On a General Class of Discrete Bivariate Distributions

Debasis Kundu *

Abstract

In this paper we develop a general class of bivariate discrete distributions. The basic idea is quite simple. The marginals are obtained by taking the random geometric sum of the baseline random variables. The proposed class of distributions is a flexible class of bivariate discrete distributions in the sense the marginals can take variety of shapes. The probability mass functions of the marginals can be heavy tailed, unimodal as well as multimodal. It can be both over dispersed as well as under dispersed. We discuss different properties of the proposed class of bivariate distributions. The proposed distribution has some interesting physical interpretations also. Further, we consider two specific base line distributions: Poisson and negative binomial distributions for illustrative purposes. Both of them are infinitely divisible. The maximum likelihood estimators of the unknown parameters cannot be obtained in closed form. They can be obtained by solving three and five dimensional non-linear optimizations problems, respectively. To avoid that we propose to use expectation maximization algorithm, and it is observed that the proposed algorithm can be implemented quite easily in practice. We have performed some simulation experiments to see how the proposed EM algorithm performs, and it works quite well in both the cases. The analysis of one real data set has been performed to show the effectiveness of the proposed class of models. Finally, we discuss some open problems and conclude the paper.

Key Words and Phrases: Discrete distributions; joint probability mass function; bivariate generating function; infinite divisibility, method of moment estimators.

AMS Subject Classifications: 62F10, 62F03, 62H12.

*Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. E-mail: kundu@iitk.ac.in, Phone no. 91-512-2597141, Fax no. 91-512-2597500.

1 Introduction

An extensive amount of work has been done introducing different bivariate discrete distributions, analyzing their properties and developing different estimation procedures. Special attention has been paid on bivariate geometric distributions and bivariate Poisson distributions, see for example Kocherlakota and Kocherlakota [10], Kocherlakota [11], Basu and Dhar [3], Kumar [13], Kemp [9], Lee and Cha [19], Nekoukhou and Kundu [24], Kundu and Nekoukhou [16] and see the references cited therein.

Recently, Lee and Cha [20] proposed two classes of discrete bivariate distributions and discussed their properties. Their idea is based on the minimum and maximum of two independent non-identically distributed random variables. The idea is quite simple, and it produces different unimodal shapes of bivariate discrete distributions. Unfortunately, because of the non-identical distributions, the joint probability mass functions (PMFs) or the marginal PMFs may not be in a convenient form. It makes it difficult to compute the estimates of the unknown parameters, and to derive different properties. Moreover, the marginals produced by the method of Lee and Cha [20] cannot have heavy tailed or multimodal probability mass functions, see also Nekoukhou and Kundu [24] in this respect.

The main aim of this paper is two fold. First of all we develop a flexible class of discrete bivariate distributions. The proposed discrete bivariate distribution has a very convenient joint probability generating function (PGF), hence the joint PMF can be obtained in a convenient form. We consider geometric random sum of independent identically distributed (i.i.d.) base line random variables. The idea is not new. It has been used in case of continuous random variables by several authors. For example, Chahkandi and Ganjali [5] used this idea when the base distribution is exponential and Barreto-Souza [2] extended the results in case of gamma distribution. The author [14] considered the case when the base distribution is univariate normal and the results have been recently generalized for multivariate normal base distribution by the author [15]. For some of the related literature, interested readers are referred to Kozubowski et al. [17] and [18]. Although, the above method seems to be a powerful method and it has been successfully used for several continuous distributions, not much attention has been paid in case of discrete distributions. For some of the related work along this direction, one may refer to Jayakumar and Mundassery [8], Kostadinova and Minkova [12], Minkova and Balakrishnan [21], Ozel [22], Ye [23] and see the references cited therein.

The main advantage of the proposed method is that it produces a flexible class of bivariate discrete distribution functions. The PMFs of the marginals can take variety of shapes. It can be heavy tailed, unimodal or multimodal. Since many well known discrete distributions like Poisson, geometric and negative binomial have convolution properties, several interesting properties of the proposed distributions can be established. It is observed that the marginals can be both under dispersed as well as over dispersed also depending on the base line distribution and the parameter values. The proposed model has some interesting physical interpretations also. We derive different properties of the proposed class of distributions in general and discuss in details two special cases.

Estimation of the unknown parameters is always an important problem in any data analysis. In this case it is observed that the maximum likelihood estimators (MLEs) of the unknown parameters cannot be obtained in closed forms. It involves solving higher dimensional optimization problems. In case of Poisson base line random variables, it involves solving three dimensional optimization problem, and in case of negative binomial base line random variables, one needs to solve a five dimensional optimization problem to compute the MLEs of the unknown parameters. It is observed that in this case the computation of the MLEs can be formulated as a missing value problem. Hence, the expectation maximization

(EM) algorithm seems to be a natural choice to compute the MLEs. But it is observed that to implement the classical EM one needs to compute two infinite sums at each M-step. To avoid that we have used an EM algorithm recently introduced by Kundu and Nekoukhou [16], mainly for discrete distribution and which is quite intuitive. Moreover, it does not involve any computation of infinite summation, and at each M-step the solutions can be obtained in explicit forms in case of Poisson base line random variables. In case of negative binomial base line random variable, at each M-step the solutions cannot be obtained in explicit forms. Due to that, at each M-step, we have used another EM algorithm following the approach of Adamidis [1], where the maximization can be obtained without solving any non-linear equations. We call this as nested EM (NEM) algorithm. The implementation of the proposed EM and NEM algorithms are quite simple. Further, it is observed that the method of moment estimators (MMEs) can be obtained quite conveniently in both the cases. Hence, the MMEs can be used as initial guesses of the EM and NEM algorithms. They work quite well. We performed some simulation experiments to see the performances of the proposed EM algorithm, and the performances are quite satisfactory. We analyze one real data set for illustrative purposes, and it is observed that the proposed model fits the data quite well and the EM and NEM algorithms also converge quite quickly from the proposed initial guesses.

The rest of the paper is organized as follows. In Section 2, we provide some motivations of the proposed model. The model formulation and some basic properties are discussed in Section 3. In Sections 4 and 5, we discuss two special cases. The EM and NEM algorithms are provided in Section 6. The results of the simulation experiments have been presented in Section 7, and the analysis of one data set is presented in Section 8. Finally, in Section 9 we conclude the paper.

2 Motivations

To motivate our proposed model we will start with an example. First let us consider the traditional construction of the bivariate Poisson distribution. It is based on a trivariate reduction technique and it can be described as follows. Suppose X_1, X_2 and X_3 are three independent Poisson random variables with mean λ_1, λ_2 and λ_3, respectively. Consider a new bivariate random variable (Y_1, Y_2), where

$$Y_1 = X_1 + X_3 \quad \text{and} \quad Y_2 = X_2 + X_3.$$

It was originally proposed by Holgate [7], see also Campbell [4] in this respect. The joint PMF of Y_1 and Y_2 can be obtained as

$$P(Y_1 = i, Y_2 = j) = e^{-(\lambda_1+\lambda_2+\lambda_3)} \sum_{k=0}^{\min\{i,j\}} \frac{\lambda_1^{i-k}\lambda_2^{j-k}\lambda_3^{k}}{(i-k)!(j-k)!k!}; \qquad i, j \in \mathbb{N}_0,$$

where $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. In this case the marginals will follow Poisson distribution. Clearly, this is an advantage of this model. But the marginal PMFs cannot be multimodal. Due to presence of the summation in the joint PMF computing the MLEs become difficult. Moreover, it may not be very easy to generalize it for other bivariate distributions.

Recently, Lee and Cha [20] introduced two very general methods to generate bivariate discrete distribution functions. The methods can be briefly described as follows. Let X_1, X_2 and X_3 be as defined before.

Lee-Cha Method 1:

$$U_1 = \max\{X_1, X_3\} \quad \text{and} \quad U_2 = \max\{X_2, X_3\}.$$

Lee-Cha Method 2:

$$V_1 = \min\{X_1, X_3\} \quad \text{and} \quad V_2 = \min\{X_2, X_3\}.$$

In both cases the generated bivariate distributions can be quite flexible. The joint PMF and the marginals PMFs can be of different shapes. Unfortunately, in this case also it has been observed, see Nekoukhou and Kundu [24], that the marginals may not be in a very convenient form. Moreover, in this case also in case of standard discrete distributions like Poisson, geometric or binomial, the marginal PMFs cannot be multimodal or heavy tailed. Most of the other bivariate distributions as proposed by Kumar [13] and Piperigou and Papageorgiou [25] have similar deficiencies.

The main aim of this paper is to propose a class of discrete bivariate distributions which has convenient joint PMF and marginal PMFs. The joint PMF and the marginal PMFs should be flexible enough and the marginals PMFs can have heavy tailed and multimodal shapes.

3 A general Class of Discrete Bivariate Distributions

In this section we develop a general class of discrete bivariate distributions and discuss its different properties. Two special cases will be discussed in the subsequent sections.

Suppose $U_1, U_2, \ldots$, are i.i.d. random variables with the probability mass function (PMF) $f_1(x)$, for $x \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}$. Further, let $V_1, V_2, \ldots$ be i.i.d. random variables with the PMF $f_2(x)$, for $x \in \mathbb{N}_0$, and let N be a geometric random variable with the PMF

$$P(N = n) = p(1-p)^{n-1}; \quad n \in \mathbb{N} = \{1, 2, \ldots\}, \tag{1}$$

for $0 < p < 1$. All the random variables are independently distributed. We define the bivariate random variable (X, Y) as follows

$$X = \sum_{i=1}^{N} U_i \quad \text{and} \quad Y = \sum_{i=1}^{N} V_i.$$

The above bivariate random variable (X, Y) has the following physical interpretations.

Accident Model: Suppose N is the number of accidents that took place in a given place during a fixed period of time. Let U_i and V_i denote the number of male and female deaths, respectively, due to the i-th accident for $i = 1, 2, \ldots, N$. Then X and Y denote the total number of male and female deaths, respectively, during that fixed period of time due to accidents in that given place.

Soccer Model: Suppose N denotes the number of soccer games played during a year between Team A and Team B. Suppose U_i and V_i denote the number of goal scored by Team A and Team B, respectively, at the i-th game, for $i = 1, 2, \ldots, N$. Then X and Y denote the total number of goals scored by Team A and Team B, respectively, against each other in that year.

From now on we will call $f_1(x)$ and $f_2(y)$ the base line PMFs. First we would like to derive different properties of the bivariate random variable (X, Y) for general base line PMFs. We use the following notation. $F_1(x)$ and $F_2(x)$ denote the cumulative distribution functions (CDFs) of $f_1(x)$ and $f_2(x)$, respectively. Moreover, $\phi_1(t)$ and $\phi_2(s)$ denote the characteristic functions (CHFs) of $f_1(x)$ and $f_2(x)$, respectively. We define $f_1^{(n)}(x)$ and $f_2^{(n)}(y)$ to be the n-fold convolutions of $f_1(x)$ and $f_2(x)$, respectively, i.e.

$$f_1^{(n)}(x) = P(U_1 + \ldots + U_n = x) \quad \text{and} \quad f_2^{(n)}(y) = P(V_1 + \ldots + V_n = y).$$

We further define $F_1^{(n)}(x)$ and $F_2^{(n)}(y)$ to be the CDFs correspond to the PMFs $f_1^{(n)}(x)$ and $f_2^{(n)}(y)$, respectively. The joint PMF of X and Y for $m = 0, 1, \ldots$ and $n = 0, 1, \ldots$, can be obtained as follows:

$$\begin{aligned} P\left(X = m, Y = n\right) &= P\left(\sum_{i=1}^{N} U_i = m, \sum_{j=1}^{N} V_j = n\right) \\ &= \sum_{k=1}^{\infty} P\left(\sum_{i=1}^{N} U_i = m, \sum_{j=1}^{N} V_j = n \middle| N = k\right) P(N = k) \end{aligned}$$

$$= \sum_{k=1}^{\infty} P\left(\sum_{i=1}^{k} U_i = m, \sum_{j=1}^{k} V_j = n\right) P(N=k)$$

$$= p\sum_{k=1}^{\infty}(1-p)^{k-1} f_1^{(k)}(m) f_2^{(k)}(n). \tag{2}$$

The joint CDF of X and Y for $m = 0, 1, \ldots$ and $n = 0, 1, \ldots$, is

$$P(X \le m, Y \le n) = p\sum_{k=1}^{\infty}(1-p)^{k-1} F_1^{(k)}(m) F_2^{(k)}(n). \tag{3}$$

From the joint PMF of X and Y, we immediately obtain the marginals PMFs of X and Y as

$$P(X=m) = p\sum_{k=1}^{\infty}(1-p)^{k-1} f_1^{(k)}(m) \quad \text{and} \quad P(Y=n) = p\sum_{k=1}^{\infty}(1-p)^{k-1} f_2^{(k)}(n). \tag{4}$$

$$P(X \le m) = p\sum_{k=1}^{\infty}(1-p)^{k-1} F_1^{(k)}(m) \quad \text{and} \quad P(Y \le n) = p\sum_{k=1}^{\infty}(1-p)^{k-1} F_2^{(k)}(n). \tag{5}$$

The joint CHF of X and Y for $t, s \in \mathbb{R}$ is

$$\begin{aligned}
\phi_{X,Y}(t,s) = E\left(e^{i(tX+sY)}\right) &= p\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{k=1}^{\infty} e^{i(tm+sn)}(1-p)^{k-1} f_1^{(k)}(m) f_2^{(k)}(n) \\
&= p\sum_{k=1}^{\infty}(1-p)^{k-1}\sum_{m=0}^{\infty} e^{itm} f_1^{(k)}(m) \sum_{n=0}^{\infty} e^{isn} f_2^{(k)}(n) \\
&= p\sum_{k=1}^{\infty}(1-p)^{k-1}\phi_1^k(t)\phi_2^k(s) \\
&= \frac{p\phi_1(t)\phi_2(s)}{1-(1-p)\phi_1(t)\phi_2(s)}.
\end{aligned} \tag{6}$$

From the joint CHF of X and Y, we obtain for $t, s \in \mathbb{R}$,

$$\phi_X(t) = \frac{p\phi_1(t)}{1-(1-p)\phi_1(t)} \quad \text{and} \quad \phi_Y(s) = \frac{p\phi_2(s)}{1-(1-p)\phi_2(s)}. \tag{7}$$

Moreover, it is immediate from (6) that X and Y are independent if and only if $p = 1$. Using the CHFs or otherwise, different moments and product moments can be easily obtained.

$$E(X) = \frac{E(U_1)}{p}, \qquad E(Y) = \frac{E(U_2)}{p},$$

$$V(X) = \frac{(1-p)(E(U_1))^2}{p^2} + \frac{V(U_1)}{p} \qquad V(Y) = \frac{(1-p)(E(V_1))^2}{p^2} + \frac{V(V_1)}{p}$$

$$\text{Cov}(X,Y) = \frac{1-p}{p^2} \; E(U_1)E(V_1).$$

Some of the points are quite clear from the above moments expressions. It is clear that as $p \to 0$, the mean and variance go to ∞. Hence, for large p, the corresponding distribution function behaves like a heavy tailed distribution. The correlation between X and Y is always positive and it goes to zero, as $p \to 1$, and it goes to 1, as $p \to 0$. The variance to mean ratio (VMR) for X and Y are

$$\text{VMR}(X) = \frac{(1-p)E(U_1)}{p} + \text{VMR}(U_1) \quad \text{and} \quad \text{VMR}(Y) = \frac{(1-p)E(V_1)}{p} + \text{VMR}(V_1).$$

It is clear that as $p \to 0$, the marginals will be over dispersed, and for large p, the marginals can be under dispersed if the base line distributions are under dispersed. Therefore, it is possible to have both over dispersed and under dispersed marginals for the proposed bivariate distributions. It is also possible to have two opposite behavior of the two marginals. Now we will consider some special cases in the subsequent sections.

4 Two Special Cases

4.1 Bivariate Poisson Geometric

In this section it is assumed that U_i follows ($\sim$) a Poisson distribution with mean λ_1 (PO (λ_1)) and $V_i \sim$ PO (λ_2). We denote this new distribution as BPG (λ_1, λ_2, p), and the marginals will be denoted by UPG (λ_1, p) and UPG (λ_2, p), respectively.

The joint PMF of X and Y for $m = 0, 1, \ldots$ and $n = 0, 1, \ldots$, can be obtained as follows:

$$P\left(X = m, Y = n\right) = C(\lambda_1 + \lambda_2 - \ln(1-p), m+n) \times \frac{p}{1-p} \times \frac{\lambda_1^m \lambda_2^n}{m!n!}. \tag{8}$$

Here, for $a > 0$ and $j = 0, 1, 2, \ldots$,

$$C(a, j) = \sum_{k=1}^{\infty} k^j \; e^{-ak}.$$

The exact expressions of $C(a, j)$ for different values of j can be obtained recursively, and they are provided in the Appendix A. The marginal PMF of X for $m = 0, 1, 2, \ldots$, and of Y for $n = 0, 1, 2, \ldots$ can be obtained as,

$$\begin{aligned} P\,(X = m) &= C(\lambda_1 - \ln(1-p), m) \times \frac{p}{1-p} \times \frac{\lambda_1^m}{m!} \\ P\,(Y = n) &= C(\lambda_2 - \ln(1-p), n) \times \frac{p}{1-p} \times \frac{\lambda_2^n}{n!}, \end{aligned}$$

respectively.

We have plotted the PMFs of the UPG distribution for different parameter values in Figures 1-4. It shows that the PMFs can take variety of shapes. It can be a decreasing function, unimodal, multimodal and heavy tailed also.

$$P(X = m | Y = n) = \frac{C(\lambda_1 + \lambda_2 - \ln(1-p), m+n)}{C(\lambda_1 - \ln(1-p), m)} \times \frac{\lambda_1^m}{m!}.$$

The joint CHF of X and Y for $t, s \in \mathbb{R}$ is

$$\phi_{X,Y}(t, s) = E(e^{i(tX+sY)}) = \frac{pe^{\lambda_1(e^{it}-1)} e^{\lambda_2(e^{is}-1)}}{1-(1-p)e^{\lambda_1(e^{it}-1)} e^{\lambda_2(e^{is}-1)}}.$$

Hence, the CHF of X and Y for $t \in \mathbb{R}$, are

$$\phi_X(t) = Ee^{itX} = \frac{pe^{\lambda_1(e^{it}-1)}}{1-(1-p)e^{\lambda_1(e^{it}-1)}} \quad \text{and} \quad \phi_Y(t) = Ee^{itY} = \frac{pe^{\lambda_2(e^{it}-1)}}{1-(1-p)e^{\lambda_2(e^{it}-1)}}.$$

Different moments and product moments are

$$E(X) = \frac{\lambda_1}{p}, \quad E(Y) = \frac{\lambda_2}{p}, \quad V(X) = \frac{(1-p)\lambda_1^2}{p^2} + \frac{\lambda_1}{p}, \quad V(Y) = \frac{(1-p)\lambda_2^2}{p^2} + \frac{\lambda_2}{p},$$

$$\text{Cov}(X, Y) = \frac{\lambda_1 \lambda_2 (1-p)}{p^2}.$$

We have the following results regarding stochastic ordering of the Poisson geometric distribution. The proofs are quite simple, hence the details are avoided.

RESULT 1: If $(X, Y) \sim \text{BPG}(\lambda_1,\, \lambda_2,\, p)$ and $\lambda_1 > \lambda_2$, then X is stochastically larger than Y.

RESULT 2: If $U \sim \text{UPG}(\lambda, p_1)$ and $V \sim \text{UPG}(\lambda, p_2)$, then for $p_1 < p_2$, U is stochastically larger than V.

THEOREM 1: If $(X, Y) \sim \text{BPG}(\lambda_1,\, \lambda_2,\, p)$, then we have the following results.

(a) $X + Y \sim \text{UPG}(\lambda_1 + \lambda_2, p)$.

(b) X given $X + Y$ follows a binomial distribution, i.e.

$$P(X = m | X + Y = n) = \binom{n}{m} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^m \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-m}; \qquad n = 0, 1, \ldots, m.$$

PROOF: (a) can be obtained from the joint CHF. (b) can be obtained using (a).

THEOREM 2: If $\{(X_i, Y_i); i = 1, 2, \ldots\}$ is a sequence of i.i.d. random variables from $\text{BPG}(\lambda_1,\, \lambda_2,\, p)$, M is a geometric random variable with parameter δ, for $0 < \delta < 1$, and for $i = 1, \ldots$, it is independent of (X_i, Y_i), then

$$\left(\sum_{i=1}^{M} X_i, \sum_{i=1}^{M} Y_i\right) \sim \text{BPG}(\lambda_1, \lambda_2, p\delta).$$

PROOF: The proof mainly follows from the joint CHF. ∎

THEOREM 3: If $(X, Y) \sim \text{BPG}(\lambda_1,\, \lambda_2,\, p)$, then (X, Y) is infinitely divisible.

PROOF: Note that we need to prove that for any positive integer n, there exists i.i.d. random variables $\{(W_{1i}^{(n)}, W_{2i}^{(n)}); i = 1, 2, \ldots, n\}$, such as $(W_{1i}^{(n)}, W_{2i}^{(n)})$ follows bivariate Poisson geometric random variables, and

$$(X, Y) \overset{d}{=} \left(\sum_{i=1}^{n} W_{1i}^{(n)}, \sum_{i=1}^{n} W_{2i}^{(n)}\right).$$

Here '$\stackrel{d}{=}$' means equal in distribution. For any fixed n, let $r = 1/n$. Consider a sequence of i.i.d. random variables $Y_1, Y_2, \ldots$, such that $Y_i \sim \text{PO}(r\lambda_1)$. Similarly, consider another sequence of i.i.d. random variables $Z_1, Z_2, \ldots$, such that $Z_i \sim \text{PO}(r\lambda_2)$. Suppose T is a negative binomial NB(r, p) random variable, with the following PMF

$$P(T = k) = \frac{\Gamma(k + r)}{k!\Gamma(r)} p^r (1 - p)^k; \quad k = 0, 1, 2, \ldots.$$

All the above random variables, namely the Y_i's the Z_i's and T are independently distributed. Consider the following bivariate random variable

$$(W_{11}^{(n)}, W_{21}^{(n)}) \stackrel{d}{=} \left(\sum_{i=1}^{1+nT} Y_i, \sum_{i=1}^{1+nT} Z_i \right).$$

The joint CHF of $(W_{11}^{(n)}, W_{21}^{(n)})$ can be obtained as

$$\begin{aligned}
\phi_{W_{11}^{(n)}, W_{12}^{(n)}}(t, s) &= E(e^{itW_{11}^{(n)} + isW_{21}^{(n)}}) = E_T \left(E \left(e^{itW_{11}^{(n)} + isW_{21}^{(n)}} | T \right) \right) \\
&= E_T \left(e^{r\lambda_1(e^{it}-1)} e^{r\lambda_2(e^{is}-1)(1+nT)} \right) \\
&= e^{r\lambda_1(e^{t}-1)} e^{r\lambda_2(e^{is}-1)} \left(\frac{p}{1 - (1-p)e^{\lambda_1(e^{it}-1)} e^{\lambda_2(e^{is}-1)}} \right)^r \\
&= \left(\frac{p e^{\lambda_1(e^{it}-1)} e^{\lambda_2(e^{is}-1)}}{1 - (1-p)e^{\lambda_1(e^{it}-1)} e^{\lambda_2(e^{is}-1)}} \right)^r.
\end{aligned}$$

Hence, the result is obtained. ∎

The following result shows that the bivariate Poisson geometric distribution has an interesting decomposition. It might have some independent interest also.

THEOREM 4: Let us assume that $(X, Y) \sim \text{BPG}(\lambda_1,\ \lambda_2,\ p)$. Suppose $Q \sim \text{PO}(\lambda)$, where $\lambda = -\ln p$, and it is independent of Z_i's, where $\{Z_i; i = 1, 2, \ldots\}$ is a sequence of i.i.d. random variables having logarithmic distribution with the probability mass function

$$P(Z_1 = k) = \frac{(1-p)^k}{\lambda k}; \quad k = 1, 2, \ldots. \tag{9}$$

Moreover, $\{W_{1i}; i = 1, 2, \ldots\}$ is a sequence of i.i.d. random variables such that $W_{1i}|Z_i \sim \text{PO}(\lambda_1 Z_i)$, similarly, $\{W_{2i}; i = 1, 2, \ldots\}$ is a sequence of i.i.d. random variables such that

$W_{2i}|Z_i \sim \text{PO}(\lambda_2 Z_i)$, and conditionally they are independently distributed. R and S are two independent random variables and they are independent of all the previous random variables, such that $R \sim \text{PO}(\lambda_1)$, $S \sim \text{PO}(\lambda_2)$. Then we have the following decomposition of (X, Y):

$$(X, Y) \stackrel{d}{=} \left(R + \sum_{i=1}^{Q} W_{1i}, S + \sum_{i=1}^{Q} W_{2i} \right). \tag{10}$$

Proof: First observe that the probability generating function of Q and Z_1 are

$$E(t^Q) = e^{\lambda(t-1)}, \quad t \in \mathbb{R} \quad \text{and} \quad E(t^{Z_1}) = \frac{\ln(1-(1-p)t)}{\ln p}, \quad t < (1-p)^{-1}.$$

The joint CHF of the right hand side of (10) can be written as

$$\begin{aligned}
\phi(u, v) &= E\left(e^{iu(R+\sum_{j=1}^{Q} W_{1j})+iv(S+\sum_{j=1}^{Q} W_{2j})}\right) \\
&= e^{\lambda_1(e^{iu}-1)} e^{\lambda_2(e^{iv}-1)} E\left(e^{iu\sum_{j=1}^{Q} W_{1j}+iv\sum_{j=1}^{Q} W_{2j}}\right) \\
&= e^{\lambda_1(e^{iu}-1)} e^{\lambda_2(e^{iv}-1)} E_Q\left[E\left(e^{iu\sum_{j=1}^{Q} W_{1j}+iv\sum_{j=1}^{Q} W_{2j}}|Q\right)\right] \\
&= e^{\lambda_1(e^{iu}-1)} e^{\lambda_2(e^{iv}-1)} E_Q E\left[E\left(e^{iu\sum_{j=1}^{Q} W_{1j}}|Q, Z_1, Z_2, \ldots\right) E\left(e^{iv\sum_{j=1}^{Q} W_{2j}}|Q, Z_1, Z_2, \ldots\right)\right] \\
&= e^{\lambda_1(e^{iu}-1)} e^{\lambda_2(e^{iv}-1)} E_Q E\left[e^{\lambda_1(e^{iut}-1)\sum_{j=1}^{Q} Z_j} e^{\lambda_2(e^{ivs}-1)\sum_{j=1}^{Q} Z_j}\right] \\
&= e^{\lambda_1(e^{iu}-1)} e^{\lambda_2(e^{iv}-1)} E_Q\left[E\left[e^{(\lambda_1(e^{iut}-1)+\lambda_2(e^{ivs}-1))Z_1}\right]\right]^Q \\
&= e^{\lambda_1(e^{iu}-1)} e^{\lambda_2(e^{iv}-1)} E_Q\left[\frac{\ln(1-(1-p)e^{(\lambda_1(e^{iut}-1)+\lambda_2(e^{ivs}-1))})}{\ln p}\right]^Q \\
&= e^{\lambda_1(e^{iu}-1)} e^{\lambda_2(e^{iv}-1)} e^{\ln p - \ln(1-(1-p)e^{(\lambda_1(e^{iut}-1)+\lambda_2(e^{ivs}-1))})} \\
&= \frac{p e^{\lambda_1(e^{iu}-1)} e^{\lambda_2(e^{iv}-1)}}{1-(1-p)e^{\lambda_1(e^{iu}-1)} e^{\lambda_2(e^{iv}-1)}}.
\end{aligned}$$

4.2 Bivariate Negative Binomial Geometric

In this section we consider another special case when $U_1 \sim \text{NB}(r_1, \theta_1)$ and $V_1 \sim \text{NB}(r_2, \theta_2)$, where $r_1 > 0, r_2 > 0, 0 < \theta_1, \theta_2 < 1$. Here, $\text{NB}(r, \theta)$, for $r > 0, 0 < \theta < 1$, means a negative binomial distribution with the PMF

$$P(X = k) = \frac{\Gamma(k+r)}{k!\,\Gamma(r)} \theta^k (1-\theta)^r; \quad k = 0, 1, 2, \ldots. \tag{11}$$

In this case we denote this bivariate distribution as BNBG$(r_1, \theta_1, r_2, \theta_2, p)$, and the marginals will be denoted by UNBG(r_1, θ_1, p) and UNBG(r_2, θ_2, p), respectively. The joint PMF of X and Y for $m = 0, 1, \ldots$ and $n = 0, 1, \ldots$, can be written as

$$P(X = m, Y = n) = D(r_1, r_2, \theta_1, \theta_2, m, n, p)\theta_1^m \theta_2^n \frac{p}{1-p}, \tag{12}$$

where

$$D(r_1, r_2, \theta_1, \theta_2, m, n, p) = \sum_{k=1}^{\infty} \frac{\Gamma(m + kr_1)}{m!\,\Gamma(kr_1)} \times \frac{\Gamma(n + kr_2)}{n!\,\Gamma(kr_2)} (1-p)^k (1-\theta_1)^{kr_1} (1-\theta_2)^{kr_2}.$$

The marginal PMFs of X and Y can be obtained as

$$\begin{aligned} P(X = m) &= D_1(r_1, \theta_1, m, p)\theta_1^m \frac{p}{1-p} \\ P(Y = n) &= D_1(r_2, \theta_2, n, p)\theta_2^m \frac{p}{1-p}, \end{aligned}$$

where

$$D_1(r, \theta, m, p) = \sum_{k=1}^{\infty} \frac{\Gamma(m + kr)}{m!\,\Gamma(kr)} (1-\theta)^{kr} (1-p)^k.$$

In this case also we have plotted the PMFs of the UNBG distribution for different parameter values in Figures 5 - 8. It shows that for UNBG also, the PMFs can take variety of shapes. The PMF can be a decreasing function, unimodal, multimodal and heavy tailed also.

The joint CHF of X and Y for $t, s \in \mathbb{R}$, is

$$\phi_{X,Y}(t, s) = E(e^{i(tX+sY)}) = \frac{p(1-\theta_1)^{r_1}(1-\theta_2)^{r_2}}{(1-\theta_1 e^{it})^{r_1}(1-\theta_2 e^{is})^{r_2} - (1-p)(1-\theta_1)^{r_1}(1-\theta_2)^{r_2}}.$$

The marginal CHFs of X and Y can be obtained as

$$\phi_X(t) = \frac{p(1-\theta_1)^{r_1}}{(1-\theta_1 e^{it})^{r_1} - (1-p)(1-\theta_1)^{r_1}}, \quad \phi_Y(s) = \frac{p(1-\theta_2)^{r_2}}{(1-\theta_2 e^{is})^{r_2} - (1-p)(1-\theta_2)^{r_2}}. \tag{13}$$

Different moments and product moments of the BNBG distribution can be easily obtained and they are as follows:

$$E(X) = \frac{\theta_1 r_1}{p(1-\theta_1)}, \quad E(Y) = \frac{\theta_2 r_2}{p(1-\theta_2)},$$

$$V(X) = \frac{(1-p)\theta_1^2 r_1^2}{p^2(1-\theta_1)^2} + \frac{\theta_1 r_1}{p(1-\theta_1)^2} \qquad V(Y) = \frac{(1-p)\theta_2^2 r_2^2}{p^2(1-\theta_2)^2} + \frac{\theta_2 r_2}{p(1-\theta_2)^2}$$

$$\text{Cov}(X,Y) = \frac{1-p}{p^2} \quad \frac{\theta_1 r_1 \theta_2 r_2}{(1-\theta_1)(1-\theta_2)}.$$

THEOREM 5: If $(X,Y) \sim$ BNBG(r_1, θ, r_2, θ, p), then $X + Y \sim$ UNBG$(r_1 + r_2, \theta, p)$.

PROOF: It can be obtained from the joint CHF. ∎

THEOREM 6: If $\{(X_i, Y_i); i = 1, 2, \ldots\}$ is a sequence i.i.d. random variables from BNBG(r_1, θ_1, r_2, θ_2, p), and M is a geometric random variable with parameter δ, for $0 < \delta < 1$, then

$$\left(\sum_{i=1}^{M} X_i, \sum_{i=1}^{M} Y_i\right) \sim \text{BNBG}(r_1, \theta_1, r_2, \theta_2, p\delta).$$

PROOF: The proof mainly follows from the joint CHF. ∎

THEOREM 7: If $(X,Y) \sim$ BNBG(r_1, θ_1, r_2, θ_2, p), then (X,Y) is infinitely divisible.

PROOF: The proof can be obtained along the same line of proof as of Theorem 3, the details are avoided. ∎

THEOREM 8: Let us assume that $(X,Y) \sim$ BNBG(r_1, θ_1, r_2, θ_2, p). Suppose $Q \sim$ PO(λ), where $\lambda = -\ln p$, and it is independent of Z_i's, where $\{Z_i; i = 1, 2, \ldots\}$ is a sequence of i.i.d. random variables having logarithmic distribution with probability mass function as defined in (9). Moreover, $\{W_{1i}; i = 1, 2, \ldots\}$ is a sequence of i.i.d. random variables such that $W_{1i}|Z_i \sim$ NB($r_1 Z_i$, θ_1), similarly, $\{W_{2i}; i = 1, 2, \ldots\}$ is a sequence of i.i.d. random variables such that $W_{2i}|Z_i \sim$ NB($r_2 Z_i$, θ_2), and conditionally they are independently distributed. R and S are two independent random variables and they are independent of all the previous random variables, such that $R \sim$ NB(r_1, θ_1), $S \sim$ NB(r_2, θ_2). Then we have the following decomposition of (X,Y):

$$(X,Y) \stackrel{d}{=} \left(R + \sum_{i=1}^{Q} W_{1i}, S + \sum_{i=1}^{Q} W_{2i}\right). \tag{14}$$

PROOF: The proof can be obtained similarly as the proof of Theorem 4. ∎

5 Maximum Likelihood Estimators

In this section we discuss the computation of the MLEs based on EM and NEM algorithms for BPG and BNBG distributions, respectively. It is assumed that we have a random sample of size m, $\mathcal{D} = \{(x_i, y_i); i = 1, \ldots, m\}$, either from BPG$(\lambda_1, \lambda_2, p)$ or from a BNBG$(r_1, \theta_1, r_2, \theta_2, p)$ depending on the situation. The problem is to estimate the unknown parameter vector $\mathbf{\Omega}$, where $\mathbf{\Omega} = (\lambda_1, \lambda_2, p)$ or $\mathbf{\Omega} = (r_1, \theta_1, r_2, \theta_2, p)$ for BPG$(\lambda_1, \lambda_2, p)$ and BNBG$(r_1, \theta_1, r_2, \theta_2, p)$, respectively, based on $\mathcal{D}$.

5.1 BPG Distribution

Based on the sample $\mathcal{D}$, the relevant part of the log-likelihood function can be written as

$$l(\mathbf{\Omega}|\mathcal{D}) = K_1(\lambda_1, \lambda_2, p) + m \ln p - m \ln(1-p) + \widetilde{x} \ln \lambda_1 + \widetilde{y} \ln \lambda_2, \tag{15}$$

where

$$K_1(\lambda_1, \lambda_2, p) = \sum_{i=1}^{m} \ln C(\lambda_1 + \lambda_2 - \ln(1-p), x_i + y_i), \quad \widetilde{x} = \sum_{i=1}^{m} x_i, \quad \widetilde{y} = \sum_{i=1}^{m} y_i.$$

Hence, the MLEs can be obtained by solving the following three non-linear equations simultaneously:

$$\dot{l}_{\lambda_1}(\mathbf{\Omega}|\mathcal{D}) = 0, \quad \dot{l}_{\lambda_2}(\mathbf{\Omega}|\mathcal{D}) = 0, \quad \dot{l}_{p}(\mathbf{\Omega}|\mathcal{D}) = 0.$$

Clearly, they cannot be obtained in explicit forms, and one needs to use the Newton-Raphson or Gauss-Newton type algorithm to solve the above non-linear equations. To avoid that we propose to use an EM algorithm which can be implemented very easily. The main idea of the proposed EM algorithm is based on the following observations. Let us assume that along with each (X, Y), we observe the associated N also. Therefore, the complete observations are $\mathcal{D}^* = \{(x_1, y_1, n_1), \ldots, (x_m, y_m, n_m)\}$. Based on the complete observations, the relevant part of the log-likelihood function can be written as

$$l_{complete}(\boldsymbol{\Omega}|\mathcal{D}^*) = m \ln p + (\widetilde{n} - m)\ln(1-p) - \widetilde{n}\lambda_1 + \widetilde{x}\ln\lambda_1 - \widetilde{n}\lambda_2 + \widetilde{y}\ln\lambda_2, \tag{16}$$

where $\widetilde{n} = \sum_{i=1}^{m} n_i$. The MLEs can be obtained by maximizing (16) with respect to the unknown parameters. In this case it can be easily seen that they are unique and they can be obtained in explicit forms as

$$\widehat{\lambda}_{1c} = \frac{\widetilde{x}}{\widetilde{n}}, \quad \widehat{\lambda}_{2c} = \frac{\widetilde{y}}{\widetilde{n}}, \quad \widehat{p}_c = \frac{m}{\widetilde{n}}. \tag{17}$$

Now to implement EM, let us assume that at the j-th stage the estimates of the parameters are denoted by $\boldsymbol{\Omega}^{(j)} = (\lambda_1^{(j)}, \lambda_2^{(j)}, p^{(j)})$. At the j-th stage we replace the unknown n_i by its estimate, and in this case we estimate it by maximizing the conditional probability $P(N = n|X = x_i, Y = y_i)$, similarly as in Kundu and Nekoukhou [16]. We denote the estimated n_i as $\widetilde{n}_i^{(j)}$, and it can be obtained as

$$\widetilde{n}_i^{(j)} = \arg\max_n P(N = n|X = x_i, Y = y_i, \boldsymbol{\Omega}^{(j)}). \tag{18}$$

The explicit expression of $\widetilde{n}_i^{(j)}$ is provided in the Appendix B. Therefore, the implementation of the proposed EM algorithm becomes quite simple. If at the j-th iteration step, we denote $\widetilde{n}^{(j)} = \sum_{i=1}^{m} \widetilde{n}_i^{(j)}$, then

$$\lambda_1^{(j+1)} = \frac{\widetilde{x}}{\widetilde{n}^{(j)}}, \quad \lambda_2^{(j+1)} = \frac{\widetilde{y}}{\widetilde{n}^{(j)}}, \quad p^{(j+1)} = \frac{m}{\widetilde{n}^{(j)}}. \tag{19}$$

The algorithm continues until it converges.

5.2 BNBG Distribution

The relevant part of the log-likelihood function can be written as

$$l(\boldsymbol{\Omega}|\mathcal{D}) = K_2(r_1, \theta_1, r_2, \theta_2, p) + \widetilde{x}\ln\theta_1 + \widetilde{y}\ln\theta_2 + m\ln p - m\ln(1-p), \tag{20}$$

where

$$K_2(r_1, \theta_1, r_2, \theta_2, p) = \sum_{i=1}^{m} \ln D(r_1, r_2, \theta_1, \theta_2, x_i, y_i),$$

$\widetilde{x}$ and $\widetilde{y}$ are same as defined before. Hence, the MLEs can be obtained by solving the following five non-linear equations simultaneously:

$$\dot{l}_{r_1}(\mathbf{\Omega}|\mathcal{D}) = 0, \quad \dot{l}_{\theta_1}(\mathbf{\Omega}|\mathcal{D}) = 0, \quad \dot{l}_{r_2}(\mathbf{\Omega}|\mathcal{D}) = 0, \quad \dot{l}_{\theta_2}(\mathbf{\Omega}|\mathcal{D}) = 0, \quad \dot{l}_{p}(\mathbf{\Omega}|\mathcal{D}) = 0.$$

In this case also, they cannot be solved explicitly. Here also we would like to use EM algorithm by treating this problem as a missing value problem. Let us assume that here also we have the complete observation as $\mathcal{D}^*$ as defined before. Based on the complete observations, the relevant part of the complete log-likelihood function can be written as follows:

$$\begin{aligned} l_{complete}(\mathbf{\Omega}|\mathcal{D}^*) \;=\; & \sum_{i=1}^{m} \ln\Gamma(x_i + n_i r_1) - \sum_{i=1}^{m} \ln\Gamma(n_i r_1) + \widetilde{x}\ln\theta_1 + \widetilde{n} r_1 \ln(1-\theta_1) + \\ & \sum_{i=1}^{m} \ln\Gamma(y_i + n_i r_2) - \sum_{i=1}^{m} \ln\Gamma(n_i r_2) + \widetilde{x}\ln\theta_2 + \widetilde{n} r_2 \ln(1-\theta_2) + \\ & m\ln p + (\widetilde{n} - m)\ln(1-p). \end{aligned} \qquad (21)$$

The MLEs of the unknown parameters can be obtained by maximizing (21) with respect to five unknown parameters. But unlike the previous case, here they cannot be obtained in explicit forms. Now to obtain the MLEs in case of complete observations, we adopt the similar procedure as suggested by Adamidis [1]. It may be mentioned that Adamidis [1] has suggested a very effective EM algorithm to compute the unknown parameters of a NB(r, θ) distribution based on a i.i.d. sample. Adamidis [1] showed that at each E-step, the corresponding M-step can be obtained in explicit forms. Although, in this case the observations are independent, they are not identically distributed. We need to solve the following problem. With the abuse of notation, suppose it is assumed that $X_i \sim$ NB$(n_i r, \theta)$, for $i = 1, \ldots, m$, where X_i's are independent, n_i's are known, then how to compute the MLEs of r and θ, based on a random sample $\{x_1, \ldots, x_m\}$. It is possible to use the method suggested by Adamidis [1], and it is observed that in our case also at each E-step, the corresponding M-step can be obtained in explicit form. The details are explained in Appendix C.

Hence, from the complete data log-likelihood function (21), we can obtain the MLE of p, and using the EM algorithm as described in Appendix C, we can obtain the MLEs of r_1, θ_1, r_2 and θ_2. Now, to perform the outer EM, we proceed exactly the same way as in the BPG case. Now to implement the NEM algorithm, at the j-th stage similarly as before, we need to compute

$$\widetilde{n}_i^{(j)} = \arg\max_n P(N = n | X = x_i, Y = y_i, \mathbf{\Omega}^{(j)}), \tag{22}$$

here $\mathbf{\Omega}^{(j)} = (r_1^{(j)}, \theta_1^{(j)}, r_2^{(j)}, \theta_2^{(j)}, p^{(j)})$ denotes the estimate of $\mathbf{\Omega} = (r_1, \theta_1, r_2, \theta_2, p)$ at the j-th step. In this case also, the explicit expression of $\widetilde{n}_i^{(j)}$ is provided in the Appendix B. Therefore, the implementation of the proposed NEM algorithm becomes quite simple.

6 Simulation Results

In this section we have presented some simulation results to show how the proposed EM algorithm performs for both the models for different sample sizes and for different parameter values.

6.1 BPG Model

In this case we have considered different sets of $(\lambda_1, \lambda_2, p)$ and different n. For a given set of $(\lambda_1, \lambda_2, p)$ and n, we generate a random sample from a BPG model. In each case we first compute the initial estimates of the unknown parameters based on the method of moment estimators, and these have been used as initial guesses for the EM algorithm. Then we use the EM algorithm to compute the MLEs of the unknown parameters. The EM algorithm stops when ever the absolute difference between the two iterates is less that 10^{-4} for all the three parameters. We replicate the process 1000 times, and report the average estimates and the mean squared errors in each case. In all these cases it has been observed that the algorithm stops within 10 iterations. The results are reported in Tables 1 to 3.

n	$\lambda_1 = 2.0$	$\lambda_2 = 2.0$	$p = 0.5$
25	2.2136	2.2181	0.5547
	(0.6133)	(0.6147)	(0.0417)
50	2.1031	2.1028	0.5313
	(0.3875)	(0.3863)	(0.0251)
75	2.0612	2.0615	0.5142
	(0.2815)	(0.2798)	(0.0187)
100	2.0501	2.0504	0.5017
	(0.2214)	(0.2238)	(0.0140)

Table 1: The average ML estimates and the associated MSEs of λ_1, λ_2 and p

n	$\lambda_1 = 2.0$	$\lambda_2 = 2.0$	$p = 0.75$
25	2.0281	2.0263	0.7591
	(0.2312)	(0.2345)	(0.0214)
50	2.0207	2.0202	0.7584
	(0.1213)	(0.1198)	(0.0145)
75	2.0178	2.0165	0.7568
	(0.0879)	(0.0865)	(0.0114)
100	2.0147	2.0138	0.7525
	(0.0675)	(0.0659)	(0.0072)

Table 2: The average ML estimates and the associated MSEs of λ_1, λ_2 and p

Some of the points are quite clear from the above simulation results. It is observed that as the sample size increases the biases and the MSEs decrease in each case. It verifies the consistency properties of the MLEs. Comparing Table 1 and 2 it is observed that as p increases the biases and the MSEs of the MLEs of λ_1 and λ_2 decrease. Similarly, comparing Table 1 and Table 3, it is observed that as λ_1 and λ_2 increase the biases and the MSEs of p decrease.

n	$\lambda_1 = 5.0$	$\lambda_2 = 5.0$	$p = 0.5$
25	5.4611	5.3912	0.5489
	(3.1237)	(3.1150)	(0.0311)
50	5.2312	5.2215	0.5216
	(1.9548)	(1.9448)	(0.0187)
75	5.1549	5.1338	0.5145
	(1.4016)	(1.3998)	(0.0122)
100	5.0217	5.0116	0.5017
	(1.1127)	(1.1099)	(0.0140)

Table 3: The average ML estimates and the associated MSEs of λ_1, λ_2 and p

6.2 BNBG Model

In this case we have considered different sets of $(r_1, r_2, \theta_1, \theta_2, p)$ and different n. For a given set of $(r_1, r_2, \theta_1, \theta_2, p)$ and n, first we generate a random sample from BNBG model and we obtain the method of moment estimates of the unknown parameters. Based on the method of moment estimates as the initial guesses, we use the EM algorithm to compute the MLEs of the unknown parameters. We have used the same stopping criterion as the BPG model. We replicate the process 1000 times, and report the average estimates and the mean squared errors in each case. In all these cases it has been observed that the algorithm stops within 18 iterations. The results are reported in Tables 4 to 7.

n	$r_1 = 2$	$r_2 = 2$	$\theta_1 = 0.15$	$\theta_2 = 0.15$	$p = 0.95$
25	2.3773	2.3225	0.1793	0.1698	0.9582
	(0.9123)	(0.9211)	(0.0061)	(0.0059)	(0.0045)
50	2.3128	2.3142	0.1712	0.1778	0.9531
	(0.6115)	(0.6098)	(0.0027)	(0.0031)	(0.0024)
75	2.2978	2.2776	0.1619	0.1624	0.9487
	(0.5051)	(0.5112)	(0.0024)	(0.0027)	(0.0018)
100	2.1013	2.0998	0.1514	0.1505	0.9502
	(0.2978)	(0.2814)	(0.0013)	(0.0014)	(0.0011)

Table 4: The average ML estimates and the associated MSEs of r_1, r_2, θ_1, θ_2 and p

n	$r_1 = 1$	$r_2 = 1$	$\theta_1 = 0.15$	$\theta_2 = 0.15$	$p = 0.95$
25	1.1454	1.1445	0.1808	0.1812	0.9212
	(0.2689)	(0.2670)	(0.0108)	(0.0115)	(0.0213)
50	1.1321	1.1365	0.1803	0.1801	0.9238
	(0.1523)	(0.1587)	(0.0048)	(0.0049)	(0.0165)
75	1.1189	1.1201	0.1789	0.1728	0.9289
	(0.1432)	(0.1387)	(0.0038)	(0.0037)	(0.0094)
100	1.1013	1.0879	0.1715	0.1689	0.9328
	(0.0769)	(0.0754)	(0.0023)	(0.0021)	(0.0078)

Table 5: The average ML estimates and the associated MSEs of r_1, r_2, θ_1, θ_2 and p

n	$r_1 = 2$	$r_2 = 2$	$\theta_1 = 0.15$	$\theta_2 = 0.15$	$p = 0.75$
25	2.2487	2.2453	0.1612	0.1601	0.7601
	(0.8987)	(0.8769)	(0.0057)	(0.0061)	(0.0807)
50	2.2455	1.2559	0.1601	0.1613	0.7589
	(0.6161)	(0.6210)	(0.0028)	(0.0029)	(0.0498)
75	2.2212	2.2788	0.1587	0.1589	0.7575
	(0.4142)	(0.4254)	(0.0020)	(0.0022)	(0.0298)
100	2.1178	2.1005	0.1571	0.1511	0.7512
	(0.3011)	(0.2987)	(0.0014)	(0.0012)	(0.0268)

Table 6: The average ML estimates and the associated MSEs of r_1, r_2, θ_1, θ_2 and p

Some of the points are quite clear from the simulation experiments. It is observed that in all the cases as the sample size increases, the biases and MSEs decrease. It indicates the consistency properties of the MLEs in this case also. Comparing Tables 4 and 5 it is observed that as r_1 and r_2 decrease the biases and MSEs of the MLEs of θ_1, θ_2 and p increase. Similarly, comparing Tables 4 and 6, it is observed that as p changes the biases and MSEs of the MLEs of the other parameters do not change significantly. Finally comparing 4 and 7, it is observed that as θ_1 and θ_2 increases, then the biases and MSEs of the MLEs of r_1, r_2 and p increase.

n	$r_1 = 2$	$r_2 = 2$	$\theta_1 = 0.25$	$\theta_2 = 0.25$	$p = 0.95$
25	2.2392	2.2488	0.2622	0.2634	0.9601
	(1.2314)	(1.2715)	(0.0098)	(0.0096)	(0.0068)
50	2.2229	2.2312	0.2601	0.2605	0.9578
	(0.8017)	(0.8119)	(0.0079)	(0.0081)	(0.0038)
75	2.2119	2.2112	0.2598	0.2578	0.9534
	(0.6110)	(0.6060)	(0.0053)	(0.0051)	(0.0024)
100	2.2013	2.2011	0.2574	0.2569	0.9501
	(0.4516)	(0.4489)	(0.0037)	(0.0033)	(0.0015)

Table 7: The average ML estimates and the associated MSEs of r_1, r_2, θ_1, θ_2 and p

7 Data Analysis

In this section we present the analysis of one data set to see how the new model and the proposed EM and NEM algorithms work in practice. The data set represents the Italian Series A football match score data between 'ACF Firontina' (X) and 'Juventus' (Y) during 1990 to 2005. The data set is presented in Table 8. It is presented in the contingency table form in Table 9.

We would like to use both BPG and BNBG to analyze this data set. Some of the basic statistics of the data set are presented below. The sample means, variances and correlation are

$$\mu_x = 1.3846, \quad \sigma_x^2 = 1.7751, \quad \mu_y = 1.6923, \quad \sigma_y^2 = 2.2130, \quad r_{x,y} = 0.1179,$$

respectively. Based on the BPG model it can be easily seen that the MMEs of the unknown parameters are

$$\widetilde{p} = 1 - r_{x,y} = 0.8821, \quad \widetilde{\lambda}_1 = \mu_x(1 - r_{x,y}) = 1.2214, \quad \widetilde{\lambda}_2 = \mu_y(1 - r_{x,y}) = 1.4928.$$

We have used these initial estimators to compute the MLEs based on the proposed EM algorithm. We have stopped the EM algorithm if the absolute difference between two consecutive log-likelihood values is less than 10^{-6}, and the algorithm stops after 11 iterations.

Obs.	ACF Firontina (X)	Juventus (Y)	Obs.	ACF Firontina (X)	Juventus (Y)
1	1	2	14	1	2
2	0	0	15	1	1
3	1	1	16	1	3
4	2	2	17	3	3
5	1	1	18	0	1
6	0	1	19	1	1
7	1	1	20	1	2
8	3	2	21	1	0
9	1	1	22	3	0
10	2	1	23	1	2
11	1	2	24	1	1
12	3	3	25	0	1
13	0	1	26	0	1

Table 8: UEFA Champion's League data

$X \downarrow Y \rightarrow$	0	1	2	3	Total
0	1	5	0	0	6
1	1	7	5	1	14
2	0	1	1	0	2
3	1	0	1	2	4
Total	3	13	7	3	26

Table 9: UEFA Champion's League data (Contingency table)

The MLEs, the associated 95% confidence intervals and the corresponding log-likelihood (LL) value are presented below.

$$\widehat{p} = 0.8315(\mp 0.0387), \quad \widehat{\lambda}_1 = 1.2174(\mp 0.2541), \quad \widehat{\lambda}_2 = 1.4928(\mp 0.3218), \quad LL = -52.1342.$$

The expected frequencies for each cell based on the fitted BPG distribution are provided in Table 10. The observed chi-square value is 16.11, with the p value greater than 0.20 for the χ^2 distribution with 14 degrees of freedom. Hence, it implies that BPG provides a good fit to the data set.

$X \downarrow Y \rightarrow$	0	1	2	3
0	1.30	2.25	1.30	0.48
1	1.68	5.75	2.45	0.90
2	0.62	1.83	2.49	0.48
3	0.28	0.69	0.43	0.76

Table 10: Expected cell frequencies based on fitted BPG

Based on the BNBG model the MMEs of the unknown parameters are

$$\widetilde{p} = 1 - \frac{r_{x,y}\sigma_x\sigma_y}{\mu_x\mu_y} = 0.9003, \quad \widetilde{\theta}_1 = 1 - \frac{\mu_x}{\sigma_x^2 - (1-\widetilde{p})\mu_x^2} = 0.1259, \quad \widetilde{r}_1 = \frac{\mu_x\widetilde{p}(1-\widetilde{\theta}_1)}{\widetilde{\theta}_1} = 8.6546,$$

$$\widetilde{\theta}_2 = 1 - \frac{\mu_y}{\sigma_y^2 - (1-\widetilde{p})\mu_y^2} = 0.1221, \quad \widetilde{r}_2 = \frac{\mu_y\widetilde{p}(1-\widetilde{\theta}_2)}{\widetilde{\theta}_2} = 10.9545.$$

We have used the NEM algorithm in this case to compute the MLEs of the unknown parameters. We have used the same stopping criterion as before, and the algorithm stops after 23 iterations. The MLEs, the associated 95% confidence intervals and the corresponding LL value are provided below.

$$\widehat{p} = 0.8723(\mp 0.0198), \quad \widehat{\theta}_1 = 0.1316(\mp 0.0027), \quad \widehat{\theta}_2 = 0.1258(\mp 0.0021)$$

$$\widehat{r}_1 = 8.7214(\mp 1.5926) \quad \widehat{r}_2 = 10.7895(\mp 2.1657), \quad LL = -48.2134.$$

The expected frequencies for each cell based on the fitted BNBG distribution are provided in Table 11. The observed chi-square value is 7.35, with the p value greater than 0.40 for the χ^2 distribution with 14 degrees of freedom. Hence, it implies that BNBG provides a good fit to the data set.

$X \downarrow Y \rightarrow$	0	1	2	3
0	1.24	3.76	1.24	0.33
1	1.28	5.98	3.47	1.03
2	0.53	1.49	1.99	0.67
3	0.48	0.38	0.48	0.96

Table 11: Expected cell frequencies based on fitted BNBG

For comparison purposes we would like to examine whether bivariate Poisson distribution provides a better fit or not to this data set. We have used the following joint PMF of the bivariate Poisson distribution with parameters λ_1, λ_2 and λ_3.

$$P(X = i, Y = j) = \sum_{k=0}^{\min\{i,j\}} \frac{e^{-\lambda_1} \lambda_1^{i-k}}{(i-k)!} \times \frac{e^{-\lambda_2} \lambda_2^{j-k}}{(j-k)!} \times \frac{e^{-\lambda_3} \lambda_3^{k}}{k!} \qquad i, j \in \mathbb{N}_0.$$

The MLEs of λ_1, λ_2 and λ_3 are: $\widehat{\lambda}_1 = 0.8089$, $\widehat{\lambda}_2 = 0.9737$ and $\widehat{\lambda}_3 = 0.5643$. The associated log-likelihood value is -53.3251. The expected frequencies for each cell based on the fitted bivariate Poisson distribution are provided in Table 12. The observed chi-square value and the p value are 21.8381 and 0.08, respectively. Hence, it implies that the bivariate Poisson distribution does not provide a good fit to the data set.

$X \downarrow Y \rightarrow$	0	1	2	3
0	2.48	2.42	1.18	0.38
1	2.01	3.36	2.32	0.98
2	0.81	1.92	1.89	1.05
3	0.22	0.67	0.87	0.64

Table 12: Expected cell frequencies based on fitted bivariate Poisson

8 Conclusions

In this paper we have proposed a very general bivariate discrete distributions which is a very flexible class of distributions. Due to presence of an extra parameter, the proposed class of distributions is more flexible than the base distribution functions. It can take variety of shapes, and it can be both over dispersed as well as under dispersed depending on the parameters. We have discussed several properties of the proposed class of distributions and consider two special cases, namely BPG and BNBG distributions. It is observed that both BPG and BNBG are infinitely divisible and they have some interesting physical interpretations also. We have proposed to use EM and NEM algorithms to compute the MLEs of the unknown parameters. The MMEs have been used to obtain initial estimates, and they perform quite well. We have analyzed one data set to see how the model and the proposed algorithms perform. It is observed that the model is quite flexible and it can be used for data analysis purposes, and also the proposed EM and NEM algorithms can be implemented easily in practice and they perform quite well. Although, in this paper we have proposed the bivariate class of distributions, the method can be extended even for the multivariate case also.

Another natural question is how to choose the correct bivariate discrete model between say BPG and BNBG models. It is not a very easy question. The chi-square test definitely can be used as we have used in our data analysis purposes. Alternatively, may be the mutual entropy be considered to explore the possibility of choosing the correct model. It has not pursued here. More work is needed along that direction.

ACKNOWLEDGEMENTS: The author would like to than the unknown reviewers for providing constructive suggestions to improve the paper significantly.

APPENDIX A: EXACT EXPRESSIONS OF $C(a, j)$

Note that for

$$C(a,0) = \sum_{k=1}^{\infty} e^{-ak} = \frac{e^{-a}}{1-e^{-a}} = \frac{1}{e^a - 1}.$$

To compute $C(a, 1)$, first observe that

$$C(a,1) = \sum_{k=1}^{\infty} e^{-ak} k = \frac{1}{e^a} + \frac{2}{e^{2a}} + \frac{3}{e^{3a}} + \cdots$$

and

$$e^a C(a,1) = 1 + \frac{2}{e^a} + \frac{3}{e^{2a}} + \frac{4}{e^{3a}} + \cdots.$$

Hence

$$(e^a-1)C(a,1) = 1 + \frac{1}{e^a} + \frac{1}{e^{2a}} + \frac{1}{e^{3a}} + \cdots = \frac{e^a}{e^a-1}.$$

Therefore

$$C(a,1) = \frac{e^a}{(e^a-1)^2}.$$

Now to compute $C(a,2)$, note that

$$C(a,2) = \sum_{k=0}^{\infty} e^{-ak}k^2 = \sum_{k=0}^{\infty} e^{-ak}k(k-1) + \sum_{k=0}^{\infty} e^{-ak}k.$$

If we denote $S = \sum_{k=0}^{\infty} e^{-ak}k(k-1)$, then

$$S = \frac{2\cdot 1}{e^{2a}} + \frac{3\cdot 2}{e^{3a}} + \frac{4\cdot 3}{e^{4a}} + \cdots$$

and

$$e^a S = \frac{2\cdot 1}{e^{a}} + \frac{3\cdot 2}{e^{2a}} + \frac{4\cdot 3}{e^{3a}} + \cdots.$$

Hence

$$S(e^a-1) = \frac{2\cdot 1}{e^{a}} + \frac{2\cdot 2}{e^{2a}} + \frac{2\cdot 3}{e^{3a}} + \cdots = 2\sum_{k=1}^{\infty} ke^{-ak} = \frac{2e^a}{(e^a-1)^2}.$$

Therefore,

$$C(a,2) = \frac{2e^a}{(e^a-1)^3} + \frac{e^a}{(e^a-1)^2} = \frac{e^{2a}+e^a}{(e^a-1)^3}.$$

We will use the following notations:

$$S_0(a) = \sum_{k=0}^{\infty} ke^{-ka}, \quad S_1(a) = \sum_{k=0}^{\infty} k(k-1)e^{-ka}, \ldots, S_m(a) = \sum_{k=0}^{\infty} k(k-1)\cdots(k-m)e^{-ka}.$$

Then using the fact

$$e^a S_m(a) = \sum_{k=0}^{\infty} k(k-1)\cdots(k-m)e^{-(k-1)a}.$$

We can easily obtain the following relation

$$S_m(a)(e^a-1) = (m+1)S_{m-1}(a).$$

Further note that if we denote

$$k^m = C_{0m}k(k-1)\cdots(k-m+1)+C_{1m}k(k-1)\cdots(k-m+1)+\cdots+C_{m-2,m}k(k-1)+C_{m-1,m}k,$$

then C_{0m}, $C_{1m}\cdots, C_{m-1,m}$ can be obtained recursively from the following set of linear equations. $C_{0m} = 1$ and

$$\begin{aligned} -C_{0m}\sum_{1\le i_1\le m-1} i_1 + C_{1m} &= 0 \\ C_{0m}\sum_{1\le i_1<i_2\le m-1} i_1i_2 - C_{1m}\sum_{1\le i_1\le m-2} i_1 + C_{2m} &= 0 \\ -C_{0m}\sum_{1\le i_1<i_2<i_3\le m-1} i_1i_2i_3 + C_{1m}\sum_{1\le i_1<i_2\le m-2} i_1i_2 - C_{2m}\sum_{1\le i_1\le m-3} i_1 + C_{3m} &= 0 \\ &\vdots \\ (-1)^{m-1}C_{0m}\prod_{i=1}^{m-1} i(-1)^{m-2}C_{1m}\prod_{i=1}^{m-2} i(-1)^{m-3}C_{2m}\prod_{i=1}^{m-3} i + \cdots - C_{m-2,m} + C_{m-1,m} &= 0. \end{aligned}$$

If we use the following notations for $n < m$;

$$a_{nm} = \sum_{1\le i_1<i_2<\ldots<i_n\le m} i_1i_2\cdots i_n,$$

$a_{mm} = \prod_{i=1}^{m} i$, then clearly

$$a_{n,m+1} = a_{n,m} + (m+1)a_{n-1,m},$$

and we obtain

$$\begin{aligned} C_{1m} &= a_{1,m-1} \\ C_{2m} &= C_{1m}a_{1,m-2} - a_{2,m-1} \\ C_{3m} &= C_{2m}a_{1,m-3} - C_{1m}a_{2,m-2} + a_{3,m-1} \\ \vdots &= \vdots \\ C_{m-1,m} &= C_{m-2,m}a_{11} - C_{m-3,m}a_{2,2} + \ldots (-1)^{m-2}a_{m-1,m-1}. \end{aligned}$$

Since we have

$$C(a,m) = C_{0m}S_{m-1}(a) + C_{1m}S_{m-2}(a) + \ldots + C_{m-1,m}S_0(a),$$

we can obtain recursively $C(a, m+1)$ from $C(a, m)$.

Appendix B: Expressions of $P(N = n|X = x, Y = y)$

In this appendix we provide the expressions of $P(N = n|X = x, Y = y)$ for both BPG and BNBG models. Suppose $(X, Y) \sim \text{BPG}(\lambda_1, \lambda_2, p)$, then

$$\begin{aligned} P(N = n|X = x, Y = y) &= \frac{P(X = x, Y = y|N = n)P(N = n)}{P(X = x, Y = y)} \\ &= \frac{n^{x+y}e^{-n(\lambda_1+\lambda_2)}(1-p)^n}{C(\lambda_1 + \lambda_2 - \ln(1-p), x+y)}. \end{aligned}$$

Now to compute $\arg\max_n P(N = n|X = x, Y = y)$, we consider

$$g(n) = \frac{P(N = n+1|X = x, Y = y)}{P(N = n|X = x, Y = y)} = \left(\frac{n+1}{n}\right)^{x+y} e^{-(\lambda_1+\lambda_2)}(1-p).$$

It is immediate that either $g(n)$ is a decreasing function or it is an unimodal function and if $n^* = \arg\max_n P(N = n|X = x, Y = y)$, then n^* is the smallest integer greater than

$$\left(\left[\frac{e^{\lambda_1+\lambda_2}}{1-p}\right]^{1/(x+y)} - 1\right)^{-1}.$$

Now suppose $(X, Y) \sim \text{BNBG}(r_1, \theta_1, r_2, \theta_2, p)$, then

$$\begin{aligned} P(N = n|X = x, Y = y) &= \frac{P(X = x, Y = y|N = n)P(N = n)}{P(X = x, Y = y)} \\ &= \frac{(1-p)^n(1-\theta_1)^{nr_1}(1-\theta_2)^{nr_2}}{D(r_1, r_2, \theta_1, \theta_2, x, y, p)} \times \frac{\Gamma(x+nr_1)}{x!\Gamma(nr_1)}\frac{\Gamma(y+nr_2)}{y!\Gamma(nr_2)}. \end{aligned}$$

Hence,

$$\begin{aligned} g(n) &= \frac{P(N = n+1|X = x, Y = y)}{P(N = n|X = x, Y = y)} \\ &= (1-p)(1-\theta_1)^{r_1}(1-\theta_2)^{r_2} \times \\ &\quad \frac{\Gamma(x+(n+1)r_1)\Gamma(y+(n+1)r_2)}{\Gamma((n+1)r_2)\Gamma((n+1)r_2)} \times \frac{\Gamma(nr_1)\Gamma(nr_2)}{\Gamma(x+nr_1)\Gamma(y+nr_2)}. \end{aligned}$$

In this case because of the complicated nature of $g(n)$, it is not possible to show that $g(n)$ has a unique maximum. But in all our numerical experiments it has been observed that $g(n)$ has a unique maximum. We have chosen n^* to be the minimum n, such that $g(n) < 1$.

Appendix C: EM Algorithm for Non-Identical NB Distribution

In this Appendix we will show that if $X_i \sim \text{NB}(n_i r, \theta)$, X_i's are independent, n_i's are known for $i = 1, \ldots, m$, then how to obtain MLEs of r and θ, based on a sample $\{x_1, \ldots, x_m\}$. In this case we will be using an EM algorithm very similar to Adamidis [1]. We use the following notation: $\alpha = -(\ln(1-\theta))^{-1}$, $r = \alpha\lambda$ and provide the algorithm to compute the MLEs of λ and θ. It is observed that in this case at each E-step, the corresponding M-step can be obtained in explicit forms. Using the same notation as in Adamidis [1], it can be easily seen that for $i = 1, \ldots, m$,

$$X_i \stackrel{d}{=} \sum_{j=1}^{M_i} Y_{ij},$$

here Y_{ij}'s are i.i.d. logarithmic series distribution (LSD) with PDF

$$f_{LSD}(y;\theta) = \frac{\alpha\theta^y}{y}; \quad y \in \mathbb{N} = \{1, 2, 3, \ldots\},$$

$M_i \sim \text{PO}(n_i\lambda)$ and all the random variables are independently distributed. Further, if Z_{ij}'s are i.i.d. random variables with PDF

$$f(z;\theta) = \alpha^{-1}\frac{(1-\theta)^z}{\theta}; \quad z \in (0,1),$$

and 0, otherwise, then the log-likelihood function of the 'complete data' $(Y_{ij}, Z_{ij}, M_i; i = 1, \ldots, m, j = 1, \ldots, M_i)$, without the additive constant can be written as

$$l^*(\alpha,\lambda) = -\lambda\widetilde{n} + \ln\lambda\sum_{i=1}^{m} m_i + \ln\theta\left[\sum_{i=1}^{m}\sum_{j=1}^{m_i} y_{ij} - \sum_{i=1}^{m} m_i\right] + \ln(1-\theta)\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i} z_{ij}\right],$$

here $\widetilde{n} = \sum_{i=1}^{m} n_i$. Hence, the MLEs of λ and θ based on the complete observations can be easily obtained as

$$\widehat{\lambda} = \frac{\sum_{i=1}^{m} m_i}{\widetilde{n}} \quad \text{and} \quad \widehat{\theta} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{m_i} y_{ij} - \sum_{i=1}^{m} m_i}{\sum_{i=1}^{m}\sum_{j=1}^{m_i} y_{ij} + \sum_{i=1}^{m}\sum_{j=1}^{m_i} z_{ij} - \sum_{i=1}^{m} m_i}.$$

Hence, following the same way as in Adamidis [1], it can be easily seen that if at the k-th stage the estimates of λ and θ are $\lambda^{(k)}$ and $\theta^{(k)}$, respectively, and if we denote

$$a_i(\alpha, \lambda) = n_i \alpha \lambda \sum_{l=1}^{x_i} (\alpha n_i \lambda + l - 1)^{-1},$$

then

$$\lambda^{(k+1)} = \frac{\sum_{i=1}^{m} a_i(\alpha^{(k)}, \lambda^{(k)})}{\widetilde{n}} \quad \text{and} \quad \theta^{(k+1)} = \frac{\widetilde{x} - \sum_{i=1}^{m} a_i(\alpha^{(k)}, \lambda^{(k)})}{\widetilde{x} + \sum_{i=1}^{m} a_i(\alpha^{(k)}, \lambda^{(k)}) \left(\frac{\alpha^{(k)}(1-\theta^{(k)})}{\theta^{(k)}} - 1\right)}.$$

Here $\widetilde{x} = \sum_{i=1}^{m} x_i$.

References

[1] Adamidis, K. (1999), "An EM algorithm for estimating negative binomial parameters", *Australian and New Zealand Journal of Statistics*, vol. 41, 213 – 221.

[2] Barreto-Souza, W. (2012), "Bivariate gamma-geometric law and its induced Levy process", *Journal of Multivariate Analysis*, vol. 109, 130 – 145

[3] Basu, A.P. and Dhar, S.K. (1995), "Bivariate geometric distribution", *Journal of Applied Statistical Sciences*, vol. 2, 33 – 34.

[4] Campbell, J.T. (1934), "The Poisson correlation function", *Proceedings of the Edinburgh Mathematical Society*, Series 2, vol. 4, 18 – 26.

[5] Chahkandi, M. and Ganjali, M. (2009), "On some lifetime distributions with decreasing failure rate", *Computational Statistics and Data Analysis*, vol. 53, 4433 – 4440.

[6] Davis, C.S. (2002), *Statistical methods for the analysis of repeated measurements*, Springer-Verlag, New York.

[7] Holgate, B. (1964), "Estimation for the bivariate Poisson distribution", *Biometrika*, vol. 51, 241 – 245.

[8] Jayakumar, K. and Mundassery, D.A. (2007), "On a bivariate geometric distribution", *Statistica*, Vol. LXVII, 389 – 404.

[9] Kemp, A.W. (2013), "New discrete Appell and Humbert distributions with relevance to bivariate accident data", *Journal of Multivariate Analysis*, vol. 113, 2–6.

[10] Kocherlakota, S. and Kocherlakota, K. (1992), *Bivariate discrete distributions*, Marcel and Dekker, New York.

[11] Kocherlakota, S. (1995), "Discrete bivariate weighted distributions under multiplicative weight function", *Communications in Statistics - Theory and Methods*, vol. 24, 533 – 551.

[12] Kostadinova, K. and Minkova, L. (2014), "On a bivariate Poisson negative binomial process", *Biomath*, vol. 3, 1 – 6.

[13] Kumar, C.S. (2008) "A unified approach to bivariate discrete data", *Metrika*, vol. 67, 113 – 123.

[14] Kundu, D. (2014), "Geometric skewed normal distribution", *Sankhya, Ser. B.*, vol. 76, 167 – 189.

[15] Kundu, D. (2017), "Multivariate geometric skew normal distribution", *Statistics*, vol. 51, 1377 – 1397.

[16] Kundu, D. and Nekoukhou, V. (2018), "Univariate and bivariate geometric discrete generalized exponential distributions", *Journal of Statistical Theory and Practice*, vol. 12, 595 - 614.

[17] Kozubowski, T.J., Panorska, A.K., Podgorski, K. (2008), "A bivariate Levy process with negative binomial and gamma marginals", *Journal of Multivariate Analysis*, vol. 199, 1418 – 1437.

[18] Kozubowski, T.J., Panorska, A.K., Qeadan, F. (2011), "A new multivariate model involving geometric sums and maxima of exponentials", *Journal of Statistical Planning and Inference*, vol. 141, 2353 – 2367.

[19] Lee, H. and Cha, J.H. (2014), "On construction of general class of bivariate distributions", *Journal of Multivariate Analysis*, vol. 127, 151 – 159.

[20] Lee, H. and Cha, J.H. (2015), "On two general classes of discrete bivariate distributions", *American Statistician*, vol. 69, 221 – 230.

[21] Minkova, L.D. and Balakrishnan, N. (2014) "Bivariate Pólya-Aeppli distribution", *Communications in Statistics - Theory and Methods*, vol. 43, 5026 – 5038.

[22] Ozel, G. (2011), "On certain properties of a class of bivariate compound Poisson distribution and an application to earthquake data", *Revista Colombiana de Estadstica*, vol. 34, 545 – 566.

[23] Ye, Y. (2015), "Likelihood inference for type-I bivariate Pólya-Aeppli distribution", Thesis, McMaster University.

[24] Nekoukhou, V. and Kundu, D. (2017), "Bivariate discrete generalized exponential distribution", *Statistics*, vol. 51, 1143 – 1158.

[25] Piperigou, V.E. and Papageorgiou, H. (2003), "On bivariate discrete distributions: a unified treatment", *Metrika*, vol. 58, 221 – 233.

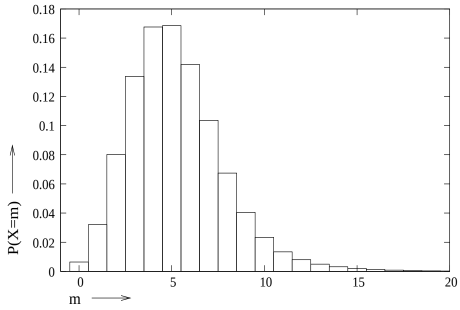

Figure 1: The PMF of a UPG distribution when $\lambda = 5$ and $p = 0.95$.

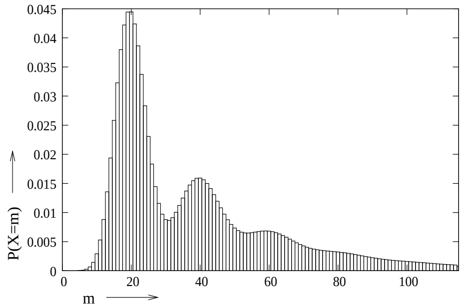

Figure 2: The PMF of a UPG distribution when $\lambda = 20$ and $p = 0.50$.

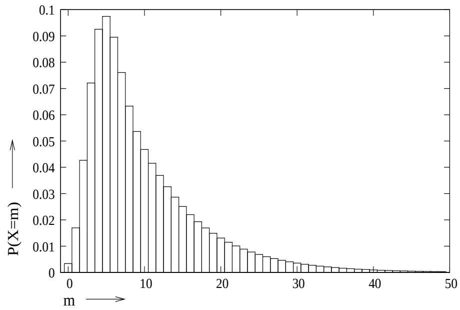

Figure 3: The PMF of a UPG distribution when $\lambda = 5$ and $p = 0.50$.

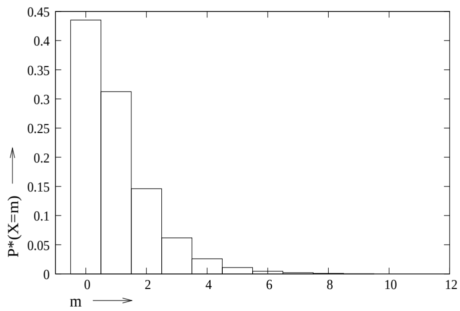

Figure 4: The PMF of a UPG distribution when $\lambda = 0.50$ and $p = 0.50$.

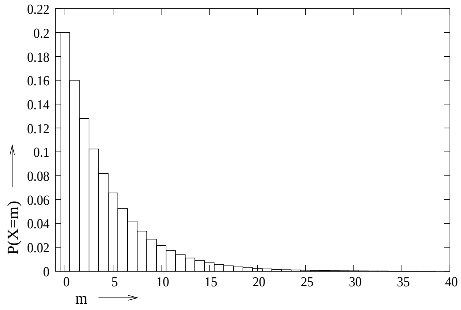

Figure 5: The PMF of a UNBG distribution when $r = 1$, $\theta = 0.50$ and $p = 0.25$.

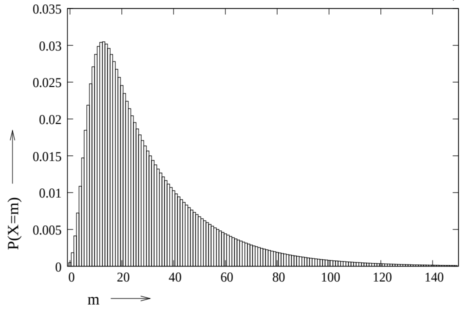

Figure 6: The PMF of a UNBG distribution when $r = 5$, $\theta = 0.75$ and $p = 0.50$.

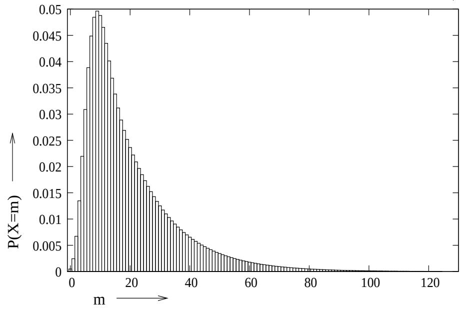

Figure 7: The PMF of a UNBG distribution when $r = 10$, $\theta = 0.50$ and $p = 0.50$.

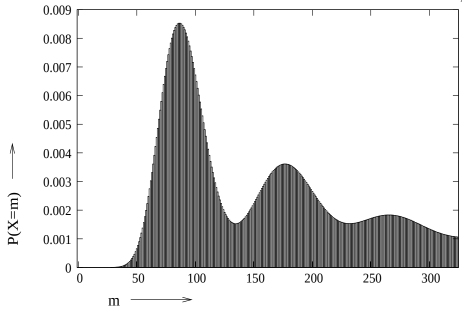

Figure 8: The PMF of a UNBG distribution when $r = 30$, $\theta = 0.75$ and $p = 0.40$.