Probability in the Engineering and Informational Sciences, **35**, 2021, 538–556.
doi:10.1017/S0269964819000433

SUPERPOSITIONED STATIONARY COUNT TIME SERIES

Yisu Jia
Department of Mathematics and Statistics, University of North Florida, 1 UNF Drive, Jacksonville, FL 32224, USA
E-mail: y.jia@unf.edu

Robert Lund
Department of Mathematical Sciences, Clemson University, O-110 Martin Hall, Box 340975, Clemson, S.C. 29634-0975, USA

James Livsey
U.S. Census Bureau, Center for Statistical Research and Methodology, 4600 Silver Hill Rd, Washington, DC 20233, USA

This paper probabilistically explores a class of stationary count time series models built by superpositioning (or otherwise combining) independent copies of a binary stationary sequence of zeroes and ones. Superpositioning methods have proven useful in devising stationary count time series having prespecified marginal distributions. Here, basic properties of this model class are established and the idea is further developed. Specifically, stationary series with binomial, Poisson, negative binomial, discrete uniform, and multinomial marginal distributions are constructed; other marginal distributions are possible. Our primary goal is to derive the autocovariance function of the resulting series.

Keywords: count-valued processes, negative binomial distribution, poisson distribution, stationary processes, time series

1. INTRODUCTION

There has been significant recent interest in modeling stationary series with prescribed discrete marginal distributions; see Kachour and Yao [16], Weiß [31,32], Davis, Holan, Lund, and Ravishanker [6]. Often, the discreteness arises in the form of counts taking values in $\{0, 1, 2, \ldots\}$. Count series arise when describing storm numbers, accident tallies, wins by a sports team, disease cases, etc. While series having Gaussian marginal distributions are often parsimoniously described by the autoregressive moving-average (ARMA) model class and its variants (Holan, Lund, and Davis [11]), no one model class dominates the count series literature. In fact, the autocovariance function of many commonly used count models is deficient in some sense (only non-negative correlations can be produced, for example). The purpose of this paper is to probabilistically develop and explore the covariance properties of a flexible class of count series built from stationary binary series.

 0269-9648/19 $25.00

A recently proposed class of count time series models involves combining binary series in various ways to achieve the desired marginal distribution (see [4,19]), extending the idea that any discrete random variable can be built from Bernoulli trials. From the flexibility of binary autocovariances, this model class inherits flexible autocovariances. Many of the series can be written in the superpositioned form

$$X_t = \sum_{i=1}^{M_t} B_{t,i}, \tag{1}$$

where X_t is the count at time t, $\{B_{t,i}\}$ are independent and identically distributed (IID) copies of stationary but time-correlated Bernoulli processes in i, and $\{M_t\}$ is a sequence of IID count-valued random variables with finite first moment that is independent of the $\{B_{t,i}\}$s. The sum in (1) is taken as zero if $M_t = 0$. Superpositioning was advocated by Blight [3] and used further in Cui and Lund [4] to build stationary series with binomial and Poisson marginal distributions. Here, we expand upon this paradigm and construct series having a wide range of marginal distributions, including negative binomial, discrete uniform, and multinomial. The autocovariances of these series can be positive or negative and/or have long-memory features.

There is no known result characterizing autocovariances of stationary count series. Elaborating, if $\gamma(\cdot)$ is a symmetric non-negative definite function on the integers, then there exists a stationary Gaussian sequence $\{X_t\}$ having $\text{cov}(X_t, X_{t+h}) = \gamma(h)$ for all integers h. Unfortunately, no analogous result exists for say, a stationary series with Poisson marginal distributions. In fact, restrictions on autocovariance functions of count time series are often more stringent than just non-negative definiteness. For example, it may not be possible to have a stationary count series with a specific marginal distribution that is highly negatively correlated at some lags.

It is common to specify the marginal distribution or other specific features in count time series models and applications. For example, Poisson ARMA models are introduced in McKenzie [23] and Weiß [30] posits Poisson marginals for IP address and claim counts. Binomial marginal distributions are adopted in Weiß [31,32]. Negative binomial applications are pursued in Hilbe [10] and Ver Hoef and Boveng [29] compare negative binomial and generalized Poisson marginal distributions. Negative binomial marginals are further discussed in Zhu and Joe [38] from a thinning-based self-decomposable approach. Integer-valued GARCH models are studied in Zhu [36]; Zhu [37] considers overdispersed count distributions such as negative binomial and generalized Poisson.

The rest of this paper proceeds as follows. The next section reviews stationary count time series, discussing several classic count time series models of the past. Section 3 introduces several ways of modeling stationary binary sequences. Section 4 superimposes these zero/one sequences to obtain our class of count time series models. Section 5 builds stationary binomial, Poisson, negative binomial, and multinomial models via superpositioning. Section 6 presents non-stationarity extensions of the methods and Section 7 closes with comments.

2. BACKGROUND

Initial attempts to model stationary count series used the discrete-valued autoregressive moving-average (DARMA) methods introduced in Jacobs and Lewis [12,13]. For example, a first order discrete autoregressive (DAR(1)) series $\{X_t\}$ is built from IID count variables $\{Y_t\}$, with marginal cumulative distribution $F(\cdot)$, and an IID Bernoulli sequence $\{V_t\}$ with

$\mathbb{P}(V_t = 1) := p \in [0, 1]$. The count series is initialized with $X_0 = Y_0$ and then recursively updated via

$$X_t = V_t X_{t-1} + (1 - V_t) Y_t, \quad t = 1, 2, \ldots. \tag{2}$$

Induction will show that X_t has distribution $F(\cdot)$ for every $t \in \{1, 2, \ldots\}$.

While one can have any marginal distribution for a DAR(1) series (in addition to discrete structures), there are some undesirable properties of DARMA series. Foremost, in a DAR(1) model, X_t and X_{t-1} are either equal or independent. Since $\mathbb{P}(X_t = X_{t-1}) \geq p$, sample paths can remain constant for long runs, especially for larger p. To resolve this 'repeating property', a recent modification is introduced in Gouveia, Möller, Weiß, and Scotto [8]. Secondly, DAR(1) models cannot produce negatively correlated series. This is because p must lie in $[0, 1]$. In fact, one can show that DAR(1) autocorrelations have the form $\text{corr}(X_t, X_{t+h}) = p^h$ for all $h \geq 0$. While higher order autoregressions can be built by introducing additional IID Bernoulli trial sequences, negative correlations cannot be achieved with any DAR formulation. See Livsey, Lund, Kechagais, and Pipiras [18] for a hurricane example where negatively correlated count series are encountered. In short, DARMA models do not span all stationary count autocovariances. For these reasons and more, DARMA models fell out of favor in the 1980s.

Another count model class, and one that is still popular today, is the integer ARMA (INARMA) models. INARMA models were introduced in Steutel and Van Harn [27] and studied further in Al-Osh and Alzaid [1] and McKenzie [21–23]. For example, a first-order integer autoregressive (INAR(1)) model for $\{X_t\}$ obeys the recursion

$$X_t = p \circ X_{t-1} + \epsilon_t. \tag{3}$$

Here, $\circ$ denotes a thinning operator that acts on a count-valued random variate Y via $p \circ Y := \sum_{i=1}^{Y} B_i$, where $\{B_i\}_{i=1}^{\infty}$ is a sequence of Bernoulli trials with $\mathbb{P}(B_i = 1) = p \in [0, 1]$ and $\{\epsilon_t\}$ is a sequence of IID count-valued random variables with a finite second moment.

Unlike DARMA series, INARMA sample paths do not tend to stay constant for long runs; however, like the DARMA class, most INARMA-based models cannot have negative correlations. In fact, the INAR(1) model has $\text{corr}(X_t, X_{t+h}) = p^h$ for any $h \geq 0$. One can construct higher order integer autoregressions and even add moving-average components as in McKenzie [21–23]. Modifications can be made in some settings to INARMA models to allow for negative autocorrelations, such as the binomial AR models in McKenzie [21] and Weiß [31,32]. Unlike DARMA series, it is not clear how to obtain any marginal distribution with INARMA methods—this may be hard or impossible, depending on the marginal distribution desired.

Another count series model type was proposed in Joe [15] and easily produces stationary series whose marginal distribution lies in the so-called convolution-closed infinitely divisible class. Suppose that F_θ is a family of marginal distributions whose convolution, denoted by $*$, satisfies $F_{\theta_1} * F_{\theta_2} = F_{\theta_1+\theta_2}$. For the first order autoregressive case, the count series $\{X_t\}$ obeys the recursion

$$X_t = A_t(X_{t-1}; \alpha) + \epsilon_t, \tag{4}$$

where $\{\epsilon_t\}$ are IID variables with a finite second moment having the marginal distribution $F_{(1-\alpha)\theta}$, $\alpha \in [0, 1]$, and ϵ_t is independent of X_j for $j < t$. The stochastic operator A_t, which is IID in time t, is defined so that $A_t(Y; \alpha)$ is a random variable whose marginal distribution is $F_{\alpha\theta}$ (see Joe [15] for details). These models capably describe many marginal distributions—both discrete and continuous—and include gamma, beta, normal, Poisson, negative binomial, and generalized Poisson. However, the marginal distribution must come from the

https://doi.org/10.1017/S0269964819000433 Published online by Cambridge University Press

convolution-closed class. Unfortunately, again the correlations of these models cannot be negative.

Recently, Kachour and Yao [16] produced negatively correlated count series models by rounding Gaussian series. For example, a rounded autoregressive model of order p with location parameter μ and autoregressive parameters $\phi_1, \ldots, \phi_p$ obeys

$$X_t = \left\langle \mu + \sum_{j=1}^{p} \phi_j X_{t-j} \right\rangle + \epsilon_t, \tag{5}$$

where $\langle x \rangle$, for $x \geq 0$, rounds x to its nearest integer (round down should this be non-unique), and $\{\epsilon_t\}$ is count-valued IID noise. While such $\{X_t\}$ can have negative correlations, due to the rounding, it is difficult to construct a pre-specified marginal distribution in this model class.

While the above models all have some attractive features and individualized drawbacks, none of them easily generate negatively correlated count time series with pre-specified marginal distributions. While autocovariances of a stationary count series can indeed be negative, there are bounds on the degree of negativity possible. Elaborating, if $\{X_t\}$ is a stationary count time series with marginal distribution $F(\cdot)$, define $\rho+ = \max\{\mathrm{corr}(X_{t_1}, X_{t_2})\}$ and $\rho- = \min\{\mathrm{corr}(X_{t_1}, X_{t_2})\}$ as the maximum and minimum achievable correlations. Theorem 2.5 of Whitt [34] shows that

$$\rho+ = \mathrm{corr}(F^{-1}(U), F^{-1}(U)) = 1, \quad \rho- = \mathrm{Corr}(F^{-1}(U), F^{-1}(1-U)), \tag{6}$$

where U is a uniform random variable on (0,1). In a pairwise sense, $\rho-$ is the most negative correlation possible for the marginal distribution $F(\cdot)$. This bound is studied further below. More about bivariate distributions and correlation bounds are discussed in Lin, Dou, Kuriki, and Huang [17].

3. STATIONARY ZERO-ONE SERIES

Our count time series will be built from a stationary zero-one (binary) random sequence $\{B_t\}$. This strategy was used in Blight [3] and further developed in Cui and Lund [4] and Lund and Livsey [19]. The idea can be viewed as the time series extension of the fact that any discrete-valued distribution can be constructed from independent fair coin flips (how to efficiently do this is another matter).

For notation, let $p_B = \mathbb{E}[B_t] \equiv \mathbb{P}(B_t = 1)$ be the mean of $\{B_t\}$ and denote its lag h autocovariance by $\gamma_B(h) = \mathrm{cov}(B_t, B_{t+h})$. Of course,

$$\gamma_B(h) = \mathbb{P}(B_t = 1 \cap B_{t+h} = 1) - p_B^2 = p_B[\mathbb{P}(B_{t+h} = 1 | B_t = 1) - p_B]. \tag{7}$$

Two quantities that will be important later are the h-step-ahead transition probabilities to a unit point: $p_{1,1}(h) := \mathbb{P}(B_{t+h} = 1 | B_t = 1)$ and $p_{0,1}(h) := \mathbb{P}(B_{t+h} = 1 | B_t = 0)$. The h-step-ahead transition probabilities to a "zero point" are $p_{1,0}(h) := \mathbb{P}(B_{t+h} = 0 | B_t = 1) = 1 - p_{1,1}(h)$ and $p_{0,0}(h) := \mathbb{P}(B_{t+h} = 0 | B_t = 0) = 1 - p_{0,1}(h)$.

In the following subsections we discuss two models for $\{B_t\}$; others are possible.

3.1. Discrete Renewal Processes

One method for a Bernoulli process construction uses the renewal times in a stationary discrete-time renewal process as in Blight [3], Cui and Lund [4], and Lund and Livsey [19]. A stationary renewal process employs an initial random "lifetime" $L_0 \in \{0, 1, \ldots\}$ (delay) and a sequence of IID aperiodic lifetimes $\{L_i\}_{i=1}^{\infty}$ supported in $\{1, 2, \ldots\}$ with $\mu_L := \mathbb{E}[L_1] < \infty$. In what follows, L will denote a lifetime whose distribution is equivalent to any of $L_1, L_2, \ldots$. The random walk $\{S_n\}_{n=0}^{\infty}$ associated with the renewal process obeys $S_n = \sum_{i=0}^{n} L_i$ for $n \in \{0, 1, \ldots\}$. A renewal is said to have occurred at time t if $S_n = t$ for any n. The zero-one process $\{B_t\}$ is simply set to unity at all renewal times: $B_t = \mathbf{1}_{[\cup_{n=0}^{\infty}\{S_n = t\}]}$. To have $\{B_t\}$ stationary, L_0 needs to have the so-called first-derived distribution from the tails of L_1 [9]:

$$\mathbb{P}(L_0 = k) = \frac{\mathbb{P}(L_1 > k)}{\mu_L}, \quad k \in \{0, 1, \ldots\}. \tag{8}$$

Stationarity of $\{B_t\}$ gives

$$p_B = \mathbb{P}(B_t = 1) = \mathbb{P}(B_0 = 1) = \mathbb{P}(L_0 = 0) = \frac{\mathbb{P}(L_1 > 0)}{\mu_L} = \frac{1}{\mu_L}. \tag{9}$$

A discrete renewal process with L_0 distributed as in (8) is called stationary-delayed. Alternatively, if $L_0 = 0$ with probability one, the process is called zero-delayed. Define u_h to be the probability of a renewal at time h in a zero-delayed process:

$$u_h := \mathbb{P}(B_t = 1 | B_0 = 1) = \mathbb{P}(B_{t+h} = 1 | B_t = 1). \tag{10}$$

Equation (7) yields

$$\gamma_B(h) = \mathrm{cov}(B_t, B_{t+h}) = \frac{1}{\mu_L}\left(u_h - \frac{1}{\mu_L}\right). \tag{11}$$

Notice that $\gamma_B(h) < 0$ if and only if $u_h < \mu_L^{-1}$, which happens for many L. For example, any lifetime with $\mathbb{P}(L = 1) = 0$ has $\gamma_B(h) = -1/\mu_L^2$, which is negative. The parameters in this model are those that describe the lifetime L. Observe that $p_{1,1}(h) = u_h$ and $p_{0,1}(h) = p_B(1 - u_h)/(1 - p_B)$.

3.2. Clipped Gaussian Sequences

Another binary $\{B_t\}$ is built from a correlated latent Gaussian process $\{Z_t\}$ as in Livsey, Lund, Kechagais, and Pipiras [18]. Specifically, let $\{Z_t\}$ be a correlated zero-mean unit-variance Gaussian random processes with $\mathrm{corr}(Z_t, Z_{t+h}) = \rho_Z(h)$. Set $B_t = \mathbf{1}_{(Z_t > \kappa)}$ for some preset real κ. Then $\{B_t\}$ is a strictly stationary binary sequence with $p_B = \mathbb{P}(B_t = 1) = 1 - \Phi(\kappa)$; here, $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal random variable. This construct is very similar to the clipping strategies in Van Vleck and Middleton [28].

The autocovariance function of $\{B_t\}$ can be derived from bivariate normal probabilities. As an illustration, suppose that $\kappa = 0$ so that $p_B = 1/2$. Then a classical multivariate normal orthant probability calculation gives

$$\mathbb{E}(B_t B_{t+h}) = \mathbb{P}(Z_t > 0 \cap Z_{t+h} > 0) = \frac{1}{4} + \frac{\arcsin(\rho_Z(h))}{2\pi}, \tag{12}$$

where Rose and Smith [26] is used. Since $\mathbb{E}[B_t] \equiv 1/2$,

$$\gamma_B(h) = \frac{\arcsin(\rho_Z(h))}{2\pi}, \quad \rho_B(h) = \frac{2\arcsin(\rho_Z(h))}{\pi}. \tag{13}$$

In this case, lag h autocovariances and autocorrelations are negative if and only if $\rho_Z(h) < 0$. Hence, this model can produce negative covariances; in fact, $\rho_B(h)$ can take on any value in $[-1, 1]$ when $p_B = 1/2$.

When $p_B \neq 1/2$, one will need to solve $p_B = 1 - \Phi(\kappa)$ for κ. In this case,

$$\mathbb{E}[B_t B_{t+h}] = \mathbb{P}(B_t = 1, B_{t+h} = 1) = \mathbb{P}(Z_t > \kappa, Z_{t+h} > \kappa). \quad \textbf{(14)}$$

Evaluating the covariance function in this case requires integrating the bivariate normal density over an infinite rectangle that is not an orthant. With $G(\kappa, \rho_Z(h)) := \mathbb{P}(Z_t > \kappa, Z_{t+h} > \kappa)$, which can be evaluated via many computer packages, we have

$$\gamma_B(h) = G(\kappa, \rho_Z(h)) - (1-\Phi(\kappa))^2, \quad \rho_B(h) = \frac{G(\kappa, \rho_Z(h)) - (1-\Phi(\kappa))^2}{\Phi(\kappa)(1-\Phi(\kappa))}. \quad \textbf{(15)}$$

Point probabilities for $\{B_t\}$ are

$$p_{1,1}(h) = \frac{G(\kappa, \rho_Z(h))}{1-\Phi(\kappa)} \quad p_{0,1}(h) = \frac{1-\Phi(\kappa) - G(\kappa, \rho_Z(h))}{\Phi(\kappa)}. \quad \textbf{(16)}$$

3.3. Binary Processes

Suppose that $\{B_t\}$ is a general zero-one stationary binary process with $\mathbb{P}[B_t = 1] = p_B$. While stationary binary process with arbitrary positive correlations exist, for a fixed p_B, straightforward computations show that lag h correlations can be no more negative than

$$\rho- = \begin{cases} 1 - \frac{1}{1-p_B}, & 0 \leq p_B \leq \frac{1}{2}; \\ 1 - \frac{1}{p_B}, & \frac{1}{2} < p_B \leq 1 \end{cases} \quad \textbf{(17)}$$

(compare to the bounds given at the end of 2).

Consider the case where $p_B \in [0, 1/2)$ and $\{B_t\}$ is generated with the renewal methods of Section 3.1. The lag h correlation is $\rho_B(h) = (u_h - p_B)/(1-p_B)$, which achieves the bound in (17) whenever $u_h = 0$. When $h = 1$, any lifetime L with $\mathbb{P}(L = 1) = 0$ and $\mu_L = 1/p_B > 2$ will have correlation $\rho-$. Cases for higher h are more involved, but are straightforward to investigate.

When $p_B = 1/2$, $h = 1$, and $\epsilon > 0$ is small, a lifetime L with $\mathbb{P}[L = 1] = \mathbb{P}[L = 3] = \epsilon$ and $\mathbb{P}[L = 2] = 1 - 2\epsilon$ gives $\mu_L = 1/p_B = 2$ and $\rho_B(1) = 2\epsilon - 1$, which comes arbitrarily close to $p- = -1$ as $\epsilon \downarrow 0$ (one cannot take $L \equiv 2$ as the support set of L is assumed aperiodic).

When $p_B \in (1/2, 1)$ and $h = 1$, a lifetime L with $\mathbb{P}(L = 1) = 2 - 1/p_B$ and $\mathbb{P}(L = 2) = 1/p_B - 1$ will achieve $\rho- = 1 - 1/p_B$. Again, cases for a higher h are straightforward to investigate.

Now suppose that $\{B_t\}$ is generated by the clipped Gaussian methods of Section 3.2. The lag h autocorrelation in $\{B_t\}$ is

$$\rho_B(h) = \frac{\mathbb{P}(Z_t > \kappa, Z_{t+h} > \kappa) - p_B^2}{p_B(1-p_B)}, \quad \textbf{(18)}$$

and $\rho_B(h)$ is "maximally negative" when $\mathbb{P}(Z_t > \kappa, Z_{t+h} > \kappa) = 0$. When $p_B \in [0, 1/2]$, $\kappa > 0$ and one again chooses $\rho_Z(h) = -1$.

When $p_B \in (1/2, 1)$, $\kappa < 0$. The autocorrelation in (18) achieves $\rho-$ when $\mathbb{P}(Z_t > \kappa, Z_{t+h} > \kappa) = 2p_B - 1$. This is equivalent to requiring that $\mathbb{P}(Z_t < \kappa, Z_{t+h} < \kappa) = 0$, which again happens by choosing $\rho_Z(h) = -1$.

4. SUPERPOSITIONING

We now move to superpositioned count series. Let $\{B_{t,i}\}$ for $i \in \{1, 2, \ldots\}$ denote IID copies of $\{B_t\}$. Our count series $\{X_t\}$ is built by adding (superimposing) a random number of IID copies of $\{B_t\}$ as in (1). Again, $\{M_t\}$ is an IID count-valued random sequence with assumed finite second moments that is independent of all $\{B_{t,i}\}$.

Let $\mathbb{E}[M_t] = \mu_M$ and $\text{var}(M_t) = \sigma_M^2$. It is obvious that $\{X_t\}$ in (1) is a count-valued strictly stationary random sequence with mean $\mathbb{E}[X_t] \equiv p_B \mu_M$. The following result establishes additional properties of $\{X_t\}$.

THEOREM 1: *Let $\{X_t\}$ be the strictly stationary count series in (1). Then*

(a) *The probability generating function of X_t has form $\psi_X(u) := \mathbb{E}[u^{X_t}] = \psi_M(1 - p_B + p_B u)$, where $\psi_M(u) := \mathbb{E}[u^{M_t}]$ is the probability generating function of M_t.*

(b) *The dispersion of $\{X_t\}$ is $D_X := var(X_t)/\mathbb{E}[X_t] = 1 + p_B(D_M - 1)$, where $D_M := \sigma_M^2/\mu_M$ is the dispersion of M_t. X_t is over/under dispersed if and only if M_t is over/under dispersed.*

(c) *The lag h autocovariance of $\{X_t\}$ has the form*

$$\gamma_X(h) = \begin{cases} \kappa\gamma_B(h), & h \neq 0; \\ \gamma_B(0)\mu_M + p_B^2\sigma_M^2, & h = 0; \end{cases} \tag{19}$$

where

$$\kappa = \mathbb{E}[\min(M_1, M_2)] = \sum_{k=0}^{\infty} \mathbb{P}(M_t > k)^2. \tag{20}$$

(d) *The lag h bivariate probability distributions of $\{X_t\}$ have form*

$$\mathbb{P}(X_t = x_t, X_{t+h} = x_{t+h}) = \sum_{m_t=0}^{\infty} \sum_{m_{t+h}=0}^{\infty} r_{m_t, m_{t+h}}(x_t, x_{t+h}) f_M(m_t) f_M(m_{t+h}),$$

where $f_M(k) = \mathbb{P}(M_t = k)$ and $r_{m_t,m_{t+h}}(x_t, x_{t+h})$ has the form identified in the Appendix.

(e) *In the renewal case, $\{X_t\}$ has long memory if and only if $\mathbb{E}[L^2] = \infty$. In the clipped Gaussian case, $\{X_t\}$ has long memory if and only if $\{Z_t\}$ has long memory.*

This theorem is proven in the Appendix. Attempts to derive higher order (beyond bivariate) joint process distributions have not produced tractable expressions to date. This is unfortunate as it precludes using the processes' joint distribution to construct likelihood-based parameter estimators; nonetheless, the bivariate distribution above allows one to compute composite likelihood estimators as in Pedeli and Karlis [25] and Ng, Joe, Karlis, and Liu [24]. Particle filtering techniques show promise in approximating process likelihoods (see Jia, Kechagias, Livsey, Lund, and Pipiras [14]). The covariance structure of the model enables pseudo-Gaussian likelihood parameter estimation as in Livsey, Lund, Kechagais, and Pipiras [18]. Estimation issues for these and other count model classes are considered in Jia, Kechagias, Livsey, Lund, and Pipiras [14].

5. CLASSICAL COUNT MARGINAL DISTRIBUTIONS

This section constructs stationary time series with the classical count marginal distributions: binomial, Poisson, negative binomial, and multinomial.

5.1. Binomial Marginals

Count time series with binomial marginal distributions with M trials and success probability p_B are easily obtained: just take M_t equal to the constant M. The binomial distribution is under-dispersed with $D_X = 1 - p_B$. This model was introduced by Blight [3] and studied further in Cui and Lund [4,5] and Weiß [31,32]. By part (c) of Theorem 4.1, when $h \neq 0$, the lag h autocovariance and autocorrelation of $\{X_t\}$ are

$$\gamma_X(h) = M\gamma_B(h), \quad \rho_X(h) = \frac{\gamma_B(h)}{p_B(1-p_B)}, \tag{21}$$

and $\gamma_X(0) = Mp_B(1-p_B)$, $\rho_X(0) = 1$. By (11), in the renewal case, the lag h autocovariance and autocorrelation of $\{X_t\}$ are

$$\gamma_X(h) = \frac{M}{\mu_L}\left(u_h - \frac{1}{\mu_L}\right), \quad \rho_X(h) = \frac{\frac{1}{\mu_L}\left(u_h - \frac{1}{\mu_L}\right)}{p_B(1-p_B)}. \tag{22}$$

From our derived expressions in the clipped Gaussian case, the lag h autocovariance and autocorrelation of $\{X_t\}$ are, for $h \neq 0$,

$$\gamma_X(h) = M\left[G(\kappa, \rho_Z(h)) - p_B^2\right], \quad \rho_X(h) = \frac{G(\kappa, \rho_Z(h)) - p_B^2}{p_B(1-p_B)}. \tag{23}$$

Figure 1 shows a simulated realization of a binomial series with $M = 5$. Here, the $B_{t,i}$ were generated from a renewal process with lifetime L having a Pareto distribution with parameter $\alpha = 2.1$: $\mathbb{P}(L = k) = C(\alpha)/k^{\alpha}$, for $k = 1, 2, \ldots$. The constant $C(\alpha)$ makes the probability mass sum to unity; there is no explicit form for $C(\alpha)$. In this case, the renewal lifetime L has a finite mean, $\mu_L \approx 3.57$, which gives $p_B \approx 0.28$. Moreover, L has an infinite second moment, which implies from part (e) of Theorem 4.1 that this series has long memory. Thus, the series has binomial marginal distributions with five trials, success probability of approximately 0.28, and long memory.

5.2. Poisson Marginals

To construct a count time series $\{X_t\}$ with Poisson marginal distributions with mean $\lambda > 0$, let $\{M_t\}$ be an IID Poisson sequence with mean 2λ and $p_B = 1/2$. It is easy to show that X_t in (1) has a Poisson distribution with mean λ (see [4]). The Poisson distribution has unit dispersion. The lag h autocovariance of this process has form $\gamma_X(h) = \kappa\gamma_B(h)$ for $h \neq 0$, where $\kappa = \mathbb{E}[\min(M_t, M_{t+h})]$ and M_t and M_{t+h} are independent Poisson random variables with mean 2λ, which is derived in Lund and Livsey [19] as

$$\kappa = 2\lambda\left\{1 - e^{-4\lambda}[I_0(4\lambda) + I_1(4\lambda)]\right\}, \tag{24}$$

where $I_i(x)$ is the modified Bessel function

$$I_i(x) = \sum_{n=0}^{\infty} \frac{(x/2)^{2n+i}}{n!(n+i)!}, \quad i \in \{0, 1\}. \tag{25}$$

When $h = 0$, $\mathbb{E}[\min(M_t, M_{t+h})] = 2\lambda$. By Theorem 4.1, the lag h autocovariance of $\{X_t\}$ is $\gamma_X(h) = \kappa\gamma_B(h)$ when $h \neq 0$ and $\gamma_X(0) = \lambda$.

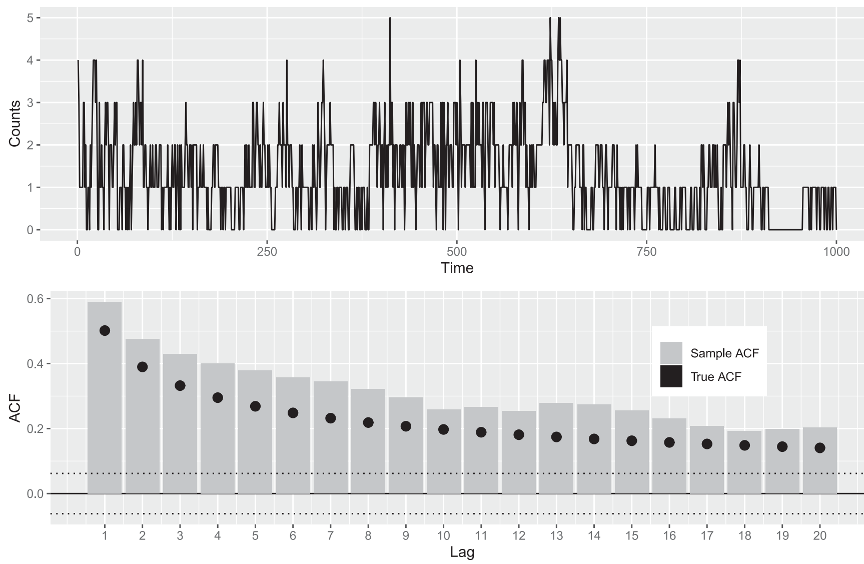

FIGURE 1. A realization of a stationary count time series with a Bin(5, 0.28) marginal distribution and long memory. Sample and true autocorrelations are shown with dotted pointwise 95% confidence bands for white noise.

In the renewal case, autocovariance and autocorrelation functions are

$$\gamma_X(h) = \frac{\kappa}{\mu_L}\left(u_h - \frac{1}{\mu_L}\right), \quad \rho_X(h) = \frac{\kappa}{\lambda\mu_L}\left(u_h - \frac{1}{\mu_L}\right). \tag{26}$$

Observe that $\gamma_X(h) < 0$ whenever $u_h < \mu_L^{-1}$, which happens for many renewal lifetime distributions. In the clipped Gaussian case,

$$\gamma_X(h) = \kappa\left[G(\kappa, \rho_Z(h)) - \frac{1}{4}\right], \quad \rho_X(h) = \frac{\kappa}{\lambda}\left[G(\kappa, \rho_Z(h)) - \frac{1}{4}\right]; \tag{27}$$

these quantities are negative at lag h when $\rho_Z(h) < 0$.

Table 1 shows some negative correlations that can be made at lag one in our Poisson series for different λ. The table shows the theoretically most negative correlation possible as quantified in (6), the most negative possible lag one correlation achievable using renewal binary series, and the most negative correlation that can be made from Gaussian clipped binary series. The renewal negative correlation is achieved by allowing L to be a three-point lifetime with $\mathbb{P}[L = 1] = \mathbb{P}[L = 3] = \epsilon$ and $\mathbb{P}[L = 2] = 1 - 2\epsilon$. Here, $u_1 = \epsilon$ and we let $\epsilon \downarrow 0$ to achieve the reported negative correlation. The most negative correlation for the clipped Gaussian series is obtained by taking $\rho_Z(1) = -1$.

The results show that some substantial negative correlations can be produced with these methods, as λ increases, the degree of negative correlation obtained increases. This said, no negative correlations smaller than -0.5 were produced from these methods. One could experiment with other choices of p_B besides $1/2$.

Figure 2 shows a simulated realization of a stationary series with Poisson marginal distributions with mean $\lambda = 5$. The $\{B_{t,i}\}$ are generated from a clipped Gaussian process: a zero-mean unit variance AR(1) series with a lag one autocorrelation of 0.9.

https://doi.org/10.1017/S0269964819000433 Published online by Cambridge University Press

Table 1. Poisson negative correlations for various λ. The second column shows the theoretical most negative correlation possible in (6); the third column is a renewal-based negative correlation obtained with the above three-point lifetime with $\epsilon \downarrow 0$; the rightmost column shows the negative correlation achieved when $\rho_Z(1) = -1$ in a clipped Gaussian binary process.

λ	$\rho-$ in (6)	Renewal binary series	Clipped Gaussian binary series
0.01	−0.0100	−0.0098	−0.0098
0.1	−0.0996	−0.0828	−0.0829
0.5	−0.5004	−0.2376	−0.2381
1	−0.7364	−0.3065	−0.3071
2	−0.8871	−0.3605	−0.3612
3	−0.9271	−0.3853	−0.3861
5	−0.9584	−0.4105	−0.4114
10	−0.9800	−0.4362	−0.4371
50	−0.9961	−0.4709	−0.4718
100	−0.9981	−0.4791	−0.4801

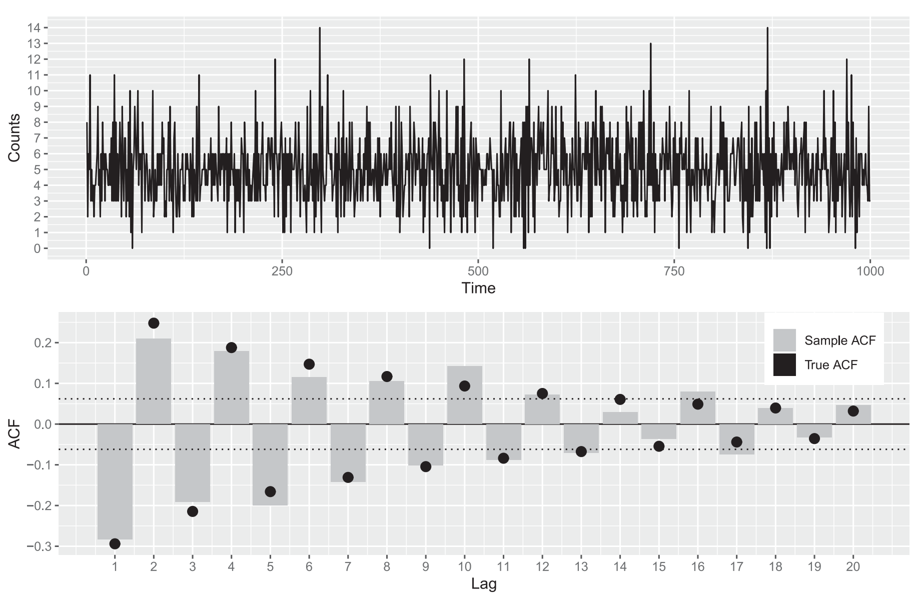

Figure 2. A realization of a stationary count time series with Poisson marginal distributions with mean 5. Sample and true autocorrelations are shown with dotted pointwise 95% confidence bands for white noise.

5.3. Negative Binomial Marginals

Count series with negative binomial marginal distributions are often used to model overdispersed count series [10,29,33]. The negative binomial distribution with parameters $r \in \{1, 2, \ldots\}$ and $p \in (0, 1)$ (NB(r, p)) has the probability mass function

$$\mathbb{P}(X_t = k) = \binom{r + k - 1}{r - 1} p^r (1 - p)^k, \quad k \in \{0, 1, \ldots\}. \tag{28}$$

https://doi.org/10.1017/S0269964819000433 Published online by Cambridge University Press

The dispersion of this distribution is $D_X = 1/p > 1$ and its probability generating function is

$$\psi_{X_t}(u) = \mathbb{E}[u^{X_t}] = \left(\frac{p}{1-(1-p)u}\right)^r, \quad |u| < (1-p)^{-1}. \tag{29}$$

To construct a negative binomial count series $\{X_t\}$ via (1), apply part 1 of Theorem 4.1 to infer that the probability generating function of M_t must satisfy

$$\psi_M(u) = \left(\frac{\frac{pp_B}{1-p+pp_B}}{1-\left(1-\frac{pp_B}{1-p+pp_B}\right)u}\right)^r. \tag{30}$$

From this, it follows that the marginal distribution of $\{M_t\}$ is again negative binomial: $M_t \sim \mathrm{NB}(r, \tilde{p})$, where $\tilde{p} = pp_B/(1-p+pp_B) \in [0,1]$.

Part (c) of Theorem 4.1 shows that the lag h autocovariance of $\{X_t\}$ has form $\gamma_X(h) = \kappa\gamma_B(h)$ when $h \neq 0$ and $\gamma_X(0) = r(1-p)/p^2$, where $\kappa = \sum_{k=0}^{\infty} \mathbb{P}(M_t > k)^2$. The tail probability $\mathbb{P}(M_t > k)$ for $M_t \sim \mathrm{NB}(r, \tilde{p})$ can be calculated via a recursion in r. Specifically, M_t has the representation $M_t = A_1 + \cdots + A_r$, where the A_i are independent with tail distribution $\mathbb{P}(A_i > k) = (1-\tilde{p})^k$ for $k \in \{0, 1, 2, \ldots\}$ (a $\mathrm{NB}(r = 1, \tilde{p})$ distribution). Let $\tilde{q} = 1 - \tilde{p}$ and condition on A_1 to get the recursion

$$\mathbb{P}(A_1 + \cdots + A_r > k) = \sum_{\ell=0}^{\infty} \mathbb{P}(A_2 + \cdots + A_r > k - \ell)\tilde{p}\tilde{q}^{\ell}. \tag{31}$$

With $\eta_r(k) = \mathbb{P}(A_1 + \cdots + A_r > k)$, we arrive at the difference equation

$$\eta_r(k) = \tilde{q}^{k+1} + \tilde{p}\sum_{\ell=0}^{k} \eta_{r-1}(k-\ell)\tilde{q}^{\ell}, \tag{32}$$

which can be evaluated recursively in r to obtain $\eta_r(k) = \mathbb{P}(M_t > k)$, starting with $\eta_1(k) = \tilde{q}^{k+1}$.

Figure 3 shows a realization of stationary count time series with a negative binomial marginal distribution with $r = 10$ and $p = 0.5$. The $\{B_{t,i}\}$s were generated from a renewal process with lifetime L supported on $\{1, 2, 3\}$, with $\mathbb{P}(L = 1) = \mathbb{P}(L = 3) = 0.1$ and $\mathbb{P}(L = 2) = 0.8$. Here, $\mathbb{E}[L] = 2$ and $p_B = 1/2$. From the plotted sample autocorrelations, it is evident that negative correlations are obtained.

Overall, the degree of negative correlations produced by superpositioning as in (1) for negative binomial marginal distributions is disappointing. In particular, no scenarios (renewal or clipping) or values of p_B produced a negative lag one correlation less than -0.12 when $r = 1$. Because of this, we will explore alternative ways of combining the binary series to achieve a negative binomial marginal.

Another way of combining the zero-one processes mimics a strategy in Cui and Lund [4]. Using that a negative binomial draw is the first time that r heads are obtained in independent coin flips (minus r to render a variable supported on $\{0, 1, \ldots\}$), set $M_t^{(r)} = \inf\{k \geq 1 : \sum_{i=1}^{k} B_{t,i} = r\}$ and $X_t = M_t^{(r)} - r$. Then X_t has a $\mathrm{NB}(r, p_B)$ distribution by construction and has the superpositioned form

$$X_t = \sum_{i=1}^{M_t^{(r)}} (1 - B_{t,i}). \tag{33}$$

The difference between (1) and (33), besides the $B_{t,i}$ versus the $1 - B_{t,i}$, is that $M_t^{(r)}$ is not independent of the $B_{t,i}$s, but is rather a stopping time constructed from them.

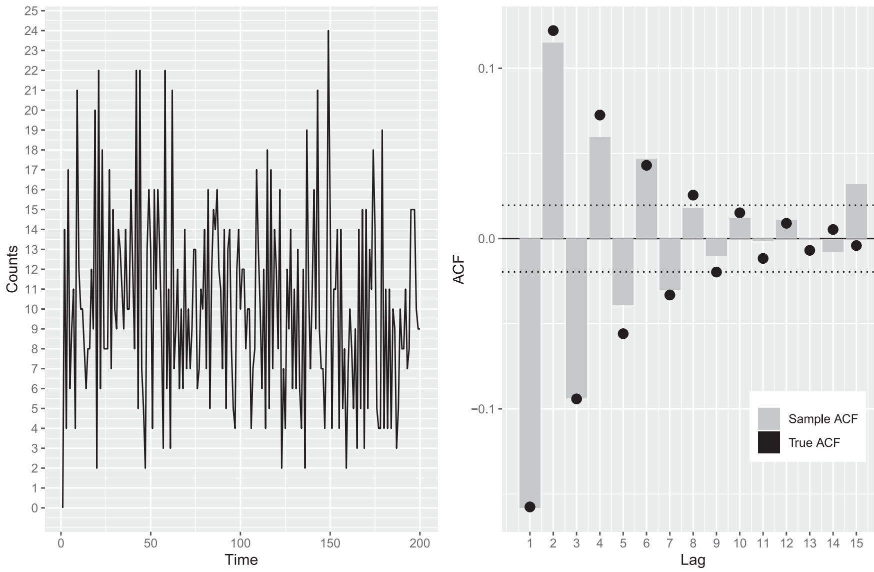

FIGURE 3. A realization of a long memory stationary count time series with NB(10,0.5) marginal distributions. The first 200 observations are plotted. Sample autocorrelations are shown with dotted pointwise 95% confidence bands for white noise.

Explicit evaluation of the autocovariance function of this model is difficult, but can be done recursively in r. Let $\xi_{i,j}(h) := \mathbb{E}[M_t^{(i)} M_{t+h}^{(j)}]$, where $M_t^{(i)} = A_{t,1} + \cdots + A_{t,i}$ and $M_{t+h}^{(j)} = A_{t+h,1} + \cdots + A_{t+h,j}$ are ordinary geometric random variables supported on $\{1, 2, \ldots\}$ (note that this geometric support set does not contain zero). Since $\gamma_X(h) = \text{cov}(M_t^{(r)}, M_{t+h}^{(r)})$, $\gamma_X(h) = \xi_{r,r}(h) - r^2/p_B^2$. The Appendix establishes the recursion

$$\begin{aligned}\xi_{i,j}(h) = &\frac{\xi_{i-1,j-1}(h)p_B p_{1,1}(h) + \xi_{i,j-1}(h)(1-p_B)p_{0,1}(h) + \xi_{i-1,j}(h)p_B p_{1,0}(h)}{1-(1-p_B)p_{0,0}(h)} \\ &+ \left(\frac{i+j-1}{p_B} - \frac{p_{1,1}(h)}{1-(1-p_B)p_{0,0}(h)}\right)\frac{1}{1-(1-p_B)p_{0,0}(h)} \\ &+ \frac{[2-(1-p_B)p_{0,0}(h)]}{[1-(1-p_B)p_{0,0}(h)]^2}, \end{aligned} \tag{34}$$

in $i, j \in \{1, 2, \ldots\}$. Of course, $\xi_{i,j}(h) = \xi_{j,i}(h)$. Boundary conditions take $\xi_{0,i}(h) = \xi_{i,0}(h) = 0$. One starts the recursion with

$$\begin{aligned}\xi_{1,1}(h) = &\frac{1}{p_B p_{0,1}(h)} - \frac{p_B p_{0,0}(h)}{[1-(1-p_B)p_{0,0}(h)]^2 p_{0,1}(h)} \\ &+ \frac{p_{1,0}(h)}{[1-(1-p_B)p_{0,0}(h)]^2}. \end{aligned} \tag{35}$$

For example, to get $\xi_{2,2}(h)$, one first uses (34) to get $\xi_{1,2}(h)$ from $\xi_{1,1}(h)$. Using the recursion again gives $\xi_{2,2}(h)$ from $\xi_{1,2}(h)$.

Table 2 is analogous to Table 1 and shows negative correlations achieved at lag one for negative binomial marginals with $r = 1$ for varying values of p_B. The correlation at lag h

https://doi.org/10.1017/S0269964819000433 Published online by Cambridge University Press

TABLE 2. Negative lag one correlations produced for negative binomial marginals. The renewal and clipped Gaussian specifications are described above.

p_B	$\rho-$ in Eq. (6)	Renewal binary series	Clipped Gaussian binary series
0.1	−0.6370	−0.0556	−0.0556
0.2	−0.6176	−0.1250	−0.1250
0.3	−0.5881	−0.2143	−0.2143
0.4	−0.5463	−0.3333	−0.3333
0.5	−0.4991	−0.4990	−0.4990
0.6	−0.4004	−0.4000	−0.4000
0.7	−0.3004	−0.3000	−0.3000
0.8	−0.2001	−0.2000	−0.2000
0.9	−0.1000	−0.1000	−0.1000

for the case is

$$\rho_X(h) = \frac{\xi_{1,1}(h) - \frac{1}{p_B^2}}{\frac{1-p_B}{p_B^2}}. \tag{36}$$

For each fixed $p = p_B$ in the table, the renewal methods chooses a lifetime L with mean $\mu_L = 1/p_B$ as follows. When $0 < p_B \leq 1/2$, we choose L with $P(L = 1) = 0$ so that $u_1 = 0$. When $1/2 < p_B < 1$, a two-point L supported on $\{1, 2\}$ is chosen. Since $p_B = 1/\mu_L$, the probabilities $P(L = 1) = 2 - 1/p_B$ and $P(L = 2) = 1/p_B - 1$ are chosen. The clipped Gaussian method again takes $\rho_Z(1) = -1$ for the latent Gaussian process $\{Z_t\}$; the clipping value of κ is set to make the success probability of the binary process p_B. From these specifications, one can compute $p_{0,0}(1)$, $p_{1,1}(1)$, $p_{0,1}(1)$, and $p_{1,0}(1)$; the lag one correlation is then computed from (36).

Observe that the degree of negative correlation produced by the methods is very close to the minimal values possible (these are in the second column) for large p_B, becoming almost optimal for $p_B \geq 1/2$. Degrees of negative correlation for smaller p_B are worse. Also, the renewal and clipped Gaussian methods produce the same values (one can verify this computationally).

5.4. Multinomial Marginals

Our goal here is to construct a J-dimensional time series with a multinomial marginal distribution with M trials and success probability vector $(p_1, p_2, \ldots, p_J)$. While any of the series in the previous subsections can be built from either renewal or clipped Gaussian binary processes, we find it more convenient to work with a latent Gaussian processes $\{Z_t\}$, but to place it into more than two categories. Partition $\mathbb{R}$ into the J sets $E_1, E_2, \ldots, E_J$ so that $\mathbb{P}(Z_1 \in E_j) = p_j$ for $j = 1, 2, \ldots, J$. Define

$$B_{t,j} = \begin{cases} 1, & Z_t \in E_j; \\ 0, & \text{otherwise} \end{cases}, \tag{37}$$

and set $\boldsymbol{B}_t = (B_{t,1}, B_{t,2}, \ldots, B_{t,J})$. Then $\boldsymbol{B}_t$ has a multinomial distribution with one trial and success probability vector $(p_1, p_2, \ldots, p_J)$. Choosing $E_1, \ldots, E_J$ such that $p_1 = \cdots = p_J = 1/J$ yields a discrete uniform variate over the categories $\{1, \ldots, J\}$.

For a fixed number of trials M, one can superposition M independent copies of $\{\boldsymbol{B}_t\}$ — call these $\{\boldsymbol{B}_{t,i}\}$ — via

$$\boldsymbol{X_t} = \sum_{i=1}^{M} \boldsymbol{B}_{t,i}. \tag{38}$$

Then $\boldsymbol{X}_t = (X_{t,1}, X_{t,2}, \ldots, X_{t,J})'$ has a multinomial distribution with M trials and success probability vector $(p_1, p_2, \ldots, p_J)$. Observe that $\mathbb{E}[X_{t,j}] = Mp_j$. For lag h autocovariances, simple computations give $\mathbb{E}(X_{t,i}X_{t+h,j}) = M\mathbb{P}(Z_t \in E_i \cap Z_{t+h} \in E_j) + (M^2 - M)p_i p_j$. It now follows that

$$\text{cov}(X_{t,i}, X_{t+h,j}) = M\left[\mathbb{P}(Z_t \in E_i \cap Z_{t+h} \in E_j) - \mathbb{P}(Z_t \in E_i)\mathbb{P}(Z_{t+h} \in E_j)\right]. \tag{39}$$

A graphic of a sample path of $\{\mathbf{X}_t\}$ and its autocorrelations is omitted, but negative autocovariances arise when $\{Z_t\}$ has negative correlations.

6. COVARIATES

Situations arise where stationarity is not reasonable, particularly when covariates are present. Suppose that there are P non-random covariates to help explain the series at time t and call these $C_{t,1}, \ldots, C_{t,P}$. This setting is easily accommodated by allowing M_t to have marginal distribution $F_{\boldsymbol{\theta}(t)}(\cdot)$, where $\boldsymbol{\theta}(t)$ is a function of the time-varying covariates and other specifications. For example, if a Poisson marginal is desired for $\{X_t\}$ in a Poisson regression setup, one simply lets $\{M_t\}$ have a Poisson marginal with mean $\lambda(t)$, where

$$\lambda(t) = \exp\left(\beta_0 + \beta_1 C_{t,1} + \cdots + \beta_P C_{t,P}\right). \tag{40}$$

Here, the exponential link function is used to guarantee that $\lambda(t)$ is positive and $\beta_0, \beta_1, \ldots, \beta_P$ are regression coefficients. In this setting, X_t has a Poisson marginal distribution for each time, but the mean $\mathbb{E}[X_t]$ changes over time.

Seasonality can be accommodated in multiple ways. One way simply allows for seasonal covariates via (40). Seasonality in the underlying Bernoulli sequences is also possible and will yield count series with seasonally varying autocorrelations; this avenue was pursued in Fralix, Livsey, and Lund [7].

7. COMMENTS

This paper presented methods to build stationary time series with some of the classical count marginal distributions that can have very general autocovariance features, including negative correlations and long-memory. The methods build the series by combining correlated zero-one binary series in various ways. A distribution family not pursued here, but worthy of further research, is the generalized Poisson.

Another strategy that could be used to generate a variety of series that is not discussed here is a copula approach. Some words on this merit mention. Suppose $F(\cdot)$ is the desired marginal distribution of a count series. If $\{Z_t\}$ is a Gaussian process with standard normal marginals and autocorrelation function $\rho_Z(\cdot)$, then $\{X_t\}$ defined pointwise by $X_t = F^{-1}(\Phi(Z_t))$ will have marginal distribution $F(\cdot)$. This definition requires taking the inverse of the count cumulative distribution function F^{-1} in the manner of Equation (14.5) of Billingsley [2]. This is the so-called normal to anything (NORTA) approach of Yahav and Shmueli [35]. This approach generally does not yield explicit autocovariance expressions (see Jia, Kechagias, Livsey, Lund, and Pipiras [14] for a series expansion). The likelihood is also difficult to obtain because of the discrete nature of F^{-1}. If F were continuous, then a simple Jacobian transformation method would suffice to evaluate the likelihood. But the discrete nature of F^{-1} makes one have to quantify a discrete joint transformations; this said, Jia, Kechagias, Livsey, Lund, and Pipiras [14] make some progress on a particle filtering likelihood approximation.

Acknowledgments

The authors acknowledge support from NSF Grant DMS-1407480. The referee and Associate Editor are thanked for helpful comments.

Disclaimer

This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

References

1. Al-Osh, M.A. & Alzaid, A.A. (1987). First order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* 8: 261–275.
2. Billingsley, P. (2008). *Probability and measure*. Hoboken, New Jersey, United States: John Wiley & Sons.
3. Blight, P.A. (1989). Time series formed from the superposition of discrete renewal processes. *Journal of Applied Probability* 26: 189–195.
4. Cui, Y. & Lund, R.B. (2009). A new look at time series of counts. *Biometrika* 96: 781–792.
5. Cui, Y. & Lund, R.B. (2010). Inference for binomial AR(1) models. *Statistics and Probability Letters* 80: 1985–1990.
6. Davis, R.A., Holan, S., Lund, R.B., & Ravishanker, N. (2016). *Handbook of discrete-valued time series*. Boca Raton, Florida, United States: CRC Press.
7. Fralix, B., Livsey, J., & Lund, R.B. (2012). Renewal sequences with periodic dynamics. *Probability in the Engineering and Informational Sciences* 26: 1–15.
8. Gouveia, S., Möller, T.A., Weiß, C.H., & Scotto, M.G. (2018). A full ARMA model for counts with bounded support and its application to rainy-days time series. *Stochastic Environmental Research and Risk Assessment* 32: 2495–2514.
9. Heathcote, C.R. (1967). Complete exponential convergence and related topics. *Journal of Applied Probability* 4: 217–256.
10. Hilbe, J.M. (2011). *Negative Binomial Regression*. Cambridge, England: Cambridge University Press.
11. Holan, S., Lund, R.B., & Davis, G. (2010). The ARMA alphabet soup: A tour of ARMA model variants. *Statistics Surveys* 4: 232–274.
12. Jacobs, P.A. & Lewis, P.A. (1978). Discrete time series generated by mixtures I: Correlational and runs properties. *Journal of the Royal Statistical Society* 40: 94–105.
13. Jacobs, P.A. & Lewis, P.A. (1978). Discrete time series generated by mixtures II: Asymptotic properties. *Journal of the Royal Statistical Society* 40: 222–228.
14. Jia, Y., Kechagias, S., Livsey, J., Lund, R.B., & Pipiras, V. (2018). Latent Gaussian count time series modeling, *eprint arXiv:1811.00203*.
15. Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability* 33: 664–677.
16. Kachour, M. & Yao, J.F. (2009). First-order rounded integer-valued autoregressive (RINAR(1)) process. *Journal of Time Series Analysis* 30: 417–448.
17. Lin, G.D., Dou, X., Kuriki, S., & Huang, J.S. (2014). Recent developments on the construction of bivariate distributions with fixed marginals. *Journal of Statistical Distributions and Applications* 1: 14SEP.
18. Livsey, J., Lund, R.B., Kechagais, S., & Pipiras, V. (2018). Multivariate count time series with flexible autocovariances and their application to hurricane counts. *Annals of Applied Probability* 12: 408–431.
19. Lund, R.B. & Livsey, J. (2015). Renewal-based count time series. In Davis, R.A., Holan, S., Lund, R.B. & Ravishanker, N. (eds.), Handbook of Discrete-Valued Time Series. New York City: CRC Press, pp. 101–120.

20. Lund, R.B., Holan, S., & Livsey, J. (2015). Long memory discrete-valued time series. In Davis, R.A., Holan, S., Lund, R.B. & Ravishanker, N. (eds.), *Handbook of Discrete-Valued Time Series*, New York City: CRC Press, pp. 447–458.
21. McKenzie, E. (1985). Some simple models for discrete variate time series. *Journal of the American Water Resources Association* 21: 645–650.
22. McKenzie, E. (1986). Autoregressive-moving average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability* 18: 822–835.
23. McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability* 20: 645–650.
24. Ng, C.T., Joe, H., Karlis, D., & Liu, J. (2011). Composite likelihood for time series models with a latent autoregressive process. *Statistica Sinica* 21: 279–305.
25. Pedeli, X. & Karlis, D. (2013). On composite likelihood estimation of a multivariate INAR(1) model. *Journal of Time Series Analysis* 34: 206–220.
26. Rose, C. & Smith, M.D. (1996). The multivariate normal distribution. *Mathematica Journal* 6: 32–37.
27. Steutel, F.W. & Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability* 7: 893–899.
28. Van Vleck, J.H. & Middleton, D. (1966). The spectrum of clipped noise. *Proceedings of the IEEE* 54: 2–19.
29. Ver Hoef, J.M. & Boveng, P.L. (2007). Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data?. *Ecological Society of America* 88: 2766–2772.
30. Weiß, C.H. (2007). Serial dependence and regression of Poisson INARMA models. *Journal of Statistical Planning and Inference* 138: 2975–2990.
31. Weiß, C.H. (2009). A new class of autoregressive models for time series of binomial counts. *Communications in Statistics: Theory and Methods* 38: 447–460.
32. Weiß, C.H. (2009). Controlling correlated processes with binomial marginals. *Journal of Applied Statistics* 36: 399–414.
33. White, G.C. & Bennetts, R.E. (1996). Analysis of frequency count data using the negative binomial distribution. *Ecological Society of America* 77: 2549–2557.
34. Whitt, W. (1976). Bivariate distributions with given marginals. *The Annals of Statistics* 4: 1280–1289.
35. Yahav, I. & Shmueli, G. (2011). On generating multivariate Poisson data in management science applications. *Applied Stochastic Models in Business and Industry* 28: 91–102.
36. Zhu, F. (2010). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis* 32: 54–67.
37. Zhu, F. (2012). Modeling time series of counts with COM-Poisson INGARCH models. *Mathematical and Computer Modelling* 56: 191–240.
38. Zhu, R. & Joe, H. (2003). A new type of discrete self-decomposability and its application to continuous-time Markov processes for modeling count data time series. *Stochastic Models* 19: 235–254.

APPENDIX

Proof of Theorem 4.1. Parts (a) and (b) follow easily from standard techniques.

To prove part (c), apply the law of total covariance to get

$$\begin{aligned}\operatorname{cov}(X_t, X_{t+h}) &= \operatorname{cov}\left(\sum_{i=1}^{M_t} B_{t,i}, \sum_{j=1}^{M_{t+h}} B_{t+h,j}\right) \\ &= \mathbb{E}\left[\operatorname{cov}\left(\sum_{i=1}^{M_t} B_{t,i}, \sum_{j=1}^{M_{t+h}} B_{t+h,j} \middle| M_t, M_{t+h}\right)\right] \\ &\quad + \operatorname{cov}\left(\mathbb{E}\left[\sum_{i=1}^{M_t} B_{t,i} \middle| M_t, M_{t+h}\right], \mathbb{E}\left[\sum_{j=1}^{M_{t+h}} B_{t+h,j} \middle| M_t, M_{t+h}\right]\right). \qquad \textbf{(A.1)}\end{aligned}$$

https://doi.org/10.1017/S0269964819000433 Published online by Cambridge University Press

For any positive integers m_t, m_{t+h},

$$\mathrm{cov}\left(\sum_{i=1}^{m_t} B_{t,i}, \sum_{j=1}^{m_{t+h}} B_{t+h,j}\right) = \min(m_t, m_{t+h})\gamma_B(h)$$

and $\mathbb{E}[\sum_{i=1}^{m_t} B_{t,i}] = m_t p_B$. Plugging these into (A.1) gives

$$\mathrm{cov}(X_t, X_{t+h}) = \mathrm{E}\left[\min(M_t, M_{t+h})\right]\gamma_B(h) + p_B^2 \mathrm{cov}(M_t, M_{t+h}). \tag{A.2}$$

When $h \neq 0$, M_t and M_{t+h} are independent and $\mathrm{cov}(M_t, M_{t+h}) = 0$, implying that $\gamma_X(h) = \kappa\gamma_B(h)$, where $\kappa = \mathbb{E}[\min(M_t, M_{t+h})]$. When $h = 0$, extracting the first two moments from the probability generating function gives $\gamma_X(0) = \gamma_B(0)\mu_M + p_B^2\sigma_M^2$ as claimed.

For part (d), when $h \neq 0$, conditioning on M_t and M_{t+h} gives

$$\mathbb{P}(X_t = x_t, X_{t+h} = x_{t+h}) = \sum_{m_t=0}^{\infty}\sum_{m_{t+h}=0}^{\infty} r_{m_t,m_{t+h}}(x_t, x_{t+h}) f_M(m_t) f_M(m_{t+h}), \tag{A.3}$$

where $r_{m_t,m_{t+h}}(x_t, x_{t+h}) := \mathbb{P}(S_t = x_t, S_{t+h} = x_{t+h})$, $S_t = \sum_{i=1}^{m_t} B_{t,i}$, $S_{t+h} = \sum_{i=1}^{m_{t+h}} B_{t+h,i}$, and m_t and m_{t+h} are fixed. This joint probability can be calculated by further conditioning on X_t:

$$r_{m_t,m_{t+h}}(x_t, x_{t+h}) = \mathbb{P}\left(S_{t+h} = x_{t+h} | S_t = x_t\right)\mathbb{P}(S_t = x_t). \tag{A.4}$$

It is easy to see that S_t has a binomial distribution with m_t trials and success probability p_B. The conditional probability above changes form with two cases.

When $m_t < m_{t+h}$, the conditional probability in (A.4) represents the sum of three independent binomial distributions: one with x_t trials and success probability $p_{1,1}(h)$, one with $m_t - x_t$ trials and success probability $p_{0,1}(h)$, and one with $m_{t+h} - m_t$ trials and success probability p_B. Hence,

$$\mathbb{P}\left(S_{t+h} = x_{t+h} | S_t = x_t\right) = \sum_{b=0}^{\min(m_t - x_t, x_{t+h})}\ \sum_{a=0}^{\min(x_t, x_{t+h}-b)} T_1 T_2 T_3,$$

where $T_1 = B(x_t, a, p_{1,1}(h))$, $T_2 = B(m_t - x_t, b, p_{0,1}(h))$, and $T_3 = B(m_{t+h} - m_t, x_{t+h} - a - b, p_B)$. Here, $B(n, k, p) := \binom{n}{k} p^k (1-p)^{n-k}$ denotes the binomial probability mass function.

The case where $m_t \geq m_{t+h}$ is more complicated. Further conditioning on $\sum_{i=1}^{m_{t+h}} B_{t,i}$ (this is neither S_t nor S_{t+h}) gives

$$\begin{aligned}\mathbb{P}\left(S_{t+h} = x_{t+h} | S_t = x_t\right) = \sum_{k=0}^{\infty} \mathbb{P}\left(S_{t+h} = x_{t+h} | S_t = x_t, \sum_{i=1}^{m_{t+h}} B_{t,i} = k\right) \\ \times \mathbb{P}\left(\sum_{i=1}^{m_{t+h}} B_{t,i} = k \,\middle|\, S_t = x_t\right).\end{aligned} \tag{A.5}$$

Conditional on $S_t = x_t$, $\sum_{i=1}^{m_{t+h}} B_{t,i}$ has a hypergeometric distribution with m_{t+h} draws from a population of containing x_t type I items and $m_t - x_t$ type II items:

$$\mathbb{P}\left(\sum_{i=1}^{m_{t+h}} B_{t,i} = k \,\middle|\, S_t = x_t\right) = \frac{\binom{x_t}{k}\binom{m_t - x_t}{m_{t+h} - k}}{\binom{m_t}{m_{t+h}}}, \tag{A.6}$$

for $k \in \{\max(0, m_{t+h} - m_t + x_t), \ldots, \min(x_t, m_{t+h})\}$.

The distribution of S_{t+h} conditioned on $S_t = x_t$ and $\sum_{i=1}^{m_{t+h}} B_{t,i} = k$ behaves as the sum of two independent binomial distributions: one with k trials and success probability $p_{1,1}(h)$, and the other with $m_{t+h} - k$ trials and success probability $p_{0,1}(h)$. This gives

$$\mathbb{P}\left(S_{t+h} = x_{t+h} \middle| S_t = x_t, \sum_{i=1}^{m_{t+h}} B_{t,i} = k\right) = \sum_{a=0}^{\min(k, x_{t+h})} B(m_{t+h} - k, x_{t+h} - a, p_{0,1}(h)) \times B(k, a, p_{1,1}(h)).$$

This identifies the two terms in (A.5). Plugging these back into (A.4) identifies the form of $r_{m_t,m_{t+h}}(x_t, x_{t+h})$.

To prove part (e), the renewal case is established in Lund, Holan, and Livsey [20]. In the clipped Gaussian case, if $\rho_Z(h) \to 0$ as $h \to \infty$, then the result follows from (13) and the limit comparison test since $\lim_{x\to 0} \sin^{-1}(x)/x = 1$. Should $\rho_Z(h) \nrightarrow 0$, then $\{Z_t\}$ must have long memory and there are an infinite number of lags h where $|\rho_Z(h)| > \delta$ for some $\delta > 0$. For these lags, we also must have $|\sin^{-1}(\rho_Z(h))| > \delta^*$ for some $\delta^* > 0$ by properties of the inverse sin function. It follows from (13) that $\{X_t\}$ must also have long memory. This proves part (e). ■

Proof of (34). To rig up a type of regeneration epoch, condition on the minimum of $A_{t,1}$ and $A_{t+h,1}$ to get

$$\xi_{i,j}(h) = \sum_{\ell=1}^{\infty} \mathbb{E}\left[M_t^{(i)} M_{t+h}^{(j)} \middle| \min(A_{t,1}, A_{t+h,1}) = \ell\right] \mathbb{P}(\min(A_{t,1}, A_{t+h,1}) = \ell). \tag{A.7}$$

When $A_{t,1} = A_{t+h,1} = \ell$, due to the memoryless property of the geometric distribution, $M_t^{(i)}$ is distributionally equivalent to $M_t^{(i-1)} + \ell$ and $M_{t+h}^{(j)}$ is equal in distribution to $M_{t+h}^{(j-1)} + \ell$. Similarly, when $\ell = A_{t,1} < A_{t+h,1}$, $M_t^{(i)}$ is equal in distribution to $M_t^{(i-1)} + \ell$ and $M_{t+h}^{(j)}$ is equal to $M_{t+h}^{(j)} + \ell$ in distribution. When $M_{t,1} > M_{t+h,1} = \ell$, $M_t^{(i)}$ equals $M_t^{(i)} + \ell$ in distribution and $M_{t+h}^{(j)}$ equals $M_{t+h}^{(j-1)} + \ell$ in distribution. Using these in (A.7) and simplifying gives

$$\begin{aligned}\xi_{i,j}(h) &= \xi_{i-1,j-1}(h)p_1(h) + \xi_{i,j-1}(h)p_2(h) + \xi_{i-1,j}(h)p_3(h) \\ &\quad + \frac{i+j-1-p_1(h)}{p_B}\mathbb{E}\left[\min(A_{t,1}, A_{t+h,1})\right] \\ &\quad + \operatorname{var}(\min(A_{t,1}, A_{t+h,1})) + \mathbb{E}^2\left[\min(A_{t,1}, A_{t+h,1})\right],\end{aligned} \tag{A.8}$$

where $i, j \in \{1, 2, \ldots\}$. Here, the $p_i(h)$s are

$$p_1(h) = \mathbb{P}(A_{t,1} = A_{t+h,1} = \ell | \min(A_{t,1}, A_{t+h,1}) = \ell) = \frac{p_B p_{1,1}(h)}{1 - (1-p_B)p_{0,0}(h)},$$

$$p_2(h) = \mathbb{P}(A_{t,1} > A_{t+h,1} = \ell | \min(A_{t,1}, A_{t+h,1}) = \ell) = \frac{(1-p_B)p_{0,1}(h)}{1 - (1-p_B)p_{0,0}(h)},$$

$$p_3(h) = \mathbb{P}(\ell = A_{t,1} < A_{t+h,1} | \min(A_{t,1}, A_{t+h,1}) = \ell) = \frac{p_B p_{1,0}(h)}{1 - (1-p_B)p_{0,0}(h)}.$$

To verify the expression in (35), notice that $M_{t+h}^{(1)}$ conditioned on $M_t^{(1)} = k$ has the distributional form shown in Table A1.

Table A1. Probability distribution of $M_{t+h}^{(1)}$ conditioned on $M_t^{(1)}$.

ℓ	$\mathbb{P}(M_{t+h}^{(1)} = \ell \vert M_t^{(1)} = k)$
1	$p_{0,1}(h)$
$\vdots$	$\vdots$
$k-1$	$p_{0,0}(h)^{k-2} p_{1,1}(h)$
k	$p_{0,0}(h)^{k-1} p_{1,1}(h)$
$k+1$	$p_B p_{0,0}(h)^{k-1} p_{1,0}(h)$
$k+2$	$(1-p_B) p_B p_{0,0}(h)^{k-1} p_{1,0}(h)$
$\vdots$	$\vdots$

Using these in the law of total expectation gives

$$\xi_{1,1}(h) = \frac{1}{p_B p_{0,1}(h)} - \frac{p_B p_{0,0}(h)}{[1-(1-p_B)p_{0,0}(h)]^2 p_{0,1}(h)} + \frac{p_{1,0}(h)}{[1-(1-p_B)p_{0,0}(h)]^2}. \quad \textbf{(A.9)}$$

Finally, we derive an explicit form for $\mathbb{E}[\min(A_{t,1}, A_{t+h,1})]$ and $\mathrm{var}(\min(A_{t,1}, A_{t+h,1}))$. The tail distribution of $\min(A_{t,1}, A_{t+h,1})$ satisfies

$$\begin{aligned}\mathbb{P}(\min(A_{t,1}, A_{t+h,1}) > \ell) &= \mathbb{P}\left(B_{t,1} = 0, B_{t+h,1} = 0, \cdots, B_{t,\ell} = 0, B_{t+h,\ell} = 0\right) \\ &= [(1-p_B)p_{0,0}(h)]^{\ell},\end{aligned}$$

which is an ordinary geometric distribution. Hence, $\mathbb{E}[\min(A_{t,1}, A_{t+h,1})] = [1-(1-p_B)p_{0,0}(h)]^{-1}$ and $\mathrm{var}(\min(A_{t,1}, A_{t+h,1})) = (1-p_B)p_{0,0}(h)/[1-(1-p_B)p_{0,0}(h)]^2$. Plugging these into (A.8) verifies the claim. ∎

https://doi.org/10.1017/S0269964819000433 Published online by Cambridge University Press