COM-Poisson Pascal Distribution

J. Priyadharshini, Research Scholar, Department of Mathematics, Nehru Memorial College, Trichy, Tamil Nadu, priyajaya895@gmail.com

V. Saavithri, Assistant Professor, Department of Mathematics, Nehru Memorial College, Trichy, Tamil Nadu, saavithriramani@gmail.com

Abstract **In the early stage of the twentieth century, very few discrete distributions (binomial and Poisson) were commonly used in modeling. These distributions fail to model over-dispersed and under-dispersed data. From 1920 to 1970, new count distributions such as 'contagious' COM-Poisson have been introduced. 'Contagious' distribution, which includes the negative binomial, Hermite, Neyman, Thomas Polya-Aeppli and Poisson-Pascal distributions can be expressed as mixture distributions or as stopped sum distributions. COM-Poisson distribution is a two parameter extension of Poisson distribution. In this paper, COM-Poisson Pascal distribution is proposed and its properties are studied. It is expressed as stopped sum distribution. It is a Compound COM-Poisson distribution with negative binomial compounding distribution. Also it is generalization of COM-Poisson Polya-Aeppli distribution. The parameters are estimated by the method of profile likelihood estimation. The traffic accidents and fatalities data are analyzed using this distribution.**

Keywords — Poisson distribution, Pascal (Negative Binomial) distribution, COM-Poisson Distribution, COM-Poisson Polya-Aeppli distribution, COM- Poisson Pascal distribution.

I. INTRODUCTION

The COM-Poisson distribution is a two-parameter extension of Poisson distribution, which is introduced by Conway & Maxwell [1] and revived by Galit Shmueli et al [10]. The COM-Poisson distribution generalizes Poisson, Bernoulli and geometric distributions and hence leads to the generalization of binomial and negative binomial distributions. It belongs to the exponential family and also to the parameter power series distribution family. This distribution is best suited for over and under dispersed data. Josemar Rodrigues et al [4] used this distribution in cure rate models.

Consider the combination of two independent distributions in a particular way. This process was called "generalization" by Feller [3]. The principal model for the process can be interpreted as the summation of observations from the distribution F_2, where the number of observations to be summed is determined by an observation from the distribution F_1. (ie, summation of F_2 observations is stopped by the value of the F_1 observation). COM-Poisson Pascal distribution comes under the stopped sum distributions.

The Polya-Aeppli distribution was described by Polya [8]; he ascribed the derivation of the distribution to his student in a Zurich thesis in 1924. The Polya – Aeppli distribution arises in a model formed by supposing that objects occur in clusters. The number of clusters has a Poisson distribution and the number of objects has a geometric distribution.

The Poisson-Pascal distribution was introduced in the context of the spatial distribution of plants by Skellam [13], who called it a generalized Polya-Aeppli distribution. Priyadharshini and Saavithri [9] proposed a new distribution called "COM-Poisson Polya-Aeppli Distribution" which is a combination of COM-Poisson and geometric distribution. In 2007, Meintanis [6] fitted certain bivariate distributions to traffic accidents data.

In this paper, COM-Poisson Pascal distribution is proposedand its properties are studied. It is a Compound COM-Poisson distribution with negative binomial compounding distribution. It is a generalization of COM-Poisson Polya–Aeppli distribution. The parameters are estimated by the method of profile likelihood estimation. The traffic accidents and fatalities data are analyzed using this distribution.

This paper is organized as follows: In section 2, mean and variance of COM-Poisson distribution using probability generating function are studied. The COM-Poisson Polya-Aeppli distribution and some of its properties are given in section 3. In section 4, the COM-Poisson Pascal distribution is defined and its mean and variance are obtained. The parameters are estimated in section 5. In section 6, traffic accidents and fatalities data are analyzed. Section 7 concludes this paper.

 DOI : 10.18231/2454-9150.2018.1069

II. COM-Poisson Distribution

2.1 Probability density function

The probability density function of COM-Poisson distribution [10] is

$$P(X=x)=\frac{\lambda^x}{(x!)^\nu}\frac{1}{Z(\lambda,\nu)}, x=0,1,2,\ldots \qquad \ldots.. (2.1.1)$$

where $Z(\lambda,\nu)=\sum_{j=0}^{\infty}\frac{\lambda^j}{(j!)^\nu}$ for $\lambda>0$ and $\nu\geq 0$.

2.2 Special Cases

- When $\nu=1$, $Z(\lambda,\nu)=e^\lambda$, and this distribution is simply the Poisson (λ)
- As $\nu\to\infty, Z(\lambda,\nu)\to 1+\lambda$, and the distribution approaches a Bernoulli distribution with

$$P(X=1)=\frac{\lambda}{1+\lambda}$$

- When $\nu=0$ and $\lambda<1, Z(\lambda,\nu)$ is a geometric sum

$$Z(\lambda,\nu)=\sum_{j=0}^{\infty}\lambda^j=\frac{1}{1-\lambda}$$

and the distribution itself is geometric

$$P(X=x|\lambda,\nu)=\lambda^x(1-\lambda) \text{ for } x=0,1,2\ldots$$

- When $\nu=0$ and $\lambda\geq 1$, $Z(\lambda,\nu)$ does not converge and hence the distribution is undefined.

2.3 Properties

The probability generating function of COM-Poisson distribution is

$$G_X(s)=\frac{Z(\lambda s,\nu)}{Z(\lambda,\nu)} \qquad \ldots\ldots (2.3.1)$$

The mean and variance of the distribution are as follows [9]

$$Mean(X)=G'_X(1)=\frac{\lambda Z_\lambda(\lambda,\nu)}{Z(\lambda,\nu)}$$

$$Var(X)=G''_X(1)+G'_X(1)-[G'_X(1)]^2$$

$$=\frac{\lambda^2 Z_{\lambda\lambda}(\lambda,\nu)}{Z(\lambda,\nu)}+\frac{\lambda Z_\lambda(\lambda,\nu)}{Z(\lambda,\nu)}-\left[\frac{\lambda Z_\lambda(\lambda,\nu)}{Z(\lambda,\nu)}\right]^2$$

where $$Z_\lambda(\lambda,\nu)\equiv\frac{\partial}{\partial\lambda}[Z(\lambda,\nu)],$$

$$Z_{\lambda\lambda}(\lambda,\nu)\equiv\frac{\partial^2}{\partial\lambda^2}[Z(\lambda,\nu)]$$

The ratio between variance and mean to be,

$$\frac{Var(X)}{Mean(X)}=\frac{\lambda Z_{\lambda\lambda}(\lambda,\nu)}{Z_\lambda(\lambda,\nu)}-\frac{\lambda Z_\lambda(\lambda,\nu)}{Z(\lambda,\nu)}+1$$

III. COM-Poisson Polya-Aeppli Distribution

Suppose that the several events can happen simultaneously at an instant, and then there is a cluster of occurrences at a point.

Assume that there are Y independent random variables of the form X, and N denotes the sum of these random variables.

(ie) $N=X_1+X_2+\ldots+X_Y$

This random variable, N formed by compounding these two random variables X and Y gives the COM-Poisson Polya-Aeppli distribution with parameters λ,ν and ρ.

Its probability generating function is

$$G_N(s)=\frac{1}{Z(\lambda,\nu)}\sum_{j=0}^{\infty}\frac{1}{(j!)^\nu}\left[\left(\lambda s(1-\rho)\right)^j(1-\rho s)^{-j}\right] \qquad \ldots.. (3.1)$$

The probability mass function of N is

$$P(N=n)=\begin{cases}\dfrac{1}{Z(\lambda,\nu)} & \text{for } n=0\\ \dfrac{1}{Z(\lambda,\nu)}\displaystyle\sum_{j=0}^{\infty}\frac{1}{(j!)^\nu}\binom{n-1}{j-1}(\lambda(1-\rho))^j\rho^{n-j} & \text{for } n=1,2,\ldots\end{cases} \qquad \ldots\ldots(3.2)$$

The mean and variance are given by

$$Mean(N)=\frac{\lambda Z_\lambda(\lambda,\nu)}{(1-\rho)Z(\lambda,\nu)}$$

$$Var(N)=\frac{1}{(1-\rho)^2 Z(\lambda,\nu)}\left[\lambda^2 Z_{\lambda\lambda}(\lambda,\nu)+(1+\rho)\lambda Z_\lambda(\lambda,\nu)-\frac{[\lambda Z_\lambda(\lambda,\nu)]^2}{Z(\lambda,\nu)}\right]$$

The expression for ratio between variance and mean is

$$\frac{Var(N)}{Mean(N)}=\frac{\lambda Z_{\lambda\lambda}(\lambda,\nu)}{(1-\rho)Z_\lambda(\lambda,\nu)}-\frac{\lambda Z_\lambda(\lambda,\nu)}{(1-\rho)Z(\lambda,\nu)}+\frac{1+\rho}{1-\rho}$$

IV. COM-Poisson Pascal Distribution

4.1 Definition

Suppose that the several events can happen simultaneously at an instant, and then there is a cluster of occurrences at a point.

Assume that there are Y independent random variables of the form X, and N denotes the sum of these random variables.

(ie) $N=X_1+X_2+\ldots+X_Y$

COM-Poisson Pascal distribution is derived by assuming that

(i) X denotes the number of objects within a cluster and X follows Negative binomial (Pascal) distribution with parameters k and P.

(ie) $X\sim NB(k,P)$

(ii) Y denotes the number of clusters and Y follows COM-Poisson distribution with parameters λ and ν

(ie) $Y\sim COM-Poisson(\lambda,\nu)$

This random variable, N formed by compounding these two random variables X and Y gives the COM-Poisson Pascal distribution with parameters λ,ν,k and P.

1. The probability generating function of X is,

$$G_X(s)=(Q-Ps)^{-k} \qquad \ldots.. (4.1.1)$$

 DOI : 10.18231/2454-9150.2018.1069

where $Q = 1 + P$.

2. The probability generating function of COM-Poisson distribution is

$$G_Y(s) = \frac{Z(\lambda s, \nu)}{Z(\lambda, \nu)} \qquad \text{...... (4.1.2)}$$

The probability generating function of the random variable N can be derived as follows

$$G_N(s) = E(s^N) = E(s^{X_1 + X_2 + \dots + X_Y})$$

$$= \sum_{y=0}^{\infty} E(s^{X_1 + X_2 + \dots + X_Y} / Y = y)\, P(Y = y)$$

$$= \sum_{y=0}^{\infty} [E(s^x)]^y\, P(Y = y)$$

$$= G_Y(\, G_X(s))$$

$$= \frac{Z(\lambda\, G_X(s), \nu)}{Z(\lambda, \nu)}$$

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

$$= \frac{1}{Z(\lambda, \nu)} \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu} (Q - Ps)^{-kj}$$

Collecting the coefficient of s^n in the above series we get

$$P(N = n) = \frac{1}{Z(\lambda,\nu)} \left(\frac{P}{Q}\right)^n \sum_{j=1}^{\infty} \frac{(kj+n-1)!}{(kj-1)!n!} \frac{(\lambda Q^{-k})^j}{(j!)^\nu}$$

The probability mass function of N is

$$P(N = 0) = \frac{Z(\lambda Q^{-k},\ \nu)}{Z(\lambda,\nu)}$$

$$P(N = n) = \frac{1}{Z(\lambda,\nu)} \left(\frac{P}{Q}\right)^n \sum_{j=1}^{\infty} \frac{(kj+n-1)!}{(kj-1)!n!} \frac{(\lambda Q^{-k})^j}{(j!)^\nu} \quad \text{.... (4.1.3)}$$

where $\lambda > 0, \nu \geq 0,\ k > 1$ *and* $P \geq 0$.

4.2 Special cases

1. Replacing k by 1, P by $\frac{p}{1-p}$, λ by $\frac{\lambda_1}{p}$ in equation (4.1.3), the PGF of COM-Poisson Polya-Aeppli distribution is obtained as

$$\frac{z\left(\frac{\lambda s(1-p)}{(1-ps)}, \nu\right)}{Z(\lambda, \nu)}$$

2. When $\nu = 1$ from equation (4.1.3), we get Generalized Polya-Aeppli (Poisson Pascal) distribution with PGF

$$exp\{\lambda[(Q - Ps)^{-k} - 1]\}$$

3. Replacing ν by 1 k by 1, P by $\frac{p}{1-p}$, λ by $\frac{\lambda_1}{p}$ in equation (4.1.3), the PGF of Polya-Aeppli distribution

$$\exp\left\{\frac{\lambda s(1-p)}{(1-ps)} - 1\right\}$$

4.3 Limiting cases of the COM-Poisson Pascal Distribution

1. When $k \to \infty, P \to 0, Pk = \mu$, from equation (4.1.3), we get COM-Poisson Neyman Type A distribution with PGF

$$\frac{Z\big(\lambda(\exp(\mu(s-1))), \nu}{Z(\lambda, \nu)}$$

2. When $\nu = 1, k \to \infty, P \to 0, Pk = \mu$, from equation (4.1.3), we get Neyman Type A distribution with PGF

$$exp[\lambda(exp(\mu(s-1)) - 1)]$$

4.4 Properties of COM-Poisson Pascal distribution

The mean and variance of the distribution are given by

$$Mean(N) = \frac{\lambda\, kP\, Z_\lambda(\lambda, \nu)}{Z(\lambda, \nu)}$$

$$Var(N) = \frac{1}{Z(\lambda, \nu)} \left[\lambda^2 k^2 P^2 \left(Z_{\lambda\lambda}(\lambda, \nu) - \frac{[Z_y(\lambda, \nu)]^2}{Z(\lambda, \nu)}\right)\right] + \frac{1}{Z(\lambda, \nu)} [\lambda kp(Q + kP)\lambda Z_\lambda(\lambda, \nu)]$$

The expression for ratio between variance and mean is

$$\frac{Var(N)}{Mean(N)} = (Q + kP) + (\lambda kP) \left[\frac{\lambda Z_{\lambda\lambda}(\lambda, \nu)}{Z_\lambda(\lambda, \nu)} - \frac{\lambda Z_\lambda(\lambda, \nu)}{Z(\lambda, \nu)}\right]$$

V. PROFILE LIKELIHOOD ESTIMATION

Let $N_1, N_2, \dots, N_m$ be the samples following the COM-Poisson Pascal distribution with parameters $\lambda > 0, \nu \geq 0$, $k > 0$ and $P \geq 0$. The likelihood function of $N_1, N_2, \dots, N_m$ is

$$L = \prod_{i=1}^{m} P(N = N_i)$$

$$= \prod_{i=1}^{m} \frac{1}{Z(\lambda, \nu)} \left(\frac{P}{Q}\right)^{n_i} \sum_{j=1}^{\infty} \frac{(kj + n_i - 1)!}{(kj-1)!\, n_i!} \frac{(\lambda Q^{-k})^j}{(j!)^\nu}$$

$$= \frac{1}{[Z(\lambda, \nu)]^m} \prod_{i=1}^{m} \left(\frac{P}{1+P}\right)^{n_i} \prod_{i=1}^{m} \left[\sum_{j=1}^{\infty} \frac{(kj + n_i - 1)!}{(kj-1)!\, n_i!} \frac{(\lambda(1+P)^{-k})^j}{(j!)^\nu}\right]$$

The log likelihood function is

$$\log(L) = l = -m\log\left[\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}\right] + [\log(P) + \log(1+P)] \sum_{i=1}^{m} n_i + \sum_{i=1}^{m} \log\left[\sum_{j=1}^{\infty} \frac{(kj + n_i - 1)!}{(kj-1)!\, n_i!} \frac{(\lambda(1+P)^{-k})^j}{(j!)^\nu}\right]$$

Partially differentiating the above expression with respect to λ, ν and P and equating to zero, the following expressions are obtained.

$$\sum_{i=1}^{m} \frac{I_2(n_i)}{I_1(n_i)} - m\frac{J_2}{J_1} = 0$$

$$m\frac{J_3}{J_1} - \sum_{i=1}^{m} \frac{I_3(n_i)}{I_1(n_i)} = 0$$

$$\sum_{i=1}^{m} \frac{I_4(n_i) + I_5(n_i)}{I_2(n_i)} = 0$$

where

$$I_1(n_i) = \sum_{j=1}^{\infty} \frac{(kj + n_i - 1)!}{(kj-1)!\, n_i!} \frac{(\lambda(1+P)^{-k})^j}{(j!)^\nu}$$

$$I_2(n_i) = \sum_{j=1}^{\infty} \frac{(kj + n_i - 1)!}{(kj-1)!\, n_i!} \frac{(j\lambda^{j-1}(1+P)^{-kj})}{(j!)^\nu}$$

 DOI : 10.18231/2454-9150.2018.1069

$$I_3(n_i) = \sum_{j=1}^{\infty} \frac{(kj+n_i-1)!}{(kj-1)!\,n_i!} \frac{(\lambda(1+P)^{-k})^j \log(j!)}{(j!)^{\nu}}$$

$$I_4(n_i) = \sum_{j=1}^{\infty} \frac{(kj+n_i-1)!}{(kj-1)!\,n_i!} \frac{(\lambda(1+P)^{-k})^j \log(1+P)}{(j!)^{\nu}}$$

$$I_5(n_i) = \sum_{j=1}^{\infty} \frac{(\lambda(1+P)^{-k})^j}{(j!)^{\nu}} \left[\frac{(kj-1)!\,\Gamma(kj+ni)\psi(kj+n_i)}{(kj-1)^2!\,n_i!} - \frac{(kj+n_i-1)!\,\Gamma(kj)\psi(kj)}{(kj-1)^2!\,n_i!} \right]$$

$$J_1 = \sum_{j=1}^{\infty} \frac{\lambda^j}{(j!)^{\nu}}$$

$$J_2 = \sum_{j=1}^{\infty} \frac{j\lambda^{j-1}}{(j!)^{\nu}}$$

$$J_3 = \sum_{j=1}^{\infty} \frac{\lambda^j \log(j!)}{(j!)^{\nu}}$$

Then

$\max_{\lambda,\nu,k,P} logL(\lambda,\nu,k,P|n) = \max_k[\max_{\lambda,\nu,,P} logL(\lambda,\nu,k,P|n)]$ is calculated.

The estimators $\hat{\lambda}, \hat{\nu}, \hat{k}, \hat{P}$ are estimated numerically using MATLAB commands.

VI. Data Analysis

In this section, two sets of traffic accidents and fatalities data are analyzed

6.1 Data set 1

The following table gives the total Sunday accidents (left entry) and the corresponding number of fatalities (right entry) recorded in the Groningen region for each month during the years 1997-2004 [6]

Month	1997		1998		1999		2000		2001		2001		2003		2004	
January	6	0	6	0	13	1	11	0	8	0	8	0	11	4	2	0
February	10	0	10	1	7	0	4	0	8	1	8	0	9	0	2	0
March	7	0	13	4	8	0	10	0	6	0	12	0	9	0	3	0
April	11	0	5	0	14	1	15	1	9	0	10	1	7	1	1	1
May	12	0	17	2	13	0	18	0	13	2	11	0	12	1	5	0
June	21	1	19	0	14	0	21	1	12	3	12	1	13	0	7	2
July	15	0	10	0	14	0	11	1	10	2	4	0	8	0	1	0
August	11	1	11	1	10	0	8	0	9	0	14	1	6	0	5	0
September	7	0	11	0	7	0	9	0	22	1	16	1	7	0	8	1
October	11	2	13	1	16	1	14	0	15	1	8	1	6	1	2	0
November	15	1	17	1	13	0	13	0	6	0	9	1	11	1	1	0
December	5	0	7	0	10	1	11	0	10	0	8	0	5	0	2	0

Let Y be the number of Sunday's that accidents occur in Groningen between the years 1997-2004.
$X_i, i = 1,2,\ldots,$ be the number of fatalities of i^{th} accident and N be the total number of fatalities from the year 1997 to 2004.

Assuming that the number of accidents follows COM-Poisson distribution, the estimated values of λ and ν are 2.6027 & 0.4320 respectively.

Number of accidents	**Observed frequency**	**Expected frequency COM-Poisson**
1	3	1.0089
2	4	1.9462
3	1	3.1509
4	2	4.5048
5	5	5.4885
6	6	7.0175
7	8	7.8775
8	10	8.3470
9	6	8.4055
10	9	8.0876
11	11	7.4677
12	5	6.6409
13	8	5.7047
14	5	4.7461
15	4	3.8325
16	2	3.0097
17	2	2.3024
18	1	1.7183
19	1	1.2528
20	0	0.8934
21	2	0.6238
22	1	0.4269
≥23	0	1.1863
Total	96	96

Fitting the Negative binomial distribution to the number of fatalities, the parameters are obtained as $k = 2.2639, P = 0.2300$

Number of fatalities	**Observed Frequency**	**Expected Frequency Negative Binomial**
0	59	60.0793
1	29	25.4346
2	5	7.7620
3	1	2.0630
4	2	0.5077
≥ 5	0	0.1534
Total	96	96

 DOI : 10.18231/2454-9150.2018.1069

International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-04, Issue-08, Nov 2018

6.2 Data set 2

Here the data is taken from fatal crashes and fatalities calendar 2016 of Texas department of transportation, Austin. The one day accidents (left entry) and the corresponding number of fatalities (right entry) for each month during the years 2016.

Date	Jan		Feb		Mar		Apr		May		June		July		Aug		Sep		Oct		Nov		Dec	
1	10	12	5	5	6	9	8	8	13	13	2	2	8	8	11	11	6	6	10	14	10	11	6	6
2	8	8	7	7	11	11	9	12	5	5	5	5	11	11	12	12	14	15	16	19	10	10	15	17
3	10	10	5	5	6	7	11	12	4	4	7	9	11	12	10	10	12	15	6	7	10	15	8	13
4	9	10	9	11	12	15	13	15	8	11	8	9	10	10	1	1	11	13	8	8	8	8	13	14
5	6	6	9	9	10	13	8	11	7	7	9	11	9	11	12	13	7	8	5	5	13	13	11	12
6	7	7	13	14	13	14	9	9	16	16	8	9	9	9	16	17	6	8	11	11	14	17	8	9
7	11	11	12	12	6	6	6	7	9	14	10	10	4	4	12	12	8	8	10	11	10	11	6	6
8	11	13	6	6	5	5	9	9	13	18	14	15	13	13	11	12	6	6	10	12	6	6	8	10
9	12	14	9	13	8	9	8	11	10	13	6	8	13	14	10	11	9	9	12	12	7	7	7	7
10	8	8	9	10	10	12	11	12	10	10	11	11	13	16	10	13	10	10	13	17	7	7	9	9
11	5	5	10	12	6	6	6	7	9	11	13	14	4	4	12	13	13	15	4	4	11	11	9	9
12	12	13	14	14	11	11	10	12	8	8	16	18	4	5	13	13	7	8	10	10	8	8	9	10
13	7	7	17	21	9	10	8	9	13	14	13	15	4	4	17	19	4	6	6	6	12	12	7	7
14	7	9	12	12	5	5	4	4	13	21	4	4	11	11	9	11	8	8	11	14	11	11	7	8
15	6	7	9	10	7	9	15	15	9	9	13	13	15	15	9	10	8	10	17	18	5	5	3	3
16	7	9	8	8	8	8	10	10	11	11	13	14	6	7	2	2	14	15	8	8	12	12	8	8
17	8	8	9	9	8	9	6	8	3	3	5	6	12	12	7	7	9	10	12	13	12	14	13	15
18	3	3	12	15	12	12	7	8	5	5	9	12	9	11	14	14	12	12	11	11	11	11	13	15
19	6	7	16	20	11	11	11	14	5	5	13	17	4	5	6	6	13	13	6	6	13	15	3	3
20	5	6	10	11	7	9	3	3	14	14	7	7	7	9	10	14	9	10	8	8	13	16	11	12
21	11	12	21	23	10	11	7	8	17	22	17	18	14	18	11	11	11	11	14	16	9	9	12	15
22	11	12	6	6	6	6	12	14	15	16	7	7	8	9	5	5	14	14	16	18	13	14	6	6
23	11	11	5	5	5	5	13	15	8	11	8	9	8	8	3	3	4	4	9	10	11	11	11	11
24	7	7	3	4	13	13	7	9	4	5	7	8	14	16	2	2	5	5	17	19	13	13	12	17
25	6	6	10	10	9	11	6	8	7	7	12	12	9	10	4	5	6	8	6	7	10	10	5	5
26	7	7	12	14	17	19	6	8	12	12	16	19	10	10	12	14	4	4	15	15	14	16	7	7
27	5	5	10	10	13	17	5	5	10	11	4	4	8	9	7	10	9	9	13	16	10	10	4	4
28	5	5	8	10	8	10	8	9	8	10	6	7	8	8	7	8	6	7	9	12	12	14	11	12
29	12	13	6	7	8	8	9	10	13	13	7	11	10	10	9	9	8	8	16	23	13	16	4	5
30	14	14			10	10	18	18	13	15	3	3	8	11	10	10	13	16	17	19	10	10	4	4
31	15	15			9	11			4	4			12	12	4	4			11	13			14	16

Let Y be the number of day's that accidents occurred at the year 2016. X_i, $i = 1,2,\ldots$, be the number of fatalities of i^{th} accident and N be the total number of fatalities from January 2016 to December 2016.

Assuming that the number of accidents follows COM-Poisson distribution, the estimated values of λ and ν are 5.1161 & 0.7385 respectively.

Number of fatal crashes	Observed frequency	Expected frequency COM-Poisson
1	1	1.1303
2	3	3.4659
3	8	7.8778
4	20	14.4785
5	21	22.5670
6	34	30.7431
7	31	37.3750
8	41	41.1707
9	36	41.5738
10	36	38.8382
11	33	33.8165
12	29	27.6117
13	35	21.2513
14	14	15.4849
15	6	10.7227
16	8	7.0795
17	8	4.4694
18	1	2.7050
19	0	1.5731
20	0	10.8808
21	1	0.4757
≥22	0	0.7091
Total	366	366

Fitting the Negative binomial distribution to the number of fatalities, the parameters are obtained as $k = 15.5646$, $P = 0.6622$.

Number of fatal crashes	Observed frequency	Expected frequency COM-Poisson
1	1	0.8336
2	3	2.7506
3	7	6.4159
4	15	11.8633
5	24	18.4937
6	19	25.2530

 DOI : 10.18231/2454-9150.2018.1069

7	26	30.9939
8	33	34.8283
9	30	36.3304
10	35	35.5548
11	40	32.9203
12	30	29.0339
13	23	24.5264
14	24	19.9367
15	19	15.6550
16	11	11.9144
17	7	8.8134
18	7	6.3523
19	6	4.4708
20	1	3.0782
21	2	2.0769
22	1	1.3752
23	2	0.8948
≥24	0	1.6340
Total	366	366

For COM-Poisson Pascal distribution, the parameters, mean, variance, ratio between variance and mean are given in the following table.

Data	**Parameters**	**Mean**	**Standard Deviation**	**Ratio between Variance and Mean**	**Dispersion**
Data Set 1	$\lambda = 2.6027$ $\nu = 0.4320$ $k = 2.2639$ $P = 0.2300$	5.1256	5.1383	5.1511	Over Dispersion
Data Set 2	$\lambda = 5.1161$ $\nu = 0.7385$ $k = 15.5646$ $P = 0.6622$	95.8676	78.6972	64.6022	Over Dispersion

As both the data sets are over dispersed, COM-Poisson Pascal distribution will be more suitable than the Poisson Pascal distribution

VII. Conclusion

In this paper, COM-Poisson Pascal distribution which is obtained by compounding COM-Poisson and Pascal distribution is proposed. This is a stopped sum distribution. The Poisson distribution is suitable for equi-dispersed data. But all real world data need not be equi-dispersed always. So it is necessary to utilize distributions which support over and under dispersed data. Com-Poisson Pascal distribution is suitable for over and under-dispersed data. In this research work, it is shown that for the over-dispersed traffic accident model COM-Poisson Pascal distribution is more suitable than Poisson Pascal distribution.

References

[1] Conway R.W, and W.L. Maxwell, “A queuing model with state dependent service rates”, Journal of Industrial Engineering, 1962, 12, pp.132-136.

[2] Feller W, “On a general class of “contagious” distributions”, Annals of Mathematical Statistics, 1943). 14, 389-400.

[3] Johnson N.L, Kotz S and Kemp A.W, “Univariate Discrete Distributions”, 3rd edition, Wiley Series in Probability and Mathematical Sciences. 2005.

[4] Josemar Rodrigues, Mario De Castro, Vicente G. Cancho and N.Balakrishnan, “COM-Poisson cure rate survival models and an application to a cutaneous melanoma data”, Journal of Statistical Planning and Inference, 2009, 139, 3605-3611

[5] Katti S.K., and Gurland J, “The Poisson Pascal distribution”, Biometrics, 1961, 17, 527-538.

[6] Meintanis S.G, “ A new goodness of fit test for certain bivariate distributions applicable to traffic accidents”, Statist. Methodol. 2007, 4, pp. 22-34

[7] Minkova L.D, “A Generalization of the Classical Discrete Distributions”, Commun. Statist - Theory and Methods, 2002, 31 (6), 871-888.

[8] Polya G, “Sur quelques points de la th'eorie des probabilities”, Annales de l'Institut H. Poincre, 1930 , 1, 117-161. [6.2.4,9.7].

[9] Priyadharshini J, Saavithri V and Seethalakshmi R, “COM-Poisson Polya-Aeppli distribution”, International Journal of Mathematics Trends and Technology, Special Issue of ICRMIT-2018,125-132.

[10] Shmeli G, Minka T.P, Kadane J.B, Borle S and Boatwright P, “A useful distribution for fitting discrete data: Revival of the COM-Poisson distribution”, J. R. Stat. Soc. Ser. C (Appl. Stat), 2005, 54, 127-142.

[11] Shumway R., and Gurland J, “A fitting procedure for some generalized Poisson distributions”, Skandivanavisk Aktuarietidskrift, 1960a, 43, 87-108.

[12] Shumway R., And Gurland J, “Fitting the Poisson binomial distributions”, Biometrics, 1960b, 16, 522-533.

[13] Skellam J.G, “Studies in statistical ecology I: Spatial pattern”, Biometrika, 1952, 39, 346-362.

 DOI : 10.18231/2454-9150.2018.1069