Binomial Autoregressive Processes with Density Dependent Thinning

Christian H. Weiß, Philip K. Pollett

Full addresses:

Christian H. Weiß **(Corresponding author)**

Department of Mathematics and Statistics

Helmut Schmidt University Hamburg

Postfach 700822

D-22008 Hamburg, Germany

Email: weissc@hsu-hh.de

Tel.: +49-40-6541-2779

Fax: +49-40-6541-2023

Philip K. Pollett

Department of Mathematics

University of Queensland

QLD 4072, Australia

Email: pkp@maths.uq.edu.au

Binomial Autoregressive Processes with Density Dependent Thinning

Christian H. Weiß* and Philip K. Pollett†

October 2, 2013

Abstract

We present an elaboration of the usual binomial AR(1) process on $\{0, 1, \ldots, N\}$ that allows the thinning probabilities to depend on the current state n only through the "density" n/N, a natural assumption in many real contexts. We derive some basic properties of the model and explore approaches to parameter estimation. Some special cases are considered that allow for over- and underdispersion, as well as positive and negative autocorrelation. We derive a law of large numbers and a central limit theorem, which provide useful large–N approximations for various quantities of interest.

KEY WORDS: binomial AR(1) model; binomial INARCH(1) model; stationarity; normal approximation; overdispersion and underdispersion; metapopulation models; parameter estimation.

*Department of Mathematics and Statistics, Helmut Schmidt University, Postfach 700822, D-22008 Hamburg, Germany. weissc@hsu-hh.de

†Department of Mathematics, University of Queensland, QLD 4072, Australia. pkp@maths.uq.edu.au

1 Introduction

In recent times there has been considerable interest in time series of count data with a fixed finite range $\{0, 1, \ldots, N\}$. Applications include the monitoring of computer pools (with N workstations) (Weiß, 2009; Weiß & Kim, 2013a), of infections (with N individuals) (Daley & Gani, 1999, Section 4.4), of metapopulations (with N patches) (Buckley & Pollett, 2009, 2010a,b; Weiß & Pollett, 2012), and of transactions in the stock market (with N listed companies) (Weiß & Kim, 2013b). A popular model is the *binomial AR(1) process*, proposed by McKenzie (1985) as an integer-valued counterpart to the standard Gaussian AR(1) process. It is defined using the *binomial thinning* operation introduced by Steutel & van Harn (1979): $\alpha \circ x := \sum_{i=1}^{x} y_i$, where the y_i are i.i.d. Bernoulli variables with success probability α, independent of the count-data random variable x (given x, $\alpha \circ x \sim \text{Bin}(x, \alpha)$). If α and β are suitable thinning probabilities, the process $(n_t,\, t = 0, 1, \ldots)$, taking values in $\{0, \ldots, N\}$, defined by the recursion

$$n_t = \alpha \circ n_{t-1} + \beta \circ (N - n_{t-1}) \qquad (t \geq 1),$$

where all thinnings are performed independently of one another, and where the thinnings at time t are independent of $(n_s,\, s < t)$, is called a *binomial AR(1) process*. (A similar model, which instead uses *hypergeometric thinning*, was developed by Al-Osh & Alzaid (1991).)

Properties of the binomial AR(1) process are now well understood. It is an ergodic Markov chain with a $\text{Bin}(N, \pi)$ stationary distribution, where $\pi = \beta/(1 - r)$ and $r = \alpha - \beta$, and its autocorrelation function is given by $\rho(k) = r^k$ $(k \geq 0)$ (analogous to the Gaussian AR(1) process). Its h-step regression properties were explored in Weiß & Pollett (2012), where it was also shown that, for large N, the process can be approximated by a Gaussian

AR(1) process. Closed-form expressions for the joint moments and cumulants were derived in Weiß & Kim (2013a), and questions concerning parameter estimation were addressed in Cui & Lund (2010); Weiß & Kim (2013a,b); Weiß & Pollett (2012).

A significant limitation of the binomial AR(1) model is that the thinning probabilities at time t do not depend on the process up to that time. For example, in the epidemic context, if n_t represents the number of infectives and $N - n_t$ the number susceptibles in a population of size N (Daley & Gani, 1999, Section 4.4), or, in the metapopulation context, if n_t represents the number of occupied habitat patches and $N - n_t$ the number unoccupied in an N-patch population (Buckley & Pollett, 2010a), it is natural to allow α and/or β to depend on n_t, reflecting the (internal) infection-recovery or colonization-extinction mechanisms. Such a state-dependence of the parameters is also plausible in an economic context, where it may arise, for example, through the interaction of supply and demand; see also our analysis of securities data in Section 4.2.

Our purpose here is to extend the binomial AR(1) model to allow this kind of state dependence, thus increasing its applicability. As we shall see, if the thinning probabilities are allowed to depend on n_t only through the "density" n_t/N (for example, the *proportion* of infectives or *proportion* of occupied patches), then many of the properties of the basic model can be extended without difficulty. Similar state-dependent thinning has been considered previously in the context of the INAR(1) model by McKenzie (1985): the FINAR(1) model (Triebsch, 2008) and the SETINAR$(2, 1)$ model (Monteiro et al., 2012).

A formal definition of a *density-dependent* binomial AR(1) process is given in Section 2. Here we derive some of its basic properties and indicate approaches to parameter estimation. In Section 3 we a derive a law of large numbers and a central limit theorem, thus providing

useful large–N approximations for various quantities of interest. Then we focus on several special cases within the family of density-dependent binomial AR(1) models: in Section 4 we study a case that is convenient for modelling over- and underdispersion (as well as positive and negative autocorrelation), and in Section 5 we give special attention to the case where α (only) is constant. Our conclusions are summarized in Section 6.

2 Density-dependent Binomial AR(1) Processes

Following the approach of Buckley & Pollett (2010a), we extend the definition of the binomial AR(1) model to allow the model parameters to be density dependent.

Definition 2.1. Let $\pi : [0;1] \to (0;1)$ and $r : [0;1] \to (0;1)$, so that the functions $\beta(y) := \pi(y)\big(1 - r(y)\big)$ and $\alpha(y) := \beta(y) + r(y)$, defined for $y \in [0;1]$, also have range $(0;1)$. Fix $N \geq 1$ and write $\pi_t := \pi(n_{t-1}/N)$, $r_t := r(n_{t-1}/N)$, $\alpha_t := \alpha(n_{t-1}/N)$ and $\beta_t := \beta(n_{t-1}/N)$, so that the thinning probabilities at time t satisfy $\beta_t := \pi_t\,(1 - r_t)$ and $\alpha_t := \beta_t + r_t$. The process $(n_t,\, t \geq 0)$, taking values in $\{0, \ldots, N\}$ and defined by the recursion

$$n_t = \alpha_t \circ n_{t-1} + \beta_t \circ (N - n_{t-1}) \qquad (t \geq 1),$$

where the thinnings at time t are performed independently of one another and of $(n_s,\, s < t)$, is called a *density-dependent binomial AR(1) process*.

A process thus defined is a time-homogeneous finite-state Markov chain with transition probabilities

$$\begin{aligned} P(k|l) \;&:=\; \Pr(n_t = k \mid n_{t-1} = l) \;=\; \textstyle\sum_{m=\max\{0,k+l-N\}}^{\min\{k,l\}} \\ &\tbinom{l}{m}\,\tbinom{N-l}{k-m}\,\big(\alpha(l/N)\big)^m\big(1-\alpha(l/N)\big)^{l-m}\,\big(\beta(l/N)\big)^{k-m}\big(1-\beta(l/N)\big)^{N-l+m-k}\;(>0), \end{aligned} \tag{1}$$

from which the following regression properties are readily obtained:

$$\begin{aligned} \mathbb{E}(n_t \mid n_{t-1}) &= r_t\, n_{t-1} + N\, \beta_t, \\ \mathrm{Var}\,(n_t \mid n_{t-1}) &= r_t(1-r_t)\,(1-2\pi_t)\, n_{t-1} + N\, \beta_t(1-\beta_t). \end{aligned} \tag{2}$$

For details, see Appendix A.1.

2.1 Unique Stationary Solution

Notice that the transition matrix $\mathbf{P} = (P(k|l)_{k,l=0,\ldots,N})$ is primitive, implying that the process is ergodic with uniquely determined stationary distribution. Whilst in some special cases a closed-form expression can be obtain for this distribution, it is not available in general, nor even are there closed-form expressions for its mean and variance. However, given a particular parametric form, these quantities can always be determined numerically. Let $\boldsymbol{p} = \big(P(0), \ldots, P(N)\big)^\top$ denote the stationary probability vector, where $P(k) = \Pr(n_t = k)$. Then $\boldsymbol{p}$ is obtained from the equation $\mathbf{P}\boldsymbol{p} = \boldsymbol{p}$, and it can be used, for example, to evaluate stationary moments of n_t. For second-order moments with time-lag h, as required to evaluate the autocorrelation function, one uses $\Pr(n_t = k, n_{t-h} = l)$, being the entries of $\mathbf{P}^h\,\mathrm{diag}(\boldsymbol{p})$. Furthermore, large–$N$ approximations are available; in Section 3.3, we obtain closed-form approximations for stationary mean, variance and autocorrelation function, the quality of which improve as N becomes large.

2.2 Approaches to Parameter Estimation

In any particular application the functions α and β will depend on a vector of model parameters $\boldsymbol{\theta}$, which lie in some open subset Θ of $\mathbb{R}^s$. Using formula (1), the conditional

log-likelihood of observing $n_1, \ldots, n_T$, given n_0, is

$$\ell(\boldsymbol{\theta}) \;=\; \sum_{t=1}^{T} \ln P_{\boldsymbol{\theta}}(n_t|n_{t-1}).$$

Maximum likelihood (ML) estimates for $\boldsymbol{\theta}$ are obtained numerically by maximizing $\ell(\boldsymbol{\theta})$. We can prove, in the stationary case, existence, consistency and asymptotic normality of the conditional ML estimator $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ of $\boldsymbol{\theta}$ using Theorems 2.1 and 2.2 of Billingsley (1961). It suffices to check Condition 5.1 of Billingsley (1961): for all $\boldsymbol{\theta} \in \Theta$,

(i) $P_{\boldsymbol{\theta}}(k|l)$ is three times continuously differentiable with respect to $\boldsymbol{\theta}$, and

(ii) the $(N+1)^2 \times s$ matrix with entries $\frac{\partial}{\partial \theta_u} P_{\boldsymbol{\theta}}(k|l)$ has rank s.

This approach is exemplified below in Sections 4 and 5, where we consider some three-parameter models ($s = 3$).

Remark 2.2.1 (Moment Estimation). For the above-mentioned numerical procedure, it is natural to use *method of moments* (MM) estimates for the initial values. For some types of stationary density-dependent binomial AR(1) model (like the one described in Section 4.1), it is possible to find closed-form expressions for moments, autocorrelation function, et cetera. In such cases, moment relations are used to obtain parameter estimates based on empirically determined moments. If no closed-form expressions are available (as is the case for the model considered in Section 5), we recommend using an *approximate* method of moments based on the large–N moment approximations derived in Section 3.3.

3 Large–N Approximations

Let $(n_t,\, t \geq 0)$ be a density-dependent binomial AR(1) process and define $(X_t^N,\, t \geq 0)$ by $X_t^N = n_t/N$. By varying N we obtain a family of Markov chains that is density-dependent in the sense of (Buckley & Pollett, 2010a, Section 3). In particular, it follows from (2) that $\mathbb{E}(X_t^N \,|X_{t-1}^N) = f(X_{t-1}^N)$, where

$$f(x) = \alpha(x)x + \beta(x)(1-x) = r(x)x + \beta(x), \tag{3}$$

and $N \operatorname{Var}(X_t^N \,|X_{t-1}^N) = v(X_{t-1}^N)$, where

$$v(x) = \alpha(x)(1-\alpha(x))x + \beta(x)(1-\beta(x))(1-x). \tag{4}$$

3.1 A Law of Large Numbers

Since $0 \leq X_t^N \leq 1$, we may apply Theorem 1 of Buckley & Pollett (2010a) to obtain the following law of large numbers.

Theorem 3.1.1. *Suppose that α and β are continuous and bounded. If $X_0 \xrightarrow{\mathrm{P}} x_0$ (a constant) as $N \to \infty$, then $X_t^N \xrightarrow{\mathrm{P}} x_t$ for all $t \geq 1$, where (x_t) is determined by $x_{t+1} = f(x_t)$ $(t \geq 0)$.*

Thus for large N our binomial AR(1) process can be approximated by a deterministic process (x_t), the form of which is determined by the thinning probabilities. Since f is not necessarily a linear function, (x_t) can potentially exhibit the full range of long-term behaviour, including limit cycles and even chaos. However, since $f(x) = x$ if and only if $\big(1 - r(x)\big)\, x = \beta(x) = \pi(x)\, \big(1 - r(x)\big)$, we can see that x^* is a fixed point of f if and only if

(a) x^* is a fixed point of π, or

(b) $r(x^*) = 1$ and $\beta(x^*) = 0$.

Example 3.1.2. In the case $r(x) = r$ (< 1) and $\pi(x) = a + bx$ (to be considered later in Section 4), there is a unique fixed point $x^* = a/(1-b)$, which is stable because $f'(x) = \kappa := r + (1-r)b < 1$. Furthermore, we obtain the explicit expression $x_t = x_0\kappa^t + x^*(1-\kappa^t)$.

Example 3.1.3. In contrast, if $r(x) = x$ and $\pi(x) = a+bx$, then $f(x) = a+(b-a)x+(1-b)x^2$. Now there are two fixed points, namely $x^* = 1$ (unstable, because $f'(1) = 2-a-b > 1$) and $x^* = a/(1-b)$ (stable, because $f'(a/(1-b)) = a+b < 1$).

3.2 A Central Limit Law

Having established conditions for convergence in probability of (X_t^N) to a limiting deterministic process (x_t), we next consider the process of scaled fluctuations about (x_t), namely $(Z_t^N\,,\, t \geq 0)$ defined by $Z_t^N = \sqrt{N}(X_t^N - x_t)$. Assuming now that $Z_0^N \xrightarrow{\mathrm{D}} z_0$ as $N \to \infty$, we can identify conditions under which (Z_t^N) converges weakly to a Gaussian Markov chain (Z_t). By writing n_t as the sum of two independent sums of independent Bernoulli random variables with parameters $\alpha(n_{t-1}/N)$ and $\beta(n_{t-1}/N)$, we can applied the methods of (Buckley & Pollett, 2010a, Section 3) to prove the following central limit law.

Theorem 3.2.1. *Suppose that both α and β are twice continuously differentiable with bounded second derivative and that $X_t^N \xrightarrow{\mathrm{P}} x_t$, where (x_t) satisfies $x_{t+1} = f(x_t)$ $(t \geq 0)$. If $Z_0^N \xrightarrow{\mathrm{D}} z_0$ (a constant), then (Z_t^N) converges weakly to a Gaussian Markov chain (Z_t) with representation*

$$Z_{t+1} = f'(x_t)Z_t + E_t \qquad (Z_0 = z_0), \tag{5}$$

where (E_t) are independent with $E_t \sim \mathrm{N}\big(0, v(x_t)\big)$.

The mean and covariance function of (Z_t) are easy to evaluate by iterating (5).

Corollary 3.2.2. *For $v > u$, let*

$$\Pi_{u,\,v} = \prod_{w=u}^{v-1} f'(x_w), \tag{6}$$

where empty products are to be interpreted as being equal to 1. *Then, for $s \geq t \geq 1$,*

$$\mu_t := \mathbb{E}\, Z_t = z_0\, \Pi_{0,\,t}, \tag{7}$$

$$V_t := \mathrm{Var}\,(Z_t) = \sum_{s=0}^{t-1} v(x_s) \Pi_{s+1,\,t}^2, \tag{8}$$

$$c_{t,\,s} := \mathrm{Cov}\,(Z_t, Z_s) = V_t\, \Pi_{t,\,s}.$$

It is clear that, for any $t \geq 1$, $Z_t^N \xrightarrow{\mathrm{D}} \mathrm{N}(\mu_t, V_t)$, and so these formulae can be used to approximate the mean and covariance function of (n_t). Indeed, the joint distribution of $n_{t_1}, \ldots, n_{t_n}$, where $t_1, \ldots, t_n$ is any finite set of times, can be approximated by an n-dimensional Gaussian distribution with $\mathbb{E} n_{t_i} \approx N x_{t_i} + \sqrt{N}\mu_{t_i}$ and $\mathrm{Cov}\,(n_{t_i}, n_{t_j}) \approx N c_{t_i,\,t_j}$.

3.3 Large–N Approximations with Stable Fixed Points

If f has a stable fixed point x^* and $X_0^N \xrightarrow{\mathrm{P}} x^*$, then, assuming α and β are continuous bounded, we will have $X_t^N \xrightarrow{\mathrm{P}} x^*$ for all t. So Theorem 3.2.1 leads to the following central limit law, which divines the relationship between the density-dependent binomial AR(1) process and the standard autoregressive AR(1) process.

Theorem 3.3.1. *Suppose that both α and β are twice continuously differentiable with bounded second derivative and that $X_t^N \xrightarrow{\mathrm{P}} x^*$, the stable fixed point of f. Set $Z_t^N = \sqrt{N}(X_t^N - x^*)$. If $Z_0^N \xrightarrow{\mathrm{D}} z_0$, then (Z_t^N) converges weakly to an autoregressive AR(1) process with representation $Z_{t+1} = f'(x^*)Z_t + E_t$, where (E_t) are independent and $E_t \sim \mathrm{N}(0, v(x^*))$.*

Also, the results of Corollary 3.2.2 further simplify. Equation (6) gives $\Pi_{u,\,v} = \kappa^{v-u}$, where $\kappa = f'(x^*)$. So in the case of Theorem 3.3.1, $Z_t^N \xrightarrow{\text{D}} \text{N}(\mu_t, V_t)$, where

$$\mu_t = z_0 \kappa^{\,t}, \qquad c_{t,\,s} = V_t\, \kappa^{|s-t|}, \qquad V_t = \begin{cases} v(x^*)t & \text{if } |\kappa| = 1, \\ v(x^*)(1-\kappa^{2t})/(1-\kappa^2) & \text{otherwise.} \end{cases}$$

Furthermore, if $Z_0^N \xrightarrow{\text{D}} z_0$, and $|\kappa| < 1$, then there will be a sequence of times (t_N) such that $Z_{t_N}^N \xrightarrow{\text{D}} \text{N}(0, V^*)$, where $V^* = v(x^*)/(1-\kappa^2)$.

Remark 3.3.2. More generally, suppose that f admits a stable limit cycle, $x_0^*, x_1^*, \ldots, x_{q-1}^*$, so that, for large t the deterministic process (x_t) can be approximated by one that follows this cycle ad infinitum. Then, if $X_0^N \xrightarrow{\text{P}} x_0^*$, we will have $Z_{nq+j}^N = \sqrt{N}(X_{nq+j}^N - x_j^*)$ $(n \geq 0,\ j = 0, \ldots, q-1)$, which measures the degree of deviation of (X_t^N) from this cycle. And, assuming $Z_0^N \xrightarrow{\text{D}} z_0$, the limit process (Z_t) will have the following representation: $(\boldsymbol{Y}_n,\, n \geq 0)$, where $\boldsymbol{Y}_n = (Z_{nq}, Z_{nq+1}, \ldots, Z_{(n+1)q-1})^\top$ with $Z_0 = z_0$, is a q-variate AR(1) process of the form $\boldsymbol{Y}_{n+1} = A\boldsymbol{Y}_n + \boldsymbol{E}_n$, where $(\boldsymbol{E}_n)$ are independent and $\boldsymbol{E}_n \sim \text{N}(\boldsymbol{0}, \Sigma)$. The distribution of $\boldsymbol{Y}_0$, and both the coefficient matrix A and the covariance matrix Σ, are determined using (7)–(8) with $x_{nq+j} = x_j^*$ $(n \geq 0,\ j = 0, \ldots, q-1)$.

As announced in Section 2, Theorem 3.3.1 permits closed-form approximations for stationary mean μ, variance σ^2 and autocorrelation function $\rho(k)$, which are obtained from the corresponding properties of the approximating AR(1) process. Defining $f(x)$ and $v(x)$ as in formulae (3) and (4) with x^* the stable fixed point of f, we use

$$\mu \;\approx\; N\,x^*, \qquad \sigma^2 \;\approx\; N\,\frac{v(x^*)}{1-\kappa^2}, \qquad \rho(k) \;\approx\; \kappa^k, \qquad \text{where } \kappa = f'(x^*). \tag{9}$$

Next, we apply these results to Examples 3.1.2 and 3.1.3, described above.

Example 3.3.3. In the case $r(x) = r$ (< 1) and $\pi(x) = a + bx$ (Example 3.1.2), both of α and β are linear functions and $f(x) = (1-r)a + \kappa x$, where recall that $\kappa := r + (1-r)b$. There is a unique stable fixed point $x^* = a/(1-b)$, implying the approximate expression $N\,x^* = Na/(1-b)$ for the stationary mean μ; in fact, we shall see below in formula (13) that this approximation is actually *equal to* the true mean. The fluctuations about x^*, $Z_t^N = \sqrt{N}(X_t^N - x^*)$, can be approximated by the AR(1) process (Z_t) with the representation $Z_{t+1} = \kappa Z_t + E_t$, where (E_t) are independent and $E_t \sim \mathrm{N}(0, v(x^*))$. Again, we shall see below that the approximation κ^k of the autocorrelation function is exact; see formula (16).

Note that $v(x^*)$ can be evaluated explicitly from (4), because $\alpha(x^*) = r + (1-r)x^*$ and $\beta(x^*) = (1-r)x^*$. We find that $v(x^*) = (1-r^2)x^*(1-x^*)$. Thus, the stationary variance of the approximating AR(1) process is

$$V^* = \frac{1-r^2}{1-\kappa^2}\, x^*(1-x^*).$$

Example 3.3.4. In the case $r(x) = x$ and $\pi(x) = a + bx$ (Example 3.1.3), α, β and f are now quadratic functions, but $x^* = a/(1-b)$ is the unique stable fixed point of f. In this case fluctuations of X_t^N about x^* can be approximated by the AR(1) process (Z_t) with the representation $Z_{t+1} = (a+b)Z_t + E_t$, where (E_t) are independent and $E_t \sim \mathrm{N}(0, v(x^*))$, noting again that $v(x^*)$ can be evaluated explicitly from (4); now $\alpha(x^*) = x^*(2 - x^*)$ and $\beta(x^*) = x^*(1-x^*)$.

4 Special Case: A Model for Binomial Overdispersion or Underdispersion

For the Poisson distribution, the mean and variance are the same, a property commonly referred to as *equidispersion*. An analogous fixed relation between the mean and the variance also holds for the binomial distribution, and this might be expressed by using the so-called *binomial index of dispersion*. For a random variable with finite support $\{0, \ldots, N\}$ and with a certain mean μ and variance σ^2, the binomial index of dispersion (as a function of N, μ and σ^2) is defined as

$$I_{\mathrm{d}} \;:=\; \frac{N\sigma^2}{\mu(N-\mu)} \quad \in\; (0;\infty). \tag{10}$$

In the case of a $\mathrm{Bin}(N,\pi)$-distributed random variable, we have $I_{\mathrm{d}} = 1$ for any $\pi \in (0;1)$. For this reason, finite-range count data random variables satisfying $I_{\mathrm{d}} > 1$ are said to show *overdispersion* with respect to the binomial distribution (sometimes extra-binomial variation), while *underdispersion* refers to the case $I_{\mathrm{d}} < 1$.

In this section, we study a special density-dependent binomial AR(1) model, where $r(y) = r$, a constant, and π is a linear function. Notice then that α and β are also linear functions. We shall see that the stationary case has very attractive features for real applications, because we will be able to deal with both over- and underdispersion (in the above sense), as well as positive and negative autocorrelation. Furthermore, this new model encapsulates a binomial counterpart of the so-called INARCH(1) model; see Section 4.3 for details.

4.1 Definition and Stochastic Properties

Suppose that $\pi(y) = a + by$, where $a, a+b \in (0;1)$. This entails $\pi(y) \in (0;1)$, but allows b to also be negative (see Figure 1 (a)). If $b > 0$, then π increases with y (corresponding to the upper triangle in Figure 1 (a)), while π decreases with increasing y if $b < 0$ (corresponding to the lower triangle in Figure 1 (a)). $b = 0$ corresponds to density *independence*.

Notice that the thinning probabilities β_t and α_t from Definition 2.1 are also linear functions that depend on the density n_t/N through

$$\beta_t = (1-r)\,(a + b\,n_t/N), \qquad \alpha_t = (1-r)\,(a + b\,n_t/N) + r.$$

So, depending on the sign of b, these probabilities increase or decrease with increasing density. In a metapopulation context as mentioned in Section 1, for instance, we would conclude that the occupation of patches becomes more attractive ($b > 0$) or less attractive ($b < 0$) if the occupation rate is already large. In an economic context, $b < 0$ could reflect the interaction of supply and demand.

Theorem 4.1.1. *Let $(n_t,\, t \geq 0)$ be a density-dependent binomial AR(1) process with $r(y) = r \in (0;1)$, a constant, and $\pi(y) = a + by$, where $a, a+b \in (0;1)$. Then, for $k = 1, 2, \ldots, t$,*

$$\mathbb{E}(n_t \mid n_{t-k}) \;=\; \big(r + (1-r)b\big)^k\, n_{t-k} \;+\; \frac{Na}{1-b}\Big(1 - \big(r + (1-r)b\big)^k\Big), \tag{11}$$

in particular

$$\mathbb{E}(n_t \mid n_{t-1}) \;= \big(r + (1-r)b\big)\, n_{t-1} + N(1-r)a,$$

and

$$\begin{aligned}\mathrm{Var}\,(n_t \mid n_{t-1}) \;=\;& N(1-r)a\,\big(1 - (1-r)a\big) - \tfrac{b(1-r)}{N}\,\big(2r + b(1-r)\big)\, n_{t-1}^2 \\ &+ (1-r)\,\big((1-2a)(b+r) + 2abr\big)\, n_{t-1}.\end{aligned} \tag{12}$$

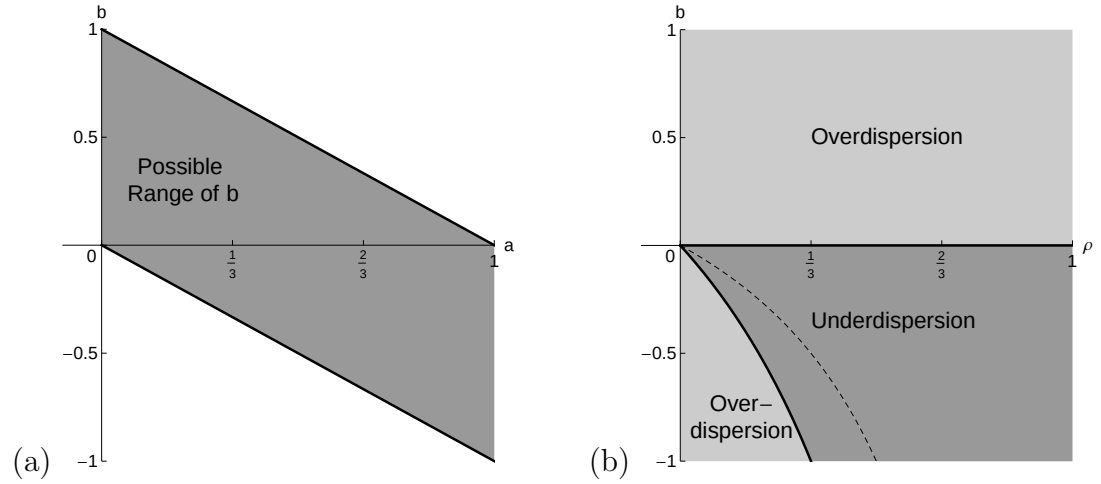

Figure 1: Attainable range of b depending on a in (a), and dispersion determined by b and r in (b).

The proof is given in Appendix A.2.

Henceforth we consider the case of a *stationary* density-dependent binomial AR(1) process. Recall that we showed in Section 2.1 that the stationary distribution exists and is uniquely determined, so if n_0 follows this distribution, then the process will be stationary. We shall see that for the particular model being considered here, we obtain closed-form expressions for the mean and variance of n_t; a large–N approximation as derived in Section 3.3 is not necessary for expressing these quantities.

Theorem 4.1.2. *Let* $(n_t,\, t \geq 0)$ *be a* stationary *density-dependent binomial AR(1) process with* $r(y) = r$ *and* $\pi(y) = a + by$. *Then the mean* $\mu = \mathbb{E}(n_t)$ *and variance* $\sigma^2 = \mathrm{Var}(n_t)$ *are given by*

$$\mu \;=\; \frac{Na}{1-b}; \tag{13}$$

$$\begin{aligned}\sigma^2 &= \frac{1-r^2}{1-\frac{r^2}{N}-(1-\frac{1}{N})\big(r+(1-r)b\big)^2}\ \mu(1-\frac{\mu}{N}) \\ &= \frac{1+r}{1+r-2(1-\frac{1}{N})rb-(1-\frac{1}{N})(1-r)b^2}\ \mu(1-\frac{\mu}{N}).\end{aligned} \tag{14}$$

The proof is given in Appendix A.2. Note that $\mu(1-\mu/N)$ is the usual binomial variance. So, the first factor in (14) is just the actual value of the binomial index of dispersion I_{d} from (10) and, hence, a measure of the deviation of the true distribution from a binomial distribution. Remembering the large–N approximation from Example 3.3.3, where $\kappa = r + (1-r)b$, we note that

$$I_{\mathrm{d}}(N,b,r) = \frac{1-r^2}{1-\frac{r^2}{N}-(1-\frac{1}{N})\kappa^2} \to \frac{1-r^2}{1-\kappa^2} \qquad (\text{as } N\to\infty),$$

in accordance with the expression for V^*.

For $N > 1$, we have

$$\begin{array}{ccccc} & > & & > & \\ I_{\mathrm{d}} = I_{\mathrm{d}}(N,b,r) & = & 1 \quad \text{iff} \quad b\big(2r+(1-r)b\big) & = & 0, \\ & < & & < & \end{array}$$

that is, we have over-/equi-/underdispersion with respect to the binomial distribution according to the following rule:

$$\begin{aligned} &N = 1 \text{ or } b = \tfrac{-2r}{1-r} \text{ or } b = 0 && \Rightarrow \quad \text{equidispersion}, \\ &N > 1 \text{ and } b < \tfrac{-2r}{1-r} \text{ or } b > 0 && \Rightarrow \quad \text{overdispersion}, \\ &N > 1 \text{ and } \tfrac{-2r}{1-r} < b < 0 && \Rightarrow \quad \text{underdispersion}. \end{aligned} \tag{15}$$

If $r > \frac{1}{3}$, then $\frac{-2r}{1-r} < -1$, that is, $b < \frac{-2r}{1-r}$ is not possible in this case. So for $r > \frac{1}{3}$ and $b < 0$, we always observe underdispersion. This is illustrated in Figure 1 (b).

Remark 4.1.3. If we take the partial derivative $\frac{\partial}{\partial b}\ I_{\mathrm{d}}(N,b,r)$ and equate it to 0 (see Appendix A.2), we see that for $N > 1$ the index of dispersion is minimal in b when $b = \frac{-r}{1-r}$ (the dashed line in Figure 1 (b)) and increases above or below this value. So the strongest underdispersion for given N and r is

$$I_{\mathrm{d}}(N, \tfrac{-r}{1-r}, r) \ = \ \frac{1-r^2}{1-\frac{r^2}{N}} \ \geq \ 1-r^2.$$

On the other hand,

$$\lim_{b\to 1}\ I_{\mathrm{d}}(N,b,r) \ = \ \frac{1-r^2}{1-\frac{r^2}{N}-(1-\frac{1}{N})} \ = \ N,$$

$$\lim_{b\to -1}\ I_{\mathrm{d}}(N,b,r) \ = \ \frac{1-r^2}{1-\frac{r^2}{N}-(1-\frac{1}{N})(1-2r)^2} \ = \ N\,\frac{1+r}{1-3r+4Nr} \ \leq \ N,$$

that is, the maximum possible overdispersion is below N; see also the discussion in Hagmark (2009). Note, however, that values of b close to ± 1 severely restrict the choice of a according to Figure 1 (a).

Finally, let us investigate the serial dependence structure. As shown in Appendix A.2, the autocorrelation function is given by

$$\rho(k) \ = \ \big(r+(1-r)b\big)^k, \tag{16}$$

analogous to the Gaussian AR(1) model. Since $r \geq 0$, we obtain negative autocorrelations (for odd k) iff $b < \frac{-r}{1-r}$, the latter being again the dashed line in Figure 1 (b). So dispersion behaviour plus the sign of $\rho(1)$ determine to which of the four regions in Figure 1 (b) the pair (r,b) belongs.

4.2 Parameter Estimation and a Real-Data Example

As already discussed in Section 2.2, to prove existence, consistency and asymptotic normality of the conditional ML estimator, we may check the conditions (i) and (ii) listed above in Section 2.2. In the present case, α and β are polynomials in the three model parameters. Hence, so is $P_{\boldsymbol{\theta}}(k|l)$, and therefore differentiable up to any order. To verify (ii), we consider the Jacobian of

$$\begin{pmatrix} P_{\boldsymbol{\theta}}(0|0) \\ P_{\boldsymbol{\theta}}(0|1) \\ P_{\boldsymbol{\theta}}(1|1) \end{pmatrix} = \begin{pmatrix} \big(1-\beta(0)\big)^N \\ \big(1-\alpha(\frac{1}{N})\big)\,\big(1-\beta(\frac{1}{N})\big)^{N-1} \\ (N-1)\,\big(1-\alpha(\frac{1}{N})\big)\,\beta(\frac{1}{N})\,\big(1-\beta(\frac{1}{N})\big)^{N-2} \;+\; \alpha(\frac{1}{N})\,\big(1-\beta(\frac{1}{N})\big)^{N-1} \end{pmatrix}, \tag{17}$$

where $\boldsymbol{\theta} = (a, b, r)^\top$, $\alpha(y) = r + (1-r)(a+by)$ and $\beta(y) = (1-r)(a+by)$. Its determinant is obtained (after tedious computations) as

$$(N-1)\,r(1-r)^2\,\big(1-a(1-r)\big)^{N-1}\,\big(1-(1-r)(a+b/N)\big)^{2(N-2)}.$$

Provided $N > 1$, this determinant is not equal to 0, implying that the Jacobian of (17) has full rank of $s = 3$, which, in turn, implies that (ii) is satisfied.

Concerning the MM approach outlined in Remark 2.2.1, we note that the three model parameters a, b, r are determined through the relations (13) for the mean, (14) for the variance, and (16) for the first-order autocorrelation.

Example 4.2.1 (Securities Counts). We consider the count-data time series presented in Section 4 of Weiß & Kim (2013b), being the number of different securities companies (among $N = 22$ such companies) traded in the Korea stock market per 5-minute period on 8th February 2011, during trading from 09:00 to 14:50 (see Figure 2). Each of the 70 counts can

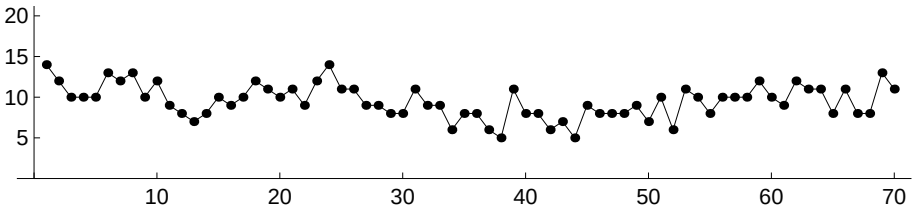

Figure 2: Plot of the securities counts discussed in Example 4.2.1.

take a value between 0 and 22. The sample mean and variance are $\bar{n} \approx 9.529$ and $s^2 \approx 4.253$, respectively. Analyzing the empirical (partial) autocorrelations, a first-order autoregressive dependence structure is apparent for the securities counts.

As shown by Weiß & Kim (2013b), these data are well described by the standard binomial AR(1) model. But, since the empirical index of dispersion $\hat{I}_{\mathrm{d}} = s^2/\big(\bar{n}(1-\bar{n}/N)\big) \approx 0.787$ indicates a slight degree of binomial underdispersion, one may ask if the data might be even better described by the *density-dependent* binomial AR(1) model of Section 4.1. We first computed moment estimates for the parameters a, b, r of this model. Substituting $\bar{n} \approx 9.529$, $s^2 \approx 4.253$ and $\hat{\rho}(1) \approx 0.402$ in equations (13), (14) and (16), respectively, we obtain the estimates

$$\hat{a}_{\mathrm{MM}} \approx 0.630, \qquad \hat{b}_{\mathrm{MM}} \approx -0.454, \qquad \hat{r}_{\mathrm{MM}} \approx 0.588.$$

These values were used to initialize the numerical procedure for computing the conditional ML estimates (we used R's `nlm`); the relevant results are summarized in Table 1. The ML estimate for b is negative, so if many of the securities companies are traded, the probabilities for trading activity become reduced. The ML-fitted density-dependent binomial AR(1) model has a dispersion index of about 0.728 and, hence, describes the empirical level of about 0.787 rather well. It also has a slightly lower value for Akaike's information criterion (AIC) than the usual binomial AR(1) model. But it performs worse in terms of the Bayesian

Table 1: Conditional ML estimates (together with the estimated standard errors in parentheses) for the securities counts discussed in Example 4.2.1.

	$\hat{\pi}_{\mathrm{ML}}$	$\hat{r}_{\mathrm{ML}}$	AIC	BIC
Bin. AR(1)	0.428	0.510	283.8	288.3
	(0.022)	(0.090)		

	$\hat{a}_{\mathrm{ML}}$	$\hat{b}_{\mathrm{ML}}$	$\hat{r}_{\mathrm{ML}}$	AIC	BIC
D.-D. Bin. AR(1)	0.693	–0.619	0.630	283.6	290.3
	(0.198)	(0.458)	(0.083)		

information criterion (BIC), and the estimate for b (the parameter that controls the density dependence) is not significantly different from 0. Both of these observations might be due to the small sample size of 70.[1] So, although the fitted density-dependent binomial AR(1) model is able to describe the empirically observed underdispersion, the more sparsely parameterized binomial AR(1) model seems adequate for the given count data time series.

4.3 Boundary Case $r \to 0$: The Binomial INARCH(1) Model

The family of INGARCH(p, q) models constitutes an *in*teger-valued counterpart to the usual *g*eneralized *a*uto*r*egressive *c*onditional *h*eteroscedastic models (Ferland et al., 2006). The model order (p, q)=(1, 0) leads to the INARCH(1) model, which gives a homogeneous Markov chain. Most commonly, this model is defined using a conditional Poisson distribution, but also versions with different types of count data distribution (always with a strictly infinite range) have been considered in the literature. The popular Poisson INARCH(1) model assumes that the process (z_t) has the property that the current observation z_t, given the past $z_{t-1}, z_{t-2}, \ldots$, stems from a conditional Poisson distribution with a mean being linear

[1]Note: Reparametrizing $b = -a \cdot d$ such that $a, d, r \in (0; 1)$, the estimate for d (≈ 0.893) is significantly different from 0 (s.e. ≈ 0.409).

in z_{t-1}. The unconditional distribution of z_t then shows overdispersion with respect to the Poisson distribution. See Weiß (2010) for a survey of the INARCH(1) model.

Let us consider the density-dependent binomial AR(1) models with $r(y) = r \in (0;1)$, a constant, and $\pi(y) = a + by$, where $a, a+b \in (0;1)$. In the boundary case $r \to 0$, we have $\alpha(y) = \beta(y) = \pi(y)$, and so the recursion of Definition 2.1 becomes

$$n_t \stackrel{\mathrm{D}}{=} \mathrm{Bin}\big(N,\ \pi(n_{t-1}/N)\big) \ = \ \mathrm{Bin}\big(N,\ a + b\,n_{t-1}/N\big) \qquad (t \geq 1), \tag{18}$$

in that the conditional distribution of n_t given the past observations is a binomial distribution with (conditional) success probability $a + b\,n_{t-1}/N$. Motivated by the analogy to the aforementioned (Poisson) INARCH(1) model, we shall refer to the model (18) as the binomial INARCH(1) model.

The stationary mean of the binomial INARCH(1) model (18) is given by formula (13), while formula (14) for the variance (taking $r \to 0$) simplifies to

$$\sigma^2 \ = \ \frac{\mu(1-\frac{\mu}{N})}{1-(1-\frac{1}{N})b^2}, \qquad \text{that is,} \quad I_{\mathrm{d}} = I_{\mathrm{d}}(N,b,0) \ = \ \frac{1}{1-(1-\frac{1}{N})b^2} \ \in \ [1;N). \tag{19}$$

Consequently, only overdispersion is possible in this case, in accordance with rule (15) and analogous to the case of the Poisson INARCH(1) model. Also, formula (16) simplifies to $\rho(k) = b^k$. Note that $\rho(k)$ might also take negative values, which is in contrast to the case of the Poisson INARCH(1) model. Another difference to the Poisson INARCH(1) model is the fact that the conditional variance is not a linear but a quadratic function in n_{t-1}:

$$\begin{aligned} \mathbb{E}(n_t \mid n_{t-1}) \ &= \ b\,n_{t-1} \ + \ N\,a, \\ \mathrm{Var}\,(n_t \mid n_{t-1}) \ &= \ N\,a\,(1-a) \ - \ \tfrac{b^2}{N}\,n_{t-1}^2 \ + \ (1-2a)\,b\,n_{t-1}. \end{aligned} \tag{20}$$

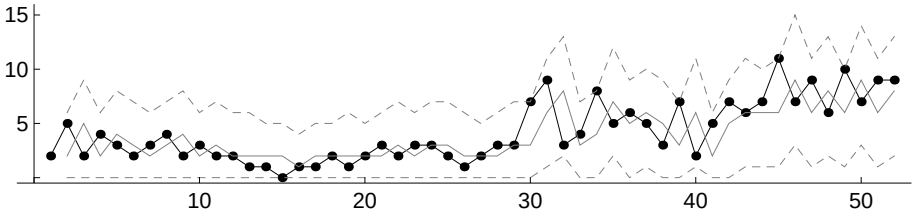

Figure 3: Plot of the infection counts discussed in Example 4.3.1, together with 1-step-ahead forecasts (median and 95 % interval).

Example 4.3.1 (Infection Counts). The online data base "SurvStat", maintained by the Robert Koch-Institut (RKI, 2013), provides a large source of epidemic data. Among others, it offers weekly data about new cases of diverse infections in Germany that have been reported to the RKI via local and state health departments. In the following, we use these data to obtain condensed information about the regional spread of an infection with time. But, instead of referring to the 16 federal states in Germany (which are very heterogeneous in terms of population), we use the classification according to the $N = 38$ districts as defined by the SurvStat data base.

For illustrative purposes, let us now concentrate on one specific infection, namely infections by the hantavirus in the year 2011 with its $T = 52$ weeks. So each count n_t represents the number of districts with a new case of hantavirus infection in week t. The observed counts vary between 0 and 11; a plot of the data is shown in Figure 3. The empirical mean and variance are given by $\bar{n} \approx 4.173$ and $s^2 \approx 7.793$, respectively. So the resulting value $\hat{I}_{\mathrm{d}} \approx 2.098$ of the index of dispersion points to a considerable degree of overdispersion. An inspection of the empirical (partial) autocorrelation function indicates use of an autoregressive model for the data; in particular, we have $\hat{\rho}(1) \approx 0.634$. Using this information, we are able to compute the moment estimates for the usual binomial AR(1) model, for the density-

Table 2: Conditional ML estimates (together with the estimated standard errors in parentheses) for the infection counts discussed in Example 4.3.1.

	$\hat{\pi}_{\mathrm{ML}}$	$\hat{r}_{\mathrm{ML}}$	AIC	BIC
Bin. AR(1)	0.115	0.535	222.8	226.7
	(0.013)	(0.071)		

	$\hat{a}_{\mathrm{ML}}$	$\hat{b}_{\mathrm{ML}}$	$\hat{r}_{\mathrm{ML}}$	AIC	BIC
D.-D. Bin. AR(1)	0.030	0.748	0.000	213.4	219.2
	(0.016)	(0.143)	(0.367)		

	$\hat{a}_{\mathrm{ML}}$	$\hat{b}_{\mathrm{ML}}$	AIC	BIC
Bin. INARCH(1)	0.030	0.748	211.4	215.3
	(0.011)	(0.108)		

dependent binomial AR(1) model of Section 4.1, as well as for the binomial INARCH(1) model.

As the next step, we computed the corresponding ML estimates; see Table 2. Although the density-dependent binomial AR(1) model clearly outperforms the usual (equidispersed) binomial AR(1) model in terms of both AIC and BIC, we realize that the estimate for r is not significantly different from 0. Since the boundary case $r \to 0$ just leads to the binomial INARCH(1) model, it is not surprising that this model shows the best model fit (in view of standard errors and the information criteria) among all considered candidate models. In particular, the value of I_{d} within the ML-fitted model is about 2.194 and, hence, very close to the empirically observed value.

As an application of the ML-fitted binomial INARCH(1) model, we evaluate the conditional distribution for the purpose of forecasting. Besides the original data, Figure 3 also shows the respective 1-step-ahead conditional median (solid line) as a coherent point forecast for $t \geq 2$, together with the limits of a 95 % prediction interval (dashed lines). None of the data points is plotted outside these prediction limits (while this happens twice

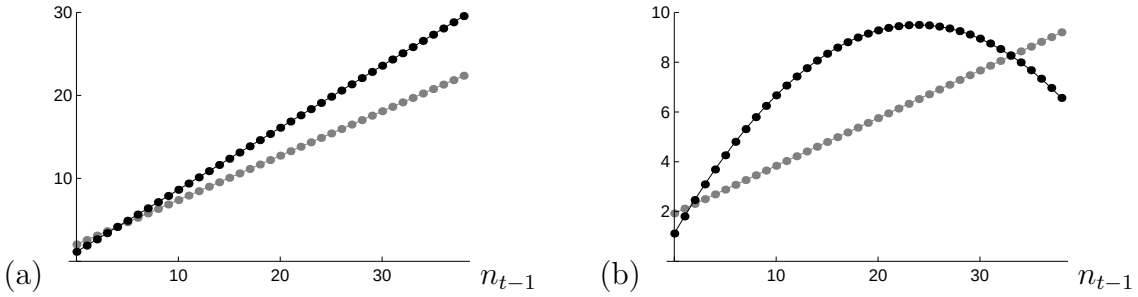

Figure 4: (a) Conditional means and (b) conditional variances of the ML-fitted binomial AR(1) model (gray) and binomial INARCH(1) model (black), respectively.

if using the ML-fitted binomial AR(1) model), again confirming the adequacy of the binomial INARCH(1) model. Also if looking at the standardized residuals $\big(n_t - \mathbb{E}(n_t \mid n_{t-1})\big) \big/ \sqrt{\mathrm{Var}(n_t \mid n_{t-1})}$, see (20), the binomial INARCH(1) model outperforms the binomial AR(1) model in terms of the residuals sum of squares (55.4 vs. 80.6). As the essential difference between the respective conditional distributions of the two models, we recognize from Figure 4 that the binomial INARCH(1) model has larger conditional means and variances for most values of n_{t-1}. An increased conditional variance can also be seen in the right part of Figure 3, where the counts reach a higher level. Although the binomial INARCH(1) model describes this behaviour reasonably well (in terms of the above criteria), an even better fit might be obtained, for example, with a self-exciting threshold version of the binomial AR(1) model, in analogy to Monteiro et al. (2012). The development of such a model appears to be an interesting issue for future research.

5 Special Case: A Model with Density-Dependent Colonisation

We concluded our earlier article (Weiß & Pollett, 2012) by referring briefly to the case $\alpha(y) = \alpha \in (0;1)$, constant, and β a continuous, increasing and concave function with range in $[0;\alpha]$. Such a model assumption is motivated by certain metapopulations with a density-dependent colonisation probability and a fixed local extinction probability, and epidemic models with a density-dependent infection probability and a fixed per-capita recovery probability.

5.1 Definition and Stochastic Properties

Let $\alpha(y) = \alpha \in (0;1)$ be constant, and let β be a linear function with range in $[0;\alpha]$, say, $\beta(y) = \alpha\,(a + by)$ with $a, a+b \in (0;1]$. Then (see Definition 2.1),

$$n_t \;=\; \alpha \circ n_{t-1} + \beta_t \circ (N - n_{t-1}),$$

which has an intuitive interpretation, for instance, in the epidemic context: if $b > 0$, the probability that a susceptible becomes infected increases if the number of infectives in the population is already large (the infection is spreading), while the recovery from the infection is independent of other infectives.

If $\alpha(y) = \alpha$ and $\beta(y) = \alpha\,(a + by)$, then the following relations hold:

$$r(y) \;=\; \alpha - \beta(y) = \alpha\,(1 - a - by), \qquad \pi(y) \;=\; \frac{\beta(y)}{1 - r(y)} \;=\; \frac{\beta(y)}{1 - \alpha + \beta(y)}. \tag{21}$$

Hence, the regression properties of formula (2) simplify to

$$\begin{aligned} \mathbb{E}(n_t \mid n_{t-1}) &= (\alpha - \beta_t)\, n_{t-1} + N\, \beta_t, \\ \mathrm{Var}\,(n_t \mid n_{t-1}) &= (\alpha - \beta_t)\,(1 - \alpha - \beta_t)\, n_{t-1} + N\, \beta_t (1 - \beta_t). \end{aligned} \tag{22}$$

If β is a linear function, then $\mathbb{E}(n_t \mid n_{t-1})$ will be a quadratic polynomial in n_{t-1}, and $\mathrm{Var}\,(n_t \mid n_{t-1})$ will be a cubic polynomial in n_{t-1}. So we cannot proceed as in Theorem 4.1.2 to obtain closed-form expressions for stationary mean or variance.

From now on, we focus on the stationary case again (see Section 2.1). While (22) does not allow us this time to obtain closed-form expressions for stationary mean or variance, we can at least derive a variance-mean relation: since $\mathbb{E}(n_t \mid n_{t-1}) = \alpha(1 - a - b\, n_{t-1}/N)\, n_{t-1} + N\alpha\,(a + b\, n_{t-1}/N)$, we have

$$\mu = N\alpha\, a + \alpha(1 - a + b)\, \mu - \alpha b/N\, (\sigma^2 + \mu^2). \tag{23}$$

Given concrete values for the model parameters a, b and α, the exact mean, variance and autocorrelation function can be computed numerically according to the procedure laid out in Section 2.1; see also the discussion below.

General closed-form approximations are now derived using formula (9). The functions $f(x)$ and $v(x)$ according to formulae (3) and (4) are obtained as

$$\begin{aligned} f(x) &= \alpha\, x \;+\; \alpha\,(a + bx)\,(1 - x) \;=\; \alpha\left(a + (1 - a + b)\, x - b\, x^2\right), \\ f'(x) &= \alpha\,(1 - a + b - 2b\, x), \\ v(x) &= \alpha(1 - \alpha)\, x \;+\; \alpha\,(a + bx)\left(1 - \alpha\,(a + bx)\right)(1 - x). \end{aligned} \tag{24}$$

Lemma 5.1.1. *If $b \neq 0$, then the unique fixed point of f in $[0;1]$ is*

$$x^* \;=\; \frac{1-\alpha\,(1-a+b)\;-\;\sqrt{\big(1-\alpha\,(1-a+b)\big)^2+4\alpha^2 ab}}{-2\alpha b} \quad \in (0;1).$$

If $b = 0$ (no density dependence), then $x^* = \alpha a/\big(1-\alpha(1-a)\big) \in (0;1)$.

The proof is given in Appendix A.3.

Using Lemma 5.1.1 and formula (24), we obtain

$$\kappa \;=\; f'(x^*) \;=\; \alpha\,(1-a+b) - 2b\alpha\, x^* \;=\; 1-\sqrt{\big(1-\alpha\,(1-a+b)\big)^2+4\alpha^2 ab}, \qquad (25)$$

which also holds for $b = 0$, since $1-\alpha(1-a) \in (0;1]$. So we can compute approximations for stationary mean μ, variance σ^2 and autocorrelation function $\rho(k)$ using formula (9).

We performed numerical experiments to test the quality of this approximation; results are summarized in Table 3. Model parametrizations were chosen so that different levels of the mean μ/N (about 0.15 or 0.40), of the index of dispersion I_{d} (about 1.1 or 1.5), and of the $\rho(1)$ (about 0.4 or 0.8) were covered. For each parametrization we computed

- numerically exact values for a range of N values (as described in Section 2.1) and
- large–N approximations (rows "$N = \infty$" in Table 3).

From the tabulated values, it is apparent that the approximations are quite accurate, at least for $N \geq 25$. If the stationary distribution is only slightly overdispersed, then even when $N = 5$ the approximation is good.

5.2 Parameter Estimation

We follow the programme of Section 4.2. For the model described in Section 5.1, the functions α and β, and hence $P_{\boldsymbol{\theta}}(k|l)$, are also polynomials in the model parameters, and therefore

Table 3: Density-dependent binomial AR(1) model of Section 5: values for mean μ, dispersion index I_{d} and autocorrelation function $\rho(k)$; determined either exactly for finite values of N or approximately by using the large–N approximation ("$N = \infty$").

a	b	α	N	μ/N	I_{d}	$\rho(1)$	$\rho(2)$
0.366	0.366	0.565	5	0.382	1.081	0.396	0.158
			10	0.391	1.091	0.398	0.160
			25	0.396	1.096	0.399	0.160
			50	0.398	1.098	0.400	0.160
			100	0.399	1.099	0.400	0.160
			∞	0.400	1.100	0.400	0.160
0.085	0.044	0.866	5	0.389	1.079	0.799	0.638
			10	0.394	1.090	0.800	0.639
			25	0.397	1.096	0.800	0.640
			50	0.398	1.098	0.800	0.640
			100	0.398	1.099	0.800	0.640
			∞	0.399	1.100	0.800	0.640
0.059	0.251	0.807	5	0.332	1.411	0.806	0.653
			10	0.364	1.475	0.807	0.653
			25	0.385	1.497	0.804	0.647
			50	0.393	1.499	0.802	0.644
			100	0.396	1.499	0.801	0.642
			∞	0.400	1.499	0.800	0.640
0.226	0.381	0.384	5	0.144	1.072	0.378	0.143
			10	0.147	1.085	0.389	0.151
			25	0.149	1.094	0.395	0.156
			50	0.149	1.097	0.397	0.158
			100	0.150	1.098	0.398	0.159
			∞	0.150	1.100	0.400	0.160
0.036	0.033	0.811	5	0.146	1.076	0.796	0.634
			10	0.148	1.088	0.798	0.638
			25	0.149	1.095	0.800	0.640
			50	0.149	1.098	0.800	0.640
			100	0.149	1.099	0.800	0.641
			∞	0.149	1.100	0.801	0.641
0.037	0.197	0.726	5	0.129	1.310	0.772	0.597
			10	0.138	1.391	0.785	0.617
			25	0.145	1.451	0.793	0.630
			50	0.147	1.474	0.796	0.634
			100	0.149	1.485	0.798	0.637
			∞	0.150	1.497	0.799	0.639

differentiable up to any order. To check (ii), we consider the Jacobian of (17) with $\boldsymbol{\theta} = (a, b, \alpha)^\top$, $\alpha(y) = \alpha$ and $\beta(y) = \alpha(a + by)$. Its determinant is obtained (after tedious computations) as

$$(N-1)\,\alpha^3\,(1-\alpha a)^{N-1}\,\big(1-(a+b/N)\big)\,\big(1-\alpha(a+b/N)\big)^{2(N-2)}.$$

Again, provided $N > 1$, this determinant is not equal to 0, implying that Jacobian of (17) has full rank, which, in turn, implies that (ii) holds.

Concerning the MM approach discussed in Remark 2.2.1, we are now in a situation where no closed-form expressions for the moments are available. However, an approximate MM approach by way of the large–N approximation of moments derived in Section 3.3 is possible: using formula (9) together with formulae (24) and (25), as well as Lemma 5.1.1, closed-form approximations for mean, variance and first-order autocorrelation follow, which can be obtained for the model parameters a, b and α.

6 Conclusion

We have considered a generalization of the binomial AR(1) process n_t with range $\{0, 1, \ldots, N\}$, where the model parameters are not constant in time, but rather depend on the current density n_t/N. We were able to derive a law of large numbers and a central limit theorem, which can be applied to obtaining large–N approximations for various quantities of interest. The utility of our approach was demonstrated by considering two particular subfamilies of the density-dependent binomial AR(1) model. The models discussed in Section 4 permit a wide range of over- and underdispersion scenarios, a valuable feature in practice, as illustrated by the real-data examples of securities counts and infection counts. The model presented

in Section 5 is motivated by metapopulations with density-dependent colonisation and a fixed extinction probability, and epidemic models with density-dependent infection and a fixed per-capita recovery probability. It was useful in helping to illustrate the quality of our large–N approximations.

An important issue for future research will be the development of statistical procedures for model identification and diagnostics for density-dependent binomial autoregressive models. Different forms of state-dependence will also be investigated, including a self-exciting threshold version of the binomial AR(1) model, in analogy to Monteiro et al. (2012) and motivated by Example 4.3.1. Future research will focus on the binomial INARCH(1) model introduced in Section 4.3. Extensions to higher-order autoregressions and time-varying population ceiling N_t will also be considered. Finally, an extension of (density-dependent) binomial autoregressive models to negative integer values will be explored, for example, by using a signed binomial thinning operation, as in Kim & Park (2008).

Acknowledgements

The work of Phil Pollett is supported by the Australian Research Council (Discovery grant DP120102398, and the Centre of Excellence for Mathematics and Statistics of Complex Systems). The authors thank the editor, the co-editor, and the two referees for carefully reading the article and for their comments, which greatly improved the article. The authors are also grateful to Prof. Hee-Young Kim, Korea University, for contributing the securities counts data of Example 4.2.1.

References

Al-Osh, M.A., Alzaid, A.A. (1991) Binomial autoregressive moving average models. *Communications in Statistics – Stochastic Models* **7**(2), 261–282.

Billingsley, P. (1961) *Statistical Inference for Markov Processes*. University of Chicago Press.

Buckley, F.M., Pollett, P.K. (2009) Analytical methods for a stochastic mainland-island metapopulation model. In *Proceedings of the 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation*, R. Anderssen, R. Braddock, and L. Newham, Eds. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, Canberra, Australia, 1767–1773.

Buckley, F.M., Pollett, P.K. (2010a) Limit theorems for discrete-time metapopulation models. *Probability Surveys* **7**, 53–83.

Buckley, F.M., Pollett, P.K. (2010b) Analytical methods for a stochastic mainland-island metapopulation model. *Ecological Modelling* **221**(21), 2526–2530.

Cui, Y., Lund, R. (2010) Inference in binomial AR(1) models. *Statistics and Probability Letters* **80**(23-24), 1985–1990.

Daley, D., Gani, J. (1999) *Epidemic Modelling: an Introduction*. Cambridge Studies in Mathematical Biology, Vol. 15. Cambridge University

Ferland, R., Latour, A., Oraichi, D. (2006) Integer-valued GARCH processes. *Journal of Time Series Analysis* **27**(6), 923–942.

Hagmark, P.-E. (2009) A new concept for count distributions. *Statistics and Probability Letters* **79**(8), 1120–1124.

Kim H.-Y., Park, Y. (2008) A non-stationary integer-valued autoregressive model. *Statistical Papers* **49**(3), 485–502.

McKenzie, E. (1985) Some simple models for discrete variate time series. *Water Resources Bulletin* **21**(4), 645–650.

Monteiro, M., Scotto, M.G., Pereira, I. (2012) Integer-valued self-exciting threshold autoregressive processes. *Communications in Statistics – Theory and Methods* **41**(15), 2717–2737.

Robert Koch-Institut (2013) SurvStat@RKI `http://www3.rki.de/SurvStat`, data status: June 12, 2013.

Steutel, F.W., van Harn, K. (1979) Discrete analogues of self-decomposability and stability. *Annals of Probability* **7**(5), 893–899.

Triebsch, L.K. (2008) *New Integer-valued Autoregressive and Regression Models with State-dependent Parameters*. Doctoral dissertation, TU Kaiserslautern, Verlag Dr. Hut, Munich.

Weiß, C.H. (2009) Monitoring correlated processes with binomial marginals. *Journal of Applied Statistics* **36**(4), 399–414.

Weiß, C.H. (2010) The INARCH(1) model for overdispersed time series of counts. *Communications in Statistics – Simulation and Computation* **39**(6), 1269–1291.

Weiß, C.H., Kim, H.-Y. (2013a) Binomial AR(1) processes: moments, cumulants, and estimation. *Statistics* **47**(3), 494–510.

Weiß, C.H., Kim, H.-Y. (2013b) Parameter estimation for binomial AR(1) models with applications in finance and industry. *Statistical Papers* **54**(3), 563–590.

Weiß, C.H., Pollett, P.K. (2012) Chain binomial models and binomial autoregressive processes. *Biometrics* **68**(3), 815–824.

A Proofs

A.1 Proofs of Section 2

Since

$$\alpha_t(1-\alpha_t) - \beta_t(1-\beta_t) \;=\; (\alpha_t-\beta_t)\,(1-\alpha_t-\beta_t) \;=\; r_t\,(1-r_t-2\beta_t),$$

we obtain

$$\begin{aligned}
\mathbb{E}(n_t \mid n_{t-1}) &= \mathbb{E}(\alpha_t \circ n_{t-1} \mid n_{t-1}) \;+\; \mathbb{E}\big(\beta_t \circ (N-n_{t-1}) \mid n_{t-1}\big) \\
&= (\alpha_t-\beta_t)\,n_{t-1} + N\,\beta_t \;=\; r_t\,n_{t-1} + N\,\beta_t, \\
\\
\mathrm{Var}\,(n_t \mid n_{t-1}) &= \mathrm{Var}\,(\alpha_t \circ n_{t-1} \mid n_{t-1}) \;+\; \mathrm{Var}\,\big(\beta_t \circ (N-n_{t-1}) \mid n_{t-1}\big) \\
&= \big(\alpha_t(1-\alpha_t) - \beta_t(1-\beta_t)\big)\,n_{t-1} + N\,\beta_t(1-\beta_t) \\
&= r_t(1-r_t)\,(1-2\pi_t)\,n_{t-1} + N\,\beta_t(1-\beta_t),
\end{aligned}$$

so formula (2) follows.

A.2 Proofs of Section 4

From formula (2), we obtain

$$\begin{aligned}
\mathbb{E}(n_t \mid n_{t-1}) &= r\, n_{t-1} + N(1-r)\,\pi_t \;=\; \big(r + (1-r)b\big)\, n_{t-1} + N(1-r)a, \\
\mathrm{Var}\,(n_t \mid n_{t-1}) &= r(1-r)\,(1-2\pi_t)\, n_{t-1} + N(1-r)\,\pi_t\,\big(1-(1-r)\,\pi_t\big) \\
&= \tfrac{r(1-r)}{N}\,(N(1-2a) - 2b\, n_{t-1})\, n_{t-1} \\
&\quad + \tfrac{1-r}{N}\,(Na + b\, n_{t-1})\,\big(N - N(1-r)a - (1-r)b\, n_{t-1}\big) \\
&= r(1-r)(1-2a)\, n_{t-1} - 2b\tfrac{r(1-r)}{N}\, n_{t-1}^2 \;+\; N(1-r)a\,\big(1-(1-r)a\big) \\
&\quad + b(1-r)\,\big(1-2(1-r)a\big)\, n_{t-1} - \tfrac{(1-r)^2\, b^2}{N}\, n_{t-1}^2 \\
&= N(1-r)a\,\big(1-(1-r)a\big) - \tfrac{b(1-r)}{N}\,\big(2r + b(1-r)\big)\, n_{t-1}^2 \\
&\quad + (1-r)\,\big((1-2a)(b+r) + 2abr\big)\, n_{t-1}.
\end{aligned}$$

For k-step regressions with $k = 1, \ldots, t$, we argue by induction:

$$\begin{aligned}
\mathbb{E}(n_t \mid n_{t-k}) &= \mathbb{E}\big(\mathbb{E}(n_t \mid n_{t-1}, \ldots, n_{t-k}) \;\big|\; n_{t-k}\big) \\
&= \big(r + (1-r)b\big)\, \mathbb{E}(n_{t-1} \mid n_{t-k}) + N(1-r)a \\
&= \ldots \;=\; \big(r + (1-r)b\big)^k\, n_{t-k} \;+\; N(1-r)a\, \textstyle\sum_{i=0}^{k-1} \;\big(r + (1-r)b\big)^i \\
&= \big(r + (1-r)b\big)^k\, n_{t-k} \;+\; N(1-r)a\, \tfrac{1-\big(r+(1-r)b\big)^k}{(1-r)(1-b)},
\end{aligned}$$

from which we obtain (11). Hence, the assertion of Theorem 4.1.1 follows.

Next, we consider Theorem 4.1.2. From (12), the mean $\mu = \mathbb{E}(n_t)$ satisfies

$$\mu \;=\; \mathbb{E}\big(\mathbb{E}(n_t \mid n_{t-1})\big) \;=\; \big(r + (1-r)b\big)\, \mu + N(1-r)a.$$

Solving for μ, we obtain formula (13).

For the variance $\sigma^2 = \mathrm{Var}\,(n_t)$, the relation

$$\begin{aligned}
\sigma^2 &= \mathbb{E}\big(\mathrm{Var}\,(n_t \mid n_{t-1})\big) + \mathrm{Var}\,\big(\mathbb{E}(n_t \mid n_{t-1})\big) \\
&= N(1-r)a\,\big(1-(1-r)a\big) \;-\; \tfrac{b(1-r)}{N}\,\big(2r+b(1-r)\big)\,(\sigma^2+\mu^2) \\
&\quad + (1-r)\,\big((1-2a)(b+r)+2abr\big)\,\mu \;+\; \big(r+(1-r)b\big)^2\,\sigma^2 \\
&= (1-r)(1-b)\,\big(1-(1-r)a\big)\,\mu \;+\; (1-r)\,\big((1-2a)(b+r)+2abr\big)\,\mu \\
&\quad - ab\tfrac{1-r}{1-b}\,\big(2r+b(1-r)\big)\,\mu \;+\; \big(r+(1-r)b\big)^2\,\sigma^2 - \tfrac{1}{N}\,\Big(\big(r+b(1-r)\big)^2 - r^2\Big)\,\sigma^2 \\
&= \tfrac{1-a-b}{1-b}(1-r^2)\,\mu \;+\; \Big(\tfrac{r^2}{N} + (1-\tfrac{1}{N})\big(r+(1-r)b\big)^2\Big)\,\sigma^2
\end{aligned}$$

holds. Using $1-\mu/N = \frac{1-a-b}{1-b}$, formula (14) follows, and the proof of Theorem 4.1.2 is complete.

For Remark 4.1.3, we consider the following partial derivative:

$$\tfrac{\partial}{\partial b}\; I_{\mathrm{d}}(N,b,r) \;=\; 0 \quad \text{iff} \quad -(1+r)\,\big(-2(1-\tfrac{1}{N})r - 2(1-\tfrac{1}{N})(1-r)b\big) \;=\; 0$$

$$\text{iff} \quad r+(1-r)b \;=\; 0 \qquad \text{iff} \quad b \;=\; \tfrac{-r}{1-r}.$$

Finally, we have to prove formula (16): we obtain from formula (12) and for $k \geq 1$ that

$$\begin{aligned}
\mathrm{Cov}\,[n_t, n_{t-k}] &= \mathrm{Cov}\,\big[\mathbb{E}(n_t \mid n_{t-1},\ldots),\; \mathbb{E}(n_{t-k} \mid n_{t-1},\ldots)\big] \\
&= \mathrm{Cov}\,\big[\big(r+(1-r)b\big)\,n_{t-1} + N(1-r)a,\; n_{t-k}\big] \\
&= \big(r+(1-r)b\big)\,\mathrm{Cov}\,[n_{t-1}, n_{t-k}] \;= \ldots = \; \big(r+(1-r)b\big)^k\,\sigma^2.
\end{aligned}$$

A.3 Proof of Lemma 5.1.1

The function $g(x) = f(x) - x = \alpha a - \big(1-\alpha\,(1-a+b)\big)\,x - \alpha b\,x^2$ is a quadratic polynomial (if $b \neq 0$) with $g(0) = \alpha a > 0$ and $g(1) = -(1-\alpha) < 0$. Hence, there is a unique root in $(0;1)$. b determines the sign of the quadratic term. If $b < 0$, the smaller one of the two roots of $g(x)$ is in $(0;1)$, for $b > 0$, it is the larger one.

The roots of $g(x)$ are given by

$$x_{1,2} \;=\; \frac{1-\alpha\,(1-a+b) \;\pm\; \sqrt{\big(1-\alpha\,(1-a+b)\big)^2 + 4\alpha^2 ab}}{-2\alpha b},$$

so

$$x_2 - x_1 \;=\; \frac{\sqrt{\left(1-\alpha\left(1-a+b\right)\right)^2+4\alpha^2 ab}}{\alpha b}.$$

Therefore, $x_2 < x_1$ if $b < 0$ and $x_2 > x_1$ if $b > 0$, that is, x_2 is the unique root in $(0;1)$.