

# The Poisson-Shukla Distribution and its Applications

Kamlesh Kumar Shukla

•  
Department of Statistics, Mainefhi College of Science, Asmara, Eritrea,  
Email: kkshukla22@gmail.com

Rama Shanker

•  
<sup>2</sup>Department of Statistics, Assam University, Silchar, India  
Email: shankerrama2009@gmail.com

## Abstract

*In this paper, a study on Poisson-Shukla distribution (PSD), a Poisson mixture of Shukla distribution introduced by Shukla and Shanker (2019), has been carried out. The expression for  $r$ th factorial moment about origin has been derived. The expressions for its mean and variance have been given. Maximum likelihood estimation for estimating the parameters have been discussed. The goodness of fit of the proposed distribution has been explained with two count datasets and its fit was found quite satisfactory over Poisson distribution (PD), Poisson-Lindley distribution (PLD), Poisson-weighted Lindley distribution (P-WLD), Generalized Poisson Lindley distribution (GPLD) and Negative Binomial distribution (NBD).*

**Keywords:** Shukla distribution, Moments based measures, Estimation of parameter, Goodness of fit

## 1. Introduction

The probability density function (pdf) and the cumulative distribution function (cdf) of Shukla distribution proposed by Shukla and Shanker (2019) are given by

$$f(x; \theta, \alpha) = \frac{\theta^{\alpha+1}}{\theta^{\alpha+1} + \Gamma(\alpha+1)} (\theta + x^\alpha) e^{-\theta x}; x > 0, \theta > 0, \alpha \geq 0 \quad (1.1)$$

$$F(x; \theta, \alpha) = 1 - \frac{\theta^\alpha (\theta + x^\alpha) e^{-\theta x} + \alpha \Gamma(\alpha, \theta x)}{\theta^{\alpha+1} + \Gamma(\alpha+1)}; x > 0, \theta > 0, \alpha \geq 0 \quad (1.2)$$

This distribution, being a convex combination of exponential ( $\theta$ ) distribution and gamma( $\alpha, \theta$ ) distribution, has been proposed for modeling lifetime data form engineering and biomedical science. The statistical properties of Shukla distribution including its shapes, moments, skewness, kurtosis, hazard rate function, mean residual life function, stochastic ordering, mean deviations along with estimation of parameters using Maximum likelihood estimation are given in Shukla and Shanker (2019). The main motivation behind this new discrete distribution arises from the fact that it has been observed by Shukla and Shanker (2019) that Shukla distribution provides a better fit for modeling lifetime data than exponential distribution, Lindley distribution of Lindley (1958), gamma distribution, generalized Lindley distribution of Zakerzadeh and Dolati (2009) and weighted Lindley distribution given by Ghitanay et al (2011), and it is expected that the Poisson mixture Shukla distribution would provide a better fit for modeling count data than the Poisson

mixture of these distributions namely Poisson-Lindley distribution (PLD) of Sankaran (1970), Negative binomial distribution (NBD) of Greenwood and Yule (1920), generalized Poisson-Lindley distribution (GPLD) of Mahmoudi and Zakerzadeh (2010) and Poisson-Weighted Lindley distribution (P-WLD) introduced by El-Monsef and Sohsah (2014) and studied by Shanker and Shukla (2019). The probability mass function of PLD, NBD, GPLD and P-WLD are presented in the following table

**Table 1:** The pmf of Poisson Lindley distribution (PLD), Negative binomial distribution (NBD), Generalized Poisson Lindley distribution (GPLD) and Poisson weighted Lindley distribution (PLWD)

Distributions	Probability mass function (pmf)	Introducer (Year)
PLD	$P_1(x; \theta) = \frac{\theta^2(x + \theta + 2)}{(\theta + 1)^{x+3}} ; x = 0, 1, 2, \dots, \theta > 0$	Sankaran (1970)
NBD	$P(X = x) = (x + 1) \left( \frac{\theta}{\theta + 1} \right)^p \left( \frac{1}{\theta + 1} \right)^x ; x = 0, 1, 2, \dots$	Greenwood and Yule (1920)
GPLD	$P(x; \theta, \alpha) = \frac{\theta^3 \{ \alpha x^2 + (\theta + 3\alpha + 1)x + (\theta^2 + 3\theta + 2\alpha + 2) \}}{(\theta^2 + \theta + 2\alpha)(\theta + 1)^{x+3}}$	Mahmoudi, and Zakerzadeh (2010)
P-WLD	$P(x; \theta, \alpha) = \frac{\Gamma(x + \alpha)}{\Gamma(x + 1)\Gamma(\alpha)} \frac{\theta^{\alpha+1}}{(\theta + \alpha)} \frac{x + \theta + \alpha + 1}{(\theta + 1)^{x+\alpha+1}} ; \theta > 0, \alpha > 0$	El-Monsef and Sohsah (2014)

In the present paper Poisson-Shukla distribution (PSD) by compounding Poisson distribution with Shukla distribution has been proposed. The paper is divided into six sections. Second section deals with derivation of the pmf of PSD and shapes of PSD for varying values of parameters. Moments and moments based measures of the proposed distribution have been discussed in section 3. Sections four and five deal the discussion of the estimation of parameters using maximum likelihood estimation and the goodness of fit of the proposed distribution, respectively. Finally, conclusions of the paper have been given in section six.

## 2. Poisson-Shukla Distribution

Assuming the parameter  $\lambda$  of the Poisson distribution following Shukla distribution (1.1), the Poisson mixture of Shukla distribution can be obtained as

$$\begin{aligned}
 P_1(x, \theta, \alpha) &= \int_0^\infty \frac{e^{-\lambda} \lambda^x}{x!} \frac{\theta^{\alpha+1}}{\theta^{\alpha+1} + \Gamma(\alpha + 1)} (\theta + \lambda^\alpha) e^{-\theta\lambda} d\lambda \quad (2.1) \\
 &= \frac{\theta^{\alpha+1}}{\{\theta^{\alpha+1} + \Gamma(\alpha + 1)\} \Gamma(x + 1)} \int_0^\infty e^{-(\theta+1)\lambda} (\theta \lambda^x + \lambda^{x+\alpha}) d\lambda \\
 &= \left( \frac{\theta}{\theta + 1} \right)^{\alpha+1} \left( \frac{1}{\theta + 1} \right)^x \frac{1}{\Gamma(x + 1)} \frac{\theta(\theta + 1)^\alpha \Gamma(x + 1) + \Gamma(x + \alpha + 1)}{\theta^{\alpha+1} + \Gamma(\alpha + 1)} ; x > 0, (\theta, \alpha, \beta) > 0 \quad (2.2)
 \end{aligned}$$

We name this distribution ‘‘Poisson-Shukla distribution (PSD)’. The behavior of the pmf of PSD for different values of parameter is shown in the figure 1

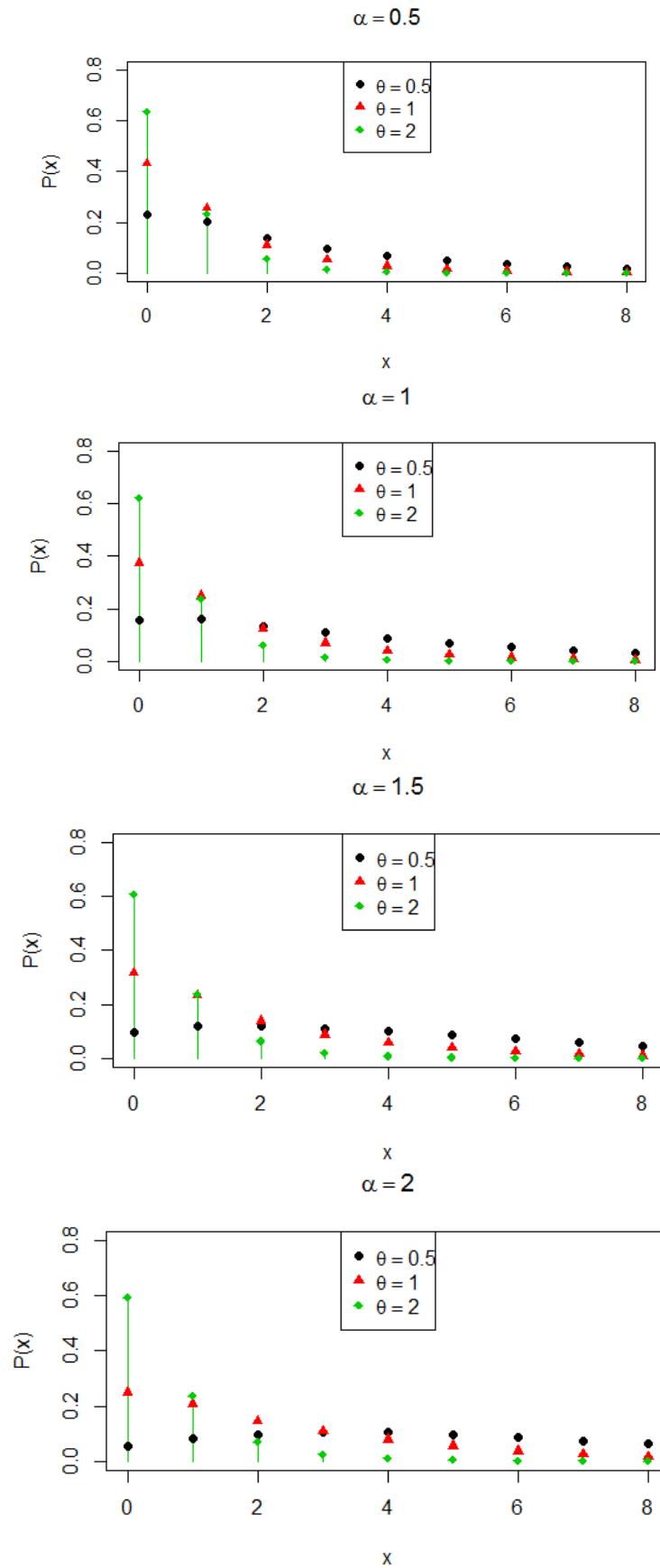


Fig.1: Behavior of the pmf of PSD for different values of the parameters  $\theta$  and  $\alpha$

### 3. Moments

The  $r^{\text{th}}$  factorial moment about origin of PSD (2.2) can be obtained as

$$\mu_{(r)}' = E\left[E\left(X^{(r)} \mid \lambda\right)\right], \quad \text{where } X^{(r)} = X(X-1)(X-2)\dots(X-r+1)$$

Using (2.1), the  $r$  th factorial moment about origin of PSD (2.2) can be obtained as

$$\begin{aligned} \mu_{(r)}' &= E\left[E\left(X^{(r)} \mid \lambda\right)\right] = \frac{\theta^{\alpha+1}}{\theta^{\alpha+1} + \Gamma(\alpha+1)} \int_0^\infty \left[ \sum_{x=0}^\infty x^{(r)} \frac{e^{-\lambda} \lambda^x}{x!} \right] (\theta + \lambda^\alpha) e^{-\theta\lambda} d\lambda \\ &= \frac{\theta^{\alpha+1}}{\theta^{\alpha+1} + \Gamma(\alpha+1)} \int_0^\infty \lambda^r \left[ \sum_{x=r}^\infty \frac{e^{-\lambda} \lambda^{x-r}}{(x-r)!} \right] (\theta + \lambda^\alpha) e^{-\theta\lambda} d\lambda \end{aligned}$$

Taking  $x+r$  in place of  $x$  within the bracket, we get

$$\mu_{(r)}' = \frac{\theta^{\alpha+1}}{\theta^{\alpha+1} + \Gamma(\alpha+1)} \int_0^\infty \lambda^r \left[ \sum_{x=0}^\infty \frac{e^{-\lambda} \lambda^x}{x!} \right] (\theta + \lambda^\alpha) e^{-\theta\lambda} d\lambda$$

The expression within the bracket is clearly unity and hence we have

$$\mu_{(r)}' = \frac{\theta^{\alpha+1}}{\theta^{\alpha+1} + \Gamma(\alpha+1)} \int_0^\infty \lambda^r (\theta + \lambda^\alpha) e^{-\theta\lambda} d\lambda$$

Using gamma integral and a little algebraic simplification, we get finally, a general expression for the  $r$  th factorial moment about origin of PSD (2.2) as

$$\mu_{(r)}' = \frac{\theta^{\alpha+1} \Gamma(r+1) + \Gamma(\alpha+r+1)}{\theta^r \{\theta^{\alpha+1} + \Gamma(\alpha+1)\}}; r = 1, 2, 3, \dots \quad (3.1)$$

The first four factorial moments about origin can thus be obtained as

$$\begin{aligned} \mu_{(1)}' &= \frac{\theta^{\alpha+1} + \Gamma(\alpha+2)}{\theta \{\theta^{\alpha+1} + \Gamma(\alpha+1)\}} \\ \mu_{(2)}' &= \frac{2\theta^{\alpha+1} + \Gamma(\alpha+3)}{\theta^2 \{\theta^{\alpha+1} + \Gamma(\alpha+1)\}} \\ \mu_{(3)}' &= \frac{6\theta^{\alpha+1} + \Gamma(\alpha+4)}{\theta^3 \{\theta^{\alpha+1} + \Gamma(\alpha+1)\}} \\ \mu_{(4)}' &= \frac{24\theta^{\alpha+1} + \Gamma(\alpha+5)}{\theta^4 \{\theta^{\alpha+1} + \Gamma(\alpha+1)\}}. \end{aligned}$$

Using the relationship between moments about origin and factorial moments about origin, the first two moments about origin of PSD can be expressed as

$$\begin{aligned} \mu_1' &= \mu_{(1)}' = \frac{\theta^{\alpha+1} + \Gamma(\alpha+2)}{\theta \{\theta^{\alpha+1} + \Gamma(\alpha+1)\}} \\ \mu_2' &= \mu_{(2)}' + \mu_{(1)}' = \frac{\theta^{\alpha+2} + 2\theta^{\alpha+1} + \theta\Gamma(\alpha+2) + \Gamma(\alpha+3)}{\theta^2 \{\theta^{\alpha+1} + \Gamma(\alpha+1)\}} \end{aligned}$$

Similarly, third and fourth moments about origin can be obtained. The variance of PSD can thus be obtained as

$$\mu_2 = \mu_2' - (\mu_1')^2 = \left( \frac{\theta^{\alpha+2} + 2\theta^{\alpha+1} + \theta\Gamma(\alpha+2) + \Gamma(\alpha+3)}{\theta^2 \{\theta^{\alpha+1} + \Gamma(\alpha+1)\}} \right) - \left( \frac{\theta^{\alpha+1} + \Gamma(\alpha+2)}{\theta \{\theta^{\alpha+1} + \Gamma(\alpha+1)\}} \right)^2.$$

The expressions for third and fourth moments about the mean are lengthy and hence are not being given. However, if they are required, can be obtained easily.

#### 4. Maximum Likelihood Estimate (mle)

Let  $(x_1, x_2, \dots, x_n)$  be a random sample of size  $n$  from the PSD (2.2) and let  $f_x$  be the observed frequency in the sample corresponding to  $X = x$  ( $x = 1, 2, 3, \dots, k$ ) 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 PSD (2.2) is given by

$$L = \left(\frac{\theta}{\theta+1}\right)^{n(\alpha+1)} \left(\frac{1}{\theta^{\alpha+1} + \Gamma(\alpha+1)}\right)^n \frac{1}{(\theta+1)^{\sum_{x=1}^k x f_x}} \prod_{x=1}^k \left[ \frac{\theta(\theta+1)^\alpha \Gamma(x+1) + \Gamma(x+\alpha+1)}{\Gamma(x+1)} \right]^{f_x}.$$

The log-likelihood function is thus obtained as

$$\begin{aligned} \log L = & n(\alpha+1) \log\left(\frac{\theta}{\theta+1}\right) - n \log[\theta^{\alpha+1} + \Gamma(\alpha+1)] - n \bar{x} \log(\theta+1) \\ & + \sum_{x=1}^k f_x \left[ \log\left\{ \theta(\theta+1)^\alpha \Gamma(x+1) + \Gamma(x+\alpha+1) - \log \Gamma(x+1) \right\} \right]' \end{aligned}$$

where  $\bar{x}$  is the sample mean.

The maximum likelihood estimates (MLE's  $(\hat{\theta}, \hat{\alpha})$ ) of the parameter  $(\theta, \alpha)$  of PSD (2.2) are the solutions of the following log-likelihood equations

$$\begin{aligned} \frac{\partial \log L}{\partial \theta} = & \frac{n(\alpha+1)}{\theta(\theta+1)} - \frac{n(\alpha+1)\theta^\alpha}{\theta^{\alpha+1} + \Gamma(\alpha+1)} - \frac{n\bar{x}}{(\theta+1)} + \sum_{x=1}^k \frac{[(\theta\alpha+1)(\theta+1)^\alpha \Gamma(x+1)] f_x}{\theta(\theta+1)^\alpha \Gamma(x+1) + \Gamma(x+\alpha+1)} = 0 \\ \frac{\partial \log L}{\partial \alpha} = & n \log\left(\frac{\theta}{\theta+1}\right) - \frac{n\theta^{\alpha+1} \log \theta + \psi(\alpha+1)}{\theta^{\alpha+1} + \Gamma(\alpha+1)} + \sum_{x=1}^k \frac{[\theta(\theta+1)^\alpha \log(\theta+1) \Gamma(x+1)] f_x}{[\theta(\theta+1)^\alpha \Gamma(x+1) + \Gamma(x+\alpha+1)]} = 0, \end{aligned}$$

where  $\psi(\alpha) = \frac{d}{d\alpha} \ln \Gamma(\alpha)$  is the digamma function. These two log likelihood equations do not seem to be solved directly because they do not have closed forms. Therefore, to find the maximum likelihood estimates of parameters an iterative method such as Fisher Scoring method, Bisection method, Regula Falsi method or Newton-Raphson method can be used. In this paper Newton-Raphson method has been used using R-software.

#### 5. Goodness of fit

The PSD has been fitted to a number of datasets to test its goodness of fit along with PD, PLD, NBD, GPLD and P-WLD., because PLD, NBD, GPLD and P-WLD are always over-dispersed and hence their comparison regarding goodness of fit is justifiable. The maximum likelihood estimate has been used to fit PD, PLD, NBD, GPLD, P-WLD and PSD for two examples of observed count datasets. The first dataset in table 2 is due to the data regarding the number of European red mites on apple leaves, available in Bliss (1953), the second dataset in table 3 is the Mammalian Cytogenetic dosimetry Lesions in Rabbit Lymphoblast induced by streptonigrin (NSC-45383), available in Catcheside et al (1946).

**Table 2:** Observed and Expected number of European red mites on Apple leaves, available in Catcheside et al (1946)

Number of European red mites per leaf	Observed frequency	Expected frequency					
		PD	PLD	NBD	GPLD	P-WLD	PSD
0	70	47.6	67.2	69.5	69.8	69.8	71.3
1	38	54.6	38.9	37.6	36.7	36.8	35.0
2	17	31.3	21.2	20.1	20.1	20.1	19.5
3	10	11.9	11.1	10.7	10.9	10.9	11.1
4	9	3.4	5.7	5.7	5.8	5.8	6.2
5	3	0.8	2.8	3.0	3.1	3.0	3.4
6	2	0.2	1.4	1.6	1.6	1.6	1.7
7	1	0.1	0.9	0.9	0.8	0.8	0.8
8	0	0.1	0.8	0.9	1.2	1.2	1.0
Total	150	150.0	150.0	150.0	150.0	150.0	150.0
ML estimate		$\hat{\theta} = 1.14666$	$\hat{\theta} = 1.26010$	$\hat{\alpha} = 1.02459$ $\hat{p} = 0.52811$	$\hat{\theta} = 1.09620$ $\hat{\alpha} = 0.78005$	$\hat{\theta} = 1.09141$ $\hat{\alpha} = 0.82194$	$\hat{\theta} = 1.8444$ $\hat{\alpha} = 3.1231$
Standard Errors		0.08743	0.11390	0.42097 0.40136	0.25400 0.31550	0.26231 0.25230	0.6409 2.0978
$\chi^2$		26.50	2.49	2.91	2.43	2.41	2.09
d.f		2	4	3	3	3	3
p-value		0.0000	0.5595	0.4057	0.4880	0.4917	0.5539
$-2\log L$		485.61	445.02	469.68	444.62	425.35	444.09
AIC		487.61	447.02	447.02	448.62	429.35	448.09

**Table 3:** Mammalian Cytogenetic dosimetry Lesions in Rabbit Lymphoblast induced by streptonigrin (NSC-45383), Exposure- 90  $\mu g | kg$

Class/Exposure ( $\mu g   kg$ )	Observed frequency	Expected frequency					
		PD	PLD	NBD	GPLD	P-WLD	PSD
0	155	127.8	158.3	155.1	155.3	155.9	158.9
1	83	109.0	77.2	80.6	80.1	80.0	76.8
2	33	46.5	35.9	36.7	36.9	36.7	35.5
3	14	13.2	16.1	15.9	16.0	15.9	16.0
4	11	2.8	7.1	6.7	6.7	6.7	7.1
5	3	0.5	3.1	2.8	2.8	2.7	3.2
6	1	0.2	2.3	2.2	2.2	2.1	2.5
Total	300	300.0	300.0	300.0	300.0	300.0	300.0
ML estimate		$\hat{\theta} = 0.85333$	$\hat{\theta} = 1.61761$	$\hat{\theta} = 1.56009$ $\hat{\alpha} = 1.33128$	$\hat{\theta} = 1.80860$ $\hat{\alpha} = 1.18743$	$\hat{\theta} = 1.82011$ $\hat{\alpha} = 1.16320$	$\hat{\theta} = 1.4384$ $\hat{\alpha} = 0.7011$
SE( $\hat{\theta}$ ) SE( $\hat{\alpha}$ )		0.05333	0.11327	0.41479 0.33752	0.40045 0.37007	0.41992 0.32483	0.3398 1.1181
$\chi^2$		24.969	1.51	1.60	1.69	1.78	1.40
d.f		2	3	2	2	2	2
p-value		0.0000	0.6799	0.4488	0.42955	0.4106	0.4965
$-2\log L$		800.92	766.10	765.86	765.79	834.51	766.32
AIC		802.92	768.10	769.86	769.79	838.51	770.32

The fitted plots of the distributions for dataset in tables 2 and 3 have been presented in figure 3 and 4 respectively.

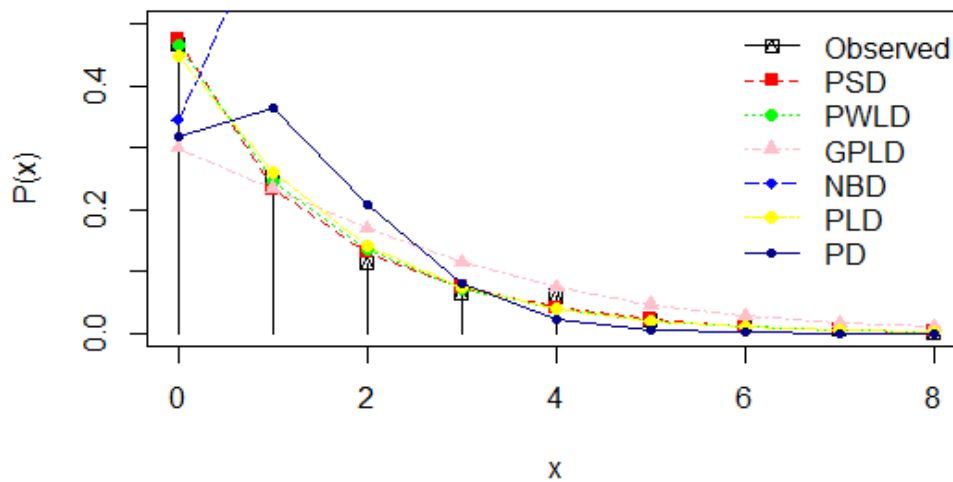


Fig.3: Fitted plot of distributions for datasets in table 2

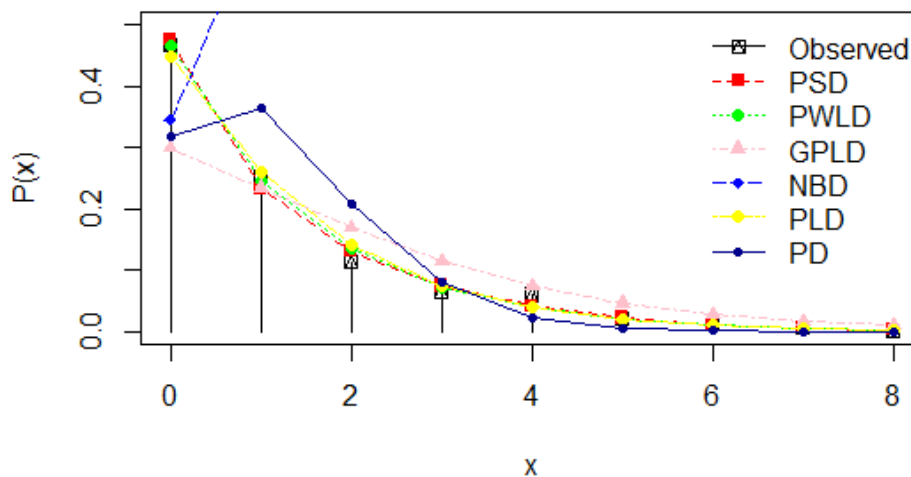


Fig. 4: Fitted plot of distributions for datasets in table 3

It is clear from the goodness of fit of the considered distribution in tables 3 and 4 and the fitted plots of the distributions that PSD is gives much closer fit.

## 6. Concluding Remarks

In this paper Poisson-Shukla distribution (PSD) has been proposed. The expression for the  $r$ th factorial moment about origin has been derived and hence its mean and variance have been obtained. The maximum likelihood estimation for estimating its parameter has been discussed. The goodness of fit of PSD over PD, PLD, NBD, GPLD and P-WLD have been discussed with two examples of observed real datasets. It is observed that PSD gives much better fit than PD, PLD, NBD, GPLD and P-WLD on all the datasets and hence it can be considered an important discrete distribution for modeling count data over these distributions.

## References

1. Bliss, C.I. (1953): Fitting negative binomial distribution to biological data, *Biometrics*, 9, 177 – 200.
2. Catcheside, D.G., Lea, D.E. & Thoday, J.M. (1946). The production of chromosome structural changes in *Tradescantia* microspores in relation to dosage, intensity and temperature, *Journal of Genetics*, 47, 137-149.
3. El-Monsef, M.M.E. and Sohsah, N.M. (2014): Poisson-Weighted Lindley Distribution, *Jokull Journal*, 64(5), 192 – 202
4. Ghitany ,M.E., Alqallaf ,F., Al-Mutairi, D.K., Husain, H.A. (2011): A two-parameter weighted Lindley distribution and its applications to survival data, *Mathematics and Computers in simulation*, 81, 1190-1201.
5. Greenwood, M. and G.U., Yule, (1920) “An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the multiple attacks of disease or of repeated accidents”, *Journal of the Royal Statistical Society*, 83(2), 115 – 121
6. Lindley, D.V. (1958): Fiducial distributions and Bayes theorem, *Journal of the Royal Statistical Society*, 20 (1), 102- 107.
7. Mahmoudi, E. and Zakerzadeh, H. (2010): Generalized Poisson-Lindley distribution, *Communications in Statistics-Theory & Methods*, 39, 1785 – 1798.
8. Sankaran, M. (1970): The discrete Poisson-Lindley distribution, *Biometrics*, 26, 145- 149.
9. Shanker, R. and Shukla, K. K. (2019): On Poisson weighted Lindley distribution and its Applications *Journal of Scientific Research*, 11(1), 1-13.
10. Shukla, K.K. and Shanker, R (2019): Shukla distribution and its Application, *R.T.&A*,14(3), 46-55
11. Zakerzadeh, H. and Dolati, A. (2009): Generalized Lindley distribution, *Journal of Mathematical extension*, 3 (2), 13 – 25.