A discrete Pranav distribution and its applications

Abstract

In the recent decades, the discretization of continuous distribution has been attracting the attention of researchers because it generates distributions that can be used for strictly discrete data. In this paper, a discrete Pranav distribution, which is a discrete analogue of continuous Pranav distribution, has been carried out. It’s important properties including coefficient of variation, skewness, kurtosis and index of dispersion have been obtained and discussed graphically. The method of maximum likelihood estimation has been used for estimating its parameter. The goodness of fit of the proposed distribution have been illustrated using some real count datasets and it was found better fit as compared to other one parameter discrete distributions.

Keywords: Pranav distribution, discretization, moment generating function, moments, estimation, goodness of fit

Introduction

In recent past years, the use of discrete analogue of a continuous distribution avoids the use of a continuous distribution in the case of strictly available discrete data.

In many cases, it is not easy to get samples from continuous distributions. The observed values, in the most of cases, are collected actually discrete in nature for the reason that they are measured to only finite number of decimal places and cannot completely presents all points in a continuum. According to Lai,1 discretization of a continuous lifetime model is an appealing approach to derive a discrete lifetime model corresponding to the continuous one. Therefore, it is reasonable and convenient to model the situation by an appropriate discrete distribution generated from the underlying continuous distribution preserving one or more important characteristics including probability density function (pdf), mean residual life function etc. and important statistical properties of the distribution.

In Statistics literature, different researchers have used different methods of discretization to propose a discrete type of distribution analogues of continuous distribution. In this study, one of the discretization methods has been used to find discrete analogues of continuous Pranav distribution introduced by Shukla.2 Infinite series method has been used to find pmf of Pranav distribution, appropriate definition is given in the next paragraph. This method was firstly used by Good3 who proposed the discrete Good distribution to model for frequencies of species. It was given as follows:

A random variable $Y$ is said to have a discrete Good distribution if its pmf can be expressed as

$$P(Y = y) = \alpha^y \beta^y \frac{\theta^y}{\sum_{j=0}^{\infty} \alpha^j \beta^j}; y = 0, 1, 2, ..., \ (1.1)$$

where $\beta \in R$ and $\alpha \in (0,1)$.

The method of infinite series is formulated by the definition which is given as below:

Definition: Let $X$ be a continuous random variable having pdf $f_X(x)$ with support on $R$. Then the corresponding discrete random variable $Y$ has pmf given by

$$P(Y = y) = P(Y = y; \theta) = \frac{f_X(y; \theta)}{\sum_{j=-\infty}^{\infty} f_X(j; \theta)}; y \in Z, \ (1.2)$$

where $\theta$ may be the vector of parameters indexing the distribution of $X$.

This method has been used by many researchers to derive discretization of continuous distribution, such as, Kulasekara & Tonkyn,4 Doray & Luong,5 Sato et al.,6 Nekoukhou et al.,7 some are among others, who proposed a version of the method when the continuous random variable of interest is defined on $R$. Thus, if the random variable $X$ is defined on $R$, the pmf of $Y$ can be defined as

$$P(Y = y) = P(Y = y; \theta) = \frac{f_X(y; \theta)}{\sum_{j=0}^{\infty} f_X(j; \theta)}; y \in Z, \ (1.3)$$

Josmar et al.,8 using infinite series method has derived a discrete Shanker distribution (DSD) with parameter $\theta > 0$ and having pmf

$$P_1(Y = y; \theta) = \frac{(\theta + y)^{\theta} e^{\theta(y+1)}}{1 + (\theta + 1)^{\theta}}; y = 0, 1, 2, ... \ (1.4)$$

They have discussed its various statistical properties including its applications to model count datasets in their paper. Which is a discrete analogue of continuous Shanker distribution introduced by Shanker9 having pdf

$$f_1(x; \theta) = \theta e^{\theta x} (\theta + x) e^{-\theta x}; x > 0, \theta > 0 \ (1.5)$$

Using same method of discretization, the pmf of discrete Lindley distribution (DLD) proposed by Berhane & Shanker10 is given by

$$P_2(Y = y; \theta) = \frac{(\theta + 1)^2 e^{\theta y}}{e^{\theta} (1 + \theta e^{-\theta y})}; y = 0, 1, 2, ... \ (1.6)$$

where the parameter $\theta > 0$.

They have discussed its important statistical properties including estimation of parameter of DLD and applied on some count datasets from engineering and biology in their paper. They showed its superiority over other discrete one parameter distributions such as Poisson Lindley distribution (PLD) proposed by Shankar,11 Poisson Akash distribution (PAD) proposed by Shanker,12 and DSD proposed by Josmer et al.8 As mentioned above, DLD is a discrete analogue of continuous Lindley distribution introduced by Lindley13 having pdf

$$f_2(x; \theta) = \theta e^{\theta x} (1 + \theta x) e^{-\theta x}; x > 0, \theta > 0 \ (1.7)$$
Recently, Berhanu & Shanker,\textsuperscript{14} proposed a discrete Akash distribution (DAD) using infinite series method, the pmf of a discrete Akash distribution is given as
\[
P_x(x; \theta) = \frac{\theta^x}{\theta^2 + 2} \frac{(1 + x^2) e^{-\theta x}}{\theta + 1}; \quad x = 0, 1, 2, \ldots \quad (1.10)
\]

Shanker\textsuperscript{15} proposed PAD, a Poisson mixture of Akash distribution, having pmf
\[
P_x(x; \theta) = \frac{\theta^x}{\theta^2 + 2} \frac{x^2 + 3x + (\theta^2 + 2\theta + 3)}{(\theta + 1)^{x+1}}; \quad x = 0, 1, 2, \ldots, \theta > 0 \quad (1.11)
\]

He has discussed its important statistical properties including estimation of parameter along with applications of PAD in his paper. PAD was applied to count datasets and showed its superiority with PLD and other distribution of one parameter.

The PLD is a Poisson mixture of Lindley distribution introduced by Sankaran\textsuperscript{13} and is defined by its pmf
\[
P_x(x; \theta) = \frac{\theta^x}{\theta^2 + 2} \frac{(x + \theta + 2)}{(\theta + 1)^{x+1}}; \quad x = 0, 1, 2, \ldots, \theta > 0 \quad (1.12)
\]

The main objective of this paper is to propose a discretization of Pranav distribution for the reason being that it was observed, Pranav distribution gives better fit than one parameter continuous Akash, Shanker, Sujatha, Lindley and Exponential distributions. Keeping this view in mind, it is hoped that it would be better than discrete Akash, discrete Shanker and discrete Lindley distributions and other one parameter discrete distributions.

In this paper, Discrete Pranav distribution has been proposed in the second section. It’s important properties including coefficients of variation, skewness, kurtosis and index of dispersion have been derived and discussed graphically in the third section. The method of maximum likelihood estimation has been discussed in fourth section. The applications of proposed distribution have been illustrated on two count datasets in fifth section. Conclusions on the basis of result have been discussed in the last section.

**A discrete Pranav distribution**

The pdf and the cdf of a random variable $X$ having Pranav distribution proposed by Shukla\textsuperscript{2} are given by
\[
f_x(x; \theta) = \frac{\theta^x}{\theta^2 + 6} \left( x + 1 \right) e^{-\theta x}; \quad x > 0, \quad \theta > 0 \quad (2.1)
\]
\[
F_x(x; \theta) = 1 - \frac{\theta x^2 + 3x + 6}{\theta^2 + 6} e^{-\theta x}; \quad x > 0, \theta > 0 \quad (2.2)
\]

Shukla\textsuperscript{2} has discussed its various statistical properties including moments based measures, hazard rate function, and other important properties along with Bonferroni and Lorenz curves and stress-strength reliability. Pranav distribution applied for modeling lifetime data from biomedical sciences and engineering and explained its superiority over Akash, Shanker, Ishita, Sujatha, and exponential distributions.

Using the above definition, the pmf of the discrete random variable $Y$, corresponding to a continuous random variable $X$ follows Pranav distribution (2.1) with parameter $\theta > 0$, can be obtained as
\[
P(y; \theta) = \frac{\theta^y}{\theta^{y+1} + 3\theta^y + 4e^{\theta} + 1} \left( e^{\theta y} + 3e^{\theta} + 4e^{\theta y} + 1 \right) \quad (2.3)
\]

We would call this distribution, a discrete Pranav distribution (DPD). The nature and behavior of DPD for varying values of its parameter $\theta$ has been shown graphically in Figure 1. From the figure it was observed that pmf of DPD is increasing as increased values of $\theta$ (Figure 1).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure1.png}
\caption{The pmf plot of DPD for varying values of the parameter $\theta$.}
\end{figure}

The survival function, $S(y; \theta)$ and the cumulative distribution function (cdf), $F(y; \theta)$ of DPD can be obtained as
\[
S(y; \theta) = \frac{\theta^y}{\theta^{y+1} + 3\theta^y + 4e^{\theta} + 1} \left( e^{\theta y} + 3e^{\theta} + 4e^{\theta y} + 1 \right) \quad (2.5)
\]
\[
F(y; \theta) = 1 - \frac{\theta^y}{\theta^{y+1} + 3\theta^y + 4e^{\theta} + 1} \left( e^{\theta y} + 3e^{\theta} + 4e^{\theta y} + 1 \right) \quad (2.6)
\]

cdf graphs of DPD are presented in Figure 2. Since $P(y; \theta)$ is log-concave and therefore, the DPD has an increasing hazard rate. Further, $[P(y; \theta)]^2 \geq P(y - 1; \theta)$ is also log-concave and $[P(y; \theta)]^2 \geq P(y - 1; \theta)$ for $y \geq 3$, which implies unimodality, by theorem 3 of Keilson & Gerber.\textsuperscript{16} The detailed about interrelationship between log-concavity, unimodality and increasing hazard rate of discrete distributions can be shown in Grandell.\textsuperscript{17}

**Mean variance and statistical constants**

The probability generating function and the moment generating function (mgf) of DPD can be obtained as
A discrete Pranav distribution and its applications

The first four moments about origin of DPD can thus be obtained

\[ \mu_4 = \frac{\theta^2 (\theta + 1) e^{2\theta} - (8\theta^3 - 47\theta^2 + \theta) e^{\theta} + (27\theta - 96\theta^2 + 15\theta) e^{2\theta}}{\theta^3 (\theta + 1) e^{2\theta} - (4\theta^3 - 342\theta^2 + 60\theta - 8) e^{\theta} + (42\theta^3 + 132\theta^2 + 300\theta + 32) e^{2\theta}} \]

\[ - \frac{(-1536\theta^2 - 3990 + 28) e^{\theta} - (42\theta^3 - 438\theta^2 + 780 - 536) e^{2\theta}}{\theta^3 (\theta + 1) e^{2\theta} - (8\theta^3 - 65\theta^2 - 15\theta + 32) e^{\theta} - (109\theta^2 - 17\theta + 8) e^{2\theta}} \]

\[ \mu_2 = \frac{\theta e^{\theta} - (59\theta - 106\theta^2 + \theta) e^{2\theta} - (179\theta^2 + 48\theta - 11\theta^2 + \theta) e^{3\theta}}{\theta e^{\theta} - (2300\theta^3 - 3184\theta^2 + 606\theta + 92\theta) e^{2\theta} - (9790\theta^2 - 12857\theta^2 + 578\theta^2 + 1136\theta) e^{3\theta}} \]

\[ + \frac{4884\theta - 4386\theta^2 + 22116\theta + 13416\theta + 8400\theta e^{2\theta}}{\theta e^{\theta} - (4125\theta^3 + 55749\theta^2 + 97886\theta + 17856\theta + 20544\theta e^{2\theta}} \]

\[ + \frac{(50\theta^3 + 2423\theta^2 + 700\theta + 280\theta + 10\theta^2 + 21\theta^2 - 25\theta + 8) e^{2\theta}}{\theta e^{\theta} - (483\theta^3 + 4100\theta^2 + 33615\theta^2 + 26646\theta + 30528\theta e^{2\theta}} \]

\[ - \frac{9790 + 13987\theta^2 + 3047\theta - 2044\theta e^{2\theta}}{\theta e^{\theta} - (2300\theta^3 + 1230\theta^2 - 5232\theta + 840\theta e^{2\theta}} \]

\[ - \frac{(176\theta^2 - 1291\theta^2 + 642\theta + 2864\theta + 322\theta e^{2\theta}}{\theta e^{\theta} - (2300\theta^3 + 352\theta^2 - 5236\theta + 840\theta e^{2\theta}} \]

The coefficient of variation (C.V), coefficient of skewness (\(\beta_1\)), coefficient of kurtosis (\(\beta_2\)) and index of dispersion (\(\gamma\)) of DPD can be obtained using the relationships below

\[ C.V = \frac{\sigma}{\mu_1} \sqrt{\beta_1} = \frac{\mu_3}{(\mu_2)^{3/2}} \beta_2 = \frac{\mu_4 - \mu_2^2}{\mu_2} \gamma = \frac{\sigma^2}{\mu_1} \]
It is clear from Table 1 that the mean, variance, and index of dispersion of DPD are decreasing as increased values of parameter $\theta$, whereas coefficient of variation, coefficient of skewness and coefficient of kurtosis of DPD are increasing as increased values of parameter $\theta$. $\sigma^2 > \mu$, indicates that DPD can be a suitable model for over-dispersed data.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Values of descriptive statistics of DPD for varying values of $\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>Mean</td>
</tr>
<tr>
<td>0.5</td>
<td>7.915</td>
</tr>
<tr>
<td>1.0</td>
<td>3.2844</td>
</tr>
<tr>
<td>1.5</td>
<td>1.1918</td>
</tr>
<tr>
<td>2.0</td>
<td>0.4147</td>
</tr>
<tr>
<td>2.5</td>
<td>0.1712</td>
</tr>
<tr>
<td>3.0</td>
<td>0.0825</td>
</tr>
<tr>
<td>3.5</td>
<td>0.0438</td>
</tr>
<tr>
<td>4.0</td>
<td>0.0245</td>
</tr>
</tbody>
</table>

### Maximum likelihood estimation

The likelihood function, $L$ of (2.3) is given by

$$L = \left(\frac{(e^\theta - 1)^4}{e^\theta(\theta e^\theta - 1)^3 + e^{2\theta} + 4e^\theta + 1}\right)^n \prod_{i=1}^{n} \left(\theta + y_i\right)e^{-\theta y_i}.$$

Table 2: Distribution of mistakes in copying groups of random digits

<table>
<thead>
<tr>
<th>No. of error per group</th>
<th>Observed Frequency</th>
<th>Expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLD</td>
<td>PAD</td>
</tr>
<tr>
<td>0</td>
<td>35</td>
<td>33.1</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>15.3</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>6.7</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2.9</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>60</td>
</tr>
</tbody>
</table>

$\hat{\theta} = 1.7434 \quad \text{d.f.} = 20 \quad \chi^2 = 1.8141 \quad 1.4185 \quad 3.3667 \quad 2.9963 \quad 1.0398 \quad 0.1712$

$p$-value = 0.178 0.492 0.066 0.0837 0.595 0.919

Table 3: Observed and expected frequencies for distribution of Pyrausta nubilalis in 1937

<table>
<thead>
<tr>
<th>No. of insects</th>
<th>Observed Frequency</th>
<th>Expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLD</td>
<td>PAD</td>
</tr>
<tr>
<td>0</td>
<td>33</td>
<td>31.5</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>14.2</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>6.1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2.5</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>56</td>
</tr>
</tbody>
</table>

$\hat{\theta} = 1.8115 \quad \text{d.f.} = 20 \quad \chi^2 = 0.4598 \quad 0.2541 \quad 1.5422 \quad 1.1516 \quad 0.1747 \quad 0.6055$

$p$-value = 0.498 0.614 0.215 0.283 0.676 0.739

A discrete Pranav distribution and its applications

Copyright: ©2019 Abebe et al.

Conclusion

In this paper, a discrete Pranav distribution (DPD) has been proposed. Its moment generating function, moments and moments based measures including statistical constants have been derived and their nature and behavior has been discussed numerically and graphically. The method of maximum likelihood estimation has been discussed for estimating its parameter. The goodness of fit of DPD has been explained using two real count datasets. The DPD gives better fit as compared to PLD, PAD, DLD, DSD and DAD in the presented datasets.

Acknowledgments

None.

Conflicts of interest

Author declares that there is no conflict of interest.

References