Introducing a Conway-Maxwell-Multinomial Distribution for Flexible Modeling of Categorical Data

Darcy Steeg Morris*,† Andrew M. Raim† Kimberly F. Sellers†,‡

Abstract
Count data commonly arise as a simple count of events in a fixed interval or the number of successes for a set of categories in a fixed number of trials. The Poisson, binomial and multinomial distributions are traditionally used to model such data, where the appropriate choice depends on the data generating mechanism. In practice, count data often exhibit over- or under-dispersion where variability observed in the data cannot be adequately captured via these standard distributions. The Conway-Maxwell (COM)-Poisson distribution supports such flexibility relative to the Poisson distribution for modeling simple count data. Shmueli et. al. (2005) present a COM-binomial distribution that permits flexibility in modeling binomial data, based on COM-Poisson conditional probabilities. We formally extend the COM-binomial distribution to the setting of more than two categories, thus defining a COM-multinomial distribution. We describe properties and illustrate the flexible characteristics of this distribution.

Key Words: count data, categorical data analysis, COM-Poisson distribution, multinomial distribution

1. Introduction

The Poisson, binomial and multinomial distributions are commonly used for modeling categorical data where either a simple count is observed in a fixed interval, or a number of successes is observed in a fixed number of trials. The nature of the data generating mechanism – specifically, the support and dimension of the count outcome – suggests the choice of distribution. Each of these standard distributions has a theoretically defined mean-variance relationship that depends only on the rate parameter for the Poisson distribution, and the number of trials and probability parameters for the binomial and multinomial distributions. However, count data often exhibit variability that violates the restrictive mean-variance assumptions of the Poisson, binomial and multinomial distribution.

There are a variety of distributions that provide flexibility via additional dispersion parameters to adequately capture variability that is inconsistent with these traditional distributions. The negative binomial distribution (i.e. Poisson-gamma compound distribution) and the Poisson-lognormal mixture model are two commonly used alternatives to the Poisson distribution that allow exclusively for over-dispersion. Models that allow exclusively for under-dispersion relative to the Poisson distribution are less common (Sellers and Morris, 2017). The generalized Poisson distribution

* Corresponding author. E-mail: darcy.steeg.morris@census.gov.

†Center for Statistical Research & Methodology, U.S. Census Bureau, Washington, DC. Disclaimer: This paper is intended to inform interested parties of research and to encourage discussion. Any views expressed on statistical, methodological, or technical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

‡ Mathematics and Statistics Department, Georgetown University, Washington, DC

(Consul and Jain, 1973) and the Conway-Maxwell-Poisson distribution (Conway and Maxwell, 1962; Shmueli et al., 2005) are two-parameter generalizations of the Poisson distribution that can flexibly support both over- and under-dispersion for modeling simple count data. Similarly, the beta-binomial compound distribution and logistic-normal mixture model are commonly used alternatives to the binomial distribution that allow exclusively for over-dispersion relative to the binomial distribution; while the Conway-Maxwell-binomial distribution supports versatility for modeling both over- and under-dispersed binomial data (Shmueli et al., 2005). The Dirichlet-multinomial compound distribution (Mosimann, 1962) and the multinomial cluster model (Morel and Nagaraj, 1993) can account for over-dispersion relative to the multinomial distribution. We formalize a Conway-Maxwell extension of the multinomial distribution – analogous to the formulation of the Conway-Maxwell-binomial distribution – to flexibly model both over- and under-dispersed categorical data with a fixed number of trials and more than two categories. We describe properties of this Conway-Maxwell-multinomial distribution and illustrate its behavior for varying degrees of dispersion relative to the multinomial distribution.

The rest of the paper proceeds as follows. Section 2 reviews basic properties of the Conway-Maxwell-Poisson and Conway-Maxwell-binomial distributions. Section 3 introduces the Conway-Maxwell-multinomial distribution and describes properties: the probability mass function, special cases, moments, generating functions, marginal and conditional distributions, and a tractable Gibbs sampler for random value generation. Section 4 concludes the paper.

2. Conway-Maxwell Extensions of the Poisson and Binomial Distributions

2.1 Conway-Maxwell-Poisson Distribution

The Conway-Maxwell-Poisson (COM-Poisson or CMP) distribution is a flexible distribution for count data that allows for over- or under-dispersion (Conway and Maxwell, 1962; Shmueli et al., 2005). The CMP probability mass function (pmf) for a single observation takes the form

$$\mathrm{P}(Y = y \mid \lambda, \nu) = \frac{\lambda^y}{(y!)^\nu Z(\lambda, \nu)}, \quad y = 0, 1, 2, \ldots$$

for a random variable Y, where $Z(\lambda, \nu) = \sum_{y=0}^{\infty} \frac{\lambda^y}{(y!)^\nu}$ is a normalizing constant. In this setting, $\lambda = \mathrm{E}(Y^\nu)$, where $\nu \geq 0$ is the dispersion parameter such that $\nu = 1$ denotes equi-dispersion, $\nu > 1$ signifies under-dispersion, and $\nu < 1$ signifies over-dispersion. The moments of the CMP distribution are not of closed form. For example,

$$\mu = \mathrm{E}(Y) = \sum_{y=0}^{\infty} \frac{y\lambda^y}{(y!)^\nu Z(\lambda, \nu)} = \lambda \frac{\partial \log Z(\lambda, \nu)}{\partial \lambda}.$$

However, Shmueli et al. (2005) note that assuming an asymptotic approximation for $Z(\lambda, \nu)$ leads to a close approximation for the mean:

$$\mathrm{E}(Y) \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu} \text{ for } \nu \leq 1 \text{ or } \lambda > 10^\nu.$$

The CMP distribution includes three well-known distributions as special cases: Poisson with rate parameter λ ($\nu = 1$); geometric with success probability $1 - \lambda$ ($\nu = 0, \lambda < 1$); and Bernoulli with success probability $\frac{\lambda}{1+\lambda}$ ($\nu \to \infty$). See Shmueli et al. (2005) and Sellers et al. (2012) for details regarding this distribution.

2.2 Conway-Maxwell-Binomial Distribution

The Conway-Maxwell-binomial (COM-binomial or CMB) distribution is a flexible generalization of the binomial distribution that captures over- and under-dispersion (Kadane, 2016). A random variable follows the CMB distribution (alternatively termed the Conway-Maxwell-Poisson-binomial distribution in Shmueli et al. (2005) and Borges et al. (2014)) according to pmf

$$\mathrm{P}(Y = y) = \frac{\binom{m}{y}^{\nu} p^{y} (1-p)^{m-y}}{C(p, \nu)}, \; y = 0, 1, \ldots, m, \tag{1}$$

where $C(p, \nu) = \sum_{y=0}^{m} \binom{m}{y}^{\nu} p^{y} (1-p)^{m-y}$ is the normalizing constant, $m \in \mathbb{N}$, $\nu \in \mathbb{R}$ and $p \in (0, 1)$. This distribution captures over- and under-dispersion relative to the binomial distribution when $\nu < 1$ and $\nu > 1$, respectively, and reduces to the usual binomial distribution for $\nu = 1$. For under-dispersion as $\nu \to \infty$, the CMB distribution concentrates at $\frac{m}{2}$ for even m or $\lceil \frac{m}{2} \rceil$ and $\lfloor \frac{m}{2} \rfloor$ for odd m. Conversely, for over-dispersion as $\nu \to -\infty$, the CMB distribution concentrates at 0 or m. Shmueli et al. (2005) recognize that the CMB distribution can be obtained as the sum of dependent Bernoulli random variables $(Z_1, \ldots Z_m)$ with pmf

$$\mathrm{P}(Z_1 = z_1, \ldots, Z_m = z_m) \propto \binom{m}{z}^{\nu-1} p^{z} (1-p)^{m-z},$$

where $z = \sum_{i=1}^{m} z_i$. The CMB distribution allows for negatively and positively correlated $(z_1, \ldots, z_m)$ corresponding to $\nu > 1$ and $\nu < 1$, respectively.[1] The extreme cases of the CMB distribution, $\nu \to -\infty$ and $\nu \to \infty$, reflect perfect postive and negative correlation of the Bernoulli components, respectively (Borges et al., 2014; Kadane, 2016).

Shmueli et al. (2005) show that the CMB distribution can be derived as an extension of the CMP distribution. Consider two independent random variables $X \sim \mathrm{CMP}(\lambda_x, \nu)$ and $Y \sim \mathrm{CMP}(\lambda_y, \nu)$. Shmueli et al. (2005) find the distribution derived by conditioning X on the sum $X + Y = m$ results in the CMB $\left(m, \frac{\lambda_x}{\lambda_x + \lambda_y}, \nu\right)$ with pmf

$$\mathrm{P}(X = x \mid X + Y = x + y) \propto \binom{x+y}{x}^{\nu} \left(\frac{\lambda_x}{\lambda_x + \lambda_y}\right)^{x} \left(\frac{\lambda_y}{\lambda_x + \lambda_y}\right)^{y},$$

in terms of the underlying CMP variates and parameters where $x = 0, 1, \ldots$ and $y = 0, 1, \ldots$. This result is a natural extension of the relationship between the Poisson and binomial distributions.

[1] Kadane and Naeshagen (2013) study an application of the CMB distribution that restricts the CMB to only allow for positive or null correlation of dependent Bernoulli random variables ($\nu \leq 1$).

Considering a random sample $Y_1, \ldots, Y_n \sim \text{CMB}(m, p, \nu)$ with common number of trials m, Kadane (2016) and Borges et al. (2014) describe exponential family properties and generating functions of the CMB distribution. The distribution can be written in exponential family form as

$$\begin{aligned}
\text{P}(Y_1 = y_1, \ldots, Y_n = y_n) &\propto \prod_{i=1}^{n} \binom{m}{y_i}^{\nu} p^{y_i} (1-p)^{m-y_i} \\
&= (m!)^{n\nu}(1-p)^{nm} \prod_{i=1}^{n} \left(\frac{p}{1-p}\right)^{y_i} \frac{1}{(y_i!(m-y_i)!)^{\nu}} \\
&\propto \exp\left(\sum_{i=1}^{n} y_i \log\left(\frac{p}{1-p}\right) - \nu \sum_{i=1}^{n} \log\left[y_i!(m-y_i)!\right]\right),
\end{aligned}$$

with sufficient statistics $\sum_{i=1}^{n} y_i$ and $\sum_{i=1}^{n} \log\left[y_i!(m-y_i)!\right]$. The probability generating function of the CMB distribution is

$$\begin{aligned}
\text{E}(t^Y) &= \frac{1}{C(p,\nu)} \sum_{y=0}^{m} t^y \binom{m}{y}^{\nu} p^y (1-p)^{m-y} \\
&= \frac{(1-p)^m}{C(p,\nu)} \sum_{y=0}^{m} \binom{m}{y}^{\nu} \left(\frac{tp}{1-p}\right)^{y} \\
&= T\left(\frac{tp}{1-p}, \nu\right) \Big/ T\left(\frac{p}{1-p}, \nu\right), \qquad (2)
\end{aligned}$$

using $C(p, \nu) = (1-p)^m T\left(\frac{p}{1-p}, \nu\right)$ where $T(w, \nu) = \sum_{y=0}^{m} w^y \binom{m}{y}^{\nu}$. Similarly, the moment generating function is

$$\text{E}(e^{tY}) = T\left(\frac{e^t p}{1-p}, \nu\right) \Big/ T\left(\frac{p}{1-p}, \nu\right). \qquad (3)$$

3. Conway-Maxwell Extension of the Multinomial Distribution

The CMB distribution naturally extends to generalize the multinomial distribution. Define $\Omega_{m,k} = \{\boldsymbol{y} = (y_1, \ldots, y_k) \in \mathbb{N}^k : \sum_{j=1}^{k} y_j = m\}$ as the multinomial sample space based on m trials and k categories, and $\binom{m}{y_1 \cdots y_k} = \frac{m!}{y_1! \cdots y_k!}$ as the multinomial coefficient. Recall that there are $\binom{m+k-1}{k-1}$ points in the sample space $\Omega_{k,m}$; see for example Feller (1968, Chapter 2). A random variable $\boldsymbol{Y} = (Y_1, \cdots, Y_k)$ is distributed according to the Conway-Maxwell-multinomial (CMM) if it has pmf

$$\text{P}(\boldsymbol{Y} = \boldsymbol{y} \mid m, \boldsymbol{p}, \nu) = \frac{1}{C(\boldsymbol{p}, \nu)} \binom{m}{y_1 \cdots y_k}^{\nu} \prod_{j=1}^{k} p_j^{y_j}, \quad \boldsymbol{y} \in \Omega_{m,k}, \qquad (4)$$

and we will write $\boldsymbol{Y} \sim \text{CMM}_k(m, \boldsymbol{p}, \nu)$. Here,

$$C(\boldsymbol{p}, \nu) = \sum_{\boldsymbol{y} \in \Omega_{m,k}} \binom{m}{y_1 \cdots y_k}^{\nu} \prod_{j=1}^{k} p_j^{y_j} \tag{5}$$

is the normalizing constant. We derive the CMM pmf using the CMP conditioning approach as in Shmueli et al. (2005). Suppose $Y_j \sim \text{CMP}(\lambda_j, \nu)$ for independent Y_j, $j = 1, \ldots, k$, and let $S = \sum_{j=1}^{k} Y_j$. First, consider the probability distribution of the sum of CMP random variables:

$$\begin{aligned}
\text{P}(S = m) &= \sum_{\boldsymbol{y} \in \Omega_{m,k}} \text{P}(Y_1 = y_1, \ldots, Y_k = y_k) \\
&= \sum_{\boldsymbol{y} \in \Omega_{m,k}} \prod_{j=1}^{k} \left(\frac{\lambda_j^{y_j}}{(y_j!)^{\nu} Z(\lambda_j, \nu)} \right) \\
&= \frac{1}{\prod_{j=1}^{k} Z(\lambda_j, \nu)} \sum_{\boldsymbol{y} \in \Omega_{m,k}} \frac{\prod_{j=1}^{k} \lambda_j^{y_j}}{\left(\prod_{j=1}^{k} y_j! \right)^{\nu}} \\
&= \frac{\left(\sum_{j=1}^{k} \lambda_j \right)^m}{(m!)^{\nu} \prod_{j=1}^{k} Z(\lambda_j, \nu)} \sum_{\boldsymbol{y} \in \Omega_{m,k}} \left(\frac{m!}{y_1! \cdots y_k!} \right)^{\nu} \prod_{j=1}^{k} \left(\frac{\lambda_j}{\sum_{h=1}^{k} \lambda_h} \right)^{y_j} \\
&= \frac{\left(\sum_{j=1}^{k} \lambda_j \right)^m}{(m!)^{\nu} \prod_{j=1}^{k} Z(\lambda_j, \nu)} \sum_{\boldsymbol{y} \in \Omega_{m,k}} \binom{m}{y_1 \cdots y_k}^{\nu} \prod_{j=1}^{k} p_j^{y_j},
\end{aligned}$$

where $p_j = \frac{\lambda_j}{\sum_{h=1}^{k} \lambda_h}$. This result is an extension of the sum-of-Conway-Maxwell-Poissons (sCMP) class of distributions (Sellers et al., 2017) which allows each CMP component Y_j a different parameter λ_j. When $\lambda_1 = \cdots = \lambda_k$ the distribution reduces to the sCMP class of distributions. Next we obtain the form of the CMM distribution by conditioning $\boldsymbol{Y}$ on the sum S:

$$\begin{aligned}
\text{P}(\boldsymbol{Y} = \boldsymbol{y} \mid S = m) &= \frac{\text{P}(\boldsymbol{Y} = \boldsymbol{y}, S = m)}{\text{P}(S = m)} = \frac{\prod_{j=1}^{k} \text{P}(Y_j = y_j)}{\text{P}(S = m)} \\
&= \frac{\prod_{j=1}^{k} \left[\lambda_j^{y_j} \,/\, (y_j!)^{\nu} Z(\lambda_j, \nu) \right]}{\frac{\left(\sum_{j=1}^{k} \lambda_j \right)^m}{(m!)^{\nu} \prod_{j=1}^{k} Z(\lambda_j, \nu)} \sum_{\boldsymbol{y} \in \Omega_{m,k}} \binom{m}{y_1 \cdots y_k}^{\nu} \prod_{j=1}^{k} \left(\frac{\lambda_j}{\sum_{h=1}^{k} \lambda_h} \right)^{y_j}} \\
&= \frac{1}{C(\boldsymbol{p}, \nu)} \frac{(m!)^{\nu}}{\prod_{j=1}^{k} (y_j!)^{\nu}} \frac{\prod_{j=1}^{k} \lambda_j^{y_j}}{\left(\sum_{j=1}^{k} \lambda_j \right)^m} \\
&= \frac{1}{C(\boldsymbol{p}, \nu)} \binom{m}{y_1 \cdots y_k}^{\nu} \prod_{j=1}^{k} p_j^{y_j},
\end{aligned} \tag{6}$$

where $\boldsymbol{p} = \left(\frac{\lambda_1}{\sum_{h=1}^k \lambda_h}, \ldots, \frac{\lambda_k}{\sum_{h=1}^k \lambda_h}\right)$ is the set of probabilities.

The CMM distribution can be parameterized in terms of the original probability parameters $\boldsymbol{p}$ or the baseline odds $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_{k-1}\} = \left\{\frac{p_1}{p_k}, \ldots, \frac{p_{k-1}}{p_k}\right\}$, where the k^{th} category is taken as the baseline. The baseline odds parameterization relies on the pmf in (6) written as

$$\mathrm{P}(\boldsymbol{Y} = \boldsymbol{y} \mid m, \boldsymbol{\theta}, \nu) = \frac{1}{T(\boldsymbol{\theta}, \nu)} \binom{m}{y_1 \cdots y_k}^{\nu} \prod_{j=1}^{k-1} \theta_j^{y_j},$$

where

$$T(\boldsymbol{\theta}, \nu) = \sum_{\boldsymbol{y} \in \Omega_{m,k}} \binom{m}{y_1 \cdots y_k}^{\nu} \prod_{j=1}^{k-1} \theta_j^{y_j} = \frac{C(\boldsymbol{p}, \nu)}{p_k^m}$$

is the normalizing constant.

3.1 Special Cases

The CMM distribution results in standard distributions for special cases of the parameters ν and $\boldsymbol{p}$: multinomial ($\nu = 1$), discrete uniform on the multinomial sample space ($\nu = 0$ and $p_1 = \cdots = p_k$), point masses at the "vertex" points ($\nu \to -\infty$), and point masses at the "center" points ($\nu \to \infty$). Figure 1 presents a matrix of density plots illustrating special cases for $m = 20$ and $k = 3$; special cases are proven in the general setting. Three sets of $\boldsymbol{p}$ are depicted in Figure 1 to illustrate the behavior of the CMM distribution with equal and unequal probability parameters. The ordering of the categories in $\boldsymbol{p}$ is chosen without loss of generality. The simplest and immediately obvious special case is the multinomial distribution when $\nu = 1$. This case, depicted in the third row of Figure 1, serves as the baseline for interpreting over- and under-dispersion of the CMM relative to the multinomial distribution.

For $\nu = 0$ and $p_1 = \cdots = p_k$, the CMM distribution reduces to a discrete uniform distribution with the probability of each outcome in the multinomial sample space equal to $\binom{m+k-1}{k-1}^{-1}$. This can be seen as follows:

$$\begin{aligned} \mathrm{P}(\boldsymbol{Y} = \boldsymbol{y} \mid m, \boldsymbol{p} = (1/k, \ldots, 1/k), \nu = 0) &= \frac{\prod_{j=1}^k (1/k)^{y_j}}{\sum_{\boldsymbol{y} \in \Omega_{m,k}} \prod_{j=1}^k (1/k)^{y_j}} \\ &= \frac{(1/k)^m}{(1/k)^m \sum_{\boldsymbol{y} \in \Omega_{m,k}} 1} = \binom{m+k-1}{k-1}^{-1}. \end{aligned}$$

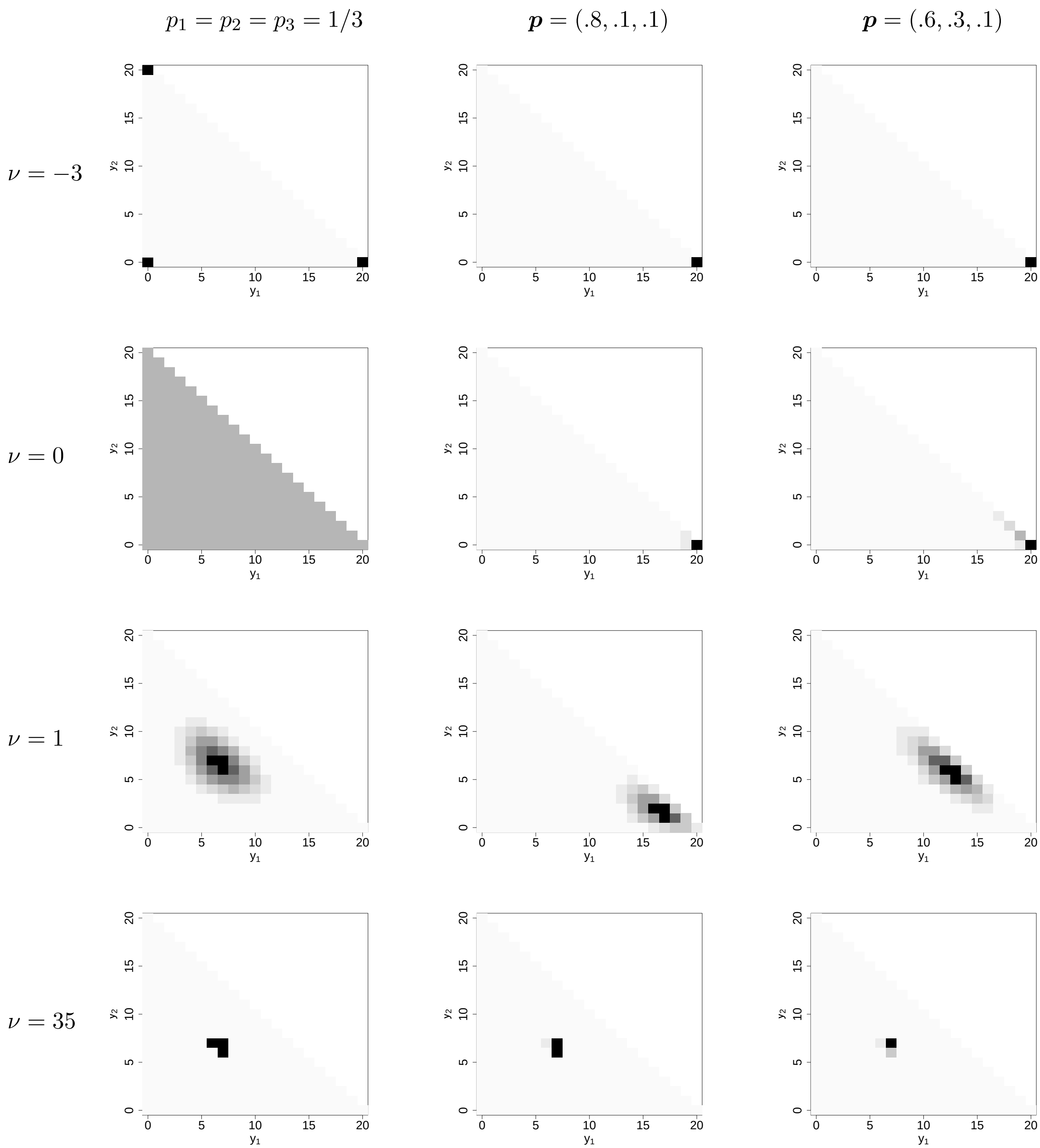

Figure 1: Special cases of the $\text{CMM}_3(m = 20, \boldsymbol{p}, \nu)$ density with varying ν and $\boldsymbol{p}$. Here, y_1 is plotted on the x-axis and y_2 is on the y-axis; $y_3 = m - y_1 - y_2$ is redundant and is therefore not shown. Darker squares represent higher probability points in the space. Density values represented by shades of gray are not consistent across plots.

Without the equality constraint on the probability parameters, the CMM does not reduce to a familiar form in this special case. For this special case with $k = 2$ (i.e. the CMB), the CMM density simplifies to

$$\begin{aligned} \mathrm{P}(Y_1 = y_1 \mid m, p_1, \nu = 0) &= \frac{p_1^{y_1}(1-p_1)^{m-y_1}}{\sum_{y_1=0}^{m} p_1^{y_1}(1-p_1)^{m-y_1}} \\ &= \frac{[p_1/(1-p_1)]^{y_1}}{\sum_{y_1=0}^{m} [p_1/(1-p_1)]^{y_1}} \\ &= \frac{\theta_1^{y_1}}{\sum_{y_1=0}^{m} \theta_1^{y_1}} = \frac{\theta_1^{y_1}(1-\theta_1)}{1-\theta_1^{m+1}} \,, \end{aligned}$$

where the last equality follows from the geometric series. However, this result is not easily generalized to larger k. Even without a standard form, the second row of Figure 1 illustrates that for $\nu = 0$ (i.e. a dispersion parameter value less than 1) the CMM distribution tends toward the behavior observed in the extreme case of over-dispersion ($\nu \to -\infty$).

To describe the CMM density as $\nu \to -\infty$, we refer to the points $m\boldsymbol{e}_j$ for $j = 1, \ldots, k$ as the vertex points of the multinomial sample space, where $\boldsymbol{e}_j$ is the j^{th} column of a $k \times k$ identity matrix. These vertex points correspond to the outcomes where all trials are assigned to the same category. For $\nu \to -\infty$, CMM becomes a distribution on vertex points $m\boldsymbol{e}_1, \ldots, m\boldsymbol{e}_k$ with probabilities proportional to $p_1, \ldots, p_k$. The multinomial coefficient drives this result because its value at the vertex points dictates the limiting probabilities. If $\boldsymbol{y} \notin \{m\boldsymbol{e}_1, \ldots, m\boldsymbol{e}_k\}$, then

$$\binom{m}{y_1 \cdots y_k} > 1 \quad \Longrightarrow \quad \binom{m}{y_1 \cdots y_k}^{\nu} \to 0 \quad \text{as } \nu \to -\infty$$

so that

$$\mathrm{P}(\boldsymbol{Y} = \boldsymbol{y} \mid m, \boldsymbol{p}, \nu) \to \begin{cases} p_j^m/(p_1^m + \cdots + p_k^m) & \text{if } \boldsymbol{y} = m\boldsymbol{e}_j \text{ for } j = 1, \ldots, k, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

as $\nu \to -\infty$. This behavior is exhibited at $\nu = -3$ for $m = 20$, which can be seen in the first row of Figure 1. Here we see that the mass is split evenly between the vertex points when $p_1 = \cdots = p_k$, but concentrates at points with larger p_j's otherwise.

For $\nu \to \infty$, the CMM distribution results in one or more point masses on the points of the multinomial support which are closest to the center of the sample space $(m/k, \ldots, m/k)$. The number of these points depends on the divisibility of m by k. Let q and r be integers such that $m = qk + r$ with $r \in \{0, 1, \ldots, k-1\}$. Consider assigning q trials to all k categories, and let each category have at most one of the remaining r trials. We define the $\binom{k}{r}$ such points of $\Omega_{m,k}$ as the center points, written as

$$\boldsymbol{y}^* \in \Omega^*_{m,k} = \{(q + r_1, \ldots, q + r_k) : r_j \in \{0, 1\}, r_1 + \cdots + r_k = r\} \,.$$

Similar to the $\nu \to -\infty$ case, the multinomial coefficient drives this special case because its value at the center points dictates the limiting probabilities. To see this, let $\boldsymbol{y} \in \Omega_{m,k} \setminus \Omega^*_{m,k}$:

$$\binom{m}{y_1^* \cdots y_k^*} > \binom{m}{y_1 \cdots y_k} \implies \binom{m}{y_1^* \cdots y_k^*}^{\nu} \Big/ \binom{m}{y_1 \cdots y_k}^{\nu} \to \infty \quad \text{as } \nu \to \infty,$$

so that

$$\mathrm{P}(\boldsymbol{Y} = \boldsymbol{y} \mid m, \boldsymbol{p}, \nu) \to \begin{cases} \frac{p_1^{y_1} \cdots p_k^{y_k}}{\sum_{y \in \Omega^*_{m,k}} p_1^{y_1} \cdots p_k^{y_k}} & \text{if } \boldsymbol{y} \in \Omega^*_{m,k}, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

as $\nu \to \infty$. This behavior is exhibited at $\nu = 35$ for $m = 20$, shown in the fourth row of Figure 1. In all scenarios varying $\boldsymbol{p}$, the three center points $(7, 7, 6)$, $(7, 6, 7)$, and $(6, 7, 7)$ constitute all the density. For a number of trials equally divisible by $k = 3$, say $m = 21$, the density would exhibit a single point mass at $(7, 7, 7)$; however, in this illustration the number of trials $m = 20$ is not equally divisible by the number of categories $k = 3$ resulting in $\binom{3}{2} = 3$ center points. In the equal probability case, the CMM density is equal at these three center points, but for unequal probabilities the density at the three points differs according to the probability mass function in (8).

These special cases of CMM suggest that $\boldsymbol{p}$ should not be interpreted as category probabilities for individual trials, as in the standard multinomial distribution, but as weights which influence where the probability mass shifts.

3.2 Intermediate Case Examples

Figure 2 presents a matrix of density plots illustrating intermediate cases of $\nu \in \{-.25, .25, 4\}$ for $m = 20$ and $k = 3$; special cases $\nu \in \{0, 1\}$ are included for reference.[2] For the equal $\boldsymbol{p}$ case shown in column 1 of Figure 2, the areas of highest density progress from concentrating around the vertex points ($\nu = -.25$) to clustering at the center points ($\nu = 4$) as ν increases. For the unequal $\boldsymbol{p}$ case shown in columns 2 and 3 of Figure 2, as ν increases to one, the areas of highest density progressively spread around the vertex point with the largest p_j and shift to the traditional multinomial distribution. As ν increases from one, the areas of highest density begin shifting to the center points and the clustering tightens as ν becomes large.

[2]Visit our R Shiny app at `https://dsteeg.shinyapps.io/CMMshinyapp` to independently explore the behavior of the CMM distribution for all values of ν, m and $\boldsymbol{p}$.

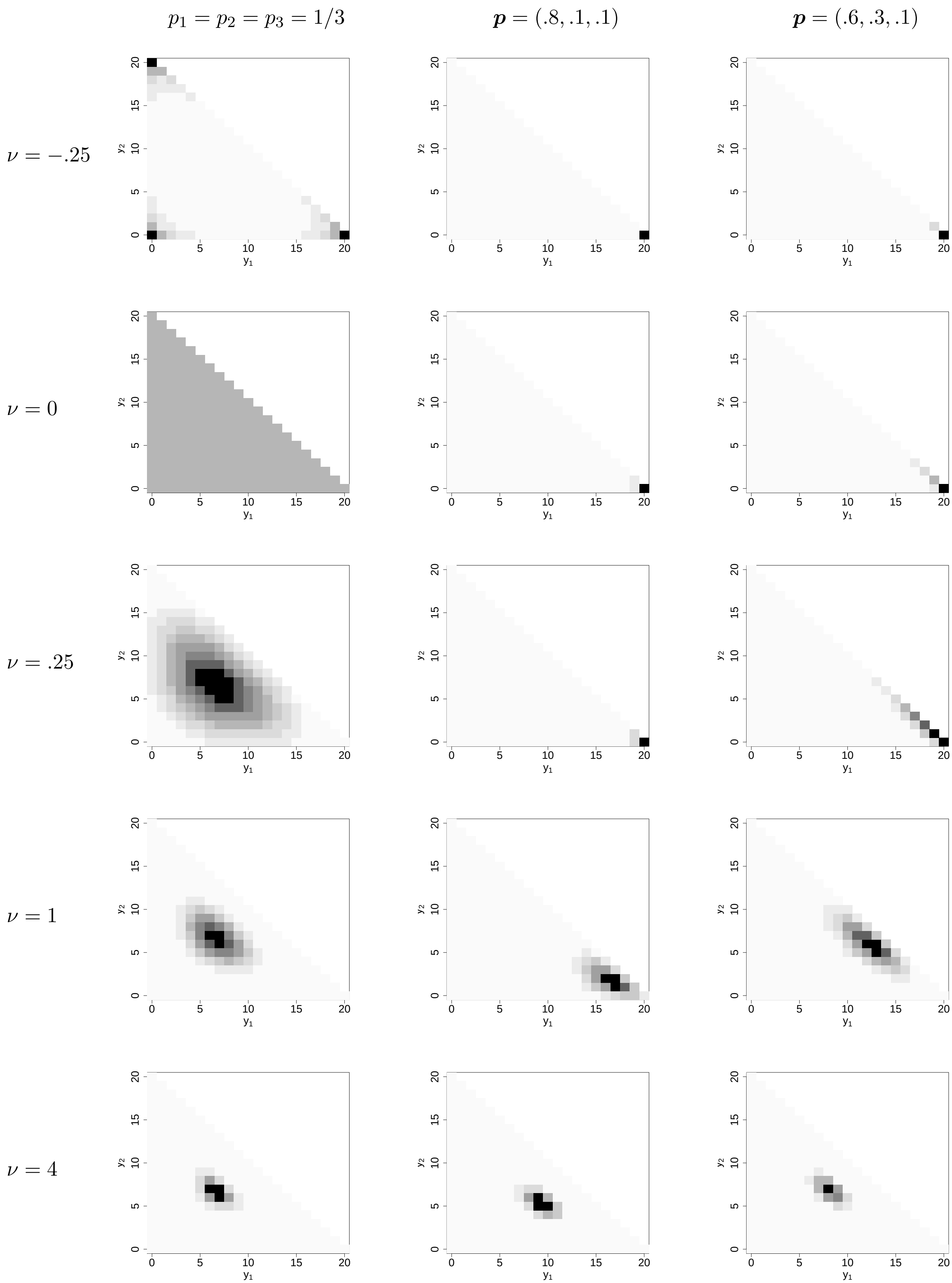

Figure 2: Intermediate cases of the $\text{CMM}_3(m = 20, \boldsymbol{p}, \nu)$ density with varying ν and $\boldsymbol{p}$. Here, y_1 is plotted on the x-axis and y_2 is on the y-axis; $y_3 = m - y_1 - y_2$ is redundant and is therefore not shown. Darker squares represent higher probability points in the space. Density values represented by shades of gray are not consistent across plots.

3.3 Likelihood in Exponential Family Form

Consider a random sample $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n \sim \text{CMM}_k(m, \boldsymbol{p}, \nu)$ with common number of trials m. The pmf can be written in exponential family form as

$$\begin{aligned}
\text{P}(\boldsymbol{Y}_1 = \boldsymbol{y}_1, \ldots, \boldsymbol{Y}_n = \boldsymbol{y}_n \mid m, \boldsymbol{p}, \nu) &= \frac{1}{C(\boldsymbol{p}, \nu)^n} \prod_{i=1}^{n} \binom{m}{y_{i1} \cdots y_{ik}}^{\nu} \prod_{j=1}^{k} p_j^{y_{ij}} \\
&= \frac{1}{T(\boldsymbol{\theta}, \nu)^n} \prod_{i=1}^{n} \left(\frac{m!}{\prod_{j=1}^{k} y_{ij}!} \right)^{\nu} \prod_{j=1}^{k-1} \theta_j^{y_{ij}} \\
&= \exp\left\{ \sum_{i=1}^{n} \sum_{j=1}^{k-1} y_{ij} \log(\theta_j) - \nu \sum_{i=1}^{n} \sum_{j=1}^{k} \log(y_{ij}!) \right\} \times \\
&\qquad \exp\left\{ \nu \sum_{i=1}^{n} \log(m!) - n \log\left(T(\boldsymbol{\theta}, \nu)\right) \right\} \\
&\propto \exp\left\{ \sum_{j=1}^{k-1} \log(\theta_j) S_{1j} - \nu S_2 \right\},
\end{aligned}$$

with sufficient statistics $S_1 = (y_{+1}, \ldots, y_{+,k-1})$ and $S_2 = \sum_{i=1}^{n} \sum_{j=1}^{k} \log(y_{ij}!)$, where $y_{+j} = \sum_{i=1}^{n} y_{ij}$ is the total count in category j. The natural parameters are ν and the set of baseline category logits, $\log(\theta_j) = \log\left(\frac{p_j}{p_k}\right)$ for $j = 1, \ldots, k-1$. This distributional form of the CMM is particularly important because exponential family distributions have many properties that are useful for statistical analysis.

3.4 Moments and Generating Functions

It is useful to consider the properties of the CMM distribution for both the probability and odds parameterizations. Let $\boldsymbol{p}_{-k} = (p_1, \ldots, p_{k-1})$, $\boldsymbol{I}$ be the $(k-1) \times (k-1)$ the identity matrix, $\boldsymbol{e}_j$ be the j^{th} column of $\boldsymbol{I}$, $\mathbf{1}$ be a vector of $k-1$ ones, and $1_{\ell=j}$ be an indicator function. The transpose of a vector $\boldsymbol{x}$ is denoted by $\boldsymbol{x}^\top$. The expected value for the j^{th} category of a CMM random variable $\boldsymbol{Y}$ for the odds and probability parameterizations are obtained in (9) and (10), respectively, as

$$\begin{aligned}
\text{E}(Y_j) &= \frac{1}{T(\boldsymbol{\theta}, \nu)} \sum_{\boldsymbol{y} \in \Omega_{m,k}} y_j \binom{m}{y_1 \cdots y_k}^{\nu} \prod_{i=1}^{k-1} \theta_i^{y_i} \\
&= \frac{\theta_j}{T(\boldsymbol{\theta}, \nu)} \frac{\partial T(\boldsymbol{\theta}, \nu)}{\partial \theta_j} \\
&= \theta_j \frac{\partial \log T(\boldsymbol{\theta}, \nu)}{\partial \theta_j} \qquad (9)
\end{aligned}$$

$$
\begin{aligned}
&= \theta_j \left[\boldsymbol{e}_j^\top \left(\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{p}_{-k}} \right)^{-1} \frac{\partial \log T(\boldsymbol{\theta}, \nu)}{\partial \boldsymbol{p}_{-k}} \right] \\
&= \theta_j \left[\boldsymbol{e}_j^\top \left(p_k^{-2} \left(p_k \boldsymbol{I} + \boldsymbol{p}_{-k} \boldsymbol{1}^\top \right) \right)^{-1} \frac{\partial \log T(\boldsymbol{\theta}, \nu)}{\partial \boldsymbol{p}_{-k}} \right] \\
&= \theta_j \left[\boldsymbol{e}_j^\top \left(p_k \boldsymbol{I} - p_k \boldsymbol{p}_{-k} \boldsymbol{1}^\top \right) \frac{\partial \log T(\boldsymbol{\theta}, \nu)}{\partial \boldsymbol{p}_{-k}} \right] \\
&= \frac{p_j}{p_k} \left[\sum_{\ell=1}^{k-1} \left(p_k 1_{\ell=j} - p_\ell p_k \right) \frac{\partial \log T(\boldsymbol{\theta}, \nu)}{\partial p_\ell} \right] \\
&= p_j \left[\sum_{\ell=1}^{k-1} \left(1_{\ell=j} - p_\ell \right) \left(\frac{m}{p_k} + \frac{\partial \log C(\boldsymbol{p}, \nu)}{\partial p_\ell} \right) \right] \\
&= p_j \left[\frac{m}{p_k} + \frac{\partial \log C(\boldsymbol{p}, \nu)}{\partial p_j} - \frac{m(1-p_k)}{p_k} - \sum_{\ell=1}^{k-1} p_\ell \frac{\partial \log C(\boldsymbol{p}, \nu)}{\partial p_\ell} \right] \\
&= m p_j + p_j \frac{\partial \log C(\boldsymbol{p}, \nu)}{\partial p_j} - \sum_{\ell=1}^{k-1} p_\ell \frac{\partial \log C(\boldsymbol{p}, \nu)}{\partial p_\ell}, \qquad (10)
\end{aligned}
$$

for $j = 1, \ldots, k-1$. The inverse of the matrix $p_k^{-2} \left(p_k \boldsymbol{I} + \boldsymbol{p}_{-k} \boldsymbol{1}^\top \right)$ can be obtained using the Sherman-Morrison matrix identity; e.g., see Meyer (2001, Section 3.8). The expected value for the k^{th} category is $\mathrm{E}(Y_k) = m - \sum_{j=1}^{k-1} \mathrm{E}(Y_j)$. For the special case $\nu = 1$, $C(\boldsymbol{p}, \nu) \equiv 1$ for all $\boldsymbol{p}$ gives $\partial \log C(\boldsymbol{p}, \nu) / \partial p_\ell = 0$, and (10) reduces to $m p_j$, the multinomial expected value for the count in category j.

To derive the variance and covariance for categories of a CMM random variable under the odds parameterization, we find that for two categories $j \neq h$

$$
\begin{aligned}
\mathrm{E}(Y_j Y_h) &= \frac{1}{T(\boldsymbol{\theta}, \nu)} \sum_{\boldsymbol{y} \in \Omega_{m,k}} y_j y_h \binom{m}{y_1 \cdots y_k}^{\nu} \prod_{i=1}^{k-1} \theta_i^{y_i} \\
&= \frac{\theta_j \theta_h}{T(\boldsymbol{\theta}, \nu)} \frac{\partial^2 \, T(\boldsymbol{\theta}, \nu)}{\partial \theta_j \partial \theta_h} \\
&= \theta_j \theta_h \left[\frac{\partial}{\partial \theta_j} \left(\frac{1}{T(\boldsymbol{\theta}, \nu)} \frac{\partial \, T(\boldsymbol{\theta}, \nu)}{\partial \theta_h} \right) \right] \\
&= \theta_j \theta_h \left[\frac{1}{T(\boldsymbol{\theta}, \nu)} \frac{\partial^2 \, T(\boldsymbol{\theta}, \nu)}{\partial \theta_j \partial \theta_h} + \frac{1}{(T(\boldsymbol{\theta}, \nu))^2} \frac{\partial \, T(\boldsymbol{\theta}, \nu)}{\partial \theta_j} \frac{\partial \, T(\boldsymbol{\theta}, \nu)}{\partial \theta_h} \right] \\
&= \theta_j \theta_h \frac{\partial^2 \log T(\boldsymbol{\theta}, \nu)}{\partial \theta_j \partial \theta_h} + \mathrm{E}(Y_j) \, \mathrm{E}(Y_h)
\end{aligned}
$$

and similarly for one category $j = h$

$$\begin{aligned}
\mathrm{E}(Y_j(Y_j-1)) &= \frac{1}{T(\boldsymbol{\theta},\nu)} \sum_{\boldsymbol{y}\in\Omega_{m,k}} y_j(y_j-1)\binom{m}{y_1 \cdots y_k}^{\nu} \prod_{i=1}^{k-1} \theta_i^{y_i} \\
&= \frac{\theta_j^2}{T(\boldsymbol{\theta},\nu)} \frac{\partial^2 T(\boldsymbol{\theta},\nu)}{\partial \theta_j^2} \\
&= \theta_j^2 \frac{\partial^2 \log T(\boldsymbol{\theta},\nu)}{\partial \theta_j^2} + [\mathrm{E}(Y_j)]^2 .
\end{aligned}$$

Therefore,

$$\mathrm{Cov}(Y_j, Y_h) = \mathrm{E}(Y_j Y_h) - \mathrm{E}(Y_j)\,\mathrm{E}(Y_h) = \theta_j \theta_h \frac{\partial^2 \log T(\boldsymbol{\theta},\nu)}{\partial \theta_j \partial \theta_h} = \theta_j \frac{\partial\, \mathrm{E}(Y_h)}{\partial \theta_j} = \theta_h \frac{\partial\, \mathrm{E}(Y_j)}{\partial \theta_h}$$

and

$$\begin{aligned}
\mathrm{Var}(Y_j) &= \mathrm{E}[Y_j(Y_j-1)] + \mathrm{E}(Y_j) - [\mathrm{E}(Y_j)]^2 = \theta_j^2 \frac{\partial^2 \log T(\boldsymbol{\theta},\nu)}{\partial \theta_j^2} + \theta_j \frac{\partial \log T(\boldsymbol{\theta},\nu)}{\partial \theta_j} \\
&= \theta_j \left[\frac{\partial}{\partial \theta_j} \theta_j \frac{\partial \log T(\boldsymbol{\theta},\nu)}{\partial \theta_j} \right] = \theta_j \frac{\partial\, \mathrm{E}(Y_j)}{\partial \theta_j}.
\end{aligned}$$

The probability generating function of the CMM distribution is

$$\begin{aligned}
\Pi_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathrm{E}\left(\prod_{j=1}^{k} t_j^{Y_j} \right) &= \frac{1}{C(\boldsymbol{p},\nu)} \sum_{\boldsymbol{y}\in\Omega_{m,k}} \left[\left(\prod_{j=1}^{k} t_j^{y_j} \right) \binom{m}{y_1 \cdots y_k}^{\nu} \prod_{j=1}^{k} p_k^{y_j} \right] \\
&= \frac{1}{C(\boldsymbol{p},\nu)} \sum_{\boldsymbol{y}\in\Omega_{m,k}} \binom{m}{y_1 \cdots y_k}^{\nu} \prod_{j=1}^{k} (t_j p_j)^{y_j} \\
&= C\left((t_1 p_1, \ldots, t_k p_k), \nu\right) \,/\, C(\boldsymbol{p},\nu)
\end{aligned}$$

in terms of the original parameters $\boldsymbol{p}$, and

$$\begin{aligned}
\Pi_{\boldsymbol{Y}}(\boldsymbol{t}) &= \frac{t_k^m}{T(\boldsymbol{\theta},\nu)} \sum_{\boldsymbol{y}\in\Omega_{m,k}} \binom{m}{y_1 \cdots y_k}^{\nu} \prod_{j=1}^{k-1} \left(\frac{t_j p_j}{t_k p_k} \right)^{y_j} \\
&= t_k^m T\left(\left(\frac{t_1}{t_k}\frac{p_1}{p_k}, \ldots, \frac{t_{k-1}}{t_k}\frac{p_{k-1}}{p_k} \right), \nu \right) \,/\, T(\boldsymbol{\theta},\nu)
\end{aligned}$$

in terms of the baseline odds $\boldsymbol{\theta}$. Note that the probability generating function for the non-baseline categories $(1, \ldots, k-1)$ is obtained by setting $t_k = 1$. Similarly the moment generating function is

$$M_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathrm{E}\left(\prod_{j=1}^{k} e^{t_j Y_j}\right) = C\left((e^{t_1}p_1, \ldots, e^{t_k}p_k), \nu\right) \,/\, C(\boldsymbol{p}, \nu)$$
$$= e^{mt_k} T\left(\left(\frac{e^{t_1}}{e^{t_k}}\frac{p_1}{p_k}, \ldots, \frac{e^{t_{k-1}}}{e^{t_k}}\frac{p_{k-1}}{p_k}\right), \nu\right) \,/\, T(\boldsymbol{\theta}, \nu).$$

The CMM moment and probability generating functions naturally reduce to the corresponding functions for the CMB distribution, (2) and (3), in the special case of $k = 2$.

3.5 Marginal and Conditional Distributions

The family of traditional multinomial distributions is closed under some useful manipulations; specifically, marginals, conditionals, and grouping coordinates together all result in another multinomial distribution (which is also binomial if two categories remain). The CMM family of distributions is closed under some of these manipulations. Some notation will be useful in the following derivations. Suppose (A, B) is a partition of the index set $\{1, \ldots, k\}$ into nonempty sets with lengths $|A|$ and $|B|$ where $\boldsymbol{Y} = (\boldsymbol{Y}_A, \boldsymbol{Y}_B) \sim \mathrm{CMM}_k(m, \boldsymbol{p}, \nu)$. Let $\boldsymbol{Y}_A = (Y_j : j \in A)$, $\boldsymbol{y}_A = (y_j : j \in A)$, $y_A^+ = \sum_{j \in A} y_j$, $\boldsymbol{p}_A = (p_j : j \in A)$, $\tilde{\boldsymbol{p}}_A = (p_j / p_A^+ : j \in A)$, $p_A^+ = \sum_{j \in A} p_j$ and $\Omega_A = \{\boldsymbol{y}_A \in \mathbb{N}^{k-|B|} : y_A^+ = m - y_B^+\}$.

Computing the marginal distribution for $\boldsymbol{Y}_A$, we have

$$\mathrm{P}(\boldsymbol{Y}_A = \boldsymbol{y}_A \mid m, \boldsymbol{p}, \nu) = \sum_{\boldsymbol{y}_B \in \Omega_B} \frac{1}{C(\boldsymbol{p}, \nu)} \binom{m}{\boldsymbol{y}_A \; \boldsymbol{y}_B}^{\nu} \prod_{j=1}^{k} p_j^{y_j}$$
$$= \frac{1}{C(\boldsymbol{p}, \nu)} \binom{m}{\boldsymbol{y}_A \; m - y_A^+}^{\nu} \prod_{j \in A} p_j^{y_j} \left[\sum_{\boldsymbol{y}_B \in \Omega_B} \binom{m - y_A^+}{\boldsymbol{y}_B}^{\nu} \prod_{j \in B} p_j^{y_j}\right]$$
$$= \frac{1}{C(\boldsymbol{p}, \nu)} \binom{m}{\boldsymbol{y}_A \; m - y_A^+}^{\nu} \prod_{j \in A} p_j^{y_j} \left(1 - \sum_{j \in A} p_j\right)^{m - y_A^+} \times$$
$$\left[\sum_{\boldsymbol{y}_B \in \Omega_B} \binom{m - y_A^+}{\boldsymbol{y}_B}^{\nu} \prod_{j \in B} p_j^{y_j} \left(\sum_{j \in B} p_j\right)^{-y_B^+}\right]$$
$$= \frac{C\left(\tilde{\boldsymbol{p}}_B, \nu; m - y_A^+\right)}{C(\boldsymbol{p}, \nu; m)} \binom{m}{\boldsymbol{y}_A \; m - y_A^+}^{\nu} \prod_{j \in A} p_j^{y_j} \left(1 - \sum_{j \in A} p_j\right)^{m - y_A^+} \tag{11}$$

We have added a third argument to the constant $C(\cdot, \cdot, \cdot)$ to emphasize the number of trials, which can now vary in different parts of the expression. Note that the number of categories is now $k' =$

$|A|+1$, where the added one comes from collapsing the coordinates B into a single count y_B^+, which is redundant in the sense that $y_B^+ = m - y_A^+$. We observe that (11) is a not a CMM distribution, as it involves a ratio of CMM normalizing constants, nor is it another recognizable family. This marginal distribution depends on the entire $\boldsymbol{p}$ and ν even when $k' < k$. For the special case of $\nu = 1$, however, the ratio of CMM normalizing constants simplifies to 1, and (11) becomes a $\text{CMM}_{k'}(m, \boldsymbol{p}_A, \nu = 1)$; this follows from the usual multinomial distribution. An important special case of (11) is the count for one category with $A = \{\ell\}$ and $k' = 2$, which gives

$$\mathrm{P}(Y_\ell = y_\ell \mid m, \boldsymbol{p}, \nu) \quad = \quad \frac{C(\tilde{\boldsymbol{p}}_B, \nu; m - y_\ell)}{C(\boldsymbol{p}, \nu; m)} \binom{m}{y_\ell \; m - y_\ell}^\nu p_\ell^{y_\ell} (1 - p_\ell)^{m - y_\ell}; \tag{12}$$

this again is not equivalent to a CMB distribution (1) because

$$C(p_\ell, \nu; m) \neq C(\boldsymbol{p}, \nu; m) / C(\tilde{\boldsymbol{p}}_B, \nu; m - y_\ell), \tag{13}$$

except in special cases such as $\nu = 1$.

The conditional distribution of $\boldsymbol{Y}_A$ given $\boldsymbol{Y}_B$ is

$$\begin{aligned}
\mathrm{P}(\boldsymbol{Y}_A = \boldsymbol{y}_A \mid \boldsymbol{Y}_B = \boldsymbol{y}_B, m, \boldsymbol{p}, \nu) &= \left[\frac{1}{C(\boldsymbol{p}, \nu)} \binom{m}{\boldsymbol{y}_A \; \boldsymbol{y}_B}^\nu \prod_{j=1}^k p_j^{y_j} \right] \left[\sum_{\boldsymbol{y}_A \in \Omega_A} \frac{1}{C(\boldsymbol{p}, \nu)} \binom{m}{\boldsymbol{y}_A \; \boldsymbol{y}_B}^\nu \prod_{j=1}^k p_j^{y_j} \right]^{-1} \\
&= \binom{m - y_B^+}{\boldsymbol{y}_A}^\nu \prod_{j \in A} p_j^{y_j} \left(\sum_{j \in A} p_j \right)^{-y_A^+} \times \\
&\qquad \left[\sum_{\boldsymbol{y}_A \in \Omega_A} \binom{m - y_B^+}{\boldsymbol{y}_A}^\nu \prod_{j \in A} p_j^{y_j} \left(\sum_{j \in A} p_j \right)^{-y_A^+} \right]^{-1} \\
&= \binom{m - y_B^+}{\boldsymbol{y}_A}^\nu \prod_{j \in A} \tilde{p}_j^{y_j} \left[\sum_{\boldsymbol{y}_A \in \Omega_A} \binom{m - y_B^+}{\boldsymbol{y}_A}^\nu \prod_{j \in A} \tilde{p}_j^{y_j} \right]^{-1} \\
&= \frac{1}{C(\tilde{\boldsymbol{p}}_A, \nu; m - y_B^+)} \binom{m - y_B^+}{\boldsymbol{y}_A}^\nu \prod_{j \in A} \tilde{p}_j^{y_j},
\end{aligned}$$

for $\boldsymbol{y}_A \in \Omega_{m - y_B^+, |A|}$. This is a $\text{CMM}_{k''}(m - y_B^+, \tilde{\boldsymbol{p}}_A, \nu)$ distribution, with $k'' = |A|$ categories which are constrained to sum to $m - y_B^+$. In particular, the conditional distribution for the count in $k'' = 2$ categories with $A = \{\ell, h\}$ is

$$\begin{aligned}
&\mathrm{P}(Y_\ell = y_\ell, Y_h = m - y_B^+ - y_\ell \mid \boldsymbol{Y}_B = \boldsymbol{y}_B, m, \boldsymbol{p}, \nu) \\
&\quad = \frac{1}{C((\tilde{p}_\ell, \tilde{p}_h), \nu; m - y_B^+)} \binom{m - y_B^+}{y_\ell \;\; m - y_B^+ - y_\ell}^\nu \tilde{p}_\ell^{y_\ell} \tilde{p}_h^{m - y_B^+ - y_\ell}
\end{aligned}$$

$$= \frac{1}{C\left(\tilde{p}_\ell, \nu; m - y_B^+\right)} \binom{m - y_B^+}{y_\ell \;\; m - y_B^+ - y_\ell}^\nu \tilde{p}_\ell^{y_\ell} \left(1 - \tilde{p}_\ell\right)^{m - y_B^+ - y_\ell}, \tag{14}$$

which corresponds to CMB$(m - y_B^+, \frac{p_\ell}{p_\ell + p_h}, \nu)$. This result is particularly useful for devising a Gibbs sampler (Robert and Casella, 2010) to draw $\boldsymbol{Y} \sim \text{CMM}_k(m, \boldsymbol{p}, \nu)$ based on a series of draws from the CMB distribution; see Algorithm 1. The tail of a sufficiently long chain will approximate draws from the desired CMM distribution. For the j^{th} step, $j = 1, \ldots, k-1$, we take $A = \{j, k\}$ and use (14) to identify the conditional distribution of $\boldsymbol{Y}_A \mid \boldsymbol{Y}_B$; the k^{th} category is always taken to be second free category, without loss of generality. A more naive method of sampling from CMM would be to precompute all probabilities for the multinomial sample space and use Algorithm 2, which is typically used to handle finite discrete distributions. However, the size of the sample space quickly becomes intractable as m and k increase. The Gibbs sampler requires exact draws from a series of CMB distributions, which can each be done efficiently as follows: compute all unnormalized probabilities (i.e. the numerators only) as in (1), divide by their sum to get normalized probabilities, and invoke Algorithm 2.

Algorithm 1 Produce a chain of R draws approximating draws from $\text{CMM}_k(m, \boldsymbol{p}, \nu)$, starting from initial value $\boldsymbol{y}^{(0)} \in \Omega_{m,k}$.

```
function GibbsSampler(R, y^(0), m, p, ν)
    Let y = y^(0).
    for r = 1, ..., R do
        Draw y_1 ~ CMB(m − y_B^+, p_1/(p_1+p_k), ν) and let y_k = m − y_B^+ − y_1.
        Draw y_2 ~ CMB(m − y_B^+, p_2/(p_2+p_k), ν) and let y_k = m − y_B^+ − y_2.
        ...
        Draw y_{k−1} ~ CMB(m − y_B^+, p_{k−1}/(p_{k−1}+p_k), ν) and let y_k = m − y_B^+ − y_{k−1}.
        Let y^(r) = y.
    return y^(1), ..., y^(R).
```

Algorithm 2 Draw an element from $\boldsymbol{s} = (s_1, \ldots, s_d)$ with probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_d)$.

```
function DrawDiscrete(s, π)
    Let Π_j = π_1 + ··· + π_j, j = 1, ... d be the cumulative probabilities.
    Draw u from Uniform(0, 1).
    if u ∈ [0, Π_1) then return s_1.
    else if u ∈ [Π_1, Π_2) then return s_2.
        ...
    else if u ∈ [Π_{d−2}, Π_{d−1}) then return s_{d−1}.
    else return s_d.
```

The CMM distribution is not closed to the grouping of categories. Consider $(A_1, \ldots, A_K)$ a partition of the index set $\{1, \dots, k\}$. The distribution of the grouped categories is

$$\begin{aligned}
&\mathrm{P}(Y^+_{A_1} = y^+_{A_1}, \dots, Y^+_{A_K} = y^+_{A_K} \mid m, \boldsymbol{p}, \nu) \\
&\quad = \sum_{\boldsymbol{y}_{A_1} \in \Omega_{y^+_{A_1}, |A_1|}} \cdots \sum_{\boldsymbol{y}_{A_K} \in \Omega_{y^+_{A_K}, |A_K|}} \frac{m^\nu}{C(\boldsymbol{p}, \nu)} \binom{1}{\boldsymbol{y}_{A_1}}^\nu \cdots \binom{1}{\boldsymbol{y}_{A_K}}^\nu \prod_{j \in A_1} p_j^{y_j} \cdots \prod_{j \in A_K} p_j^{y_j} \\
&\quad = \frac{1}{C(\boldsymbol{p}, \nu)} \binom{m}{y^+_{A_1} \cdots y^+_{A_K}}^\nu \left[\sum_{\boldsymbol{y}_{A_1} \in \Omega_{y^+_{A_1}, |A_1|}} \binom{y^+_{A_1}}{\boldsymbol{y}_{A_1}}^\nu \prod_{j \in A_1} p_j^{y_j} \right] \times \quad \cdots \\
&\qquad\qquad\qquad\qquad \times \left[\sum_{\boldsymbol{y}_{A_K} \in \Omega_{y^+_{A_K}, |A_K|}} \binom{y^+_{A_K}}{\boldsymbol{y}_{A_K}}^\nu \prod_{j \in A_K} p_j^{y_j} \right] \\
&\quad = \frac{C\left(\tilde{\boldsymbol{p}}_{A_1}, \nu; y^+_{A_1}\right) \cdots C\left(\tilde{\boldsymbol{p}}_{A_K}, \nu; y^+_{A_K}\right)}{C(\boldsymbol{p}, \nu)} \binom{m}{y^+_{A_1} \cdots y^+_{A_K}}^\nu \prod_{\ell=1}^K \left(p^+_{A_\ell}\right)^{y^+_{A_\ell}}. \qquad (15)
\end{aligned}$$

Similar to the marginal distribution, we observe that (15) is not a CMM distribution because the term involving the normalizing constants does not reduce. For the special case of $\nu = 1$, however, the term involving the normalizing constants simplifies to 1 and (15) becomes $\text{CMM}_K(m, \boldsymbol{p}^+, \nu = 1)$ with $\boldsymbol{p}^+ = (p^+_{A_1}, \dots, p^+_{A_K})$; this follows from the usual multinomial distribution.

4. Conclusion

This paper introduces a Conway-Maxwell-multinomial (CMM) distribution as a flexible alternative to the traditional multinomial distribution. We formally extend the CMB distribution to the setting of more than two categories by deriving the CMM from a combination of more than two CMP random variables. Akin to the versatility of the CMP relative to the Poisson distribution and the CMB relative to the binomial distribution, CMM allows for both over- and under-dispersion as compared to the multinomial distribution. The CMM captures extreme cases of dispersion that exhibit mass points on the support of the multinomial sample space, as well as all intermediate cases that exhibit all varieties of distributional spread. CMM is an exponential family distribution which naturally yields properties such as moments and generating functions. Conditional distributions of a CMM random variable are again in the CMM family, similarly to conditionals of the standard multinomial distribution being multinomial. However, unlike the standard multinomial, marginal and grouping distributions of CMM are not CMM distributions.

The flexibility offered by CMM comes at the cost of additional computational challenges. Complete enumeration of the terms in the normalizing constant and related quantities is a reasonable strategy for CMB and CMP (with appropriate truncation), but becomes intractable for CMM as the multinomial sample space grows. Addressing these challenges to make CMM practical for statistical applications is a topic of future work.

References

P. Borges, J. Rodrgues, N. Balakrishnan, and J. Bazan. A COM-Poisson type generalization of the binomial distribution and its properties and applications. *Statistics and Probability Letters*, 87:158–166, 2014.

P. C. Consul and G. C. Jain. A generalization of the Poisson distribution. *Technometrics*, 15:791–799, 1973.

R. W. Conway and W. L. Maxwell. A queuing model with state dependent service rates. *Journal of Industrial Engineering*, 12:132–136, 1962.

W. Feller. *An Introduction to Probability Theory and Its Applications*. Wiley, 3rd edition, 1968.

J. B. Kadane. Sums of possibly associated Bernoulli variables: the Conway-Maxwell-binomial distribution. *Bayesian Analysis*, 11(2):403–420, 2016.

J. B. Kadane and F. L. Naeshagen. The number of killings in southern rural Norway, 1300–1569. *The Annal of Applied Statistics*, 7(2):846–859, 2013.

C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2001.

J. G. Morel and N. K. Nagaraj. A finite mixture distribution for modelling multinomial extra variation. *Biometrika*, 80:363–371, 1993.

J. E. Mosimann. On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. *Biometrika*, 49(1/2):65–82, 1962.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2nd edition, 2010.

K. F. Sellers and D. S. Morris. Underdispersion models: Models that are "under the radar". *Communications in Statistics - THeory and Methods*, 46:12075–12086, 2017.

K. F. Sellers, S. Borle, and G. Shmueli. The COM-Poisson model for count data: a survey of methods and applications. *Applied Stochastic Models in Business and Industry*, 28:104–116, 2012.

K. F. Sellers, A. W. Swift, and K. S. Weems. A flexible distribution class for count data. *Journal of Statistical Distributions and Applications*, 4, 2017.

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. *Applied Statistics*, 54:127–142, 2005.