

Article

A Flexible Multivariate Distribution for Correlated Count Data

Kimberly F. Sellers ^{1,2,*} , Tong Li ¹, Yixuan Wu ¹ and Narayanaswamy Balakrishnan ³ 

¹ Department of Mathematics and Statistics, Georgetown University, Washington, DC 20057, USA; tl728@georgetown.edu (T.L.); wy117@georgetown.edu (Y.W.)

² Center for Statistical Research and Methodology, U. S. Census Bureau, Washington, DC 20233, USA

³ Department of Mathematics and Statistics, McMaster University, Hamilton, ON L8S 4K1, Canada; bala@mcmaster.ca

* Correspondence: kfs7@georgetown.edu

Abstract: Multivariate count data are often modeled via a multivariate Poisson distribution, but it contains an underlying, constraining assumption of data equi-dispersion (where its variance equals its mean). Real data are oftentimes over-dispersed and, as such, consider various advancements of a negative binomial structure. While data over-dispersion is more prevalent than under-dispersion in real data, however, examples containing under-dispersed data are surfacing with greater frequency. Thus, there is a demonstrated need for a flexible model that can accommodate both data types. We develop a multivariate Conway–Maxwell–Poisson (MCMP) distribution to serve as a flexible alternative for correlated count data that contain data dispersion. This structure contains the multivariate Poisson, multivariate geometric, and the multivariate Bernoulli distributions as special cases, and serves as a bridge distribution across these three classical models to address other levels of over- or under-dispersion. In this work, we not only derive the distributional form and statistical properties of this model, but we further address parameter estimation, establish informative hypothesis tests to detect statistically significant data dispersion and aid in model parsimony, and illustrate the distribution’s flexibility through several simulated and real-world data examples. These examples demonstrate that the MCMP distribution performs on par with the multivariate negative binomial distribution for over-dispersed data, and proves particularly beneficial in effectively representing under-dispersed data. Thus, the MCMP distribution offers an effective, unifying framework for modeling over- or under-dispersed multivariate correlated count data that do not necessarily adhere to Poisson assumptions.

Keywords: multivariate Poisson; multivariate Bernoulli; multivariate geometric; Conway–Maxwell–Poisson; confounding; over-dispersion; under-dispersion; dependence



Citation: Sellers, K.F.; Li, T.; Wu, Y.; Balakrishnan, N. A Flexible Multivariate Distribution for Correlated Count Data. *Stats* **2021**, *4*, 308–326. <https://doi.org/10.3390/stats4020021>

Academic Editor:
Célestin C. Kokonendji

Received: 17 March 2021
Accepted: 11 April 2021
Published: 15 April 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There exists a rich history of research regarding multivariate discrete distributions [1]. Krishnamoorthy [2] introduced a multivariate binomial (MB) distribution for the d -dimensional vector $\mathbf{B} = (B_1, B_2, \dots, B_d)'$ from a 2^d table with a factorial moment generating function (fmgf)

$$G_{\mathbf{B}}(\boldsymbol{\alpha}) = \left[1 + \sum_{i=1}^d p_i^* \alpha_i + \sum_{1 \leq i < j \leq d} p_{ij}^* \alpha_i \alpha_j + \dots + p_{12\dots d}^* \alpha_1 \alpha_2 \dots \alpha_d \right]^n, \quad (1)$$

where p_i^* is the probability of B_i , $i = 1, \dots, d$; p_{ij}^* denotes the probability of $B_i B_j$; and so on. Utilizing this form, Krishnamoorthy [2] further introduced the multivariate Poisson (MP) distribution as the limiting distribution of the multivariate binomial distribution wherein all of the probabilities appearing in Equation (1) have order $O(1/n)$ and $np_o^* \rightarrow \lambda_o$.

as $n \rightarrow \infty$, where $\circ$ denotes the corresponding probability subscripts in Equation (1). Accordingly, the fmfg of the MP distribution for a random vector $\mathbf{X} = (X_1, X_2, \dots, X_d)'$ is

$$G_{\mathbf{X}}(\boldsymbol{\alpha}) = \exp \left[\sum_{i=1}^d \lambda_i \alpha_i + \sum_{1 \leq i < j \leq d} \lambda_{ij} \alpha_i \alpha_j + \dots + \lambda_{12\dots d} \alpha_1 \alpha_2 \dots \alpha_d \right]. \quad (2)$$

Mahamunulu [3] noted that the MP distribution can likewise be derived by defining $\mathbf{X} = (X_1, X_2, \dots, X_d)$ as the sum of independent $\text{Poisson}(A_*)$ random variables Y_* , where $*$ denotes all subscripts involving $* = i \in \{1, 2, \dots, d\}$ with Y_* distributed as $\text{Poisson}(A_*)$, and X_* as $\text{Poisson}(\mu_*)$ where μ_* denotes the sum of the associated A_* parameters with subsets $j_1, j_2, \dots, j_s$ for $s \in \{1, 2, \dots, d\}$ such that $j_1 < j_2 < \dots < j_s$. The corresponding joint probability generating function (pgf) has the form

$$\Pi_{\mathbf{X}}(\mathbf{s}) = \exp \left[\sum_{i=1}^d A_i s_i + \sum_{1 \leq i < j \leq d} A_{ij} s_i s_j + \dots + A_{12\dots d} s_1 s_2 \dots s_d - A \right], \quad (3)$$

where $A = \sum_{i=1}^d A_i + \sum_{1 \leq i < j \leq d} A_{ij} + \dots + A_{12\dots d}$ [3,4]. From (3), it is evident that the variables $X_1, X_2, \dots, X_d$ have marginal Poisson distributions, and it can be further shown that all pairs of variables X_i 's are positively correlated.

While the MP distribution is a popular model for describing correlated discrete random variables, it is well known that Poisson models are constrained by their underlying assumption of equi-dispersion; analogous negative binomial (NB) models serve as a popular alternative due to their ability to address data over-dispersion [5]. Doss [6] discussed a multivariate negative binomial (MNB) distribution with joint pgf

$$\Pi(\mathbf{s}) = \left[a_0 + \sum_{i=1}^d a_i s_i + \sum_{1 \leq i < j \leq d} a_{ij} s_i s_j + \dots + a_{12\dots d} s_1 s_2 \dots s_d \right]^{-k} \quad (4)$$

for $k > 0$. From (4), it is evident that the variables $X_1, X_2, \dots, X_d$ have marginal NB distributions which are known to be over-dispersed. For this reason, the MNB distribution can only accommodate data over-dispersion; accordingly, correlated under-dispersed data structures are only at best fitted by a MP model where the associated model parameters will still be biased. Therefore, in this work, we introduce the reader to the Conway–Maxwell–Poisson (CMP) distribution and develop a multivariate CMP (MCMP) distribution as a flexible alternative distribution for modeling correlated discrete count data. Section 2 introduces the reader to the CMP distribution and its bivariate analog as motivation. Section 3 develops the MCMP distribution and discusses its associated properties, and also introduces approaches for parameter estimation and hypothesis testing. Section 4 demonstrates the model flexibility by means of simulated and real data examples. Finally, Section 5 concludes the manuscript with discussion, while the appendices contain more detailed derivations and the datasets referenced in this work.

2. Conway–Maxwell–Poisson Distribution

The CMP(λ, ν) distribution [7] has the probability mass function (pmf)

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

for a random variable Y , where $\nu \geq 0$ is the dispersion parameter, $\lambda = E(Y^\nu)$ generalizes the Poisson rate parameter, and $Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$ denotes the normalizing constant. Equi-dispersion relative to the Poisson distribution is represented when $\nu = 1$ while data over-dispersion (under-dispersion) occurs when $\nu < (>)1$. The CMP(λ, ν) distribution contains three well-known distributions as special cases: Poisson with rate parameter λ

when $\nu = 1$; geometric with success probability $1 - \lambda$ when $\nu = 0$ and $\lambda < 1$; and Bernoulli with success probability $\frac{\lambda}{1+\lambda}$ when $\nu \rightarrow \infty$ [8].

The distribution's moments can be represented recursively as

$$E(Y^{r+1}) = \begin{cases} \lambda [E(Y + 1)]^{1-\nu}, & r = 0 \\ \lambda \frac{\partial}{\partial \lambda} E(Y^r) + E(Y)E(Y^r), & r > 0; \end{cases}$$

in particular, its expected value and variance are

$$E(Y) = \frac{\partial \log Z(\lambda, \nu)}{\partial \log \lambda} \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu}, \quad \text{and} \quad (5)$$

$$\text{Var}(Y) = \frac{\partial E(Y)}{\partial \log \lambda} \approx \frac{1}{\nu} \lambda^{1/\nu}, \quad (6)$$

where the approximations provided in Equations (5) and (6) hold for $\nu \leq 1$ or $\lambda > 10^\nu$ [9,10].

Further, the CMP has the moment generating function (mgf) $M_Y(t) = E(e^{Yt}) = \frac{Z(\lambda e^t, \nu)}{Z(\lambda, \nu)}$ and pgf $E(t^Y) = \frac{Z(\lambda t, \nu)}{Z(\lambda, \nu)}$.

Sellers et al. [11] construct a bivariate CMP model by means of the compounding method, wherein the joint conditional distribution of $\{(X_1, X_2) \mid n\}$ has a bivariate binomial distribution and the number of trials n is $\text{CMP}(\lambda, \nu)$ distributed. The pmf of (X_1, X_2) is

$$P(X_1 = x_1, X_2 = x_2) = \frac{1}{Z(\lambda, \nu)} \sum_{n=0}^{\infty} \frac{\lambda^n}{(n!)^\nu} \times \sum_{a=n-x_1-x_2}^n \binom{n}{a, n-a-x_2, n-a-x_1, x_1+x_2+a-n} p_{00}^a p_{10}^{n-a-x_2} p_{01}^{n-a-x_1} p_{11}^{x_1+x_2+a-n},$$

where $\binom{n}{a, n-a-x_2, n-a-x_1, x_1+x_2+a-n}$ is the multinomial coefficient, and it has the joint pgf

$$\Pi(t_1^*, t_2^*) = \frac{Z(\lambda[1 + p_{1+}(t_1^* - 1) + p_{+1}(t_2^* - 1) + p_{11}(t_1^* - 1)(t_2^* - 1)], \nu)}{Z(\lambda, \nu)} \quad (7)$$

for some parameters, λ, ν , and probabilities $p_{00}, p_{10}, p_{01}, p_{11}$ such that $p_{00} + p_{10} + p_{01} + p_{11} = 1$, $p_{i+} = p_{i0} + p_{i1}$ for $i = 0, 1$, and $p_{+j} = p_{0j} + p_{1j}$ for $j = 0, 1$. This bivariate CMP distribution yields the three special bivariate cases that are achieved in their univariate analogs: for $\nu = 1$, the bivariate CMP distribution reduces to the bivariate Poisson [12,13]; when $\nu \rightarrow \infty$, we obtain the bivariate Bernoulli distribution [14]; and, for $\nu = 0$, $\lambda < 1$, and $\lambda\{p_{1+}(t_1^* - 1) + p_{+1}(t_2^* - 1) + p_{11}(t_1^* - 1)(t_2^* - 1)\} < 1$, the bivariate CMP distribution reduces to a bivariate geometric model [11].

3. Multivariate Conway–Maxwell–Poisson Distribution

Generalizing the compounding approach in [11], we develop a convenient form for the MCMP distribution. Consider d random variables $\mathbf{X} = (X_1, X_2, \dots, X_d)$ that, given some number of trials n , jointly have a conditional MB distribution with pgf

$$\Pi(t_1^*, \dots, t_d^* | n) = \left(\sum_{x_1=0}^1 \sum_{x_2=0}^1 \cdots \sum_{x_d=0}^1 p_{x_1 x_2 \cdots x_d} \prod_{i=1}^d (t_i^*)^{x_i} \right)^n$$

(Equation (37.71) of [1]), where n is a $\text{CMP}(\lambda, \nu)$ random variable. The compounding technique that is formulated as a CMP-stopped MB (i.e., where the MB index parameter is CMP distributed) can then be applied, resulting in the corresponding MCMP distribution's pgf as

$$\begin{aligned}\Pi(t_1^*, \dots, t_d^*) &= \sum_{n=0}^{\infty} \frac{\lambda^n}{(n!)^\nu Z(\lambda, \nu)} \Pi(t_1^*, \dots, t_d^* | n) \\ &= \frac{Z\left[\lambda \left(\sum_{x_1=0}^1 \cdots \sum_{x_d=0}^1 p_{x_1 x_2 \dots x_d} \prod_{i=1}^d (t_i^*)^{x_i}\right), \nu\right]}{Z(\lambda, \nu)},\end{aligned}\quad (8)$$

where $Z(\psi, \nu) = \sum_{s=0}^{\infty} \frac{\psi^s}{(s!)^\nu}$ for some $\psi > 0$. Equation (8) contains $2^d + 2$ parameters, but its degrees of freedom equals $2^d + 1$ due to the restriction, $\sum_{x_1=0}^1 \cdots \sum_{x_d=0}^1 p_{x_1 x_2 \dots x_d} = 1$; this adds difficulty in the determination of model parameter maximum likelihood estimates (MLEs). We circumvent this issue through the reparametrization, $\lambda_{x_1 x_2 \dots x_d} = \lambda p_{x_1 x_2 \dots x_d}$. Each variable is independent under this parameterization, where

$$\lambda = \lambda \left(\sum_{x_1=0}^1 \cdots \sum_{x_d=0}^1 p_{x_1 x_2 \dots x_d} \right) = \sum_{x_1=0}^1 \cdots \sum_{x_d=0}^1 \lambda p_{x_1 x_2 \dots x_d} = \sum_{x_1=0}^1 \cdots \sum_{x_d=0}^1 \lambda_{x_1 x_2 \dots x_d}$$

and $p_{x_1 x_2 \dots x_d} = \frac{\lambda_{x_1 x_2 \dots x_d}}{\lambda}$. For simplicity, we use λ to denote $\sum_{x_1=0}^1 \cdots \sum_{x_d=0}^1 \lambda_{x_1 x_2 \dots x_d}$ but recognize that λ is no longer an independent parameter in the ensuing discussion. The pgf of the MCMP distribution can now be parameterized as

$$\Pi(t_1^*, \dots, t_d^*) = \frac{Z\left[\left(\sum_{x_1=0}^1 \cdots \sum_{x_d=0}^1 \lambda_{x_1 x_2 \dots x_d} \prod_{i=1}^d (t_i^*)^{x_i}\right), \nu\right]}{Z(\lambda, \nu)}.\quad (9)$$

As is the case of the univariate and bivariate CMP, this MCMP includes the MP, multivariate geometric, and multivariate Bernoulli distributions all as special cases, where ν maintains representation as the dispersion parameter. When $\nu = 1$, this MCMP pgf reduces to the form of the MP joint pgf (see Equation (3)). When $\nu = 0$ and $\lambda < 1$, its pgf becomes

$$\Pi(t_1^*, \dots, t_d^*; \nu = 0, \lambda < 1) = \left(a_0 + \sum_{i=1}^d a_i t_i^* + \sum_{1 \leq i < j \leq d} a_{ij} t_i^* t_j^* + \dots + a_{12 \dots d} t_1^* t_2^* \dots t_d^* \right)^{-1},$$

where $a_0 = \frac{1 - \lambda_{00 \dots 0}}{1 - \lambda}$; $a_i = \frac{-\lambda_{0 \dots 01_i 0 \dots 0}}{1 - \lambda}$ where 1_i denotes a 1 in the i th position, $i = 1, 2, \dots, d$; $a_{ij} = \frac{-\lambda_{0 \dots 01_i 0 \dots 01_j 0 \dots 0}}{1 - \lambda}$ where $1_i, 1_j$ denote 1s in the $1 \leq i \neq j \leq d$ locations; $\dots$; $a_{12 \dots d} = \frac{-\lambda_{11 \dots 1}}{1 - \lambda}$. This is the pgf of a multivariate geometric distribution (i.e., the MNB distribution pgf in Equation (4) with $k = 1$). Finally, when $\nu \rightarrow \infty$, this MCMP becomes a multivariate Bernoulli (i.e., the MB in [2] with $n = 1$) with $p_{00 \dots 0}^* = \frac{1 + \lambda_{00 \dots 0}}{1 + \lambda}$ and all remaining probabilities are $p_{x_1 x_2 \dots x_d}^* = \frac{\lambda_{x_1 x_2 \dots x_d}}{1 + \lambda}$ where at least one of x_i equals 1, $i = 1, 2, \dots, d$. More broadly, $\nu = 1$ denotes the equi-dispersion case while $\nu < (>)1$ reflects data over-dispersion (under-dispersion), both for the joint distribution and the respective marginal distributions.

Given the joint pgf in Equation (8), this MCMP model has the joint mgf

$$\begin{aligned}M(t_1, t_2, \dots, t_d) &= \Pi(e^{t_1}, e^{t_2}, \dots, e^{t_d}) \\ &= \frac{Z\left[\left(\sum_{x_1=0}^1 \cdots \sum_{x_d=0}^1 \lambda_{x_1 x_2 \dots x_d} e^{\sum_{i=1}^d t_i x_i}\right), \nu\right]}{Z(\lambda, \nu)},\end{aligned}\quad (10)$$

and the joint fmgf as

$$\begin{aligned}G(t_1, t_2, \dots, t_d) &= \Pi(t_1 + 1, t_2 + 1, \dots, t_d + 1) \\ &= \frac{Z\left[\left(\sum_{x_1=0}^1 \cdots \sum_{x_d=0}^1 \lambda_{x_1 x_2 \dots x_d} \prod_{i=1}^d (t_i + 1)^{x_i}\right), \nu\right]}{Z(\lambda, \nu)}.\end{aligned}\quad (11)$$

We derive the MCMP pmf by taking partial derivatives of the pgf, i.e.,

$$p(x_1, x_2, \dots, x_d) = \frac{1}{x_1!x_2! \cdots x_d!} \frac{\partial^{x_1+x_2+\dots+x_d}}{\partial t_1^{x_1} \partial t_2^{x_2} \cdots \partial t_d^{x_d}} \Pi(t_1, t_2, \dots, t_d) \Big|_{t_1=t_2=\dots=t_d=0}; \quad (12)$$

see Appendix A for pertinent details. Moving forward, we shall illustrate the MCMP results using the trivariate case as motivation, where the joint fmgf reduces to $G(t_1, t_2, t_3)$ from which we can obtain the moments and product moments, respectively; see Appendix B for all relevant details. These results confirm that the dispersion parameter ν denotes the type of data-dispersion for the joint and marginal distributions, and the correlations between any two random variables is non-negative with $0 \leq \rho_{X_i X_j} \leq 1$ for any variables $i, j = 1, 2, 3$.

3.1. Parameter Estimation

We perform parameter estimation by the method of maximum likelihood (ML). Considering the trivariate case of Equations (9) and (12), there are nine parameters required to specify a trivariate CMP distribution, namely, $\lambda_{x_1 x_2 x_3}$ for $x_i = 0, 1; i = 1, 2, 3$; and ν . Accordingly, the log-likelihood has the form

$$\ln L(\lambda_{000}, \dots, \lambda_{111}, \nu; (x_1, x_2, x_3)) = \ln \prod_{j=1}^n p(x_{1j}, x_{2j}, x_{3j}) = \sum_{j=1}^n \ln p(x_{1j}, x_{2j}, x_{3j}), \quad (13)$$

where x_{ij} denotes the j th observation in the i th data dimension, x_i denotes the vector of the entire data set in the i th dimension; the precise form of $p(x_{1j}, x_{2j}, x_{3j})$ is provided in Equation (A4) in Appendix A. The resulting score equations, however, do not have a closed form solution. For this reason, we carry out the statistical computations by using optimizing routines in R [15].

To perform the parameter estimation, we use the `optim` function where the negated form of the log-likelihood (Equation (13)) serves as the function to be optimized, and the L-BFGS-B method and its default convergence criteria are applied. Additionally, we approximate the standard errors of the estimated parameters by calculating the square root of the diagonal of the inverse Hessian matrix based on the approximate form obtained from `optim`. The complexity of the MCMP distribution, however, brings with it some computational difficulties when applying `optim`. The resulting MLE can vary considerably depending on the choice of starting values. To avert this, we consider several starting points including an exhaustive search in order to potentially improve the estimation result. Meanwhile, the resulting Hessian matrix provided from `optim` sometimes produces an inverse matrix containing negative diagonal elements; this violates the presumed positive semidefinite form of the Fisher information matrix. For these reasons, we recommend utilizing a parametric bootstrap method as an alternative approach for quantifying variability in the parameter estimates.

3.2. Hypothesis Testing

To check if a multivariate count data set suffers from any statistically significant data dispersion such that the MP distribution is unsuitable (favoring the MCMP distribution), we conduct the hypothesis test, $H_0: \nu = 1$ versus $H_1: \nu \neq 1$. We do not concern ourselves with the direction of the data dispersion because the MCMP distribution can accommodate both over- and under-dispersion. Nonetheless, the resulting statistical inference, along with the estimate for ν , offers guidance regarding the type of dispersion present in the data. We use the likelihood ratio test (LRT) statistic, $\Lambda_{\nu=1} = \frac{\sup_{\nu=1} \ln L}{\sup \ln L}$, where $\sup_{\nu=1} \ln L$ and $\sup \ln L$, respectively, denote the maximum log-likelihoods associated with the MP and MCMP models. Theoretically, $-2 \log(\Lambda_{\nu=1})$ follows a χ_1^2 distribution and thus can be used to assess whether the data are reasonably distributed as a MP distribution, or if statistically significant dispersion exists such that it warrants using the MCMP model. In a similar vein, one can consider hypothesis tests, $H_0: \nu = 0$ or $H_0: \nu \rightarrow \infty$ (versus $H_1:$

otherwise) to determine whether the multivariate data satisfy a multivariate geometric or multivariate Bernoulli distribution, respectively; their associated LRTs have adjusted distributional forms based on a mixture involving χ_1^2 to account for being at the respective boundaries for ν [16].

4. Examples

This section considers various simulated and real data examples to illustrate the flexibility of the MCMP model. For the real data sets, we compare model performance via the respective log-likelihood and Akaike Information Criterion (AIC) values. We particularly consider $\Delta_i = \text{AIC}_i - \text{AIC}_{\min}$ as introduced in Burnham and Anderson [17], where AIC_i denotes the AIC associated with Model i , and $\text{AIC}_{\min}$ is the minimum AIC among the considered models. [17] provides model support levels based on recommended Δ_i ranges; see Table 1 for details.

Table 1. Model support levels based on AIC difference values, $\Delta_i = \text{AIC}_i - \text{AIC}_{\min}$, for Model i [17].

Δ_i	Empirical Support Level for Model i
$[0, 2]$	Substantial
$[4, 7]$	Considerably less
$(10, \infty)$	Essentially none

4.1. Simulated Data

Here, we provide simulated data examples to illustrate the MCMP model's ability to correctly distinguish the MP distribution. Without loss of generality, we proceed with the use of the trivariate case. To evaluate the robustness of the simulation process, we consider data simulations of size {100, 250, 500, 1000} and simulate data 500 times at each size level.

We first consider a simulated trivariate Poisson distribution where the joint pgf (Equation (3)) is defined with $A_1 = 1.5$, $A_2 = 2$, $A_3 = 1.3$, $A_{12} = 1.2$, $A_{13} = 0.6$, $A_{23} = 0.7$ and $A_{123} = 0.3$, and obtain the MLEs for the trivariate CMP under two conditions: the unconstrained case, and the restricted case where $\nu = 1$. The latter case serves to reflect the trivariate Poisson model with $\lambda_{100} = A_1$, $\lambda_{010} = A_2$, $\lambda_{001} = A_3$, $\lambda_{110} = A_{12}$, $\lambda_{101} = A_{13}$, $\lambda_{011} = A_{23}$, $\lambda_{111} = A_{123}$; thus, the value of λ_{000} no longer affects the model. Table 2 displays the proportion of $-2 \log \Lambda$ statistics that fall within the respective 95% or 99% confidence bounds across the simulations. As expected, the proportion of $-2 \log \Lambda$ values that are within the respective bounds, $\chi_1^2(0.95) = 3.841$ and $\chi_1^2(0.99) = 6.635$, is quite close to their respective nominal levels, regardless of size level.

To assess the power of the test, $H_0: \nu = 1$ versus $H_1: \nu \neq 1$, we further generate data from the trivariate geometric ($a_1 = -0.3$, $a_2 = -0.5$, $a_3 = -0.7$, $a_{12} = -0.2$, $a_{13} = -1$, $a_{23} = -1.3$, $a_{123} = -0.1$), the trivariate Bernoulli ($p_{000} = 0.1$, $p_{001} = 0.07$, $p_{010} = 0.23$, $p_{100} = 0.15$, $p_{011} = 0.08$, $p_{101} = 0.11$, $p_{110} = 0.17$, $p_{111} = 0.09$), and the trivariate CMP expressing over-dispersion ($\lambda_{000} = 0.07$, $\lambda_{100} = 0.13$, $\lambda_{010} = 0.22$, $\lambda_{001} = 0.15$, $\lambda_{110} = 0.08$, $\lambda_{101} = 0.06$, $\lambda_{011} = 0.12$, $\lambda_{111} = 0.11$, $\nu = 0.5$) and under-dispersion ($\lambda_{000} = 2$, $\lambda_{100} = 0.8$, $\lambda_{010} = 1.4$, $\lambda_{001} = 1.7$, $\lambda_{110} = 2.2$, $\lambda_{101} = 1.3$, $\lambda_{011} = 0.6$, $\lambda_{111} = 0.9$, $\nu = 2$), respectively. All simulation results obtained are presented in Table 3. As the generating distribution has a measure of dispersion that moves away from 1 (i.e., the data deviate from the Poisson), we see the power increase in both directions. Meanwhile, the power likewise increases with the sample size in association with all of the respective distributions.

Table 2. Proportion of $-2 \log \Lambda$ values that lie within the $\chi_1^2(0.95)$ and $\chi_1^2(0.99)$ bounds, respectively, given various sample sizes, {100, 250, 500, 1000}.

Sample Size	Within $\chi_1^2(0.95)$	Within $\chi_1^2(0.99)$
100	93.0%	99.2%
250	93.4%	99.4%
500	94.2%	98.8%
1000	94.8%	98.6%

Table 3. Power of the likelihood ratio test (at 5% level) when data are generated from the trivariate geometric, the trivariate CMP with $\nu = 0.5$, the trivariate CMP with $\nu = 2$ and the trivariate Bernoulli, respectively, for various sample sizes, {100, 250, 500, 1000}.

Sample Size	Geometric	CMP ($\nu = 0.5$)	CMP ($\nu = 2$)	Bernoulli
100	100%	37.0%	82.8%	100%
250	100%	62.6%	99.4%	100%
500	100%	75.4%	100.0%	100%
1000	100%	99.2%	100.0%	100%

4.2. Real Data: Corporación Favorita Grocery Sales

The Corporación Favorita grocery sales data [18] include information regarding the number of unit items sold daily to more than 4000 items in 35 different stores over a five-year period. To illustrate the MCMP distribution's flexibility for describing real count data, we consider the unit sales regarding a particular item (Item ID:103665) over 100 days in each of three stores (Stores 1, 2 and 3, respectively); the data are provided in Table A2 in Appendix D. This dataset is over-dispersed due to the weekly and monthly periodic fluctuation; the number of sales often tend to be high at the beginning of each month as well as on weekends. Table 4 summarizes the results that stem from considering various trivariate models to describe the data, namely, the trivariate Poisson, trivariate NB [6], trivariate geometric, and trivariate CMP. For each of the assumed models, this table provides the respective MLEs, resulting log-likelihood, and AIC values.

Although the trivariate Poisson distribution has the least number of parameters (i.e., 7) among the models considered, it has the largest AIC (1748.5), suggesting its unsuitability for these data [17]. Meanwhile, the MLEs of the trivariate CMP model include $\hat{\nu} \approx 0.25$, implying that the data are over-dispersed. The trivariate CMP produces a considerably smaller AIC (1627.9) relative to the trivariate Poisson; this further demonstrates the apparent data over-dispersion that should be addressed, but with $\Delta = 6.2$ relative to the AIC from the trivariate NB, the trivariate CMP (while second best among the four considered models) still has model support that is “considerably less” than that of the trivariate NB (AIC = 1621.7); this result is still substantially better than the difference between the trivariate NB and Poisson models ($\Delta = 126.8$), clearly inferring no support for the trivariate Poisson. Further, applying the trivariate CMP model introduces consideration of the trivariate geometric and NB models, respectively, as possible parsimonious models. The respective LRT statistics, $-2 \log \Lambda_{\nu=1} = 124.6$ for the test $H_0 : \nu = 1$ and $-2 \log \Lambda_{\nu=0} = 79.4$ for $H_0 : \nu = 0$, both have p-values smaller than 0.005 which indicate that neither the trivariate Poisson nor the trivariate geometric fits the data well. Even still, $\hat{\nu} \approx 0.25$ serves as an indication of data over-dispersion, hence consideration of the general MNB distribution as a possible model.

Table 4. Estimation results associated with the Corporación Favorita grocery sales data based on various assumed trivariate models: Conway–Maxwell–Poisson (CMP), trivariate Poisson, trivariate geometric and trivariate negative binomial (NB). Respective log-likelihood and Akaike Information Criterion (AIC) values are also provided, along with the number of free parameters for AIC determination.

Model	Estimated Parameters			Log Likelihood	No. of Free Parameters	AIC
CMP	$\hat{\lambda}_{000} = 0.00$	$\hat{\lambda}_{111} = 0.00$	$\hat{\lambda}_{100} = 0.32$	−804.9	9	1627.9
	$\hat{\lambda}_{010} = 0.47$	$\hat{\lambda}_{001} = 1.17$	$\hat{\lambda}_{110} = 0.00$			
	$\hat{\lambda}_{101} = 0.00$	$\hat{\lambda}_{011} = 0.00$	$\hat{\nu} = 0.25$			
Poisson	$\hat{A}_1 = 2.32$	$\hat{A}_2 = 2.97$	$\hat{A}_3 = 8.94$	−867.3	7	1748.5
	$\hat{A}_{12} = 0.19$	$\hat{A}_{13} = 0.00$	$\hat{A}_{23} = 0.62$			
		$\hat{A}_{123} = 0.11$				
geometric	$\hat{a}_1 = -2.62$	$\hat{a}_2 = -3.89$	$\hat{a}_3 = -9.67$	−844.6	7	1703.2
	$\hat{a}_{12} = 0.00$	$\hat{a}_{13} = 0.00$	$\hat{a}_{23} = 0.00$			
	$\hat{a}_{123} = 0.00$	$\hat{a}_0 = 17.18$				
NB	$\hat{a}_1 = -0.44$	$\hat{a}_2 = -0.65$	$\hat{a}_3 = -1.61$	−802.8	8	1621.7
	$\hat{a}_{12} = 0.00$	$\hat{a}_{13} = 0.00$	$\hat{a}_{23} = 0.00$			
	$\hat{a}_{123} = 0.00$	$\hat{a}_0 = 3.69$	$\hat{k} = 6.01$			

Table 4 further shows that $\hat{\lambda}_{110}$, $\hat{\lambda}_{101}$, $\hat{\lambda}_{011}$ and $\hat{\lambda}_{111}$ for the CMP model are all 0; a similar situation appears on the estimation of the geometric and NB models, where $\hat{a}_{12}$, $\hat{a}_{13}$, $\hat{a}_{23}$ and $\hat{a}_{123}$ are also all 0. This indicates that there is no significant correlation within the data; this is true because the correlation coefficients between Stores 1 and 2, Stores 1 and 3, and Stores 2 and 3 are 0.15, 0.02, 0.21, respectively.

Figure 1 compares the marginal pmfs associated with each of the four models with the marginal relative frequencies associated with the number of unit sales for each of the three stores (Stores 1, 2, 3). These images show that the trivariate CMP and NB models produce very similar estimated marginal distributions with modes that are close to the observed mode, and have sufficiently wide tails to reflect the observed marginal frequencies, particularly for Store 2. Goodness-of-fit tests are likewise performed for comparing the aforementioned models to assess how well their marginal pmfs fit the marginal data frequencies. Following [19], we modify our observed frequencies by grouping observations greater than 8 on Store 1, greater than 9 on Store 2, and observations greater than 21 on Store 3. This allows the respective tail bins associated with each store to have a sufficiently large observed frequency to allow for the goodness-of-fit test to be conducted and the associated asymptotic chi-square distribution to be used. As a result, resulting statistics for the goodness-of-fit tests are expected to follow the chi-square distribution with 10, 11 and 23 degrees of freedom, respectively, for Stores 1, 2 and 3.

Table 5 summarizes the goodness-of-fit test statistics for each of the stores and models. While the trivariate geometric model best fits the Store 1 marginal distribution, the goodness-of-fit scores for the trivariate CMP and NB models are considerably better and outperform their peers for Stores 2 and 3. Table 5 confirms these assertions with $\chi^2_{10}(0.95) = 18.3$, $\chi^2_{11}(0.95) = 19.7$ and $\chi^2_{23}(0.95) = 35.2$, respectively, for Stores 1, 2 and 3; we again see that the geometric model fits the data better for Store 1, and the trivariate CMP and NB models produce closer fits for Stores 2 and 3.

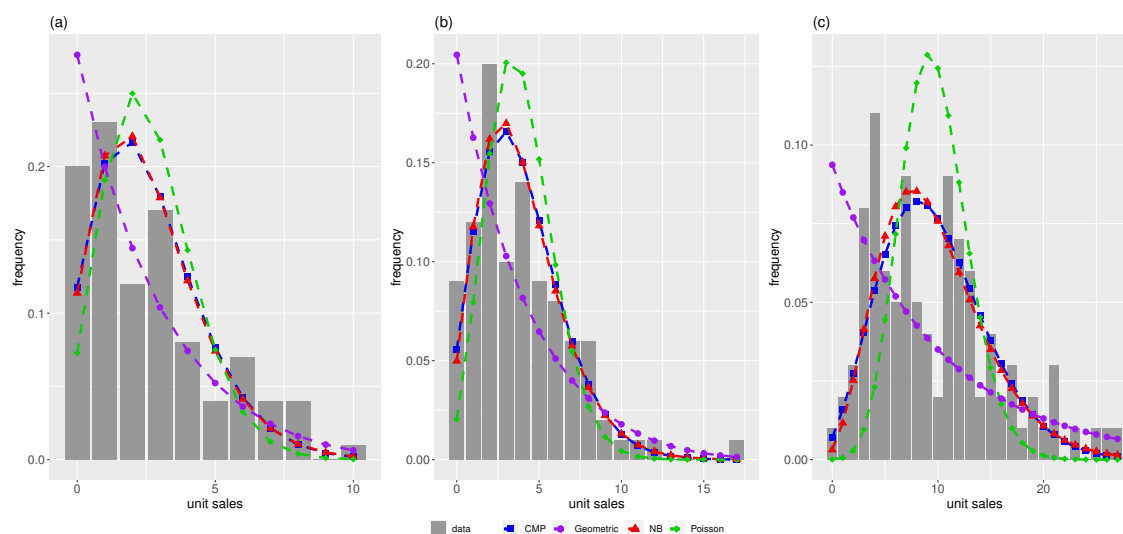


Figure 1. Estimated marginal distributions associated with the trivariate CMP (blue/square), the trivariate geometric (purple/circle), the trivariate negative binomial (red/triangle), and the trivariate Poisson (green/diamond) compared with the original data relative frequencies (histogram) regarding the number of unit sales for: (a) Store 1, (b) Store 2, (c) Store 3.

Table 5. The goodness-of-fit test measures to compare the considered trivariate model (i.e., trivariate CMP, trivariate negative binomial (NB), trivariate Poisson, and trivariate geometric) marginal pmfs to the marginal data regarding the number of unit sales for Stores 1, 2 and 3, respectively.

	Store 1	Store 2	Store 3
CMP	26.0	8.9	30.6
NB	26.4	9.7	30.2
Poisson	82.7	57.7	857.0
Geometric	16.1	22.7	51.6

4.3. Real Data: NBA All-Star

To demonstrate that the trivariate CMP can also be suitable for under-dispersed data, we consider data from the National Basketball Association (NBA) All-Star game rosters from 2000 to 2016 and seek to model the distribution of the number of players selected for the All-Star game each year in various positions [20]. For simplicity, we focus on the number of players that can play as Center (C), Forward (F), or Forward-center (FC); the data are provided in Table A3 of Appendix D. We again consider the trivariate CMP, the trivariate Poisson, and the trivariate NB distributions as possible models to describe this dataset. Table 6 contains a summary of the results including the respective MLEs, the resulting maximized log-likelihood, the number of free parameters, and the associated AIC for each of the three considered models.

The trivariate CMP model performs the best among the considered models, attaining a maximum likelihood equaling -68.2 and $AIC_{min} = 154.4$. The trivariate Poisson and NB models meanwhile produce respective AICs equaling 180.6 and 182.7 such that both respective difference values as defined in [17] (Table 1) are greater than 26, indicating no empirical support in favor of either model. The difference between the respective AIC values for the trivariate Poisson and NB models stems from the difference in the number of free parameters while they attain the same maximized log-likelihood value (-83.3). Neither of these models can accommodate data under-dispersion, and consequently the optimal trivariate NB distribution is that model which converges to the trivariate Poisson as $k \rightarrow \infty$. Accordingly, the trivariate NB MLEs that best address data under-dispersion are those under the constraint of data equi-dispersion.

The trivariate CMP model successfully detects the data under-dispersion ($\hat{\nu} = 38.4 \gg 1$). In fact, such a large $\hat{\nu}$ suggests that we should consider modeling the data via a trivariate Bernoulli model. This would normally be true because the resulting CMP denominator includes $(m!)^{\nu}$ which becomes considerably large for $m > 1$ given large $\hat{\nu}$. This makes $p_{(x,y,z)}$ vanish for any (x, y, z) such that at least one of the random variables exceeds 1. This is not the case here, however, because this data example likewise produces extremely large $\hat{\lambda}$ estimates. Reviewing the raw data likewise suggests clearly that the trivariate Bernoulli distribution is not appropriate because there exist count data that are larger than 1, thus violating the multivariate Bernoulli structure. Therefore, the use of the trivariate CMP to analyze these data is duly justified.

Table 6. Estimation results associated with the NBA All-Star data based on various assumed trivariate models: Conway–Maxwell–Poisson (CMP), Poisson, and negative binomial (NB). Respective log-likelihood and Akaike Information Criterion (AIC) values are also provided.

Model	Estimated Parameters			Log Likelihood	No. of Free Params	AIC
CMP	$\hat{\lambda}_{000} = 0.00 \times 10^0$	$\hat{\lambda}_{111} = 0.36 \times 10^{28}$	$\hat{\lambda}_{100} = 1.53 \times 10^{28}$	−68.2	9	154.4
	$\hat{\lambda}_{010} = 0.00 \times 10^0$	$\hat{\lambda}_{001} = 1.38 \times 10^{28}$	$\hat{\lambda}_{110} = 2.17 \times 10^{28}$			
	$\hat{\lambda}_{101} = 3.77 \times 10^{28}$	$\hat{\lambda}_{011} = 4.52 \times 10^{28}$	$\hat{\nu} = 3.84 \times 10^1$			
Poisson	$\hat{A}_1 = 1.15$	$\hat{A}_2 = 0.94$	$\hat{A}_3 = 1.94$	−83.3	7	180.6
	$\hat{A}_{12} = 0.00$	$\hat{A}_{13} = 0.12$	$\hat{A}_{23} = 0.04$			
		$\hat{A}_{123} = 1.66$				
NB	$\hat{a}_1 = -3.3 \times 10^{-5}$	$\hat{a}_2 = -2.5 \times 10^{-5}$	$\hat{a}_3 = -5.5 \times 10^{-5}$	−83.3	8	182.7
	$\hat{a}_{12} = -2.6 \times 10^{-6}$	$\hat{a}_{13} = -4.3 \times 10^{-6}$	$\hat{a}_{23} = -4.1 \times 10^{-6}$			
	$\hat{a}_{123} = -4.5 \times 10^{-5}$	$\hat{a}_0 = 1.00017$	$\hat{k} = 34856$			

5. Discussion

In this paper, we present a MCMP model that is developed via the compounding method. The distribution is established as a CMP-stopped multivariate binomial distribution, i.e., a multivariate binomial distribution where the associated index parameter is CMP distributed. Along with an introduction to this resulting distribution, we discussed its statistical properties which aid in better model interpretation. The CMP model can flexibly accommodate both over- and under-dispersed count data, and it includes the Poisson, Bernoulli, and geometric distributions all as special cases. Accordingly, the MCMP model serves as a reliable tool for model determination because it can successfully recognize these three multivariate special cases, and serve as an overarching distributional structure connecting them. One can determine if significant data dispersion exists by calculating the LRT statistic Λ discussed in Section 3.2, and analogous tests can be considered to determine whether the data effectively approximate either of the other two special case distributions (i.e., $H_0 : \nu = 0$ or $H_0 : \nu \rightarrow \infty$, respectively). The MCMP distribution is particularly useful for modeling under-dispersed count data, as demonstrated through the simulated and real data examples.

A limitation of the MCMP model is that the correlation between any two of the d random variables comprising the MCMP is constrained to be non-negative, and so it may not be appropriate to consider this model to analyze multivariate count data containing negative correlations. This is true, however, of several multivariate discrete distributions, e.g., the [3] multivariate Poisson distribution. Meanwhile, this MCMP construct involves only one parameter (ν) to describe data dispersion. Hence, this MCMP model is suitable only for data with similar levels of data dispersion in each dimension, however the model can be broadened to allow for dynamic dispersion. Future work will seek to define a broader generalization of the MP distribution (or modification of this MCMP model) that allows for a broader range of correlation and possesses greater flexibility with regard to data dispersion. One proposed approach, for example, is to consider using copulas to develop a multivariate CMP distribution, as described in [21]. Though this is a standard

method for multivariate continuous variables, its use for modeling multivariate count data has its own limitations, most importantly that copulas for discrete outcomes are not identifiable, especially when those discrete outcomes follow count distributions [21–23].

Table 4 demonstrates another limitation of the MCMP model, namely that it cannot accommodate as much data over-dispersion as the MNB; the MCMP distribution at best contains the multivariate geometric distribution (which is a special case of MNB). The MNB model, however, can be viewed as the convolution of independent and identically distributed (iid) multivariate geometric distributions. This convolution structure will then be able to capture greater over-dispersion. More broadly, the same idea can be used to consider a multivariate version of the sum of CMPs (MSCMP) model [24] as a generalization to accommodate broader dispersion, and use its trivariate form to revisit the Corporación Favorita grocery sales dataset. Unfortunately, due to computational issues, we were only able to perform parameter estimations for the trivariate SCMP model under the restriction $m = 1, 2, 3$. Future work will further study the MSCMP model, for example, to determine how to optimally and directly compute the MSCMP pmf, and more efficiently determine the MLEs of model parameters. See Appendix C for more information about the MSCMP.

Author Contributions: Conceptualization (K.F.S.); Methodology (K.F.S., T.L., Y.W., N.B.); Formal analysis (K.F.S., T.L., Y.W., N.B.); Investigation (K.F.S., T.L., Y.W., N.B.); Software (K.F.S., T.L., Y.W.); Supervision (K.F.S.); Writing—original draft preparation (K.F.S., T.L., Y.W., N.B.); Writing—review and editing (K.F.S., T.L., Y.W., N.B.). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the appendix.

Acknowledgments: The authors thank Richard Sellers for insightful discussions that aided in better comprehension and understanding of the NBA data set. This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MB	multivariate binomial
fmgf	factorial moment generating function
MP	multivariate Poisson
pgf	probability generating function
NB	negative binomial
MNB	multivariate negative binomial
CMP	Conway–Maxwell–Poisson
MCMP	multivariate Conway–Maxwell–Poisson
mgf	moment generating function
MLEs	maximum likelihood estimates
pmf	probability mass function
ML	maximum likelihood
LRT	likelihood ratio test
AIC	Akaike Information Criterion
MLE	maximum likelihood estimate
NBA	National Basketball Association

C	Center
F	Forward
FC	Forward-center
sCMP	sum of CMPs
MSCMP	multivariate version of the sum of CMPs

Appendix A. Deriving the Probability Mass Function

In order to derive the general form of the pmf, we first introduce some notation. Assuming a MCMP distribution with d dimensions, there exist 2^d distinct probabilities $p_{x_1 x_2 \dots x_d}$ in the pgf, where $x_i = \{0, 1\}$ for all $i = 1, 2, \dots, d$. The first derivation relies on the identity

$$\left(\sum_{i=1}^d a_i \right)^n = \sum_{s^*} \binom{n}{l_1, l_2, \dots, l_d} a_1^{l_1} a_2^{l_2} \dots a_d^{l_d}, \quad (\text{A1})$$

where $s^* = \{(l_1, \dots, l_d) : 0 \leq l_i \leq n \text{ for } i = 1, \dots, d, \text{ and } \sum_{i=1}^d l_i = n\}$ and $\binom{n}{l_1, l_2, \dots, l_d} = \frac{n!}{l_1! l_2! \dots l_d!}$ is the multinomial coefficient. We can express

$$\begin{aligned} G_n(t_1, t_2, \dots, t_d) &= \left\{ \sum_{x_1=0}^1 \dots \sum_{x_d=0}^1 p_{x_1 \dots x_d} \prod_{i=1}^d t_i^{x_i} \right\}^n \\ &= \left\{ p_{00 \dots 0} + \sum_{i_1=1}^d p_{0 \dots 010 \dots 0} t_{i_1} \right. \\ &\quad \left. + \sum_{1 \leq i_1 < i_2 \leq d} p_{0 \dots 010 \dots 01 \dots 0} t_{i_1} t_{i_2} + \dots + p_{11 \dots 1} t_1 t_2 \dots t_d \right\}^n \\ &\doteq \{s_0 + s_1 + s_2 + \dots + s_d\}^n \\ &= \sum_{s^*} \binom{n}{l_0, l_1, \dots, l_d} s_0^{l_0} s_1^{l_1} \dots s_d^{l_d}, \end{aligned} \quad (\text{A2})$$

where $s_0 = p_{00 \dots 0}$, $s_1 = \sum_{i_1=1}^d p_{0 \dots 010 \dots 0} t_{i_1}$, $s_2 = \sum_{1 \leq i_1 < i_2 \leq d} p_{0 \dots 010 \dots 010 \dots 0} t_{i_1} t_{i_2}$, $\dots$, $s_d = p_{11 \dots 1} t_1 \dots t_d$, and l_i denotes the number of times out of n trials where we obtain exactly i of the required elements, $i = 1, \dots, d$. Observe that the number of respective s_i elements is $d_0 = \#s_0 = 1$, $d_1 = \#s_1 = \binom{d}{1}$, $d_2 = \#s_2 = \binom{d}{2}$, $d_3 = \#s_3 = \binom{d}{3}$, $\dots$, $d_d = \#s_d = \binom{d}{d} = 1$. Then,

$$\begin{aligned} s_0^{l_0} &= p_{00 \dots 0}^{l_0} \\ s_1^{l_1} &= \left(\sum_{i_1=1}^d p_{0 \dots 010 \dots 0} t_{i_1} \right)^{l_1} = \left(\sum_{i_1=1}^{d_1} p_{(i_1)} t_{i_1} \right)^{l_1} \\ &= \sum_{s_1^*} \binom{l_1}{j_{11}, j_{12}, \dots, j_{1d_1}} p_{(1)}^{j_{11}} p_{(2)}^{j_{12}} \dots p_{(d_1)}^{j_{1d_1}} t_1^{j_{11}} t_2^{j_{12}} \dots t_{d_1}^{j_{1d_1}}, \end{aligned}$$

where we use the simplified notation, $p_{(i_1)} \doteq p_{0 \dots 010 \dots 0}$ with 1 being in the i_1 th position, and $s_1^* = \{(j_{11}, j_{12}, \dots, j_{1d_1}) : 0 \leq j_{1i_1} \leq l_1 \text{ for } i_1 = 1, \dots, d_1, \text{ and } \sum_{i_1=1}^{d_1} j_{1i_1} = l_1\}$. Similarly,

$$\begin{aligned} s_2^{l_2} &= \left(\sum_{1 \leq i_2 < i_2 \leq d} p_{0 \dots 10 \dots 010 \dots 0} t_{i_1} t_{i_2} \right)^{l_2} = \left(\sum_{1 \leq i_2 < i_2 \leq d_2} p_{(i_1, i_2)} t_{i_1} t_{i_2} \right)^{l_2} \\ &= \sum_{s_2^*} \binom{l_2}{j_{21}, j_{22}, \dots, j_{2d_2}} p_{(1,2)}^{j_{21}} p_{(1,3)}^{j_{22}} \dots p_{(d_2, d_2)}^{j_{2d_2}} (t_1 t_2)^{j_{21}} (t_1 t_3)^{j_{22}} \dots (t_{d-1} t_d)^{j_{2d_2}}, \end{aligned}$$

where $s_2^* = \{(j_{21}, j_{22}, \dots, j_{2,d_2}) : 0 \leq j_{2i_2} \leq l_2 \text{ for } i_2 = 1, \dots, d_2, \text{ and } \sum_{i_2=1}^{d_2} j_{2,i_2} = l_2\}$, etc.; finally,

$$s_d^{l_d} = (p_{11\dots 1} t_1 t_2 \dots t_d)^{l_d} = p_{1\dots 1}^{l_d} t_1^{l_d} t_2^{l_d} \dots t_d^{l_d}. \quad (\text{A3})$$

To illustrate this, consider the case when $d = 3$. In this case,

$$\begin{aligned} s_0^{l_0} &= p_{000}^{l_0} \\ s_1^{l_1} &= \sum_{s_1^*} \binom{l_1}{j_{11}, j_{12}, j_{13}} p_{(1)}^{j_{11}} p_{(2)}^{j_{12}} p_{(3)}^{j_{13}} t_1^{j_{11}} t_2^{j_{12}} t_3^{j_{13}} = \sum_{s_1^*} \binom{l_1}{j_{11}, j_{12}, j_{13}} p_{100}^{j_{11}} p_{010}^{j_{12}} p_{001}^{j_{13}} t_1^{j_{11}} t_2^{j_{12}} t_3^{j_{13}} \\ s_2^{l_2} &= \sum_{s_2^*} \binom{l_2}{j_{21}, j_{22}, j_{23}} p_{(12)}^{j_{21}} p_{(13)}^{j_{22}} p_{(23)}^{j_{23}} (t_1 t_2)^{j_{21}} (t_1 t_3)^{j_{22}} (t_2 t_3)^{j_{23}} \\ &= \sum_{s_2^*} \binom{l_2}{j_{21}, j_{22}, j_{23}} p_{110}^{j_{21}} p_{101}^{j_{22}} p_{011}^{j_{23}} t_1^{j_{21}+j_{22}} t_2^{j_{22}+j_{23}} t_3^{j_{22}+j_{23}} \\ s_3^{l_3} &= (p_{111} t_1 t_2 t_3)^{l_3} = p_{111}^{l_3} t_1^{l_3} t_2^{l_3} t_3^{l_3}, \end{aligned}$$

where

$$\begin{aligned} s_1^* &= \left\{ (j_{11}, j_{12}, j_{13}) : 0 \leq j_{1i_1} \leq l_1 \text{ for } i_1 = 1, 2, 3, \text{ and } \sum_{i_1=1}^3 j_{1i_1} = l_1 \right\}, \\ s_2^* &= \left\{ (j_{21}, j_{22}, j_{23}) : 0 \leq j_{2i_2} \leq l_2 \text{ for } i_2 = 1, 2, 3, \text{ and } \sum_{i_2=1}^3 j_{2i_2} = l_2 \right\} \end{aligned}$$

and $\sum_{i=0}^3 l_i = n$. We then find that

$$\begin{aligned} G_n(t_1, t_2, t_3) &= \sum_{s^*} \binom{n}{l_0, l_1, l_2, l_3} s_0^{l_0} s_1^{l_1} s_2^{l_2} s_3^{l_3} \\ &= \sum_{s^*} \binom{n}{l_0, l_1, l_2, l_3} \sum_{s_1^*} \binom{l_1}{j_{11}, j_{12}, j_{13}} \left[\sum_{s_2^*} \binom{l_2}{j_{21}, j_{22}, j_{23}} p_{000}^{l_0} p_{100}^{j_{11}} p_{010}^{j_{12}} p_{001}^{j_{13}} \right. \\ &\quad \left. p_{110}^{j_{21}} p_{101}^{j_{22}} p_{011}^{j_{23}} p_{111}^{l_3} t_1^{j_{11}+j_{21}+j_{22}+l_3} t_2^{j_{12}+j_{21}+j_{23}+l_3} t_3^{j_{13}+j_{22}+j_{23}+l_3} \right], \end{aligned}$$

where

$$\begin{aligned} s^* &= \left\{ (l_0, l_1, l_2, l_3) : 0 \leq l_i \leq n \text{ for } i = 0, 1, 2, 3, \text{ and } \sum_{i=0}^3 l_i = n \right\}, \\ s_1^* &= \left\{ (j_{11}, j_{12}, j_{13}) : 0 \leq j_{1i} \leq l_1 \text{ for } i = 1, 2, 3, \text{ and } \sum_{i=1}^3 j_{1i} = l_1 \right\}, \\ s_2^* &= \left\{ (j_{21}, j_{22}, j_{23}) : 0 \leq j_{2i} \leq l_2 \text{ for } i = 1, 2, 3, \text{ and } \sum_{i=1}^3 j_{2i} = l_2 \right\}. \end{aligned}$$

The second derivation utilizes the direct approach of differentiating the pgf to obtain the pmf. To simplify the notation, let $\lambda = \sum_{x_1=0}^1 \dots \sum_{x_d=0}^1 \lambda_{x_1 x_2 \dots x_d}$ as before and let $q_1 = \frac{\lambda_{000\dots 01}}{\lambda}$, $q_2 = \frac{\lambda_{000\dots 10}}{\lambda}$, $q_3 = \frac{\lambda_{000\dots 11}}{\lambda}$, $\dots$, $q_{2^d-1} = \frac{\lambda_{111\dots 1}}{\lambda}$. As a result, the general joint pmf is given by

$$p(\mathbf{x}) = \frac{1}{Z(\lambda, \nu)} \sum_{a=0}^{\sum x_i - \max(\mathbf{x})} \left(\frac{\partial^a Z(\lambda_{000\dots 0}, \nu)}{\partial (\lambda_{000\dots 0})^a} \sum_{\substack{k_1, k_2, \dots, k_{(2^d-1)} \in N \\ \sum_j b_{ij} k_j = x_i \\ \sum_j (\sum_i b_{ij} - 1) k_j = a}} \frac{q_1^{k_1} q_2^{k_2} \dots q_{(2^d-1)}^{k_{(2^d-1)}}}{k_1! k_2! \dots k_{(2^d-1)}!} \right).$$

where $\mathbf{x} = (x_1, x_2, \dots, x_d)$ and $\max(\mathbf{x}) = \max(x_1, x_2, \dots, x_d)$. In the trivariate case, the model simplifies to

$$p(\mathbf{x}) = \frac{1}{Z(\lambda, \nu)} \sum_{a=0}^{\sum x_i - \max(\mathbf{x})} \left(\frac{\partial^a Z(\lambda_{000}, \nu)}{\partial (\lambda_{000})^a} \sum_{\substack{k_1, k_2, \dots, k_7 \in N \\ \sum_j b_{ij} k_j = x_i \\ \sum_j (\sum_i b_{ij} - 1) k_j = a}} \frac{q_1^{k_1} q_2^{k_2} \dots q_7^{k_7}}{k_1! k_2! \dots k_7!} \right) \quad (\text{A4})$$

where $\mathbf{x} = (x_1, x_2, x_3)$ and $\max(\mathbf{x}) = \max(x_1, x_2, x_3)$.

Appendix B. Derivations of Moments

Let $\lambda_{ij+} = \lambda_{ij1} + \lambda_{ij0}$ for $i = 0, 1$; and $j = 0, 1$ and $\lambda_{i++} = \lambda_{i00} + \lambda_{i01} + \lambda_{i10} + \lambda_{i11}$ for $i = 0, 1$; λ_{i+j} , λ_{+ij} , λ_{+i+} , λ_{++i} are similarly defined. By differentiating the fmfg with respect to t_1, t_2, t_3 , and then setting $t_1 = t_2 = t_3 = 0$, we obtain the joint factorial moments of the trivariate form, (X_1, X_2, X_3) . Accordingly, letting $Z^{(k)}(\cdot) = \frac{\partial^k Z}{\partial \lambda^k}$, the initial marginal and product moments are obtained as

$$\begin{aligned} \mu_{X_1} &= E(X_1) = \left\{ \frac{\partial \ln Z(\lambda, \nu)}{\partial \lambda} \right\} \lambda_{1++}, \\ \mu_{X_2} &= E(X_2) = \left\{ \frac{\partial \ln Z(\lambda, \nu)}{\partial \lambda} \right\} \lambda_{+1+}, \\ \mu_{X_3} &= E(X_3) = \left\{ \frac{\partial \ln Z(\lambda, \nu)}{\partial \lambda} \right\} \lambda_{++1}, \\ \mu_{X_1 X_2} &= E(X_1 X_2) = \frac{Z''(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{1++} \lambda_{+1+} + \frac{Z'(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{11+}, \\ \mu_{X_1 X_3} &= E(X_1 X_3) = \frac{Z''(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{1++} \lambda_{++1} + \frac{Z'(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{1+1}, \\ \mu_{X_2 X_3} &= E(X_2 X_3) = \frac{Z''(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{+1+} \lambda_{++1} + \frac{Z'(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{+11}, \\ \sigma_{X_1 X_2} &= Cov(X_1, X_2) = \left[\frac{Z''(\lambda, \nu) Z(\lambda, \nu) - \{Z'(\lambda, \nu)\}^2}{\{Z(\lambda, \nu)\}^2} \right] \lambda_{1++} \lambda_{+1+} + \frac{Z'(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{11+}, \\ \sigma_{X_1 X_3} &= Cov(X_1, X_3) = \left[\frac{Z''(\lambda, \nu) Z(\lambda, \nu) - \{Z'(\lambda, \nu)\}^2}{\{Z(\lambda, \nu)\}^2} \right] \lambda_{1++} \lambda_{++1} + \frac{Z'(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{1+1}, \\ \sigma_{X_2 X_3} &= Cov(X_2, X_3) = \left[\frac{Z''(\lambda, \nu) Z(\lambda, \nu) - \{Z'(\lambda, \nu)\}^2}{\{Z(\lambda, \nu)\}^2} \right] \lambda_{+1+} \lambda_{++1} + \frac{Z'(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{+11}, \\ \sigma_{X_1}^2 &= Var(X_1) = \left[\frac{Z''(\lambda, \nu) Z(\lambda, \nu) - \{Z'(\lambda, \nu)\}^2}{\{Z(\lambda, \nu)\}^2} \right] \lambda_{1++}^2 + \frac{Z'(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{1++}, \\ \sigma_{X_2}^2 &= Var(X_2) = \left[\frac{Z''(\lambda, \nu) Z(\lambda, \nu) - \{Z'(\lambda, \nu)\}^2}{\{Z(\lambda, \nu)\}^2} \right] \lambda_{+1+}^2 + \frac{Z'(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{+1+}, \\ \sigma_{X_3}^2 &= Var(X_3) = \left[\frac{Z''(\lambda, \nu) Z(\lambda, \nu) - \{Z'(\lambda, \nu)\}^2}{\{Z(\lambda, \nu)\}^2} \right] \lambda_{++1}^2 + \frac{Z'(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{++1}. \end{aligned}$$

The above moments demonstrate that dispersion definitions are maintained via ν for the marginal distributions, where $\nu = 1$ denotes equi-dispersion and $\nu < (>)1$ capture over- (under-) dispersion. For example, when $\nu = 1$, $Z(\lambda, \nu = 1) = Z'(\lambda, \nu = 1) = Z''(\lambda, \nu = 1) = \exp(\lambda)$ and $\ln(Z) = \lambda$, thus one can see that $\mu_{X_i} = \sigma_{X_i}^2$ for $i = 1, 2, 3$ (i.e., marginal equi-dispersion holds), while $\mu_{X_i} < (>) \sigma_{X_i}^2$ for $i = 1, 2, 3$ when $\nu < (>)1$.

Using the notation introduced in Section 3, we derive the expression below. Let

$$h(t_1, t_2, t_3) = \left[\lambda_{000} + \lambda_{100}(t_1 + 1) + \lambda_{010}(t_2 + 1) + \lambda_{001}(t_3 + 1) + \lambda_{110}(t_1 + 1)(t_2 + 1) \right. \\ \left. + \lambda_{101}(t_1 + 1)(t_3 + 1) + \lambda_{011}(t_2 + 1)(t_3 + 1) + \lambda_{111}(t_1 + 1)(t_2 + 1)(t_3 + 1) \right].$$

We then obtain the derivatives,

$$\begin{aligned} \frac{\partial G_{x_1, x_2, x_3}(t_1, t_2, t_3)}{\partial t_1} &= \frac{Z'\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} \frac{\partial h(t_1, t_2, t_3)}{\partial t_1} \\ &= \frac{Z'\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} [\lambda_{100} + \lambda_{110}(t_2 + 1) + \lambda_{101}(t_3 + 1) + \lambda_{111}(t_2 + 1)(t_3 + 1)], \\ \frac{\partial G_{x_1, x_2, x_3}(t_1, t_2, t_3)}{\partial t_2} &= \frac{Z'\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} \frac{\partial h(t_1, t_2, t_3)}{\partial t_2} \\ &= \frac{Z'\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} [\lambda_{010} + \lambda_{110}(t_1 + 1) + \lambda_{011}(t_3 + 1) + \lambda_{111}(t_1 + 1)(t_3 + 1)], \\ \frac{\partial G_{x_1, x_2, x_3}(t_1, t_2, t_3)}{\partial t_3} &= \frac{Z'\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} \frac{\partial h(t_1, t_2, t_3)}{\partial t_3} \\ &= \frac{Z'\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} [\lambda_{001} + \lambda_{101}(t_1 + 1) + \lambda_{011}(t_2 + 1) + \lambda_{111}(t_1 + 1)(t_2 + 1)], \\ \frac{\partial^2 G_{x_1, x_2, x_3}(t_1, t_2, t_3)}{\partial t_1 \partial t_2} &= \frac{Z''\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} [\lambda_{100} + \lambda_{110}(t_2 + 1) + \lambda_{101}(t_3 + 1) + \lambda_{111}(t_2 + 1)(t_3 + 1)] \\ &\quad \cdot [\lambda_{010} + \lambda_{110}(t_1 + 1) + \lambda_{011}(t_3 + 1) + \lambda_{111}(t_1 + 1)(t_3 + 1)] \\ &\quad + \frac{Z'\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} [\lambda_{110} + \lambda_{111}(t_3 + 1)], \\ \frac{\partial^2 G_{x_1, x_2, x_3}(t_1, t_2, t_3)}{\partial t_1 \partial t_3} &= \frac{Z''\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} [\lambda_{100} + \lambda_{110}(t_2 + 1) + \lambda_{101}(t_3 + 1) + \lambda_{111}(t_2 + 1)(t_3 + 1)] \\ &\quad \cdot [\lambda_{001} + \lambda_{101}(t_1 + 1) + \lambda_{011}(t_2 + 1) + \lambda_{111}(t_1 + 1)(t_2 + 1)] \\ &\quad + \frac{Z'\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} [\lambda_{101} + \lambda_{111}(t_2 + 1)], \\ \frac{\partial^2 G_{x_1, x_2, x_3}(t_1, t_2, t_3)}{\partial t_2 \partial t_3} &= \frac{Z''\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} [(\lambda_{010} + \lambda_{110}(t_1 + 1) + \lambda_{011}(t_3 + 1) + \lambda_{111}(t_1 + 1)(t_3 + 1))] \\ &\quad \cdot [\lambda_{001} + \lambda_{101}(t_1 + 1) + \lambda_{011}(t_2 + 1) + \lambda_{111}(t_1 + 1)(t_2 + 1)] \\ &\quad + \frac{Z'\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} [\lambda_{011} + \lambda_{111}(t_1 + 1)], \\ \frac{\partial^2 G_{x_1, x_2, x_3}(t_1, t_2, t_3)}{\partial t_1^2} &= \frac{Z''\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} [\lambda_{100} + \lambda_{110}(t_2 + 1) + \lambda_{101}(t_3 + 1) + \lambda_{111}(t_2 + 1)(t_3 + 1)]^2, \\ \frac{\partial^2 G_{x_1, x_2, x_3}(t_1, t_2, t_3)}{\partial t_2^2} &= \frac{Z''\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} [\lambda_{010} + \lambda_{110}(t_1 + 1) + \lambda_{011}(t_3 + 1) + \lambda_{111}(t_1 + 1)(t_3 + 1)]^2, \\ \frac{\partial^2 G_{x_1, x_2, x_3}(t_1, t_2, t_3)}{\partial t_3^2} &= \frac{Z''\{h(t_1, t_2, t_3), \nu\}}{Z(\lambda, \nu)} [\lambda_{001} + \lambda_{101}(t_1 + 1) + \lambda_{011}(t_2 + 1) + \lambda_{111}(t_1 + 1)(t_2 + 1)]^2. \end{aligned}$$

Then, the expected value of X_1 is

$$\begin{aligned}\mu_{X_1} &= \frac{Z'\{h(t_1, t_2, t_3), v\}}{Z(\lambda, v)} [\lambda_{100} + \lambda_{110}(t_2 + 1) + \lambda_{101}(t_3 + 1) + \lambda_{111}(t_2 + 1)(t_3 + 1)] \Big|_{t_1=t_2=t_3=0} \\ &= \frac{Z'(\lambda, v)}{Z(\lambda, v)} \lambda_{1++} \\ &= \left\{ \frac{\partial \ln Z(\lambda, v)}{\partial \lambda} \right\} \lambda_{1++};\end{aligned}$$

similarly, $\mu_{X_2} = \left\{ \frac{\partial \ln Z(\lambda, v)}{\partial \lambda} \right\} \lambda_{+1+}$, and $\mu_{X_3} = \left\{ \frac{\partial \ln Z(\lambda, v)}{\partial \lambda} \right\} \lambda_{++1}$. Meanwhile,

$$\begin{aligned}\mu_{X_1 X_2} = E(X_1 X_2) &= \frac{\partial^2 G_{x_1, x_2, x_3}(t_1, t_2, t_3)}{\partial t_1 \partial t_2} \Big|_{t_1=t_2=t_3=0} \\ &= \frac{Z''\{h(t_1, t_2, t_3), v\}}{Z(\lambda, v)} [\lambda_{100} + \lambda_{110}(t_2 + 1) + \lambda_{101}(t_3 + 1) + \lambda_{111}(t_2 + 1)(t_3 + 1)] \\ &\quad \cdot [\lambda_{010} + \lambda_{110}(t_1 + 1) + \lambda_{011}(t_3 + 1) + \lambda_{111}(t_1 + 1)(t_3 + 1)] \\ &\quad + \frac{Z'\{h(t_1, t_2, t_3), v\}}{Z(\lambda, v)} [\lambda_{110} + \lambda_{111}(t_3 + 1)] \Big|_{t_1=t_2=t_3=0} \\ &= \frac{Z''(\lambda, v)}{Z(\lambda, v)} \lambda_{1++} \lambda_{+1+} + \frac{Z'(\lambda, v)}{Z(\lambda, v)} \lambda_{11+};\end{aligned}$$

similarly, $\mu_{X_1 X_3} = \frac{Z''(\lambda, v)}{Z(\lambda, v)} \lambda_{1++} \lambda_{++1} + \frac{Z'(\lambda, v)}{Z(\lambda, v)} \lambda_{1+1}$ and $\mu_{X_2 X_3} = \frac{Z''(\lambda, v)}{Z(\lambda, v)} \lambda_{+1+} \lambda_{++1} + \frac{Z'(\lambda, v)}{Z(\lambda, v)} \lambda_{+11}$. Consequently, we obtain the covariances to be

$$\begin{aligned}\sigma_{X_1 X_2} = \text{Cov}(X_1, X_2) &= E(X_1 X_2) - E(X_1)E(X_2) \\ &= \left\{ \frac{Z''(\lambda, v)}{Z(\lambda, v)} \lambda_{1++} \lambda_{+1+} + \frac{Z'(\lambda, v)}{Z(\lambda, v)} \lambda_{11+} \right\} - \left\{ \frac{Z'(\lambda, v)}{Z(\lambda, v)} \lambda_{1++} \frac{Z'(\lambda, v)}{Z(\lambda, v)} \lambda_{+1+} \right\} \\ &= \left\{ \frac{Z''(\lambda, v)}{Z(\lambda, v)} \lambda_{1++} \lambda_{+1+} + \frac{Z'(\lambda, v)}{Z(\lambda, v)} \lambda_{11+} \right\} - \left\{ \left(\frac{Z'(\lambda, v)}{Z(\lambda, v)} \right)^2 \lambda_{1++} \lambda_{+1+} \right\} \\ &= \left[\frac{Z''(\lambda, v)Z(\lambda, v) - \{Z'(\lambda, v)\}^2}{\{Z(\lambda, v)\}^2} \right] \lambda_{1++} \lambda_{+1+} + \frac{Z'(\lambda, v)}{Z(\lambda, v)} \lambda_{11+};\end{aligned}$$

similarly,

$$\begin{aligned}\sigma_{X_1 X_3} &= \left[\frac{Z''(\lambda, v)Z(\lambda, v) - \{Z'(\lambda, v)\}^2}{\{Z(\lambda, v)\}^2} \right] \lambda_{1++} \lambda_{++1} + \frac{Z'(\lambda, v)}{Z(\lambda, v)} \lambda_{1+1}, \text{ and} \\ \sigma_{X_2 X_3} &= \left[\frac{Z''(\lambda, v)Z(\lambda, v) - \{Z'(\lambda, v)\}^2}{\{Z(\lambda, v)\}^2} \right] \lambda_{+1+} \lambda_{++1} + \frac{Z'(\lambda, v)}{Z(\lambda, v)} \lambda_{+11}.\end{aligned}$$

Finally, noting that $E\{X_1(X_1 - 1)\} = \frac{\partial^2 G_{x_1, x_2, x_3}(t_1, t_2, t_3)}{\partial t_1^2} \Big|_{t_1=t_2=t_3=0}$, we obtain the variances as

$$\begin{aligned}\sigma_{X_1}^2 &= \text{Var}(X_1) = E\{X_1(X_1 - 1)\} + E(X_1) - \{E(X_1)\}^2 \\ &= \frac{Z''(\lambda, v)}{Z(\lambda, v)} \lambda_{1++}^2 + \frac{Z'(\lambda, v)}{Z(\lambda, v)} \lambda_{1++} - \left\{ \frac{Z'(\lambda, v)}{Z(\lambda, v)} \right\}^2 \lambda_{1++}^2 \\ &= \left[\frac{Z''(\lambda, v)Z(\lambda, v) - \{Z'(\lambda, v)\}^2}{\{Z(\lambda, v)\}^2} \right] \lambda_{1++}^2 + \frac{Z'(\lambda, v)}{Z(\lambda, v)} \lambda_{1++};\end{aligned}$$

similarly,

$$\begin{aligned}\sigma_{X_2}^2 &= \left[\frac{Z''(\lambda, \nu)Z(\lambda, \nu) - \{Z'(\lambda, \nu)\}^2}{\{Z(\lambda, \nu)\}^2} \right] \lambda_{+1+}^2 + \frac{Z'(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{+1+}, \text{ and} \\ \sigma_{X_3}^2 &= \left[\frac{Z''(\lambda, \nu)Z(\lambda, \nu) - \{Z'(\lambda, \nu)\}^2}{\{Z(\lambda, \nu)\}^2} \right] \lambda_{++1}^2 + \frac{Z'(\lambda, \nu)}{Z(\lambda, \nu)} \lambda_{++1}.\end{aligned}$$

Appendix C. Introduction to the Multivariate sCMP Model

A multivariate form of the sum of CMPs (MSCMP) distribution is an extension of the sCMP model in [24]. This section applies various MSCMP forms as alternative models to analyze the Corporación Favorita grocery sales data. The MSCMP distribution is defined as follows: given iid random variables $W_1, W_2, \dots, W_m$ that are $\text{MCMP}(\lambda_{00\dots 0}, \dots, \lambda_{11\dots 1}, \nu)$ distributed, $Y = \sum_{i=1}^m W_i$ has a $\text{MSCMP}(\lambda_{00\dots 0}, \dots, \lambda_{11\dots 1}, \nu, m)$ distribution with pgf

$$\Pi(t_1^*, \dots, t_d^*) = \left(\frac{Z \left[\left(\sum_{x_1=0}^1 \cdots \sum_{x_d=0}^1 \lambda_{x_1 x_2 \dots x_d} \prod_{i=1}^d (t_i^*)^{x_i} \right), \nu \right]}{Z(\lambda, \nu)} \right)^m. \quad (\text{A5})$$

Though the sCMP is defined as the sum of multiple CMPs, m need not be integer-valued since the pgf is valid for all $m \geq 0$. This MSCMP distribution includes the MNB (when $\nu = 0$), the MP ($\nu = 1$), and the MB ($\nu \rightarrow \infty$) distributions all as special cases, and serves as an over-arching distribution that connects the three special cases.

Given the difficulties associated with calculating the pmf for the SCMP model, we pursue an alternative approach to compute its pmf for a positive integer, m . For simplicity, we illustrate this approach in the trivariate case and consider $m = 2$: let $W_1 = (W_{11}, W_{12}, W_{13})$, $W_2 = (W_{21}, W_{22}, W_{23})$ be iid trivariate $\text{CMP}(\lambda_{000}, \dots, \lambda_{111}, \nu)$, and let $Y = (Y_1, Y_2, Y_3) = W_1 + W_2$; then Y has a trivariate $\text{SCMP}(\lambda_{000}, \dots, \lambda_{111}, \nu, m = 2)$ distribution, and

$$\begin{aligned}p(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) &= \sum_{a,b,c} p(W_{11} = a, W_{12} = b, W_{13} = c) \\ &\quad \cdot p(W_{21} = y_1 - a, W_{22} = y_2 - b, W_{23} = y_3 - c), (\text{A6})\end{aligned}$$

where a, b, c are non-negative integers such that $y_1 - a \geq 0$, $y_2 - b \geq 0$, and $y_3 - c \geq 0$; similarly, we can determine the pmf of the trivariate $\text{SCMP}(\lambda_{000}, \dots, \lambda_{111}, \nu, m = 3)$ distribution.

We fitted the Corporación Favorita grocery sales dataset with the trivariate SCMP model so as to demonstrate its capability of dealing with trivariate count data. Table A1 provides the resulting trivariate SCMP estimates with $m = 2$ and $m = 3$ where, for comparison, we also include the results of the trivariate NB and CMP models. Accordingly, we find that the trivariate SCMP models fit the data better than the trivariate CMP, with improvement growing with m . More precisely, we note that, as m increases, the log-likelihood increases while the AIC decreases. In particular, $\hat{\nu}$ likewise decreases toward 0 (which results in the trivariate NB model) as m increases. Unfortunately, current computational issues prevent us from providing SCMP results for $m > 3$, but these results do illustrate that the SCMP model will produce a log-likelihood no worse than that from the trivariate NB as it is a special case of bivariate SCMP model.

Table A1. Estimation results associated with the Corporación Favorita grocery sales data based on various assumed trivariate models: CMP, sCMP ($m = 2$), sCMP ($m = 3$), and NB. Respective log-likelihood and Akaike Information Criterion (AIC) values are also provided.

Model	Estimated Parameters			Log Likelihood	No. of Parameters	AIC
CMP	$\hat{\lambda}_{000} = 0.00$	$\hat{\lambda}_{111} = 0.00$	$\hat{\lambda}_{100} = 0.32$	−804.9	9	1627.9
	$\hat{\lambda}_{010} = 0.47$	$\hat{\lambda}_{001} = 1.17$	$\hat{\lambda}_{110} = 0.00$			
	$\hat{\lambda}_{101} = 0.00$	$\hat{\lambda}_{011} = 0.00$	$\hat{\nu} = 0.25$			
sCMP ($m = 2$)	$\hat{\lambda}_{000} = 0.05$	$\hat{\lambda}_{111} = 0.00$	$\hat{\lambda}_{100} = 0.22$	−804.0	9	1626.0
	$\hat{\lambda}_{010} = 0.33$	$\hat{\lambda}_{001} = 0.82$	$\hat{\lambda}_{110} = 0.00$			
	$\hat{\lambda}_{101} = 0.00$	$\hat{\lambda}_{011} = 0.00$	$\hat{\nu} = 0.19$			
sCMP ($m = 3$)	$\hat{\lambda}_{000} = 0.12$	$\hat{\lambda}_{111} = 0.00$	$\hat{\lambda}_{100} = 0.16$	−803.5	9	1625.0
	$\hat{\lambda}_{010} = 0.24$	$\hat{\lambda}_{001} = 0.61$	$\hat{\lambda}_{110} = 0.00$			
	$\hat{\lambda}_{101} = 0.00$	$\hat{\lambda}_{011} = 0.00$	$\hat{\nu} = 0.13$			
NB	$\hat{a}_1 = -0.44$	$\hat{a}_2 = -0.65$	$\hat{a}_3 = -1.61$	−802.8	8	1621.7
	$\hat{a}_{12} = 0.00$	$\hat{a}_{13} = 0.00$	$\hat{a}_{23} = 0.00$			
	$\hat{a}_{123} = 0.00$	$\hat{a}_0 = 3.69$	$\hat{k} = 6.01$			

Appendix D. Real Datasets

Table A2. Corporación Favorita sales data: Unit sales of an item in each of three stores over 100 days.

Day	Store 1	Store 2	Store 3	Day	Store 1	Store 2	Store 3	Day	Store 1	Store 2	Store 3
1	2	5	6	35	2	3	3	68	1	3	5
2	3	8	23	36	4	2	6	69	1	2	3
3	2	8	21	37	3	2	16	70	0	4	4
4	4	5	8	38	7	12	12	71	1	4	10
5	2	7	13	39	6	0	13	72	1	0	14
6	1	4	15	40	2	2	4	73	0	1	14
7	0	0	11	41	1	0	13	74	0	2	19
8	1	6	2	42	3	6	4	75	0	4	6
9	6	6	7	43	2	0	3	76	7	2	12
10	3	8	13	44	7	2	4	77	1	4	7
11	3	8	16	45	5	7	3	78	1	0	7
12	0	1	7	46	8	1	19	79	3	1	8
13	0	5	11	47	1	6	17	80	4	3	17
14	6	8	7	48	1	2	5	81	4	7	7
15	1	2	4	49	3	6	5	82	3	3	13
16	8	4	4	50	5	2	9	83	1	0	9
17	3	10	20	51	10	1	0	84	1	1	11
18	6	6	12	52	4	4	11	85	0	2	8
19	0	1	6	53	1	5	25	86	6	5	11
20	3	7	10	54	3	5	4	87	0	0	9
21	0	2	5	55	1	7	3	88	0	4	8
22	3	4	1	56	2	1	4	89	1	2	15
23	3	4	3	57	1	5	2	90	1	6	4
24	7	17	13	58	0	3	7	91	0	2	11
25	2	4	11	59	6	4	12	92	0	2	5
26	0	5	9	60	8	1	11	93	8	2	11
27	3	1	2	61	3	4	12	94	3	3	21
28	2	3	1	62	1	2	5	95	3	9	21
29	2	9	4	63	5	0	3	96	0	8	8
30	4	2	7	64	2	1	3	97	4	5	27
31	6	4	12	65	0	3	7	98	5	3	15
32	1	6	18	66	1	2	26	99	1	2	4
33	0	11	15	67	0	1	17	100	4	3	6
34	2	7	12								

Table A3. NBA All-Star data: Number of players selected for the NBA all-star game each year in each position (C = Center; F = Forward; FC = Forward-center).

Year	C	F	FC	Year	C	F	FC	Year	C	F	FC
2000	6	1	4	2006	3	3	4	2012	3	2	4
2001	3	4	3	2007	2	3	3	2013	2	2	4
2002	4	3	4	2008	3	2	3	2014	2	2	5
2003	4	2	3	2009	2	3	5	2015	2	4	4
2004	3	4	4	2010	2	2	5	2016	3	4	2
2005	2	2	5	2011	4	2	2				

References

- Johnson, N.; Kotz, S.; Balakrishnan, N. *Discrete Multivariate Distributions*; John Wiley & Sons: New York, NY, USA, 1997.
- Krishnamoorthy, A.S. Multivariate binomial and Poisson distributions. *Sankhyā Indian J. Stat.* **1951**, *11*, 117–124.
- Mahamunulu, D.M. A note on regression in the multivariate Poisson distribution. *J. Am. Stat. Assoc.* **1967**, *62*, 251–258. [[CrossRef](#)]
- Teicher, H. On the Multivariate Poisson distribution. *Skand. Aktuarietidskr.* **1954**, *37*, 1–9. [[CrossRef](#)]
- Hilbe, J.M. *Modeling Count Data*; Cambridge University Press: New York, NY, USA, 2014.
- Doss, D.C. Definition and characterization of multivariate negative binomial distribution. *J. Multivar. Anal.* **1979**, *9*, 460–464. [[CrossRef](#)]
- Conway, R.W.; Maxwell, W.L. A queuing model with state dependent service rates. *J. Ind. Eng.* **1962**, *12*, 132–136.
- Sellers, K.F.; Shmueli, G.; Borle, S. The COM-Poisson model for count data: A survey of methods and applications. *Appl. Stoch. Model. Bus. Ind.* **2011**, *28*, 104–116. [[CrossRef](#)]
- Shmueli, G.; Minka, T.P.; Kadane, J.B.; Borle, S.; Boatwright, P. A useful distribution for fitting discrete data: Revival of the Conway-Maxwell-Poisson distribution. *Appl. Stat.* **2005**, *54*, 127–142. [[CrossRef](#)]
- Guikema, S.D.; Coffelt, J.P. A Flexible Count Data Regression Model for Risk Analysis. *Risk Anal.* **2008**, *28*, 213–223. [[CrossRef](#)] [[PubMed](#)]
- Sellers, K.F.; Morris, D.S.; Balakrishnan, N. Bivariate Conway-Maxwell-Poisson distribution: Formulation, properties, and inference. *J. Multivar. Anal.* **2016**, *150*, 152–168. [[CrossRef](#)]
- Kocherlakota, S.; Kocherlakota, K. *Bivariate Discrete Distributions*; Marcel Dekker: New York, NY, USA, 1992.
- Lai, C.D. Constructions of discrete bivariate distributions. In *Advances in Distribution Theory, Order Statistics and Inference, Part I*; Balakrishnan, N., Sarabia, J.M., Castillo, E., Eds.; Birkhauser: Boston, MA, USA, 2006; pp. 29–58.
- Marshall, A.W.; Olkin, I. A family of bivariate distributions generated by the bivariate Bernoulli distribution. *J. Am. Stat. Assoc.* **1985**, *80*, 332–338. [[CrossRef](#)]
- R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017.
- Balakrishnan, N.; Pal, S. Lognormal lifetimes and likelihood-based inference for flexible cure rate models based on COM-Poisson family. *Comput. Stat. Data Anal.* **2013**, *67*, 41–67. [[CrossRef](#)]
- Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodel Inference*; Springer: New York, NY, USA, 2002.
- Corporación Favorita. Grocery Sales Data. 2018. Available online: <https://www.kaggle.com/c/favorita-grocery-sales-forecasting/data> (accessed on 26 April 2020).
- Voinov, V.; Nikulin, M.; Balakrishnan, N. *Chi-Squared Goodness of Fit Tests with Applications*; Academic Press: Boston, MA, USA, 2013.
- NBA. NBA All-Star Game, 2000–2016. Available online: <https://www.kaggle.com/fmeja21/nba-all-star-game-20002016?> (accessed on 22 April 2020).
- Inouye, D.I.; Yang, E.; Allen, G.I.; Ravikumar, P. A review of multivariate distributions for count data derived from the Poisson distribution. *WIREs Comput. Stat.* **2017**, *9*, e1398. [[CrossRef](#)] [[PubMed](#)]
- Genest, C.; Nešlehová, J. A Primer on Copulas for Count Data. *ASTIN Bull.* **2007**, *37*, 475–515. [[CrossRef](#)]
- Trivedi, P.; Zimmer, D. A Note on Identification of Bivariate Copulas for Discrete Count Data. *Econometrics* **2017**, *5*, 10. [[CrossRef](#)]
- Sellers, K.F.; Swift, A.W.; Weems, K.S. A flexible distribution class for count data. *J. Stat. Distrib. Appl.* **2017**, *4*, 1–21. [[CrossRef](#)]