever. The next result clarifies what AR(1) autocovariance functions can be produced with the renewal model.

THEOREM 7. a) *For any* AR(1) *process with* $\phi \in [0, 1)$ *and any* $\sigma^2 > 0$, *there exists a renewal process with the* AR(1) *autocovariance function in* (3.2). *This process can be chosen to have either binomial or Poisson marginals.*

b) *For an* AR(1) *process with* $\phi \in (-1, 0)$ *and any* $\sigma^2 > 0$, *there exists a renewal process with Poisson marginals and the* AR(1) *autocovariance function in*

(3.2). *When an* M *exists such that* $-\phi(1-\phi)^{-2} < M^{-1}\sigma^2(1-\phi^2)^{-1} \leq 1/4$, *then there exists a renewal process that has the* AR(1) *autocovariance function in* (3.2) with Binomial marginals.

When M is a fixed integer and h_k is constant for $k \geq 2$, it is proved in Theorem 4 that $\{Y_t\}$ is Markov (this is slightly stronger than the Theorem 5 conclusion that $\{Y_t\}$ is AR(1)). We also know that $P(X_{i,t} = 1|X_{i,t-1} = 1) = f_1 = h_1$ and $P(X_{i,t} = 1|X_{i,t-1} = 0) = 1 - r = h_2$ for all i. Since $P(X_{i,t} = 1|X_{i,t-1} = 1) = f_1$ and $P(X_{i,t} = 1|X_{i,t-1} = 0) = h_2$ for each i, Y_t be can represented as

$$Y_t = h_1 \circ Y_{n-1} + h_2 \circ (M - Y_{n-1}).$$

This is a special case of a stationary series with binomial marginal distributions discussed in McKenzie [1985].

3.2 Estimation

Suppose that $\{Y_t\}$ is an AR(1) renewal process and that M is known. This section considers estimation of the process parameters from the data $Y_0, \dots, Y_n$. The parameters in the model are those governing the lifetime L. By Theorem 5, the lifetime L must have constant hazard rates past lag 1. Hence, the parameters to be estimated are simply h_1 and h_2.

It is natural to first explore likelihood estimation methods. Since $\{Y_t\}$ is Markov, the likelihood function can be expressed as

$$L(h_1, h_2) = P(Y_0) \prod_{t=1}^{n} P(Y_t|Y_{t-1})$$

$$= \begin{pmatrix} M \\ Y_0 \end{pmatrix} \mu^{-Y_0}(1-\mu^{-1})^{M-Y_0} \prod_{t=1}^{n} P(Y_t|Y_{t-1}), \tag{3.4}$$

where $P(Y_t|Y_{t-1})$ is calculated from (2.15) as

$$P(Y_n = j|Y_{n-1} = i) = \sum_{\ell=\max(0,i+j-M)}^{\min(i,j)} \binom{i}{\ell}\binom{M-i}{j-\ell} h_1^{\ell}(1-h_1)^{i-\ell} h_2^{j-\ell}(1-h_2)^{M+\ell-j-i}.$$

This likelihood function is somewhat unwieldy: no simple closed-form expressions for h_1 and h_2 that maximize the likelihood function are evident to us. Maximization of the likelihood appears to be a numerical task.

Since our AR(1) model is ergodic and stationary, the asymptotic properties of conditional least squares estimators can be quantified (see Klimko and Nelson, 1978). We first consider the parameters ϕ and $\eta = M\mu^{-1}(1-\phi)$ as their asymptotic properties can be easily quantified. A multivariate delta method with the relations

$$h_1 = \phi + \frac{\eta}{M}, \qquad h_2 = \frac{\eta}{M} \tag{3.5}$$

will then identify the asymptotic distributions of the conditional least squares estimates of h_1 and h_2.

Let $g(\phi, \eta) = E(Y_t|Y_{t-1}, \cdots, Y_0) = E(Y_t|Y_{t-1})$. From (2.7), it is known that $g(\phi, \eta) = \phi Y_{t-1} + \eta$. The conditional least squares estimators for ϕ and η, denoted by $\hat{\phi}_{\text{CLS}}$ and $\hat{\eta}_{\text{CLS}}$, are minimizers of

$$Q_n(\phi, \eta) = \sum_{t=0}^{n} (Y_t - \phi Y_{t-1} - \eta)^2.$$

By solving the equations

$$\frac{\partial Q_n(\phi,\eta)}{\partial \phi} = 0 \quad \text{and} \quad \frac{\partial Q_n(\phi,\eta)}{\partial \eta} = 0,$$

the conditional least squares estimators are identified explicitly as

$$\hat{\phi}_{\text{CLS}} = \frac{\sum_{t=1}^{n} Y_t Y_{t-1} - \left(\frac{1}{n}\sum_{t=1}^{n} Y_t\right)\left(\sum_{t=1}^{n} Y_{t-1}\right)}{\sum_{t=1}^{n} Y_{t-1}^2 - \left(\frac{1}{n}\sum_{t=1}^{n} Y_{t-1}\right)\left(\sum_{t=1}^{n} Y_{t-1}\right)}; \quad \hat{\eta}_{\text{CLS}} = \frac{1}{n}\left(\sum_{t=1}^{n} Y_t - \hat{\phi}_{\text{CLS}} \sum_{t=1}^{n} Y_{t-1}\right).$$

Since the conditions of Theorem 3.1 and 3.2 of Klimko and Nelson [1978] are satisfied, $\hat{\phi}_{\text{CLS}}$ and $\hat{\eta}_{\text{CLS}}$ are consistent and asymptotically normal with

$$\begin{pmatrix} \hat{\phi}_{\text{CLS}} \\ \hat{\eta}_{\text{CLS}} \end{pmatrix} \sim \text{AN}_2\left(\begin{pmatrix} \phi \\ \frac{M}{\mu}(1-\phi) \end{pmatrix}, \frac{V^{-1}WV^{-1}}{n} \right),$$

where V is the 2×2 matrix

$$V = \begin{pmatrix} E[Y_{t-1}^2] & E[Y_{t-1}] \\ E[Y_{t-1}] & 1 \end{pmatrix} = \begin{pmatrix} \frac{M}{\mu}(1-\frac{1}{\mu}) + \frac{M^2}{\mu^2} & \frac{M}{\mu} \\ \frac{M}{\mu} & 1 \end{pmatrix},$$

and W is the 2×2 matrix

$$W = \begin{pmatrix} E[(Y_t - \phi Y_{t-1} - \frac{M}{\mu}(1-\phi))^2 Y_{t-1}^2] & E[(Y_t - \phi Y_{t-1} - \frac{M}{\mu}(1-\phi))^2 Y_{t-1}] \\ E[(Y_t - \phi Y_{t-1} - \frac{M}{\mu}(1-\phi))^2 Y_{t-1}] & E[(Y_t - \phi Y_{t-1} - \frac{M}{\mu}(1-\phi))^2] \end{pmatrix}.$$

Since $(Y_t, Y_{t-1})'$ has a bivariate binomial distribution, the generating function of $(Y_t, Y_{t-1})'$ can be computed explicitly (see equation (2.4)). From this, one can calculate all entires in W and identify the asymptotic information matrix $C = V^{-1}WV^{-1}$.

Specifically, tedious algebra identifies the entries of C as

$$
\begin{aligned}
C_{11} &= \frac{(1-\phi)[-4\phi\mu^{-1}(1-\mu^{-1}) + M\mu^{-1}(1+\phi)(1-\mu^{-1}) + \phi]}{M\mu^{-1}(1-\mu^{-1})}; \\
C_{12} &= C_{21} = \frac{(1-\phi)[M\mu^{-1}(1+\phi) - M(1+\phi) - 2\phi\mu^{-1}(1+\phi) + \phi]\mu^{-1}}{1-\mu^{-1}}; \\
C_{22} &= \frac{(1-\phi)[\mu^{-2}(1+\phi) - 2\mu^{-1} - M\mu^{-2}(1+\phi) + M\mu^{-1}(1+\phi) + 1]M\mu^{-1}}{1-\mu^{-1}}.
\end{aligned}
$$

The conditional least squares estimate of $E[Y_t] = M/\mu$ is simply taken as $\hat{\mu}_{Y,\mathrm{CLS}} = \hat{\eta}_{\mathrm{CLS}}/(1-\hat{\phi}_{\mathrm{CLS}})$. A multivariate delta theorem (Brockwell and Davis, [1991], Proposition 6.4.3) now gives

$$
\begin{pmatrix} \hat{\phi}_{\mathrm{CLS}} \\ \hat{\mu}_{Y,\mathrm{CLS}} \end{pmatrix} \sim \mathrm{AN}_2 \left(\begin{pmatrix} \phi \\ \frac{M}{\mu} \end{pmatrix}, \; \frac{ACA'}{n} \right),
$$

where A is the 2×2 matrix

$$
A = \begin{pmatrix} 1 & 0 \\ \frac{M}{\mu(1-\phi)} & \frac{1}{(1-\phi)} \end{pmatrix}.
$$

Let $\Sigma = ACA'$, then we have

$$
\begin{aligned}
\Sigma_{11} &= \frac{(1-\phi)[-4\phi\mu^{-1}(1-\mu^{-1}) + M\mu^{-1}(1+\phi)(1-\mu^{-1}) + \phi]}{M\mu^{-1}(1-\mu^{-1})}; \\
\Sigma_{12} &= \Sigma_{21} = \phi(1-2\mu^{-1}); \\
\Sigma_{22} &= \frac{M\mu^{-1}(1-\mu^{-1})(1+\phi)}{1-\phi}.
\end{aligned}
$$

Conditional least squares estimates of the hazard rates in (3.5) are obtained by $\hat{h}_{1,\mathrm{CLS}} = \hat{\phi}_{\mathrm{CLS}} + \hat{\eta}_{\mathrm{CLS}}/M$ and $\hat{h}_{2,\mathrm{CLS}} = \hat{\eta}_{\mathrm{CLS}}/M$. Using a multivariate delta method again gives

$$\begin{pmatrix} \hat{h}_{1,\text{CLS}} \\ \hat{h}_{2,\text{CLS}} \end{pmatrix} \sim \text{AN}_2\left(\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}, \frac{BCB'}{n} \right), \tag{3.6}$$

where B is the 2×2 matrix

$$B = \begin{pmatrix} 1 & \frac{1}{M} \\ 0 & \frac{1}{M} \end{pmatrix}.$$

With $\Delta = BCB'$, then

$$\begin{aligned}
\Delta_{11} &= \frac{\mu^{-1}(1-\phi^2)[M(1-\mu^{-1})^2 - \mu^2] - (1-\phi)[\phi(3\mu^{-1}-1) - \mu^{-2}(3\phi+1)]}{M\mu^{-1}}; \\
\Delta_{12} &= \Delta_{21} = \frac{(1-M)\mu^{-1}(1-\mu^{-1})(1-\phi^2)}{M}; \\
\Delta_{22} &= \frac{\mu^{-2}(1-\phi^2)(-M+1) + M\mu^{-1}(1-\phi^2) + (1-\phi)(1-2\mu^{-1})}{M(1-\mu^{-1})}.
\end{aligned}$$

3.3 Examples

This section considers several simulation issues, including how the parameter estimators in the last section perform. We first address how to simulate a count series with a given autocorrelation/autocovariance structure.

Example 1. a) *Suppose we want to generate a stationary* AR(1) *count series with binomial marginals that has the autocorrelation function* $\rho(h) = \phi^h$ *for* $h \geq 0$.

As in the proof of Theorem 6, pick any integer $M \geq 1$. The two hazard rates of L, h_1 and h_2, can be chosen as any numbers in $(0, 1)$ that satisfy $h_1 - h_2 = \phi$. The renewal series generated by (2.3) with L having hazard rates h_1 and $h_k = h_2$ for $k \geq 2$ has the given autocorrelation function.

b) *Suppose we want to generate a stationary* AR(1) *count series with Poisson*

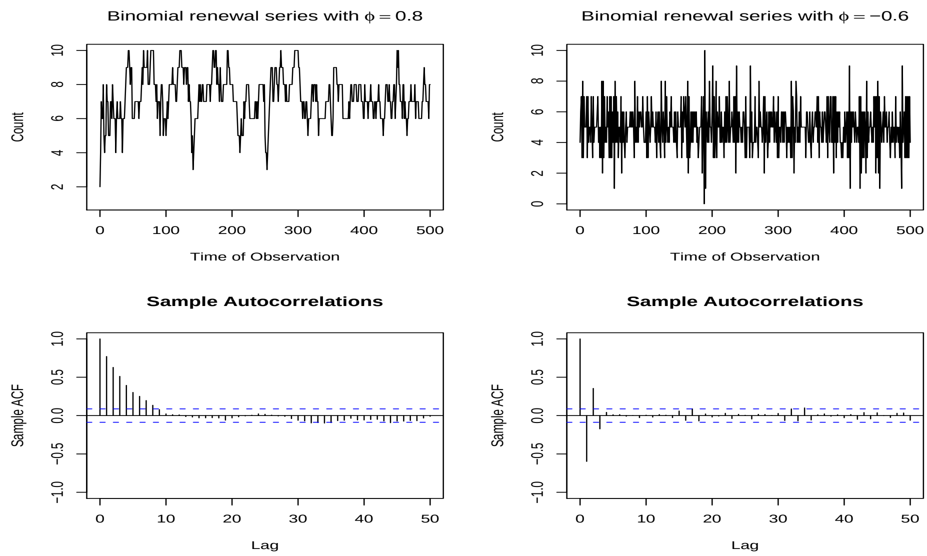

Figure 3.1: Top: A sample path of a stationary series with binomial marginals and $\rho(h) = \phi^h$ for $h \geq 0$, when a) $\phi = 0.8$ and b) $\phi = -0.6$. Bottom: Sample autocorrelations of the series.

marginals and the autocovariance function $\gamma(h) = \sigma^2\phi^h/(1-\phi^2)$ *for* $h \geq 0$ *and* $\phi < 0$.

By (3.16), it is possible to find a λ such that $-\phi(1-\phi)^{-2} < \lambda^{-1}(1-\phi^2)^{-1} < 1/4$. With such a λ, we then solve the second equation of (3.13) for μ^{-1}. The parameters in the lifetime distribution of L are now directly calculated from (3.12). The hazard rates are $h_1 = \phi - \mu^{-1}(\phi - 1)$ and $h_2 = \mu^{-1}(1 - \phi)$.

c) *Suppose we want to generate a stationary* AR(1) *count series with binomial marginals that has the autocovariance function* $\gamma(h) = (-4/5)^h/(1-(4/5)^2)$ *for* $h \geq 0$.

This is an example that cannot be produced with our model class. To see this, note that by (3.15) we would need to find and integer $M \geq 1$ such that $-\phi(1-\phi)^{-2} < M^{-1}(1-\phi^2)^{-1} < 1/4$ with $\phi = -4/5$. However, no such M exists.

d) *Suppose we want to generate a stationary* AR(1) *count series with binomial marginals that has the autocovariance function* $\gamma(h) = (-1/5)^h/(1-(1/5)^2)$ *for* $h \geq 0$.

This is an example with negative ϕ that we can handle. First, select M such that $-\phi(1-\phi)^{-2} < M^{-1}(1-\phi^2)^{-1} < 1/4$ with $\phi = -1/5$. There are two choices: $M = 5$ or $M = 6$. With $M = 5$, $\mu_1^{-1} = (12/5)(1+\sqrt{1/6})$ and $\mu_2^{-1} = (12/5)(1-\sqrt{1/6})$. Direct calculation gives two possible distributions for L: $f_1 = 2/5 + \sqrt{6}/10$ and $f_n = (3/10)(2/5 - \sqrt{6}/10)^{n-2}$ for $n \geq 2$, or $f_1 = 2/5 - \sqrt{6}/10$ and $f_n = (3/10)(2/5 + \sqrt{6}/10)^{n-2}$ for $n \geq 2$. Similar analyses hold with $M = 6$.

Example 2.

This example uses simulation to study the performance of the conditional least squares estimators of h_1 and h_2. For each pair of h_1 and h_2 shown in Table 1, we simulated binomial series of lengths $n = 100$, $n = 500$, and $n = 1000$ respectively. Ten thousand simulations were performed in all cases. We have taken $M = 10$ in all simulations. The variance of the white noise term in the AR(1) model is determined by M, h_1, and h_2 (recall that $h_1 - h_2 = \phi$). Each simulation run first generates a sample series with the noted properties. Next, the estimates $\hat{h}_{1,\mathrm{CLS}}$ and $\hat{h}_{2,\mathrm{CLS}}$ are computed. Reported are sample averages of the 10000 estimates $\hat{h}_{1,\mathrm{CLS}}$ and $\hat{h}_{2,\mathrm{CLS}}$ and their sample standard deviation. Observe that the sample averages become closer to their true values as n increases. Moreover, the sample standard deviations agree with those listed in equation (3.6). In particular, when $h_1 = 0.9$, $h_2 = 0.1$ and $n = 1000$, (3.6) gives $[\mathrm{Var}(\hat{h}_{1,\mathrm{CLS}})]^{\frac{1}{2}} \approx 9.95 \times 10^{-3}$, which is close to the sample standard deviation of 1.015×10^{-2} listed in Table 1. Overall, the conditional least squares estimates appear to work well.

3.4 Proofs

Proof of Theorem 5. Suppose that L has a constant h_k over $k \geq 2$. Letting

Table 3.1: Simulation results for $\hat{h}_{1,\mathrm{CLS}}$ and $\hat{h}_{2,\mathrm{CLS}}$

		$n = 100$	$n = 500$	$n = 1000$
$\phi = 0.8$				
$h_1 = 0.9$	$\overline{\hat{h}_{1,\mathrm{CLS}}}$	$0.8831\ (3.551 \times 10^{-2})$	$0.8970\ (1.432 \times 10^{-2})$	$0.8984\ (1.015 \times 10^{-2})$
$h_2 = 0.1$	$\overline{\hat{h}_{2,\mathrm{CLS}}}$	$0.1168\ (3.517 \times 10^{-2})$	$0.1029\ (1.444 \times 10^{-2})$	$0.1017\ (1.014 \times 10^{-2})$
$\phi = 0.5$				
$h_1 = 0.8$	$\overline{\hat{h}_{1,\mathrm{CLS}}}$	$0.7907(3.776 \times 10^{-2})$	$0.7980(1.626 \times 10^{-2})$	$0.7991(1.162 \times 10^{-2})$
$h_2 = 0.3$	$\overline{\hat{h}_{2,\mathrm{CLS}}}$	$0.3144(5.585 \times 10^{-2})$	$0.3029(2.404 \times 10^{-2})$	$0.3013(1.735 \times 10^{-2})$
$\phi = 0$				
$h_1 = 0.6$	$\overline{\hat{h}_{1,\mathrm{CLS}}}$	$0.5967(4.288 \times 10^{-2})$	$0.5994(1.908 \times 10^{-2})$	$0.5995(1.355 \times 10^{-2})$
$h_2 = 0.6$	$\overline{\hat{h}_{2,\mathrm{CLS}}}$	$0.6051(6.187 \times 10^{-2})$	$0.6012(2.747 \times 10^{-2})$	$0.6009(1.963 \times 10^{-2})$
$\phi = -0.5$				
$h_1 = 0.2$	$\overline{\hat{h}_{1,\mathrm{CLS}}}$	$0.2019(4.705 \times 10^{-2})$	$0.2005(2.114 \times 10^{-2})$	$0.1998(1.493 \times 10^{-2})$
$h_2 = 0.7$	$\overline{\hat{h}_{2,\mathrm{CLS}}}$	$0.6983(4.267 \times 10^{-2})$	$0.6996(1.920 \times 10^{-2})$	$0.7000(1.359 \times 10^{-2})$
$\phi = -0.8$				
$h_1 = 0.1$	$\overline{\hat{h}_{1,\mathrm{CLS}}}$	$0.1057(3.316 \times 10^{-2})$	$0.1011(1.412 \times 10^{-2})$	$0.1006(1.004 \times 10^{-2})$
$h_2 = 0.9$	$\overline{\hat{h}_{2,\mathrm{CLS}}}$	$0.8944(3.316 \times 10^{-2})$	$0.8989(1.416 \times 10^{-2})$	$0.8994(1.003 \times 10^{-2})$

$W_t = Y_t - E(Y_t|Y_{t-1}\ldots,Y_0)$, from (2.7) and §2.2 we know $\{W_t\}$ is white noise and

$$Y_t - \frac{M}{\mu} = \phi\left(Y_{t-1} - \frac{M}{\mu}\right) + W_t, \tag{3.7}$$

with

$$\phi = \frac{f_1 - \mu^{-1}}{1-\mu^{-1}}. \tag{3.8}$$

Therefore, $\{Y_t\}$ satisfies an AR(1) difference equation.

By (3.3) and (3.8), $\phi = h_1 - h_2$. Since $0 < h_1, h_2 < 1$, it is easy to see that $-1 < \phi < 1$. The variance of the white noise process $\{W_t\}$ is seen to be

$$\begin{aligned}\mathrm{Var}(\mathrm{Y}_t - \phi\mathrm{Y}_{t-1}) &= \mathrm{Cov}(\mathrm{Y}_t,\mathrm{Y}_t) + \phi^2\mathrm{Cov}(\mathrm{Y}_{t-1},\mathrm{Y}_{t-1}) - 2\phi\mathrm{Cov}(\mathrm{Y}_t,\mathrm{Y}_{t-1})\\ &= M\frac{1}{\mu}(1-\frac{1}{\mu})(1-\phi^2).\end{aligned}$$

Now suppose that M is constant and that $\{Y_t\}$ satisfies the AR(1) difference equation (3.1). Then $\{Y_t\}$ has the same autocovariance function as (3.2). Thus,

$$\begin{aligned}\frac{\sigma^2\phi^h}{(1-\phi^2)} &= M\mu^{-1}(u_h - \mu^{-1}),\ h \geq 1,\\ \frac{\sigma^2}{(1-\phi^2)} &= M\mu^{-1}(1-\mu^{-1}).\end{aligned} \tag{3.9}$$

Combining the two equations in (3.9) gives

$$u_h = (1-\mu^{-1})\phi^h + \mu^{-1}. \tag{3.10}$$

Now take generating functions of both sides of (3.10) to get

$$U(s) = \Sigma_{n=0}^{\infty} u_n s^n = \frac{\mu - (\mu+\phi-1)s}{\mu(1-s)(1-\phi s)}. \tag{3.11}$$

The relationship between the generating function of the lifetime L — say $F(s) = E[S^L] = \Sigma_{n=1}^{\infty} f_n s^n$ — and $U(s)$ is $U(s) = (1 - F(s))^{-1}$(Feller, 1968). Equation (3.11) now gives

$$F(s) = \frac{1}{1 - \left(1 + \frac{\phi - 1}{\mu}\right) s} \left[\left(\phi - \frac{\phi - 1}{\mu} \right) s - \phi s^2 \right].$$

Inverting this generating function, one gets, with $r = 1 + \mu^{-1}(\phi - 1)$,

$$\begin{aligned} f_1 &= \phi - \mu^{-1}(\phi - 1); \\ f_2 &= (1 - r)(r - \phi); \\ f_n &= f_2 r^{n-2}, \ \forall\, n \geq 2. \end{aligned} \tag{3.12}$$

Now suppose that M has a Poisson distribution and that $\{Y_t\}$ satisfies the AR(1) difference equation (3.1). Then (3.9) holds with M replaced by λ:

$$\begin{aligned} \frac{\sigma^2 \phi^h}{(1 - \phi^2)} &= \lambda \mu^{-1}(u_h - \mu^{-1}), \ h \geq 1, \\ \frac{\sigma^2}{(1 - \phi^2)} &= \lambda \mu^{-1}(1 - \mu^{-1}). \end{aligned} \tag{3.13}$$

Arguing as above, it can be shown that L has a constant failure rate after lag 1. One again obtains (3.12). This finished our work. □

Proof of Theorem 6. Given the AR(1) autocorrelation function $\rho(h) = \tau^h$ for $h \geq 0$, $-1 < \tau < 1$, pick any integer $M \geq 1$ (or $\lambda > 0$ for the Poisson marginals) and any lifetime distribution L that has constant hazard rates after lag 1 with $\tau = h_1 - h_2$. Then as shown in the last proof, the series $\{Y_t\}$ generated from (2.3) satisfies an AR(1) difference equation with $\phi = \tau$. Thus, $\{Y_t\}$ has the autocorrelation function $\rho(h) = \tau^h$ for $h \geq 0$. □

Proof of Theorem 7. Arguing as in the proof of Theorem 5, if a renewal AR(1) series has the autocovariance function $\gamma(h) = \sigma^2\phi^h/(1-\phi^2)$ at lag h then (3.9) holds for the case of binomial marginals and (3.13) holds for the case of Poisson marginals. The distribution of L is given in (3.12). Hence, our work consists of identifying μ, and showing that $\mu > 1$ and the probabilities in (3.12) are legitimate lifetime probabilities. This entails

$$\begin{aligned} \mu > 1, \qquad & \text{if} \quad 0 < \phi < 1 \qquad \text{and} \\ 1 - \phi < \mu < 1 - \phi^{-1}, \qquad & \text{if} \quad -1 < \phi < 0. \end{aligned} \tag{3.14}$$

Using quadratic equations, it can be shown that any μ solving the second equation in (3.9) or (3.13) will satisfy (3.14) if there exists an integer M such that

$$\begin{aligned} M^{-1}\sigma^2(1-\phi^2)^{-1} \le 1/4, \qquad & \text{if} \quad 0 < \phi < 1; \\ -\phi(1-\phi)^{-2} < M^{-1}\sigma^2(1-\phi^2)^{-1} \le 1/4, \qquad & \text{if} \quad -1 < \phi < 0 \end{aligned} \tag{3.15}$$

for the case of binomial marginals; or if there exists a real λ such that

$$\begin{aligned} \lambda^{-1}\sigma^2(1-\phi^2)^{-1} \le 1/4, \qquad & \text{if} \quad 0 < \phi < 1; \\ -\phi(1-\phi)^{-2} < \lambda^{-1}\sigma^2(1-\phi^2)^{-1} \le 1/4, \qquad & \text{if} \quad -1 < \phi < 0 \end{aligned} \tag{3.16}$$

for the case of Poisson marginals.

It is easy to show that $-\phi(1-\phi)^{-2} < 1/4$ since $-1 < \phi < 1$. So for the Poisson case, a $\lambda > 0$ can be found that satisfies (3.16). Hence, one can always find a renewal AR(1) series with Poisson marginals that has the autocovariance function $\gamma(h) = \sigma^2\phi^h/(1-\phi^2)$ for any $\sigma^2 > 0$ and $\phi \in (-1, 1)$. This proves part a).

For part b), if the AR(1) model has $\phi > 0$, then an M satisfying (3.15) always exists (take M as large as necessary in the first equation of (3.15)). When $\phi < 0$, a proper integer as required by (3.15) is not always available. For example, when $\sigma^2(1-\phi^2)^{-1} < -\phi(1-\phi)^{-2}$, we cannot find an $M \geq 1$ such that (3.15) is satisfied, since M is at least one. That is, there is no renewal AR(1) series with binomial marginals having the autocovariance function $\gamma(h) = \sigma^2\phi^h/(1-\phi^2)$. $\square$

Remark: For each feasible value of M or λ, there are two feasible roots in step 2. The lifetime corresponding to these two roots are related: if one distribution has f_1, $f_n = f_2 r^{n-2}$ for $n \geq 2$, then the other distribution has $f_1 = r$, $f_n = f_2 f_1^{n-2}$ for $n \geq 2$.

We now summarize how to generate a renewal process with the AR(1) autocovariance function $\gamma(h) = \sigma^2\phi^h/(1-\phi^2)$ for $h \geq 0$.

1. For the binomial case, find an integer $M \geq 1$ such that: if $0 < \phi < 1$, $M^{-1}\sigma^2(1-\phi^2)^{-1} \leq 1/4$; if $-1 < \phi < 0$, $-\phi(1-\phi)^{-2} < M^{-1}\sigma^2(1-\phi^2)^{-1} \leq 1/4$. For the Poisson case, find a real $\lambda > 0$ such that: if $0 < \phi < 1$, $\lambda^{-1}\sigma^2(1-\phi^2)^{-1} \leq 1/4$; if $-1 < \phi < 0$, $-\phi(1-\phi)^{-2} < \lambda^{-1}\sigma^2(1-\phi^2)^{-1} \leq 1/4$.

2. In the binomial case solve $\mu^{-1}(1-\mu^{-1}) = M^{-1}\sigma^2(1-\phi^2)^{-1}$ for μ. In the Poisson case, solve $\mu^{-1}(1-\mu^{-1}) = \lambda^{-1}\sigma^2(1-\phi^2)^{-1}$ for μ.

3. The lifetime distribution parameters are then given by (3.12).

Bibliography

M. Al-Osh and A. Alzaid. First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8:261–275, 1987.

M. Al-Osh and A. Alzaid. An integer-valued pth-order autoregressive (INAR(p)) process. *Journal of Applied Probability*, 27:314–324, 1990.

M. Al-Osh and A. Alzaid. Some gamma processes based on the dirichlet-gamma transformation. *Communications in Statistics - Stochastic Models*, 9:123–143, 1993.

P. Albert. A Markov model for sequences of ordinal data from a relapsing-remitting disease. *Biometrics*, 50:51–60, 1994.

M. Benjamin, R. Rigby, and D. Stasinopoulos. Generalized autoregressive moving average models. *Journal of Applied Probability*, 98:214–223, 2003.

K. Berenhaut and R. Lund. Geometric renewal convergence rates from hazard rates. *Journal of Applied Probability*, 38:180–194, 2001.

P. Blight. Time series formed from the superposition of discrete renewal processes. *Journal of Applied Probability*, 26:189–195, 1989.

U. Böckenholt. Autoregressive negative multinomial regression models. *Journal of the American Statistical Association*, 94:757–765, 1999a.

U. Böckenholt. Mixed INAR(1) poisson regression models: analyzing heterogeneity and serial dependencies in longitudinal count data. *Journal of Econometrics*, 89: 317–338, 1999b.

U. Böckenholt. Analysing state dependences in emotional experiences by dynamic count data models. *Journal of the Royal Statistical Society Series C*, 52:213–226, 2003.

K. Brännäs. Estimation and testing in integer valued AR(1) models. Umeå Economic Studies, 355, 1994.

K. Brännäs and J. Hellström. Generalized integer-valued autoregression. *Econometric Reviews*, 20:425–443, 2001.

K. Brännäs and P. Johansson. Time series count regression. *Communication in Statistics: Theory and Methods*, 23:2907–2925, 1994.

K. Brännäs and Q. Shahiduzzaman. Integer-valued moving average modelling of the number of transactions in stocks. *Umeå Economic Studies*, 637, 2004.

P. Brockwell and R. Davis. *Time Series: Theory and Methods, 2nd edition*. Springer, New York, 1991.

P. Brockwell and R. Davis. *Introduction to Time Series and Forecasting, 2nd edition*. Springer, New York, 2002.

A. Cameron and L. Leon. Markov regression models for time series data. Presented at Western Economic Association Meetings, Lake Tahoe, NV, 1993.

A. Cameron and P. Trivedi. *Regression Analysis of Count Data*. Cambridge University Press, London, 1998.

M. Campbell. Time series regression for counts: an investigation into the relationship between sudden infant death syndrome and environmental temperature. *Journal of the Royal Statistical Society Series A*, 157:191–120, 1994.

J. Du and Y. Li. The integer valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, 12:129–142, 1991.

L. Fahrmeir and G. Tutz. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag, New York, 1994.

W. Feller. *An Introduction to Probability Theory and Its Applications, Volume I, 3rd edition*. Wiley, New York, 1968.

K. Fokianos and B. Kedem. Regression theory for categorical time series. *Statistical Science*, 18:357–376, 2003.

J. Franke and T. Seligmann. Conditional maximum-likelihood estimates for INAR(1) processes and their applications to modelling epileptic seizure counts. In T. Edited by Rao, editor, *Developments in Time Series*, pages 310–330. Chapman & Hall, London, 1993.

R. Freeland and B. McCabe. Analysis of low count time series data by poisson autoregression. *Journal of Time Series Analysis*, 25:701–722, 2004.

G. Gauthier and A. Latour. Convergence forte des estimateurs des paramètres d'un processus GENAR(p). *Annales des Sciences Mathematiques du Québec*, 18:49–71, 1994.

B. Hansen and J. Frenk. Some monotonicity properties of the delayed renewal function. *Journal of Applied Probability*, 28:811–821, 1991.

A. Harvey. *Forecasting, Structural Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, 1989.

A. Harvey and C. Fernandes. Time series models for count or qualitative observations. *Journal of Business and Economic Statistics*, 7:407–417, 1989.

P. Holgate. Estimation for the bivariate poisson distribution. *Biometrika*, 51:241–245, 1964.

P. Jacobs and P. Lewis. Discrete time series generated by mixtures I: correlational and runs properties. *Journal of the Royal Statistical Society Series B*, 40:94–105, 1978a.

P. Jacobs and P. Lewis. Discrete time series generated by mixtures II: asymptotic properties. *Journal of the Royal Statistical Society Series B*, 40:222–228, 1978b.

P. Jacobs and P. Lewis. Stationary discrete autoregressive-moving average time series generated by mixtures. *Journal of Time Series Analysis*, 4:19–36, 1983.

P. Johansson. Speed limitation and motorway casualties: a time series count data regression approach. *Accident Analysis and Prevention*, 28:73–87, 1996.

D. Kendall. Unitary dilations of Markov transition operators, and the corresponding integral representations for transition-probability matrices. In U. Grenander, editor, *Probability and Statistics*, pages 139–161. John Wiley, New York, 1959.

L. Klimko and P. Nelson. On conditional least squares estimation for stochastic processes. *Annals of Statistics*, 6:629–642, 1978.

S. Kocherlakota and K. Kocherlakota. *Bivariate Discrete Distributions*. Marcel Dekker, New York, 1992.

A. Latour. The multivariate GINAR(p) processes. *Advances in Applied Probability*, 29:228–248, 1997.

A. Latour. Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, 19:439–455, 1998.

A. Lawrance and P. Lewis. An exponential moving average sequence and point process, EMA(1). *Journal of Applied Probability*, 14:98–113, 1977a.

A. Lawrance and P. Lewis. The exponential autoregressive-moving average EARMA(p, q) process. *Journal of the Royal Statistical Society Series B*, 42:150–161, 1977b.

W. Li. Testing model adequacy for some Markov regression models for time series. *Biometrika*, 78:83–89, 1991.

W. Li. Time series models based on generalized linear models: some further results. *Biometrics*, 50:506–511, 1994.

T. Lindvall. *Lectures on the Coupling Method*. John Wiley & Sons, New York, 1992.

S. Loukas, C. Kemp, and H. Papageorgiou. Even point estimation for the bivariate poisson distribution. *Biometrika*, 73:222–223, 1986.

I. MacDonald and W. Zucchini. *Hidden Markov and Other Models for Discrete-Valued Time Series*. Chapman and Hall, London, 1997.

A. Marshall and I. Olkin. A family of bivariate distributions generated by the bivariate bernoulli distribution. *Journal of the American Statistical Association*, 80:332–338, 1985.

E. McKenzie. Some simple models for discrete variate time series. *Water Resources Bulletin*, 21:645–650, 1985.

E. McKenzie. Autoregressive-moving average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability*, 18:679–705, 1986.

E. McKenzie. Some ARMA models for dependent sequences of poisson counts. *Advances in Applied Probability*, 20:822–835, 1988.

E. McKenzie. Discrete variate time series. In C. Rap. and D. Shanbhag, editors, *Stochastic Processes: Modelling and Simulation, Handbook of Statistics*, volume 21, pages 573–606. Elsevier Science B.V., North-Holland, Amsterdam, 2003.

J. Pitman. Uniform rate of convergence for Markov chain transition probabilities. *Probability Theory and Related Fields*, 29:193–227, 1974.

G. Ronning and R. Jung. Estimation of a first order autoregressive process with poisson marginals for count data. In L. Fahrmeir, editor, *Advances in GLIM and Statistical Modelling*. Springer-Verlag, New York, 1992.

S. Ross. *Simulation, 4th edition*. Academic Press, Boston, 2006.

N. Rudholm. Entry and the number of firms in the Swedish pharmaceuticals market. *Review of Industrial Organization*, 19:351–364, 2001.

M. Sørensen. Prediction-based estimating functions. *Econometrics Journal*, 3:123–137, 2000.

F. Steutel and K. V. Harn. Discrete analogues of self-decomposability and stability. *Annals of Probability*, 7:893–899, 1979.

C. Weiβ. A new class of autoregressive models for time series of binomial counts. *Communications in Statistics, Theory and Methods*, 38(447-460), 2009.

S. Zeger. A regression model for time series of counts. *Biometrika*, 75:621–629, 1988.

S. Zeger and B. Qaqish. Markov regression models for time series quasi-likelihood approach. *Biometrics*, 44:1019–1031, 1988.