

RESEARCH

Open Access



# A flexible distribution class for count data

Kimberly F. Sellers<sup>1\*</sup> , Andrew W. Swift<sup>2</sup> and Kimberly S. Weems<sup>3</sup>

\*Correspondence:

kfs7@georgetown.edu

<sup>1</sup>Department of Mathematics and Statistics, Georgetown University, 306 St. Mary's Hall, Washington, DC, 20057, USA

Full list of author information is available at the end of the article

## Abstract

The Poisson, geometric and Bernoulli distributions are special cases of a flexible count distribution, namely the Conway-Maxwell-Poisson (CMP) distribution – a two-parameter generalization of the Poisson distribution that can accommodate data over- or under-dispersion. This work further generalizes the ideas of the CMP distribution by considering sums of CMP random variables to establish a flexible class of distributions that encompasses the Poisson, negative binomial, and binomial distributions as special cases. This sum-of-Conway-Maxwell-Poissons (sCMP) class captures the CMP and its special cases, as well as the classical negative binomial and binomial distributions. Through simulated and real data examples, we demonstrate this model's flexibility, encompassing several classical distributions as well as other count data distributions containing significant data dispersion.

**Keywords:** Conway-Maxwell-Poisson (CMP), Negative binomial, Poisson, Binomial, Geometric, Bernoulli, Over-dispersion, Under-dispersion

**Mathematics Subject Classification:** 60E05; 62F10

## 1 Introduction

The Poisson distribution is one of the most popular discrete distributions, serving as a natural, classical distribution to model count data. It is well-known that a random variable  $Y$  that is Poisson distributed with rate parameter  $\mu_*$  has a probability mass function (pmf) of the form,

$$P(Y = y) = \frac{\mu_*^y e^{-\mu_*}}{y!}, y = 0, 1, 2, \dots,$$

where  $\mu_*$  equals both the mean and variance of the distribution. The relationship between the mean and variance implies a goodness-of-fit index,  $GOF = \frac{Var(Y)}{E(Y)} = 1$ , i.e. equi-dispersion is established. This constraining assumption, however, does not oftentimes hold true for real data – an issue that has significant implications affecting numerous applications.

Over-dispersion relative to the Poisson distribution (i.e. where the variance is greater than the mean) is a common feature among real data. The most popular distribution to model over-dispersion is the negative binomial distribution; for such a random variable  $Y$  with a negative binomial( $n, p$ ) distribution, its pmf is

$$P(Y = y) = \binom{y+n-1}{y} (1-p)^y p^n, y = 0, 1, 2, \dots,$$

where  $y$  denotes the number of failures before the  $n$ th success in a series of Bernoulli trials with success probability,  $0 \leq p \leq 1$ . The geometric distribution with success probability

$p$  is a special case of negative binomial  $(n, p)$  where  $n = 1$ . The mean and variance of this random variable are  $E(Y) = \frac{n(1-p)}{p}$  and  $Var(Y) = \frac{n(1-p)}{p^2}$ , respectively, thus the goodness-of-fit index for dispersion is

$$GOF = \frac{Var(Y)}{E(Y)} = \frac{1}{p} \geq 1. \quad (1)$$

This dispersion index motivates considering the negative binomial distribution as a viable option for addressing data over-dispersion. In fact, this distribution is a popular choice for modeling over-dispersion in various statistical methods (e.g. regression (Hilbe 2008)) and is well studied with statistical computational ability in many softwares (e.g. SAS, R, etc.). The negative binomial distribution, however, is unable to address data under-dispersion, as demonstrated in Eq. (1). This result further illustrates that the Poisson GOF is the boundary case of the negative binomial distribution; the Poisson distribution is known to be the limiting case of the negative binomial distribution where  $n \rightarrow \infty$ .

The binomial distribution (while arguably a truncated count distribution) is an under-dispersed count distribution relative to the Poisson model. A binomially distributed random variable  $Y$  with  $b$  Bernoulli trials and success probability  $p_*$  has the pmf,

$$P(Y = y) = \binom{b}{y} p_*^y (1 - p_*)^{b-y}, \quad y = 0, 1, 2, \dots, b.$$

Naturally, the bernoulli  $(p_*)$  distribution is a special case of the binomial distribution where  $b = 1$ . The associated mean and variance of this random variable equal  $E(Y) = bp_*$  and  $Var(Y) = bp_*(1 - p_*)$ , respectively, thus the goodness-of-fit index for dispersion is  $GOF = 1 - p_* \leq 1$ . The Poisson, negative binomial, and binomial distributions are popular, classical tools for modeling count data of a particular (in)finite form. What is most interesting about these distributions is that they each represent sums of other classical distributions, namely the Poisson, geometric, and Bernoulli distributions, respectively. The Poisson, geometric and Bernoulli distributions are themselves special cases of the Conway-Maxwell-Poisson (CMP) distribution – a two-parameter flexible count distribution that generalizes the Poisson distribution to accommodate data over- or under-dispersion. This work introduces and thus considers the sum of CMP random variables to establish the flexible class of distributions that encompass the Poisson, geometric, Bernoulli, negative binomial, binomial, and CMP distributions as special cases.

The paper is outlined as follows. Section 2 acquaints the reader with the CMP distribution in order to motivate and introduce the sum-of-Conway-Maxwell-Poissons (sCMP) class in Section 3, including discussion of the statistical properties associated with this larger class of count distributions. Section 4 addresses parameter estimation and statistical computing procedures. Section 5 illustrates the flexibility of this class of distributions via simulated and real data examples. Finally, Section 6 concludes the manuscript with discussion.

## 2 The Conway-Maxwell-Poisson distribution

The CMP distribution is a viable two-parameter count distribution that generalizes the Poisson distribution in light of data dispersion. Conway and Maxwell (1962) derive the distributional form, motivated by considering a queuing system with a flexible state-

dependent service rate where  $\nu$  describes the degree to which the system service rate is affected by the system state. The resulting pmf has the form

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

for a random variable  $X$ , where  $\lambda = E(X^\nu)$  denotes a generalized form of the Poisson rate parameter,  $\nu \geq 0$  is a dispersion parameter, and  $Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$  normalizes the distribution such that the distribution satisfies the basic probability axioms. The dispersion parameter,  $\nu$ , accounts for the amount of data over- or under-dispersion relative to the Poisson distribution:  $\nu = 1$  implies that data equi-dispersion exists, while  $\nu > (<)1$  denotes under- (over-) dispersion relative to the Poisson model. The CMP distribution includes three well-known distributions as special cases: the Poisson( $\mu_* = \lambda$ ) distribution when  $\nu = 1$ , the geometric distribution with success probability  $p = 1 - \lambda$  when  $\nu = 0$  and  $\lambda < 1$ , and the Bernoulli distribution with success probability  $p_* = \frac{\lambda}{1+\lambda}$  as  $\nu \rightarrow \infty$ ; see Table 1 for details.

Shmueli et al. (2005) provide the moments for the CMP distribution via the recursion,

$$E(X^{r+1}) = \begin{cases} \lambda [E(X + 1)]^{1-\nu} & r = 0 \\ \lambda \frac{\partial}{\partial \lambda} E(X^r) + E(X)E(X^r) & r > 0. \end{cases} \quad (3)$$

The expected value and variance can alternatively be represented as

$$E(X) = \lambda \frac{\partial \log Z(\lambda, \nu)}{\partial \lambda} \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu}, \text{ and} \quad (4)$$

$$\text{Var}(X) = \frac{\partial E(X)}{\partial \log \lambda} \approx \frac{1}{\nu} \lambda^{1/\nu}, \quad (5)$$

where the approximation holds for  $\nu \leq 1$  or  $\lambda > 10^\nu$  (Sellers et al. 2011); see Minka et al. (2003) for details. More generally, the associated moment generating function of  $X$  is  $M_X(t) = \frac{Z(\lambda e^t, \nu)}{Z(\lambda, \nu)}$ , from which the higher moments can be obtained for  $X$ .

The CMP distribution satisfies several nice properties. The distribution has an exponential family form with joint sufficient statistics  $\{\sum_{i=1}^n x_i, \sum_{i=1}^n \log(x_i!)\}$  for  $\{\lambda, \nu\}$ . Further, the ratio between probabilities of two consecutive values is nonlinear in  $x$ , namely,

$$\gamma_{X,x} = \frac{P(X = x - 1)}{P(X = x)} = \frac{x^\nu}{\lambda}. \quad (6)$$

The linear relation among probabilities of two consecutive values is achieved when  $\nu = 1$ , i.e. given data equi-dispersion associated with the Poisson( $\lambda$ ) model. Meanwhile, for  $\nu = 0$  and  $\lambda < 1$  (i.e. the geometric distribution with success probability  $1 - \lambda$ ), we confirm that the ratio between probabilities of two consecutive values is constant, equaling  $\frac{1}{\lambda} > 1$ .

The CMP distribution has quickly grown in popularity because of its ability to model count data in a flexible manner. Methodological developments are vast, including works in distribution theory (Sellers 2012; Sellers and Shmueli 2013; Borges et al. 2014), regression analysis (Sellers and Shmueli 2009; 2010; Sellers and Raim 2016), control chart

**Table 1** Well-known distributions associated with the Conway-Maxwell-Poisson (CMP) distribution for special cases of  $\lambda$  and  $\nu$

Case	$Z(\lambda, \nu)$	pmf	Distribution
$\nu = 1$	$e^\lambda$	$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, 2, \dots$	Poisson( $\lambda$ )
$\nu = 0, \lambda < 1$	$\frac{1}{1-\lambda}$	$P(X = x) = (1 - \lambda) \lambda^x, x = 0, 1, 2, \dots$	Geom( $1 - \lambda$ )
$\nu \rightarrow \infty$	$1 + \lambda$	$P(X = 0) = \frac{1}{1+\lambda}; P(X = 1) = \frac{\lambda}{1+\lambda}$	Bernoulli( $\frac{\lambda}{1+\lambda}$ )

theory (Sellers 2012; Saghir and Lin 2014a; 2014b), stochastic processes (Zhu et al. 2017), and multivariate data analysis (Sellers et al. 2016). The model has further been applied for various data problems including fitting word lengths (Wimmer et al. 1994), modeling online sales (Boatwright et al. 2003; Borle et al. 2006) and customer behavior (Borle et al. 2007), analyzing traffic accident data (Lord et al. 2008), and for use as a disclosure limitation procedure to protect individual privacy (Kadane et al. 2006). See Sellers et al. (2011) for additional overview and discussion.

### 3 The sum of Conway-Maxwell-Poissons (sCMP) class of distributions and its statistical properties

The sum of  $m$  independent and identically distributed (iid) CMP variables leads to what will be termed a *sum of Conway-Maxwell-Poissons* (sCMP)( $\lambda, \nu, m$ ) class of distributions. Theorem 1 defines the three-parameter structure for some generalized rate parameter ( $\lambda$ ), dispersion parameter ( $\nu$ ), and number of underlying CMP random variables ( $m$ ).

**Theorem 1.** *The sCMP( $\lambda, \nu, m$ ) distribution has the following pmf for a random variable  $Y = \sum_{i=1}^m X_i$ , where  $X_i \stackrel{iid}{\sim} \text{CMP}(\lambda, \nu)$ :*

$$P(Y = y | \lambda, \nu, m) = \frac{\lambda^y}{(y!)^\nu Z^m(\lambda, \nu)} \sum_{\substack{x_1, \dots, x_m=0 \\ x_1 + \dots + x_m = y}}^y \binom{y}{x_1 \dots x_m}^\nu, \quad y = 0, 1, 2, \dots, \quad (7)$$

where  $\binom{y}{x_1 \dots x_m} = \frac{y!}{x_1! \dots x_m!}$  is the multinomial coefficient.

*Proof* We prove this result by induction. For  $m = 2$ , let  $X_i \sim \text{CMP}(\lambda, \nu)$ ,  $i = 1, 2$  and  $Y = X_1 + X_2$ . Then, the result holds by the transformation technique (see Chapter 2 of Casella and Berger (2002)). Similarly, given that the result is true for  $m = k - 1$ , we can likewise apply the transformation technique to show that

$$\begin{aligned} P(Y = y) &= \sum_{x_k} \frac{\lambda^{y-x_k}}{[(y-x_k)!]^\nu Z^{k-1}(\lambda, \nu)} \cdot \sum_{\substack{x_1, \dots, x_{k-1}=0 \\ x_1 + \dots + x_{k-1} = y-x_k}}^{y-x_k} \binom{y-x_k}{x_1 \dots x_{k-1}}^\nu \frac{\lambda^{x_k}}{(x_k!)^\nu Z(\lambda, \nu)} \\ &= \frac{\lambda^y}{Z^k(\lambda, \nu)} \sum_{\substack{x_1, \dots, x_k=0 \\ x_1 + \dots + x_k = y}}^y \frac{[(y-x_k)!]^\nu}{[(y-x_k)!]^\nu (x_k!)^\nu (x_1! \dots x_{k-1}!)^\nu} \cdot \left(\frac{y!}{y!}\right)^\nu \\ &= \frac{\lambda^y}{(y!)^\nu Z^k(\lambda, \nu)} \sum_{\substack{x_1, \dots, x_k=0 \\ x_1 + \dots + x_k = y}}^y \binom{y}{x_1 \dots x_k}^\nu. \end{aligned}$$

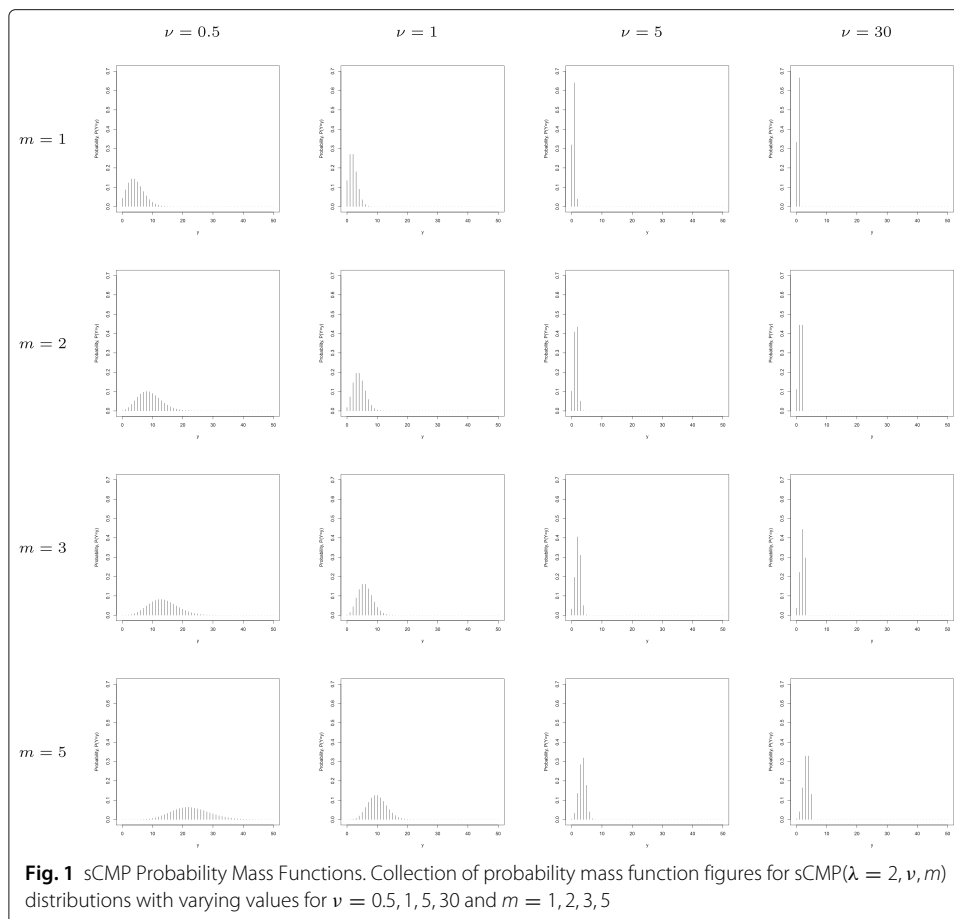
□

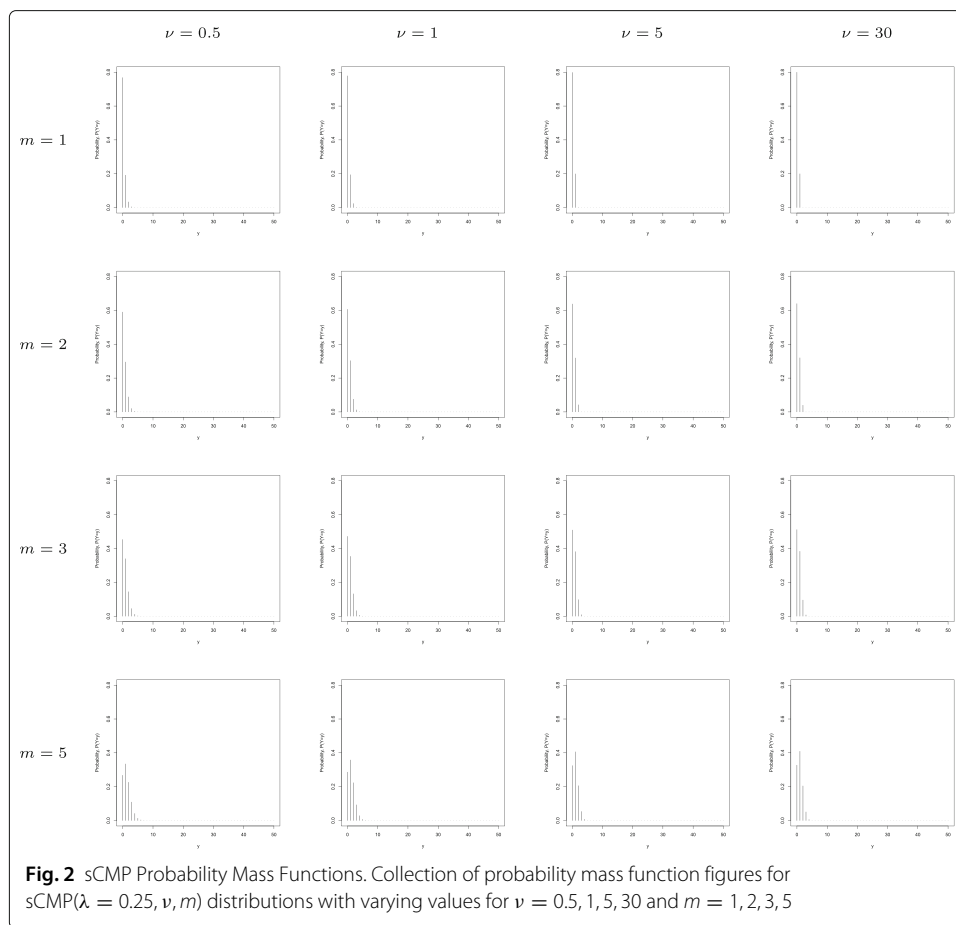
The sCMP( $\lambda, \nu, m$ ) class encompasses the Poisson distribution with rate parameter  $\mu_* = m\lambda$  (for  $\nu = 1$ ), negative binomial( $m, 1 - \lambda$ ) distribution (for  $\nu = 0$  and  $\lambda < 1$ ), and Binomial( $m, p$ ) distribution (as  $\nu \rightarrow \infty$  with success probability  $p = \frac{\lambda}{\lambda+1}$ ) as special cases. Further, for  $m = 1$ , the sCMP( $\lambda, \nu, m = 1$ ) is the CMP( $\lambda, \nu$ ) distribution. Accordingly, the sCMP class further captures the special case distributions of the CMP model: a geometric distribution with success probability,  $p = 1 - \lambda$ , when  $m = 1$ ,  $\nu = 0$ , and  $\lambda < 1$ ; and a Bernoulli distribution with success probability  $p_* = \frac{\lambda}{1+\lambda}$  when  $m = 1$  and  $\nu \rightarrow \infty$ .

Figures 1 and 2 display the sCMP class for different values of  $m = 1, 2, 3, 5$  and  $\nu = 0.5, 1, 5, 30$  for  $\lambda = 2$  (Fig. 1) and  $\lambda = 0.25$  (Fig. 2), respectively. Both figures illustrate the right skewness of the distribution, and show that the range of the data decreases as  $\nu$  increases. Figure 1 more clearly demonstrates how  $m$  and  $\nu$  influence the centrality and shape of the distribution when  $\lambda = 2$ . As previously discussed, the special case where  $\nu = 1$  simplifies the  $\text{sCMP}(\lambda, \nu, m)$  model to the  $\text{Poisson}(m\lambda)$  distribution. This illustrative example thus displays Poisson models with respective means equaling 4, 6, and 10. The increased variation in the respective figures is consistent with the increased shift in the distribution mean; recall that the Poisson mean and variance equal each other. Relative to the Poisson model, we see that (for a given  $m$ ) increasing  $\nu$  associates with decreasing variation. Figure 1 displays longer tail distributions relative to the Poisson distribution for  $\nu < 1$  and shorter tails relative to the Poisson model when  $\nu > 1$ . Meanwhile (for a given  $\nu$ ), increasing  $m$  clearly associates with increased shifts in the measures of distributional centrality (i.e. mean, median, and mode).

The ratio between probabilities of two consecutive values is

$$\gamma_{Y,y} = \frac{P(Y = y - 1)}{P(Y = y)} = \frac{y^\nu}{\lambda} \cdot \frac{\sum_{a_1, \dots, a_m=0}^{y-1} (a_1, \dots, a_m)^\nu}{\sum_{b_1, \dots, b_m=0}^y (b_1, \dots, b_m)^\nu}, \quad (8)$$





where the ratio of sums drops out in the special case where  $m = 1$  (i.e. one  $\text{CMP}(\lambda, \nu)$  random variable); clearly, this produces the special case shown in Eq. (6). For the special case where  $\nu = 1$ ,  $\gamma_{Y,y} = \frac{y}{m\lambda}$ , which is the linear form property of the Poisson random variable with parameter  $m\lambda$  (i.e. the distribution of the sum of  $m$  Poisson random variables). Meanwhile, for  $\nu = 0$ ,  $\gamma_{Y,y} = \frac{y}{\lambda(m+y-1)}$ , namely the form associated with a negative binomial distribution (i.e. the sum of  $m$  geometric random variables). Equation (8) implies that the sCMP model has a mode at 0 when  $\gamma_{Y,y} > 1$ , i.e.  $\lambda <$

$$y^v \cdot \frac{\sum_{a_1, \dots, a_m=0}^{y-1} (a_1, \dots, a_m)^v}{\sum_{b_1, \dots, b_m=0}^y (b_1, \dots, b_m)^v}. \text{ In particular, } \gamma_{Y,1} = \frac{P(Y=0)}{P(Y=1)} = \frac{1}{m\lambda}, \text{ thus sCMP}(\lambda, \nu, m)$$

models where  $\lambda < \frac{1}{m}$  have a mode at 0. Figure 2 displays the sCMP( $\lambda = 0.25, \nu, m$ ) distributions for  $\nu = 0.5, 1, 5, 30$  and  $m = 1, 2, 3, 5$ . Given that  $\lambda = 0.25 = \frac{1}{4}$ , we expect sCMP distributions where  $m < 4$  to have the mode at 0. This is illustrated accordingly in Fig. 2; sCMP( $\lambda = 0.25, \nu, m$ ) distributions for  $m = 2, 3$  and any  $\nu \geq 0$  have the mode at 0, while the sCMP( $\lambda = 0.25, \nu, m = 5$ ) distribution has the mode at 1 for all  $\nu \geq 0$ .

The moment-generating function  $M_Y(t)$ , probability generating function  $\Pi_Y(t)$ , and characteristic function  $\phi_Y(t)$  of a sCMP( $\lambda, \nu, m$ ) random variable  $Y$  are given, respectively, as

$$\begin{aligned}
 M_Y(t) &= \left( \frac{Z(\lambda e^t, \nu)}{Z(\lambda, \nu)} \right)^m, \\
 \Pi_Y(t) &= \left( \frac{Z(\lambda t, \nu)}{Z(\lambda, \nu)} \right)^m, \text{ and} \\
 \phi_Y(t) &= \left( \frac{Z(\lambda e^{it}, \nu)}{Z(\lambda, \nu)} \right)^m.
 \end{aligned} \tag{9}$$

The moment generating function technique can be used to show that, given the same parameters  $\lambda$  and  $\nu$ , the sum of independent sCMP distributions is invariant under addition (i.e. the sum of sCMP random variables has a sCMP distribution). For two independent random variables,  $Y_1 \sim \text{sCMP}(\lambda, \nu, m_1)$  and  $Y_2 \sim \text{sCMP}(\lambda, \nu, m_2)$ ,

$$\begin{aligned}
 M_{Y_1+Y_2}(t) &= M_{Y_1}(t) \cdot M_{Y_2}(t) \\
 &= \frac{Z^{m_1}(\lambda e^t, \nu)}{Z^{m_1}(\lambda, \nu)} \cdot \frac{Z^{m_2}(\lambda e^t, \nu)}{Z^{m_2}(\lambda, \nu)} \\
 &= \frac{Z^{m_1+m_2}(\lambda e^t, \nu)}{Z^{m_1+m_2}(\lambda, \nu)},
 \end{aligned}$$

which is the mgf of a  $\text{sCMP}(\lambda, \nu, m_1 + m_2)$  distribution, therefore  $Y_1 + Y_2$  has a  $\text{sCMP}(\lambda, \nu, m_1 + m_2)$  distribution. This result is logically sound because  $Y_1$  and  $Y_2$  respectively represent the sum of  $m_1$  and  $m_2$  iid CMP  $(\lambda, \nu)$  random variables; thus,  $Y_1 + Y_2$  defines the sum of  $m_1 + m_2$  iid CMP random variables, which precisely has a  $\text{sCMP}(\lambda, \nu, m_1 + m_2)$  distribution. This distinction is key between the CMP distribution and the larger sCMP class – the CMP distribution does not have the invariance property under addition.

### 3.1 Moments of the distribution

One can differentiate Eq. (9) to obtain the moments of the sCMP model, with the help of the following relation.

**Theorem 2.** For a normalizing function of the form,  $Z(\lambda e^t, \nu)$ , where  $Z(\cdot, \cdot)$  is as defined following Eq. (2), the  $k$ th ( $k = 1, 2, 3, \dots$ ) derivative is

$$\frac{\partial^k Z(\lambda e^t, \nu)}{\partial t^k} = \sum_{j=0}^{\infty} j^k \frac{(\lambda e^t)^j}{(j!)^\nu}. \tag{10}$$

*Proof* This proof is straightforward, given the differentiation formula for exponential functions.  $\square$

This result proves helpful in showing that the  $\text{sCMP}(\lambda, \nu, m)$  has mean  $E(Y) = mE(X)$  and variance  $V(Y) = mV(X)$ , where  $E(X)$  and  $V(X)$  (provided in Eqs. (4)-(5), respectively) are the mean and variance of a  $\text{CMP}(\lambda, \nu)$  random variable  $X$ .

### 3.2 Introducing the generalized Conway-Maxwell-Binomial (gCMB) distribution

Conditioning a CMP random variable on a sum of two independent CMP random variables produces a random variable whose distribution is Conway-Maxwell-Binomial (CMB) (Kadane 2016) (alternatively termed as “Conway-Maxwell-Poisson-Binomial” in Shmueli et al. (2005) and Borges et al. (2014)). The CMB random variable  $X$  has the pmf,

$$P(X = x \mid r, p, \nu) = \frac{\binom{r}{x}^\nu p^x (1-p)^{r-x}}{\sum_{k=0}^r \binom{r}{k}^\nu p^k (1-p)^{r-k}}, \quad x = 0, 1, 2, \dots, r, \tag{11}$$

for  $r \in \mathcal{Z}^+$ ,  $p \in (0, 1)$ , and  $\nu \in \mathcal{R}$  such that  $\nu = 1$  produces the usual binomial( $r, p$ ) distribution,  $\nu > 1$  corresponds to data under-dispersion relative to a binomial distribution while  $\nu < 1$  corresponds to data over-dispersion relative to the binomial( $r, p$ ) model. Extreme distribution cases hold where, for  $\nu \rightarrow \infty$ , the pmf is concentrated at the point,  $rp$  and, for  $\nu \rightarrow -\infty$ , it is concentrated at 0 or  $r$  (Borges et al. 2014).

Analogously, conditioning a sCMP variable on the sum of two independent sCMP variables produces a generalized form of the CMB distribution; we denote this as the gCMB distribution. Letting  $S = Y_1 + Y_2$  where  $Y_1 \sim \text{sCMP}(\lambda_1, \nu, m_1)$  and  $Y_2 \sim \text{sCMP}(\lambda_2, \nu, m_2)$  are independent, and given  $S = s$ ,

$$P(Y_1 = y_1 | S = s) = \left\{ \binom{s}{y_1}^\nu \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{y_1} \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{s-y_1} \left[ \sum_{\substack{a_1, \dots, a_{m_1}=0 \\ a_1 + \dots + a_{m_1} = y_1}} \binom{y_1}{a_1, \dots, a_{m_1}}^\nu \right] \right. \\ \left. \cdot \left[ \sum_{\substack{b_1, \dots, b_{m_2}=0 \\ b_1 + \dots + b_{m_2} = s-y_1}} \binom{s-y_1}{b_1, \dots, b_{m_2}}^\nu \right] \right\} / G\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}, \nu, s, m_1, m_2\right) \quad (12)$$

where

$$G(p, \nu, s, m_1, m_2) = \sum_{k=0}^s \binom{s}{k}^\nu p^k (1-p)^{s-k} \left[ \sum_{\substack{a_1, \dots, a_{m_1}=0 \\ a_1 + \dots + a_{m_1} = k}} \binom{k}{a_1, \dots, a_{m_1}}^\nu \right] \\ \cdot \left[ \sum_{\substack{b_1, \dots, b_{m_2}=0 \\ b_1 + \dots + b_{m_2} = s-k}} \binom{s-k}{b_1, \dots, b_{m_2}}^\nu \right] \quad (13)$$

is a normalizing constant and  $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$ . Thus, the conditional probability of a  $\text{sCMP}(\lambda_1, \nu, m_1)$  random variable given the value of a sum of sCMP random variables as described above is

$$P(Y_1 = y_1 | S = s) \propto \binom{s}{y_1}^\nu p^{y_1} (1-p)^{s-y_1} \left[ \sum_{\substack{a_1, \dots, a_{m_1}=0 \\ a_1 + \dots + a_{m_1} = y_1}} \binom{y_1}{a_1, \dots, a_{m_1}}^\nu \right] \\ \cdot \left[ \sum_{\substack{b_1, \dots, b_{m_2}=0 \\ b_1 + \dots + b_{m_2} = s-y_1}} \binom{s-y_1}{b_1, \dots, b_{m_2}}^\nu \right]. \quad (14)$$

This generalized CMB distribution [denoted as  $\text{gCMB}(p, \nu, s, m_1, m_2)$ ] contains several special cases. When  $m_1 = m_2 = 1$ , the gCMB distribution reduces to the CMB( $s, p, \nu$ ) distribution. For  $\nu = 1$ , the probability reduces to

$$P(Y_1 = y_1 | S = s) = \frac{\binom{s}{y_1} (m_1 p)^{y_1} [m_2 (1-p)]^{s-y_1}}{[m_1 p + m_2 (1-p)]^s} \\ = \binom{s}{y_1} \left( \frac{m_1 p}{m_1 p + m_2 (1-p)} \right)^{y_1} \left( \frac{m_2 (1-p)}{m_1 p + m_2 (1-p)} \right)^{s-y_1},$$



i.e. given data equi-dispersion, we have a binomial distribution with  $s$  trials and  $p^* = \frac{m_1 p}{m_1 p + m_2(1-p)} = \frac{m_1 \lambda_1}{m_1 \lambda_1 + m_2 \lambda_2}$  success probability. In particular, for  $m_1 = m_2 = 1$  and  $v = 1$ , the gCMB distribution reduces to the  $\text{Bin}(s, p) = \text{Bin}\left(s, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$  distribution. For the special case where  $\lambda_1 = \lambda_2 = \lambda$ , Eq. (12) reduces to the following for  $y_1 = 0, \dots, s$ :

$$P(Y_1 = y_1 \mid S = s) = \frac{\binom{s}{y_1}^v \left[ \sum_{\substack{a_1, \dots, a_{m_1}=0 \\ a_1 + \dots + a_{m_1} = y_1}}^{y_1} \binom{y_1}{a_1, \dots, a_{m_1}}^v \right] \left[ \sum_{\substack{b_1, \dots, b_{m_2}=0 \\ b_1 + \dots + b_{m_2} = s - y_1}}^{s - y_1} \binom{s - y_1}{b_1, \dots, b_{m_2}}^v \right]}{\sum_{\substack{c_1, \dots, c_{m_1+m_2}=0 \\ c_1 + \dots + c_{m_1+m_2} = s}}^s \binom{s}{c_1, \dots, c_{m_1+m_2}}^v},$$

i.e. a gCMB(1/2,  $v, s, m_1, m_2$ ) distribution. In particular, this implies that

$$\sum_{\substack{c_1, \dots, c_{m_1+m_2}=0 \\ c_1 + \dots + c_{m_1+m_2} = s}}^s \binom{s}{c_1, \dots, c_{m_1+m_2}}^v = \sum_{y_1=0}^s \binom{s}{y_1}^v \left[ \sum_{\substack{a_1, \dots, a_{m_1}=0 \\ a_1 + \dots + a_{m_1} = y_1}}^{y_1} \binom{y_1}{a_1, \dots, a_{m_1}}^v \right] \cdot \left[ \sum_{\substack{b_1, \dots, b_{m_2}=0 \\ b_1 + \dots + b_{m_2} = s - y_1}}^{s - y_1} \binom{s - y_1}{b_1, \dots, b_{m_2}}^v \right].$$

The probability generating function for the gCMB distribution is

$$\begin{aligned} \phi_{Y_1|S=s}(t) &= \frac{\sum_{y_1=0}^s \binom{s}{y_1}^v (tp)^{y_1} (1-p)^{s-y_1} \left[ \sum_{a_1, \dots, a_{m_1}=0}^{y_1} \binom{y_1}{a_1, \dots, a_{m_1}}^v \right] \left[ \sum_{b_1, \dots, b_{m_2}=0}^{s-y_1} \binom{s-y_1}{b_1, \dots, b_{m_2}}^v \right]}{G(p, v, s, m_1, m_2)} \\ &= \frac{H\left(\frac{tp}{1-p}, v, s, m_1, m_2\right)}{H\left(\frac{p}{1-p}, v, s, m_1, m_2\right)}, \end{aligned}$$

where

$$H(\theta, v, s, m_1, m_2) = \sum_{y_1=0}^s \binom{s}{y_1}^v \theta^{y_1} \left[ \sum_{a_1, \dots, a_{m_1}=0}^{y_1} \binom{y_1}{a_1, \dots, a_{m_1}}^v \right] \left[ \sum_{b_1, \dots, b_{m_2}=0}^{s-y_1} \binom{s-y_1}{b_1, \dots, b_{m_2}}^v \right]. \quad (15)$$

#### 4 Parameter estimation and statistical computing

Because  $m$  is a natural number, we consider a series of sCMP estimations for a given  $m$ , from which we can determine an optimal sCMP( $\lambda, v, m$ ) model. Given  $m$ , estimates for  $\lambda$  and  $v$  are obtained via maximum likelihood estimation (MLE), where we consider the log-likelihood,

$$\log \mathcal{L}(\lambda, v \mid m) = \sum_{i=1}^N \log P(Y_i = y_i \mid \lambda, v, m)$$

for a random sample  $Y_1, \dots, Y_N$ , where  $P(Y_i = y_i \mid \lambda, v, m)$ ,  $i = 1, \dots, N$  is defined in Eq. (7). Given the complex nature of the log-likelihood function and the corresponding score equations, as well as the constrained parameter space for  $\lambda > 0$  and  $v \geq 0$ , maximum likelihood estimates are determined via the `nllminb` function in *R* (R Core Team 2017) which is used to identify the parameters that minimize the negated log-likelihood function (thus determining the MLE values). Meanwhile, parameter robustness is quantified through the corresponding standard errors for the stated estimates obtained via the information matrix,

$$I(\lambda, \nu) = -N \cdot E \left( \begin{array}{cc} \frac{\partial^2 \ln P(Y=y)}{\partial \lambda^2} & \frac{\partial^2 \ln P(Y=y)}{\partial \lambda \partial \nu} \\ \frac{\partial^2 \ln P(Y=y)}{\partial \lambda \partial \nu} & \frac{\partial^2 \ln P(Y=y)}{\partial \nu^2} \end{array} \right),$$

where  $P(Y = y)$  is defined in Eq. (7). The information matrix is computed (via the hessian function in the `numDeriv` package (Gilbert and Varadhan 2016) in *R* (R Core Team 2017)) and inverted, and the square root of the resulting diagonal elements contain the standard errors of the parameter estimates.

Optimal  $\text{sCMP}(\lambda, \nu, m)$  models are determined by comparing potential conditional sCMP models where  $m$  is assumed known and identifying the conditional model with the largest log-likelihood value. Section 5 illustrates this procedure via simulated and real data examples.

Statistical computing for the Poisson and negative binomial distributions are conducted in *R* (R Core Team 2017) via the function, `fitdistr`, contained in the *MASS* package (Venables and Ripley 2002). This package uses an alternative parametrization for the negative binomial model, namely  $\theta = n$  and  $\mu = \frac{n(1-p)}{p}$ , hence we can back-solve for  $p = \frac{\theta}{\mu + \theta}$ . Estimates for  $\theta$  and  $\mu$  are reported in the discussions provided in Section 5.

## 5 Examples

### 5.1 Simulation study

The sCMP distribution is a generalizable distribution that encompasses five classical distributions: the Bernoulli, binomial, Poisson, geometric, and negative binomial distributions; more broadly, for a general  $m$ , the sCMP distribution captures the binomial, Poisson, and negative binomial distributions. To demonstrate this general flexibility, data samples of size 100 were generated from a binomial( $b = 3, p_* = 0.667$ ), Poisson( $\mu_* = 6$ ), and negative binomial( $n = 3, p = 0.333$ ) distribution, respectively. To assess model performance, we compare model estimation via the sCMP distribution (assuming  $m = 1, 2, 3, 4$ ) with estimations assuming a Poisson and negative binomial distribution, respectively; the CMP distribution is the sCMP( $m = 1$ ) case. Table 2 provides the parameter estimates and standard errors associated with the various models considered for model comparisons via the log-likelihood ( $\log(L)$ ), Akaike and Bayes Information Criteria (AIC and BIC), respectively.

For model comparison via AIC, Burnham and Anderson (2002) suggest considering  $\Delta_i = AIC_i - AIC_{\min}$ , where  $AIC_{\min}$  is the minimum of the model AIC values being compared, thus inferring that the best model has  $\Delta = 0$  and the other models have  $\Delta > 0$ . Model comparisons are thus determined via these difference measures in that “models having  $\Delta_i \leq 2$  have substantial support (evidence), those in which  $4 \leq \Delta_i \leq 7$  have considerably less support, and models having  $\Delta_i > 10$  have essentially no support” in comparison with the best model; see p. 70-71 of Burnham and Anderson (2002). We will apply this approach for model comparison accordingly, and can analogously apply this method using BIC.

The sCMP class of distributions appears to offer a consistent ability to properly model all of the simulated classical data structures. What is interesting to see is the distribution’s resulting parameter estimations as  $m$  increases. For the binomial example (i.e. the case of extreme under-dispersion), we see that  $\lambda$  decreases and  $\nu$  increases for  $m \leq 3$ . While that

**Table 2** Simulated data example

		Simulated data distribution		
		Bin( $b = 3, p_* = 0.667$ )	Pois( $\mu_* = 6$ )	NB( $n = 3, p = 0.333$ )
Pois.	$\hat{\mu}_*$ (SE)	2.0000(0.1414)	6.1100(0.2472)	5.3000(0.2302)
	log(L)	-144.8109	-228.1811	-288.0710
	AIC	291.6218	458.3623	578.1419
	BIC	294.2270	460.9675	580.7471
NB	$\hat{\mu}$ (SE)	1.9999(0.1419)	6.1100(0.2503)	5.3001(0.3632)
	$\hat{\theta}$ (SE)	276.5396(394.3489)	239.7812(421.0122)	3.5599(0.8564)
	log(L)	-145.0563	-228.3236	-258.7486
	AIC	294.1126	460.6472	521.4971
	BIC	299.3229	465.8575	526.7075
CMP/sCMP( $m = 1$ )	$\hat{\lambda}$ (SE)	18.7071(8.9855)	6.9145(2.1193)	1.5576(0.2079)
	$\hat{\nu}$ (SE)	3.3931(0.5024)	1.0653(0.1603)	0.3150(0.0708)
	log(L)	-123.2624	-228.0950	-260.3649
	AIC	250.5248	460.1900	524.7298
	BIC	255.7351	465.4003	529.9402
sCMP( $m = 2$ )	$\hat{\lambda}$ (SE)	4.2531(0.9494)	3.4046(0.8152)	0.9309(0.1109)
	$\hat{\nu}$ (SE)	4.2854(0.4998)	1.0838(0.1826)	0.1674(0.0825)
	log(L)	-120.8816	-228.0722	-259.3193
	AIC	245.7632	460.1444	522.6386
	BIC	250.9735	465.3547	527.8489
sCMP( $m = 3$ )	$\hat{\lambda}$ (SE)	2.0000(0.2450)	2.2683(0.4822)	0.6709(0.0576)
	$\hat{\nu}$ (SE)	33.6942(12536.57)	1.1093(0.2127)	0.0392(0.0698)
	log(L)	-116.2486	-228.0418	-258.8683
	AIC	236.4972	460.0836	521.7366
	BIC	241.7075	465.2939	526.9469
sCMP( $m = 4$ )	$\hat{\lambda}$ (SE)	1.0000(0.1000)	1.7044(0.3382)	0.5700(0.0486)
	$\hat{\nu}$ (SE)	32.5126(12322.57)	1.1381(0.2469)	0.0000(0.0826)
	log(L)	-124.7123	-228.0175	-258.8470
	AIC	253.4246	460.0350	521.6940
	BIC	258.6349	465.2453	526.9043

True model parameters versus estimated parameters (and associated standard errors provided in parentheses) for various assumed distributions. For model comparisons, the log-likelihood, Akaike and Bayes Information Criteria (AIC and BIC, respectively) are provided

pattern does not continue for  $m = 4$ , we see that the log-likelihood value is maximized (and the AIC and BIC values minimized) with the sCMP( $m = 3$ ) case. We see that the sCMP( $\hat{\lambda} = 2.0000$ ,  $\hat{\nu} = 33.6942$ ,  $m = 3$ ) distribution is the best model, when compared with the other considered distributions. In fact, all other models produce a difference  $\Delta$  that associates with considerably less support to essentially no support.

The binomial case can be viewed as the summation of three Bernoulli trials, thus we expect the corresponding sCMP estimates to be  $\hat{\lambda} \approx 2$  and  $\hat{\nu} \geq 30$ ; recall that the special CMP case that corresponds with a Bernoulli distribution occurs when  $\nu \rightarrow \infty$  with probability  $\frac{\lambda}{1+\lambda}$ , where empirical evidence shows that dispersion parameter estimation is sufficiently achieved when  $\hat{\nu} \approx 30$  or more (see Sellers et al. (2016) and Sellers and Raim (2016) for examples). In fact, for the simulated Binomial dataset, we obtain  $\hat{\lambda} = 2.0000$ ,  $\hat{\nu} = 33.6942$ ; the obtained estimate for  $\nu$  implies extreme under-dispersion, thus we have sufficient evidence implying that the estimates approximate a Bernoulli distribution with

success probability,  $\hat{p}_* = \frac{2.0000}{1+2.0000} = 0.6667$ . The sCMP( $m = 3$ ) distribution best models the binomial data, producing the largest log-likelihood ( $\log(L) = -116.2486$ ), and the smallest AIC and BIC (236.4972 and 241.7075, respectively). In comparison, the Poisson and negative binomial models produce comparable log-likelihoods (both that are considerably less than those from the sCMP class) because they are unable to effectively model the under-dispersion present in this dataset. The large negative binomial parameter ( $\hat{\theta} = 276.5396$ ) shows that the model is converging to a Poisson model (i.e. towards data equi-dispersion) to estimate this data. While the CMP model is able to recognize the dataset as being under-dispersed ( $\hat{\nu} = 3.3931 > 1$ ), the form of the distribution still limits the amount of model flexibility it can address.

For the Poisson example, we see that all of the considered models perform comparably well. While the best model is naturally Poisson, this is true moreso because the distribution only requires estimating one parameter. All of the models considered produced log-likelihoods equalling approximately  $-228$ , thus the associated difference measures imply that the other models (in particular, the sCMP class of distributions) show substantial support for model consideration. The negative binomial estimates ( $\hat{\theta} = 239.7812$  and  $\hat{\mu} = 6.1100$ ) demonstrate the convergence of the negative binomial distribution to the Poisson model as  $\theta \rightarrow \infty$  in order to address the limiting case of analyzing equi-dispersed data.

Because the simulation reflects a Poisson(6) dataset, we expect to obtain sCMP parameter estimates  $\hat{\lambda} \approx 6/m$  and  $\hat{\nu} \approx 1$  for all  $m = 1, 2, 3, 4$ . The obtained estimates for  $\lambda$  and  $\nu$  are consistently larger than their projected values where  $\hat{\nu}$  increases slightly with  $m$ . The corresponding parameter standard errors, however, suggest that none of these estimates is statistically significantly different from their hypothesized values.

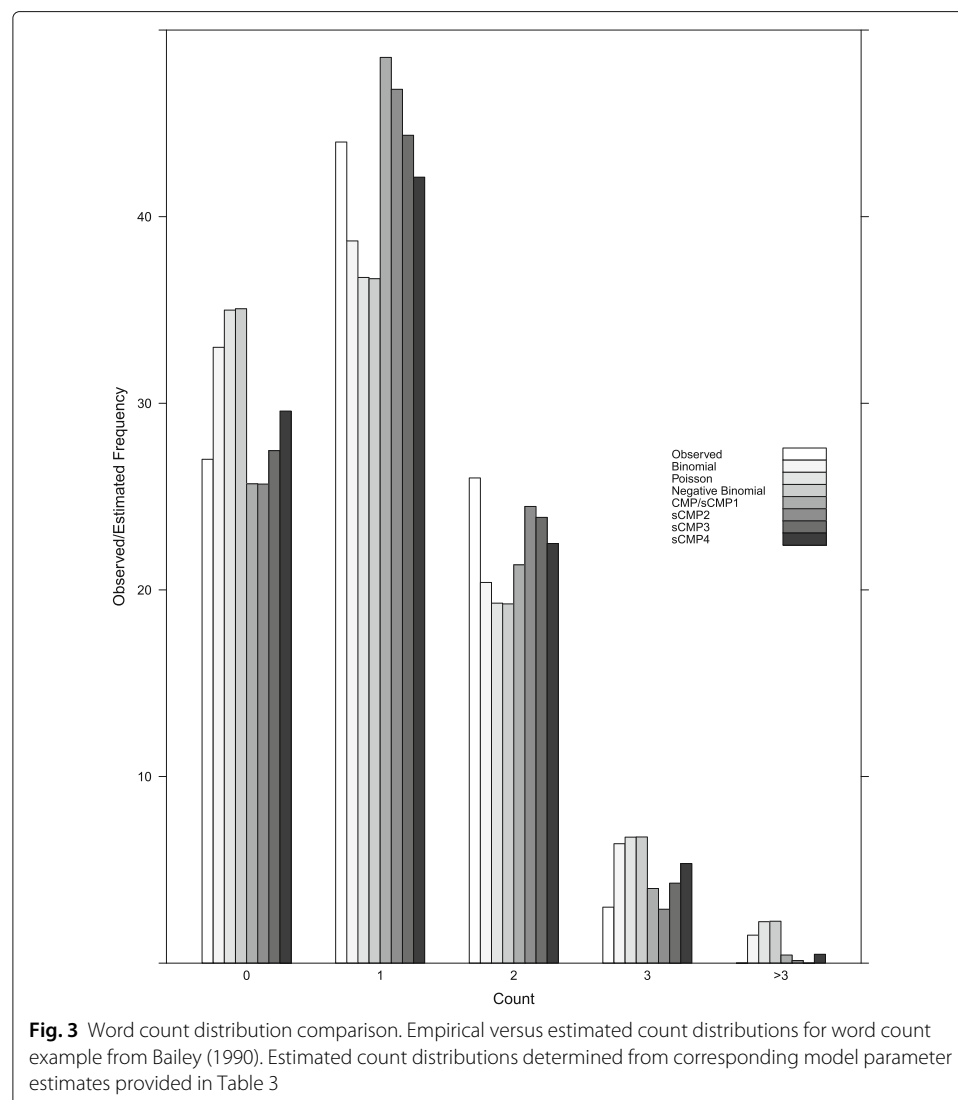
For the negative binomial example, we see that the sCMP class of distributions again performs well in estimating the form of the simulated dataset. The true parameter values associated with the negative binomial model imply that  $\mu = \frac{n(1-p)}{p} = 6$  and  $\theta = 3$ . The negative binomial distribution ( $\hat{\theta} = 3.5599$ ,  $\hat{\mu} = 5.3001$ ) is the best model among the distributions considered (AIC = 521.4971), however, the sCMP class of distributions performs more optimally as  $m$  increases (among those values for  $m$  considered). Larger values for  $m$  were not considered here because the sCMP models for  $m = 3, 4$  produce approximately equal log-likelihood values, thus likewise producing comparable AIC and BIC values; this makes sense because the negative binomial estimate is  $3 < \hat{\theta} = 3.5599 < 4$ . Meanwhile, even for the sCMP models where  $m = 1, 2$ , the difference in AIC when compared with the best model still implies that these models show considerable support.

With the sCMP class of distributions, we see that  $\hat{\nu}$  decreases as  $m$  increases. Interestingly here, because we know the data are simulated from a negative binomial( $n = 3, p = 0.333$ ) distribution, we expect the sCMP( $m = 3$ ) distribution to produce estimates  $\hat{\lambda} = 0.667$  and  $\nu \approx 0$ . In fact, the observed estimates ( $\hat{\lambda} = 0.6709$  and  $\hat{\nu} = 0.0392$ ) are within one standard error of the projected estimates. The CMP (i.e. the sCMP( $m = 1$ )) model does reasonably well, as evidenced by the resulting log-likelihood and AIC values ( $-260.3649$  and  $524.7298$ , respectively); the CMP estimated dispersion parameter,  $\hat{\nu} = 0.3150$ , indicates recognized over-dispersion in the dataset. The Poisson model is the worst performer (with  $\log(L) = -288.0710$ ) because of its constraining equi-dispersion requirement.

## 5.2 Under-dispersed real data example: word count

Bailey (1990) studies the frequency of articles in 10-word samples from Macaulay's "Essay on Milton", counting the number of occurrences of articles 'the', 'a', and 'an' as a means to infer the author's style. The provided dataset contains 100 observations where the number of occurrences of these articles in the 10-word samples range from 0 to 3; see Fig. 3.

We consider the Poisson, negative binomial, and sCMP( $m$ ) models where  $m = 1, 2, 3, 4$  to describe the data distribution; Bailey (1990) previously considered a binomial model to describe the data. Table 3 provides the sCMP parameter estimates and standard errors (in parentheses), along with the log-likelihood, AIC, and BIC values for model comparison. The sCMP( $m = 2$ ) model is the optimal choice, producing a log-likelihood equaling  $-117.327$ , and AIC and BIC equaling 238.6546 and 243.8649, respectively. Because this dataset is under-dispersed (with a sample mean and sample variance equaling 1.05 and 0.654, respectively), all models considered from the sCMP family outperform the Poisson and negative binomial models. The Poisson model produces an estimated sample mean and standard error, 1.0500 (0.1025), with log-likelihood  $-123.2741$ . The negative



**Table 3** Word count model comparisons

	CMP/sCMP( $m = 1$ )	sCMP( $m = 2$ )	sCMP( $m = 3$ )	sCMP( $m = 4$ )
$\hat{\lambda}$ (SE)	1.8897 (0.4219)	0.9120 (0.1511)	0.5385 (0.0652)	0.3559 (0.0404)
$\hat{\nu}$ (SE)	2.1033 (0.3858)	3.7750 (1.0049)	3.0900 (15045)	29.7650 (13118)
$\log(L)$	-118.319	-117.327	-117.331	-118.521
AIC	240.638	238.655	238.662	241.041
BIC	245.848	243.865	243.873	246.252

Model comparison for the word count data from Bailey (1990), where sCMP with  $m = 1, 2, 3, 4$  distributions are considered. For model comparisons, the log-likelihood, Akaike and Bayes Information Criteria (AIC and BIC, respectively) are provided. All sCMP family distributions outperform the Poisson model which produces an estimated sample mean,  $\mu_* = 1.0500$  (0.1025), with log-likelihood  $-123.2741$ . The negative binomial model likewise converges to a Poisson model with estimates,  $\hat{\theta} = 269.9607$  (702.1046),  $\hat{\mu} = 1.0500$  (0.1027),  $\log(L) = -123.3487$

binomial model meanwhile produces estimates ( $\hat{\theta} = 269.9607$  (702.1046),  $\hat{\mu} = 1.0500$  (0.1027),  $\log(L) = -123.3487$ ) comparable to the Poisson. Because the data are under-dispersed, the negative binomial model can only perform as well as the Poisson model. As demonstrated, the estimated size parameter is large and the estimated mean equals that from the Poisson model.

Figure 3 provides the empirical and estimated distributions for this data based on the various considered models, including the estimated binomial frequencies provided in Bailey (1990). This figure confirms the results provided in Table 3, namely that the sCMP( $m = 2$ ) best represents the shape of the observed distribution for the number of occurrences of an article in 10-word samples from Macaulay's 'Essay on Milton'. In particular, we see the small estimated sCMP( $m = 2$ ) frequency associated with more than 3 articles; recall that the observed number of articles is zero. Meanwhile, the number of occurrences as determined via the Poisson and negative binomial are visibly over- or under-estimated, including a sizable estimated frequency associated with more than 3 articles. The estimated frequencies determined by the binomial distribution, while better than those from the Poisson and negative binomial models, still deviate considerably in comparison to the sCMP class. Finally, while Table 3 shows that the sCMP( $m = 3$ ) distribution performs comparably well, we nonetheless determine the sCMP( $\hat{\lambda} = 0.9120, \hat{\nu} = 3.7750, m = 2$ ) model to be the best choice to estimate the observed distribution, based on the resulting estimated frequencies shown in Fig. 3.

### 5.3 Over-dispersed real data example: fetal lamb movement

Guttorp (1995) provides data on the number of movements by a fetal lamb observed by ultrasound and counted in successive 5-second intervals. The dataset contains 225 observations ranging in value from 0 to 7, and are over-dispersed with dispersion index  $\widehat{\text{Var}}(Y)/\widehat{E}(Y) = 0.693/0.382 = 1.8119$ ; summary information regarding the distribution is provided in Table 4(a). Assuming no knowledge of the data dispersion type, we consider various count data model parameter estimations to describe this real data distribution: Poisson, negative binomial, and sCMP at various levels of  $m = 1, 2, 3, 4$ . Table 5 provides the resulting estimation output (including the corresponding log-likelihood, AIC, and BIC) associated with the various distributions considered to model the original 5-second movement data summarized in Table 4(a).

**Table 4** Fetal lamb summary statistic information

Summary statistic	(a)	(b)
	5-second Data value	15-second Data value
No. of obs.	225	75
Minimum	0.000	0.000
1st Quartile	0.000	0.000
Median	0.000	1.000
Mean	0.382	1.147
3rd Quartile	1.000	2.000
Maximum	7.000	12.000
Variance	0.693	3.694
Std. Deviation	0.832	1.922

Summary statistics associated with fetal lamb data set: (a) based on original data, where fetal lamb are observed by ultrasound and counted in successive 5-second intervals, and (b) based on reconstructed (from original) data, where fetal lamb are observed by ultrasound and counted in successive 15-second intervals. Full original data are contained in Guttorp (1995)

The sCMP class consistently recognizes this real count distribution to be extremely over-dispersed ( $\hat{\nu} = 0.000$  and  $\lambda < 1$ ), implying that the sCMP class interprets the data as being represented as sums of size  $m$  from geometrically distributed data with some success probability,  $1 - \hat{\lambda}$ . In fact, the estimates for  $\lambda$  decrease as  $m$  increases yet the corresponding log-likelihood value decreases, thus providing a sense of the contour of the larger log-likelihood space that is determined by  $\lambda$ ,  $\nu$ , and  $m$ . Because  $m$  is a natural number, we find that the optimal sCMP( $m$ ) class for modeling the 5-second fetal lamb dataset occurs for  $m = 1$ , i.e. the CMP distribution with  $\hat{\lambda} = 0.277$  and  $\hat{\nu} = 0.000$ . Continuing with this logic, however, we recognize then that one should thus consider the special case of a geometric (i.e. the CMP distribution where  $\nu = 0$ ) distribution with approximate success probability,  $\hat{p} = 1 - 0.277 = 0.723$ . Indeed, estimating the observed count distribution via a geometric model produces the estimated success probability,  $\hat{p} = 0.723$  (with standard error, 0.025). While this estimation procedure determines a geometric model to be the best model within the sCMP class, the negative binomial distribution is another viable model, as determined by Burnham and Anderson (2002); see Table 5.

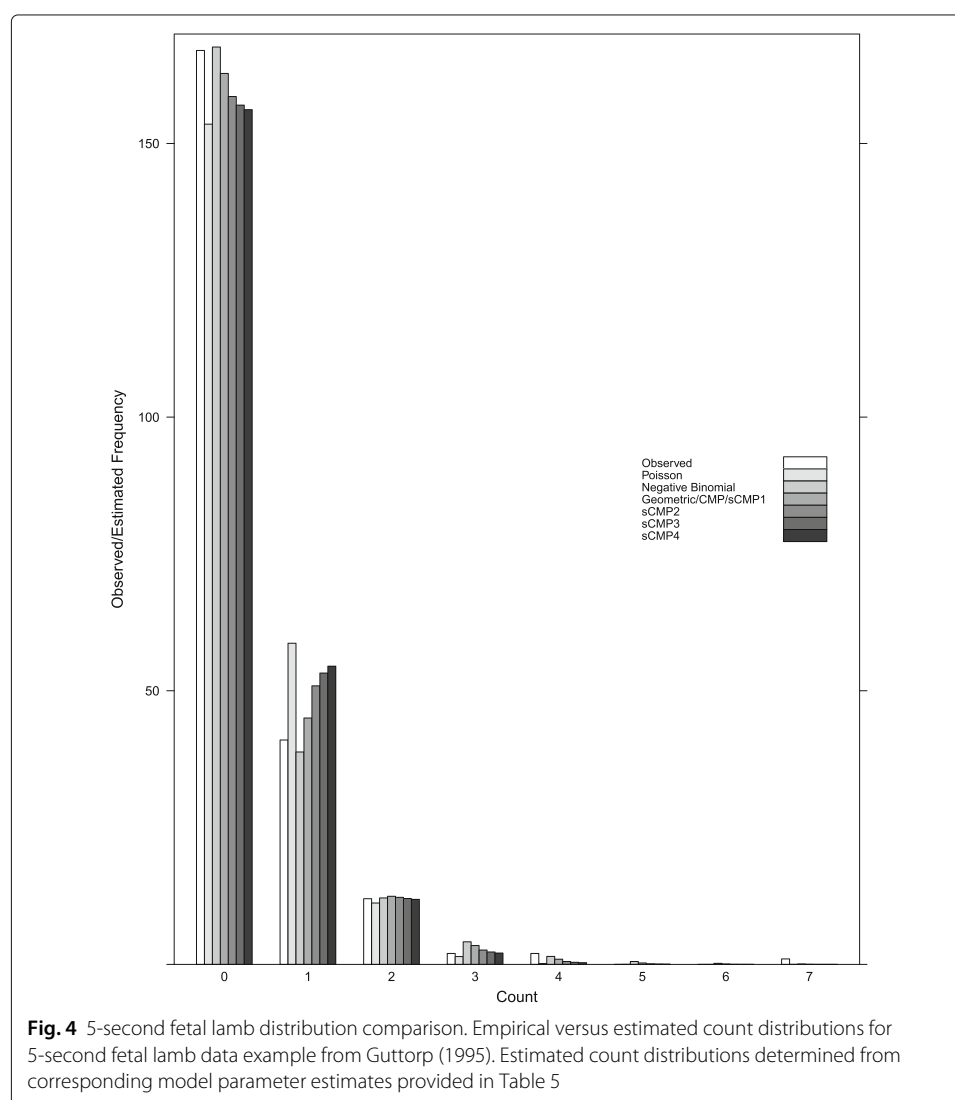
**Table 5** 5-second fetal lamb data model comparisons

	Est.	SE	log(L)	AIC	BIC
Poisson	$\hat{\mu}_* = 0.382$	(0.041)	-195.4933	394.9866	401.8188
Geom	$\hat{p} = 0.723$	(0.025)	-183.3791	368.7583	372.1744
NB	$\hat{\mu} = 0.382$ $\hat{\theta} = 0.587$	(0.053) (0.200)	-182.3702	368.7404	375.5726
CMP/sCMP( $m = 1$ )	$\hat{\lambda}_* = 0.277$ $\hat{\nu}_* = 0.000$	(0.040) (0.264)	-183.3791	370.7582	377.5904
sCMP( $m = 2$ )	$\hat{\lambda} = 0.160$ $\hat{\nu} = 0.000$	(0.019) (0.263)	-186.3873	376.7746	383.6068
sCMP( $m = 3$ )	$\hat{\lambda} = 0.113$ $\hat{\nu} = 0.000$	(0.013) (0.294)	-188.3006	380.6012	387.4334
sCMP( $m = 4$ )	$\hat{\lambda} = 0.087$ $\hat{\nu} = 0.000$	(0.009) (0.213)	-189.5569	383.1138	389.9460

Model comparison for the fetal lamb 5-second movement data, where Poisson, negative binomial, CMP, and sCMP distributions are considered. For model comparisons, the log-likelihood, Akaike and Bayes Information Criteria (AIC and BIC, respectively) are provided

Figure 4 provides a comparison of the empirical versus estimated count distributions associated with the different models. While the negative binomial best fits the observed count distribution, we see that the geometric ( $\hat{p} = 0.723$ ) (i.e. the sCMP( $m = 1$ )/CMP model with  $\hat{\lambda} = 0.277$  and  $\hat{\nu} = 0.000$ ) model likewise performs reasonably. More generally, as  $m$  increases, the sCMP class appears to underestimate the number of zeroes and overestimate the number of ones. The estimated frequencies for counts greater than or equal to two, however, appear comparable for all distributions.

To further illustrate the utility of the sCMP family, we consider a condensed representation of the Guttorp (1995) data by summing successive triples of data, thus representing fetal lamb data observed by ultrasound and counted in successive 15-second intervals. Table 4(b) provides the resulting summary information – 75 observations now range in value from 0 to 12, where the dispersion index is now  $3.694/1.147 = 3.221$ , maintaining apparent data over-dispersion. Again, assuming no knowledge regarding the type of the data dispersion, we consider the Poisson, negative binomial, and sCMP distributions at  $m = 1, 2, 3, 4$  and estimate the corresponding model parameters via maximum likelihood





estimation. Table 6 provides the resulting estimation output (including the corresponding log-likelihood, AIC, and BIC) associated with the various distributions considered to model the 15-second movement data summarized in Table 4(b).

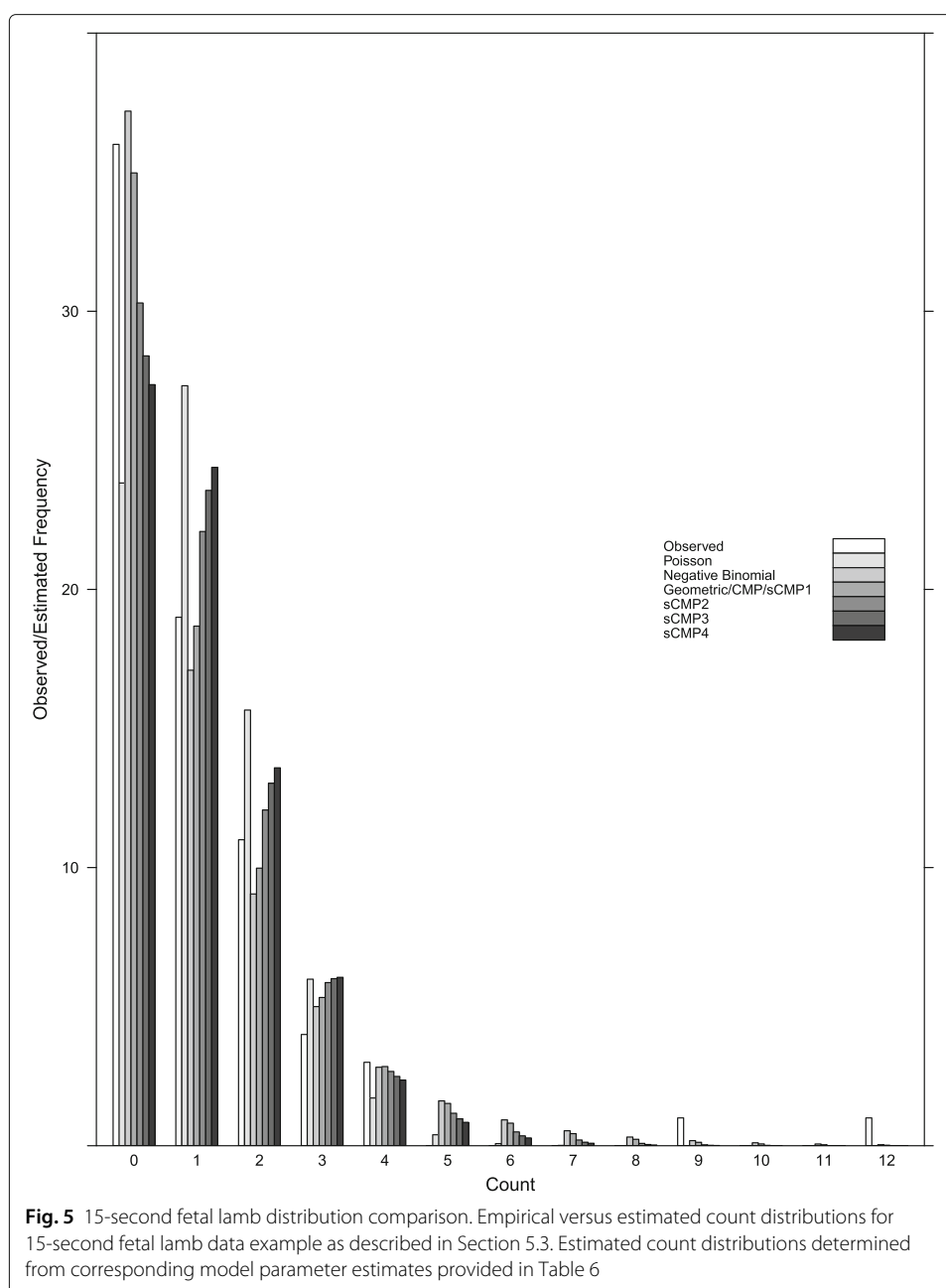
Table 6 again displays an interesting trend with respect to the sCMP class estimators. Again, the estimations for  $\lambda$  decrease as  $m$  increases, while the dispersion parameter consistently estimates to be  $\hat{\nu} = 0$  (indicating consideration of an appropriate negative binomial model structure). Further, as  $m$  increases, the corresponding log-likelihood associated with each of the models decreases. Hence, the optimal sCMP model is again the CMP model (i.e. sCMP when  $m = 1$ ). Meanwhile, the CMP model with estimated dispersion parameter,  $\hat{\nu} = 0$ , again suggests to consider a geometric model with success probability  $1 - \hat{\lambda} = 0.466$ . In fact, the estimated success probability for the geometric model is  $\hat{p} = 0.466$  (0.039). The negative binomial model slightly outperforms the CMP model, although both models perform comparably well, based on their respective AIC values (Burnham and Anderson 2002). The slight outperformance in the negative binomial model relative to the CMP/geometric model (based on log-likelihood comparisons) stems from the negative binomial estimation procedure's allowance for real  $\theta$ , thus obtaining a more precise estimation of the data over-dispersion. However, because we recognize that this special case of the sCMP class where  $m = 1$  and  $\nu = 0$  corresponds to a geometric model, the geometric model is deemed better than the negative binomial model, given the reduction in the number of estimated parameters and thus the reduced AIC and BIC (224.441 and 226.759 for the geometric model, versus 225.819 and 230.454 for the negative binomial model); see Table 6.

Figure 5 provides a comparison of the empirical versus estimated count distributions for the different models associated with the 15-second fetal lamb data. Here, we can see that the geometric( $\hat{p} = 0.466$ ) distribution (i.e. the sCMP( $m = 1$ )/CMP model where  $\hat{\lambda} = 0.534$ ,  $\hat{\nu} = 0.000$ ) best estimates the observed count distribution, given that the geometric model requires only one parameter. Meanwhile, the negative binomial distribution performs comparably well to the geometric/CMP( $\hat{\nu} = 0$ ). This makes sense, given the relationship between the geometric and negative binomial distributions. The estimated geometric and negative binomial distributions are so close because the negative binomial

**Table 6** 15-s fetal lamb data model comparisons

	Est.	SE	log(L)	AIC	BIC
Poisson	$\hat{\mu}_* = 1.147$	(0.124)	-131.3450	264.690	267.010
Geom	$\hat{p} = 0.466$	(0.039)	-111.2206	224.441	226.759
NB	$\hat{\mu} = 1.147$ $\hat{\theta} = 0.767$	(0.195) (0.251)	-110.9094	225.819	230.454
CMP/sCMP( $m = 1$ )	$\hat{\lambda}_* = 0.534$ $\hat{\nu}_* = 0.000$	(0.093) (0.161)	-111.2206	226.442	231.076
sCMP( $m = 2$ )	$\hat{\lambda} = 0.364$ $\hat{\nu} = 0.000$	(0.065) (0.259)	-114.2996	232.599	237.234
sCMP( $m = 3$ )	$\hat{\lambda} = 0.277$ $\hat{\nu} = 0.000$	(0.071) (0.554)	-116.9413	237.882	242.518
sCMP( $m = 4$ )	$\hat{\lambda} = 0.223$ $\hat{\nu} = 0.000$	(0.071) (0.554)	-118.9119	241.824	246.459

Model comparison for the fetal lamb 15-second movement data, where Poisson, negative binomial, CMP, and sCMP (with variable values for  $m$ ) distributions are considered. For model comparisons, the log-likelihood, Akaike and Bayes Information Criteria (AIC and BIC, respectively) are provided



size estimate,  $\hat{\theta} = 0.767$ , is close to one, while the corresponding probability estimate,  $\hat{p} = 0.401$ , is close to that from the geometric model ( $\hat{p} = 0.466$ ).

Notice that sCMP( $m = 3$ ) parameter estimates associated with the 15-second fetal lamb data equal the sCMP( $m = 1$ )/CMP parameter estimates for the 5-second fetal lamb example. This result is logically sound, given the means by which the sCMP distribution is derived; conducting estimations over an interval that is three times its original period is akin to “summing” the three CMP random variables to consider the sCMP model.

This example demonstrates the suitability of the sCMP class to serve as an exploratory tool for count data modeling. For over-dispersed data examples, the negative binomial

distribution is generally expected to be a good model to describe the distribution. The sCMP class of distributions contains the negative binomial (and geometric) distribution as a special case; accordingly, it is not necessarily expected for the sCMP distribution to outperform simpler distributions but rather to demonstrate that the sCMP distribution offers insights regarding model considerations. Indeed, applying the sCMP model to these over-dispersed examples motivated consideration of the geometric distribution, which turned out to be an optimal model. Accordingly, while one may not consider the geometric distribution to be a viable model a priori, the sCMP showed why the geometric model is viable.

## 6 Discussion

The sum-of-Conway-Maxwell-Poissons (sCMP) class of distributions is a flexible construct for modeling count data that captures several well-known distributions as special cases: the Poisson, negative binomial, binomial, geometric, Bernoulli, and Conway-Maxwell-Poisson (CMP). Just as the CMP distribution bridges the gap between the Poisson, geometric, and Bernoulli distributions through the addition of a dispersion parameter, the sCMP distribution sums over  $m$  CMP random variables, producing an encompassing distributional form that has an even greater containment of numerous count distributions.

The provided examples illustrate the flexibility of the sCMP class for handling over- or under-dispersed data. These examples, however, consider only the marginal distribution through unconditional means and variances (and hence unconditional dispersion), thus the true significance of the sCMP class is subdued. In actuality, it is not necessarily straightforward to determine if observed dispersion is true or "apparent". In a regression setting, for example, dispersion is measured via conditional means and variances, and exploratory data analysis may not detect the true complexity of the data (Sellers and Shmueli 2013). Under such circumstances, the sCMP class can aid with detecting dispersion when a more sophisticated approach is required.

As noted in the over-dispersed data example, we are limited in our ability to estimate  $m$  because it is a natural number. We opt for this formulation as it holds true to the form that generalizes the construction of the three special case models (negative binomial, Poisson, and binomial) as sums of their respective special case distributions associated with the CMP distribution (namely, the geometric, Poisson, and Bernoulli models). For example, the negative binomial pmf is often described as the probability of observing  $y$  failures before the  $n$ th success in a series of Bernoulli trials, or as a sum of  $n$  geometric random variables. Yet, the negative binomial distribution can alternatively be derived via a Poisson-gamma mixture, in which case the parameter  $n$  is a real number. As Hilbe (2008) notes, "there is no compelling mathematical reason to limit this parameter to integers." (page 82). Future work considers broadening the sCMP formulation to likewise allow for real-valued  $m$  and any associated implications from such a definition.

While we estimate the standard errors of the parameter estimates via the approximate information matrix as described in Section 4, the sampling distributions associated with  $\lambda$  and  $\nu$  are known to possess skewness (Sellers and Shmueli 2013). Thus, an alternative approach is non-parametric bootstrapping. To compute parameter estimates and associated variation in this manner, one can (for example) randomly draw 1000 samples with

replacement from the data using the `boot` package (Canty and Ripley 2015) in *R* (R Core Team 2017).

#### Acknowledgements

Support for Kimberly Sellers was provided in part by the American Statistical Association (ASA)/National Science Foundation (NSF)/Census Research Program, U. S. Census Bureau Contract #YA1323-14-SE-0122. Support for Andrew W. Swift was provided by a grant from the Simons Foundation (#359536). Support for Kimberly S. Weems was provided in part by the NSF grant #1700235. The authors thank the reviewers and Drs. Darcy S. Morris (U.S. Census Bureau) and Derek Young (University of Kentucky) for helpful comments regarding the manuscript.

#### Authors' contributions

KFS conceived the study, KFS, AS, and KSW contributed to statistical methods and computational development, and analyses. All authors read and approved the final manuscript.

#### Competing interests

The authors declare no competing interests.

#### Publisher's Note

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

#### Author details

<sup>1</sup>Department of Mathematics and Statistics, Georgetown University, 306 St. Mary's Hall, Washington, DC, 20057, USA.

<sup>2</sup>Department of Mathematics, University of Nebraska - Omaha, 6001 Dodge Street, Omaha 68182, NE, USA. <sup>3</sup>Department of Mathematics and Physics, North Carolina Central University, 2211 Mary M. Townes Science Building, 1801 Fayetteville Street, Durham 27707, NC, USA.

Received: 23 December 2016 Accepted: 11 September 2017

Published online: 26 September 2017

#### References

- Bailey, BJR: A model for function word counts. *J. R. Stat. Soc. Ser. C.* **39**(1), 107–114 (1990)
- Boatwright, P, Borle, S, Kadane, JB: A model of the joint distribution of purchase quantity and timing. *J. Am. Stat. Assoc.* **98**, 564–572 (2003)
- 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)
- Borle, S, Boatwright, P, Kadane, JB: The timing of bid placement and extent of multiple bidding: An empirical investigation using ebay online auctions. *Stat. Sci.* **21**(2), 194–205 (2006)
- Borle, S, Dholakia, U, Singh, S, Westbrook, R: The impact of survey participation on subsequent behavior: An empirical investigation. *Mark. Sci.* **26**(5), 711–726 (2007)
- Burnham, KP, Anderson, DR: *Model Selection and Multimodel Inference*. Springer, New York (2002)
- Canty, A, Ripley, B: *boot*: Bootstrap Functions. 1.3-15 edn (2015). <http://cran.r-project.org/web/packages/boot/index.html>
- Casella, G, Berger, RL: *Statistical Inference*, Second Edition. Duxbury, Pacific Grove (2002)
- Conway, RW, Maxwell, WL: A queuing model with state dependent service rates. *J. Ind. Eng.* **12**, 132–136 (1962)
- Gilbert, P, Varadhan, R: *NumDeriv*: Accurate Numerical Derivatives. 2016.8-1 edn (2016). <https://cran.r-project.org/web/packages/numDeriv/index.html>
- Guttorm, P: *Stochastic Modeling of Scientific Data*. Chapman & Hall/CRC, Boca Raton (1995)
- Hilbe, JM: *Negative Binomial Regression*. Cambridge University Press, United Kingdom (2008)
- Kadane, JB, Krishnan, R, Shmueli, G: A data disclosure policy for count data based on the COM-Poisson distribution. *Manag. Sci.* **52**(10), 1610–1617 (2006)
- Kadane, JB: Sums of possibly associated Bernoulli variables: The Conway-Maxwell-Binomial distribution. *Bayesian Anal.* **11**(2), 403–420 (2016)
- Lord, D, Guikema, SD, Geedipally, SR: Application of the Conway-Maxwell-Poisson generalized linear model for analyzing motor vehicle crashes. *Accid. Anal. Prev.* **40**(3), 1123–1134 (2008)
- Minka, TP, Shmueli, G, Kadane, JB, Borle, S, Boatwright, P: *Computing with the COM-Poisson distribution*. Technical Report 776, Dept. of Statistics, Carnegie Mellon University (2003). <http://www.stat.cmu.edu/tr/tr776/tr776.pdf>
- R Core Team: *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna (2017). <https://www.R-project.org/>
- Saghir, A, Lin, Z: Cumulative sum charts for monitoring the COM-Poisson processes. *Comput. Ind. Eng.* **68**, 65–77 (2014)
- Saghir, A, Lin, Z: A flexible and generalized exponentially weighted moving average control chart for count data. *Qual. Reliab. Eng. Int.* **30**(8), 1427–1443 (2014)
- Sellers, KF, Shmueli, G, Borle, S: The COM-Poisson model for count data: a survey of methods and applications. *Appl. Stoch. Model. Bus. Ind.* **28**, 104–116 (2011)
- Sellers, K: A distribution describing differences in count data containing common dispersion levels. *Adv. Appl. Stat. Sci.* **7**(3), 35–46 (2012)
- Sellers, KF, Shmueli, G: A regression model for count data with observation-level dispersion. In: Booth, JG (ed.) *Proceedings of the 24th International Workshop on Statistical Modelling*, pp. 337–344. Cornell University Press, Ithaca, (2009)
- Sellers, KF, Shmueli, G: Data dispersion: Now you see it... now you don't. *Commun. Stat. Theory Methods.* **42**, 1–14 (2013)
- Sellers, KF: A generalized statistical control chart for over- or under-dispersed data. *Qual. Reliab. Eng. Int.* **28**(1), 59–65 (2012)
- Sellers, KF, Raim, A: A flexible zero-inflated model to address data dispersion. *Comput. Stat. Data Anal.* **99**, 68–80 (2016)

- Sellers, KF, Shmueli, G: A flexible regression model for count data. *Ann. Appl. Stat.* **4**(2), 943–961 (2010)
- Sellers, KF, Morris, DS, Balakrishnan, N: Bivariate Conway-Maxwell-Poisson distribution: Formulation, properties, and inference. *J. Multivar. Anal.* **150**, 152–168 (2016)
- Shmueli, G, Minka, TP, Kadane, JB, Borle, S, Boatwright, P: A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Appl. Stat.* **54**, 127–142 (2005)
- Venables, WN, Ripley, BD: *Modern Applied Statistics with S*. 4th edn. Springer, New York (2002)
- Wimmer, G, Köhler, R, Frotjahn, R, Altmann, G: Towards a theory of word length distribution. *J. Quant. Linguist.* **1**(1), 98–106 (1994)
- Zhu, L, Sellers, KF, Morris, DS, Shmuéli, G: Bridging the gap: A generalized stochastic process for count data. *Am. Stat.* **71**(1), 71–80 (2017)

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---