Inference for Bivariate Conway-Maxwell-Poisson Distribution and Its Application in Modeling Bivariate Count Data

INFERENCE FOR BIVARIATE CONWAY-MAXWELL-POISSON DISTRIBUTION AND ITS APPLICATION IN MODELING BIVARIATE COUNT DATA

BY

XINYI WANG, B.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF MATHEMATICS & STATISTICS

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Master of Science (2019)	McMaster University
(Mathematics & Statistics)	Hamilton, Ontario, Canada

TITLE:	Inference for Bivariate Conway-Maxwell-Poisson Distribution and Its Application in Modeling Bivariate Count Data
AUTHOR:	Xinyi Wang B.Sc., (Actuarial and Financial Math) McMaster University, Hamilton, Canada
SUPERVISOR:	Dr. Narayanaswamy Balakrishnan
NUMBER OF PAGES:	x, 59

To my parents, Hong and Hao.

Abstract

In recent actuarial literature, the bivariate Poisson regression model has been found to be useful for modeling paired count data. However, the basic assumption of marginal equi-dispersion may be quite restrictive in practice. To overcome this limitation, we consider here the recently developed bivariate Conway–Maxwell–Poisson (CMP) distribution. As a distribution that allows data dispersion, the bivariate CMP distribution is a flexible distribution which includes the bivariate Poisson, bivariate Bernoulli and bivariate Geometric distributions all as special cases. We discuss inferential methods for this CMP distribution. An application to automobile insurance data demonstrates its usefulness as an alternative framework to the commonly used bivariate Poisson model.

Acknowledgements

First and most importantly, I would like to express my great appreciation to my supervisor, Professor Narayanaswamy Balakrishnan, for his guidance and encouragement throughout my Master study. It is a great honor to get to work with someone who cares not only care about my research work but also inspires me every day.

I would also like to thank Dr. Shui Feng and Dr. Anas Abdallah, who were members of my examination committee, for making the defense enjoyable.

My sincere thanks also go to Dr. Xiaojun Zhu and Nikola Počuča for their valuable help on the theoretical and computational work involved in this thesis.

Thanks to all my friends for being my side. I am especially appreciative of Kai Liu who has been a wonderful teacher and a friend of mine in the past few years. I am grateful for all the suggestions she gave.

Finally, I would like to thank my dear parents, Hao Wang and Hong Cao, for shaping my life with positivity and passion, and encouraging me to chase my dreams. Because of their support, both financially and emotionally, my life seems full of happiness.

Notation and abbreviations

CMP	Conway–Maxwell–Poisson distribution
LRT	Likelihood Ratio Test
MLE	Maximum Likelihood Estimation
PGF	Probability Generating Function

Contents

List of Tables

Chapter 1

Introduction

Bermúdez and Karlis (2017) proposed an application of the bivariate Poisson distribution to a Spanish automobile data which provides a relatively accurate prediction of two different types of claim. As a well-know distribution, bivariate Poisson is useful for modeling paired count data. This thesis takes consideration of data dispersion and proposed the bivariate Conway–Maxwell–Poisson (CMP) distribution, which includes the bivariate Poisson, bivariate Bernoulli and bivariate Geometric all as three special cases, is an alternate approach of the bivariate Poisson distribution.

To start, Chapter 3 outlines the fundamental properties of distributions that are used in this work. Basic properties of the Bernoulli distribution are discussed in Kocherlakota and Kocherlakota (1992), Marshall and Olkin (1985). The Poisson distribution was introduced by M'Kendrick (1926). It is originally derived as a solution to a differential equation for a biological application. Later, Campbell (1934) proposed that it can be obtained from the Bernoulli distribution by taking limits in nth power of the factorial moment generating function. Moreover, the number of failure before the first successful trial has a Geometric distribution (Hawkes, 1972). Due to the natural bivariate property of Bernoulli distribution, all these basic facts can be extended to two dimensions.

The CMP distribution was first introduced in Conway and Maxwell (1962) and Shmueli *et al.* (2005). As a flexible distribution, it has also been applied as a bivariate distribution to numerous insurance and biology papers such as Sellers *et al.* (2016), Balakrishnan and Pal (2013). In this work, the usefulness has been revealed by applying it to the auto insurance dataset discussed above as an alternative approach to the bivariate Poisson distribution.

After a brief description of the automobile data in Chapter 2, an introductory theoretical framework of three classical bivariate distributions and bivariate CMP distribution has been described in Chapter 3. Chapter 4 addresses the method of maximum likelihood estimation and model selection criterion such as likelihood ratio test, AIC, and BIC. Results of some illustrating examples are presented in Chapter 5, followed by some conclusion and remarks have been discussed in the last Chapter.

Chapter 2

Data description

Data that have been used in this thesis originally contained a ten percent automobile portfolio of private used cars from an insurance company in Spain in 1995. Originally, the collected data has 80,994 profiles of customers who had been insured with the company for seven or more years.

The collected information included the number of annual accidents reported along with twelve exogenous variables that are listed in Table 2.1. Bermúdez and Karlis (2017) defined the third-party liability as type N_1 claims, and basic guarantees, comprehensive coverage and collision coverage as type N_2 claims. Third-party liability included damage caused by the policyholder to someone else's property. Basic guarantees here included emergency roadside assistance or legal and medical assistance. In addition, comprehensive coverage contained the damage of vehicle resulting from theft, fire or flood. The collision coverage was defined as damage resulting from a collision when the policyholder is at fault.

In this thesis, we have focused over attention on policyholders were clients for three or more years. Thus, v_7 is removed with restriction applied to v_8. Now, the

Table 2.1: Tweleve exogenous variables in the data set

Variable	**1**	**0**
v_1	women	men
v_2	driving in urban area	otherwise
v_3	zone is medium risk (Madrid and Catalonia)	
v_4	when zone is high risk (Northern Spain)	
v_5	if the driving license is between 4 and 14 years old	
v_6	if the driving license is 15 or more years old	
v_7	if the client is in the company between 3 and 5 years	
v_8	if the client is in the company for more than 5 years	
v_9	if the insured is 30 years old or younger	
v_{10}	if includes comprehensive coverage (exceptfire)	
v_{11}	if includes comprehensive and collision coverages	
v_{12}	if horsepower is greater than or equal to 5500cc	

variable v_8 equals 0 if the customer is insured with the company for less than 5 years. Also, variable v_{10} and v_{11} are classified as type N_2 claim. Therefore, there are nine exogenous variables that have been considered in the later analysis of the thesis. Similar to Table 1 of Bermúdez and Karlis (2017), Table C.2 in Appendix C shows the number of claims for types N_1 and N_2 obtained from these data. The total number of claims is 5986. The number of claims for N_1 is 3564 and for N_2 is 2422. The average frequency of N_1 and N_2 are 0.085 and 0.125, respectively.

Chapter 3

Background

In the Bivariate Conway-Maxwell-Poisson (CMP) distribution, there are three special cases that have been mentioned in Sellers *et al.* (2016), which are the bivariate Poisson, bivariate Bernoulli, and bivariate geometric distributions. The following sections focus on introducing the fundamental concepts of all four distributions starting with the bivariate Bernoulli distribution and followed by the bivariate Poisson and bivariate geometric distribution since they can be derived from a sequence of independent Bernoulli random variables. Specifically, the number of failures before the first successful results in a geometric distribution, while the number of failed trials before the rth success has a negative binomial distribution. The Poisson distribution can be obtained as a limit of the binomial and the negative binomial distributions. Finally, the general theoretical framework of the bivariate CMP, that encompasses all these three models as special cases, is described.

3.1 Three Prominent Bivariate Distributions

3.1.1 Bivariate Bernoulli Distribution

To start from the simplest case, we recall the probability mass function of Bernoulli random variable X is $f_X(x) = p^x(1-p)^{1-x}$, where x is the realization of X taking values 0 or 1. Similarly, consider another Bernoulli random variable Y. Assume they form a random pair (X, Y), random possessing the bivariate Bernoulli distribution (Marshall and Olkin, 1985; Kocherlakota and Kocherlakota, 1992). Possible outcomes of (**X**,**Y**) are (0,0), (0, 1), (1, 0), (1, 1) in the Cartesian product space $\{0,1\}^2 = \{0,1\} \times \{0,1\}$ with probabilities p_{00}, p_{01}, p_{10} and p_{11} respectively.

According to the above definition of the bivariate Bernoulli distribution, it is easy to see that the marginal distribution of X is $\mathrm{P}(X=0) = p_{00} + p_{01}$ and $\mathrm{P}(X=1) = p_{10} + p_{11}$. Then X has an univariate Bernoulli distribution with parameter $p_{10} + p_{11}$, and Y follows another univariate Bernoulli distribution with parameter $p_{01} + p_{11}$ since $\mathrm{P}(Y=1) = p_{01} + p_{11}$ and $\mathrm{P}(Y=0) = p_{00} + p_{10}$. The means of X and Y are then $p_{10} + p_{11}$ and $p_{01} + p_{11}$, respectively. It is also easy to see that

$$
\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) \\
&= p_{11} - (p_{10} + p_{11})(p_{01} + p_{11}) \\
&= p_{00}p_{11} - p_{10}p_{01}, \\
\mathrm{Corr}(X, Y) &= \mathrm{Cov}(X, Y)/\sqrt{p_{1+}p_{0+}p_{+1}p_{+0}}
\end{aligned}
$$

where $p_{1+} = p_{10} + p_{11}$, $p_{0+} = p_{00} + p_{01}$, $p_{+1} = p_{01} + p_{11}$, and $p_{+0} = p_{00} + p_{10}$. Due to the close relation between probabilities, a 2×2 contingency table is always used to illustrate this distribution.

The probability mass function of a bivariate Bernoulli distribution can be written as

$$\begin{aligned} P(x,y) &= p_{11}^{xy} p_{10}^{x(1-y)} p_{01}^{(1-x)y} p_{00}^{(1-x)(1-y)} \\ &= exp\left\{ log(p_{00}) + x\, log\left(\frac{p_{10}}{p_{00}}\right) + y\, log\left(\frac{p_{01}}{p_{00}}\right) + xy\, log\left(\frac{p_{11}p_{00}}{p_{10}p_{01}}\right) \right\}, \end{aligned} \tag{3.1}$$

for $x, y = 0, 1$, where the condition $p_{00} + p_{01} + p_{10} + p_{11} = 1$ is satisfied. Eq. (3.1) can be simplified using the natural parameters N's where

$$N_1 = log\left(\frac{p_{10}}{p_{00}}\right), \quad N_2 = log\left(\frac{p_{01}}{p_{00}}\right) \quad and \quad N_3 = log\left(\frac{p_{11}p_{00}}{p_{10}p_{01}}\right) \tag{3.2}$$

Then, the original mass function can be rewritten in a log-linear formulation as

$$P(x,y) = exp\{log(p_{00}) + x\, N_1 + y\, N_2 + xy\, N_3\}, \tag{3.3}$$

which is a member of the exponential family of distributions.

As stated in Dai *et al.* (2013), the bivariate Bernoulli distribution has properties similar to those of the bivariate Gaussian distribution. More specifically, the marginal and conditional distributions of the bivariate Bernoulli distribution are still Bernoulli.

Finally, we have the generating function, given in Kocherlakota and Kocherlakota (1992), as

$$G(t_1, t_2) = E(t_1^x, t_2^y) = p_{00} + p_{01}t_2 + p_{10}t_1 + p_{11}t_1t_2. \tag{3.4}$$

3.1.2 Bivariate Poisson Distribution

In this section, we consider the bivariate Poisson distribution which can be obtained by taking limits in the nth power (as n$\rightarrow\infty$) of the factorial moment generating function of the bivariate Bernoulli distribution; see Campbell (1934) for details.

We specifically recall the trivariate reduction method, which is a classic approach to construct the bivariate Poisson distribution by setting $X = P_1+P_3$ and $Y = P_2+P_3$, where independent P_i follow $Poisson(\lambda_i)$, for $i = 1, 2, 3$, see Johnson *et al.* (1997) for all pertinent details. It can be denoted by $(X, Y) \sim BP(\lambda_1, \lambda_2, \lambda_3)$, where the two marginal distribution follows two Poisson distributions. In this case,

$$\begin{aligned}
\mathrm{E}(X) &= \lambda_1 + \lambda_3, \\
\mathrm{E}(Y) &= \lambda_2 + \lambda_3, \\
\mathrm{Cov}(X, Y) &= \lambda_3, \\
\text{and } \mathrm{Corr}(X, Y) &= \lambda_3/\sqrt{(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3)}
\end{aligned}$$

. The correlation here is always non-negative. Further, the probability mass function of (X, Y) can be expressed as

$$\mathrm{P}(X = x, Y = y; \lambda_1, \lambda_2, \lambda_3) = \mathrm{e}^{-(\lambda_1+\lambda_2+\lambda_3)} \frac{\lambda_1^x}{x!} \frac{\lambda_2^y}{y!} \sum_{i=0}^{min(x,y)} \binom{x}{i} \binom{y}{i} i! \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^i, \quad (3.5)$$

where x and y are the realizations of X and Y, respectively. Then, we can show the probability generating function (pgf) of the bivariate Poisson vector (X, Y) to be

$$\mathrm{G}(t_1, t_2) = \exp\big\{(\lambda_1 + \lambda_3)(t_1 - 1) + (\lambda_2 + \lambda_3)(t_2 - 1) + \lambda_3(t_1 - 1)(t_2 - 1)\big\}, \quad (3.6)$$

as given in Johnson *et al.* (1997), for example. Alternatively, it can be constructed through a conditional bivariate binomial distribution, conditional on n^*, with joint pgf

$$G(t_1, t_2 \mid n^*) = \{1 + p_{1+}(t_1 - 1) + p_{+1}(t_2 - 1) + p_{11}(t_1 - 1)(t_2 - 1)\}^{n^*}, \tag{3.7}$$

where n^* is the number of trials such that $n^* \sim Poisson(\lambda^*)$. Then the unconditional joint pgf of (X, Y) is given by

$$\begin{aligned} G(t_1, t_2) &= \sum_{n^*=0}^{\infty} \frac{(\lambda^*)^{n^*} e^{-\lambda^*}}{n^*!} G(t_1, t_2 \mid n^*) \\ &= exp\big\{\lambda^* p_{1+}(t_1 - 1) + \lambda^* p_{+1}(t_2 - 1) + \lambda^* p_{11}(t_1 - 1)(t_2 - 1)\big\}, \end{aligned} \tag{3.8}$$

where $p_{1+} = p_{10} + p_{11}$, and $p_{+1} = p_{01} + p_{11}$. From (3.6) and (3.8), the relation between parameters λ_i, λ^* and p are readily seen to be

$$\lambda_1 + \lambda_3 = \lambda^* p_{1+}, \tag{3.9}$$

$$\lambda_2 + \lambda_3 = \lambda^* p_{+1}, \tag{3.10}$$

$$\lambda_3 = \lambda^* p_{11}. \tag{3.11}$$

The bivariate Poisson family is useful when applying it to the distribution of claims in insurance. Because the sum of two bivariate Poisson vectors is still a bivariate Poisson vector, the claims in more than one time periods follow the same distribution with only different parameters, which is a useful and convenient property to have.

3.1.3 Bivariate Geometric Distribution

Consider the bivariate geometric framework discussed in Basu and Dhar (1995). Suppose a system has two components and three causes of failure which are the failure of the first component only, failure of the second component only, and failure of both components simultaneously. Here, the three events possess Binomial distributions with failure probabilities p_1, p_2, and p_3, respectively, and can be denoted by $B(x, 1-p_1)$, $B(y, 1-p_2)$ and $B(x \vee y, 1-p_3)$, where $x \vee y = max(x, y)$. Now consider $X \in \mathbb{Z}^+$ and $Y \in \mathbb{Z}^+$ that have discrete lifetime distribution of first and second components of the considered system. The survival function for this bivariate system can then be expressed as

$$\begin{aligned} P(X > x, Y > y) &= P(B(x, 1-p_1) = 0, B(y, 1-p_2) = 0, B(x \vee y, 1-p_3) = 0) \\ &= p_1^x p_2^y p_3^{max(x,y)} \end{aligned} \tag{3.12}$$

and satisfies $0 < p_1, p_2 < 1$ and $0 < p_3 \leq 1$ since p_1, p_2 and p_3 are probabilities. Moreover, it also satisfies $x \leq 1$, $y \in \mathbb{Z}^+$ where $\mathbb{Z}^+$ indicates positive intergers. Since the survival function has the loss of memory property,

$$P(X > x + k, Y > y + k | X > x, Y > y) = P(X > k, Y > k) = (p_1\, p_2\, p_3)^k,$$

for some positive interger k. We expend (3.12) and obtain

$$\begin{aligned} P(X = x, Y = y) = P(X > x-1, Y > y-1) - P(X > x, Y > y-1) \\ - P(X > x-1, Y > y) + P(X > x, Y > y-1). \end{aligned}$$

It is similar as treat $P(X = x, Y = y)$ as three cases under $x < y$, $x = y$ and $x > y$.

$$P(X = x, Y = y) = \begin{cases} p_1^{x-1}(p_2p_3)^{y-1}\,(1-p_1)\,(1-p_2p_3) & \text{for } x < y, \\ (p_1\,p_2\,p_3)^{x-1}(1-p_1p_3-p_2p_3+p_1p_2p_3) & \text{for } x = y, \\ p_2^{y-1}(p_1p_3)^{x-1}(1-p_2)(1-p_1p_3) & \text{for } x > y. \end{cases} \tag{3.13}$$

The probability generating function is given by

$$\begin{aligned} G(t_1,t_2) &= E(t_1^x, t_2^y) \\ &= \sum_{(x,y)\in T} t_1^x t_2^y f(x,y) \\ &= \sum_{x<y} t_1^x t_2^y f(x,y) + \sum_{x>y} t_1^x t_2^y f(x,y) + \sum_{x=y} t_1^x t_2^y f(x,y) \\ &= \sum_{x<y} t_1^x t_2^y p_1^{x-1}(p_2p_3)^{y-1} q_1(1-p_2p_3) + \sum_{x>y} t_1^x t_2^y p_2^{y-1}(p_1p_3)^{x-1} q_2(1-p_1p_3) \\ &\quad + \sum_{x=y} t_1^x t_2^y (p_1p_2p_3)^(x-1)(1-p_1p_3-p_2p_3+p_1p_2p_3) \\ &= \frac{t_1t_2q_1(1-p_2p_3)(t_2p_2p_3)}{(1-t_1t_2p_1p_2p_3)} + \frac{t_1t_2q_2(1-p_1p_3)(t_1p_1p_3)}{(1-t_1t_2p_1p_2p_3)(1-t_1p_1p_3)} \\ &\quad + \frac{t_1t_2(1-p_1p_3-p_2p_3+p_1p_2p_3)}{1-t_!t_2p_1p_2p_3}, \end{aligned}$$

where $0 < p_i < 1, i = 1, 2, 3, |t_1| < 1/p_1, |t_2| < 1/p_2$ and $|t_1t_2| < 1/p_1p_2p_3$.

3.2 Univariate Conway–Maxwell–Poisson (CMP) Distribution

The Conway–Maxwell–Poisson (CMP) distribution is a model that naturally generalizes the Poisson distribution and allows for over-dispersion and under-dispersion in a given dataset using dispersion parameter ν. It was introduced by Conway and Maxwell (1962), and its probability mass function is given by

$$P(\mathbf{K} = k) = \frac{\lambda^k}{(k!)^\nu} \frac{1}{Z(\lambda, \nu)}, \qquad k = 0, 1, 2, \ldots, \tag{3.14}$$

where

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

with $\nu \geq 0, \lambda > 0$.

The dispersion parameter ν indicates equi-dispersion when $\nu = 1$, over-dispersion when $\nu < 1$, and under-dispersion when $\nu > 1$. The CMP distribution possesses three special cases. When $\nu = 1$, it is a Poisson distribution with parameter λ and in this case the dataset has equi-dispersion, and the normalizing constant becomes $Z(\lambda, \nu) = exp(\lambda)$. Because $\nu \to \infty$ implies $Z(\lambda, \nu) \to 1 + \lambda$, the CMP in this case becomes a Bernoulli distribution with parameter $\frac{\lambda}{1+\lambda}$. When $\nu = 0$ and $\lambda < 1$, it reduces a geometric distribution with success probability $1 - \lambda$. In this case, $Z(\lambda, \nu) = \sum_{i=0}^{\infty} \lambda^i = \frac{1}{1-\lambda}$ is the normalizing constant, and the mass function can be written as $\mathrm{P}(K = k; \lambda) = \lambda^k (1 - \lambda)$, i.e., a geometric distribution. When $\lambda \geq 1$ and $\nu = 0$, $Z(\lambda, \nu)$ does not coverage, and so in this case, the distribution is undefined.

Moments of the CMP distribution can be derived using recursive methods since it is from a family of two-parameter power series distribution; see Johnson *et al.* (1992) for details. It has the form of

$$E[\mathbf{K}^{r+1}] = \begin{cases} \lambda E[\mathbf{K}+1]^{1-\nu} & \text{for } r = 0, \\ \lambda \dfrac{\partial}{\partial \lambda} E(\mathbf{K}^r) + E(\mathbf{K})E(\mathbf{K}^r) & \text{for } r > 0. \end{cases} \tag{3.15}$$

The approximation of $E(\mathbf{K})$ can be obtained using asymptotic approximation for the normalizing constant $Z(\lambda, \nu)$. Indeed, the mean is given by

$$\begin{aligned} E(\mathbf{K}) &= \lambda \frac{\partial}{\partial \lambda} log(Z(\lambda, \nu)) \\ &\approx \lambda^{\frac{1}{\nu}} - \frac{\nu - 1}{2\nu}. \end{aligned} \tag{3.16}$$

Similarly, the variance, moment generating function and pgf of $\mathbf{K}$ are given by

$$Var(\mathbf{K}) = \lambda \frac{\partial E(\mathbf{G})}{\partial Log\lambda} \approx \frac{1}{\nu} \lambda^{\frac{1}{\nu}}, \tag{3.17}$$

$$M_{\mathbf{K}}(t) = E(e^{\mathbf{K}t}) = \frac{Z(\lambda e^t, \nu)}{Z(\lambda, \nu)}, \tag{3.18}$$

$$G_{\mathbf{K}}(t) = E(t^{\mathbf{K}}) = \frac{Z(\lambda t, \nu)}{Z(\lambda, \nu)}. \tag{3.19}$$

3.3 Bivariate Conway–Maxwell–Poisson (CMP) Distribution

In order to derive the distribution of the bivariate CMP distribution, we consider the method stated in Sellers *et al.* (2016) which use bivariate Poisson distribution while the number of trials is modeled via $CMP(\lambda, \nu)$. Applying the compounding method which enables us to rewrite the joint pgf of the CMP distribution along with (3.7) as

$$
\begin{aligned}
G\left(t_1, t_2\right) &= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!^{\nu} Z(\lambda, \nu)}\, G(t_1, t_2|n) \\
&= \sum_{n=0}^{\infty} \frac{\lambda^n \{1 + p_{1+}(t_1 - 1) + p_{+1}(t_2 - 1) + p_{11}(t_1 - 1)(t_2 - 1)\}^n}{n!^{\nu} Z(\lambda, \nu)} \qquad (3.20) \\
&= \frac{Z\left[\lambda\{1 + p_{1+}(t_1 - 1) + p_{+1}(t_2 - 1) + p_{11}(t_1 - 1)(t_2 - 1)\}, \nu\right]}{Z(\lambda, \nu)} \qquad (3.21) \\
&= \frac{1}{Z(\lambda, \nu)} \sum_{n=0}^{\infty} \frac{1}{(n!)^{\nu}} (\mathcal{A} + \mathcal{B}t_1 + \mathcal{C}t_2 + \mathcal{D}t_1 t_2)^n \qquad (3.22) \\
&= \frac{1}{Z(\lambda, \nu)} \sum_{n=0}^{\infty} \frac{1}{(n!)^{\nu}} \times \mathcal{W}, \qquad (3.23)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{A} &= \lambda p_{00} = \lambda(1 - p_{+1} - p_{1+} + p_{11}), \\
\mathcal{B} &= \lambda p_{10} = \lambda(p_{1+} - p_{11}), \\
\mathcal{C} &= \lambda p_{01} = \lambda(p_{+1} - p_{11}), \\
\mathcal{D} &= \lambda p_{11}, \\
\mathcal{W} &= (\mathcal{A} + \mathcal{B}t_1 + \mathcal{C}t_2 + \mathcal{D}t_1 t_2)^n. \qquad (3.24)
\end{aligned}
$$

We use the multinomial expansion, then $\mathcal{W}$ has the following forms

$$\begin{aligned} \mathcal{W} &= (\mathcal{A} + \mathcal{B}t_1 + \mathcal{C}t_2 + \mathcal{D}t_1t_2)^n \\ &= \sum_{a,b,c=0;a+d+c\leq n}^{n} \binom{n}{a, b, c, n-a-b-c} \mathcal{A}^a \mathcal{B}^b \mathcal{C}^c \mathcal{D}^{n-a-b-c} t_1^{n-a-c} t_2^{n-a-b}, \end{aligned}$$

From (3.23), if we assume $x = n - a - c$, $y = n - a - b$, the joint pmf of (X, Y) can be derived as $P(X = x, Y = y) = \frac{1}{Z(\lambda,\nu)} \sum_{n=0}^{\infty} \frac{\lambda^n}{(n!)^\nu} \times \mathcal{W}^*(x, y)$, and

$$\mathcal{W}^*(x,y) = \sum_{a=n-x-y}^{n} \binom{n}{a, n-a-y, n-a-x, x+y+a-n} p_{00}^a p_{10}^{n-a-y} p_{01}^{n-a-x} p_{11}^{x+y+a-n},$$

as showed in Sellers *et al.* (2016). Its moment generating function, factorial moment generating function, and cumulant generating function are given by

$$\begin{aligned} M(t_1, t_2) &= G\left(e^{t_1}, e^{t_2}\right) \\ &= \frac{Z\left[\lambda\{1 + p_{1+}(e^{t_1} - 1) + p_{+1}(e^{t_2} - 1) + p_{11}(e^{t_1} - 1)(e^{t_2} - 1)\}, \nu\right]}{Z(\lambda, \nu)}, \\ M^*(t_1, t_2) &= G(t_1 + 1, t_2 + 1) \\ &= \frac{Z\left[\lambda\{1 + p_{1+}t_1 + p_{+1}t_2 + p_{11}t_1t_2\}, \nu\right]}{Z(\lambda, \nu)}, \\ K(t_1, t_2) &= log M(t_1, t_2) \\ &= log\, Z\left[\lambda\{1 + p_{1+}(e^{t_1} - 1) + p_{+1}(e^{t_2} - 1) + p_{11}(e^{t_1} - 1)(e^{t_2} - 1)\}, \nu\right] - log\, Z(\lambda, \nu). \end{aligned}$$

Apply the knowledge of the bivariate Fisher index of dispersion, as defined in

(Minkova and Balakrishnan, 2014),

$$FI(X,Y) = \left\{\frac{Var(X)}{E(X)} + \frac{Var(Y)}{E(Y)} - 2Corr(XY)\frac{Cov(XY)}{\sqrt{E(Y)}\sqrt{E(Y)}}\right\}(1 - Corr(XY)^2)^{-1},$$

we have FI(X, Y) = 2 for bivariate Poisson. Then, when $\nu = 1$, the pgf of the CMP distribution can be written as

$$G\left(t_1, t_2\right) = \exp\{\lambda p_{1+}(t_1 - 1) + \lambda p_{+1}(t_2 - 1) + \lambda p_{11}(t_1 - 1)(t_2 - 1)\}\}.$$

Because $\nu \to \infty$, the CMP in this case is given by

$$G\left(t_1, t_2\right) = 1 + \tfrac{\lambda}{\lambda+1} p_{1+}(t_1 - 1) + \tfrac{\lambda}{\lambda+1} p_{+1}(t_2 - 1) + \tfrac{\lambda}{\lambda+1} p_{11}(t_1 - 1)(t_2 - 1).$$

When $\nu = 0$ and $\lambda < 1$, it reduces a geometric distribution with success probability $1 - \lambda$. In this case, we obtain

$$G\left(t_1, t_2\right) = \{1 - \frac{\lambda}{1-\lambda}\{p_{1+}(t_1 - 1) + p_{+1}(t_2 - 1) + p_{11}(t_1 - 1)(t_2 - 1)\}\}^{-1},$$

and $\frac{\lambda}{1-\lambda}\{p_{1+}(t_1-1)+p_{+1}(t_2-1)+p_{11}(t_1-1)(t_2-1)\} < 1$. Assume the parameters of the bivariate CMP's are λ^*, ν^*, p^*_{00}, p^*_{10}, p^*_{01} and p^*_{11}. The relations between bivariate CMP and bivariate Poisson are

$$p^*_{00} = 1 - p^*_{11} - p^*_{10} - p^*_{01}, \qquad p^*_{10} = \frac{\lambda_1}{\lambda^*}, \qquad p^*_{01} = \frac{\lambda_2}{\lambda^*}, \qquad p^*_{11} = \frac{\lambda_3}{\lambda^*}. \tag{3.25}$$

The parameter values for comparison of the bivariate CMP and bivariate Bernoulli are presented in Table B.1. The parameter values for comparison of the bivariate CMP and bivariate Geometric distribution and the proof of this result is also presented in Appendix B.

The approximations perform well for $\lambda > 10^{\nu}$ or $\nu \leq 1$. It is important to note that CMP belongs to the exponential family, which makes the estimation easier. There are few methods that can be used to estimate the parameters of the bivariate CMP distribution from a dataset. Although maximum likelihood estimation (MLE) is said to be more complex and computationally intensive for the CMP distribution, we apply it in this thesis due to its efficiency and asymptotic properties. Details of its application are discussed in subsequent chapters.

Chapter 4

Methodology

The method proposed in this thesis is the maximum likelihood estimation (MLE) for the four models discussed in Chapter 3, and their performance are then evaluated. If one is interested in selecting a suitable model for a given data, the selection of model can be carried out in three ways, by applying the likelihood ratio test (LRT), by using Akaike information criterion (AIC), or by using the Bayesian Information Criterion (BIC). The bivariate CMP distribution includes the three special cases of bivariate Poisson, bivariate Bernoulli, and bivariate Geometric distributions. In this work, we denote these three special cases by M_0, M_1 and M_2, respectively, and the bivariate case of CMP distribution by M_g.

4.1 Maximum Likelihood Estimation (MLE)

The Maximum likelihood estimation (MLE) is a well-known method to estimate the model parameters. The concept of MLE was first introduced by R.A. Fisher in 1920s and has since become the most prominent model-fitting method.

As stated in Myung (2003), let $L(\theta|\mathbf{K})$ be the likelihood function of a probability distribution, where the data vector $\mathbf{K} = (k_1, \ldots, k_n)$ is random sample from a population and the parameter vector θ is equal to $(\theta_1, \ldots, \theta_{n^*})$. The ML estimates are obatined by maximizing the log-likelihood function, $l(\theta|\mathbf{K})$, due to the monotonicity. If $l(\theta|\mathbf{K})$ is differentiable and provided that the ML estimates exist, the likelihood equation at $\theta_i = \hat{\theta}_i$ is given by $\partial\, l(\theta)/\partial_{\theta_i} = 0$ for $i = 1, \ldots, n^*$. Note here that $\hat{\theta}_i$ is the MLE of parameter θ_i. In addition, to ensure the log-likelihood function is convex and that θ_i is the maximum instead of minimum, we can check whether the second derivative of $l(\theta|\mathbf{K})$ is negative, or not.

However, an optimization algorithm may generate a local maximum instead of a global maximum depending on the choice of the starting values given for the algorithm. In this thesis, different initial values were chosen over multiple iterating processes to overcome this difficulty in the applications, discussed hereon.

4.2 MLEs for Bivariate Poisson (M_0)

To compute the MLE of the bivariate Poisson, let us recall the probability function $P(x, y)$ in (3.5) given by

$$\mathrm{P}(X = x, Y = y; \lambda_1, \lambda_2, \lambda_3) = \mathrm{e}^{-(\lambda_1+\lambda_2+\lambda_3)} \frac{\lambda_1^x}{x!} \frac{\lambda_2^y}{y!} \sum_{i=0}^{min(x,y)} \binom{x}{i} \binom{y}{i} i! \left(\frac{\lambda_3}{\lambda_1 \lambda_2} \right)^i .$$

We now apply the recurrence relations introduced in Teicher (1954) to simplify the calculation. It can be shown that

$$\begin{aligned} xP(x,y) &= \lambda_1 P(x-1,y) + \lambda_3 P(x-1,y-1), \\ yP(x,y) &= \lambda_2 P(x,y-1) + \lambda_3 P(x-1,y-1). \end{aligned} \tag{4.1}$$

Following the methods proposed in Holgate (1964), we take the probability mass function of bivariate Poisson distribution in (3.5) and differentiated with respect to λ_1, λ_2 and λ_3 to get

$$\begin{aligned} \frac{\partial P(x,y)}{\partial \lambda_1} &= P(x-1,y) - P(x,y) \\ \frac{\partial P(x,y)}{\partial \lambda_2} &= P(x,y-1) - P(x,y) \\ \frac{\partial P(x,y)}{\partial \lambda_3} &= P(x,y) - P(x,y-1) - P(x-1,y) + P(x-1,y-1) \end{aligned} \tag{4.2}$$

Combining these with the recurrence relations for Poisson distribution given in (4.1), the three likelihood equations $\sum \frac{1}{P}\frac{\partial P(x,y)}{\partial \lambda_1} = \sum \frac{1}{P}\frac{\partial P(x,y)}{\partial \lambda_2} = \sum \frac{1}{P}\frac{\partial P(x,y)}{\partial \lambda_3} = 0$ can be rewritten as follows:

$$\frac{\bar{x}}{\lambda_1} - \frac{\lambda_3}{\lambda_1}\,\overline{W}_1 - 1 = 0, \tag{4.3}$$

$$\frac{\bar{y}}{\lambda_2} - \frac{\lambda_2}{\lambda_2}\,\overline{W}_2 - 1 = 0, \tag{4.4}$$

$$\frac{\bar{x}}{\lambda_1} + \frac{\bar{y}}{\lambda_2} - (1 + \frac{\lambda_3}{\lambda_1} + \frac{\lambda_2}{\lambda_2})\,\overline{W}_3 - 1 = 0, \tag{4.5}$$

where

$$\overline{W}_1 = \frac{1}{n}\sum_{i=1}^{n} \frac{P(x_i - 1, y_i)}{P(x_i, y_i)}, \tag{4.6}$$

$$\overline{W}_2 = \frac{1}{n}\sum_{i=1}^{n} \frac{P(x_i, y_i - 1)}{P(x_i, y_i)}, \tag{4.7}$$

$$\text{and} \quad \overline{W}_3 = \frac{1}{n}\sum_{i=1}^{n} \frac{P(x_i - 1, y_i - 1)}{P(x_i, y_i)}. \tag{4.8}$$

Simplification of Eqs. (4.3)-(4.5) lead to $\overline{W}_1 = \overline{W}_2 = \overline{W}_3 = 1$, which implies that $\lambda_1 + \lambda_3 = \sum_{i=1}^{n} \frac{x_i}{n} = \bar{x}$ and $\lambda_2 + \lambda_3 = \sum_{i=1}^{n} \frac{y_i}{n} = \bar{y}$. Therefore, the MLEs $\hat{\lambda}_1$, $\hat{\lambda}_2$ and $\hat{\lambda}_3$ of the bivariate Poisson satisfies $\hat{\lambda}_1 + \hat{\lambda}_3 = \bar{x}$ and $\hat{\lambda}_2 + \hat{\lambda}_3 = \bar{y}$. The MLEs for individual parameters can then be estimated using an iterative process.

4.3 MLEs for Bivariate Bernoulli (M_1)

One of the most important properties as mentioned in the last Chapter is that the two-dimensional Bernoulli distribution possesses good properties similar to those of Gaussian distibution as discussed in Dai *et al.* (2013). Recall a 2-dimensional Bernoulli pair (X, Y) has its probability mass function as $p(x, y) = p_{11}^{xy} p_{10}^{x(1-y)} p_{01}^{(1-x)y} p_{00}^{(1-x)(1-y)} = exp\{log(p_{00}) + x\, N_1 + y\, N_2 + xy\, N_3\}$, as in (3.1).

In order to generate the MLE for the bivariate Bernoulli distribution, we treat bivariate Bernoulli as a special case of the multivariate Bernoulli distribution which is of the form

$$l(x, y, \mathbf{N}) = -ln\{p(x, y)\} = -\Big\{ \sum_{x,y} \Big(\sum_{x \leq y} N^{x\,y} \mathbb{C}^{(x\,y)} \Big) - b(\mathbf{N}) \Big\}, \tag{4.9}$$

where the natural parameter is $\mathbf{N} = (N_1, N_2, N_3)^T$, the interaction term $\mathbb{C}^{(x\,y)} = xy$, and the normalizing factor is $b(\mathbf{N})$. The normalizaing factor, also known as the log partition function, is given by

$$b(\mathbf{N}) = ln\big\{1 + \sum_{x \leq y} exp(N_1 + N_2 + N_3)\big\} \tag{4.10}$$

This is a term to ensure the disrtribution is normalized properly. Since the bivariate Bernoulli distribution is a member of the exponential family, there are relations between the natural and general parameters. Finally, the log determinant relaxation of the log partition function can be used to calculate the MLE of the bivariate binary data. For details, see Banerjee *et al.* (2008) and Wainwright and Jordan (2006).

4.4 MLEs for Bivariate Geometric (M_2)

Then, we can write the likelihood function of the bivariate Geometirc according to Eq. (3.13) which is of the form

$$\begin{aligned} L(x, y\,;\, p_1, p_2, p_3) = {} & p_1^a p_2^b p_3^e (1 - p_1)^d (1 - p_2)^c (1 - p_1 p_3)^c \\ & \times (1 - p_2 p_3)^d (1 - p_1 p_3 - p_2 p_3 + p_1 p_2 p_3)^g, \end{aligned} \tag{4.11}$$

where the letter a, b, c, d, e and g satisfied the following conditions

$$a = \sum_{i=1}^{n}(x_i - n), \qquad b = \sum_{i=1}^{n}(y_i - n),$$

$$c = \sum_{i=1}^{n} I[y < x], \qquad d = \sum_{i=1}^{n} I[x < y],$$

$$e = \sum_{i=1}^{n} \big\{(y-1)I[x < y] + \sum_{i=1}^{n}(x-1)I[x \le y]\big\},$$

$$g = \sum_{i=1}^{n} I[x = y].$$

Once the formula for likelihood has been defined, the logarithm, $l(x, y; p_1, p_2, p_3)$, can be passed on to find the maximum likelihood estimators $\hat{p}_1$, $\hat{p}_2$ and $\hat{p}_3$. It can be achieved simply by solving the following score functions (Li and Dhar, 2013), where

$$\frac{\partial l}{\partial p_1} = \frac{a}{p_1} - \frac{d}{1-p_1} - \frac{cp_3}{1-p_1p_3} + \frac{g(-p_3 + p_2p_3)}{1 - p_1p_3 - p_2p_3 + p_1p_2p_3} = 0, \tag{4.12}$$

$$\frac{\partial l}{\partial p_2} = \frac{b}{p_2} - \frac{c}{1-p_2} - \frac{dp_3}{1-p_2p_3} + \frac{g(-p_3 + p_1p_3)}{1 - p_1p_3 - p_2p_3 + p_1p_2p_3} = 0, \tag{4.13}$$

$$\frac{\partial l}{\partial p_3} = \frac{c}{p_3} - \frac{cp_1}{1-p_1p_3} - \frac{dp_2}{1-p_2p_3} + \frac{g(-p_1 - p_2 + p_3)}{1 - p_1p_3 - p_2p_3 + p_1p_2p_3} = 0. \tag{4.14}$$

Due to the complex calculation involved, Eqs.(4.12) - (4.14) are not easy to solve. Fortunately, when dataset (X, Y) is given, the values of $a, b, c, d, e,$ and g can be calculated directly since they are associated with the dataset only. Given values of a to g have been calculated, we can obtain an explicit form of the equations which make estimation easier. Thus obtained the MLE's for parameters p_1, p_2 and p_3 can be achieved through iterating process. The illustrative examples based on simulated datasets for this method is given in Chapter 5.

4.5 MLEs for CMP (M_g)

There are few approaches to determine the MLE for the CMP distribution as suggested in Sellers and Shmueli (2010). In this thesis, we consider the method stated in Sellers *et al.* (2016) which uses bivariate Poisson distribution while the number of trials is modeld via $CMP(\lambda, \nu)$. Then according the definition of MLE, we maximize the log-likelihood function which is given by $l(\lambda, \nu, \mathbf{p}) = \sum_x \sum_y n_{xy}\, lnP(X = x, Y = y)$. The MLE of parameters, λ, , ν, and $\mathbf{p}$, can be estimated using an iterative scheme with deffient starting values to aviod the local maximum problem discussed before. More details are explained in Shmueli *et al.* (2005).

4.6 Model Discrimination

There are a couple of criteria that have been used in this thesis for the evaluation of parameters, the maximum likelihood estimation (MLE) and model selection. The first model selection statistic that has been considered is the likelihood ratio test (LRT) (King, 1998), followed by Akaike information criterion (Akaike, 1974) and Bayesian information criterion (Schwarz, 1978).

4.6.1 Likelihood-Based Method

The idea of Likelihood Ratio Test (LRT) is to compare the unrestricted model and the restricted model provided the simpler model (unrestricted model) is a special case of the restricted model. It also called comparison of two nested models. In terms of hypothesis testing, the null hypothesis is defined as the test parameters are all zero which means the restricted model is the correct model to choose. The alternative hypothesis is defined as the test parameters are not zero which means the unrestricted model is correct. In order to conduct the LRT, the likelihoods of both models are needed to be estimated.

In simulation and illustrative section, we apply the LRT to investigate the null hypothesis that is the distribution can be defined by one of bivariate Poisson($H_0 : \nu = 1$), bivariate Bernoulli ($H_0 : \nu \to \infty$) and bivariate Geometric ($H_0 : \nu \to 0$) distribution versus the alternative hypothesis the distribution can be described by a bivariate CMP distribution other than one of the null hypothesis.

Then let us assume the notation of the maximum likelihood of the unrestricted model and restricted model are $\hat{L}_u$ and $\hat{L}_r$ which maximized log-likelihood function values, respectively. The likelihood ratio test statistic is defined as a function of the ratrio of the two likelihoods:

$$LRT = -2ln\left(\frac{\hat{L}_r}{\hat{L}_u}\right) = 2(ln(\hat{L}_u) - ln(\hat{L}_r)) \sim \chi^2(g), \tag{4.15}$$

where g is the number of test parameters. LRT is always non-negative since $\hat{L}_u \geq \hat{L}_r$. If the random error is considered to be zero while LRT is equal to zero, then the restricted model performs better than the unrestricted model. If take the random error into account, a much larger observed value of LRT than the expected value of g indicates the difference between the two models is not only due to random differences. In this case, one should select the unrestricted model. More details of LRT can be found in King (1998) and Buse (1982). Moreover, the standard null hypothesis is a central chi-square distribution. However, if the parameters are valued on the boundary of the parameter space, such as $H_0 : \phi \to \infty$ and $H_0 : \phi \to 0$, we consider the asymptotic null distribution when conducting the LRT.

4.6.2 Information-Based Criterion

As we know that the increase of the likelihood can be achieved by increasing the parameters. In order to avoid overfitting, the Akaike information criterion (AIC, Akaike (1974)) and Bayesian information criterion (BIC, Schwarz (1978)) are introduced to resolve this problem by including a term to penalize free parameters. The AIC is defined as

$$AIC(\hat{\theta}) = -2\, ln(\hat{L}) + 2k, \tag{4.16}$$

where $\hat{L}$ represents the value of the maximum log-likelihood of the model and k is the number of parameters used in the model. It evaluates the performance of the models by the goodness of fits while penalizes the increases of the parameters. The model gives the minimum AIC value is the best model that should be selected.

The Bayesian information criterion (BIC) has been widely used as a criterion for model selection. It is an alternative to AIC and the criterion is given by

$$BIC = -2\, ln(\hat{L}) + k\, ln(n), \tag{4.17}$$

where $\hat{L}$ represents the value of the maximum log-likelihood for the estimated model, k is the number of free parameters to be estimated and n is the number of observations. The model with the lowest value of BIC is the one preferred. The increase the variation in the dependent variable and the number of explanatory variables increase the value of BIC. The lower value of BIC indicates the model has few variables, generate a better fit of the data, or both. Again, although the BIC criterion depends on the data size n and k, it is more strict with free parameters than AIC. Also, unlike LRT, BIC does not require the test models to be nested.

Chapter 5

Simulation and Illustrative Examples

5.1 Simulation study

In the simulation study, we investigate the performance of AIC and BIC in selecting one of the M_0, M_1 and M_2 distribution given the true distribution is one of the three distributions. With specified parameters, we conduct three simulated studies where each case has data generated from one of M_0, M_1 and M_2. Then the AIC and BIC values have been calculated after fitted bivariate Poisson, bivariate Bernoulli, and bivariate Geometric distribution.

Moreover, to compare the performance of bivariate Poisson and bivariate CMP under data dispersion, we calculated AIC and BIC values of the fitted bivariate CMP distribution to the three simulated datasets as well.

5.1.1 Case 1

In this section, we generated a dataset with 500 data pairs that follow the bivariate Poisson distribution (M_0 model) using the rpois function in Stats package in R version 1.1.456. Applying the method of trivariate reduction discussed in Section 3.1.2, we set $P_i \sim Poisson(\lambda_i)$, where the parameter $\lambda_i = 1, 4, 5$, $X = P_1 + P_3$ and $Y = P_2 + P_3$. The range observed for X and Y are 1 to 14 and 2 to 18. The mean and variance for X are 5.8 and 5.2, while the mean and variance for Y are 8.7 and 8.0. The empirical dispersions are 0.9 and 0.92 for X and Y, respectively.

Table 5.1: The MLEs of bivariate Poisson and bivariate CMP models on the simulated bivariate Poisson dataset (500 pairs)

	M_0	M_g
$\hat{\lambda}_1$	1.044	-
$\hat{\lambda}_2$	4.00	-
$\hat{\lambda}_3$	4.89	-
$\hat{\lambda}$	-	9.046
$\hat{\nu}$	-	**0.956**
$\hat{p}_{00}$	-	0.007
$\hat{p}_{10}$	-	0.106
$\hat{p}_{01}$	-	0.401
$\hat{p}_{11}$	-	0.486

In Table 5.1, the generated MLEs of M_0 are $\hat{\lambda}_1 = 1.044$, $\hat{\lambda}_2 = 4.00$, and $\hat{\lambda}_3 = 4.89$ which is close to the true values of parameters. The parameters $\hat{\lambda}$ and $\hat{\nu}$ from CMP distribution are 9.046 and 0.956, respectively. The $\hat{\nu} = 9.046$ indicates the simulated data are nearly equi-dispersion. The other four MLEs of M_g model are $\hat{p}_{00} = 0.007$, $\hat{p}_{10} = 0.106$, $\hat{p}_{01} = 0.401$, and $\hat{p}_{11} = 0.486$. According to the formulas derived in Section 3.2, we can calculate CMP estimates, while assuming $\nu \cong 1$, which are $\hat{\lambda}_1 = \hat{p}_{10} \times \hat{\lambda} = 0.959$, $\hat{\lambda}_2 = \hat{p}_{01} \times \hat{\lambda} = 3.627$, and $\hat{\lambda}_3 = \hat{p}_{11} \times \hat{\lambda} = 4.993$. It is clear that the estimates of the bivariate Poisson distribution are close to the true values under M_0.

Then, we fit the dataset to M_0, M_1, M_2 and M_g and resulting values of log-likelihood and AIC are presented in Table 5.2. The log-likelihood of M_0 and M_g are similar. However, according to the AIC and BIC formulas in Section 3.2, AIC, and BIC of M_0 are lower than those of M_g due to lower penalty based on the number of parameters. The likelihood ratio test confirms that the data are from M_0 model since $LRT = -2ln(L^*) = 0.40$ with a p-value of 0.53, where $L^* = \frac{\hat{L}_r}{\hat{L}_u}$.

Table 5.2: The Log-likelihood, AIC and BIC of bivariate Poisson(M_0), bivariate Bernoulli (M_1), bivariate Geometric (M_2) and bivariate CMP (M_g) models on the simulated bivariate Poisson dataset (500 pairs)

	M_0	M_1	M_2	M_g
LnL	**-2259.120**	-2833.026	-2828.070	**-2258.919**
AIC	**4524.240**	5674.052	5662.130	4525.838
BIC	**4536.884**	5641.194	5674.774	4542.696

5.1.2 Case 2

In the second case of our simulation study, we used rbinom2.or in package VGAM in R to generate 500 data pairs from a bivariate Bernoulli model (M_1). We set the marginal probability for the first component to be 0.6, odds ratio equal to $exp(1.5)$ and the two marginal probabilities are constrained to be equal. The dataset has zeros and ones with means of X and Y as 0.63 and 0.64, variances as 0.23 and 0.24, and the empirical dispersions are 0.37 and 0.40, respectively. It is equivalent to generating a dataset with probabilities $p_{00} = 0.244$, $p_{01} = 0.156$, $p_{10} = 0.156$, and $p_{11} = 0.444$.

Table 5.3: The MLEs of bivariate Poisson (M_0) and bivariate CMP (M_g) models on the simulated bivariate Bernoulli dataset (500 pairs)

	M_0	M_g
$\hat{\lambda}_1$	0.158	-
$\hat{\lambda}_2$	0.202	-
$\hat{\lambda}_3$	0.410	-
$\hat{\lambda}$	-	6.000
$\hat{\nu}$	-	**30.000**
$\hat{p}_{00}$	-	0.153
$\hat{p}_{10}$	-	0.306
$\hat{p}_{01}$	-	0.268
$\hat{p}_{11}$	-	0.273

In this case, as shown in Table 5.3, the estimates $\hat{\lambda}$ and $\hat{\nu}$ of M_g are 6 and 30 with $\hat{p}_{00} = 0.153$, $\hat{p}_{10} = 0.306$, $\hat{p}_{01} = 0.268$, and $\hat{p}_{11} = 0.273$. The under dispersion in the data is naturally detected by the estimate $\hat{\nu} = 30$. According to the relation between M_g and M_1, the estimated probabilities of the bivariate Bernoulli, as a special case of the CMP distribution, are given by

$$\hat{p}^*_{00} = 1 - \frac{\hat{\lambda}}{\hat{\lambda}+1}\left(\hat{p}_{01} + \hat{p}_{10} + \hat{p}_{11}\right) \cong 0.232, \tag{5.1}$$

$$\hat{p}^*_{01} = \frac{\hat{\lambda}}{\hat{\lambda}+1}\hat{p}_{01} \cong 0.133, \tag{5.2}$$

$$\hat{p}^*_{10} = \frac{\hat{\lambda}}{\hat{\lambda}+1}\hat{p}_{10} \cong 0.163, \tag{5.3}$$

$$\hat{p}^*_{11} = \frac{\hat{\lambda}}{\hat{\lambda}+1}\hat{p}_{11} \cong 0.469. \tag{5.4}$$

Compare results from (5.1) - (5.4) with true probabilities of the data. We noticed that p_{00} and p_{01} are slightly underestimated by the second decimal place and p_{10} and p_{11} are slightly overestimated . In general, the overall estimation is quite accurate since $1 - \mu = 0.4$, and $\hat{p}^*_{00} + \hat{p}^*_{10} = 0.394$.

Table 5.4: The Log-likelihood, AIC and BIC of bivariate Poisson (M_0), bivariate Bernoulli (M_1), and bivariate CMP (M_g) models on the simulated bivariate Bernoulli dataset (500 pairs)

	M_0	M_1	M_g
LnL	-831.110	**-629.819**	**-629.819**
AIC	1668.220	**1265.638**	1267.638
BIC	1680.864	**1278.282**	1284.496

The M_1 generates the lowest AIC and BIC values among M_0, M_1, and M_g. In Table 5.4, the difference in AIC values are small for M_g and M_1 while M_1's value is slightly lower. We performed likelihood ratio test between M_0 and M_g with $-2ln(L^*) \cong 403$ and a p-value around 0, which indicates M_0 is not the correct model and the existence of significant data dispersion. Again, the LRT between M_g and M_1 confirms that the simulated data is from M_1 model since $-2ln(L^*) \cong 0$ with p-value greater than significance level. Thus, the bivariate CMP model detects the under-dispersion in the data, and provides a better fit.

5.1.3 Case 3

For the third case, we used rbivgeo2 in package BivGeo in R to generate 500 data pairs from a bivariate Geometric model (M_2). We set failure probabilities equal to 0.9. The ranges for X and Y are 1 to 21 and 1 to 20, with means 3.12 to 3.73, and variances 7.00 and 10.25, respectively. The empirical dispersions are 2.24 and 2.75. Applying the method as in pervious cases, as shown in Table 5.5, the MLEs for M_0 model are $\hat{\lambda}_1 = 1.92$, $\hat{\lambda}_2 = 2.53$, $\hat{\lambda}_3 = 1.2$, the MLE for M_g model are $\hat{\lambda} = 0.942$, $\hat{\nu} = 0.000$, $\hat{p}_{00} = 0.389$, $\hat{p}_{01} = 0.339$, $\hat{p}_{10} = 0.270$, and $\hat{p}_{11} = 0.002$. Dispersion parameter here is smaller than 1, which implies that the data are over-dispersion.

Table 5.5: The MLEs of bivariate Poisson (M_0) and bivariate CMP (M_g) models on the simulated bivariate Geometric dataset (500 pairs)

	M_0	M_g
$\hat{\lambda}_1$	1.920	-
$\hat{\lambda}_2$	2.530	-
$\hat{\lambda}_3$	1.200	-
$\hat{\lambda}$	-	0.942
$\hat{\nu}$	-	**0.000**
$\hat{p}_{00}$	-	0.389
$\hat{p}_{10}$	-	0.339
$\hat{p}_{01}$	-	0.270
$\hat{p}_{11}$	-	0.002

Model M_2 generates the lowest AIC and BIC values among M_0, M_2, and M_g. From Table 5.6, the log-likelihood value for M_g ($\cong -2583.909$) is significantly larger than the log-likelihood value for M_0 ($\cong -3142.720$). It can also be shown by LRT where $-2ln(L^*) = 1117.622$ with p-value significantly small ($\cong 0$). It implies the existence of data dispersion and M_0 is not the correct model to use. Moreover, M_g generated the smaller AIC and BIC value than M_0 model in this case.

Again, the LRT between M_g and M_1 confirms that the simulated data is from M_1 model since $-2ln(L^*) \cong 1.242$ with a p-value of 0.265. Here again, the bivariate CMP model outperforms the bivariate Poisson model by considering the dispersion in the data.

Table 5.6: The Log-likelihood, AIC and BIC of bivariate Poisson (M_0), bivariate Geometric (M_2), and bivariate CMP (M_g) models on the simulated bivariate Geometric dataset (500 pairs)

	M_0	M_2	M_g
LnL	-3142.720	-2584.530	**-2583.909**
AIC	6291.442	**5175.060**	5175.818
BIC	6304.255	**5187.704**	5192.676

From the previous three cases with different dispersion of the datasets, we observe that the bivariate CMP always outperforms the bivariate Poisson model if data dispersion is presented. It is agreed with the theoretical derivation as in Chapter 4.

5.1.4 Illustrative real data analysis

In this section, we compare and contrast the performance of the bivariate Poisson model and the CMP model using the data introduced in Chapter 2. According to Bermúdez and Karlis (2017), the bivariate Poisson model is approved to perform well with the auto insurance data. In this section, we fit the data using the CMP model while monitoring data dispersion using the dispersion parameter ν.

Recall X and Y are annual number of claims for N_1 type and N_2 type, respectively. We observed values of X and Y with ranges of 1-6 and 1-6, means of 0.085 and 0.125, variances of 0.130 and 0.178, and empirical dispersions are 1.5 and 1.4. Fitting M_0 model produces parameters $\hat{\lambda}_1 = 0.04$, $\hat{\lambda}_2 = 0.5$ and $\hat{\lambda}_3 = 1.13$. The MLEs for the M_g model are $\hat{\lambda} = 0.112$, $\hat{\nu} = 0.000$, $\hat{p}_{00} = 0.011$, $\hat{p}_{01} = 0.315$, $\hat{p}_{10} = 0.001$, and $\hat{p}_{11} = 0.672$, as shown in Table 5.7. Here, the dispersion parameter is equal to zero, which indicates that the data are over-dispersed.

Table 5.7: The MLEs of bivariate Poisson (M_0) and bivariate CMP (M_g) models on the automobile insurance dataset

	M_0	M_g
$\hat{\lambda}_1$	0.040	-
$\hat{\lambda}_2$	0.500	-
$\hat{\lambda}_3$	1.130	-
$\hat{\lambda}$	-	0.112
$\hat{\nu}$	-	**0.000**
$\hat{p}_{00}$	-	0.011
$\hat{p}_{10}$	-	0.001
$\hat{p}_{01}$	-	0.315
$\hat{p}_{11}$	-	0.672

Table 5.8: The Log-likelihood, AIC and BIC of bivariate Poisson (M_0), and bivariate CMP (M_g) models on the automobile insurance dataset

	bivariate Poisson (M_0)	bivariate CMP (M_g)
LnL	-13690.590	**-13303.180**
AIC	27387.180	**26614.360**
BIC	27411.867	**26647.403**

From Table 5.8, the log-likelihood for M_g is -13303.180, which is larger than the value for M_0. When the M_g model is fitted to the data, it produces lower AIC value than when the M_0 model is fitted ($AIC_{M_g} = 26614 < AIC_{M_0} = 27387$). So, M_g outperforms M_0 based on AIC. We then compare the BIC values of the two models. The BIC for M_0 and M_g are 27412 and 26617, respectively. Based on these BIC values, we find that the CMP model fit the data better than the bivariate Poisson model.

When we apply the LRT to M_0 and M_g, we also notice that $-2ln(L^*) \cong 774.820$ with a p-value of 1.610×10^{-170}. Hence, we reject the null hypothesis of the bivariate Poisson. Since the data are over-dispersed, we tried to fit it using a negative bivariate binomial model which is a more general form of the bivariate geometric distribution, and it has log-likelihood value, AIC, and BIC of -13365.810, 26762.400, 26762.400. Although its values show the model performs better than bivariate Poisson, the bivariate CMP still provides the best model among these.

These examples demonstrate that the bivariate CMP model outperforms the bivariate Poisson for modeling the auto insurance dataset given by Bermúdez and Karlis (2017) due to the presence of dispersion in the data.

Chapter 6

Conclusions and Future Work

In this thesis, we have demonstrated the usefulness of the CMP model as a more general case of the bivariate Poisson distribution. It closely monitors the over and under dispersion for bivariate count data which allows the analysis to be more accurate. As a more general model, the CMP captures bivariate Poisson, bivariate Bernoulli and bivariate geometric as three special cases, and can be used as a tool for analyzing other structure using its dispersion parameter. We have applied the theoretical framework to the simulation study and real data from an insurance company in Chapter 5. The results fully demonstrated bivariate CMP is a powerful overarching model with lower AIC, BIC values than bivariate Poisson model with and without data dispersion. Along with model discrimination, we also performed parameter estimation and hypothesis testing for simulated and real datasets.

As a powerful model, the bivariate CMP distribution has ensured a more general model fitting results. However, as a complex distribution, it does not have a closed-form. Because of the complexity of its calculation, the difficulty of applying it raises.

In this work, we have used the *R* package from Sellers *et al.* (2018) where it uses the finite sum to approximate the distribution. Although it is a relatively reliable approach, more research has to be done to reveal the impact of infinite summation to the iterating process when generating the bivariate CMP model. The goodness of fit tests and Bayesian inference can also be performed. The extension of bivariate CMP to multivariate CMP is another topic can be focused on. Its accuracy of modeling can be investigated if the dataset has been grouped into more than two categories.

Appendix A

R code

```
#Case 1.
```{r}
#set up bivairate Poisson data with lamb= 1,4,5
Pois1<-rpois(500,1)
Pois2<-rpois(500,4)
Pois3<-rpois(500,5)

#set up the X and Y to be bivariate distribution
X<-Pois1+Pois3
Y<-Pois2+Pois3
XY <- matrix(0,length(X),2)
XY[,1] <- X
XY[,2] <- Y
```
```

```
#range, mean, variance, and empeirical dispersion of X and Y
c(range(X),range(Y),mean(X),mean(Y) ,var(X) ,var(Y))

#start computation;
#form BP
form.BP01 <- function(l1,l2,l3){
-sum(dbvpois(XY,a=l1,b=l2,c=l3,log = TRUE))
}

#MLE
BP01 <- mle2(form.BP01,start=list(l1=2,l2=3,l3=5),data=data.frame(XY))
summary(BP01)
c(AIC(BP01),BIC(BP01))

#form BB
Blog<-function(d){
lex <- paste(d[,1],d[,2],sep = "")
counts <- table(lex)

poo_hat <- counts[1]/sum(counts)
p1o_hat <- counts[2]/sum(counts)
po1_hat <- counts[3]/sum(counts)
p11_hat <- counts[4]/sum(counts)
```

```
props <- c(poo_hat, p1o_hat, po1_hat, p11_hat)

L <- sum(counts*log(props))
return(L)
}
#MLE BB
Blog(XY)
(-2)*Blog(XY)+2*4 #AIC.BB
(-2)*Blog(XY)+2*log(500)#BIC.BB

#log pmg of BG
dbg<-function (x, p = c()) {
x<-as.matrix(x)
x0 <- x[, 1]
y0 <- x[, 2]
inter1 <- p[1] * p[3]
inter2 <- p[2] * p[3]
inter3 <- p[1] * p[2] * p[3]
#x<y
ter1 <- p[1]^(x0 - 1) * (inter2)^(y0 - 1)
ter2 <- (1 - inter2) * (1 - p[1])
for1 <- ter1 * ter2
#x=y
```

```
ter3 <- (inter3)^(x0 - 1)
ter4 <- (1 - inter1 - inter2 + inter3)
for2 <- ter3 * ter4
#x>y
ter5 <- p[2]^(y0 - 1) * (inter1)^(x0 - 1)
ter6 <- (1 - inter1) * (1 - p[2])
for3 <- ter5 * ter6
pmf.BG <- ifelse(x0 < y0, for1,
ifelse(x0 > y0, for3, for2))
return(log(pmf.BG))
}

#form BG
form.BG01 <- function(t1,t2,t3){
f<-dbg(XY,p=c(t1,t2,t3))
return(-sum(as.numeric(f)))
}

#MLE BG
BG01 <- mle2(form.BG01,start=list(t1=0.5,t2=0.5,t3=0.7),data=data.frame(
summary(BG01)
c(AIC(BG01),BIC(BG01))

#MLE CMP
```

```
COM.MLE.BP<-multicmpests(XY, startvalues = c(9, 1, 0, 0.1, 0.4, 0.5))

#check if the lambda calculated base on parameters generated from COM m
est.l1<-COM.MLE.BP$par[1]*COM.MLE.BP$par[5]
est.l2<-COM.MLE.BP$par[1]*COM.MLE.BP$par[4]
est.l3<-COM.MLE.BP$par[1]*COM.MLE.BP$par[6]

(2)*(COM.MLE.BP$negll)+2*6 #AIC CMP
(2)*(COM.MLE.BP$negll)+6*log(500) #BIC CMP



##Case 2.
#generate BB data
BB.data<-rbinom2.or(500,mu1=0.6,oratio=exp(1.5),exchangeable = TRUE)

#check initial values of p[ij]
dbinom2.or(mu1=0.6,exchangeable = TRUE, oratio = exp(1.5))
BB.data<-as.matrix(BB.data0)

range(BB.data[,1])
range(BB.data[,2])
mean(BB.data[,1])
mean(BB.data[,2])
var(BB.data[,1])
```

```
var(BB.data[,2])

#start computation;
#form BP
BP2 <- function(l1,l2,l3){
-sum(dbvpois(BB.data,a=l1,b=l2,c=l3,log = TRUE))
}
#MLE BP
BP02 <- mle2(BP2,start=list(l1=0.1,l2=0.03,l3=0.3),data=data.frame(BB.da
summary(BP02)
AIC(BP02)
BIC(BP02)

#MLE BB
Blog(BB.data)
(-2)*Blog(BB.data)+2*4 #AIC.BB
(-2)*Blog(BB.data)+2*log(500)#BIC.BB

#form BG
BG2 <- function(t1,t2,t3){
f<-dbg(BB.data,p=c(t1,t2,t3))
return(-sum(as.numeric(f)))
}
#MLE BG
```

```
BG02 <- mle2(BG2,start=list(t1=0.5,t2=0.5,t3=0.5),data=data.frame(BB.dat
summary(BG02)
c(AIC(BG02),BIC(BG02))

#MLE CMP
COM.MLE.BB<-multicmpests(BB.data, startvalues = c(6, 30, 0.2, 0.1, 0.1,
(2)*(COM.MLE.BB$negll)+2*6 #AIC
(2)*(COM.MLE.BB$negll)+6*log(500) #BIC




##Case 3.
#data
G<-rbivgeo1(500,theta=c(0.84,0.912,0.814))
G<-as.matrix(G)
c(range(G[,1]),range(G[,2]),mean(G[,1]),mean(G[,2]),var(G[,1]),var(G[,2]

BP3 <- function(l1,l2,l3){
-sum(dbvpois(G,a=l1,b=l2,c=l3,log = TRUE))
}

BP03 <- mle2(BP3,start=list(l1=10,l2=10,l3=1),data=data.frame(G),method
#-3488.5
AIC(BP03)
```

```
BIC(BP03)



BG3 <- function(t1, t2, t3){
f<-dbg(G, p=c(t1, t2, t3))
return(-sum(as.numeric(f)))
}

#MLE
BG03 <- mle2(BG3, start=list(t1=0.9, t2=0.9, t3=0.9), data=data.frame(G), met
summary(BG03)
AIC(BG03)
BIC(BG03)

#MLE CMP
COM.MLE.BG<-multicmpests(G, startvalues = c(1, 0, 0, 0.4, 0.4, 0.2))
(2)*(COM.MLE.BG$negll)+2*6
(2)*(COM.MLE.BG$negll)+6*log(20)



##Real Data.
#data
x0<-c(24408,1916,296,69,12,6,0)
x1<-c(1068, 317, 61 ,21 ,6, 2 ,2)
```

```
x2<-c(203, 71, 18, 6 ,2, 1, 1)
x3<-c(49,14,8,3,3,1,0)
x4<-c(11, 6, 2, 0, 1, 0,0,0)
x5<-c(2 ,0 ,0, 0 ,0, 0, 1)
x6<-c(1,0,0,1,0,0,0)
x8<-c(0,0,1,0,0,0,0)
DR<-rbind(x0,x1,x2,x3,x4,x5,x6,x8)

#N1:single row sum
fr<-function(x){
sum(DR[x,1:7])
}
sum.1<-rbind(fr(1),fr(2),fr(3),fr(4),fr(5),fr(6),fr(7),fr(8))
#N2: signle col sum
fc<-function(x){
sum(DR[1:8,x])
}
sum.2<-cbind(fc(1),fc(2),fc(3),fc(4),fc(5),fc(6),fc(7))
#check if N1=N2?
sum(sum.1)-sum(sum.2)
#total observations: 28590

#func to generating N1 vector
N1.f<-function(x){
```

```
rep(x,times=sum.1[x+1])
}

#func to generating N2 vector
N2.f<-function(y){
rep(y,times=sum.2[y+1])
}

N1<-c(N1.f(0),N1.f(1),N1.f(2),N1.f(3),N1.f(4),N1.f(5),N1.f(6),rep(8,time
N2<-c(N2.f(0),N2.f(1),N2.f(2),N2.f(3),N2.f(4),N2.f(5),N2.f(6))

DR.bi<-cbind(N1,N2)
#dim(DR.bi)

as.matrix(DR.bi)
#mean, variance and dispersion of N1 and N2
c(mean(DR.bi[,1]),mean(DR.bi[,2]) ,var(DR.bi[,1]) ,var(DR.bi[,2]))

#start computations;
#form BP
f.br <- function(l1,l2,l3){
-sum(dbg(DR.bi,a=l1,b=l2,c=l3)
}
#MLE
```

```
BP.r <- mle2(f.br,start=list(l1=0.09,l2=0.11,l3=0.01),data=data.frame(DR
summary(BP.r)



#MLE CMP
COM.dr<-multicmpests(DR.bi, startvalues = c(1, 1, 0.25, 0.25, 0.25, 0.25
D1<-(2)*(COM.dr$negll)+2*6  #AIC
D2<-(2)*(COM.dr$negll)+6*log(28590)  #BIC

c(AIC(BP.r),BIC(BP.r),D1,D2)
```

Appendix B

Derivation of the three special cases via pgf

Similar as Section 4.5, we consider the method stated in Sellers *et al.* (2016) which use bivariate Poisson distribution while the number of trials is modelled via $CMP(\lambda, \nu)$. Applying the compounding method which enables us to rewrite the joint pgf of the CMP distribution along with (3.7) as

$$
\begin{aligned}
G\left(t_1, t_2\right) &= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!^{\nu} Z(\lambda, \nu)} G(t_1, t_2|n) \\
&= \sum_{n=0}^{\infty} \frac{\lambda^n \{1 + p_{1+}(t_1 - 1) + p_{+1}(t_2 - 1) + p_{11}(t_1 - 1)(t_2 - 1)\}^n}{n!^{\nu} Z(\lambda, \nu)} \\
&= \frac{Z\left[\lambda\{1 + p_{1+}(t_1 - 1) + p_{+1}(t_2 - 1) + p_{11}(t_1 - 1)(t_2 - 1)\}, \nu\right]}{Z(\lambda, \nu)} \qquad \text{(B.1)}
\end{aligned}
$$

Its moment generating function is

$$
\begin{aligned}
M\left(t_1, t_2\right) &= G\left(e^{t_1}, e^{t_2}\right) \\
&= \frac{Z\left[\lambda\{1 + p_{1+}(e^{t_1} - 1) + p_{+1}(e^{t_2} - 1) + p_{11}(e^{t_1} - 1)(e^{t_2} - 1)\}, \nu\right]}{Z(\lambda, \nu)}
\end{aligned} \tag{B.2}
$$

The factorial moment generating function is

$$
\begin{aligned}
M^*\left(t_1, t_2\right) &= G\left(t_1 + 1, t_2 + 1\right) \\
&= \frac{Z\left[\lambda\{1 + p_{1+}(t_1 - 1 + 1) + p_{+1}(t_2 - 1 + 1) + p_{11}(t_1 - 1 + 1)(t_2 - 1 + 1)\}, \nu\right]}{Z(\lambda, \nu)}
\end{aligned} \tag{B.3}
$$

$$
= \frac{Z\left[\lambda\{1 + p_{1+}t_1 + p_{+1}t_2 + p_{11}t_1t_2\}, \nu\right]}{Z(\lambda, \nu)} \tag{B.4}
$$

The cumulant generating function is

$$
\begin{aligned}
K\left(t_1, t_2\right) &= log M\left(t_1, t_2\right) \\
&= log\ Z\left[\lambda\{1 + p_{1+}(e^{t_1} - 1) + p_{+1}(e^{t_2} - 1) + p_{11}(e^{t_1} - 1)(e^{t_2} - 1)\}, \nu\right] - log\ Z(\lambda, \nu)
\end{aligned} \tag{B.5}
$$

The general case of marginal pgfs for X and Y are

$$
G_X\left(t\right) = \frac{Z\{\lambda p_{1+}(t - 1), \nu\}}{Z(\lambda, \nu)} \tag{B.6}
$$

$$
G_Y\left(t\right) = \frac{Z\{\lambda p_{+1}(t - 1), \nu\}}{Z(\lambda, \nu)} \tag{B.7}
$$

For the Poisson case, set $\nu = 1$, then marginal pgfs for X and Y are

$$G_X^*\left(t;\nu=1\right) = \exp\left\{\lambda p_{1+}(t-1)\right\} \sim Poisson(\lambda p_{+1}) \tag{B.8}$$

$$G_Y^*\left(t;\nu=1\right) = \exp\left\{\lambda p_{+1}(t-1)\right\} \sim Poisson(\lambda p_{1+}) \tag{B.9}$$

When $\nu = 1$ pgf of (X, Y) becomes (3.8)

$$\begin{aligned} G\left(t_1,t_2\right) &= \exp\{\lambda[1+p_{1+}(t_1-1)+p_{+1}(t_2-1)+p_{11}(t_1-1)(t_2-1)]-\lambda\} \quad \text{(B.10)} \\ &= \exp\{\lambda p_{1+}(t_1-1)+\lambda p_{+1}(t_2-1)+\lambda p_{11}(t_1-1)(t_2-1)\}\} \end{aligned}$$

and the relations between parameters of trivariate reduction method and the parameters from CMP are

$$\begin{aligned} \lambda_1+\lambda_3 &= \lambda p_{1+} \\ \lambda_2+\lambda_3 &= \lambda p_{+1} \\ \lambda_3 &= \lambda p_{11} \end{aligned}$$

For the bivariate Bernoulli, set $\nu \to \infty$, and then pgf of (X, Y) becomes

$$\begin{aligned} G\left(t_1,t_2\right) &= \frac{1+\lambda\{1+p_{1+}(t_1-1)+p_{+1}(t_2-1)+p_{11}(t_1-1)(t_2-1)\}}{1+\lambda} \\ &= 1+\frac{\lambda}{\lambda+1}p_{1+}(t_1-1)+\frac{\lambda}{\lambda+1}p_{+1}(t_2-1)+\frac{\lambda}{\lambda+1}p_{11}(t_1-1)(t_2-1) \end{aligned} \tag{B.11}$$

compare the (B.11) and (3.4), we generate the relation between parameters of bivariate Bernoulli and the Bernoulli case under the CMP model. The details are in Table

Table B.1: The probability table of bivariate Bernoulli distribution as a special case of CMP distribution as $\nu \to \infty$

		Y 0	1	
X	0	$p^*_{00} = 1 - \frac{\lambda}{\lambda+1}(p_{01} + p_{10} + p_{11})$	$p^*_{01} = \frac{\lambda}{\lambda+1}p_{01}$	$p^*_{0+} = 1 - \frac{\lambda}{1+\lambda}p_{1+}$
	1	$p^*_{10} = \frac{\lambda}{1+\lambda}p_{10}$	$p^*_{11} = \frac{\lambda}{1+\lambda}p_{11}$	$p^*_{1+} = \frac{\lambda}{1+\lambda}p_{1+}$
		$p^*_{+0} = 1 - \frac{\lambda}{1+\lambda}p_{+1}$	$p^*_{+1} = \frac{\lambda}{1+\lambda}p_{+1}$	

B.1.

For the Geometric case, when $\nu = 0$, $\lambda < 1$ the $Z(\lambda, \nu) = \frac{1}{1-\lambda}$. The pgf becomes a bivariate Geometric distribution as

$$G\left(t_1, t_2\right) = \frac{1-\lambda}{1 - \lambda\{1 + p_{1+}(t_1 - 1) + p_{+1}(t_2 - 1) + p_{11}(t_1 - 1)(t_2 - 1)\}} \tag{B.12}$$

$$= \frac{1}{1 - \frac{\lambda}{1-\lambda}\{p_{1+}(t_1 - 1) + p_{+1}(t_2 - 1) + p_{11}(t_1 - 1)(t_2 - 1)\}} \tag{B.13}$$

for both numerator and denominator of (B.12) are greater than 0. It is similar as set α as an $exp(\theta)$ with probability density function as $f(\alpha) = \frac{1}{\theta}e^{-\alpha/\theta}$ to the pgf of Poisson case, where $\theta > 0$ and $\alpha > 0$. Recall the pgf of a bivariate poisson as in (B.10) where $G\left(t_1, t_2\right) = \exp\{\alpha c_1(t_1 - 1) + \alpha c_2(t_2 - 1) + \alpha c_3(t_1 - 1)(t_2 - 1)\}$. Then combine adjusted (B.10), the unconditional joint pfg of (X, Y) is

$$\begin{aligned}G\left(t_1, t_2\right) &= \int_0^{\infty} G\left(t_1, t_2 \mid \alpha\right) f(\alpha) d\alpha \\ &= \int_0^{\infty} \frac{1}{\theta} e^{-\alpha\left[\frac{1}{\theta}-c_1\left(t_1-1\right)-c_2\left(t_2-1\right)-c_3\left(t_1-1\right)\left(t_2-1\right)\right]} d\alpha \\ &= \frac{1/\theta}{1/\theta-\left[c_1\left(t_1-1\right)+c_2\left(t_2-1\right)+c_3\left(t_1-1\right)\left(t_2-1\right)\right]} \\ &= \frac{1}{1-\theta\left[c_1\left(t_1-1\right)+c_2\left(t_2-1\right)+c_3\left(t_1-1\right)\left(t_2-1\right)\right]}\end{aligned} \tag{B.14}$$

where the denominator of (B.14) is great than 0. Hence, compare (B.11) and (B.14), the relations between parameter of bivariate Geometric and as a special case of the bivariate CMP are

$$\theta c_1 = \frac{\lambda}{1-\lambda} p_{1+}, \qquad \theta c_2 = \frac{\lambda}{1-\lambda} p_{+1}, \qquad \theta c_3 = \frac{\lambda}{1-\lambda} p_{11}.$$

Appendix C

Real data

Table C.2: Cross-tablation of grouped data

Types	$\mathbf{N_1}$						
$\mathbf{N_2}$	0	1	2	3	4	5	6
0	24,408	1916	296	69	12	6	0
1	1068	317	61	21	6	2	2
2	203	71	18	6	2	1	1
3	49	14	8	3	3	1	0
4	11	6	2	0	1	0	0
5	2	0	0	0	0	0	1
6	1	0	0	1	0	0	0

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**(6), 716–723.

Balakrishnan, N. and Pal, S. (2013). Lognormal lifetimes and likelihood-based inference for flexible cure rate models based on com-poisson family. *Computational Statistics and Data Analysis*, **67**, 41–67.

Banerjee, O., Ghaoui, L. E., and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *The Journal of Machine Learning Research*, **9**, 485–516.

Basu, A. P. and Dhar, S. (1995). Bivariate geometric distribution. *Journal of Applied Statistical Science*, **2**(1), 33–34.

Bermúdez, L. and Karlis, D. (2017). A posteriori ratemaking using bivariate poisson models. *Scandinavian Actuarial Journal*, **2017**(2), 148–158.

Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. *The American Statistician*, **36**(3), 153–157.

Campbell, J. T. (1934). The poisson correlation function. *Proceedings of the Edinburgh Mathematical Society*, **2**(4), 18–26.

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

Dai, B., Ding, S., and Wahba, G. (2013). Multivariate bernoulli distribution models. *Bernoulli*, **19**(4), 1465–1483.

Hawkes, A. G. (1972). A bivariate exponential distribution with applications to reliability. *Journal of the Royal Statistical Society*, **34**, 129–1131.

Holgate, P. (1964). Estimation for the bivariate polsson distribution. *Biometrika*, **51**, 241–245.

Johnson, N. L., Kotz, S., and Kemp, A. W. (1992). *Univariate Discrete Distributions*. Wiley, New York.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. Wiley, New York.

King, G. (1998). *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. University of Michigan Press, Ann Arbor.

Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions*. Marcel Dekker, New York.

Li, J. and Dhar, S. (2013). Modeling with bivariate geometric distributions. *Journal of Applied Statistical Science*, **42**(2), 252–266.

Marshall, A. W. and Olkin, I. (1985). A family of bivariate distributions generated by the bivariate bernoulli distribution. *Journal of the American Statistical Association*, **80**(390), 332–338.

Minkova, L. D. and Balakrishnan, N. (2014). Type II bivariate pólya–Aeppli distribution. *Statistics and Probability Letters*, **88**, 40–49.

M'Kendrick, A. G. (1926). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, **44**, 98–130.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, **47**, 90–100.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**(2), 461–464.

Sellers, K. F. and Shmueli, G. (2010). A flexible regression model for count data. *The Annals of Applied Statistics*, **4**(2), 943–961.

Sellers, K. F., Morris, D. S., and Balakrishnan, N. (2016). Bivariate Conway–Maxwell–Poisson distribution: Formulation, properties, and inference. *Journal of Multivariate Analysis*, **150**, 152–168.

Sellers, K. F., Morris, D. S., Balakrishnan, N., and Davenport, D. (2018). R:flexible modeling of multivariate count data via the multivariate conway-maxwell-poisson distribution.

Shmueli, G., Minka, T. P., Kadane, J, B., Borle, S., and Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the conway-maxwell-poisson distribution. *Applied Statistics*, **54**, 127–142.

Teicher, H. (1954). On the multivariate Poisson distribution. *Skandinauisk Aktttarietidskrift*, **37**, 1–9.

Wainwright, M. J. and Jordan, M. I. (2006). Log-determinant relaxation for approximate inference in discrete markov random fields. *IEEE Transactions on Signal Processing*, **54**(6), 485–516.