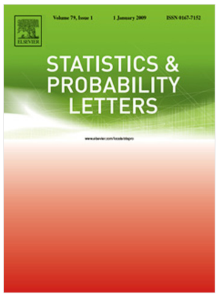
Accepted Manuscript



The max-BARMA models for counts with bounded support

Christian H. Weiß, Manuel G. Scotto, Tobias A. Möller, Sónia Gouveia

PII: S0167-7152(18)30255-4
DOI: https://doi.org/10.1016/j.spl.2018.07.011
Reference: STAPRO 8290

To appear in: *Statistics and Probability Letters*

Received date : 4 April 2017
Accepted date : 15 July 2018

The max-BARMA models for Counts with Bounded Support

Christian H. Weiß[a,*], Manuel G. Scotto[b], Tobias A. Möller[c], Sónia Gouveia[d]

[a]Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany.
[b]CEMAT and Department of Mathematics, IST University of Lisbon, Lisbon, Portugal.
[c]Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany.
[d]IEETA and CIDMA, University of Aveiro, Portugal.

Abstract

In this note, we introduce a discrete counterpart of the conventional max-autoregressive moving-average process of Davis & Resnick (1989), based on the binomial thinning operator and driven by a sequence of i. i. d. nonnegative integer-valued random variables with a finite range of counts. Basic probabilistic and statistical properties of this new class of models are discussed in detail, namely the existence of a stationary distribution, and how observations' and innovations' distributions are related to each other. Furthermore, parameter estimation is also addressed.

Keywords: thinning operator; autoregressive moving-average processes; finite counts.

1. Introduction

Modeling the temporal dependence of integer-valued time series defined on a finite range of counts, say $\{0, 1, \ldots, n\}$, is nowadays a topic of research which is gaining importance in time series analysis. Note that traditional integer-valued ARMA-type models are useless in this context, since such models are defined over unbounded sets. To tackle this limitation, McKenzie (1985) introduced the *binomial AR(1) model* based on the so-called *binomial thinning* operator (Steutel & van Harn, 1979), in which X_t takes the form

$$X_t \;=\; \alpha\circ X_{t-1} \;+\; \beta\circ(n-X_{t-1}), \tag{1.1}$$

where the binomial thinning operator "$\phi\,\circ$" is defined as follows: if X is a discrete random variable with range $\mathbb{N}_0$ and if $\phi\in(0;1)$, then the random variable $\phi\circ X$ is defined by

*Corresponding author
Email address: weissc@hsu-hh.de (Christian H. Weiß)

the conditional binomial distribution $\mathrm{Bin}(X, \phi)$, given the value of X; the boundary values $\phi = 0$ and $\phi = 1$ are included by setting $0 \circ X := 0$ and $1 \circ X := X$. Note that by construction, X_t preserves the boundedness of the range. The binomial AR(1) model was extended to a pth-order AR-like model by Weiß (2009), who combined the data generating mechanism of the binomial AR(1) model with a random mixture. The approach by Weiß (2009), however, cannot be extended beyond a purely autoregressive model, and also a pure moving-average model cannot be constructed in this way.

Our aim in this paper is to construct a *full* integer-valued time series model with finite support, in the sense that the model should include *both* an autoregressive-type and a moving-average-type component (therefore a "full" counterpart), and it should certainly contain both a purely autoregressive-type and a purely moving-average-type model as a special case. To this extent we introduce a discrete counterpart of the conventional max-autoregressive moving-average process of Davis & Resnick (1989) which will be referred to as the maximum BARMA (in short max-BARMA(p, q)) model.

The novel max-BARMA(p, q) model proposed in Section 2 includes both an autoregressive and a moving-average part, and a stationary solution exists under rather weak conditions. After having discussed stochastic properties and important special cases, we present an approach for parameter estimation in Section 3, and we conclude in Section 4.

2. The max-BARMA Model

Motivated by the max-ARMA models (Davis & Resnick, 1989), we define an integer-valued counterpart of this model for counts having the bounded support $\{0, \ldots, n\}$ with a fixed upper limit $n \in \mathbb{N}$. We refer to these models as the max-BARMA(p, q) models ("B" like "bounded"), and they are defined by the recursion

$$X_t \ = \ \max\{\alpha_1 \circ X_{t-1}, \ldots, \alpha_\mathrm{p} \circ X_{t-\mathrm{p}},\ \epsilon_t,\ \beta_1 \circ \epsilon_{t-1}, \ldots, \beta_\mathrm{q} \circ \epsilon_{t-\mathrm{q}}\}, \tag{2.1}$$

where the i. i. d. innovations are assumed to have the bounded range $\{0, \ldots, n\}$. Note that binomial thinning always has a non-increasing effect, i. e., $\alpha \circ x \leq x$ holds, so (2.1) preserves the boundedness of the range $\{0, \ldots, n\}$. The thinnings in (2.1) are assumed to be executed each time anew: although we omit this in the sequel for the sake of readability, it would be more correct to add a time index t to the thinnings,

i. e., $X_t = \max\{\alpha_1 \circ_t X_{t-1}, \dots, \beta_{\mathrm{q}} \circ_t \epsilon_{t-\mathrm{q}}\}$. In addition, all thinnings are assumed to be executed completely independent of each other, i. e., in particular, we assume that $\alpha_1 \circ_{t+1} X_t, \dots, \alpha_{\mathrm{p}} \circ_{t+\mathrm{p}} X_t$ as well as $\beta_1 \circ_{s+1} \epsilon_s, \dots, \beta_{\mathrm{q}} \circ_{s+\mathrm{q}} \epsilon_s$ are conditionally independent, given the values of X_t and ϵ_s, respectively.

Important special cases (to be discussed in some more detail in the later Sections 2.1 and 2.2) are the purely moving-average max-BMA(q) model (i. e., p = 0) defined by

$$X_t \;=\; \max\{\epsilon_t,\ \beta_1 \circ \epsilon_{t-1}, \dots, \beta_{\mathrm{q}} \circ \epsilon_{t-\mathrm{q}}\},$$

and the purely autoregressive max-BAR(p) model (i. e., q = 0) defined by

$$X_t \;=\; \max\{\alpha_1 \circ X_{t-1}, \dots, \alpha_{\mathrm{p}} \circ X_{t-\mathrm{p}},\ \epsilon_t\}.$$

The max-BARMA(p, q) process (2.1) can be represented as a (p + q)-dimensional homogeneous finite Markov chain,

$$\boldsymbol{Z}_t \;:=\; \big(X_t; X_{t-1}, \dots, X_{t-\mathrm{p}+1}, \epsilon_t, \dots, \epsilon_{t-\mathrm{q}+1}\big)^\top, \tag{2.2}$$

which is understood to reduce to $\big(X_t; X_{t-1}, \dots, X_{t-\mathrm{p}+1}\big)^\top$ in the purely autoregressive case (q = 0), and to $\big(X_t; \epsilon_t, \dots, \epsilon_{t-\mathrm{q}+1}\big)^\top$ in the purely moving-average case (p = 0). Partition the possible outcomes $\boldsymbol{z} \in \{0, \dots, n\}^{\mathrm{p}+\mathrm{q}}$ of $\boldsymbol{Z}_t$ as $\boldsymbol{z} = (z_1, \dots, z_{\mathrm{p}}, z_{-1}, \dots, z_{-\mathrm{q}})^\top$. For the cases with $\mathrm{q} \geq 1$, the range $\mathcal{Z}$ of $\boldsymbol{Z}_t$ is only a subset of $\{0, \dots, n\}^{\mathrm{p}+\mathrm{q}}$, since z_1 can never exceed the maximum of the remaining components, since z_1 can never fall below z_{-1}, and since the ϵ's distribution might have probability 0 for some counts in $\{0, \dots, n\}$. For $\mathrm{q} = 0$, $\mathcal{Z} = \{0, \dots, n\}^{\mathrm{p}}$, i. e., the full range is possible (provided that $P(\epsilon = 0), P(\epsilon = n) > 0$).

Note that the 1-step-ahead transition probabilities $p_{\boldsymbol{z}|\boldsymbol{y}} := P(\boldsymbol{Z}_t = \boldsymbol{z} \mid \boldsymbol{Z}_{t-1} = \boldsymbol{y})$ are

$$\begin{aligned} p_{\boldsymbol{z}|\boldsymbol{y}} \;=\; & \delta_{z_2,y_1} \cdots \delta_{z_{\mathrm{p}},y_{\mathrm{p}-1}} \cdot \delta_{z_{-2},y_{-1}} \cdots \delta_{z_{-\mathrm{q}},y_{-\mathrm{q}+1}} \cdot P(\epsilon_t = z_{-1}) \cdot \\ & P(X_t = z_1 \mid X_{t-1} = y_1, \dots, X_{t-\mathrm{p}} = y_{\mathrm{p}}, \epsilon_t = z_{-1}, \epsilon_{t-1} = y_{-1}, \dots, \epsilon_{t-\mathrm{q}} = y_{-\mathrm{q}}), \end{aligned} \tag{2.3}$$

with the last factor being equal to

$$\begin{aligned} & P\Big(\max\{\alpha_1 \circ y_1, \dots, \alpha_{\mathrm{p}} \circ y_{\mathrm{p}},\ z_{-1},\ \beta_1 \circ y_{-1}, \dots, \beta_{\mathrm{q}} \circ y_{-\mathrm{q}}\} \;=\; z_1\Big) \;= \\ & \quad =\; F_{\mathrm{Bin}(y_1,\alpha_1)}(z_1) \cdots F_{\mathrm{Bin}(y_{\mathrm{p}},\alpha_{\mathrm{p}})}(z_1) \cdot \mathbb{1}(z_{-1} \leq z_1) \cdot F_{\mathrm{Bin}(y_{-1},\beta_1)}(z_1) \cdots F_{\mathrm{Bin}(y_{-\mathrm{q}},\beta_{\mathrm{q}})}(z_1) \\ & \qquad -\; F_{\mathrm{Bin}(y_1,\alpha_1)}(z_1 - 1) \cdots \mathbb{1}(z_{-1} \leq z_1 - 1) \cdots F_{\mathrm{Bin}(y_{-\mathrm{q}},\beta_{\mathrm{q}})}(z_1 - 1), \end{aligned} \tag{2.4}$$

where $F_{\text{Bin}(n,\pi)}(x)$ abbreviates the cumulative distribution function (cdf) of the binomial distribution $\text{Bin}(n,\pi)$. This factorization follows immediately from the well-known equality

$$P\big(\max\{Y_1,\ldots,Y_m\}\le x\big) \;=\prod_{i=1}^{m} P(Y_i\le x), \tag{2.5}$$

for independent random variables $Y_1,\ldots,Y_m$.

The Markov-chain representation (2.2) and (2.3) is used to prove the existence of the max-BARMA(p,q) process. The main result of this section is given below.

2.1 Theorem *If all thinning probabilities satisfy $\alpha_1,\ldots,\beta_{\mathrm{q}}\in(0;1)$, and if the innovations satisfy $P(\epsilon=k)>0$ for all $k\in\{0,\ldots,n\}$, then the max-BARMA(p,q) process (2.1) is ergodic with a unique stationary solution, which is also ϕ-mixing with geometrically decreasing weights. For $\mathrm{p}=0$ or $\mathrm{q}=0$, it suffices to require that $P(\epsilon=0),P(\epsilon=n)>0$.*

Proof: The proof starts by showing that the $(\mathrm{p}+\mathrm{q}+1)$-step-ahead transition probabilities equal $p_{\boldsymbol{z}|\boldsymbol{y}}(\mathrm{p}+\mathrm{q}+1)$ of the Markov chain $(\boldsymbol{Z}_t)_{\mathbb{Z}}$ are truly positive, thus implying that $(\boldsymbol{Z}_t)_{\mathbb{Z}}$ constitutes a primitive finite Markov chain.

Case $\mathrm{q}=0$:

For $\boldsymbol{Z}_t=\big(X_t,\ldots,X_{t-\mathrm{p}+1}\big)^\top$, the $(\mathrm{p}+1)$-step-ahead transition probabilities equal

$$\begin{aligned}
p_{\boldsymbol{z}|\boldsymbol{y}}(\mathrm{p}+1) \;&=\; P\big(X_t=z_1,\ldots,X_{t-\mathrm{p}+1}=z_{\mathrm{p}}\mid X_{t-\mathrm{p}-1}=y_1,\ldots,X_{t-2\mathrm{p}}=y_{\mathrm{p}}\big)\\
&=\; \textstyle\sum_{z_{\mathrm{p}+1}} P\big(X_t=z_1,\ldots,X_{t-\mathrm{p}+1}=z_{\mathrm{p}},X_{t-\mathrm{p}}=z_{\mathrm{p}+1}\mid X_{t-\mathrm{p}-1}=y_1,\ldots,X_{t-2\mathrm{p}}=y_{\mathrm{p}}\big)\\
&=\; \textstyle\sum_{z_{\mathrm{p}+1}} P(z_1\,|\,z_2,\ldots,z_{\mathrm{p}+1})\cdots P(z_{\mathrm{p}}\,|\,z_{\mathrm{p}+1},y_1,\ldots,y_{\mathrm{p}-1})\cdot P(z_{\mathrm{p}+1}\,|\,y_1,\ldots,y_{\mathrm{p}}),
\end{aligned}$$

which is truly positive if one of the summands is truly positive, i. e., if one of the summands consists of only truly positive factors. This happens at least for the summand $z_{\mathrm{p}+1}=n$:

- If one of the past observations $x_1,\ldots,x_{\mathrm{p}}$ is equal to n, say $x_i=n$, then we have $P(x_0\,|\,x_1,\ldots,x_{\mathrm{p}})>0$, because $\alpha_i\circ n$ has the full support $\{0,\ldots,n\}$ since all $\alpha_1,\ldots,\alpha_{\mathrm{p}}\in(0;1)$, and because $P(\epsilon=0)>0$ is assumed.
- Conditional probabilities of the form $P(n\,|\,x_1,\ldots,x_{\mathrm{p}})$ are truly positive, since $P(\epsilon=n)>0$ is assumed.

Case $\mathrm{p} = 0$:

For $\boldsymbol{Z}_t = \big(X_t; \epsilon_t, \ldots, \epsilon_{t-\mathrm{q}+1}\big)^\top$, the (q+1)-step-ahead transition probabilities equals, with the same arguments as before,

$$
\begin{aligned}
& p_{\boldsymbol{z}|\boldsymbol{y}}(\mathrm{q}+1) \\
&= P\big(X_t = z_1, \epsilon_t = z_{-1}, \ldots, \epsilon_{t-\mathrm{q}+1} = z_{-\mathrm{q}} \mid X_{t-\mathrm{q}-1} = y_1, \epsilon_{t-\mathrm{q}-1} = y_{-1}, \ldots, \epsilon_{t-2\mathrm{q}} = y_{-\mathrm{q}}\big) \\
&= \textstyle\sum_{z_2,\ldots,z_{\mathrm{q}+1},z_{-\mathrm{q}-1}} P(\epsilon_{t-\mathrm{q}} = z_{-\mathrm{q}-1}) \prod_{j=0}^{\mathrm{q}-1} P(\epsilon_{t-j} = z_{-j-1}) \\
&\quad \cdot P\big(X_t = z_1 \mid \epsilon_t = z_{-1}, \ldots, \epsilon_{t-\mathrm{q}} = z_{-\mathrm{q}-1}\big) \\
&\quad \cdot \textstyle\prod_{i=1}^{\mathrm{q}} P\big(X_{t-i} = z_{i+1} \mid \epsilon_{t-i} = z_{-i-1}, \ldots, \epsilon_{t-i-\mathrm{q}} = y_{-i}\big).
\end{aligned}
$$

In analogy to the case $\mathrm{q} = 0$, if $P(\epsilon = n) > 0$ and if $\beta_1, \ldots, \beta_{\mathrm{q}} \in (0;1)$, then every summand with $z_{\mathrm{q}+1} = z_{-\mathrm{q}-1} = n$ and $z_{i+1} = z_{-i-1}$ for $i = 1, \ldots, \mathrm{q}-1$ is truly positive. Note that $\boldsymbol{z} \in \mathcal{Z}$ excludes "impossible states" such as $z_1 < z_{-1}$.

Case $\mathrm{p}, \mathrm{q} \geq 1$:

Analogous arguments as before are applied to the $(\mathrm{p}+\mathrm{q}+1)$-step-ahead transition probabilities:

$$
\begin{aligned}
& p_{\boldsymbol{z}|\boldsymbol{y}}(\mathrm{p}+\mathrm{q}+1) \\
&= P\big(X_t = z_1, \ldots, X_{t-\mathrm{p}+1} = z_{\mathrm{p}}, \epsilon_t = z_{-1}, \ldots, \epsilon_{t-\mathrm{q}+1} = z_{-\mathrm{q}} \\
&\qquad \mid X_{t-\mathrm{p}-\mathrm{q}-1} = y_1, \ldots, X_{t-\mathrm{q}-2\mathrm{p}} = y_{\mathrm{p}}, \epsilon_{t-\mathrm{p}-\mathrm{q}-1} = y_{-1}, \ldots, \epsilon_{t-\mathrm{p}-2\mathrm{q}} = y_{-\mathrm{q}}\big) \\
&= \textstyle\sum_{z_{\mathrm{p}+1},\ldots,z_{\mathrm{p}+\mathrm{q}+1},z_{-\mathrm{q}-1},\ldots,z_{-\mathrm{p}-\mathrm{q}-1}} \prod_{j=0}^{\mathrm{q}-1} P(\epsilon_{t-j} = z_{-j-1}) \prod_{j=\mathrm{q}}^{\mathrm{q}+\mathrm{p}} P(\epsilon_{t-j} = z_{-j-1}) \\
&\quad \cdot \textstyle\prod_{i=0}^{\mathrm{p}+\mathrm{q}} P\big(X_{t-i} = z_{i+1} \mid X_{t-i-1} = z_{i+2}, \ldots, X_{t-i-\mathrm{p}} = y_{i-\mathrm{q}}, \epsilon_{t-i} = z_{-i-1}, \ldots, \epsilon_{t-i-\mathrm{q}} = y_{-i}\big).
\end{aligned}
$$

Here, we focus on summands with $z_{i+1} = z_{-i-1}$ for $i = \mathrm{p}, \ldots, \mathrm{p}+\mathrm{q}$ and with $z_{-i-1} = z_{i+1}$ for $i = \mathrm{q}, \ldots, \mathrm{p}+\mathrm{q}$; if both z_{i+1} and z_{-i-1} can be selected, we consider the summands with $z_{i+1} = \cdots = z_{-i-1} = n$.

Now we follow the proof in Section 2.2 of Weiß (2013) step-by-step to obtain that $(\boldsymbol{Z}_t)_{\mathbb{Z}}$ is a primitive finite Markov chain, which, in turn, leads to $(X_t)_{\mathbb{Z}}$ being stationary, ergodic and ϕ-mixing with geometrically decreasing weights. #

2.1. The max-BMA(q) Models

The purely moving-average max-BMA(q) model,

$$X_t \;=\; \max\{\epsilon_t,\ \beta_1 \circ \epsilon_{t-1}, \ldots, \beta_{\mathrm{q}} \circ \epsilon_{t-\mathrm{q}}\}, \tag{2.6}$$

constitutes q-dependent model for bounded counts, i. e., X_t and X_{t-k} are independent of each other for lag $k > \mathrm{q}$. In particular, this implies that also the autocorrelation function (ACF) satisfies $\rho(k) = 0$ for $k > \mathrm{q}$. Furthermore, since the innovations (ϵ_t) are i. i. d., and since thinnings are executed independently, the components $\epsilon_t, \beta_1 \circ \epsilon_{t-1}, \ldots, \beta_{\mathrm{q}} \circ \epsilon_{t-\mathrm{q}}$ within the max-operator are independent. Therefore, the marginal cdf satisfies

$$F_X(x) \;= F_\epsilon(x) \prod_{i=1}^{q} F_{\beta_i \circ \epsilon}(x), \tag{2.7}$$

where the factorization follows again from the equality (2.5). Using the closed-form formula (2.7), further marginal properties are easily computed, e. g., the marginal mean as $E[X] = \sum_{x=0}^{n-1} \big(1 - F(x)\big)$, or the probability mass function (pmf) $p_X(x) := P(X = x)$ by differencing.

2.2 Example (Binomial Innovations) Let the innovations be binomially distributed, $\epsilon \sim \mathrm{Bin}(n, \pi)$, then $\beta \circ \epsilon \sim \mathrm{Bin}(n, \pi\, \beta)$. So the marginal distribution according to (2.7) equals

$$F_X(x) \;=\; F_{\mathrm{Bin}(n,\pi\,\beta_1)}(x) \cdots F_{\mathrm{Bin}(n,\pi\,\beta_{\mathrm{q}})}(x) \cdot F_{\mathrm{Bin}(n,\pi)}(x),$$

which is a product of binomial cdfs. Here, $F_{\mathrm{Bin}(n,\pi)}(x) \;=\; \sum_{m=0}^{x} \binom{n}{m} \pi^m\, (1-\pi)^{n-m}$. In particular, the zero probability $p_X(0) = F_X(0)$ becomes

$$p_X(0) \;=\; (1 - \pi\,\beta_1)^n \cdots (1 - \pi\,\beta_{\mathrm{q}})^n \cdot (1-\pi)^n.$$

2.3 Example (Uniform Innovations) If the marginal distribution of a max-BMA(q) process with uniform innovations, i. e., with pmf $P(\epsilon = l) = \frac{1}{n+1}$, has to be computed according to (2.7), then the distribution of $\beta \circ \epsilon$ is required. We compute

$$F_{\beta\circ\epsilon}(x) \;=\; \textstyle\sum_{l=0}^{n}\; P(\beta \circ l \le x)\, P(\epsilon = l) \;=\; \frac{1}{n+1} \sum_{l=0}^{n}\; F_{\mathrm{Bin}(l,\beta)}(x),$$

which is the mean of the binomial cdfs $F_{\mathrm{Bin}(0,\beta)}(x), \ldots, F_{\mathrm{Bin}(n,\beta)}(x)$. In particular, since

$$F_{\beta\circ\epsilon}(0) \;=\; \tfrac{1}{n+1} \textstyle\sum_{l=0}^{n}\; (1-\beta)^l \;=\; \tfrac{1}{n+1}\, \tfrac{1-(1-\beta)^{n+1}}{\beta},$$

the zero probability becomes

$$p_X(0) \;=\; \tfrac{1}{(n+1)^{\mathrm{q}+1}}\, \textstyle\prod_{j=1}^{\mathrm{q}}\; \tfrac{1-(1-\beta_j)^{n+1}}{\beta_j}.$$

2.4 Example (Two-Point Innovations) Next, we consider a max-BMA(q) process with innovations stemming from a two-point distribution on $\{0, \ldots, n\}$, i. e., with $P(\epsilon_t = n) = \pi =$

$1 - P(\epsilon_t = 0)$. If the marginal distribution has to be computed according to (2.7), then the distribution of $\beta \circ \epsilon$ is required. We compute

$$F_{\beta\circ\epsilon}(x) \;=\; \sum_{k=0}^{x} \sum_{l=0}^{n} \; P(\beta \circ l = k)\, P(\epsilon = l) \;=\; (1-\pi) + \pi\, F_{\beta\circ n}(x),$$

which is the zero-inflated binomial (ZIB) distribution $\mathrm{ZIB}(n, \beta, 1-\pi)$. So (2.7) leads to

$$F_X(x) \;=\; (1-\pi)\,\big(1-\pi+\pi\, F_{\mathrm{Bin}(n,\beta_1)}(x)\big)\cdots\big(1-\pi+\pi\, F_{\mathrm{Bin}(n,\beta_{\mathrm{q}})}(x)\big)$$

for all $x < n$; certainly, $F_X(n) = 1$. The zero probability becomes

$$p_X(0) \;=\; (1-\pi)\,\big(1-\pi+\pi\,(1-\beta_1)^n\big)\cdots\big(1-\pi+\pi\,(1-\beta_{\mathrm{q}})^n\big).$$

*2.2. The max-BAR(*p*) Models*

The purely autoregressive max-BAR(p) model, defined by

$$X_t \;=\; \max\{\alpha_1 \circ X_{t-1}, \ldots, \alpha_{\mathrm{p}} \circ X_{t-\mathrm{p}},\; \epsilon_t\}, \tag{2.8}$$

is a p-order Markov model. Its conditional cdf is computed

$$F_{X_t|x_{t-1},\ldots,x_{t-\mathrm{p}}}(x) \;=\; F_{\mathrm{Bin}(x_{t-1},\alpha_1)}(x)\cdots F_{\mathrm{Bin}(x_{t-\mathrm{p}},\alpha_{\mathrm{p}})}(x)\cdot F_\epsilon(x), \tag{2.9}$$

utilizing again the general equality (2.5). In particular, the conditional zero probability equals

$$P(X_t = 0 \mid x_{t-1}, \ldots, x_{t-\mathrm{p}}) \;=\; p_\epsilon(0) \prod_{i=1}^{p} (1-\alpha_i)^{x_{t-i}}.$$

The most simple instance of the max-BAR(p) model is the max-BAR(1) model (with autoregressive parameter $\alpha_1 = \alpha$) defined by

$$X_t \;=\; \max\{\alpha \circ X_{t-1},\; \epsilon_t\}, \tag{2.10}$$

which constitutes a Markov chain and serves as an alternative to the binomial AR(1) model (1.1). Its transition probabilities are computed from its conditional cdf

$$F_{X_t|X_{t-1}=l}(k) \;=\; F_\epsilon(k) \sum_{m=0}^{\min\{k,l\}} \binom{l}{m} \alpha^m\, (1-\alpha)^{l-m}, \tag{2.11}$$

see (2.9). Conditional factorial moments are computed as

$$\begin{aligned} E[(X_t)_{(r)} \mid X_{t-1} = l] \;&=\; E\big[\max\{\alpha \circ X_{t-1},\, \epsilon_t\}_{(r)} \mid X_{t-1} = l\big] \\ &=\; \sum_{m=0}^{l} \binom{l}{m} \alpha^m\, (1-\alpha)^{l-m} E\big[\max\{m,\, \epsilon_t\}_{(r)}\big], \end{aligned} \tag{2.12}$$

where

$$E\big[\max\{m,\ \epsilon_t\}_{(r)}\big] \;=\; m_{(r)}\,P(\epsilon_t \le m) \;+\; \textstyle\sum_{k>m}\ k_{(r)}\,P(\epsilon_t = k)$$
$$\;=\; m_{(r)}\,F_\epsilon(m) \;+\; \mu_{\epsilon,(r)} - \textstyle\sum_{k\le m}\ k_{(r)}\,P(\epsilon_t = k).$$

2.5 Example (Uniform Innovations) Let the ϵ_t be uniformly distributed on $\{0,\dots,n\}$, i. e., with cdf $F_\epsilon(k) = \frac{k+1}{n+1}$. The conditional mean follows from (2.11) as

$$\begin{aligned}
E[X_t \mid X_{t-1}=l] &= \textstyle\sum_{k=0}^{n} \big(1 - F_{X_t|X_{t-1}=l}(k)\big) \\
&= n+1-\textstyle\sum_{k=0}^{n} \frac{k+1}{n+1} \sum_{m=0}^{k} \binom{l}{m}\,\alpha^m\,(1-\alpha)^{l-m} \\
&= n+1-\textstyle\sum_{m=0}^{n} \binom{l}{m}\,\alpha^m\,(1-\alpha)^{l-m} \sum_{k=m}^{n} \frac{k+1}{n+1} \\
&= n+1-\textstyle\sum_{m=0}^{l} \binom{l}{m}\,\alpha^m\,(1-\alpha)^{l-m}\,\frac{1}{n+1}\big(\frac{(n+1)(n+2)}{2} - \frac{m(m+1)}{2}\big) \\
&= n+1-\frac{n+2}{2}+\frac{1}{2(n+1)}\textstyle\sum_{m=0}^{l} \binom{l}{m}\,\alpha^m\,(1-\alpha)^{l-m}\,\big(m(m-1)+2m\big) \\
&= \frac{\alpha^2}{2(n+1)}\,l^2 \;+\; \frac{\alpha\,(2-\alpha)}{2(n+1)}\,l \;+\; \frac{n}{2},
\end{aligned}$$

where in the second last step, we used the formula for the factorial moments of the binomial distribution. So the conditional mean is a quadratic function in the previous observation. Alternatively, we might have used (2.12) to derive this formula, or to derive expressions for higher-order factorial moments.

2.6 Example (Two-Point Innovations) Let ϵ_t follow a two-point distribution on $\{0,\dots,n\}$ with $P(\epsilon_t = n) = \pi = 1-P(\epsilon_t = 0)$. By the definition of the Max-BAR(1) model, the generated sample paths will show some kind of saw-tooth pattern: at each time when $\epsilon = n$, a shock is generated and shifts X to the value n, while afterwards, during the period until the next shock (when $\epsilon = 0$), the thinning $\alpha \circ X$ causes the observations to decay monotonically.

For the two-point distribution, we obtain

$$F_\epsilon(k) \;=\; 1-\pi+\delta_{k,n}\,\pi, \quad \mu_{\epsilon,(r)} := E[(\epsilon_t)_{(r)}] \;=\; n_{(r)}\,\pi, \quad \textstyle\sum_{k=0}^{m}\ k_{(r)}\,P(\epsilon_t = k) \;=\; \delta_{m,n}\,n_{(r)}\,\pi.$$

So altogether, see Appendix Appendix A for detailed derivations,

$$E\big[\max\{m,\ \epsilon_t\}_{(r)}\big] \;=\; m_{(r)}\,(1-\pi) + n_{(r)}\,\pi.$$

Hence, using (2.12) together with the formula for the binomial factorial moments,

$$E[(X_t)_{(r)} \mid X_{t-1}=l] \;=\; \alpha^r\,(1-\pi)\,l_{(r)} \;+\; n_{(r)}\,\pi,$$

which is an rth-order polynomial in the last observation with leading coefficient $\alpha^r\,(1-\pi)$. In particular, the conditional mean

$$E[X_t \mid X_{t-1}] \;=\; \alpha\,(1-\pi)\,X_{t-1} \;+\; n\,\pi,$$

is linear in the previous observation, so a max-BAR(1) model with two-point innovations is a CLAR(1) model (Grunwald et al., 2000). As a result, we obtain the marginal mean as

$$\mu := E[X] \;=\; \tfrac{n\,\pi}{1-\alpha\,(1-\pi)}.$$

Similarly, the conditional variance

$$V[X_t \mid X_{t-1}] \;=\; X_{t-1}^2\,\alpha^2\,\pi(1-\pi) \;-\; X_{t-1}\,\alpha\,(1-\pi)\,(2\,n\,\pi-1+\alpha) \;+\; n^2\,\pi(1-\pi),$$

is a quadratic polynomial in the last observation. So

$$\sigma^2 := V[X_t] \;=\; \frac{\alpha^2\,\pi(1-\pi)\,\mu^2 \;-\; \alpha\,(1-\pi)\,(2\,n\,\pi+\alpha)\,\mu \;+\; n^2\,\pi(1-\pi)}{1-\alpha^2\,\pi(1-\pi)}.$$

The transition probabilities $p_{k|l}$ implied by (2.11) take a very simple form. For $k=0,\ldots,n-1$, we have

$$F_{X_t|X_{t-1}=l}(k) = (1-\pi)\sum_{m=0}^{k}\binom{l}{m}\,\alpha^m\,(1-\alpha)^{l-m},$$

so by differencing, we obtain $p_{k|l}=(1-\pi)\,\binom{l}{k}\,\alpha^k\,(1-\alpha)^{l-k}$, which equals 0 for $k>l$. Altogether, we obtain

$$p_{k|l} \;=\; \delta_{k,n}\,\pi \;+\; (1-\pi)\,P(\alpha\circ l=k),$$

where $P(\alpha\circ l=k)=0$ if $l<k$, i.e., the transition matrix exhibits a triangular-type structure.

So far, we specified the max-BAR(1) model by specifying its innovations' distribution. In applications, one sometimes prefers to specify the observations' marginal distribution instead and to derive the corresponding distribution of the innovations. The marginal cdfs of the observations and innovations of a stationary max-BAR(1) process are related to each other by the equation

$$F_X(x) \;=\; F_{\alpha\circ X}(x)\cdot F_\epsilon(x). \tag{2.13}$$

According to Theorem 1.A.4 in Shaked & Shanthikumar (2007), $\alpha\circ X\;\leq_{\mathrm{st}}\;X$, i.e., we always have that $F_X(x)\leq F_{\alpha\circ X}(x)$. Thus, the quotient $F_X/F_{\alpha\circ X}$ always satisfies $0\leq F_X(x)/F_{\alpha\circ X}(x)\leq 1$ for all x. So the cdf $F_X(x)$ is a valid cdf for a max-BAR(1) model iff $F_X/F_{\alpha\circ X}$ is monotonically increasing (and this would then be the cdf of the innovations), which is equivalent to requiring that

$$\frac{p_X(x+1)}{p_{\alpha\circ X}(x+1)} \;\geq\; \frac{F_X(x)}{F_{\alpha\circ X}(x)} \qquad \text{for all } x=0,\dots,n-1, \tag{2.14}$$

where $p.(\cdot)$ denotes the respective pmf. Obviously, a sufficient condition for (2.14) is

$$\frac{p_X(x+1)}{p_{\alpha\circ X}(x+1)} \;\geq\; \frac{p_X(j)}{p_{\alpha\circ X}(j)} \qquad \text{for all } j\leq x. \tag{2.15}$$

The subsequent example demonstrates the application of condition (2.15) to prove that the binomial distribution is a possible marginal distribution of a max-BAR(1) process.

2.7 Example (Binomial Observations) Let $X \sim \mathrm{Bin}(n,\pi)$, then $\alpha\circ X \sim \mathrm{Bin}(n,\pi\,\alpha)$. Therefore,

$$\tfrac{p_X(x+1)}{p_{\alpha\circ X}(x+1)}\,p_{\alpha\circ X}(j) \;=\; \tfrac{(1-\pi\,\alpha)^{x+1-j}}{\alpha^{x+1-j}\,(1-\pi)^{x+1-j}}\,\tbinom{n}{j}\,\pi^j\,(1-\pi)^{n-j} \;=\; \left(\tfrac{1-\pi\,\alpha}{\alpha-\pi\,\alpha}\right)^{x+1-j} p_X(j) \;>\; p_X(j),$$

for all $j\leq x$, so condition (2.15) is satisfied. So any binomial distribution is a possible marginal distribution of a max-BAR(1) process. The innovations' cdf follows from (2.13) as $F_\epsilon(k) = F_{\mathrm{Bin}(n,\pi)}(k)/F_{\mathrm{Bin}(n,\pi\,\alpha)}(k)$, in particular, we have

$$p_\epsilon(0) \;=\; \frac{(1-\pi)^n}{(1-\pi\,\alpha)^n} \qquad \text{and} \quad \mu_\epsilon \;=\; \sum_{k=0}^{n-1}\big(1-F_\epsilon(k)\big) \;=\; n \;-\; \sum_{k=0}^{n-1}\frac{F_{\mathrm{Bin}(n,\pi)}(k)}{F_{\mathrm{Bin}(n,\pi\,\alpha)}(k)}.$$

An example of a bounded distribution, which cannot be used as a marginal for a max-BAR(1) process, is the two-point distribution.

2.8 Example (Two-Point Distribution) Let us assume that $B \sim \mathrm{Bin}(1,\pi)$ is a Bernoulli random variable, and define $X := n\cdot B$. For $\pi=0$ or $\pi=1$, this implies X to be constant, either constantly equal to 0 or to n, respectively. In the first case, it follows that also $\alpha\circ X$ and ϵ are constantly 0. In the second case, $\alpha\circ X$ is just binomially distributed according to $\mathrm{Bin}(n,\alpha)$ such that

$$\frac{F_X(x)}{F_{\alpha\circ X}(x)} \;=\; \frac{\mathbb{1}_{\{x\geq n\}}(x)}{F_{\mathrm{Bin}(n,\alpha)}(x)} \;=\; \mathbb{1}_{\{x\geq n\}}(x).$$

So again condition (2.14) is satisfied with ϵ being constantly n.

Now, let us look at the case $\pi\in(0;1)$, where X follows a non-degenerate two-point distribution on $\{0,n\}$. If $n>1$, $\alpha\circ X$ follows the zero-inflated binomial distribution with p.g.f. $\mathrm{pgf}_{\alpha\circ X}(s) \;=\; 1-\pi\,+\,\pi\,(1-\alpha+\alpha\,s)^n$, and we have

$$\frac{p_X(1)}{p_{\alpha\circ X}(1)} \;=\; \frac{0}{p_{\alpha\circ X}(1)} \;<\; \frac{1-\pi}{1-\pi+\pi\,(1-\alpha)^n} \;=\; \frac{F_X(0)}{F_{\alpha\circ X}(0)},$$

thus violating (2.14). So a non-degenerate two-point distribution cannot be the marginal distribution of a max-INAR(1) model. In fact, remembering the data generating mechanism (2.10), if $P(\epsilon = n) > 0$ and $P(\epsilon < n) > 0$ for some $n \in \mathbb{N}$, then the thinning $\alpha \circ X_{t-1}$ implies that $P(X = k) > 0$ for any $k \leq n$ ("no gaps in the range of X").

3. Parameter Estimation

Conditional maximum likelihood (CML) estimation, given the time series $x_1, \ldots, x_T$, can be implemented in analogy to the approach for Hidden-Markov models as described in Chapter 3 of Zucchini & MacDonald (2009). The idea is to compute the conditional likelihood function

$$\begin{aligned} L(\boldsymbol{\theta}) &:= P(x_T, \ldots, x_{\mathrm{p}+1} \mid x_{\mathrm{p}}, \ldots, x_1,\ \boldsymbol{\theta}) \\ &:= P(X_T = x_T, \ldots, X_{\mathrm{p}+1} = x_{\mathrm{p}+1} \mid X_{\mathrm{p}} = x_{\mathrm{p}}, \ldots, X_1 = x_1,\ \boldsymbol{\theta}) \end{aligned}$$

for the max-BARMA(p, q) model with parameters $\boldsymbol{\theta} = (\alpha_1, \ldots, \beta_{\mathrm{q}}, \boldsymbol{\theta}_\epsilon)^\top$ in a recursive way, where $\boldsymbol{\theta}_\epsilon$ abbreviates the parameters of the ϵ_t's distribution, and to maximize it numerically by using a routine for constrained optimization. To describe this recursive scheme, let us introduce some notations:

- $\boldsymbol{\epsilon}_t := (\epsilon_t, \ldots, \epsilon_{t-\mathrm{q}})^\top$ with range $\mathcal{E} := \{0, \ldots, n\}^{\mathrm{q}+1}$ having the cardinality $d := |\mathcal{E}| = (n+1)^{\mathrm{q}+1}$.
- The elements of $\mathcal{E}$ are arranged in a certain lexicographic ordering: $\mathcal{E} = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$.
- The vector-valued process $(\boldsymbol{\epsilon}_t)$ constitutes a homogeneous Markov chain with sparse transition matrix $\mathbf{A} = (a_{i|j})$ given by

  $$a_{i|j} := P(\boldsymbol{\epsilon}_t = \boldsymbol{e}_i \mid \boldsymbol{\epsilon}_{t-1} = \boldsymbol{e}_j) \ = \ \delta_{e_{i,1},\, e_{j,0}} \cdots \delta_{e_{i,\mathrm{q}},\, e_{j,\mathrm{q}-1}} \cdot P(\epsilon = e_{i,0}).$$

  Note that $\mathbf{A}$ only depends on $\boldsymbol{\theta}_\epsilon$.
- For all $x \leq \max\{x_1, \ldots, x_T\}$, it will be necessary to compute the d-cumulative probabilities

  $$c_i(x) \ = \ \mathbb{1}(e_{i,0} \leq x) \prod_{j=1}^{q} F_{\mathrm{Bin}(e_{i,j}, \beta_j)}(x),\ i = 1, \ldots, d.$$

  Note that $c_i(x)$ only depends on the parameters $\beta_1, \ldots, \beta_{\mathrm{q}}$.

To compute the conditional likelihood function $L(\boldsymbol{\theta})$ recursively, we consider the probabilities (as functions of $\boldsymbol{\theta}$)

$$\gamma_{t,i} \;:=\; P(\boldsymbol{\epsilon}_t = \boldsymbol{e}_i,\ x_t,\ldots,x_{\mathrm{p}+1} \mid x_{\mathrm{p}},\ldots,x_1),\ t \geq \mathrm{p}+1,\ i = 1,\ldots,d,$$

since $L(\boldsymbol{\theta}) \;=\; \sum_{i=1}^{d} \gamma_{T,i}$. Further, note that $\gamma_{t+1,i}$ can be factorized as follows:

$$\gamma_{t+1,i} \;=\; P(x_{t+1} \mid \boldsymbol{\epsilon}_{t+1} = \boldsymbol{e}_i,\ x_t,\ldots,x_{t-\mathrm{p}+1}) \cdot P(\boldsymbol{\epsilon}_{t+1} = \boldsymbol{e}_i,\ x_t,\ldots,x_{\mathrm{p}+1} \mid x_{\mathrm{p}},\ldots,x_1),$$

where the latter factor is computed as

$$\begin{aligned}
&P(\boldsymbol{\epsilon}_{t+1} = \boldsymbol{e}_i,\ x_t,\ldots,x_{\mathrm{p}+1} \mid x_{\mathrm{p}},\ldots,x_1) \\
&\quad = \textstyle\sum_{\boldsymbol{e}_j\in\mathcal{E}} P(\boldsymbol{\epsilon}_{t+1} = \boldsymbol{e}_i, \boldsymbol{\epsilon}_t = \boldsymbol{e}_j,\ x_t,\ldots,x_{\mathrm{p}+1} \mid x_{\mathrm{p}},\ldots,x_1) \;=\; \textstyle\sum_{\boldsymbol{e}_j\in\mathcal{E}} a_{i|j}\cdot\gamma_{t,j}.
\end{aligned}$$

The first factor of $\gamma_{t+1,i}$, in turn, is computed from the parameters $\alpha_1,\ldots,\beta_{\mathrm{q}}$ via (2.4), i. e.,

$$\begin{aligned}
P(x_{t+1} \mid \boldsymbol{\epsilon}_{t+1} = \boldsymbol{e}_i,\ x_t,\ldots,x_{t-\mathrm{p}+1}) \;&=\; F_{\mathrm{Bin}(x_t,\alpha_1)}(x_{t+1})\cdots F_{\mathrm{Bin}(x_{t-\mathrm{p}+1},\alpha_{\mathrm{p}})}(x_{t+1})\cdot c_i(x_{t+1}) \\
&\quad - F_{\mathrm{Bin}(x_t,\alpha_1)}(x_{t+1}-1)\cdots F_{\mathrm{Bin}(x_{t-\mathrm{p}+1},\alpha_{\mathrm{p}})}(x_{t+1}-1)\cdot c_i(x_{t+1}-1).
\end{aligned}$$

Overall, we update the vector $\boldsymbol{\gamma}_t = (\gamma_{t,1},\ldots,\gamma_{t,d})^\top$ to $\boldsymbol{\gamma}_{t+1}$ as follows:

$$\boldsymbol{\gamma}_{t+1} \;=\; \mathrm{diag}\Big(\big(P(x_{t+1} \mid \boldsymbol{\epsilon}_{t+1} = \boldsymbol{e}_i,\ x_t,\ldots,x_{t-\mathrm{p}+1})\big)_{i=1,\ldots,d}\Big)\,\mathbf{A}\,\boldsymbol{\gamma}_t.$$

The procedure is initialized by

$$\begin{aligned}
\gamma_{\mathrm{p}+1,i} \;&=\; P(\boldsymbol{\epsilon}_{\mathrm{p}+1} = \boldsymbol{e}_i,\ x_{\mathrm{p}+1} \mid x_{\mathrm{p}},\ldots,x_1) \\
&=\; P(x_{\mathrm{p}+1} \mid \boldsymbol{\epsilon}_{\mathrm{p}+1} = \boldsymbol{e}_i,\ x_{\mathrm{p}},\ldots,x_1)\cdot P(\boldsymbol{\epsilon}_{\mathrm{p}+1} = \boldsymbol{e}_i \mid x_{\mathrm{p}},\ldots,x_1) \\
&\approx\; P(x_{\mathrm{p}+1} \mid \boldsymbol{\epsilon}_{\mathrm{p}+1} = \boldsymbol{e}_i,\ x_{\mathrm{p}},\ldots,x_1)\cdot P(\boldsymbol{\epsilon}_{\mathrm{p}+1} = \boldsymbol{e}_i)
\end{aligned}$$

for $i = 1,\ldots,d$. Note that the scheme leads to the exact full likelihood for $\mathrm{p} = 0$ (i. e., for a max-BMA(q) model). In the purely autoregressive max-BAR(p) case ($\mathrm{q} = 0$), it essentially simplifies to the usual way of computing the conditional likelihood function of a Markov model, see (2.9):

$$\begin{aligned}
L(\boldsymbol{\theta}) \;&=\; \textstyle\prod_{t=\mathrm{p}+1}^{T} P(X_t = x_t | X_{t-1} = x_{t-1},\ldots,X_{t-\mathrm{p}} = x_{t-\mathrm{p}}) \\
&=\; \textstyle\prod_{t=\mathrm{p}+1}^{T} \Big(F_{\mathrm{Bin}(x_{t-1},\alpha_1)}(x_t)\cdots F_{\mathrm{Bin}(x_{t-\mathrm{p}},\alpha_{\mathrm{p}})}(x_t)\cdot F_\epsilon(x_t) \\
&\qquad\qquad - F_{\mathrm{Bin}(x_{t-1},\alpha_1)}(x_t-1)\cdots F_{\mathrm{Bin}(x_{t-\mathrm{p}},\alpha_{\mathrm{p}})}(x_t-1)\cdot F_\epsilon(x_t-1)\Big).
\end{aligned}$$

4. Conclusions

We proposed the max-BARMA(p, q) model as an integer-valued counterpart to the conventional max-ARMA(p, q) models, applicable to time series of bounded counts. This model includes both an autoregressive and a moving-average part. Basic probabilistic and statistical properties of this new class of models were discussed, namely the existence of a stationary distribution, and how observations' and innovations' distributions are related to each other. We also presented an approach for parameter estimation.

Acknowledgements

This research was supported by the German Academic Exchange Service (DAAD) and the Fundação para a Ciência e a Tecnologia (FCT), under the program "Ações Integradas Luso-Alemãs" and the Grants 57212119 and A-38/16. S. Gouveia acknowledges the postdoctoral grant by FCT (ref. SFRH/BPD/87037/2012). This work was also partially supported by the Portuguese FCT, with national (MEC) and European structural funds through the programs FEDER, under the partnership agreement PT2020 - within IEETA/UA project UID/CEC/00127/2013 (Instituto de Engenharia Electrónica e Informática de Aveiro, IEETA/UA, Aveiro) and CIDMA/UA project UID/MAT/04106/2013 (Centro de Investigação e Desenvolvimento em Matemática e Aplicações, CIDMA/UA, Aveiro).

References

Davis, R.A., Resnick, S.I. (1989) Basic properties and prediction of Max-ARMA processes. *Advances in Applied Probability* **21**(4), 781–803.

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

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

Shaked, M., Shanthikumar, J.G. (2007) *Stochastic Orders*. Springer-Verlag, New York.

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

Weiß, C.H. (2009) A new class of autoregressive models for time series of binomial counts. *Communications in Statistics — Theory and Methods* **38**(4), 447–460.

Weiß, C.H. (2013) Serial dependence of NDARMA processes. *Computational Statistics and Data Analysis* **68** (2013), 213–238.

Zucchini, W., MacDonald, I.L. (2009) *Hidden Markov Models for Time Series: An Introduction Using R*. Chapman & Hall/CRC, London.

Appendix A. Proof of Example 2.6

From

$$F_\epsilon(k) \;=\; 1-\pi+\delta_{k,n}\,\pi, \quad \mu_{\epsilon,(r)} := E[(\epsilon_t)_{(r)}] \;=\; n_{(r)}\,\pi, \quad \textstyle\sum_{k=0}^{m}\; k_{(r)}\,P(\epsilon_t=k) \;=\; \delta_{m,n}\,n_{(r)}\,\pi,$$

we obtain

$$\begin{aligned} E\big[\max\{m,\ \epsilon_t\}_{(r)}\big] \;&=\; m_{(r)}\,F_\epsilon(m) \;+\; \mu_{\epsilon,(r)} - \textstyle\sum_{k=0}^{m}\; k_{(r)}\,P(\epsilon_t=k) \\ &=\; m_{(r)}\,(1-\pi+\delta_{m,n}\,\pi) \;+\; (1-\delta_{m,n})\,n_{(r)}\,\pi \\ &=\; m_{(r)}\,(1-\pi)+n_{(r)}\,\pi \;-\; \delta_{m,n}\,(n_{(r)}-m_{(r)})\,\pi \;=\; m_{(r)}\,(1-\pi)+n_{(r)}\,\pi. \end{aligned}$$

Hence, using (2.12) together with the formula for the binomial factorial moments,

$$\begin{aligned} E[(X_t)_{(r)} \mid X_{t-1}=l] \;&=\; \textstyle\sum_{m=0}^{l}\; \binom{l}{m}\,\alpha^m\,(1-\alpha)^{l-m}\cdot\big(m_{(r)}\,(1-\pi) \;+\; n_{(r)}\,\pi\big) \\ &=\; l_{(r)}\,\alpha^r\,(1-\pi) \;+\; n_{(r)}\,\pi. \end{aligned}$$

The conditional variance becomes

$$\begin{aligned} V[X_t \mid X_{t-1}] \;&=\; E[(X_t)_{(2)} \mid X_{t-1}] + E[X_t \mid X_{t-1}] - E^2[X_t \mid X_{t-1}] \\ &=\; (X_{t-1})_{(2)}\,\alpha^2\,(1-\pi) \;+\; n_{(2)}\,\pi \;+\; X_{t-1}\,\alpha\,(1-\pi) \;+\; n\,\pi \;-\; \big(X_{t-1}\,\alpha\,(1-\pi) \;+\; n\,\pi\big)^2 \\ &=\; X_{t-1}^2\,\alpha^2\,\pi(1-\pi) \;-\; X_{t-1}\,\alpha\,(1-\pi)\,(2\,n\,\pi-1+\alpha) \;+\; n^2\,\pi(1-\pi). \end{aligned}$$

So

$$\begin{aligned} \sigma^2 := V[X_t] \;&=\; E\big[V[X_t \mid X_{t-1}]\big] \;+\; V\big[E[X_t \mid X_{t-1}]\big] \\ &=\; \alpha^2\,\pi(1-\pi)\,(\sigma^2+\mu^2) \;-\; \alpha\,(1-\pi)\,(2\,n\,\pi+\alpha)\,\mu \;+\; n^2\,\pi(1-\pi), \end{aligned}$$

which leads to the formula given in Example 2.6.