The Conway-Maxwell-Poisson distribution: Distributional theory and approximation

Fraser Daly and Robert E. Gaunt

Department of Actuarial Mathematics and Statistics
and the Maxwell Institute for Mathematical Sciences
Heriot-Watt University,
Edinburgh EH14 4AS, UK.
E-mail address: f.daly@hw.ac.uk
URL: http://www.macs.hw.ac.uk/~fd78

Department of Statistics, University of Oxford
24-29 St. Giles',
Oxford OX1 3LB, UK.
E-mail address: gaunt@stats.ox.ac.uk
URL: http://www.stats.ox.ac.uk/people/research_staff/robert_gaunt

Abstract. The Conway-Maxwell-Poisson (CMP) distribution is a natural two-parameter generalisation of the Poisson distribution which has received some attention in the statistics literature in recent years by offering flexible generalisations of some well-known models. In this work, we begin by establishing some properties of both the CMP distribution and an analogous generalisation of the binomial distribution, which we refer to as the CMB distribution. We also consider some convergence results and approximations, including a bound on the total variation distance between a CMB distribution and the corresponding CMP limit.

1. Introduction

A two-parameter generalisation of the Poisson distribution was introduced by Conway and Maxwell (1962) as the stationary number of occupants of a queuing system with state dependent service or arrival rates. This distribution has since become known as the Conway-Maxwell-Poisson (CMP) distribution. Beginning with the work of Boatwright et al. (2003) and Shmueli et al. (2005), the CMP distribution has received recent attention in the statistics literature on account of the flexibility it offers in statistical models. For example, the CMP distribution

can model data which is either under- or over-dispersed relative to the Poisson distribution. This property is exploited by Sellers and Shmueli (2010), who use the CMP distribution to generalise the Poisson and logistic regression models. Kadane et al. (2006) considered the use of the CMP distribution in Bayesian analysis, and Wu et al. (2013) use the CMP distribution as part of a Bayesian model for spatio-temporal data. The CMP distribution is employed in a flexible cure rate model formulated by Rodrigues et al. (2009) and further analysed by Balakrishnan and Pal (2012).

Our purpose in this work is twofold. Motivated by the use of the CMP distribution in the statistical literature, we firstly aim (in Section 2) to derive explicit distributional properties of the CMP distribution and an analogous generalisation of the binomial distribution, the CMB distribution. Our second aim is to consider the CMP distribution as a limiting distribution. We give conditions under which sums of dependent Bernoulli random variables will converge in distribution to a CMP random variable, and give an explicit bound in total variation distance between the CMB distribution and the corresponding CMP limit. These convergence results are detailed in Sections 3 and 4.

We use the remainder of this section to introduce the CMP and CMB distributions and collect some straightforward properties which will prove useful in the sequel. We also introduce some further definitions that we will need in the work that follows.

1.1. *The CMP distribution.* The CMP distribution is a natural two-parameter generalisation of the Poisson distribution. We will write $X \sim \mathrm{CMP}(\lambda, \nu)$ if

$$\mathbb{P}(X = j) = \frac{1}{Z(\lambda, \nu)} \frac{\lambda^j}{(j!)^\nu}, \qquad j \in \mathbb{Z}^+ = \{0, 1, 2, \ldots\}, \tag{1.1}$$

where $Z(\lambda, \nu)$ is a normalizing constant defined by

$$Z(\lambda, \nu) = \sum_{i=0}^{\infty} \frac{\lambda^i}{(i!)^\nu}.$$

The domain of admissible parameters for which (1.1) defines a probability distribution is $\lambda, \nu > 0$, and $0 < \lambda < 1$, $\nu = 0$.

The introduction of the second parameter ν allows for either sub- or super-linear growth of the ratio $\mathbb{P}(X = j - 1)/\mathbb{P}(X = j)$, and allows X to have variance either less than or greater than its mean. Of course, the mean of $X \sim \mathrm{CMP}(\lambda, \nu)$ is not, in general, λ. In Section 2 we will consider further distributional properties of the CMP distribution, including expressions for its moments.

Clearly, in the case where $\nu = 1$, $X \sim \mathrm{CMP}(\lambda, 1)$ has the Poisson distribution $\mathrm{Po}(\lambda)$ and the normalizing constant $Z(\lambda, 1) = e^\lambda$. As noted by Shmueli et al. (2005), other choices of ν also give rise to well-known distributions. For example, in the case where $\nu = 0$ and $0 < \lambda < 1$, X has a geometric distribution, with $Z(\lambda, 0) = (1 - \lambda)^{-1}$. In the limit $\nu \to \infty$, X converges in distribution to a Bernoulli random variable with mean $\lambda(1 + \lambda)^{-1}$ and $\lim_{\nu \to \infty} Z(\lambda, \nu) = 1 + \lambda$.

In general, of course, the normalizing constant $Z(\lambda, \nu)$ does not permit such a neat, closed-form expression. Asymptotic results are available, however. Gillispie

and Green (2015) prove that, for fixed ν,

$$Z(\lambda, \nu) \sim \frac{\exp\left\{\nu\lambda^{1/\nu}\right\}}{\lambda^{(\nu-1)/2\nu}(2\pi)^{(\nu-1)/2}\sqrt{\nu}} \left(1 + O\left(\lambda^{-1/\nu}\right)\right), \qquad (1.2)$$

as $\lambda \to \infty$, confirming a conjecture made by Shmueli et al. (2005). This asymptotic result may also be used to obtain asymptotic results for the probability generating function of $X \sim \text{CMP}(\lambda, \nu)$, since it may be easily seen that

$$\mathbb{E}s^X = \frac{Z(s\lambda, \nu)}{Z(\lambda, \nu)} . \qquad (1.3)$$

1.2. *The CMB distribution.* Just as the CMP distribution arises naturally as a generalisation of the Poisson distribution, we may define an analogous generalisation of the binomial distribution. We refer to this as the Conway-Maxwell-binomial (CMB) distribution and write that $Y \sim \text{CMB}(n, p, \nu)$ if

$$\mathbb{P}(Y = j) = \frac{1}{C_n} \binom{n}{j}^\nu p^j (1-p)^{n-j}, \qquad j \in \{0, 1, \ldots, n\},$$

where $n \in \mathbb{N} = \{1, 2, \ldots\}$, $0 \leq p \leq 1$ and $\nu \geq 0$. The normalizing constant C_n is defined by

$$C_n = \sum_{i=0}^{n} \binom{n}{i}^\nu p^i (1-p)^{n-i} .$$

The dependence of C_n on p and ν is suppressed for notational convenience. Of course, the case $\nu = 1$ is the usual binomial distribution $Y \sim \text{Bin}(n, p)$, with normalizing constant $C_n = 1$. Shmueli et al. (2005) considered the CMB distribution and derived some of its basic properties, referring to it as the CMP-binomial distribution. We, however, consider it more natural to refer to this as the CMB distribution (a similar convention is also followed by Kadane (2016)); we shall also later refer to an analogous generalisation of the Poisson binomial distribution as the CMPB distribution.

There is a simple relationship between CMP and CMB random variables, which generalises a well-known result concerning Poisson and binomial random variables. If $X_1 \sim \text{CMP}(\lambda_1, \nu)$ and $X_2 \sim \text{CMP}(\lambda_2, \nu)$ are independent, then $X_1 \mid X_1 + X_2 = n \sim \text{CMB}(n, \lambda_1/(\lambda_1 + \lambda_2), \nu)$ (see Shmueli et al., 2005).

It was also noted by Shmueli et al. (2005) that $Y \sim \text{CMB}(n, p, \nu)$ may be written as a sum of exchangeable Bernoulli random variables $Z_1, \ldots, Z_n$ satisfying

$$\mathbb{P}(Z_1 = z_1, \ldots, Z_n = z_n) = \frac{1}{C_n} \binom{n}{k}^{\nu-1} p^k (1-p)^{n-k}, \qquad (1.4)$$

where $k = z_1 + \cdots + z_n$. Note that $\mathbb{E}Z_1 \neq p$ in general, unless $\nu = 1$. However, $\mathbb{E}Z_1 = n^{-1}\mathbb{E}Y$ may be either calculated explicitly or estimated using some of the properties of the CMB distribution to be discussed in the sequel.

From the mass functions given above, it can be seen that if $Y \sim \text{CMB}(n, \lambda/n^\nu, \nu)$, then Y converges in distribution to $X \sim \text{CMP}(\lambda, \nu)$ as $n \to \infty$. We return to this convergence in Section 3 below, where we give an explicit bound on the total variation distance between these distributions.

1.3. *Power-biasing.* In what follows, we will need the definition of power-biasing, as used by Peköz et al. (2016). For any non-negative random variable W with finite ν-th moment, we say that $W^{(\nu)}$ has the ν-power-biased distribution of W if

$$(\mathbb{E}W^\nu)\mathbb{E}f(W^{(\nu)}) = \mathbb{E}\left[W^\nu f(W)\right] , \qquad (1.5)$$

for all $f : \mathbb{R}^+ \mapsto \mathbb{R}$ such that the expectations exist. In this paper, we will be interested in the case that W is non-negative and integer-valued. In this case, the mass function of $W^{(\nu)}$ is given by

$$\mathbb{P}(W^{(\nu)} = j) = \frac{j^\nu \mathbb{P}(W = j)}{\mathbb{E}W^\nu} , \qquad j \in \mathbb{Z}^+ .$$

Properties of a large family of such transformations, of which power-biasing is a part, are discussed by Goldstein and Reinert (2005). The case $\nu = 1$ is the usual size-biasing, which has often previously been employed in conjunction with the Poisson distribution: see Barbour et al. (1992); Daly et al. (2012); Daly and Johnson (2013), and references therein for some examples. The power-biasing we employ here is the natural generalisation of size-biasing that may be applied in the CMP case.

2. Distributional properties of the CMP and CMB distributions

In this section we collect some distributional properties of the CMP and CMB distributions. Some will be required in the sequel when considering approximations and convergence to the CMP distribution, and all of are of some interest, either independently or for statistical applications.

2.1. *Moments, cumulants, and related results.* We begin this section by noting, in Proposition 2.1 below, that some moments of the CMP distribution may be easily and explicitly calculated. The simple formula $\mathbb{E}X^\nu = \lambda$ was already known to Sellers and Shmueli (2010). We also note the corresponding result for the CMB distribution.

Here and in the sequel we let

$$(j)_r = j(j-1)\cdots(j-r+1)$$

denote the falling factorial.

Proposition 2.1.   (i). *Let $X \sim CMP(\lambda, \nu)$, where $\lambda, \nu > 0$. Then*

$$\mathbb{E}[((X)_r)^\nu] = \lambda^r ,$$

  for $r \in \mathbb{N}$.
  (ii). *Let $Y \sim CMB(n, p, \nu)$, where $\nu > 0$. Then*

$$\mathbb{E}[((Y)_r)^\nu] = \frac{C_{n-r}}{C_n}((n)_r)^\nu p^r ,$$

  for $r = 1, \ldots, n-1$.

Proof: We have

$$\mathbb{E}[((X)_r)^\nu] = \frac{1}{Z(\lambda, \nu)} \sum_{k=0}^{\infty} ((k)_r)^\nu \frac{\lambda^k}{(k!)^\nu} = \frac{\lambda^r}{Z(\lambda, \nu)} \sum_{k=r}^{\infty} \frac{\lambda^{k-r}}{((k-r)!)^\nu}$$

$$= \frac{\lambda^r}{Z(\lambda, \nu)} \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu} = \lambda^r,$$

and

$$\mathbb{E}[((Y)_r)^\nu] = \frac{1}{C_n} \sum_{k=0}^{n} ((k)_r)^\nu \binom{n}{k}^\nu p^k (1-p)^{n-k}$$

$$= \frac{1}{C_n} \left(\frac{n!}{(n-r)!} \right)^\nu \sum_{k=r}^{n} \binom{n-r}{k-r}^\nu p^k (1-p)^{n-k}$$

$$= \frac{1}{C_n} ((n)_r)^\nu p^r \sum_{j=0}^{n-r} \binom{n-r}{j}^\nu p^j (1-p)^{n-r-j} = \frac{C_{n-r}}{C_n} ((n)_r)^\nu p^r.$$

$\square$

Remark 2.2. It is well-known that the factorial moments of $Z \sim \mathrm{Po}(\lambda)$ are given by $\mathbb{E}[(X)_r] = \lambda^r$. We therefore have the attractive formula $\mathbb{E}[((X)_r)^\nu] = \mathbb{E}[(Z)_r]$, for $X \sim \mathrm{CMP}(\lambda, \nu)$.

Such simple expressions do not exist for moments of $X \sim \mathrm{CMP}(\lambda, \nu)$ which are not of the form $\mathbb{E}[((X)_r)^\nu]$. Instead, we use (1.2) to give asymptotic expressions for such moments.

Proposition 2.3. *Let $X \sim CMP(\lambda, \nu)$. Then, for $k \in \mathbb{N}$,*

$$\mathbb{E}X^k \sim \lambda^{k/\nu} \left(1 + O\left(\lambda^{-1/\nu} \right) \right),$$

as $\lambda \to \infty$.

Proof: It is clear that, for $k \in \mathbb{N}$,

$$\mathbb{E}[(X)_k] = \frac{\lambda^k}{Z(\lambda, \nu)} \frac{\partial^k}{\partial \lambda^k} Z(\lambda, \nu).$$

Differentiating (1.2) (see Remark 2.4 for a justification) we have that

$$\frac{\partial^k}{\partial \lambda^k} Z(\lambda, \nu) \sim \lambda^{k/\nu - k} \cdot \frac{\exp\left\{ \nu \lambda^{1/\nu} \right\}}{\lambda^{(\nu-1)/2\nu} (2\pi)^{(\nu-1)/2} \sqrt{\nu}} \left(1 + O\left(\lambda^{-1/\nu} \right) \right), \quad (2.1)$$

as $\lambda \to \infty$, and hence

$$\mathbb{E}[(X)_k] \sim \lambda^{k/\nu} \left(1 + O\left(\lambda^{-1/\nu} \right) \right),$$

as $\lambda \to \infty$. We now exploit the following connection between moments and factorial moments:

$$\mathbb{E}X^k = \sum_{r=1}^{k} \left\{ {k \atop r} \right\} \mathbb{E}[(X)_r], \quad (2.2)$$

for $k \in \mathbb{N}$, where the Stirling numbers of the second kind $\left\{ {k \atop r} \right\}$ are given by $\left\{ {k \atop r} \right\} = \frac{1}{r!} \sum_{j=0}^{r} (-1)^{r-j} \binom{r}{j} j^k$ (see Olver et al., 2010). Using (2.2), and noting that $\left\{ {k \atop k} \right\} = 1$, completes the proof. $\square$

Remark 2.4. In the above proof, we differentiated the asymptotic formula (1.2) in the naive sense by simply differentiating the leading term k times. We shall also do this below in deriving the variance formula (2.4), and in Proposition 2.6, in which we differentiate an asymptotic series for $\log(Z(\lambda e^t, \nu))$ with respect to t in an analogous manner. However, as noted by Hinch (1991, p. 23), asymptotic approximations cannot be differentiated in this manner in general. Fortunately, in the case of the asymptotic expansion (1.2) for $Z(\lambda, \nu)$ we can do so. This is because we have the following asymptotic formula for the CMP normalising constant that is more precise than (1.2). For fixed ν,

$$Z(\lambda, \nu) \sim \frac{\exp\left\{\nu\lambda^{1/\nu}\right\}}{\lambda^{(\nu-1)/2\nu}(2\pi)^{(\nu-1)/2}\sqrt{\nu}}\left(1 + \sum_{k=1}^{\infty} a_k \lambda^{-k/\nu}\right), \qquad (2.3)$$

as $\lambda \to \infty$, where the a_k are constants that do not involve λ. The m-th derivative of the asymptotic series (2.3) is dominated by the m-th derivative of the leading term of (2.3), meaning that one can naively differentiate the asymptotic series, as we did in the proof of Proposition 2.3.

The leading term in the asymptotic expansion (2.3) was obtained for integer ν by Shmueli et al. (2005), and then for all $\nu > 0$ by Gillispie and Green (2015). When stating their results, Shmueli et al. (2005) and Gillispie and Green (2015) did not include the lower order term $\sum_{k=1}^{\infty} a_k \lambda^{-k/\nu}$, but it can be easily read off from their analysis. For integer ν, Shmueli et al. (2005) gave an integral representation for $Z(\lambda, \nu)$ and then applied Laplace's approximation to write down the leading order term in its asymptotic expansion. Laplace's approximation gives that (see Shun and McCullagh, 1995, p. 750), for infinitely differentiable $g : \mathbb{R}^d \to \mathbb{R}$,

$$\int_{\mathbb{R}^d} \exp\{-ng(x)\}\, dx \sim \left(\frac{n\det(\hat{g}'')}{2\pi}\right)^{-1/2} \exp\{-n\hat{g}\}\left(1 + \sum_{k=1}^{\infty} b_k n^{-k}\right), \quad \text{as } n \to \infty\,,$$

where the b_k do not involve n, and $\hat{g}$ and $\hat{g}''$ denote g and the matrix of second order derivatives of g, respectively, evaluated at the value $\hat{x}$ that minimises g. It is now clear that the lower order term in (2.3) has the form $\sum_{k=1}^{\infty} a_k \lambda^{-k/\nu}$. For general $\nu > 0$, Gillispie and Green (2015) obtained an expression for the leading term in the asymptotic expansion of $Z(\lambda, \nu)$ by using Laplace's approximation, as well as several other simpler asymptotic approximations. In each of these approximations, the lower order term is of the form $\sum_{k=1}^{\infty} c_k \lambda^{-k/\nu}$, from which it follows that $Z(\lambda, \nu)$ has an asymptotic expansion of the form (2.3).

We also have the following relationship between moments of $X \sim \mathrm{CMP}(\lambda, \nu)$:

$$\mathbb{E}X^{r+1} = \lambda\frac{d}{d\lambda}\mathbb{E}X^r + \mathbb{E}X\mathbb{E}X^r\,,$$

for $r > 0$. See equation (6) of Shmueli et al. (2005). With $r = 1$ we obtain

$$\mathrm{Var}(X) = \lambda\frac{d}{d\lambda}\mathbb{E}X \sim \frac{1}{\nu}\lambda^{1/\nu} + O(1)\,, \qquad (2.4)$$

as $\lambda \to \infty$, from Proposition 2.3. This also gives the following corollary.

Corollary 2.5. *Let m be the median of $X \sim CMP(\lambda, \nu)$. Then*

$$m \sim \lambda^{1/\nu} + O\left(\lambda^{1/2\nu}\right)\,,$$

as $\lambda \to \infty$.

Proof: From above, $\mathbb{E}X \sim \lambda^{1/\nu} + O(1)$ and $\sigma = \sqrt{\text{Var}(X)} \sim \frac{1}{\sqrt{\nu}}\lambda^{1/2\nu} + O(1)$. Since $\sigma < \infty$, we may use a result of Mallows (1991), who showed that $|\mathbb{E}X - m| \leq \sigma$. The result follows. $\qquad\square$

As with the moments above, we may also find asymptotic expressions for the cumulants of the CMP distribution.

Proposition 2.6. *For $n \geq 1$, let κ_n be the nth cumulant of $X \sim CMP(\lambda, \nu)$. Then*

$$\kappa_n \sim \frac{1}{\nu^{n-1}}\lambda^{1/\nu} + O(1),$$

as $\lambda \to \infty$.

Proof: From (1.3), the cumulant generating function of $X \sim \text{CMP}(\lambda, \nu)$ is

$$g(t) = \log(\mathbb{E}[e^{tX}]) = \log(Z(\lambda e^t, \nu)) - \log(Z(\lambda, \nu)).$$

The cumulants are given by

$$\kappa_n = g^{(n)}(0) = \frac{\partial^n}{\partial t^n} \log(Z(\lambda e^t, \nu))\Big|_{t=0}.$$

From (1.2),

$$\log(Z(\lambda e^t, \nu)) \sim \nu \lambda^{1/\nu} e^{t/\nu},$$

as $\lambda \to \infty$. The expression for the leading term in the asymptotic expansion of κ_n now easily follows, and a straightforward analysis, which is omitted, shows that the second term is $O(1)$ for all $n \geq 1$. The result now follows. $\qquad\square$

Note that as a corollary to this result, the skewness γ_1 of $X \sim \text{CMP}(\lambda, \nu)$ satisfies

$$\gamma_1 = \frac{\kappa_3}{\sigma^3} \sim \frac{1}{\sqrt{\nu}}\lambda^{-1/2\nu} + O\left(\lambda^{-3/2\nu}\right),$$

as $\lambda \to \infty$, where $\sigma^2 = \text{Var}(X) \sim \frac{1}{\nu}\lambda^{1/\nu}$ from (2.4). Similarly, the excess kurtosis γ_2 of X satisfies

$$\gamma_2 = \frac{\kappa_4}{\sigma^4} \sim \frac{1}{\nu}\lambda^{-1/\nu} + O\left(\lambda^{-2/\nu}\right),$$

as $\lambda \to \infty$. For comparison, recall that in the Poisson case ($\nu = 1$), $\gamma_1 = \lambda^{-1/2}$ and $\gamma_2 = \lambda^{-1}$.

We conclude this section with two further results of a similar flavour. We begin by giving expressions for the modes of the CMP and CMB distributions. The expression for the mode of CMP distribution for non-integral $\lambda^{1/\nu}$ was known to Guikema and Goffelt (2008), but for clarity and completeness we state the result and give the simple proof. The expression for the mode of the CMB distribution is new. Here and in the sequel we will let $\lfloor \cdot \rfloor$ denote the floor function.

Proposition 2.7. (i). *Let $X \sim CMP(\lambda, \nu)$. Then the mode of X is $\lfloor \lambda^{1/\nu} \rfloor$ if $\lambda^{1/\nu}$ is not an integer. Otherwise, the modes of X are $\lambda^{1/\nu}$ and $\lambda^{1/\nu} - 1$.*
(ii). *Let $Y \sim CMB(n, p, \nu)$ and define*

$$a = \frac{n+1}{1 + \left(\frac{1-p}{p}\right)^{1/\nu}}.$$

Then the mode of Y is $\lfloor a \rfloor$ if a is not an integer. Otherwise, the modes of Y are a and $a - 1$.

Proof: (i). Writing

$$\mathbb{P}(X = j) = \frac{1}{Z(\lambda, \nu)} \left(\frac{(\lambda^{1/\nu})^j}{j!} \right)^\nu ,$$

the result now follows as in the Poisson case, for which the result is well-known.

(ii). This is a straightforward generalisation of the derivation of the mode of a binomial distribution given by Kaas and Buhrman (1980). Consideration of the ratio

$$\frac{\mathbb{P}(Y = k + 1)}{\mathbb{P}(Y = k)} = \left(\frac{n - k}{k + 1} \right)^\nu \frac{p}{1 - p} = \left(\frac{n - k}{k + 1} \left(\frac{p}{1 - p} \right)^{1/\nu} \right)^\nu ,$$

shows that $\mathbb{P}(Y = k)$ increases as a function of k if $k < a$ and decreases for $k > a - 1$. Therefore, if a is not an integer, $\mathbb{P}(Y = k)$ increases for $k \leq \lfloor a \rfloor$ and decreases for $k \geq \lfloor a \rfloor$, giving $\lfloor a \rfloor$ as the mode. If a is an integer then $\mathbb{P}(Y = k)$ increases for $k \leq a - 1$ and decreases for $k \geq a$ and so $a - 1$ and a are neighbouring modes.

$\square$

We also give, in Proposition 2.8 below, an expression for the mean deviation of X^ν, where $X \sim \mathrm{CMP}(\lambda, \nu)$, as usual. This generalises a result of Crow (1958), who showed that if $Z \sim \mathrm{Po}(\lambda)$ has a Poisson distribution, then

$$\mathbb{E}|Z - \lambda| = 2e^{-\lambda} \frac{\lambda^{\lfloor \lambda \rfloor + 1}}{\lfloor \lambda \rfloor!} = 2\lambda S ,$$

where $S = e^{-\lambda} \frac{\lambda^{\lfloor \lambda \rfloor + 1}}{\lfloor \lambda \rfloor!}$ is the maximum value of the mass function of Z.

Proposition 2.8. *Let $X \sim CMP(\lambda, \nu)$. Then*

$$\mathbb{E}|X^\nu - \lambda| = 2Z(\lambda, \nu)^{-1} \frac{\lambda^{\lfloor \lambda^{1/\nu} \rfloor + 1}}{\lfloor \lambda^{1/\nu} \rfloor!} = 2\lambda T ,$$

where $T = Z(\lambda, \nu)^{-1} \frac{\lambda^{\lfloor \lambda^{1/\nu} \rfloor + 1}}{\lfloor \lambda^{1/\nu} \rfloor!}$, the maximum value of the mass function of X.

Proof:

$$\mathbb{E}|X^\nu - \lambda| = \sum_{k=0}^\infty |k^\nu - \lambda| Z(\lambda, \nu)^{-1} \frac{\lambda^k}{(k!)^\nu}$$

$$= Z(\lambda, \nu)^{-1} \left[\sum_{k=0}^{\lfloor \lambda^{1/\nu} \rfloor} (\lambda - k^\nu) \frac{\lambda^k}{(k!)^\nu} + \sum_{k=\lfloor \lambda^{1/\nu} \rfloor + 1}^\infty (k^\nu - \lambda) \frac{\lambda^k}{(k!)^\nu} \right]$$

$$= 2Z(\lambda, \nu)^{-1} \frac{\lambda^{\lfloor \lambda^{1/\nu} \rfloor + 1}}{\lfloor \lambda^{1/\nu} \rfloor!} .$$

$\square$

2.2. *Characterisations.* In Section 3, we will use Stein's method for probability approximations (see, for example, Stein, 1972; Chen, 1975) to give bounds on the convergence of the CMB distribution to a suitable CMP limit. Stein's method relies on linear operators characterising distributions of interest. In the following lemma, we present such characterisations for the CMP and CMB distributions. These will also prove useful in deriving several other properties of these distributions in the work that follows.

Lemma 2.9. *We have the following characterisations for the CMP and CMB distributions.*

(i). *Let $X \sim CMP(\lambda, \nu)$, and suppose that $f : \mathbb{Z}^+ \mapsto \mathbb{R}$ is such that $\mathbb{E}|f(X + 1)| < \infty$ and $\mathbb{E}|X^\nu f(X)| < \infty$. Then*

$$\mathbb{E}[\lambda f(X + 1) - X^\nu f(X)] = 0. \qquad (2.5)$$

Conversely, suppose now that W is a real-valued random variable supported on $\mathbb{Z}^+$ such that $\mathbb{E}[\lambda f(W + 1) - W^\nu f(W)] = 0$ for all bounded $f : \mathbb{Z}^+ \mapsto \mathbb{R}$. Then $W \sim CMP(\lambda, \nu)$.

(ii). *Let $Y \sim CMB(n, p, \nu)$, and suppose that $f : \mathbb{Z}^+ \mapsto \mathbb{R}$ is such that $\mathbb{E}|f(Y + 1)| < \infty$ and $\mathbb{E}|Y^\nu f(Y)| < \infty$. Then*

$$\mathbb{E}[p(n - Y)^\nu f(Y + 1) - (1 - p)Y^\nu f(Y)] = 0. \qquad (2.6)$$

Proof: The characterising equations (2.5) and (2.6) may be obtained directly from straightforward manipulations, or from the work of Brown and Xia (2001), for example, who consider such characterisations in the more general setting of the equilibrium distribution of a birth-death process. To prove the converse statement for the CMP distribution in part (i), we consider the so-called Stein equation for the CMP distribution:

$$I(x \in A) - \mathbb{P}(X \in A) = \lambda f_A(x + 1) - x^\nu f_A(x), \qquad (2.7)$$

where $X \sim \mathrm{CMP}(\lambda, \nu)$, $A \subseteq \mathbb{Z}^+$ and $f_A : \mathbb{Z}^+ \mapsto \mathbb{R}$. In Lemma 3.7 below, we will obtain the unique solution to (2.7) and prove that it is bounded. Since f_A is bounded, evaluating both sides of (2.7) at W and taking expectations gives that, for any $A \subseteq \mathbb{Z}^+$,

$$\mathbb{P}(W \in A) - \mathbb{P}(X \in A) = \mathbb{E}[\lambda f_A(W + 1) - W^\nu f_A(W)] = 0,$$

from which it follows that $W \sim \mathrm{CMP}(\lambda, \nu)$. $\qquad\qquad \square$

We do not give a converse statement, and thus a complete characterisation, for the CMB distribution. We do not need the converse in this paper, and giving a proof analogous given to that for CMP distribution would be tedious, because we would need to solve the corresponding CMB Stein equation and then derive bounds for the solution.

2.3. *Stochastic ordering and related results.* In this section we will explore some properties of CMP and CMB distributions that may be obtained by considering various stochastic orderings. In particular, we will use the usual stochastic order and the convex order. For random variables U and V, we say that U is smaller than V in the usual stochastic order (which we denote $U \leq_{st} V$) if $\mathbb{E}f(U) \leq \mathbb{E}f(V)$ for all increasing functions f. Equivalently, $U \leq_{st} V$ if $\mathbb{P}(U > t) \leq \mathbb{P}(V > t)$ for all t. For random variables U and V with $\mathbb{E}U = \mathbb{E}V$, we will say that U is smaller

than V in the convex order (written $U \leq_{cx} V$) if $\mathbb{E}f(U) \leq \mathbb{E}f(V)$ for all convex functions f. Many further details on these orderings may be found in the book by Shaked and Shanthikumar (2007), for example.

We begin with two lemmas that make use of the power biasing introduced in Section 1.3.

Lemma 2.10. *Let W be a non-negative random variable and $0 \leq \alpha < \beta$. Suppose that $\mathbb{E}W^\alpha$ and $\mathbb{E}W^\beta$ exist. Then $W^{(\alpha)} \leq_{st} W^{(\beta)}$.*

Proof: It is easily checked using the definition (1.5) that for $\alpha, \delta \geq 0$, $\left(W^{(\alpha)}\right)^{(\delta)} =_{st} W^{(\alpha+\delta)}$. Taking $\delta = \beta - \alpha > 0$, it therefore suffices to prove the lemma with $\alpha = 0$. That is, we need to show that $\mathbb{E}f(W) \leq \mathbb{E}f(W^{(\beta)})$ for all $\beta > 0$ and all increasing functions f. From the definition (1.5) this is immediate, since $\mathrm{Cov}(W^\beta, f(W)) \geq 0$ for all increasing f. $\square$

Lemma 2.11. *Let $Y \sim CMB(n, p, \nu)$. Then $Y^{(\nu)} \leq_{st} Y + 1$.*

Proof: Note that

$$\mathbb{P}(Y + 1 = j) = \frac{1}{C_n}\binom{n}{j-1}^\nu p^{j-1}(1-p)^{n-j+1}, \quad j = 1, \dots, n+1;$$

$$\mathbb{P}(Y^{(\nu)} = j) = \frac{1}{C_n\mathbb{E}Y^\nu}\frac{p(n-j+1)^\nu}{1-p}\binom{n}{j-1}^\nu p^{j-1}(1-p)^{n-j+1}, \quad j = 1, \dots, n.$$

Hence, the condition that $\mathbb{E}f(Y^{(\nu)}) \leq \mathbb{E}f(Y+1)$ (for all increasing $f : \mathbb{Z}^+ \mapsto \mathbb{R}$) is equivalent to the non-negativity of

$$\mathbb{E}\left[f(Y+1)\left\{1 - \frac{p(n-Y)^\nu}{(1-p)\mathbb{E}Y^\nu}\right\}\right], \tag{2.8}$$

for all f increasing. Noting that, by Lemma 2.9 (ii), $\mathbb{E}\left[1 - \frac{p(n-Y)^\nu}{(1-p)\mathbb{E}Y^\nu}\right] = 0$, (2.8) is the covariance of two increasing functions, and is hence non-negative. $\square$

2.3.1. *Ordering results for CMP distributions.* Throughout this section, let $X \sim \mathrm{CMP}(\lambda, \nu)$. It is clear that $X^{(\nu)} =_{st} X + 1$ and hence, for $\nu \geq 1$, Lemma 2.10 gives $X^{(1)} \leq_{st} X + 1$. This is the negative dependence condition employed by Daly et al. (2012). Some consequences of this stochastic ordering are given in Proposition 2.12 below. Before we can state these, we define the total variation distance between non-negative, integer-valued random variables U and V:

$$d_{TV}(\mathcal{L}(U), \mathcal{L}(V)) = \sup_{A \subseteq \mathbb{Z}^+} |\mathbb{P}(U \in A) - \mathbb{P}(V \in A)|.$$

We will also need to define the Poincaré (inverse spectral gap) constant R_U for a non-negative, integer-valued random variable U:

$$R_U = \sup_{g \in \mathcal{G}(U)}\left\{\frac{\mathbb{E}[g(U)^2]}{\mathbb{E}[\{g(U+1) - g(U)\}^2]}\right\},$$

where the supremum is take over the set

$$\mathcal{G}(U) = \left\{g : \mathbb{Z}^+ \mapsto \mathbb{R} \text{ with } \mathbb{E}[g(U)^2] < \infty \text{ and } \mathbb{E}g(U) = 0\right\}.$$

Proposition 2.12. *Let $X \sim CMP(\lambda, \nu)$ with $\nu \geq 1$. Let $\mu = \mathbb{E}X$. Then*

(i).

$$d_{TV}(\mathcal{L}(X), Po(\mu)) \leq \frac{1}{\mu}(\mu - Var(X)) \sim \frac{\nu - 1}{\nu} + O\left(\lambda^{-1/\nu}\right),$$

as $\lambda \to \infty$.

(ii).

$$Var(X) \leq R_X \leq \mu.$$

(iii). $X \leq_{cx} Z$, where $Z \sim Po(\mu)$. In particular,

$$\mathbb{P}(X \geq \mu + t) \leq e^t \left(1 + \frac{t}{\mu}\right)^{-(\mu+t)},$$

$$\mathbb{P}(X \leq \mu - t) \leq e^{-t} \left(1 - \frac{t}{\mu}\right)^{t-\mu},$$

where the latter bound applies if $t < \mu$.

Proof: The upper bound in (i) follows from Proposition 3 of Daly et al. (2012). The asymptotic behaviour of the upper bound is a consequence of our Proposition 2.3 and (2.4). In (ii), the lower bound is standard and the upper bound is from Theorem 1.1 of Daly and Johnson (2013). (iii) follows from Theorem 2.2 and Corollary 2.8 of Daly (2016+). □

On the other hand, if $\nu < 1$ we have that $X + 1 \leq_{st} X^{(\nu)}$. In that case, Proposition 3 of Daly et al. (2012) gives the upper bound

$$d_{TV}(\mathcal{L}(X), Po(\mu)) \leq \frac{1}{\mu}(Var(X) - \mu) \sim \frac{1 - \nu}{\nu} + O(\lambda^{-\nu}),$$

as $\lambda \to \infty$, where $\mu = \mathbb{E}X$ and the asymptotics of the upper bound again follow from our Proposition 2.3 and (2.4).

However, in the case $\nu < 1$ we cannot adapt the proof of Theorem 1.1 of Daly and Johnson (2013) to give an analogue of Proposition 2.12 (ii). By suitably modifying the proof of Theorem 2.2 of Daly (2016+), we may note that $Z \leq_{cx} X$ in this case, where $Z \sim Po(\mu)$. There is, however, no concentration inequality corresponding to that given in Proposition 2.12 (iii).

2.3.2. *Ordering results for CMB distributions.* Now let $Y \sim CMB(n, p, \nu)$. In the case $\nu \geq 1$ we may combine Lemmas 2.10 and 2.11 to see that $Y^{(1)} \leq_{st} Y + 1$. That is, the negative dependence condition holds. We thus have the following analogue of Proposition 2.12, which may be proved in the same way as that result.

Proposition 2.13. *Let $Y \sim CMB(n, p, \nu)$ with $\nu \geq 1$. Let $\mu = \mathbb{E}Y$. Then*

(i).

$$d_{TV}(\mathcal{L}(Y), Po(\mu)) \leq \frac{1}{\mu}(\mu - Var(Y)).$$

(ii).

$$Var(Y) \leq R_Y \leq \mu.$$

(iii). $Y \leq_{cx} X$, where $X \sim Po(\mu)$. In particular,

$$\mathbb{P}(Y \geq \mu + t) \leq e^t \left(1 + \frac{t}{\mu}\right)^{-(\mu+t)} ,$$

$$\mathbb{P}(Y \leq \mu - t) \leq e^{-t} \left(1 - \frac{t}{\mu}\right)^{t-\mu} ,$$

where the latter bound applies if $t < \mu$.

There is no corresponding result in the case $\nu < 1$, since the stochastic ordering $Y^{(\nu)} \leq_{st} Y + 1$ holds regardless of the sign of $1 - \nu$, and so we cannot use our previous lemmas to make a stochastic comparison between $Y^{(1)}$ and $Y + 1$ when $Y \sim \mathrm{CMB}(n, p, \nu)$ with $\nu < 1$.

2.4. *Example: $\nu = 2$.* For some illustration of our results, consider the case $\nu = 2$ and let $X \sim \mathrm{CMP}(\lambda, 2)$. Let $I_r(x)$ be the modified Bessel function of the first kind defined by

$$I_r(x) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(r + k + 1)} \left(\frac{x}{2}\right)^{r+2k} .$$

Note that, by definition, the normalizing constant $Z(\lambda, 2) = I_0(2\sqrt{\lambda})$. Hence, for $m \in \mathbb{N}$,

$$\mathbb{E}[(X)_m] = \frac{1}{I_0(2\sqrt{\lambda})} \sum_{k=m}^{\infty} \frac{\lambda^k}{k!(k-m)!}$$

$$= \frac{\lambda^{m/2} I_m(2\sqrt{\lambda})}{I_0(2\sqrt{\lambda})} .$$

From (2.2) we therefore have

$$\mathbb{E}X^m = \sum_{k=1}^{m} \left\{{m \atop k}\right\} \frac{\lambda^{k/2} I_k(2\sqrt{\lambda})}{I_0(2\sqrt{\lambda})} .$$

In particular, the mean of X is given by

$$\mathbb{E}X = \frac{\sqrt{\lambda} I_1(2\sqrt{\lambda})}{I_0(2\sqrt{\lambda})} .$$

Also, since $\mathbb{E}X^2 = \lambda$, the variance is given by

$$\mathrm{Var}(X) = \lambda \left(1 - \frac{I_1(2\sqrt{\lambda})^2}{I_0(2\sqrt{\lambda})^2}\right) .$$

Formulas for the cumulants, skewness and excess kurtosis of X can also be obtained, but their expressions are more complicated and are omitted.

Note that the asymptotic formula $I_r(x) \sim \frac{1}{\sqrt{2\pi x}} e^x$ as $x \to \infty$ (see Olver et al., 2010) easily allows one to to verify (2.4) in this case. Writing $\mathrm{Var}(X) = \mathbb{E}[X(X - 1)] + \mathbb{E}X - (\mathbb{E}X)^2$, the Turán inequality $I_r(x)^2 > I_{r+1}(x)I_{r-1}(x)$ (see Amos, 1974) also allows direct verification that $\mathrm{Var}(X) < \mathbb{E}X$, which follows from the convex

ordering in Proposition 2.12. The total variation bound in that same result may be expressed as

$$d_{TV}(\mathcal{L}(X), \mathrm{Po}(\mathbb{E}X)) \leq \sqrt{\lambda} \left(\frac{I_1(2\sqrt{\lambda})}{I_0(2\sqrt{\lambda})} - \frac{I_2(2\sqrt{\lambda})}{I_1(2\sqrt{\lambda})} \right).$$

3. Convergence and approximation for CMB distributions

In this section we will use Stein's method for probability approximation to derive an explicit bound on the convergence of $Y \sim \mathrm{CMB}(n, \lambda/n^\nu, \nu)$ to $X \sim \mathrm{CMP}(\lambda, \nu)$ as $n \to \infty$. This convergence is the analogue of the classical convergence of the binomial distribution to a Poisson limit, which corresponds to the case $\nu = 1$ here.

Stein's method was first developed by Stein (1972) in the context of normal approximation. The same techniques were applied to Poisson approximation by Chen (1975). An account of the method for Poisson approximation, together with a wealth of examples, is given by Barbour et al. (1992). Stein's method has also found a large number of applications beyond the classical normal and Poisson approximation settings. For an introduction to Stein's method and discussion of its wide applicability, the reader is referred to Barbour and Chen (2005) and references therein.

For future use, we define

$$g_\nu(\lambda) = \begin{cases} \min\left\{1, \max\left\{1 + \frac{1}{\nu}, \left(\frac{3}{2}\right)^\nu\right\} \left(1 - \lambda^{-1/2\nu}\right)^{1/\nu - 1} \lambda^{1/2\nu - 1}\right\} & \text{if } \nu \geq 1, \lambda > 1, \\ \left(1 + \frac{1}{\nu}\right) \left(1 + \lambda^{-1/2\nu}\right)^{1/\nu - 1} \lambda^{1/2\nu - 1} & \text{if } \nu \leq 1, \lambda \geq 1, \\ 1 & \text{if } \nu \geq 1, \lambda \leq 1, \\ \left(1 - \lambda^{1-\nu}\right)^{-1} & \text{if } \nu < 1, \lambda < 1. \end{cases}$$

We use much of the remainder of this section to prove the following.

Theorem 3.1. *Let $Y \sim CMB(n, \lambda/n^\nu, \nu)$ for some $0 < \lambda < n^\nu$, and let $X \sim CMP(\lambda, \nu)$. Then*

$$d_{TV}(\mathcal{L}(Y), \mathcal{L}(X)) \leq \lambda \left(\frac{\lambda}{n^\nu - \lambda} + \frac{\nu n^\nu \mathbb{E}Y}{n(n^\nu - \lambda)} \right) \left(g_\nu(\lambda) + (1 + \mathbb{E}Y) \min\left\{1, \lambda^{-1}\right\} \right)$$
$$+ \frac{\lambda c_\nu}{n} \left(\mathbb{E}Y + \mathbb{E}Y^2 \right) \min\left\{1, \lambda^{-1}\right\},$$

where $c_\nu = \max\{1, \nu\}$ and $g_\nu(\lambda)$ is as defined above.

Remark 3.2. Taking $\lambda = n^\nu p$ in Theorem 3.1 gives the following bound for the total variation distance between the laws of $Y \sim \mathrm{CMB}(n, p, \nu)$ and $X \sim \mathrm{CMP}(n^\nu p, \nu)$:

$$d_{TV}(\mathcal{L}(Y), \mathcal{L}(X))$$
$$\leq n^\nu p \left(\frac{p}{1-p} + \frac{\nu \mathbb{E}Y}{n(1-p)} \right) \left(g_\nu(n^\nu p) + (1 + \mathbb{E}Y) \min\left\{1, (n^\nu p)^{-1}\right\} \right)$$
$$+ n^{\nu-1} p c_\nu \left(\mathbb{E}Y + \mathbb{E}Y^2 \right) \min\left\{1, (n^\nu p)^{-1}\right\}.$$

However, we prefer to work with the parameters given in Theorem 3.1, because in that theorem the CMP distribution (which we regard as the limit distribution) does not depend on n.

Remark 3.3. For large n, the bound of Theorem 3.1 is of order $n^{-\min\{1,\nu\}}$, which is in agreement with the order of upper bound in the classical case $\nu = 1$ that is given by Barbour et al. (1992). In fact, Barbour and Hall (1984) obtained a lower bound of the same order:

$$\frac{1}{32}\min\left\{\lambda, \lambda^2\right\}n^{-1} \leq d_{TV}(\mathrm{Bin}(n, \lambda/n), \mathrm{Po}(\lambda)) \leq \min\left\{\lambda, \lambda^2\right\}n^{-1}. \qquad (3.1)$$

It would be desirable to obtain a corresponding lower bound for all $\nu > 0$, although their method of proof does not generalise easily to $\nu \neq 1$. We have, however, been able to get a good indication of the 'true' rate of convergence via a simple numerical study. We fixed $\lambda = 1$ and considered a number of different values of ν. For each ν, we used Mathematica to evaluate $d_{\nu,n} := d_{TV}(\mathrm{CMB}(n, 1/n^\nu, \nu), \mathrm{CMP}(1, \nu))$ for different values of n. The values of $\log(d_{\nu,n})$ were plotted against $\log(n)$ and the gradient of the line of best fit was used to estimate the exponent of n. The results of this study strongly suggest that the convergence is indeed of order $n^{-\min\{1,\nu\}}$ (for general λ, but see Remark 3.5 below for a choice of λ that gives a faster rate). For example, in the case $\nu = 1/2$ the fitted gradient was -0.502, and the fitted gradient was -0.974 for $\nu = 3/2$. A direction for future research is to verify this assertion theoretically by obtaining a lower bound of this order.

Remark 3.4. For general ν, we do not have closed-form formulas for the moments $\mathbb{E}Y$ and $\mathbb{E}Y^2$. However, $\mathbb{E}Y^k \approx \mathbb{E}X^k$ for large n, and so we can use the asymptotic formula $\mathbb{E}X^k \approx \lambda^{k/\nu}$ to see that, for large λ, the upper bound of Theorem 3.1 is of order

$$\frac{\lambda^{2/\nu}}{n} + \frac{\lambda^{1/\nu+1}}{n^\nu}.$$

For $\nu = 1$, this dependence on λ is not as good as the $O(\lambda)$ rate of (3.1).

Remark 3.5. In the special case $\lambda = \mathbb{E}Y^\nu$, the rate improves to order n^{-1}:

$$d_{TV}(\mathcal{L}(Y), \mathcal{L}(X)) \leq \min\left\{1, \lambda\right\}\frac{c_\nu}{n}\left(\mathbb{E}Y + \mathbb{E}Y^2\right). \qquad (3.2)$$

This bound can be easily read off from the proof of Theorem 3.1.

Remark 3.6. As we shall see, the proof of Theorem 3.1 relies on a stochastic ordering argument. Before arriving at this proof, we considered generalising the classical Stein's method proofs of the Poisson approximation of the binomial distribution. These approaches involve local couplings or size bias couplings (see Barbour et al., 1992). However, neither of these approaches generalise easily to the CMP approximation of the CMB distribution. The first step in generalising the classical proofs is to write $Y \sim \mathrm{CMB}(n, \lambda/n^\nu, \nu)$ as a sum of Bernoulli random variables. However, these Bernoulli random variables are strongly dependent (see (1.4)), and so local couplings are not applicable. Also, the natural generalisation of the size-bias coupling approach involves the construction of the power-bias distribution of Y, which we found resulted in intractable calculations.

The starting point for applying Stein's method is the characterisation of the CMP distribution given by Lemma 2.9 (i). Using that, we have the representation

$$d_{TV}(\mathcal{L}(Y), \mathcal{L}(X)) = \sup_{A \subseteq \mathbb{Z}^+} |\lambda\mathbb{E}f_A(Y + 1) - \mathbb{E}[Y^\nu f_A(Y)]|$$

$$= \sup_{A \subseteq \mathbb{Z}^+} \left|\lambda\mathbb{E}f_A(Y + 1) - \mathbb{E}Y^\nu\mathbb{E}f_A(Y^{(\nu)})\right|, \qquad (3.3)$$

where $f_A : \mathbb{Z}^+ \mapsto \mathbb{R}$ solves the Stein equation

$$I(x \in A) - \mathbb{P}(X \in A) = \lambda f_A(x+1) - x^\nu f_A(x). \qquad (3.4)$$

Thus, in proving Theorem 3.1 we find a bound on $\left| \lambda \mathbb{E} f_A(Y+1) - (\mathbb{E}Y^\nu) \mathbb{E} f_A(Y^{(\nu)}) \right|$ which holds uniformly in $A \subseteq \mathbb{Z}^+$. In order to do this, we will need bounds on the functions f_A solving (3.4). These are given in Lemma 3.7 below, whose proof is deferred until Section 3.2.

Lemma 3.7. *The unique solution of the CMP(λ, ν) Stein equation (3.4) is given by $f_A(0) = 0$ and, for $j \geq 0$,*

$$f_A(j+1) = \frac{(j!)^\nu}{\lambda^{j+1}} \sum_{k=0}^{j} \frac{\lambda^k}{(k!)^\nu} [I(k \in A) - \mathbb{P}(X \in A)], \qquad (3.5)$$

for $A \subseteq \mathbb{Z}^+$. The solution satisfies the bounds

$$\sup_{A \subseteq \mathbb{Z}^+} \sup_{j \in \mathbb{Z}^+} |f_A(j)| \leq g_\nu(\lambda), \qquad (3.6)$$

$$\sup_{A \subseteq \mathbb{Z}^+} \sup_{j \in \mathbb{Z}^+} |f_A(j+1) - f_A(j)| \leq \lambda^{-1} \left(1 - Z(\lambda, \nu)^{-1} \right) \leq \min\left\{ 1, \lambda^{-1} \right\}, \qquad (3.7)$$

where $g_\nu(\lambda)$ is as defined above.

Remark 3.8. The value of $f_A(0)$ is in fact irrelevant, and we follow the usual convention and set it equal to zero (see Barbour et al., 1992, p. 6). Therefore, to be precise, the function $f_A(j)$, as given by (3.5), is the unique solution of (3.4) for $j \geq 1$.

For use in what follows, we define the forward difference operator Δ by $\Delta f(j) = f(j+1) - f(j)$, and the supremum norm $\|\cdot\|$ by $\|f\| = \sup_j |f(j)|$ for all $f : \mathbb{Z}^+ \mapsto \mathbb{R}$.

Now, we have that

$$\lambda \mathbb{E} f_A(Y+1) - \mathbb{E}Y^\nu \mathbb{E} f_A(Y^{(\nu)})$$
$$= \mathbb{E}Y^\nu \left(\mathbb{E} f_A(Y+1) - \mathbb{E} f_A(Y^{(\nu)}) \right) + (\lambda - \mathbb{E}Y^\nu) \mathbb{E} f_A(Y+1).$$

Recall from Lemma 2.11 that $Y^{(\nu)} \leq_{st} Y+1$. Hence, we may follow the methods of Daly et al. (2012) and obtain

$$|\lambda \mathbb{E} f_A(Y+1) - \mathbb{E}Y^\nu \mathbb{E} f_A(Y^\nu)|$$
$$\leq (\mathbb{E}Y^\nu) \|\Delta f_A\| \left(1 + \mathbb{E}Y - \mathbb{E}Y^{(\nu)} \right) + \|f_A\| |\lambda - \mathbb{E}Y^\nu|$$
$$= \|\Delta f_A\| \left((1 + \mathbb{E}Y)\mathbb{E}Y^\nu - \mathbb{E}Y^{\nu+1} \right) + \|f_A\| |\lambda - \mathbb{E}Y^\nu|,$$

where we used (1.5) to note that $\mathbb{E}Y^{\nu+1} = \mathbb{E}Y^\nu \mathbb{E}Y^{(\nu)}$. We may then combine the representation (3.3) with Lemma 3.7 to get

$$d_{TV}(\mathcal{L}(Y), \mathcal{L}(X)) \leq \min\left\{ 1, \lambda^{-1} \right\} \left((1 + \mathbb{E}Y)\mathbb{E}Y^\nu - \mathbb{E}Y^{\nu+1} \right) + g_\nu(\lambda) |\lambda - \mathbb{E}Y^\nu|. \qquad (3.8)$$

To complete the proof of Theorem 3.1, we use Lemmas 3.9 and 3.10 below. These make use of the characterisation of the CMB distribution. The idea of combining characterisations of two distributions when using Stein's method has previously been employed by Goldstein and Reinert (2013) and Döbler (2015).

Lemma 3.9. *Let $Y \sim CMB(n, \lambda/n^\nu, \nu)$. Then*

$$\mathbb{E}Y^{\nu+1} \geq \lambda \left(1 + \mathbb{E}Y - \frac{c_\nu}{n} \left(\mathbb{E}Y + \mathbb{E}Y^2 \right) \right),$$

where $c_\nu = \max\{1, \nu\}$.

Proof: We use the characterisation of the CMB distribution given in Lemma 2.9 (ii) to note that

$$\mathbb{E}Y^{\nu+1} = \lambda \left(1 - \frac{\lambda}{n^\nu} \right)^{-1} \mathbb{E} \left[(Y+1) \left(1 - \frac{Y}{n} \right)^\nu \right]$$

$$\geq \lambda \left(1 - \frac{\lambda}{n^\nu} \right)^{-1} \mathbb{E} \left[(Y+1) \left(1 - \frac{c_\nu Y}{n} \right) \right]$$

$$\geq \lambda \mathbb{E} \left[(Y+1) \left(1 - \frac{c_\nu Y}{n} \right) \right].$$

$\square$

Lemma 3.10. *Let $Y \sim CMB(n, \lambda/n^\nu, \nu)$. Then*

$$|\lambda - \mathbb{E}Y^\nu| \leq \lambda \left(\frac{\lambda}{n^\nu - \lambda} + \frac{\nu n^\nu \mathbb{E}Y}{n(n^\nu - \lambda)} \right).$$

Proof: Let $p = \lambda/n^\nu$. Using Lemma 2.9 (ii),

$$\lambda - \mathbb{E}Y^\nu = \lambda \left(1 - \frac{\mathbb{E} \left(1 - \frac{Y}{n} \right)^\nu}{1 - p} \right).$$

The result then follows by applying Taylor's theorem to the function $(1 - y)^\nu$. $\square$

Substituting the bounds of Lemmas 3.9 and 3.10 into (3.8) completes the proof of Theorem 3.1.

3.1. *Remarks on Lemma 3.10.* We use this section to give some remarks related to Lemma 3.10. Firstly, note that the upper bound given in that lemma is of order $n^{-\min\{1,\nu\}}$, which can in fact easily be seen to be the optimal rate. Using Lemma 3.11 below, we show that a better bound is possible when λ/n^ν is small, although this improved bound will of course still be of the same order as the bound given in Lemma 3.10.

Lemma 3.11. *Let $Y \sim CMB(n, p, \nu)$ with $n > 1$. There exists $p^\star \in (0, 1]$ such that for $p \leq p^\star$*

$$\mathbb{E}Y^\nu \begin{cases} \leq n^\nu p & \text{if } \nu \geq 1, \\ \geq n^\nu p & \text{if } \nu < 1. \end{cases}$$

Proof: Let

$$h(p) = \mathbb{E}Y^\nu = \frac{\sum_{j=0}^n j^\nu \binom{n}{j}^\nu p^j (1-p)^{n-j}}{\sum_{j=0}^n \binom{n}{j}^\nu p^j (1-p)^{n-j}}.$$

Elementary calculations show that $h(0) = 0$, $h'(0) = n^\nu$ and

$$h''(0) = -2n^\nu \left(n^\nu - (n-1)^\nu - 1 \right).$$

Note that (since $n > 1$), $h''(0) < 0$ for $\nu \geq 1$ and $h''(0) > 0$ for $\nu < 1$. Using the continuity of h and Taylor's theorem applied to h, the result follows. $\square$

Consider now the case $Y \sim \mathrm{CMB}(n, \lambda/n^\nu, \nu)$ with $\nu \geq 1$. By Lemma 3.11, for n sufficiently large we have that $\lambda - \mathbb{E}Y^\nu \geq 0$, and we may then follow the proof of Lemma 3.10 to get the bound

$$\lambda - \mathbb{E}Y^\nu \leq \lambda \left(\frac{\nu n^\nu \mathbb{E}Y}{n(n^\nu - \lambda)} - \frac{\lambda}{n^\nu - \lambda} \right),$$

which improves upon Lemma 3.10.

A similar argument in the case $\nu < 1$ gives that, for n sufficiently large, $\lambda - \mathbb{E}Y^\nu \leq 0$ and

$$\mathbb{E}Y^\nu - \lambda \leq \frac{\lambda^2}{n^\nu - \lambda}.$$

3.2. *Proof of Lemma 3.7.* It is straightforward to verify that (3.5), denoted by $f_A(j)$, solves the Stein equation (3.4). To establish uniqueness of the solution, we take $j = 0$ in (3.4), from which it follows that any function $h_A(j)$ that solves the Stein equation (3.4) must satisfy $h_A(1) = f_A(1)$. By iteration on $\lambda h_A(j+1) - j^\nu h_A(j) = \lambda f_A(j+1) - j^\nu f_A(j)$ it follows that $h_A(j) = f_A(j)$ for all $j \geq 1$, which confirms the uniqueness of the solution.

We now establish (3.7). By constructing $X \sim \mathrm{CMP}(\lambda, \nu)$ as the equilibrium distribution of a birth-death process with birth rates $\alpha_j = \lambda$ and death rates $\beta_j = j^\nu$, the first inequality of (3.7) follows from Corollary 2.12 of Brown and Xia (2001). Since $Z(\lambda, \nu) \geq 1$ for all λ and ν, it follows that $\lambda^{-1}\left(1 - Z(\lambda, \nu)^{-1}\right) \leq \lambda^{-1}$. Finally,

$$\lambda^{-1}(1 - Z(\lambda, \nu)^{-1}) = \frac{\sum_{j=0}^{\infty} \frac{\lambda^j}{((j+1)!)^\nu}}{\sum_{k=0}^{\infty} \frac{\lambda^k}{(k!)^\nu}} \leq 1.$$

This completes the proof of (3.7).

Remark 3.12. The upper bound $\lambda^{-1}\left(1 - Z(\lambda, \nu)^{-1}\right)$ for the forward difference is attained by $f_{\{1\}}(2) - f_{\{1\}}(1)$.

It remains to establish (3.6). We do this by considering separately four cases. Our strategy is to suitably generalise the proof of Lemma 1.1.1 of Barbour et al. (1992), which gives analogous bounds in the Poisson case ($\nu = 1$).

Firstly, note that from (3.7) and the choice $f_A(0) = 0$,

$$|f_A(1)| \leq \min\{1, \lambda^{-1}\}, \tag{3.9}$$

for each $A \subseteq \mathbb{Z}^+$. Given (3.9), we need only to show the stated bound on $|f_A(j+1)|$ for $j \geq 1$ in each of the four cases detailed below.

Case I: $\nu \geq 1$ and $\lambda > 1$. Note that by examining the proof of Lemma 1.1.1 of Barbour et al. (1992), it is clear that $|f_A(j+1)| \leq 5/4$ for all $j \geq 1$ whenever $\nu \geq 1$. We can, however, do a little better. Barbour and Eagleson (1983, Lemma 4) obtained the bound $|f_A(j+1)| \leq 1$ when $\nu = 1$. By examining their proof we see that the bound also holds for all $\nu \geq 1$.

Now, let $U_m = \{0, 1, \ldots, m\}$. It is easily verified that the solution f_A to the Stein equation (3.4) is given by

$$f_A(j+1) = \lambda^{-j-1}(j!)^\nu Z(\lambda, \nu)\big(\text{CMP}(\lambda, \nu)\{A \cap U_j\}$$
$$- \text{CMP}(\lambda, \nu)\{A\}\text{CMP}(\lambda, \nu)\{U_j\}\big)$$
$$= \lambda^{-j-1}(j!)^\nu Z(\lambda, \nu)\big(\text{CMP}(\lambda, \nu)\{A \cap U_j\}\text{CMP}(\lambda, \nu)\{U_j^c\}$$
$$- \text{CMP}(\lambda, \nu)\{A \cap U_j^c\}\text{CMP}(\lambda, \nu)\{U_j\}\big),$$

where $\text{CMP}(\lambda, \nu)\{A\} = \mathbb{P}(X \in A)$. Hence

$$|f_A(j+1)| \leq \lambda^{-j-1}(j!)^\nu Z(\lambda, \nu)\text{CMP}(\lambda, \nu)\{U_j\}\text{CMP}(\lambda, \nu)\{U_j^c\}, \qquad (3.10)$$

with equality for $A = U_j$.

Equation (3.10) gives us two ways of bounding $|f_A(j+1)|$. Firstly, note that

$$|f_A(j+1)| \leq \lambda^{-j-1}(j!)^\nu Z(\lambda, \nu)\text{CMP}(\lambda, \nu)\{U_j\} = \lambda^{-1}\sum_{r=0}^{j}\lambda^{-r}\left(\frac{j!}{(j-r)!}\right)^\nu, \tag{3.11}$$

and when $j^\nu < \lambda$, this may be bounded to give

$$|f_A(j+1)| \leq \lambda^{-1}\sum_{r=0}^{j}\left(\frac{j^\nu}{\lambda}\right)^r \leq \frac{1}{\lambda - j^\nu}. \tag{3.12}$$

Secondly, we also have

$$|f_A(j+1)| \leq \lambda^{-j-1}(j!)^\nu Z(\lambda, \nu)\text{CMP}(\lambda, \nu)\{U_j^c\} = \lambda^{-1}\sum_{r=j+1}^{\infty}\lambda^{r-j}\left(\frac{j!}{r!}\right)^\nu, \tag{3.13}$$

and when $(j+2)^\nu > \lambda$, this may be bounded to give

$$|f_A(j+1)| \leq \frac{1}{(j+1)^\nu}\sum_{r=0}^{\infty}\left(\frac{\lambda}{(j+2)^\nu}\right)^r = \frac{(j+2)^\nu}{(j+1)^\nu\left((j+2)^\nu - \lambda\right)}. \tag{3.14}$$

Note that the bounds (3.11)–(3.14) hold for all values of ν and λ. We will also make use of these bounds in the other cases we consider below.

Now, for $j^\nu \leq \lambda - \lambda^{1-1/2\nu}$, we use (3.12) to get that

$$|f_A(j+1)| \leq \lambda^{1/2\nu - 1}. \tag{3.15}$$

Similarly, when $(j+2)^\nu \geq \lambda + \lambda^{1-1/2\nu}$, we use (3.14) to get that

$$|f_A(j+1)| \leq (3/2)^\nu \lambda^{1/2\nu - 1}, \tag{3.16}$$

noting that $(j+2)^\nu > j^\nu$. It remains only to treat the case $|j^\nu - \lambda| < \lambda^{1-1/2\nu}$.

To that end, let $\lambda - \lambda^{1-1/2\nu} < j^\nu < \lambda$, and use (3.11) to note that

$$|f_A(j+1)| \leq \lambda^{-1}\left(\sum_{r=0}^{\lfloor B \rfloor}a_r + \sum_{r=\lfloor B \rfloor+1}^{j}a_r\right),$$

for any $B \leq j$, where

$$a_r = \lambda^{-r}\left(\frac{j!}{(j-r)!}\right)^\nu.$$

Note that $|a_r| < 1$ for each $r \in \mathbb{Z}^+$. We choose

$$B = \lambda^{1/\nu}\left[1 - (1 - \lambda^{-1/2\nu})^{1/\nu}\right],$$

so that $a_{r+1}/a_r < 1 - \lambda^{-1/2\nu}$ for all $r > B$. Hence, we have

$$|f_A(j+1)| \leq \lambda^{-1}\left(B + 1 + \frac{1 - \lambda^{-1/2\nu}}{\lambda^{-1/2\nu}}\right) = \lambda^{1/\nu-1}\left[1 - (1 - \lambda^{-1/2\nu})^{1/\nu}\right] + \lambda^{1/2\nu-1}.$$
(3.17)

Note that, by Taylor's theorem and since $\nu \geq 1$,

$$1 - (1 - \lambda^{-1/2\nu})^{1/\nu} \leq \frac{1}{\nu}\left(1 - \lambda^{-1/2\nu}\right)^{1/\nu-1}\lambda^{-1/2\nu}.$$

Hence,

$$|f_A(j+1)| \leq \left(1 + \frac{1}{\nu}\right)\left(1 - \lambda^{-1/2\nu}\right)^{1/\nu-1}\lambda^{1/2\nu-1}.$$
(3.18)

Finally, we consider the case $\lambda < j^\nu < \lambda + \lambda^{1-1/2\nu}$. From (3.13) we have

$$|f_A(j+1)| \leq \lambda^{-1}\left(\sum_{r=j+1}^{\lfloor C \rfloor} b_r + \sum_{r=\lfloor C \rfloor+1}^{\infty} b_r\right),$$

for $C \geq j$, where

$$b_r = \lambda^{r-j}\left(\frac{j!}{r!}\right)^\nu.$$

Analogously to before, we note that $|b_r| < 1$ for each r, and we make the choice $C = \lambda^{1/\nu}\left(1 + \lambda^{-1/2\nu}\right)^{1/\nu}$ so that $b_{r+1}/b_r \leq (1 + \lambda^{-1/2\nu})$ for $r > C$. We then get the bound

$$|f_A(j+1)| \leq \lambda^{1/\nu-1}\left[(1 + \lambda^{-1/2\nu})^{1/\nu} - 1\right] + \lambda^{1/2\nu-1}.$$
(3.19)

Using Taylor's theorem,

$$(1 + \lambda^{-1/2\nu})^{1/\nu} - 1 \leq \frac{1}{\nu}\lambda^{-1/2\nu},$$

since $\nu \geq 1$, and so

$$|f_A(j+1)| \leq \left(1 + \frac{1}{\nu}\right)\lambda^{1/2\nu-1}.$$
(3.20)

Combining the bounds (3.15), (3.16), (3.18) and (3.20) we obtain the stated bound on $\|f_A\|$ in this case.

Remark 3.13. Recall (3.10). Taking $j \approx \lambda^{1/\nu}$, and using Stirling's formula and (1.2), gives

$$|f_A(j+1)| \approx \frac{\lambda^{1/2\nu-1}}{(2\pi)^{(\nu-1)/2}\sqrt{\nu}},$$

for $j \approx \lambda^{1/\nu}$ and large λ. Hence, a bound of order $\lambda^{1/2\nu-1}$ is the best that we can expect for $\|f_A\|$ for large λ. This order is achieved by Lemma 3.7. This remark also applies to Case II considered below.

Case II: $\nu \leq 1$ and $\lambda \geq 1$. Here we use an analogous argument to that employed in Case I. The bounds (3.15) and (3.16) still apply; the only changes to our argument come for the cases where $|j^\nu - \lambda| < \lambda^{1-1/2\nu}$.

When $\lambda - \lambda^{1-1/2\nu} < j^\nu < \lambda$, we again use (3.17). Since $\nu \leq 1$ in this case, Taylor's theorem gives

$$1 - (1 - \lambda^{-1/2\nu})^{1/\nu} \leq \frac{1}{\nu}\lambda^{-1/2\nu},$$

from which it follows that

$$|f_A(j+1)| \leq \left(1 + \frac{1}{\nu}\right)\lambda^{1/2\nu-1}.$$

When $\lambda < j^\nu < \lambda + \lambda^{1-1/2\nu}$, we use (3.19), noting that, since $\nu < 1$,

$$(1 + \lambda^{-1/2\nu})^{1/\nu} - 1 \leq \frac{1}{\nu}\left(1 + \lambda^{-1/2\nu}\right)^{1/\nu-1}\lambda^{-1/2\nu},$$

giving

$$|f_A(j+1)| \leq \left(1 + \frac{1}{\nu}\right)\left(1 + \lambda^{-1/2\nu}\right)^{1/\nu-1}\lambda^{1/2\nu-1}.$$

The stated bound follows.

Case III: $\nu \geq 1$ and $\lambda \leq 1$. As before, we may use the proof of Lemma 4 of Barbour and Eagleson (1983) to obtain the bound $|f_A(j+1)| \leq 1$ for all $\nu \geq 1$.

Case IV: $\nu < 1$ and $\lambda < 1$. Here we again use (3.13). That bound gives us

$$|f_A(j+1)| \leq \lambda^{-1}\sum_{r=j+1}^{\infty}\left(\frac{\lambda^{r-j}}{(r-j)!}\right)^\nu \binom{r}{j}^{-\nu}\left(\lambda^{1-\nu}\right)^{r-j} \leq \lambda^{\nu-1}\sum_{r=1}^{\infty}\left(\lambda^{1-\nu}\right)^r.$$

Since $\lambda < 1$ and $\nu < 1$, we have $\lambda^{1-\nu} < 1$. Hence we get the bound

$$|f_A(j+1)| \leq \frac{1}{1-\lambda^{1-\nu}}.$$

Remark 3.14. Consider the case $\nu = 0$ and $\lambda < 1$. We then have that our CMP random variable X has a geometric distribution, supported on $\mathbb{Z}^+$, with parameter $\mathbb{P}(X = 0) = 1 - \lambda$. Note that in this case, Lemma 3.7 gives the bound $\|f_A\| \leq (1 - \lambda)^{-1}$. This was shown, in Remark 4.1 of Daly (2010), to be the correct dependence on λ for such a bound.

4. Other convergence and approximation results

In this section we consider other convergence and approximation results related to CMP distributions.

4.1. *Sums of Bernoulli random variables.* In Section 3 we have considered the convergence of the CMB distribution to an appropriate CMP limit. In this case we were able to derive an explicit bound on this convergence. Recalling (1.4), we are able to write a CMB distribution as a sum of Bernoulli random variables (having a particular dependence structure), with each Bernoulli summand having the same marginal distribution. In this section we consider how we may generalise (1.4) to a sum of Bernoulli random variables which are no longer exchangeable and yet give a CMP limiting distribution in an analogous way to the limit considered in Section

3. In this case, although we are able to prove convergence in distribution, we are unable to give an explicit bound on the convergence rate; further discussion is given in Remark 4.2.

Consider the following generalisation of (1.4). Let $X_1, \ldots, X_n$ be Bernoulli random variables with joint distribution given by

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \frac{1}{C'_n} \binom{n}{k}^{\nu-1} \prod_{j=1}^{n} p_j^{x_j} (1 - p_j)^{1-x_j},$$

where $k = x_1 + \cdots + x_n$ and the normalizing constant C'_n is given by

$$C'_n = \sum_{k=0}^{n} \binom{n}{k}^{\nu-1} \sum_{A \in F_k} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j),$$

where

$$F_k = \{A \subseteq \{1, \ldots, n\} : |A| = k\} .$$

We consider the convergence of the sum $W = X_1 + \cdots + X_n$. It is easy to see that W has mass function

$$\mathbb{P}(W = k) = p_{n,\nu}(k; p_1, \ldots, p_n) = \frac{1}{C'_n} \binom{n}{k}^{\nu-1} \sum_{A \in F_k} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j), \qquad (4.1)$$

for $k = 0, 1, \ldots, n$. This distribution generalises the Poisson binomial distribution in a way analogous to the CMP and CMB generalisations of the Poisson and binomial distributions. We therefore say that a random variable with mass function (4.1) follows the Conway-Maxwell-Poisson binomial (CMPB) distribution. Of course, the case $\nu = 1$ is the usual Poisson binomial distribution and the case $p_1 = \cdots = p_n = p$ reduces to the $\mathrm{CMB}(n, p, \nu)$ distribution.

Theorem 4.1. *Let $W = X_1 + \cdots + X_n$ be as above, with probability mass function $p_{n,\nu}(k; p_1, \ldots, p_n)$ given by (4.1) with $p_i = \frac{\lambda_i}{n^\nu}$ for $i = 1, \ldots, n$, where the λ_i are positive constants that do not involve n. Then W converges in distribution to $X \sim CMP(\lambda, \nu)$ as $n \to \infty$, where*

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \lambda_i .$$

Proof: Firstly, note that in the case $\nu = 1$ the result is known. It is the classical convergence of a sum of independent Bernoulli random variables to a Poisson distribution. That immediately gives us the limit

$$\frac{1}{n^k} \sum_{A \in F_k} \prod_{i \in A} \lambda_i \prod_{j \in A^c} \left(1 - \frac{\lambda_j}{n}\right) \to e^{-\lambda} \frac{\lambda^k}{k!} , \qquad (4.2)$$

as $n \to \infty$. As a consequence of (4.2), we have that

$$\frac{1}{n^k} \sum_{A \in F_k} \prod_{i \in A} \lambda_i \to \frac{\lambda^k}{k!} , \qquad (4.3)$$

as $n \to \infty$, since

$$\lim_{n \to \infty} \prod_{j \in A^c} \left(1 - \frac{\lambda_j}{n}\right) = \lim_{n \to \infty} \prod_{j=1}^{n} \left(1 - \frac{\lambda_j}{n}\right) \cdot \lim_{n \to \infty} \prod_{l \in A} \left(1 - \frac{\lambda_l}{n}\right)^{-1} = e^{-\lambda} .$$

Now, in the present case we may write the mass function (4.1) in the form

$$p_{n,\nu}(k; \lambda_1/n^\nu, \ldots, \lambda_n/n^\nu)$$

$$= \frac{1}{C'_n} \left(\frac{n!}{(n-k)!n^k} \right)^{\nu-1} \frac{1}{(k!)^{\nu-1}} \sum_{A \in F_k} \prod_{i \in A} \frac{\lambda_i}{n} \prod_{j \in A^c} \left(1 - \frac{\lambda_j}{n^\nu} \right).$$

Clearly

$$\lim_{n \to \infty} \frac{n!}{(n-k)!n^k} = 1,$$

and

$$\prod_{j \in A^c} \left(1 - \frac{\lambda_j}{n^\nu} \right) = \prod_{j=1}^n \left(1 - \frac{\lambda_j}{n^\nu} \right) \prod_{l \in A} \left(1 - \frac{\lambda_l}{n^\nu} \right)^{-1}.$$

Note that

$$\lim_{n \to \infty} \prod_{l \in A} \left(1 - \frac{\lambda_l}{n^\nu} \right)^{-1} = 1,$$

and that the product

$$\prod_{j=1}^n \left(1 - \frac{\lambda_j}{n^\nu} \right)$$

and the normalizing constant C'_n do not depend on k. Combining these observations with (4.3), we have that

$$\lim_{n \to \infty} p_{n,\nu}(k; \lambda_1/n^\nu, \ldots, \lambda_n/n^\nu) = \frac{C\lambda^k}{(k!)^\nu},$$

where C does not depend on k. The result follows. $\square$

Remark 4.2. It would be desirable to extend Theorem 4.1 to include an explicit bound on the convergence rate, as was achieved in Theorem 3.1. Such a bound could, in principle, be established by generalising the proof of that theorem. This approach would require one to obtain a Stein equation for the CMPB distribution, a generalisation of the stochastic ordering result of Lemma 2.11 to the CMPB distribution, and an appropriate extension of the moment estimates of Lemmas 3.9 and 3.10. This is a possible direction for future research.

4.2. *Mixed CMP distributions.* Finally, we also consider the case of a mixed CMP distribution. For a non-negative, real-valued random variable ξ, we say that $W \sim$ CMP(ξ, ν) has a mixed CMP distribution if

$$\mathbb{P}(W = j) = \frac{1}{(j!)^\nu} \mathbb{E} \left[\frac{\xi^j}{Z(\xi, \nu)} \right], \quad j \in \mathbb{Z}^+.$$

We assume throughout that ξ is such that this expectation exists.

Following the proof of Theorem 1.C (for mixed Poisson approximation) in the book by Barbour et al. (1992), we use the characterisation in Lemma 2.9 (i), along with the bounds on the solution to the Stein equation given in Lemma 3.7, to obtain the following.

Theorem 4.3. *Let ξ be a non-negative random variable. Then*

$$d_{TV}(CMP(\xi, \nu), CMP(\lambda, \nu)) \le g_\nu(\lambda)\mathbb{E}|\xi - \lambda|,$$

where $g_\nu(\lambda)$ is as defined in Section 3.

Acknowledgements

References

D. E. Amos. Computation of modified Bessel functions and their ratios. *Math. Comp.* **28**, 239–251 (1974). MR0333287.

N. Balakrishnan and S. Pal. EM algorithm-based likelihood estimation for some cure rate models. *J. Stat. Theory Pract.* **6** (4), 698–724 (2012). MR3196574.

A. D. Barbour and Louis H. Y. Chen, editors. *An introduction to Stein's method*, volume 4 of *Lecture Notes Series. Institute for Mathematical Sciences. National University of Singapore.* Singapore University Press, Singapore; World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ (2005). ISBN 981-256-330-X. MR2235447.

A. D. Barbour and G. K. Eagleson. Poisson approximation for some statistics based on exchangeable trials. *Adv. in Appl. Probab.* **15** (3), 585–600 (1983). MR706618.

A. D. Barbour and P. Hall. On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** (3), 473–480 (1984). MR755837.

A. D. Barbour, L. Holst and S. Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York (1992). ISBN 0-19-852235-5. Oxford Science Publications. MR1163825.

P. Boatwright, S. Borle and J. B. Kadane. A model of the joint distribution of purchase quantity and timing. *J. Amer. Statist. Assoc.* **98** (463), 564–572 (2003). MR2011672.

T. C. Brown and A. Xia. Stein's method and birth-death processes. *Ann. Probab.* **29** (3), 1373–1403 (2001). MR1872746.

L. H. Y. Chen. Poisson approximation for dependent trials. *Ann. Probability* **3** (3), 534–545 (1975). MR0428387.

R. W. Conway and W. L. Maxwell. A queueing model with state dependent service rate. *J. Ind. Engineering* **12**, 132–136 (1962).

E. L. Crow. The mean deviation of the poisson distribution. *Biometrika* **45** (3/4), 556–559 (1958). DOI: 10.2307/2333201.

F. Daly. Stein's method for compound geometric approximation. *J. Appl. Probab.* **47** (1), 146–156 (2010). MR2654764.

F. Daly. Negative dependence and stochastic orderings (2016+). To appear in ESAIM: Probab. StatDLMF.

F. Daly and O. Johnson. Bounds on the Poincaré constant under negative dependence. *Statist. Probab. Lett.* **83** (2), 511–518 (2013). MR3006983.

F. Daly, C. Lefèvre and S. Utev. Stein's method and stochastic orderings. *Adv. in Appl. Probab.* **44** (2), 343–372 (2012). MR2977399.

C. Döbler. Stein's method for the half-normal distribution with applications to limit theorems related to the simple symmetric random walk. *ALEA Lat. Am. J. Probab. Math. Stat.* **12** (1), 171–191 (2015). MR3343481.

S. B. Gillispie and C. G. Green. Approximating the Conway-Maxwell-Poisson distribution normalization constant. *Statistics* **49** (5), 1062–1073 (2015). MR3378031.

L. Goldstein and G. Reinert. Distributional transformations, orthogonal polynomials, and Stein characterizations. *J. Theoret. Probab.* **18** (1), 237–260 (2005). MR2132278.

L. Goldstein and G. Reinert. Stein's method for the beta distribution and the Pólya-Eggenberger urn. *J. Appl. Probab.* **50** (4), 1187–1205 (2013). MR3161381.

S. D. Guikema and J. P. Goffelt. A flexible count data regression model for risk analysis. *Risk Analysis* **28** (1), 213–223 (2008). DOI: 10.1111/j.1539-6924.2008.01014.x.

E. J. Hinch. *Perturbation methods.* Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge (1991). ISBN 0-521-37310-7; 0-521-37897-4. MR1138727.

R. Kaas and J. M. Buhrman. Mean, median and mode in binomial distributions. *Statist. Neerlandica* **34** (1), 13–18 (1980). MR576005.

J. B. Kadane. Sums of Possibly Associated Bernoulli Variables: The Conway–Maxwell-Binomial Distribution. *Bayesian Anal.* **11** (2), 403–420 (2016). MR3471996.

J. B. Kadane, G. Shmueli, T. P. Minka, S. Borle and P. Boatwright. Conjugate analysis of the Conway-Maxwell-Poisson distribution. *Bayesian Anal.* **1** (2), 363–374 (electronic) (2006). MR2221269.

C. Mallows. Another comment on O'Cinneide. *Am. Stat.* **45**, 257 (1991).

F. Olver, D. Lozier, R. Boisvert and C. Clark. *NIST handbook of mathematical functions.* U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC; Cambridge University Press, Cambridge (2010). ISBN 978-0-521-14063-8. MR2723248.

E. A. Peköz, A. Röllin and N. Ross. Generalized gamma approximation with rates for urns, walks and trees. *Ann. Probab.* **44** (3), 1776–1816 (2016). MR3502594.

J. Rodrigues, M. de Castro, V. G. Cancho and N. Balakrishnan. COM-Poisson cure rate survival models and an application to a cutaneous melanoma data. *J. Statist. Plann. Inference* **139** (10), 3605–3611 (2009). MR2549108.

K. F. Sellers and G. Shmueli. A flexible regression model for count data. *Ann. Appl. Stat.* **4** (2), 943–961 (2010). MR2758428.

M. Shaked and J. G. Shanthikumar. *Stochastic orders.* Springer Series in Statistics. Springer, New York (2007). ISBN 978-0-387-32915-4; 0-387-32915-3. MR2265633.

G. Shmueli, T. P. Minka, J. B. Kadane, S. Borle and P. Boatwright. A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *J. Roy. Statist. Soc. Ser. C* **54** (1), 127–142 (2005). MR2134602.

Z. Shun and P. McCullagh. Laplace approximation of high-dimensional integrals. *J. Roy. Statist. Soc. Ser. B* **57** (4), 749–760 (1995). MR1354079.

C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602. Univ. California Press, Berkeley, Calif. (1972). MR0402873.

G. Wu, S. H. Holan and C. K. Wikle. Hierarchical Bayesian spatio-temporal Conway-Maxwell Poisson models with dynamic dispersion. *J. Agric. Biol. Environ. Stat.* **18** (3), 335–356 (2013). MR3110897.