THE POISSON - PRATIBHA PROBABILITY MODEL WITH STATISTICAL PROPERTIES AND APPLICATIONS

.

ABSTRACT

The Poisson-Pratibha distribution is derived by compounding the Poisson distribution with the Pratibha distribution and the proposed distribution has the capability to capture the skewness and the over-dispersion of the dataset. The distribution has a tendency to accommodate the right tail and tends to zero at faster rate. A general expression for the th factorial moment of Poisson-Pratibha distribution has been obtained and hence its first four moments about origin and central moments have been derived. The proposed distribution is unimodal, has increasing hazard rate and over-dispersed. Moments based descriptive measures have been derived and studied. The reliability properties including hazard function, reverse hazard function, cumulative hazard function, second rate of failure and Mills ratio of the proposed probability model have been discussed. A simulation study has been done to test the performance of maximum likelihood estimates. Finally, the goodness of fit of the proposed distribution and its comparison with other one parameter over-dispersed discrete distributions including Poisson-Lindley distribution (PLD), Poisson-Garima distribution (PGD) and Poisson-Sujatha distribution (PSD) on two datasets are discussed and presented. The result shows that the PPD has greater flexibility and applicability in modeling real over-dispersed count data and thus provides its suitability for practical applications.

Keywords: Pratibha distribution, compounding, moments, statistical properties, Maximum likelihood estimation, Simulation, goodness of fit.

1. INTRODUCTION

The Poisson distribution is the first classical count distribution and is suitable for modeling for equi-dispersed (mean equal to variance) count data. Count data appear in several fields of knowledge including biological sciences, insurance, medicine and agriculture, some among others. But in real life situation, it has been observed that most of the datasets being stochastic in nature are either over-dispersed (variance greater than mean) or under-dispersed (variance less than mean). Various statistical techniques are proposed to deal with over- dispersed count data such as weighted discrete distributions and the mixture of discrete distributions. A well-known and widely used technique to capture over-dispersion in count data is the mixed Poisson distribution. During recent decades an attempt has been made by different researchers to derive over-dispersed one parameter discrete distribution by compounding Poisson distribution with one parameter positively skewed continuous lifetime distributions. One of the important characteristics of the Poisson mixture of lifetime distribution is that the resultant distribution follows some characteristics of its mixing distribution. A popular one parameter over-dispersed discrete distribution is the Poisson-Lindley distribution (PLD) proposed by Sankaran (1970). The PLD is the Poisson mixture of the Lindley distribution introduced by Lindley (1958). Some statistical properties and different methods of estimation of the parameter of PLD have been discussed by Ghitany and Al-Mutairi (2009). Further, it has been observed that this one parameter discrete distributions are not suitable for some over-dispersed datasets due to their levels of over-dispersion. Shanker and Hagos (2015) have detailed discussion on applications of PLD for data arising

.

from biological sciences, as the data from biological sciences are, in general, over-dispersed. It has been observed by Shanker and Hagos (2015) that there are data from biological sciences where PLD does not provide better fit and hence there is a need for another over-dispersed discrete distribution. To overcome the problem of goodness of fit by PLD, Shanker (2017) proposed Poisson-Garima distribution (PGD), the Poisson compound of the Garima distribution introduced by Shanker (2016a)). Shanker (2016b) also proposed Poisson-Sujatha distribution (PSD), the Poisson compound of the Sujatha distribution of Shanker (2016c) to model over-dispersed data. Further, it has also been observed by Shanker (2017) and Shanker and Hagos (2016) while testing the goodness of fit by PGD and PSD on count data arising from various fields of knowledge that there were some datasets where both PGD and PSD failed to provide satisfactory fit. This necessitates the search for another one parameter over-dispersed count distribution which would provide better fit over PLD, PGD and PSD and for this firstly we have to search one parameter positively skewed continuous distribution. Keeping this point in mind, Shanker (2023) introduced a one parameter lifetime distribution, named Pratibha distribution to model positively skewed data defined by its probability density function (pdf) and cumulative distribution function (cdf)

$$f(x;\theta)=\frac{\theta^3}{\theta^3+\theta+2}\left(\theta+x+x^2\right)e^{-\theta x};x>0,\theta>0 \qquad \ldots(1.1)$$

$$F(x;\theta)=1-\left[1+\frac{\theta x\left(\theta x+\theta+2\right)}{\theta^3+\theta+2}\right]e^{-\theta x};x>0,\theta>0 \qquad \ldots(1.2)$$

Pratibha distribution is also a convex combination of exponential (θ) distribution, gamma $(2,\theta)$ distribution and gamma $(4,\theta)$ distribution with respective mixing proportions $\frac{\theta^3}{\theta^3+\theta+2},\frac{\theta}{\theta^3+\theta+2}$ and $\frac{2}{\theta^3+\theta+2}$,respectively. Its hazard function is monotonically increasing which makes it suitable for representing scenarios where the probability of failure increases over time. The positive skewness of Pratibha distribution makes it suitable for modeling phenomena where the majority of values are clustered towards the lower end of the range, with a tail extending towards higher values. Prodhani and Shanker (2024a, 2024b) have proposed weighted Pratibha distribution and power Pratibha distribution and discussed their statistical properties and applications in different fields of knowledge. Pratibha distribution and its related forms offer more flexibility compared to simpler distributions like exponential distribution, Lindley distribution and Sujatha distribution. They can better capture the shape and characteristics of various datasets, leading to more accurate models. Pratibha and its related distributions have applications in diverse fields including life sciences for modeling survival time data, in reliability engineering for modeling component failure times and other areas where positively skewed data is encountered.

The main purpose of this paper is to derive an over-dispersed discrete distribution which is the compound of Poisson and Pratibha distribution because it has the capability to capture both the skewness and over-dispersion of the dataset. Descriptive statistical constants including coefficients of variation, skewness, kurtosis and index of dispersion have been studied. Over-dispersion, unimodality and increasing hazard rate of the derived distribution has been discussed. Important reliability functions expressions including hazard function, reverse hazard function, second rate of failure, cumulative hazard rate function and Mills ratio of the proposed distribution has been derived and discussed. Method of moments and the method of maximum likelihood estimation have been explained to estimate parameter of the proposed distribution. Simulation has been presented to examine the consistency of maximum likelihood estimate. Goodness of fit of the proposed probability model and its comparison with other one parameter over-dispersed discrete distributions are presented.

.

2. POISSON-PRATIBHA PROBABILITY MODEL

Definition1: A random variable X is said to be Poisson-Pratibha distribution (PPD) if it follows the stochastic representation

$X\mid\lambda\sim\text{Poisson}(\lambda)$ and $\lambda\mid\theta\sim\text{Pratibha}(\theta)$ for $\lambda>0,\theta>0$.

We would denote the unconditional distribution of the stochastic representation as PPD (θ).

Theorem 1: If $X\sim\text{PPD}(\theta)$, then the pmf of X can be expressed as

$$P(X=x)=P(x;\theta)=\frac{\theta^3}{\theta^3+\theta+2}\frac{x^2+(\theta+4)x+\left(\theta^3+2\theta^2+2\theta+3\right)}{(\theta+1)^{x+3}};x=0,1,2,...,\theta>0$$

Proof: If $X|\lambda \sim \text{Poisson}(\lambda)$ distribution and $\lambda|\theta \sim \text{Pratibha}(\theta)$ distribution, then the probability mass function (pmf) of the unconditional random variable X can be obtained as

$$P(X=x)=\int_0^\infty P(X=x|\lambda)\,f(\lambda,\theta)\,d\lambda,$$

where $f(\lambda,\theta)$ is the Pratibha distribution with parameter θ.

We have

$$P(X=x)=P(x;\theta)=\int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!}\frac{\theta^3}{\theta^3+\theta+2}\left(\theta+\lambda+\lambda^2\right)e^{-\theta\lambda}d\lambda \qquad \text{... (2.1)}$$

$$=\frac{\theta^3}{\left(\theta^3+\theta+2\right)x!}\int_0^\infty e^{-(\theta+1)\lambda}\left(\theta\lambda^x+\lambda^{x+1}+\lambda^{x+2}\right)d\lambda$$

$$=\frac{\theta^3}{\left(\theta^3+\theta+2\right)x!}\left[\frac{\theta\Gamma(x+1)}{(\theta+1)^{x+1}}+\frac{\Gamma(x+2)}{(\theta+1)^{x+2}}+\frac{\Gamma(x+3)}{(\theta+1)^{x+3}}\right]$$

$$=\frac{\theta^3}{\theta^3+\theta+2}\frac{x^2+(\theta+4)x+\left(\theta^3+2\theta^2+2\theta+3\right)}{(\theta+1)^{x+3}};x=0,1,2,...,\theta>0 \qquad \text{... (2.2)}$$

Since this is the compound of the Poisson with the Pratibha distribution, we would call this probability model as Poisson-Pratibha distribution (PPD). The pmf of PPD for different values of parameter θ are presented in figure 1. As the value of θ increases, the pmf of PPD becomes highly positively skewed.

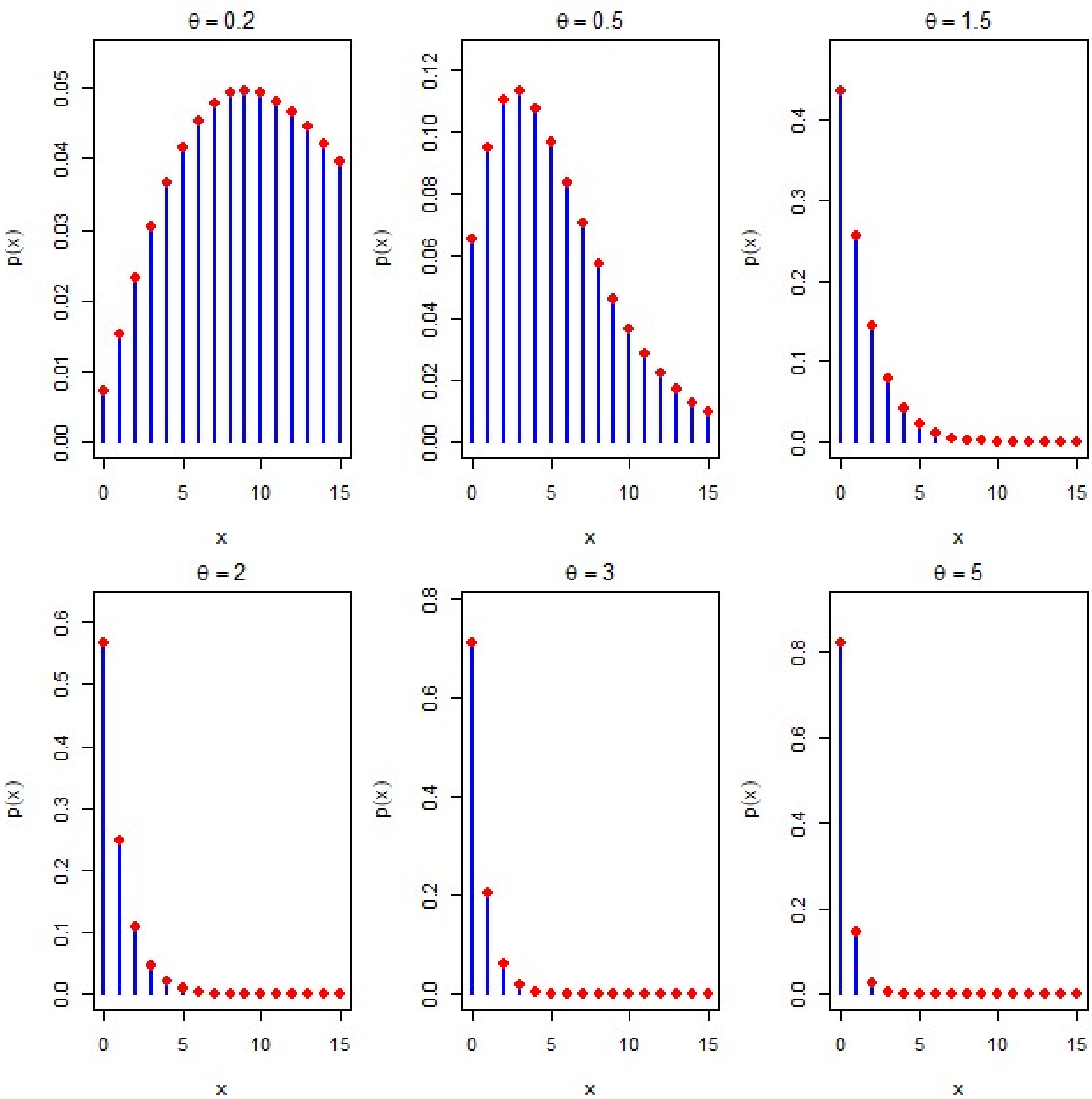


Fig.1: Pmf of PPD for varying values of parameter

The PPD distribution is skewed to the right, unimodal and decreasing which is supported by the nature of the pmf of the PPD and mathematically shown in theorems 2 and 3. In theorem 4, it has been shown that PPD is also a two-component mixture of negative binomial distributions in fixed proportions with different parameter (number of successes) and for the same probability of success. Theorem 5 is useful for deriving moments from probability generating function and moment generating function.

It can be easily shown that PPD has increasing hazard rate (IHR) and is unimodal. Since

$$Q(x;\theta)=\frac{P(x+1;\theta)}{P(x;\theta)}=\frac{1}{\theta+1}\left[1+\frac{2x+\theta+5}{x^2+(\theta+4)x+(\theta^3+2\theta^2+2\theta+3)}\right]$$ is a decreasing function of x for a given θ, $P(x;\theta)$ is log-concave. This implies that PPD has an increasing hazard rate and is unimodal. Grandell (1997) has detailed discussion about relationship between log-concavity, IHR and Unimodality of discrete distributions.

Theorem 2: The $Q(x;\theta)$ is decreasing function of x for given θ.

Proof: We have

$$Q(x;\theta)=\frac{P(x+1;\theta)}{P(x;\theta)}=\frac{1}{\theta+1}\left[1+\frac{2x+\theta+5}{x^2+(\theta+4)x+(\theta^3+2\theta^2+2\theta+3)}\right].$$

Differentiating it partialy with respect to x, we get

$Q'(x;\theta)=\dfrac{-(2x^2+2\theta x+10x-2\theta^3-3\theta^2+5\theta+14)}{(\theta+1)\left[x^2+(\theta+4)x+(\theta^3+2\theta^2+2\theta+3)\right]^2}$. Since $Q'(x;\theta)<0$, $Q(x;\theta)$ is decreasing function of x for given θ.

Theorem 3: The pmf $P(x;\theta)$ of PPD is log-concave

Proof: We have

$$P(x;\theta)=\frac{\theta^3}{\theta^3+\theta+2}\frac{x^2+(\theta+4)x+(\theta^3+2\theta^2+2\theta+3)}{(\theta+1)^{x+3}}.$$

This gives $\log P(x;\theta)=3\log\theta+\log\left[x^2+(\theta+4)x+(\theta^3+2\theta^2+2\theta+3)\right]-\log(\theta^3+\theta+2)-(x+3)\log(\theta+1)$ Assuming $g(x;\theta)=\log P(x;\theta)$ and differentiating it partially with respect to x, we have

$$g'(x;\theta)=\frac{2x+\theta+4}{x^2+(\theta+4)x+(\theta^3+2\theta^2+2\theta+3)}-\log(\theta+1)$$

and

$$g''(x;\theta)=\frac{-(2x^2+2\theta x+8x-2\theta^3-3\theta^2+4\theta+10)}{\left[x^2+(\theta+4)x+(\theta^3+2\theta^2+2\theta+3)\right]^2}<0.$$

This means that the pmf of PKD is log-concave.

Theorem 4: The PPD is a three-component mixture of negative binomial distributions and can be expressed as

$$P(x;\theta)=p_1P_1(x;\theta)+p_2P_2(x;\theta)+p_3P_3(x;\theta)\quad;p_1+p_2+p_3=1,$$

where $P_i(x;\theta)$ is the pmf of the negative binomial distribution(NBD) with parameter the number of successes i and $p_1=\dfrac{\theta^3}{\theta^3+\theta+2}$, $p_2=\dfrac{\theta}{\theta^3+\theta+2}$, $p_3=\dfrac{2}{\theta^3+\theta+2}$ with $P_1(x;\theta)=\dfrac{\theta}{(\theta+1)^{x+1}}$ as $NBD\left(1,\dfrac{\theta}{\theta+1}\right)$ and $P_2(x;\theta)=\dfrac{(x+1)\theta^2}{(\theta+1)^{x+2}}$ as the $NBD\left(2,\dfrac{\theta}{\theta+1}\right)$ and $P_3(x;\theta)=\dfrac{(x+1)(x+2)\theta^3}{2(\theta+1)^{x+3}}$ as the $NBD\left(3,\dfrac{\theta}{\theta+1}\right)$,respectively.

Proof: We have

$$P(x;\theta)=\int_0^\infty\frac{e^{-\lambda}\lambda^x}{x!}\frac{\theta^3}{\theta^3+\theta+2}(\theta+\lambda+\lambda^2)e^{-\theta\lambda}d\lambda$$

$$=\int_0^\infty\frac{e^{-\lambda}\lambda^x}{x!}\left\{\frac{\theta^3}{\theta^3+\theta+2}(\theta e^{-\theta\lambda})+\frac{\theta}{\theta^3+\theta+2}\left(\frac{\theta^2}{\Gamma(2)}e^{-\theta\lambda}\lambda\right)+\frac{2}{\theta^3+\theta+2}\left(\frac{\theta^3}{\Gamma(3)}e^{-\theta\lambda}\lambda^2\right)\right\}d\lambda$$

$$=\frac{\theta^3}{\theta^3+\theta+2}\left[\frac{\theta}{x!}\int_0^\infty e^{-(\theta+1)\lambda}\lambda^x d\lambda\right]+\frac{\theta}{\theta^3+\theta+2}\left[\frac{\theta^2}{x!\Gamma(2)}\int_0^\infty e^{-(\theta+1)\lambda}\lambda^{x+1}d\lambda\right]$$
$$+\frac{2}{\theta^3+\theta+2}\left[\frac{\theta^2}{x!\Gamma(3)}\int_0^\infty e^{-(\theta+1)\lambda}\lambda^{x+2}d\lambda\right]$$

$$=\frac{\theta^3}{\theta^3+\theta+2}\left[\frac{\theta}{x!}\frac{\Gamma(x+1)}{(\theta+1)^{x+1}}\right]+\frac{\theta}{\theta^3+\theta+2}\left[\frac{\theta^2}{x!\Gamma(2)}\frac{\Gamma(x+2)}{(\theta+1)^{x+2}}\right]+\frac{2}{\theta^3+\theta+2}\left[\frac{\theta^3}{x!\Gamma(3)}\frac{\Gamma(x+3)}{(\theta+1)^{x+3}}\right]$$

$$=\frac{\theta^3}{\theta^3+\theta+2}\left[\frac{\theta}{(\theta+1)^{x+1}}\right]+\frac{\theta}{\theta^3+\theta+2}\left[\frac{(x+1)\theta^2}{(\theta+1)^{x+2}}\right]+\frac{2}{\theta^3+\theta+2}\left[\frac{(x+1)(x+2)\theta^3}{2(\theta+1)^{x+3}}\right]$$

$$=\frac{\theta^3}{\theta^3+\theta+2}\left[\binom{x+1-1}{x}\left(\frac{\theta}{\theta+1}\right)^1\left(\frac{1}{\theta+1}\right)^x\right]+\frac{\theta}{\theta^3+\theta+2}\left[\binom{x+2-1}{x}\left(\frac{\theta}{\theta+1}\right)^2\left(\frac{1}{\theta+1}\right)^x\right]$$
$$+\frac{2}{\theta^3+\theta+2}\left[\binom{x+3-1}{x}\left(\frac{\theta}{\theta+1}\right)^3\left(\frac{1}{\theta+1}\right)^x\right]$$

$$=\frac{\theta^3}{\theta^3+\theta+2}\left[NBD\left(1,\frac{\theta}{\theta+1}\right)\right]+\frac{\theta}{\theta^3+\theta+2}\left[NBD\left(2,\frac{\theta}{\theta+1}\right)\right]+\frac{2}{\theta^3+\theta+2}\left[NBD\left(3,\frac{\theta}{\theta+1}\right)\right].$$

This completes the proof.

Although the PPD is a three-component mixture of negative binomial distribution but the existence of the three modes cannot be observed in any of the pmf's in the figure 1for the selected values of the parameter θ. This suggests that the three modes which come from the three sub-populations must be located very close to each other. As observed by Tajuddin (2022) that if the modes of the sub-populations are located very close to each other, the population will have single mode. This suggests that if the existence of the modes of the sub-populations is certain, then the true distribution can be considered as one of the candidates to model over-dispersed count data.

Theorem 5: The probability generating function and the moment generating function of PPD are given by

$$P_X(t)=\frac{\theta^3}{(\theta^3+\theta+2)(\theta+1)^2}\left[\frac{2t^2}{(\theta+1-t)^3}+\frac{(\theta+5)t}{(\theta+1-t)^2}+\frac{\theta^3+2\theta^2+2\theta+3}{(\theta+1-t)}\right] \text{, and}$$

$$M_X(t)=\frac{\theta^3}{(\theta^3+\theta+2)(\theta+1)^2}\left[\frac{2e^{2t}}{(\theta+1-e^t)^3}+\frac{(\theta+5)e^t}{(\theta+1-e^t)^2}+\frac{\theta^3+2\theta^2+2\theta+3}{(\theta+1-e^t)}\right].$$

Proof: We have

$$P_X(t)=E\left(t^X\right)=\sum_{x=0}^{\infty}t^x\frac{\theta^3}{\theta^3+\theta+2}\frac{x^2+(\theta+4)x+\left(\theta^3+2\theta^2+2\theta+3\right)}{(\theta+1)^{x+3}}$$

$$=\frac{\theta^3}{(\theta^3+\theta+2)(\theta+1)^3}\sum_{x=0}^{\infty}\left[x^2+(\theta+4)x+\left(\theta^3+2\theta^2+2\theta+3\right)\right]\left(\frac{t}{\theta+1}\right)^x$$

$$=\frac{\theta^3}{(\theta^3+\theta+2)(\theta+1)^3}\left[\sum_{x=0}^{\infty}x^2\left(\frac{t}{\theta+1}\right)^x+(\theta+4)\sum_{x=0}^{\infty}x\left(\frac{t}{\theta+1}\right)^x+\left(\theta^3+2\theta^2+2\theta+3\right)\left(\frac{t}{\theta+1}\right)^x\right]$$

$$=\frac{\theta^3}{(\theta^3+\theta+2)(\theta+1)^3}\left[\left\{\frac{(\theta+1)t}{(\theta+1-t)^2}+\frac{2(\theta+1)t^2}{(\theta+1-t)^3}\right\}+\frac{(\theta+1)(\theta+4)t}{(\theta+1-t)^2}+\frac{(\theta+1)\left(\theta^3+2\theta^2+2\theta+3\right)}{(\theta+1-t)}\right]$$

$$=\frac{\theta^3}{(\theta^3+\theta+2)(\theta+1)^3}\left[\frac{2t^2}{(\theta+1-t)^3}+\frac{t+(\theta+4)t}{(\theta+1-t)^2}+\frac{\left(\theta^3+2\theta^2+2\theta+3\right)}{(\theta+1-t)}\right]$$

$$=\frac{\theta^3}{(\theta^3+\theta+2)(\theta+1)^3}\left[\frac{2t^2}{(\theta+1-t)^3}+\frac{(\theta+5)t}{(\theta+1-t)^2}+\frac{\left(\theta^3+2\theta^2+2\theta+3\right)}{(\theta+1-t)}\right].$$

Taking $t=e^t$ in the RHS, the moment generating function of PPD can thus be obtained as

$$M_X(t)=\frac{\theta^3}{\left(\theta^3+\theta+2\right)\left(\theta+1\right)^2}\left[\frac{2e^{2t}}{\left(\theta+1-e^t\right)^3}+\frac{\left(\theta+5\right)e^t}{\left(\theta+1-e^t\right)^2}+\frac{\theta^3+2\theta^2+2\theta+3}{\left(\theta+1-e^t\right)}\right].$$

This completes the proof.

3. DESCRIPTIVE STATISTICS BASED ON MOMENTS

It is very tedious and cumbersome to find the moments of PPD directly. However, using the result (2.1), the factorial moments can be obtained easily and then using relationship between factorial moments and moments about the origin, moments about the origin can be obtained. Finally, using the relationship between moments about the mean and the moments about the origin, moments about the mean can be obtained. In theorem 6, a general expression for the factorial moment has been presented. The theorem 7 shows that the PPD is always over-dispersed and thus can be one of the important discrete distributions to model over-dispersed count data.

Theorem 6: The r th factorial moment about origin $\mu_{(r)}'$ of PPD is given by

$$\mu_{(r)}'=\frac{r!\left\{\theta^3+\left(r+1\right)\theta+\left(r+1\right)\left(r+2\right)\right\}}{\theta^r\left(\theta^3+\theta+2\right)};r=1,2,3,...$$

Proof: Using (2.1), $\mu_{(r)}'$ can be obtained as

$$\mu_{(r)}'=E\left[E\left(X^{(r)}\mid\lambda\right)\right]=\int_0^\infty\left[\sum_{x=0}^\infty x^{(r)}\frac{e^{-\lambda}\lambda^x}{x!}\right]\frac{\theta^3}{\theta^3+\theta+2}\left(\theta+\lambda+\lambda^2\right)e^{-\theta\lambda}d\lambda$$

$$=\int_0^\infty\left[\lambda^r\sum_{x=r}^\infty\frac{e^{-\lambda}\lambda^{x-r}}{\left(x-r\right)!}\right]\frac{\theta^3}{\theta^3+\theta+2}\left(\theta+\lambda+\lambda^2\right)e^{-\theta\lambda}d\lambda$$

Taking $x-r=y$, we get

$$\mu_{(r)}'=\int_0^\infty\lambda^r\left[\sum_{y=0}^\infty\frac{e^{-\lambda}\lambda^y}{y!}\right]\frac{\theta^3}{\theta^3+\theta+2}\left(\theta+\lambda+\lambda^2\right)e^{-\theta\lambda}d\lambda$$

$$=\frac{\theta^3}{\theta^3+\theta+2}\int_0^\infty\lambda^r\left(\theta+\lambda+\lambda^2\right)e^{-\theta\lambda}d\lambda$$

$$=\frac{r!\left\{\theta^3+\left(r+1\right)\theta+\left(r+1\right)\left(r+2\right)\right\}}{\theta^r\left(\theta^3+\theta+2\right)};r=1,2,3,.... \qquad ...(3.1)$$

The first four factorial moments of PPD are thus obtained as

$$\mu_{(1)}'=\frac{\theta^3+2\theta+6}{\theta\left(\theta^3+\theta+2\right)},\quad \mu_{(2)}'=\frac{2\left(\theta^3+3\theta+12\right)}{\theta^2\left(\theta^3+\theta+2\right)}$$

$$\mu_{(3)}'=\frac{6\left(\theta^3+4\theta+20\right)}{\theta^3\left(\theta^3+\theta+2\right)},\quad \mu_{(4)}'=\frac{24\left(\theta^3+5\theta+30\right)}{\theta^4\left(\theta^3+\theta+2\right)}$$

Using the relationship between factorial moments and moments about the origin, the first four moment about the origin of the PPD are given by

$$\mu_1'=\frac{\theta^3+2\theta+6}{\theta\left(\theta^3+\theta+2\right)}\qquad \mu_2'=\frac{\theta^4+2\theta^3+2\theta^2+12\theta+24}{\theta^2\left(\theta^3+\theta+2\right)}$$

$$\mu_3' = \frac{\theta^5 + 6\theta^4 + 8\theta^3 + 24\theta^2 + 96\theta + 120}{\theta^3\left(\theta^3 + \theta + 2\right)}$$

$$\mu_4' = \frac{\theta^6 + 14\theta^5 + 38\theta^4 + 72\theta^3 + 312\theta^2 + 840\theta + 720}{\theta^4\left(\theta^3 + \theta + 2\right)}$$

The moments about the mean, using relationship between moments about the origin and moments about the mean, of PPD can thus be obtained as

$$\mu_2 = \frac{\theta^7 + \theta^6 + 3\theta^5 + 12\theta^4 + 18\theta^3 + 12\theta^2 + 24\theta + 12}{\theta^2\left(\theta^3 + \theta + 2\right)^2}$$

$$\mu_3 = \frac{\theta^{11} + 3\theta^{10} + 6\theta^9 + 25\theta^8 + 71\theta^7 + 102\theta^6 + 150\theta^5 + 236\theta^4 + 156\theta^3 + 168\theta^2 + 144\theta + 48}{\theta^3\left(\theta^3 + \theta + 2\right)^3}$$

$$\mu_4 = \frac{\begin{pmatrix}\theta^{15} + 10\theta^{14} + 23\theta^{13} + 81\theta^{12} + 323\theta^{11} + 758\theta^{10} + 1331\theta^9 + 2716\theta^8 + 4110\theta^7 \\ +4812\theta^6 + 6600\theta^5 + 6744\theta^4 + 5328\theta^3 + 4512\theta^2 + 2880\theta + 720\end{pmatrix}}{\theta^4\left(\theta^3 + \theta + 2\right)^4}.$$

The moments based descriptive constants including coefficient of variation (CV), coefficient of skewness (CS), coefficient of kurtosis (CK) and the index of dispersion (ID) of PPD are thus obtained as

$$CV = \frac{\sqrt{\mu_2}}{\mu_1'} = \frac{\sqrt{\theta^7 + \theta^6 + 3\theta^5 + 12\theta^4 + 18\theta^3 + 12\theta^2 + 24\theta + 12}}{\theta^3 + 2\theta + 6}$$

$$CS = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\theta^{11} + 3\theta^{10} + 6\theta^9 + 25\theta^8 + 71\theta^7 + 102\theta^6 + 150\theta^5 + 236\theta^4 + 156\theta^3 + 168\theta^2 + 144\theta + 48}{\left(\theta^7 + \theta^6 + 3\theta^5 + 12\theta^4 + 18\theta^3 + 12\theta^2 + 24\theta + 12\right)^{3/2}}$$

$$CK = \frac{\mu_4}{\mu_2^{\,2}} = \frac{\begin{pmatrix}\theta^{15} + 10\theta^{14} + 23\theta^{13} + 81\theta^{12} + 323\theta^{11} + 758\theta^{10} + 1331\theta^9 + 2716\theta^8 + 4110\theta^7 \\ +4812\theta^6 + 6600\theta^5 + 6744\theta^4 + 5328\theta^3 + 4512\theta^2 + 2880\theta + 720\end{pmatrix}}{\left(\theta^7 + \theta^6 + 3\theta^5 + 12\theta^4 + 18\theta^3 + 12\theta^2 + 24\theta + 12\right)^2}$$

$$ID = \frac{\mu_2}{\mu_1'} = \frac{\theta^7 + \theta^6 + 3\theta^5 + 12\theta^4 + 18\theta^3 + 12\theta^2 + 24\theta + 12}{\theta\left(\theta^3 + \theta + 2\right)\left(\theta^3 + 2\theta + 6\right)}.$$

Behaviour of coefficient of variation (CV), coefficient of skewness (CS), coefficient of kurtosis (CK) and index of dispersion (ID) of PPD for changing values of parameter are shown in figure 2. The CV, CS and CK are increasing and the ID is decreasing for increasing values of the parameter θ.

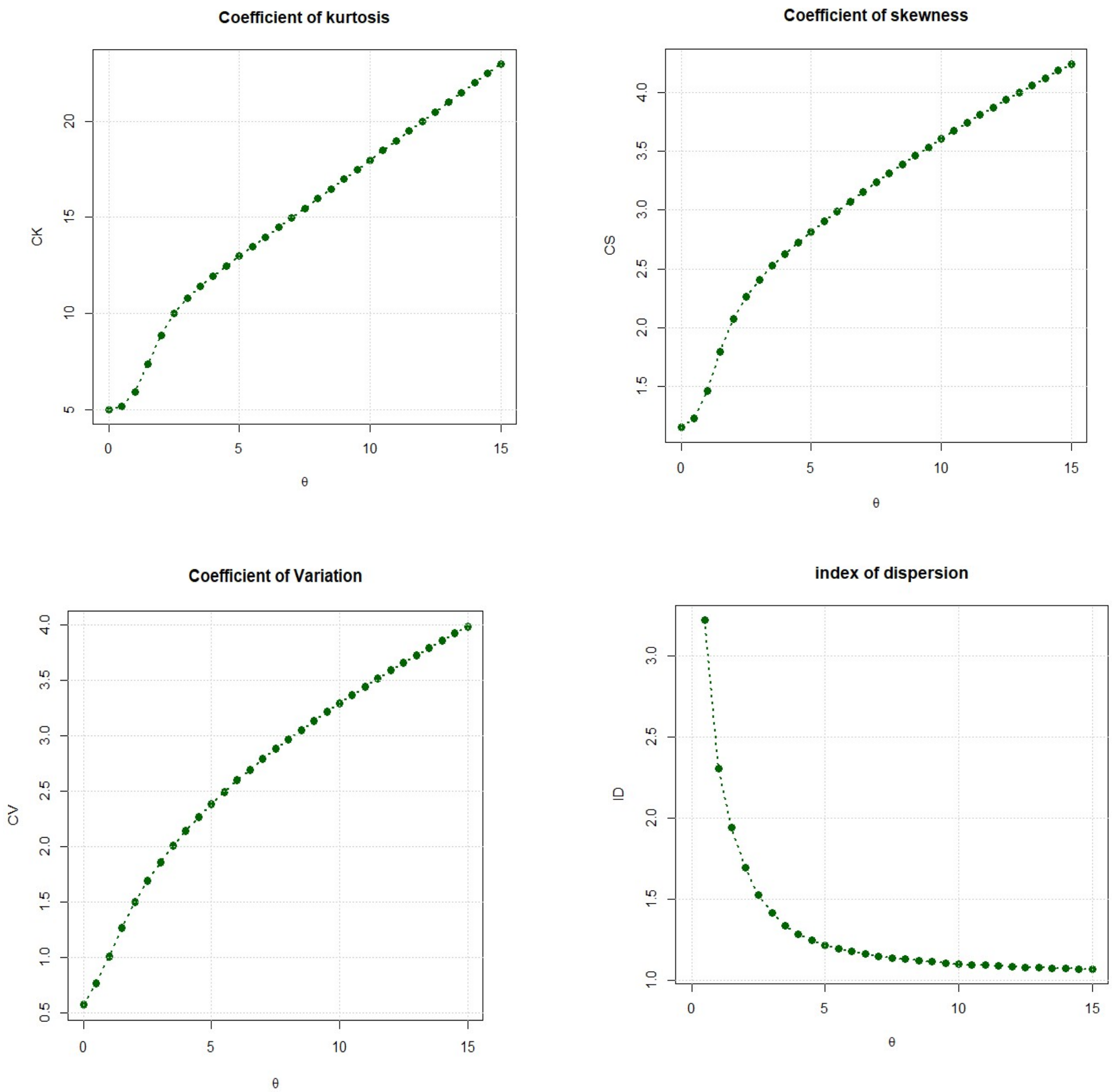


Fig. 2. CV, CS, CK and ID of PPD for varying values of parameter

Theorem 7: The PPD is over-dispersed, that is, $\mu_2 > \mu_1'$

Proof: We have

$$\mu_2 = \frac{\theta^7+\theta^6+3\theta^5+12\theta^4+18\theta^3+12\theta^2+24\theta+12}{\theta^2\left(\theta^3+\theta+2\right)^2}$$

$$= \frac{\theta^3+2\theta+6}{\theta\left(\theta^3+\theta+2\right)}\left[\frac{\theta^7+\theta^6+3\theta^5+12\theta^4+18\theta^3+12\theta^2+24\theta+12}{\theta\left(\theta^3+\theta+2\right)\left(\theta^3+2\theta+6\right)}\right]$$

$$= \mu_1' \left[\frac{\theta^7 + \theta^6 + 3\theta^5 + 12\theta^4 + 18\theta^3 + 12\theta^2 + 24\theta + 12}{\theta\left(\theta^3 + \theta + 2\right)\left(\theta^3 + 2\theta + 6\right)} \right]$$

$$= \mu_1' \left[1 + \frac{\theta^6 + 4\theta^4 + 16\theta^3 + 2\theta^2 + 12\theta + 12}{\theta\left(\theta^3 + \theta + 2\right)\left(\theta^3 + 2\theta + 6\right)} \right].$$

This gives $\mu_2 > \mu_1'$. This completes the proof.

4. RELIABILITY PROPERTIES

Various interesting and useful reliability properties including reverse hazard rate function, second rate of failure, cumulative hazard function and Mills ratio of a distribution depends on cumulative distribution function, survival function and hazard function of the distribution. The following theorem 8 deals with the cumulative distribution function (cdf), survival function and the hazard function of PPD. The expression for reverse hazard rate function, second rate of failure, cumulative hazard function and Mills ratio of PPD have also been obtained.

Theorem 8: The cumulative distribution function (cdf), survival function and the hazard function of PPD are given by

$$F(x) = F(x;\theta) = 1 - \frac{\theta^2 x^2 + \left(\theta^3 + 6\theta^2 + 2\theta\right)x + \left(\theta^5 + 2\theta^4 + 3\theta^3 + 5\theta^2 + 7\theta + 2\right)}{\left(\theta^3 + \theta + 2\right)\left(\theta + 1\right)^{x+3}}$$

$$S(x) = S(x;\theta) = \frac{\theta^2 x^2 + \left(\theta^3 + 6\theta^2 + 2\theta\right)x + \left(\theta^5 + 2\theta^4 + 3\theta^3 + 5\theta^2 + 7\theta + 2\right)}{\left(\theta^3 + \theta + 2\right)\left(\theta + 1\right)^{x+3}}$$, and

$$h(x) = h(x;\theta) = \frac{\theta^3 \left\{x^2 + \left(\theta + 4\right)x + \left(\theta^3 + 2\theta^2 + 2\theta + 3\right)\right\}}{\theta^2 x^2 + \left(\theta^3 + 6\theta^2 + 2\theta\right)x + \left(\theta^5 + 2\theta^4 + 3\theta^3 + 5\theta^2 + 7\theta + 2\right)}$$

Proof: We have

$$F(x) = F(x,\theta) = P(X \le x) = 1 - P(X \ge x+1)$$

$$= 1 - \sum_{t=x+1}^{\infty} \int_0^{\infty} P(X = t \mid \lambda) f(\lambda;\theta)\, d\lambda$$

$$= 1 - \sum_{t=x+1}^{\infty} \int_0^{\infty} \frac{e^{-\lambda}\lambda^t}{t!} \frac{\theta^3}{\theta^3 + \theta + 2}\left(\theta + \lambda + \lambda^2\right) e^{-\theta\lambda}\, d\lambda$$

$$= 1 - \sum_{t=x+1}^{\infty} \frac{\theta^3}{\left(\theta^3 + \theta + 2\right)} \frac{\left\{t^2 + \left(\theta + 4\right)t + \left(\theta^3 + 2\theta^2 + 2\theta + 3\right)\right\}}{\left(\theta + 1\right)^{t+3}}$$

$$= 1 - \frac{\theta^3}{\left(\theta^3 + \theta + 2\right)\left(\theta + 1\right)^3} \sum_{t=x+1}^{\infty} \frac{t^2 + \left(\theta + 4\right)t + \left(\theta^3 + 2\theta^2 + 2\theta + 3\right)}{\left(\theta + 1\right)^t}$$

$$= 1 - \frac{\theta^2 x^2 + \left(\theta^3 + 6\theta^2 + 2\theta\right)x + \left(\theta^5 + 2\theta^4 + 3\theta^3 + 5\theta^2 + 7\theta + 2\right)}{\left(\theta^3 + \theta + 2\right)\left(\theta + 1\right)^{x+3}}.$$

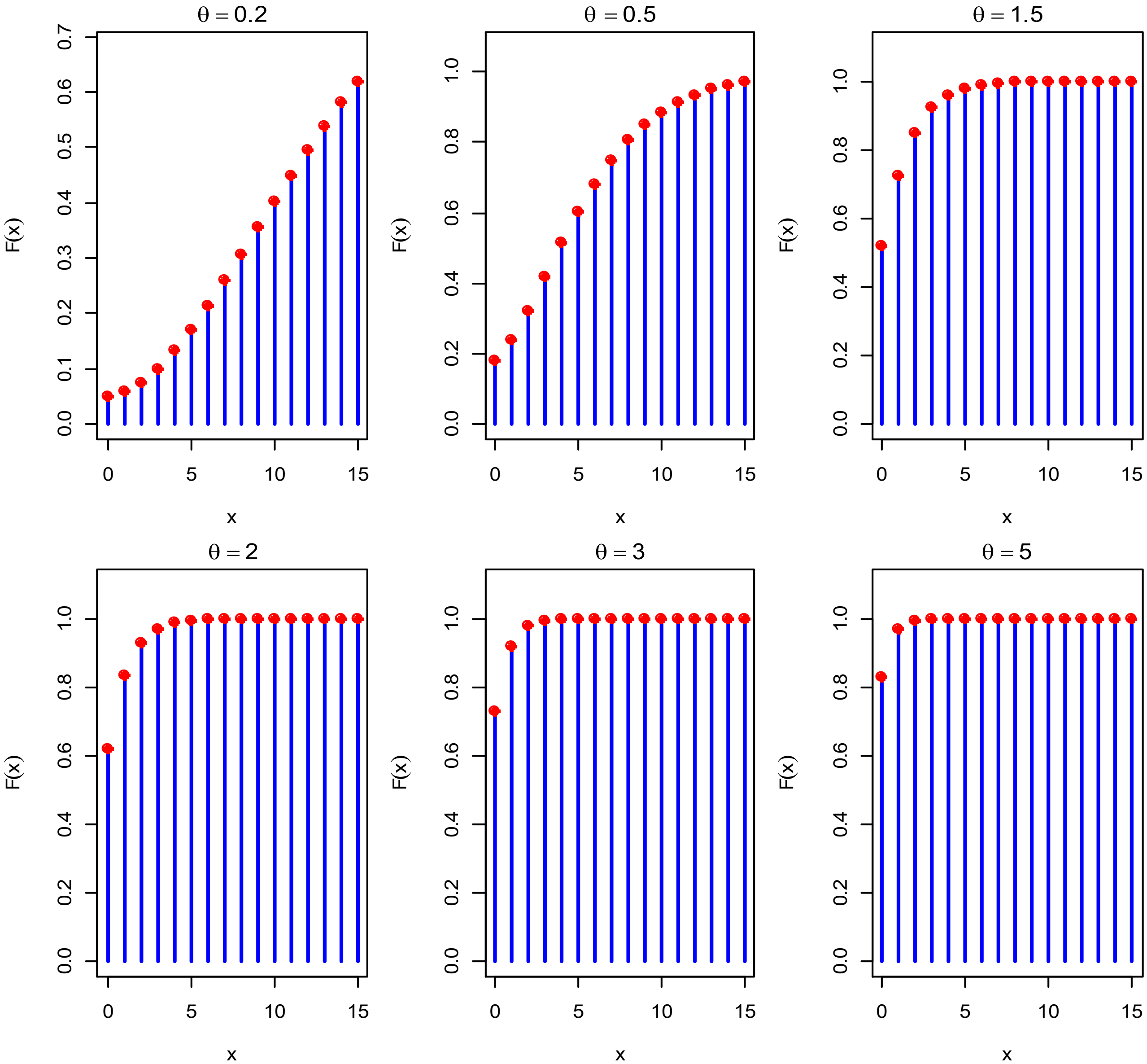
The survival function of PPD can be obtained as

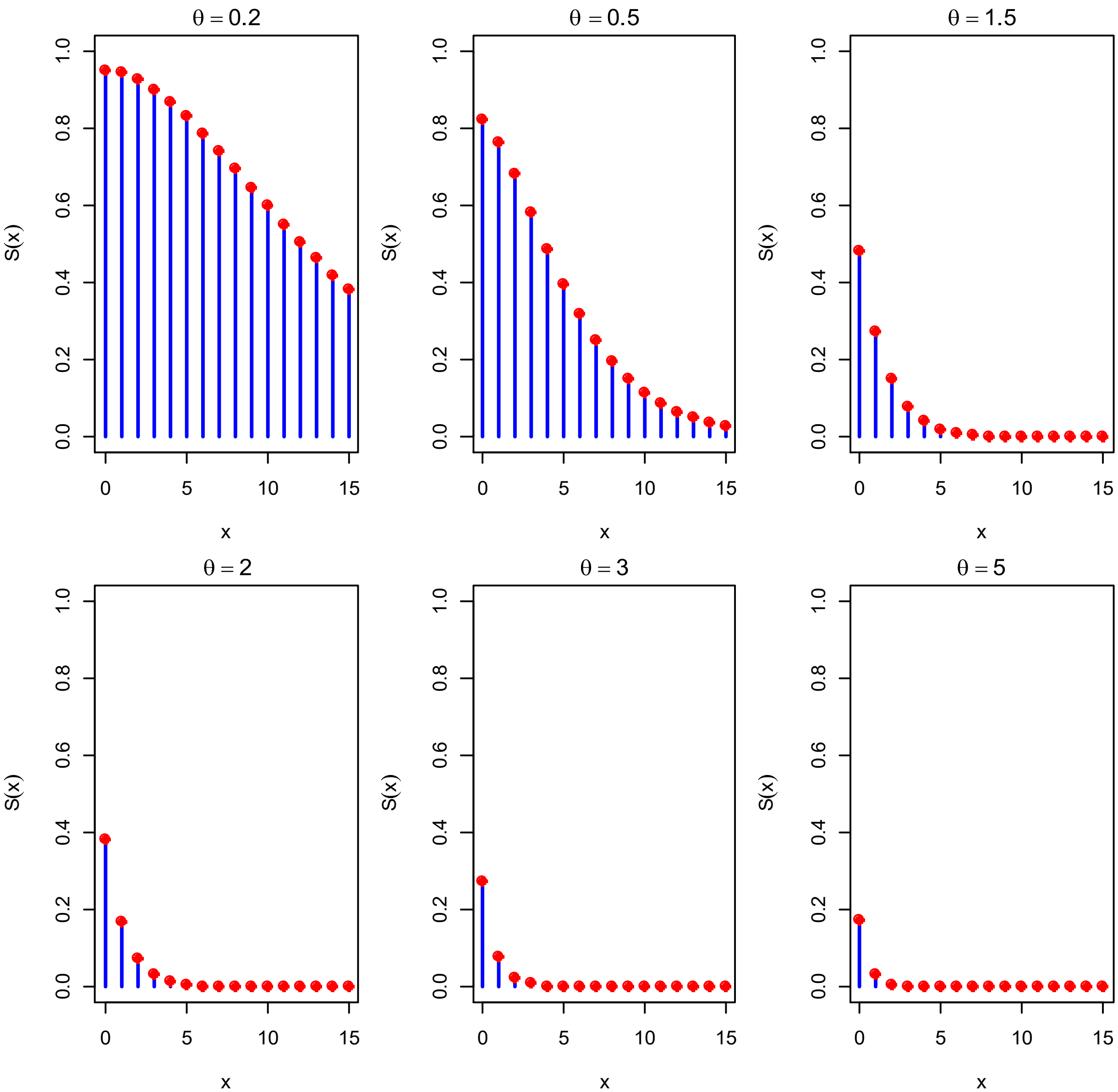
$$S(x) = S(x,\theta) = 1 - F(x,\theta) = \frac{\theta^2 x^2 + \left(\theta^3 + 6\theta^2 + 2\theta\right)x + \left(\theta^5 + 2\theta^4 + 3\theta^3 + 5\theta^2 + 7\theta + 2\right)}{\left(\theta^3 + \theta + 2\right)\left(\theta + 1\right)^{x+3}}.$$

The hazard function of PPD can be expressed as

$$h(x)=h(x,\theta)=\frac{P(x,\theta)}{S(x,\theta)}=\frac{\theta^3\left\{x^2+(\theta+4)x+\left(\theta^3+2\theta^2+2\theta+3\right)\right\}}{\theta^2x^2+\left(\theta^3+6\theta^2+2\theta\right)x+\left(\theta^5+2\theta^4+3\theta^3+5\theta^2+7\theta+2\right)}$$

The natures of cdf, survival function and hazard function of PPD for varying values of parameter are shown in the following figure 3 and it is obvious from the figure that the PPD has a valid cdf since $F(x)\to 1$ as $x\to\infty$. Further, the hazard rate function shows an increasing pattern with a limiting value of θ, which means that $\lim_{x\to\infty} h(x)=\theta$.





θ = 0.2
θ = 0.5
θ = 1.5
θ = 2
θ = 3
θ = 5
S(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15

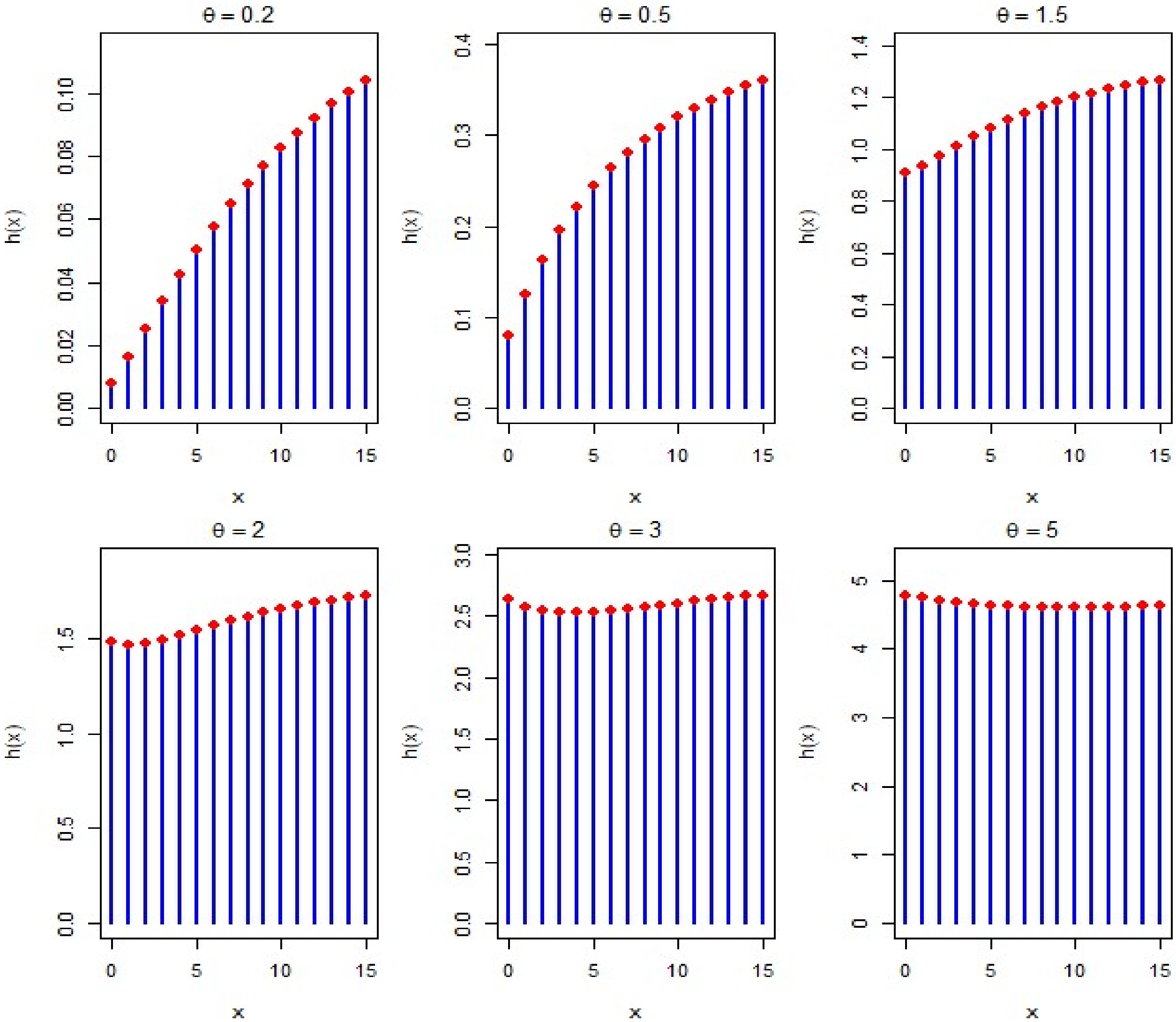


Fig. 3. cdf, survival function and hazard function of PPD for varying values of parameter

The reverse hazard rate function $Rh(x;\theta)$ and the second rate of failure $SRF(x;\theta)$ of the PPD can be obtained as

$$Rh(x;\theta)=\frac{P(x;\theta)}{F(x;\theta)}$$

$$=\frac{\theta^3\left[x^2+(\theta+4)x+\left(\theta^3+2\theta^2+2\theta+3\right)\right]}{\left[\left(\theta^3+\theta+2\right)(\theta+1)^{x+3}-\left\{\theta^2x^2+\left(\theta^3+6\theta^2+2\theta\right)x+\left(\theta^5+2\theta^4+3\theta^3+5\theta^2+7\theta+2\right)\right\}\right]}\text{ and}$$

$$SRF(x;\theta)=\ln\left[\frac{S(x;\theta)}{S(x+1;\theta)}\right]$$

$$=\ln\left[\frac{(\theta+1)\left\{\theta^2x^2+\left(\theta^3+6\theta^2+2\theta\right)x+\left(\theta^5+2\theta^4+3\theta^3+5\theta^2+7\theta+2\right)\right\}}{\theta^2(x+1)^2+\left(\theta^3+6\theta^2+2\theta\right)(x+1)+\left(\theta^5+2\theta^4+3\theta^3+5\theta^2+7\theta+2\right)}\right]$$

The cumulative hazard function $H(x;\theta)$ and Mills ratio $M(x;\theta)$ of PPD are given by

$$H(x;\theta)=-\ln S(x;\theta)=-\ln\left[\frac{\theta^2x^2+\left(\theta^3+6\theta^2+2\theta\right)x+\left(\theta^5+2\theta^4+3\theta^3+5\theta^2+7\theta+2\right)}{\left(\theta^3+\theta+2\right)(\theta+1)^{x+3}}\right],$$

and

$$M(x;\theta)=\frac{S(x;\theta)}{P(x;\theta)}=\frac{\theta^2x^2+\left(\theta^3+6\theta^2+2\theta\right)x+\left(\theta^5+2\theta^4+3\theta^3+5\theta^2+7\theta+2\right)}{\theta^3\left[x^2+(\theta+4)x+\left(\theta^3+2\theta^2+2\theta+3\right)\right]}.$$

5. ESTIMATION OF PARAMETER

5.1. Method of Moment Estimation

Let $(x_1,x_2,\ldots,x_n)$ be a random sample of size *n* from the PPD. Since PPD has one parameter, equating the population mean with the corresponding sample mean, the method of moment estimate (MOME) of PPD is the solution of the following fourth degree polynomial equation in θ

$$\bar{x}\,\theta^4-\theta^3+\bar{x}\theta^2+2(\bar{x}-1)\theta-6=0$$, where $\bar{x}$ being the sample mean.

This fourth-degree polynomial equation in θ can be solved using Newton-Raphson formula

$$\theta_{n+1}=\theta_n-\frac{f(\theta_n)}{f'(\theta_n)};n=0,1,2,3,...$$

The Newton Raphson formula has quadratic convergent where the initial value of θ_0 can be selected as follow: Suppose $f(\theta)=\bar{x}\,\theta^4-\theta^3+\bar{x}\theta^2+2(\bar{x}-1)\theta-6$, where $\bar{x}$ is the sample mean of the dataset for which we are estimating the value of the parameter. Now we have to guess two values of θ, say θ_1 and θ_2 such that $f(\theta_1)f(\theta_2)<0$. Then, we can select any value of θ say θ_0 between θ_1 and θ_2 as initial value of θ in the Newton-Raphson formula.

5.2. Method of Maximum Likelihood Estimation

Let $(x_1,x_2,\ldots,x_n)$ be a random sample of size *n* from the PPD. Let f_x be the observed frequency in the sample corresponding to $X=x$ ($x=1,2,3,...,k$) such that $\sum_{x=1}^{k}f_x=n$, where k is the largest observed value having non-zero frequency. The likelihood function, L, of the PPD is given by

$$L=\left(\frac{\theta^3}{\theta^3+\theta+2}\right)^n\frac{1}{(\theta+1)^{\sum_{x=1}^{k}(x+3)f_x}}\prod_{x=1}^{k}\left[x^2+(\theta+4)x+\left(\theta^3+2\theta^2+2\theta+3\right)\right]^{f_x}$$

The log likelihood function and the log-likelihood equation are thus given by

$$\begin{aligned}\text{Log}L=3n\log\theta-n\log\left(\theta^3+\theta+2\right)-\sum_{x=1}^{k}(x+3)f_x\log(\theta+1)\\+\sum_{x=1}^{k}f_x\log\left[x^2+(\theta+4)x+\left(\theta^3+2\theta^2+2\theta+3\right)\right]\end{aligned}$$

$$\frac{\partial\log L}{\partial\theta}=\frac{3n}{\theta}-\frac{n\left(3\theta^2+1\right)}{\theta^3+\theta+2}-\frac{1}{\theta+1}\sum_{x=1}^{k}(x+3)f_x+\sum_{x=1}^{k}\frac{\left(x+3\theta^2+4\theta+2\right)f_x}{x^2+(\theta+4)x+\left(\theta^3+2\theta^2+2\theta+3\right)}=0$$

The maximum likelihood estimate, $\hat{\theta}$ of θ is the solution of the equation $\frac{\partial\log L}{\partial\theta}=0$ and is given by the solution of the following non-linear equation

$$\frac{2n(\theta+3)}{\theta\left(\theta^3+\theta+2\right)}-\frac{n(\bar{x}+3)}{\theta+1}+\sum_{x=1}^{k}\frac{\left(x+3\theta^2+4\theta+2\right)f_x}{x^2+(\theta+4)x+\left(\theta^3+2\theta^2+2\theta+3\right)}=0$$

where $\bar{x}$ is the sample mean. Since the log-likelihood equation is non-linear and cannot be expressed in closed form and it is tedious to solve by direct method. Therefore, the MLE of the parameter θ can be computed iteratively by solving log-likelihood equation using Newton-Raphson iteration available in R-software, until sufficiently close values of the parameter θ is obtained. The initial value of the parameter θ can be taken as the value given by MOME.

6. A SIMULATON STUDY

To assess the effectiveness of the maximum likelihood estimator (MLE) for the PPD, we conducted an extensive simulation analysis. Using the inverse transform method, we generated random samples based on the distribution. The simulations were repeated 10,000 times for each sample size tested (50, 100, 200, 300, 400 and 500) to ensure robust statistical evaluation of the estimator's properties. We measured both the bias and the mean squared error (MSE) to examine how accurately and consistently the estimator performs. Simulation results, summarized in table 1 confirmed that both the bias and the MSE decline as sample size increases which indicates improved reliability of the MLE with increasing sample size. Additional simulations using different true parameter values (0.5, 1.5, 2.5, and 3.5) showed that the estimator remained consistently accurate across all tested scenarios. The formulas for the bias and the MLE are

$$bias(\hat{\theta}) = E\left[\hat{\theta}\right] - \theta = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i - \theta \qquad MSE\left(\hat{\theta}\right) = E\left[\left(\hat{\theta}-\theta\right)^2\right] = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_i - \theta\right)^2$$

Table1: Simulation result of PPD

	n	BIAS	MSE
$\theta = $ **0.5**	50	0.0423	0.0053
	100	0.0402	0.0033
	200	0.0380	0.0023
	300	0.0374	0.0019
	400	0.0376	0.0018
	500	0.0372	0.0017
	n	**BIAS**	**MSE**
$\theta = $ **1.5**	50	0.1656	0.0741
	100	0.1491	0.0439
	200	0.1399	0.0299
	300	0.1371	0.0255
	400	0.1354	0.0235
	500	0.1344	0.0220
	n	**BIAS**	**MSE**
$\theta = $ **2.5**	50	0.2652	0.3790
	100	0.2044	0.1662
	200	0.1688	0.0844
	300	0.1579	0.0619
	400	0.1559	0.0516
	500	0.1520	0.0456
	n	**BIAS**	**MSE**
$\theta = $ **3.5**	50	0.4166	1.3932
	100	0.2724	0.5375
	200	0.1897	0.2296
	300	0.1620	0.1534
	400	0.1578	0.1162
	500	0.1451	0.0973

7. APPLICATIONS

As we know that there are two conditions for the applications of Poisson distribution for count data, namely, the independence of events and equi-dispersion. But in real life situations these two conditions rarely satisfied because, in reality, events are dependent and the data are either over-dispersed or under-dispersed. For example, in biological science and medical science, the occurrence of successive events is dependent. The negative binomial distribution is a possible alternative to the Poisson distribution when successive events are possibly dependent and the data are over-dispersed. NBD, being two-parameter distribution and having lower index of dispersion does not provide better fit in most of the over-dispersed datasets. The PLD, PGD and PSD are three important over-dispersed one parameter distribution proposed for count data and it has been observed that these discrete distributions also do not provide satisfactory fit. The PPD has been found to provide quite satisfactory fit over PLD, PGD and PSD. The theoretical and empirical justification for the selection of the PPD to describe biological science and medical science data is that PDD is over dispersed ($\mu < \sigma^2$) and is suitable for data arising from mechanism where events are dependent. For testing the goodness of fit of PPD over PLD, PGD and PSD, two count datasets have been considered and the parameter of these considered distributions are estimated using maximum likelihood estimation. The mean and the variance of dataset in table 2 and 3 are (0.75, 1.31) and (0.78, 1.24) respectively and it is quite obvious that the datasets are over-dispersed. The goodness of fit measures in table 2 and 3 shows that PPD provides much better fit over PD, PLD, PGD and PSD and thus PPD can be considered as one of the important distributions for count over-dispersed data where events are dependent.

Table 2: **The distribution of Pyrausta nublilalis in 1937 and reported by Beall (1940)**

No of insects	Observed frequency	PD	PLD	PGD	PSD	PPD
0	33	26.45	31.52	31.68	31.47	31.84
1	12	19.45	14.15	13.98	14.17	13.82
2	6	7.44	6.08	6.01	6.13	5.98
3	3	1.86	2.53	2.54	2.55	2.55
4	1	0.35	1.03	1.06	1.03	1.07
5	1	0.06	0.69	0.73	0.65	0.84
total	56	56	56	56	56	56
	$\hat{\theta}(SE)$	0.7500 (0.1157)	1.81153 (0.3068)	1.6950 (0.3912)	2.2415 (0.3167)	2.0031 (0.2487)
	-2logL	143.1647	133.9691	133.8999	133.9588	133.8232
	χ^2	4.6119	0.4396	0.3776	0.4462	0.3147
	d.f	1	1	1	1	1
	P value	0.09966	0.8027	0.8280	0.8000	0.8544

Table3: **Distribution of mistakes in copying groups of random digits and available in Kemp and Kemp (1965)**

No of error per group	Observed frequency	PD	PLD	PGD	PSD	PPD
0	35	28.34	33.06	33.27	32.97	33.35
1	11	21.26	15.27	15.07	15.31	14.93
2	8	7.97	6.74	6.65	6.82	6.65
3	4	1.99	2.88	2.88	2.91	2.92
4	2	0.37	2.05	2.13	1.99	2.15
total	60	60	60	60	60	60
	$\hat{\theta}(SE)$	0.7833 (0.1143)	1.7434 (0.2809)	1.6284 (0.2831)	2.1678 (0.2907)	1.944 (0.2282)
	-2logL	155.0912	146.7021	146.6855	146.6046	146.5718
	χ^2	7.8112	1.7731	1.6588	1.7819	1.5608
	d.f	1	1	1	1	1
	P value	0.0201	0.4121	0.6461	0.4103	0.6683

8. CONCLUSION

In this paper the Poisson compound of the Pratibha distribution called Poisson-Pratibha distribution (PPD) has been suggested. The expressions of statistical constants including coefficients of variation, skewness, kurtosis and index of dispersion have been obtained and their behavior for varying values of parameter has been studied. It is observed that the PPD is unimodal, has increasing hazard rate and over-dispersed. Various reliability properties of the PPD are derived and discussed. Both the method of moment and maximum likelihood estimation has been discussed for the estimation of the parameter of the PPD. A simulation study has been done to test the performance of maximum likelihood estimates of the

PPD. Finally, the goodness of fit of the PPD and its comparison with the goodness of fit of other one parameter over-dispersed discrete distributions including Poisson-Lindley distribution (PLD) and Poisson-Garima distribution (PGD) and Poisson-Sujatha distribution (PSD) on two datasets have been presented. The goodness of fit result shows that the PPD provides greater flexibility in modeling over-dispersed count data and hence can be considered an important over-dispersed discrete distribution.

REFERENCES:

Beall, G. (1940). The fit and significance of contagious distributions when applied to observations on larval insects. *Ecology*, *21*(4), 460-474.

Ghitany, M. E. and Al-Mutairi, D.K. (2009): Estimation Methods for the discrete Poisson-Lindley distribution, *Journal of Statistical Computation and Simulation*, 79 (1), 1 – 9.

Grandell, J. Mixed Poisson Processes, Chapman &Hall, London, (1997).

Kemp, C.D. and Kemp, A.W. (1965): Some properties of the Hermite distribution, *Biometrika*, 52, 381-394.

Lindley, D. V. (1958): Fiducial distributions and Bayes theorem, *Journal of Royal Statistical Society,*20(1),102-107.

Sankaran, M. (1970): The discrete Poisson-Lindley distribution, *Biometrics,*26(1), 145-149.

Shanker, R. (2016a): Garima distribution and Its Application to model behavioral science data, *Biometrics & Biostatistics International journal*, 4(7), 275 – 281

Shanker, R. (2016b): The Discrete Poisson-Sujatha Distribution, *International Journal of Probability and Statistics*, 5(1), 1 – 9

Shanker, R. (2016c): Sujatha Distribution and Its Applications, *Statistics in Transition-New series,* 17(3), 391- 410

Shanker, R. (2017): The discrete Poisson-Garima distribution, *Biometrics & Biostatistics International Journal*, 5(2), 48-53

Shanker, R. (2023): Pratibha distribution with Statistical Properties and Application, *Biometrics and Biostatistics International Journal*, 12 (5), 136 – 142.

Shanker, R. and Hagos, F. (2015): On Poisson-Lindley distribution and its Applications to Biological Sciences, *Biometrics & Biostatistics International Journal*, 2(4), 103 – 107.

Shanker, R. and Hagos, F. (2016): On Poisson-Sujatha Distribution and Its Applications to Model count data from Biological Sciences, *Biometrics & Biostatistics International Journal*, 3(4), 100 – 106

Prodhani, H.R. and Shanker, R. (2024a): Weighted Pratibha distribution with properties and Application in flood dataset, *Biometrics & Biostatistics International Journal*, 13(2), 52 – 57.

Prodhani, H.R. and Shanker, R. (2024b): Power Pratibha Distribution with properties and Applications, *International Journal of Statistics and Reliability Engineering*, 11 (2), 217 -226.

Tajuddin, R.R.M., Ismail, N. and Ibrahim, K. (2022): Several two-component mixture distributions for count data, *Communication in Statistics-Simulation and Computation*, 51, 3760 – 3771.