

RESEARCH

Open Access



A flexible univariate moving average time-series model for dispersed count data

Kimberly F. Sellers^{1,2*} , Ali Arab¹, Sean Melville¹ and Fanyu Cui¹

*Correspondence:

kfs7@georgetown.edu

¹Department of Mathematics and Statistics, Georgetown University, Washington, DC, USA

²Center for Statistical Research and Methodology Division, U.S. Census Bureau, Washington, DC, USA

Abstract

Al-Osh and Alzaid (1988) consider a Poisson moving average (PMA) model to describe the relation among integer-valued time series data; this model, however, is constrained by the underlying equi-dispersion assumption for count data (i.e., that the variance and the mean equal). This work instead introduces a flexible integer-valued moving average model for count data that contain over- or under-dispersion via the Conway-Maxwell-Poisson (CMP) distribution and related distributions. This first-order sum-of-Conway-Maxwell-Poissons moving average (SCMPMA(1)) model offers a generalizable construct that includes the PMA (among others) as a special case. We highlight the SCMPMA model properties and illustrate its flexibility via simulated data examples.

Keywords: Over-dispersion, Under-dispersion, Conway-Maxwell-Poisson (COM-Poisson or CMP), Sum-of-Conway-Maxwell-Poisson (sCMP)

Introduction

Integer-valued thinning-based models have been proposed to model time series data represented as counts. Al-Osh and Alzaid (1988) introduce a generally defined integer-valued moving average (INMA) process as an analog to the moving average (MA) model for continuous data which assumes an underlying Gaussian distribution. This INMA process instead utilizes a thinning operator that maintains an integer-valued range of possible outcomes. To form such a model, they consider the “survivals” of independent and identically distributed (iid) non-negative integer valued random innovations to maintain and ensure discrete data outcomes (Weiss 2021). Al-Osh and Alzaid (1988) particularly consider a first-order Poisson moving average (PMA(1)), i.e. a stationary sequence U_t of the form $U_t = \gamma \circ \epsilon_{t-1} + \epsilon_t$ where $\{\epsilon_t\}$ is a sequence of iid Poisson(η) random variables and $(\gamma \circ \epsilon) = \sum_{i=1}^{\epsilon} B_i$ for a sequence of iid Bernoulli(γ) random variables $\{B_i\}$ independent of $\{\epsilon\}$. By design, the PMA(1) is an INMA whose maximum stay time in the sequence is two time units. Consequently, components of U_t are dependent, while the components of ϵ_t and $(\gamma \circ \epsilon_{t-1})$ are independent.

Given the PMA(1) structure,

$$E(U_t) = \text{Var}(U_t) = (1 + \gamma)\eta, \quad (1)$$

and the covariance of consecutive variables $Cov(U_{t-1}, U_t) = \gamma\eta$; this implies that the correlation is

$$\rho_U(r) = Corr(U_{t-r}, U_t) = \begin{cases} \frac{\gamma}{1+\gamma} & r = 1 \\ 0 & r > 1. \end{cases} \tag{2}$$

Meanwhile, the probability generating function (pgf) of U_t is $\Phi_{U_t}(u) = e^{-\eta(1+\gamma)(1-u)}$, the joint pgf of $\{U_1, \dots, U_r\}$ is $\Phi_r(u_1, \dots, u_r) = \exp(-\eta[r + \gamma - (1-\gamma)\sum_{i=1}^r u_i - \gamma(u_1 + u_r) - \gamma\sum_{i=1}^{r-1} u_i u_{i+1}])$ (which infers that time reversibility holds for the PMA), and the pgf of $T_{U,r} = \sum_{i=1}^r U_i$ is

$$\Phi_{T_{U,r}}(u) = \exp(-\eta[(1-\gamma)r + 2\gamma](1-u) - \eta\gamma(r-1)(1-u^2)).$$

Al-Osh and Alzaid (1988) note that $T_{U,r}$ does not have a Poisson distribution, which is in contrast to the standard MA(1) process. The conditional mean and variance of U_{t+1} given $U_t = u$ are both linear in U_t , namely

$$E(U_{t+1} | U_t = u) = \eta + \gamma u / (1 + \gamma), \text{ and} \tag{3}$$

$$Var(U_{t+1} | U_t = u) = \eta + \gamma u / (1 + \gamma)^2. \tag{4}$$

The PMA is a natural choice for modeling an integer-valued process, in part because of its tractability (Al-Osh and Alzaid 1988). This model, however, is limited by its constraining equi-dispersion property, i.e. the assumption that the mean and variance of the underlying process equal. Real data do not generally conform to this construct (Hilbe 2014; Weiss 2018); they usually display over-dispersion relative to the Poisson model (i.e. where the variance is greater than the mean), however integer-valued data are surfacing with greater frequency that express data under-dispersion relative to Poisson (i.e. the variance is less than the mean). Accordingly, it would be fruitful to instead consider a flexible time series model that can accommodate data over- and/or under-dispersion.

Alzaid and Al-Osh (1993) introduce a first-order generalized Poisson moving average (GPMA(1)) process as an alternative to the PMA. The associated model has the form,

$$W_t = Q_t^*(\epsilon_{t-1}^*) + \epsilon_t^*, \quad t = 0, 1, 2, \dots, \tag{5}$$

where $\{\epsilon_t^*\}$ is a sequence of iid generalized Poisson $GP(\mu^*, \theta)$, and $\{Q_t^*(\cdot)\}$ is a sequence of quasi-binomial $QB(p^*, \theta/\mu^*, \cdot)$ random operators independent of $\{\epsilon_t^*\}$. As with the PMA, W_{t+r} and W_t are independent for $|r| > 1$. The marginal distribution of W_t is $GP((1+p^*)\mu^*, \theta)$. Recognizing the relationship between moving average and autoregressive models, Alzaid and Al-Osh (1993) equate terms in this GPMA(1) model to their first-order generalized Poisson autoregressive (GPAR(1)) counterpart,

$$W_t = Q_t(W_{t-1}) + \epsilon_t, \quad t = 0, 1, 2, \dots, \tag{6}$$

where $\{\epsilon_t\}$ is a sequence of iid $GP(q\mu, \theta)$ random variables where $q = 1 - p$, and $\{Q_t(\cdot)\}$ is a sequence of $QB(p, \theta/\mu, \cdot)$ random operators, independent of $\{\epsilon_t\}$; i.e., they let $\mu = (1+p^*)\mu^*$ and $p = \frac{p^*}{1+p^*}$. The bivariate pgf of W_{t+1} and W_t can thus be represented as

$$\Phi_{W_{t+1}, W_t}(u_1, u_2) = \exp[\mu^*(A_\theta(u_1) + A_\theta(u_2) - 2) + \mu^*p^*(A_\theta(u_1u_2) - 1)] \tag{7}$$

$$= \exp[\mu q(A_\theta(u_1) + A_\theta(u_2) - 2) + \mu p(A_\theta(u_1u_2) - 1)], \tag{8}$$

where $A_\theta(s)$ is the inverse function that satisfies $A_\theta(se^{-\theta(s-1)}) = s$; see Alzaid and Al-Osh (1993). This substitution in Eq. (7) to obtain Eq. (8) further illustrates the relationship

between the GPMA(1) and GPAR(1) models such that they have the same joint pgf. Eq. (8) and the related GPAR work of Alzaid and Al-Osh (1993) therefore show that

$$E(W_t | W_{t-1} = w) = pw + \frac{q\mu}{1-\theta}. \quad (9)$$

The joint pgf of $(W_t, W_{t-1}, \dots, W_{t-r+1})$ is given by

$$\Phi(u_1, \dots, u_r) = \exp \left[\mu q \sum_{i=1}^r (A_\theta(u_i) - 1) + \mu p \sum_{i=1}^r (A_\theta(u_i u_{i+1}) - 1) \right]. \quad (10)$$

From the joint pgf, we see that the GPMA(1) is also time-reversible, because it has the same dynamics if time is reversed. Further, the pgf associated with the total counts occurring during time lag r (i.e. $T_{W,r} = \sum_{i=1}^r W_{t-r+i}$) is $\Phi_{T_{W,r}}(u) = \exp [\mu q r (A_\theta(u) - 1) + \mu p (r-1) (A_\theta(u^2) - 1)]$. Alzaid and Al-Osh (1993) note that this result extends the analogous PMA result to the broader GPMA(1) model. Finally, the GPMA autocorrelation function is

$$\rho_W(r) = \text{Corr}(W_t, W_{t+r}) = \begin{cases} p & |r| = 1 \\ 0 & |r| > 1, \end{cases}$$

where $p = p^*(1 + p^*)^{-1}$; by definition, $\rho_W(r) \in [0, 0.5]$ (Alzaid and Al-Osh 1993).

Even though the GPMA can be considered to model over- or under-dispersed count time series, it may not be a viable option for count data that express extreme under-dispersion; see, e.g. Famoye (1993). This work instead introduces another alternative for modeling integer-valued time series data. The subsequent writing proceeds as follows. We first provide background regarding the probability distributions that motivate the development of our flexible INMA model. Then, we introduce the SCMPMA(1) model to the reader and discuss its statistical properties. The subsequent section illustrates the model flexibility through simulated and real data examples. Finally, the manuscript concludes with discussion.

Motivating distributions

While the above constructs show increased ability and improvement towards modeling integer-valued time series data with various forms of dispersion, each of the models suffers from respective limitations. In order to develop and describe our SCMPMA(1), we first introduce its underlying motivating distributions: the CMP distribution and its generalized sum-of-CMPs distribution (sCMP), as well as the Conway-Maxwell-Binomial (CMB) along with a generalized CMB (gCMB) distribution.

The Conway-Maxwell-Poisson distribution and its generalization

The Conway-Maxwell-Poisson (CMP) distribution (introduced by Conway and Maxwell (1962), and revived by Shmueli et al. (2005)) is a viable count distribution that generalizes the Poisson distribution in light of potential data dispersion. The CMP probability mass function (pmf) takes the form

$$P(X = x | \lambda, \nu) = \frac{\lambda^x}{(x!)^\nu \zeta(\lambda, \nu)}, \quad x = 0, 1, 2, \dots, \quad (11)$$

for a random variable X , where $\lambda = E(X^\nu) \geq 0$, $\nu \geq 0$ is the associated dispersion parameter, and $\zeta(\lambda, \nu) = \sum_{s=0}^{\infty} \frac{\lambda^s}{(s!)^\nu}$ is the normalizing constant. The CMP distribution includes

three well-known distributions as special cases, namely the Poisson ($\nu = 1$), geometric ($\nu = 0, \lambda < 1$), and Bernoulli ($\nu \rightarrow \infty$ with probability $\frac{\lambda}{1+\lambda}$) distributions.

The associated pgf of X is $\Phi_X(u) = E(u^X) = \frac{\zeta(\lambda u, \nu)}{\zeta(\lambda, \nu)}$, and its moment generating function (mgf) is $M_X(u) = E(e^{Xu}) = \frac{\zeta(\lambda e^u, \nu)}{\zeta(\lambda, \nu)}$. The moments can meanwhile be represented recursively as

$$E(X^{g+1}) = \begin{cases} \lambda [E(X+1)]^{1-\nu}, & g = 0 \\ \lambda \frac{\partial}{\partial \lambda} E(X^g) + E(X)E(X^g), & g > 0. \end{cases} \quad (12)$$

In particular, the expected value and variance can be written in the form and approximated respectively as

$$E(X) = \frac{\partial \ln \zeta(\lambda, \nu)}{\partial \ln \lambda} \approx \lambda^{1/\nu} - \frac{\nu-1}{2\nu}, \quad (13)$$

$$\text{Var}(X) = \frac{\partial E(X)}{\partial \ln \lambda} \approx \frac{1}{\nu} \lambda^{1/\nu}, \quad (14)$$

where the approximations are especially good for $\nu \leq 1$ or $\lambda > 10^\nu$ (Shmueli et al. 2005). This distribution is a member of the exponential family, where the joint pmf of the random sample $\mathbf{x} = (x_1, \dots, x_N)$ is

$$P(\mathbf{x} \mid \lambda, \nu) = \frac{\prod_{i=1}^N \lambda^{x_i}}{\prod_{i=1}^N x_i!} \cdot \zeta^{-N}(\lambda, \nu) = \lambda^{S_1} \exp(-\nu S_2) \zeta^{-N}(\lambda, \nu),$$

where $S_1 = \sum_{i=1}^N x_i$ and $S_2 = \sum_{i=1}^N \log(x_i!)$ are joint sufficient statistics for λ and ν . Further, because the CMP distribution belongs to the exponential family, the conjugate prior distribution has the form, $h(\lambda, \nu) = \lambda^{a-1} e^{-\nu b} \zeta^{-c}(\lambda, \nu) \delta(a, b, c)$, where $\lambda > 0$, $\nu \geq 0$, and $\delta(a, b, c)$ is a normalizing constant such that $\delta^{-1}(a, b, c) = \int_0^\infty \int_0^\infty \lambda^{a-1} e^{-\nu b} \zeta^{-c}(\lambda, \nu) d\lambda d\nu < \infty$.

Meanwhile, letting $X_* = \sum_{i=1}^n X_i$ for iid random variables $X_i \sim \text{CMP}(\lambda, \nu), i = 1, \dots, n$, we say that X_* is distributed as a sum-of-CMPs [denoted sCMP(λ, ν, n)] variable, and has the pmf

$$P(X_* = x_*) = \frac{\lambda^{x_*}}{(x_*!)^\nu \zeta^n(\lambda, \nu)} \sum_{\substack{a_1, \dots, a_n=0 \\ a_1 + \dots + a_n = x_*}}^{x_*} \binom{x_*}{a_1, \dots, a_n}^\nu, \quad x_* = 0, 1, 2, \dots,$$

where $\zeta^n(\lambda, \nu)$ is the n th power of $\zeta(\lambda, \nu)$, and $\binom{x_*}{a_1, \dots, a_n} = \frac{x_*!}{a_1! \dots a_n!}$ is a multinomial coefficient. The sCMP(λ, ν, n) distribution encompasses the Poisson distribution with rate parameter $n\lambda$ (for $\nu = 1$), negative binomial($n, 1 - \lambda$) distribution (for $\nu = 0$ and $\lambda < 1$), and Binomial(n, p) distribution (as $\nu \rightarrow \infty$ with success probability $p = \frac{\lambda}{\lambda+1}$) as special cases. Further, for $n = 1$, the sCMP($\lambda, \nu, n = 1$) is simply the CMP(λ, ν) distribution.

The mgf and pgf for a sCMP(λ, ν, n) random variable X_* are

$$M_{X_*}(t) = \left(\frac{\zeta(\lambda e^t, \nu)}{\zeta(\lambda, \nu)} \right)^n \quad \text{and} \quad \Phi_{X_*}(t) = \left(\frac{\zeta(\lambda t, \nu)}{\zeta(\lambda, \nu)} \right)^n,$$

respectively; accordingly, the sCMP(λ, ν) has mean $E(X_*) = nE(X)$ and variance $V(X_*) = nV(X)$, where $E(X)$ and $V(X)$ are defined in Eqs. (13)-(14), respectively. Invariance under addition holds for two independent sCMP distributions with the same rate and dispersion parameters. See Sellers et al. (2017) for additional information regarding the sCMP distribution.

The Conway-Maxwell-Binomial distribution and its generalization

The Conway-Maxwell-Binomial distribution of Kadane (2016) (also known as the Conway-Maxwell-Poisson-Binomial distribution by Borges et al. (2014)) is a three-parameter generalization of the Binomial distribution. Denoted as $CMB(d, p, \nu)$ distributed, its pmf is

$$P(Y = y) = \frac{\binom{d}{y}^\nu p^y (1-p)^{d-y}}{\chi(p, \nu, d)}, \quad y = 0, \dots, d \tag{15}$$

for some random variable Y where $0 \leq p \leq 1$, $\nu \in \mathbb{R}$, and $\chi(p, \nu, d) = \sum_{y=0}^d \binom{d}{y}^\nu p^y (1-p)^{d-y}$ is the associated normalizing constant. The Binomial(d, p) distribution is the special case of the $CMB(d, p, \nu)$ where $\nu = 1$. Meanwhile, $\nu > (<)1$ corresponds to under-dispersion (over-dispersion) relative to the Binomial distribution. For $\nu \rightarrow \infty$, the pmf is concentrated on the point dp while, for $\nu \rightarrow -\infty$, the pmf is concentrated at 0 or d . For independent $X_i \sim CMP(\lambda_i, \nu)$, $i = 1, 2$, the conditional distribution of X_1 given that $X_1 + X_2 = d$ has a $CMB\left(d, \frac{\lambda_1}{\lambda_1 + \lambda_2}, \nu\right)$ distribution.

The pgf and mgf of Y have the form,

$$\Phi_Y(u) = E(u^Y) = \frac{\tau\left(\frac{up}{1-p}, \nu, d\right)}{\tau\left(\frac{p}{1-p}, \nu, d\right)} \quad \text{and} \quad M_Y(u) = \frac{\tau\left(\frac{pe^{u\theta}}{1-p}, \nu, d\right)}{\tau\left(\frac{p}{1-p}, \nu, d\right)}, \tag{16}$$

respectively, where $\tau(\theta_*, \nu, d) = \sum_{y=0}^d \binom{d}{y}^\nu \theta_*^y$ for some θ_* . The CMB distribution is a member of the exponential family whose joint pmf of the random sample $\mathbf{y} = \{y_1, \dots, y_N\}$ is

$$\begin{aligned} P(\mathbf{y} | p, \nu) &\propto (1-p)^{dN} \prod_{i=1}^N \left(\frac{p}{1-p}\right)^{y_i} \frac{d^{N\nu}}{[y_i! (d-y_i)!]^\nu} \\ &\propto \exp\left(S_{*1} \log\left(\frac{p}{1-p}\right) - \nu S_{*2}\right), \end{aligned}$$

where $S_{*1} = \sum_{i=1}^N y_i$ and $S_{*2} = \sum_{i=1}^N \log[y_i! (d-y_i)!]$ are the joint sufficient statistics for p and ν . Further, its existence as a member of the exponential family implies that a conjugate prior family exists of the form,

$$h(\theta_*, \nu) = \theta_*^{a-1} e^{-\nu b} \omega^{-c}(\theta_*, \nu) \psi(a, b, c), \quad 0 < \theta_* < \infty, 0 < \nu < \infty,$$

where $\omega(\theta_*, \nu) = \sum_{y=0}^d \theta_*^y / [y! (d-y)!]^\nu$, $\psi^{-1}(a, b, c) = \int_0^\infty \int_0^\infty \theta_*^{a-1} e^{-\nu b} \omega^{-c}(\theta_*, \nu) d\theta_* d\nu < \infty$ (Kadane 2016).

Sellers et al. (2017) further introduce a generalized Conway-Maxwell-Binomial (gCMB) distribution whose pmf is

$$P(Z = z) \propto \binom{s}{z}^\nu p^z (1-p)^{s-z} \left[\sum_{\substack{a_1, \dots, a_{n_1}=0 \\ a_1 + \dots + a_{n_1} = z}} \binom{z}{a_1, \dots, a_{n_1}}^\nu \right] \left[\sum_{\substack{b_1, \dots, b_{n_2}=0 \\ b_1 + \dots + b_{n_2} = s-z}} \binom{s-z}{b_1, \dots, b_{n_2}}^\nu \right] \tag{17}$$

for a random variable Z with parameters (p, ν, s, n_1, n_2) . As with the conditional probability of a CMP random variable given the sum of it and another independent CMP random variable sharing the same dispersion parameter, a special case of a gCMB distribution can be derived as the conditional distribution of X_{*1} , given the sum $X_{*1} + X_{*2} = d$ for

independent sCMP random variables, $X_{*i} \sim \text{sCMP}(\lambda_i, \nu, n_i)$, $i = 1, 2$; the resulting distribution is analogously a gCMB $\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}, \nu, d, n_1, n_2\right)$ distribution. The gCMB distribution contains several special cases, including the CMB (d, p, ν) distribution (for $n_1 = n_2 = 1$); the Binomial (d, p) distribution (when $n_1 = n_2 = 1$ and $\nu = 1$); and, for $\lambda_1 = \lambda_2 = \lambda$, the hypergeometric distribution when $\nu \rightarrow \infty$ and the negative hypergeometric distribution when $\nu = 0$ and $\lambda < 1$.

First-order sCMP time series models

This section highlights two first-order models for discrete time series data that have a sCMP marginal distribution, namely the first-order sCMP autoregressive (SCMPAR(1)) model, and a first-order SCMP moving average (SCMPMA(1)) model with the same marginal distribution structure.

First-order sCMP autoregressive (SCMPAR(1)) model

Sellers et al. (2020) introduce a first-order sCMP autoregressive (SCMPAR(1)) model to describe count data correlated in time that express over- or under-dispersion. Based on the sCMP and gCMB distributions, respectively (as described in the “Motivating distributions” section with more detail available in Sellers et al. (2017)), we use the sCMP distribution to model the marginals of the first-order integer-valued autoregressive (INAR(1)) process as

$$X_t = C_t(X_{t-1}) + \epsilon_t \quad t = 1, 2, \dots, \tag{18}$$

where $\epsilon_t \sim \text{sCMP}(\lambda, \nu, n_2)$, and $\{C_t(\bullet) : t = 1, 2, \dots\}$ is a sequence of independent gCMB $\left(\frac{1}{2}, \nu, \bullet, n_1, n_2\right)$ operators, independent of $\{\epsilon_t\}$. This flexible INAR(1) model contains the first-order Poisson autoregressive (PAR(1)) as described in several references (Al-Osh and Alzaid 1987; McKenzie 1988; Weiss 2008), and the first-order binomial autoregressive model of Al-Osh and Alzaid (1991) as special cases. It likewise contains an INAR(1) model that allows for negative binomial marginals with a thinning operator whose pmf is negative hypergeometric.

The SCMPAR(1) model is yet another special case of the infinitely divisible convolution-closed class of first-order autoregressive (AR(1)) models described in Joe (1996), and satisfies the Markov property with the transition probability,

$$\begin{aligned}
 P(X_t | X_{t-1}) = & \sum_{k=0}^{\min(x_t, x_{t-1})} \frac{\binom{x_{t-1}}{k}^\nu \left[\sum_{\substack{a_1, \dots, a_{n_1}=0 \\ a_1 + \dots + a_{n_1} = k}}^k \binom{k}{a_1, \dots, a_{n_1}}^\nu \right] \left[\sum_{\substack{b_1, \dots, b_{n_2}=0 \\ b_1 + \dots + b_{n_2} = x_{t-1} - k}}^{x_{t-1} - k} \binom{x_{t-1} - k}{b_1, \dots, b_{n_2}}^\nu \right]}{\sum_{\substack{c_1, \dots, c_{n_1+n_2}=0 \\ c_1 + \dots + c_{n_1+n_2} = x_{t-1}}}^{x_{t-1}} \binom{x_{t-1}}{c_1, \dots, c_{n_1+n_2}}^\nu} \\
 & \times \frac{\lambda^{x_t - k}}{[(x_t - k)!]^\nu Z^{n_2}(\lambda, \nu)} \sum_{\substack{d_1, \dots, d_{n_2}=0 \\ d_1 + \dots + d_{n_2} = x_t - k}}^{x_t - k} \binom{x_t - k}{d_1, \dots, d_{n_2}}^\nu.
 \end{aligned} \tag{19}$$

The SCMPAR(1) model has an ergodic Markov chain, thus X_t has a stationary sCMP $(\lambda, \nu, n_1 + n_2)$ distribution that is unique. The joint pgf associated with the SCMPAR(1) model is

$$\phi_{X_{t+1}, X_t}(u, l) = \frac{\zeta^{n_2}(\lambda u, \nu) \zeta^{n_1}(\lambda ul, \nu) \zeta^{n_2}(\lambda l, \nu)}{\zeta^{n_2}(\lambda, \nu) \zeta^{n_1}(\lambda, \nu) \zeta^{n_2}(\lambda, \nu)} = \frac{(\zeta(\lambda u, \nu)\zeta(\lambda l, \nu))^{n_2} \zeta^{n_1}(\lambda ul, \nu)}{\zeta^{n_1+2n_2}(\lambda, \nu)}, \quad (20)$$

where the pgf is symmetric in u and l , and hence the joint distribution of X_{t+1} and X_t is time reversible. The regression form for the SCMPAR(1) process can be determined, and the general autocorrelation function for the process $\{X_t\}$ is $\rho_r = \text{Corr}(X_t, X_{t-r}) = \left(\frac{n_1}{n_1+n_2}\right)^r$ for $r = 0, 1, 2, \dots$. Parameter estimation can be conducted via conditional maximum likelihood with statistical computation tools (e.g. in R); see Sellers et al. (2020) for details.

Introducing the sCMPMA(1) model

Motivated by the SCMPAR(1) model of Sellers et al. (2020), we introduce a first-order sum-of-CMPs moving average (SCMPMA(1)) process X_t by

$$X_t = C_t^*(\epsilon_{t-1}^*) + \epsilon_t^*, \quad t = 1, 2, \dots, \quad (21)$$

where ϵ_t^* is a sequence of iid sCMP($\lambda, \nu, m_1 + m_2$) random variables and $C_t^*(\bullet)$ is a sequence of independent gCMB($1/2, \nu, \bullet, m_1, m_2$) operators independent of ϵ_t^* . By definition, X_t is a stationary process with the sCMP($\lambda, \nu, 2m_1 + m_2$) distribution, and X_{t+r} and X_t are independent for $|r| > 1$. While this model can analogously be viewed as a special case of the infinitely divisible convolution-closed class of discrete MA models (Joe 1996), unlike the sCMPAR(1) process, the sCMPMA(1) process is not Markovian.

The autocorrelation between X_t and X_{t+1} is

$$\begin{aligned} \rho_1 = \text{Corr}(X_t, X_{t+1}) &= \frac{\text{Cov}(C_t^*(\epsilon_{t-1}^*) + \epsilon_t^*, C_{t+1}^*(\epsilon_t^*) + \epsilon_{t+1}^*)}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+1})}} \\ &= \frac{\text{Cov}(\epsilon_t^*, C_{t+1}^*(\epsilon_t^*))}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+1})}} \text{ by the independence assumptions,} \end{aligned}$$

where $C_{t+1}^*(\epsilon_t^*) = \sum_{i=1}^{m_1} Y_i$ and $\epsilon_t^* = \sum_{i=1}^{m_1+m_2} Y_i$, respectively, are sCMP(λ, ν, m_1) and sCMP($\lambda, \nu, m_1 + m_2$) random variables; i.e. each sCMP random variable can be viewed as respective sums of iid CMP(λ, ν) random variables, Y_i . Thus,

$$\text{Cov}(\epsilon_t^*, C_{t+1}^*(\epsilon_t^*)) = \text{Cov}\left(\sum_{i=1}^{m_1+m_2} Y_i, \sum_{i=1}^{m_1} Y_i\right) = \text{Var}\left(\sum_{i=1}^{m_1} Y_i\right) = m_1 \text{Var}(Y),$$

where, without loss of generality, we let Y denote any of the iid Y_i random variables. Meanwhile, because $\{X_t\}$ is a sCMP($\lambda, \nu, 2m_1 + m_2$) distributed stationary process, we can likewise represent $\text{Var}(X_t) = \text{Var}\left(\sum_{i=1}^{2m_1+m_2} Y_i\right) = \sum_{i=1}^{2m_1+m_2} \text{Var}(Y_i) = (2m_1 + m_2)\text{Var}(Y)$ for all t . We therefore find that

$$\rho_1 = \text{Corr}(X_t, X_{t+1}) = \frac{\text{Cov}(\epsilon_t^*, C_{t+1}^*(\epsilon_t^*))}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+1})}} = \frac{m_1 \text{Var}(Y)}{(2m_1 + m_2)\text{Var}(Y)} = \frac{m_1}{2m_1 + m_2}. \quad (22)$$

Because $m_1, m_2 \geq 1$, the one-step range of possible correlation values is $0 \leq \rho_1 \leq 0.5$. In particular, for $m_1 = m_2$, we have the special case where $\rho_1 = 1/3$. Meanwhile, $\rho_k = 0$ for all $k > 1$ because, by definition of the SCMPMA(1) model assumptions, there is no dependent structure between X_t and X_{t+r} for $r > 1$.

Recall from the “[The Conway-Maxwell-Poisson distribution and its generalization](#)” section that $\Phi_G(w) = \left(\frac{\zeta(\lambda w, \nu)}{\zeta(\lambda, \nu)}\right)^\pi$ is the pgf for a sCMP(λ, ν, π) distributed random variable, (say) G . Using this knowledge along with Eq. (21), the joint pgf can be derived as

$$\begin{aligned}
 \phi_{X_{t+1}, X_t}(u, l) &= E\left(u^{X_{t+1}} l^{X_t}\right) \\
 &= E\left(u^{C^*(\epsilon_t^*) + \epsilon_{t+1}^*} l^{X_t}\right) \\
 &= E\left(u^{C^*(\epsilon_t^*) + \epsilon_{t+1}^*} l^{X_t - C^*(\epsilon_t^*)} l^{C^*(\epsilon_t^*)}\right) \\
 &= E\left((ul)^{C^*(\epsilon_t^*)} u^{\epsilon_{t+1}^*} l^{X_t - C^*(\epsilon_t^*)}\right) \text{ where } X_t - C^*(\epsilon_t^*) \stackrel{d}{=} \epsilon_t^* \\
 &= E\left((ul)^{C^*(\epsilon_t^*)}\right) E\left(u^{\epsilon_{t+1}^*}\right) E\left(l^{\epsilon_t^*}\right) \text{ by independence} \\
 &= \phi_{C^*(\epsilon_t^*)}(ul) \phi_{\epsilon_{t+1}^*}(u) \phi_{\epsilon_t^*}(l) \\
 &= \left(\frac{\zeta(\lambda ul, \nu)}{\zeta(\lambda, \nu)}\right)^{m_1} \left(\frac{\zeta(\lambda u, \nu)}{\zeta(\lambda, \nu)}\right)^{m_1 + m_2} \left(\frac{\zeta(\lambda l, \nu)}{\zeta(\lambda, \nu)}\right)^{m_1 + m_2} \\
 &= \frac{(\zeta(\lambda u, \nu) \zeta(\lambda l, \nu))^{m_1 + m_2} (\zeta(\lambda ul, \nu))^{m_1}}{(\zeta(\lambda, \nu))^{3m_1 + 2m_2}}, \tag{23}
 \end{aligned}$$

where Eq. (23) is equivalent to Eq. (20) (i.e. the SCMPMA(1) process is comparable to the SCMPAR(1) process) when $m_1 = n_1 = n_2 - m_2$. Given this comparison, we can easily determine the conditional mean $E(X_{t+1} \mid X_t = x)$ and conditional variance $\text{Var}(X_{t+1} \mid X_t = x)$. Eq. (23) further demonstrates that the SCMPMA(1) model is time-reversible.

Parameter estimation via maximum likelihood (ML) is a difficult task with INMA models given the complex form of the underlying distributions. Even a conditional least squares approach does not appear to be feasible “because of the thinning operators, unless randomization is used” (Brännäs and Hall 2001). We therefore instead consider the following ad hoc procedure for parameter estimation. Given a data set with an observed correlation ρ_1 , we first propose values for $m_1, m_2 \in \mathbb{N}$ that satisfy the constraint, $\rho_1 \approx \frac{m_1}{2m_1 + m_2}$. Given m_1 and m_2 and recognizing that X_t is stationary with a sCMP($\lambda, \nu, 2m_1 + m_2$) distribution, we proceed with ML estimation to determine $\hat{\lambda}$ and $\hat{\nu}$ as described in Zhu et al. (2017) for conducting sCMP($\lambda, \nu, s = 2m_1 + m_2$) parameter estimation with regard to a CMP process over an interval of length $s \geq 1$. The corresponding variation for $\hat{\lambda}$ and $\hat{\nu}$ can be quantified via the Fisher information matrix or nonparametric bootstrapping. While the sampling distribution for $\hat{\lambda}$ is approximately symmetric, the sampling distribution for $\hat{\nu}$ is considerably right-skewed, hence analysts are advised to quantify estimator variation via nonparametric bootstrapping. While this is a means to an end, it only achieves in determining an appropriate distributional form regarding the data; it does not fully address the nature of the time series.

Data examples

To illustrate the flexibility of our INMA model, we consider various data simulations and a real data example. Below contains the respective details and associated commentary.

Simulated data examples

Table 1 reports the estimated mean, variance, and autocorrelation that result from

Table 1 Estimated mean, variance, and autocorrelation for various SCMPMA(1) data simulations of length 10,000 given parameters, (λ, ν, m_1, m_2) ; $\lambda = 0.5$ for all simulations

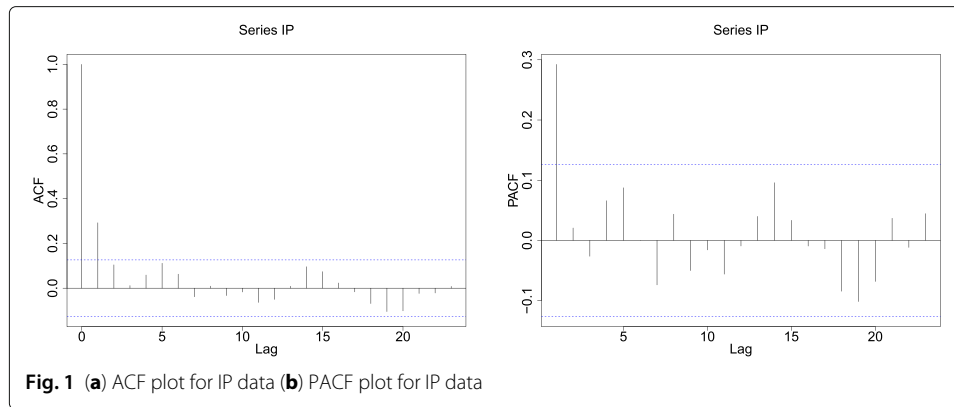
m_1	m_2	ν	Est. Mean	True Mean	Est. Var.	True Var.	$\hat{\rho}$	ρ
1	1	0	3.0073	3.00	5.9076	6.00	0.333	0.333
		0.5	1.8846		2.3395		0.335	0.333
		1	1.4842	1.50	1.4507	1.50	0.328	0.333
		2	1.2359		1.0358		0.338	0.333
		35	0.9826	1.00	0.6568	0.67	0.333	0.333
1	2	0	4.0027	4.00	8.0309	8.00	0.248	0.250
		0.5	2.4891		3.0010		0.253	0.250
		1	1.9861	2.00	1.9755	2.00	0.252	0.250
		2	1.6427		1.3284		0.257	0.250
		35	1.3355	1.33	0.8964	0.89	0.254	0.250
2	1	0	5.0408	5.00	10.2256	10.00	0.404	0.400
		0.5	3.1414		3.9010		0.402	0.400
		1	2.4823	2.50	2.4689	2.50	0.396	0.400
		2	2.0094		1.6561		0.401	0.400
		35	1.6583	1.67	1.1135	1.11	0.390	0.400
2	2	0	6.0019	6.00	12.2343	12.00	0.331	0.333
		0.5	3.7195		4.5267		0.338	0.333
		1	2.9873	3.00	3.0130	3.00	0.329	0.333
		2	2.4488		2.0178		0.336	0.333
		35	1.9861	2.00	1.2964	1.33	0.326	0.333

For the special cases (i.e. $\nu = 0, 1, 35$, where $\nu = 35$ sufficiently represents performance as $\nu \rightarrow \infty$), we likewise provide the expected/true mean and variance. Along with the estimated autocorrelation $\hat{\rho}$ that results from the data, the table reports the true autocorrelation, $\rho_1 = \frac{m_1}{2m_1+m_2}$, for all $\{m_1, m_2\}$ and any ν (Eq. (22))

various data simulations of SCMPMA(1) data given parameters (λ, ν, m_1, m_2) . In all examples, we let $\lambda = 0.5$, $m_1, m_2 \in \{1, 2\}$, and $\nu = \{0, 0.5, 1, 2, 35\}$, where $\nu = 0$ captures the case of extreme over-dispersion, $\nu = 1$ denotes equi-dispersion, and $\nu = 35$ sufficiently illustrates the case computationally of utmost under-dispersion where $\nu \rightarrow \infty$.

For all examples, we find that the associated mean and variance compare with each other as expected, i.e. the variance is greater than the mean when $\nu < 1$ (i.e. the data are over-dispersed), the variance and mean are approximately equal when $\nu = 1$ (i.e. equi-dispersion holds), and the variance is less than the mean (i.e. the data are under-dispersed) when $\nu > 1$. In particular, we can easily verify that the three special case models perform as expected. For the Poisson cases ($\nu = 1$), we expect the mean and variance to both equal $(2m_1 + m_2)\lambda$, while the binomial cases (i.e. $\nu \rightarrow \infty$ and $p = \frac{\lambda}{\lambda+1}$) produce a mean equal to $(2m_1 + m_2)\frac{\lambda}{\lambda+1}$ and variance equaling $(2m_1 + m_2)\frac{\lambda}{\lambda+1}\left(1 - \frac{\lambda}{\lambda+1}\right)$, and the negative binomial cases ($\nu = 0$ with $p = 1 - \lambda$) have a mean of $\frac{(2m_1+m_2)\lambda}{1-\lambda}$ and variance equaling $\frac{(2m_1+m_2)\lambda}{(1-\lambda)^2}$. In fact, even with the $\nu \rightarrow \infty$ case approximated by letting $\nu = 35$, we still obtain reasonable estimates for the mean and variance for all of the associated cases of m_1 and m_2 .

For each $\{m_1, m_2\}$ pair, the mean and variance both decrease as ν increases while, for all of the considered examples, we obtain estimated correlation values $\hat{\rho}$ that approximately equal the true correlation, ρ . In particular, for those cases where $m_1 = m_2$, we obtain $\hat{\rho} \approx 1/3$ as expected (see Eq. (22)).



Real data example: IP address counts

Weiss (2007) considers a modified dataset regarding the number of unique IP-addresses which access the University of Würzburg Department of Statistics’s webpages in 240 two-minute intervals. Collected on November 29, 2005 (from 10:00:00 to 18:00:00), these data have an associated mean and variance equaling 1.286 and 1.205, respectively. Weiss (2007) considers a PAR(1) model, noting that “the empirical partial autocorrelation function indicates that a first order [autoregressive] model may be an appropriate choice” with $\hat{\rho}_1 = 0.292$; Sellers et al. (2020), following suit, consider a SCMPAR(1) model as a flexible alternative to the PAR(1) model. The ACF and PACF plots of these data, however, do not clearly distinguish between considering a first-order autoregressive or a moving average model; see Fig. 1a-b. Further, recognizing that the data express apparent under-to equi-dispersion, we therefore consider the SCMPMA(1) as an illustrative model for analysis.

We perform ML estimation assuming various combinations for (m_1, m_2) (i.e. $\{(1,1), (1,2), (2,2)\}$) as these values contain the observed correlation, $0.25 = \frac{1}{4} < \hat{\rho}_1 < \frac{1}{3} \approx 0.33$. Table 2 contains the resulting parameter estimates for λ and ν , along with the respective Akaike Information Criterion (AIC). While the SCMPMA(1) model with $m_1 = m_2 = 2$ has the lowest AIC among the four models considered, all of these models produce approximately equal AIC values (i.e. 695.2) where the increasing m_1 and m_2 values associate with decreasing $\hat{\lambda}$ and increasing $\hat{\nu}$. This makes sense because the resulting estimates rely solely on the assumed underlying sCMP($\lambda, \nu, 2m_1 + m_2$) distributional form for the data.

The dispersion estimates in Table 2 are all greater than 1, thus implying a perceived level of data under-dispersion. These results naturally stem from the reported mean of the data (1.286) being greater than its corresponding variance (1.205). Their associated

Table 2 Estimated parameters, the 95% confidence intervals for λ and ν derived from nonparametric bootstrapping, and Akaike Information Criterion (AIC) values for various SCMPMA(1) models for the IP data

m_1	m_2	λ		ν		AIC
		Est.	95% CI	Est.	95% CI	
1	1	0.461	(0.374, 0.585)	1.285	(0.650, 2.291)	695.24
1	2	0.346	(0.278, 0.433)	1.387	(0.582, 2.937)	695.21
2	1	0.277	(0.225, 0.348)	1.493	(0.500, 4.594)	695.20
2	2	0.231	(0.188, 0.289)	1.612	(0.429, 25.802)	695.19

95% confidence intervals (determined via nonparametric bootstrapping; also supplied in Table 2), however, are sufficiently large such that they contain $\nu = 1$. This suggests that the apparent data under-dispersion is not statistically significant, thus instead suggesting that the data can be analyzed via the Al-Osh and Alzaid (1988) PMA(1) model. It is further striking to see that the respective 95% confidence intervals associated with the dispersion parameter increase with the size of the underlying sCMP($2m_1 + m_2$) model. This is an artifact of the (s)CMP distribution, namely that the distribution of ν is a right-skewed distribution (as discussed in Zhu et al. (2017)). This approach confirms interest in the PMA(1) model where Eqs. (1)-(2) imply that associated estimated parameters are $\hat{\gamma} \approx 0.4124$ and $\hat{\eta} \approx 0.9105$. Thus, we benefit from the SCMPMA(1) as a tool for parsimonious model determination.

Discussion

This work utilizes the sCMP distribution of Sellers et al. (2017) to develop a SCMPMA(1) model that serves as a flexible moving average time series model for discrete data where data dispersion is present. The SCMPMA(1) model captures the PMA(1), as well as versions of a negative binomial and binomial MA(1) structure, respectively, as special cases. This along with the flexible SCMPAR(1) can be used further to derive broader auto-regressive moving average (ARMA) and auto-regressive integrated moving average (ARIMA) models based on the sCMP distribution.

The SCMPMA(1) shares many properties with the analogous SCMPAR(1) model by Sellers et al. (2020). The presented models rely on predefining discrete values (i.e. m_1, m_2 for the SCMPMA(1)) for parameter estimation. As done in Sellers et al. (2017) and Sellers and Young (2019), we utilize a profile likelihood approach where, given m_1 and m_2 , we estimate the remaining model coefficients and then identify that collection of parameter estimates that produces the largest likelihood, thus identifying these parameter estimates as the MLEs. While this profile likelihood approach is acceptable as demonstrated in other applications, directly estimating m_1, m_2 along with the other SCMPMA(1) model estimates would likewise prove beneficial, as would redefining the model to allow for real-valued estimators for m_1 and m_2 . These generalizations and estimation approaches can be explored in future work.

Simulated data examples illustrate that the SCMPMA(1) model can obtain unbiased estimates, and the model demonstrates potential for accurate forecasts given data containing any measure of data dispersion. The real data illustration, however, highlights the complexities that come with parameter estimation. While we nonetheless present a means towards achieving this goal, this approach does not perform but so strongly with regard to prediction and forecasting. It nonetheless serves as a starting point for parameter estimation that we will continue to investigate in future work. Moreover, the flexibility of the SCMPMA(1) aids in determining a parsimonious model form as appropriate.

Abbreviations

AR(1): First-order autoregressive; ARIMA: Auto-regressive integrated moving average; ARMA: Auto-regressive moving average; CMB: Conway-Maxwell-Binomial; CMP: Conway-Maxwell-Poisson; gCMB: Generalized Conway-Maxwell-Binomial; GPAR(1): First-order generalized Poisson autoregressive; GPMA(1): First-order generalized Poisson moving average; INAR(1): First-order integer-valued autoregressive; INMA: Integer-valued moving average; MA: Moving average; mgf: Moment generating function; PAR(1): First-order Poisson autoregressive; pgf: Probability generating function; PMA: Poisson moving average; PMA(1): First-order Poisson moving average; QB: Quasi-binomial; sCMP: Sum-of-Conway-Maxwell-Poisson; SCMPAR(1): First-order sum-of-Conway-Maxwell-Poisson autoregressive; SCMPMA(1): First-order sum-of-Conway-Maxwell-Poissons moving average

Acknowledgements

This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. SM and FC thank the Georgetown Undergraduate Research Opportunities Program (GUROP) for their support. All authors thank Dr. Christian Weiss for use of the IP dataset, and the reviewers for their feedback and comments.

Authors' contributions

KFS developed the research idea. All authors contributed towards the literature review, theoretical developments, and statistical computing. The author(s) read and approved the final manuscript.

Funding

SM was funded in part by the GUROP.

Availability of data and materials

Simulated data can vary given the generation process. Simulation code(s) can be supplied upon request. IP data set obtained from Dr. Christian Weiss of Helmut Schmidt University.

Competing interests

No authors have competing interests relating to this work.

Received: 22 April 2020 Accepted: 26 January 2021

Published online: 21 February 2021

References

- Al-Osh, M. A., Alzaid, A. A.: First-order integer valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* **8**(3), 261–275 (1987)
- Al-Osh, M. A., Alzaid, A. A.: Integer-valued moving average (INMA) process. *Stat. Pap.* **29**(1), 281–300 (1988)
- Al-Osh, M. A., Alzaid, A. A.: Binomial autoregressive moving average models. *Commun. Stat. Stoch. Model.* **7**(2), 261–282 (1991)
- Alzaid, A. A., Al-Osh, M. A.: Some autoregressive moving average processes with generalized Poisson marginal distributions. *Ann. Inst. Stat. Math.* **45**(2), 223–232 (1993)
- Borges, P., Rodrigues, J., Balakrishnan, N., Bazán, J.: A COM-Poisson type generalization of the binomial distribution and its properties and applications. *Stat. Probab. Lett.* **87**, 158–166 (2014)
- Brännäs, K., Hall, A.: Estimation in integer-valued moving average models. *Appl. Stoch. Model. Bus. Ind.* **17**, 277–291 (2001)
- Conway, R. W., Maxwell, W. L.: A queuing model with state dependent service rates. *J. Ind. Eng.* **12**, 132–136 (1962)
- Famoye, F.: Restricted generalized Poisson regression model. *Commun. Stat. Theory Methods.* **22**(5), 1335–1354 (1993)
- Hilbe, J. M.: *Modeling Count Data*. Cambridge University Press, New York, NY (2014)
- Joe, H.: Time series models with univariate margins in the convolution-closed infinitely divisible class. *J. Appl. Probab.* **33**(3), 664–677 (1996)
- Kadane, J. B.: Sums of possibly associated Bernoulli variables: The Conway-Maxwell-Binomial distribution. *Bayesian Anal.* **11**(2), 403–420 (2016)
- McKenzie, E.: ARMA models for dependent sequences of Poisson counts. *Adv. Appl. Probab.* **20**(4), 822–835 (1988)
- Sellers, K. F., Peng, S. J., Arab, A.: A flexible univariate autoregressive time-series model for dispersed count data. *J. Time Ser. Anal.* **41**(3), 436–453 (2020). <https://doi.org/10.1111/jtsa.12516>
- Sellers, K. F., Swift, A. W., Weems, K. S.: A flexible distribution class for count data. *J. Stat. Distrib. Appl.* **4**(22), 1–21 (2017). <https://doi.org/10.1186/s40488-017-0077-0>
- Sellers, K. F., Young, D. S.: Zero-inflated sum of Conway-Maxwell-Poissons (ZISCMP) regression. *J. Stat. Comput. Simul.* **89**(9), 1649–1673 (2019)
- Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., Boatwright, P.: A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Appl. Stat.* **54**, 127–142 (2005)
- Weiss, C. H.: Controlling correlated processes of Poisson counts. *Qual. Reliab. Eng. Int.* **23**(6), 741–754 (2007)
- Weiss, C. H.: Thinning operations for modeling time series of counts—a survey. *Adv. Stat. Anal.* **92**, 319–341 (2008)
- Weiss, C. H.: *An Introduction to Discrete-Valued Time Series*. John Wiley & Sons, Inc., Hoboken, NJ (2018)
- Weiss, C. H.: Stationary count time series models. *Wiley Interdiscip. Rev. Comput. Stat.* **13**(1), 1502 (2021). <https://doi.org/10.1002/wics.1502>
- Zhu, L., Sellers, K. F., Morris, D. S., Shmueli, G.: Bridging the gap: A generalized stochastic process for count data. *Am. Stat.* **71**(1), 71–80 (2017)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.