Generalised Poisson-Modified Mishra distribution

1*Binod Kumar Sah, 2Suresh Kumar Sahani
1*Department of Statistics, R.R.M. Campus, Janakpur Dham, Tribhuvan University, Nepal.
2Department of Mathematics, MIT Campus, Janakpur Dham, Nepal.
1*sah.binod01@gmail.com
2sureshkumarsahani35@gmail.com

Abstract:
This distribution is a compound probability distribution constructed by compounding generalised Poisson distribution of Consul and Jain (1973) with Modified Mishra distribution of Sah (2022). All the important characteristics required for this distribution have been obtained and explained. In order to check the validity of the theoretical work, goodness of fit has been applied to some secondary data which have been used by others. It has been observed that this distribution gives a better fit than Poisson-Lindley distribution of Sankaran (1970) as well as Poisson-Modified Mishra distribution of Sah and Sahani (2023).

Keywords: Poisson-Modified Mishra distribution, Modified Mishra distribution, Moments, Distribution, Generalised, Mixing.

Introduction:
The word compounding is also called mixing. We can see that the use of this word in every field such as in literature, in Chemistry, in Mathematics, in Statistics and many more social and physical sciences. Let us consider two distributions \( f_A \) and \( f_B \). Let \( f_A \) denotes generalised Poisson distribution (GPD) of Consul and Jain (Consul and Jain, 1973). It is generalised case of Poisson distribution having parameters \( \lambda \) and \( \theta \). Here \( \theta \) is an additional parameter and it lies between -1 to +1. Probability mass function (pmf) of GPD is given by

\[
P_Y(y; \lambda, \theta) = \frac{\lambda(\lambda + \theta y)^{y-1} e^{-(\lambda + \theta y)}}{y!}; \lambda > 0, |\theta| \leq 1
\]

(1)

Let \( f_B \) denotes Modified Mishra distribution (MMD) of Sah (Sah, 2022). It is a continuous probability distribution having a single parameter \( \alpha \) and its probability density function (pdf) is given by

\[
f_Y(y; \alpha) = \frac{\alpha^\lambda}{(2 + \alpha + \pi \alpha^2)}(\pi + y + y^2)e^{-\alpha y}; \alpha > 0, y > 0
\]

(2)

\( f_B \) mixture of \( f_A \) can be written as \( f_A \sim f_B \). Let \( \lambda \) follows \( f_B \) distribution. It is also called mixing distribution. Here, mixing distribution we consider is MMD which is distribution of \( \lambda \). Hence, Modified Mishra mixture of GPD can be symbolically written as
Poisson-Modified Mishra distribution (PMMD) was obtained (Sah and Sahani, 2023) by mixing Poisson distribution with MMD and its pmf was given by

\[ P(Y; \alpha) = \left( \frac{\alpha^3}{(2+\alpha + \pi \alpha^2)} \right) \frac{(1+\alpha)(1+\pi + \pi\alpha + y) + (1+y)(2+y)}{(1+\alpha)^{\alpha+1}} ; y = 0, 1, 2, ...; \alpha > 0 \]

(3)

It is a particular case of the proposed distribution at \( \theta = 0 \). Sah obtained (a) On a generalised exponential-Lindley mixture of generalised Poisson distribution (Sah and Mishra, 2020) and it was obtained by mixing GPD with generalized exponential-Lindley distribution (Mishra and Sah, 2015). (b) A generalised negative binomial distribution and its important features (Sah, 2018). (c) A generalised Neyman Type-A distribution (Sah and Mishra, 2021) and (d) A two-parameter quasi-Lindley mixture of generalised Poisson distribution (Sah, 2015). It was obtained by mixing GPD with Two-Parameter quasi-Lindley distribution (Sah, 2015). All these distributions are based on three parameters and play a very important role for statistical modeling of over-dispersed count data-sets. The important point is that here we are not going to compare GPMMD with these distributions because GPMMD has only two parameters.

generalised Poisson-Lindley distribution, (GPLD), (Sah, 2013) and generalised Poisson-Mishra distribution, (GPMD), (Sah, 2018) both are based on only two parameters. We may compare GPLD and GPMD with GPMMD.

It has been observed that the proposed distribution gives a very good fit to the data sets which are over dispersed and under-dispersed in nature because of an additional parameter \( \theta \).

The first section of this paper includes an introduction part and a literature review. The second section of this paper discusses the materials and methods required for this distribution. In the third and main section, which is generally named as results section, results obtained have been presented. This section is further divided into the following sub-sections.

- Probability Mass Function of GPMMD
- Statistical Moments of GPMMD
- Estimation of the Parameters of GPMMD and
- Applications of GPMMD

In the last section of this paper, the conclusions about this distribution have been included.

Material and Methods:

Work of this paper is to formulate theoretical concepts and move towards experimentation and it has been conducted in the following steps.

- Construction of probability mass function of GPMMD
- Estimation of parameters
- Test of validity of theoretical work.
For using goodness of fit, the data used are secondary in nature and have already been used by others.

Results:
The work of this paper is divided into the following sub-topics for better understanding and recall of the work.

- Probability Mass Function of GPMMD
- Statistical Moments of GPMMD
- Estimation of the Parameters of GPMMD and
- Applications of GPMMD

**Probability Mass Function of GPMMD:**

The proposed distribution is a continuous mixture of generalized Poisson distribution. Let $\alpha$ be a parameter of Modified Mishra distribution. Let $\theta$ and $\lambda$ the parameters of the GPD. The parameter $\lambda$ of GPD is continuous in nature and it follows MMD. The proposed distribution can be obtained by mixing GPD with MMD as follows

$$P(Y; \alpha, \theta) = \left( \frac{\alpha e^{-\theta y}}{y(2 + \alpha + \pi y^2)} \right) \int_0^{\infty} \lambda^{r-1} e^{-(1+r)\lambda} \left( \pi e^{-\theta y} \right) e^{-\lambda^2 y} d\lambda ; y = 0, 1, 2, ...; 1 > 0; \alpha > 0$$

$$= \left( \frac{\alpha^3 e^{-\theta y}}{y(2 + \alpha + \pi y^2)} \right) \int_0^{\infty} \left( \frac{\theta y}{\lambda} \right)^{r-1} e^{-(1+r)\lambda^2} e^{-\lambda^2 y} d\lambda$$

$$= \left( \frac{\alpha^3 e^{-\theta y}}{y(2 + \alpha + \pi y^2)} \right) \left[ \sum_{i=0}^{\infty} \theta (y-i) y^{-1} \pi (1+\alpha)^{y-i} + (1+\alpha)(y-i+1) + (y-i+1) \right]$$

$$= \left( \frac{\alpha^3 e^{-\theta y}}{y(2 + \alpha + \pi y^2)} \right) \left[ \sum_{i=0}^{\infty} \theta (y-i) y^{-1} \pi (1+\alpha)^{y-i} + (1+\alpha)(y-i+1) + (y-i+1) \right]$$

$$= \left( \frac{\alpha^3 e^{-\theta y}}{y(2 + \alpha + \pi y^2)} \right) \left[ \sum_{i=0}^{\infty} \theta (y-i) y^{-1} \pi (1+\alpha)^{y-i} + (1+\alpha)(y-i+1) + (y-i+1) \right]$$

The expression (4) gives pmf of GPMMD and it reduces to pmf of PMMD at $\theta = 0$.

**Statistical Moments of GPMMD:**
The $r^{th}$ moment about origin of GPMMD can be obtained as
\[ \mu' = E\left[ \frac{\alpha^3}{(2 + \alpha + \pi \alpha^2)} \int_0^\infty \sum_{y=0}^\infty \frac{y^\prime \lambda (\lambda + \theta y)^{-1} e^{-(\lambda + \theta y)}}{y!} (\pi + \lambda + \lambda^2) e^{-\alpha y} d\lambda \right] \]  

(5)

The mean of GPMMD can be obtained as

\[ \mu'(r) = \frac{\alpha^3}{(2 + \alpha + \pi \alpha^2)} \int_0^\infty \frac{\lambda}{(1 - \theta)^3} (\pi + \lambda + \lambda^2) e^{-\alpha y} d\lambda \]

\[ = \frac{\alpha^3}{(2 + \alpha + \pi \alpha^2)(1 - \theta)} \left[ \frac{\pi}{\alpha^2} + \frac{2}{\alpha^2} + \frac{6}{\alpha^2} \right] = \frac{(\pi \alpha^2 + 2\alpha + 6)}{\alpha(2 + \alpha + \pi \alpha^2)(1 - \theta)} \]  

(6)

Substituting \( r = 2 \) in the equation (5), the second moment about the origin of GPMMD has been obtained as follows

\[ \mu''(2) = \frac{\alpha^3}{(2 + \alpha + \pi \alpha^2)} \int_0^\infty \left( \sum_{y=0}^\infty \frac{y^2 \lambda (\lambda + \theta y)^{-1} e^{-(\lambda + \theta y)}}{y!} \right) (\pi + \lambda + \lambda^2) e^{-\alpha y} d\lambda \]

\[ = \frac{\alpha^3}{(2 + \alpha + \pi \alpha^2)(1 - \theta)^2} \left[ \frac{\lambda^2}{(1 - \theta)^2} \right] (\pi + \lambda + \lambda^2) e^{-\alpha y} d\lambda \]

\[ = \frac{\alpha^3}{(2 + \alpha + \pi \alpha^2)(1 - \theta)^2} \left[ \frac{1}{(1 - \theta)^2} \right] \left[ \frac{\pi}{\alpha^2} + \frac{2}{\alpha^2} + \frac{6}{\alpha^2} \right] + \frac{1}{(1 - \theta)^2} \left[ \frac{2\pi}{\alpha^2} + \frac{6}{\alpha^2} + \frac{24}{\alpha^2} \right] \]

\[ = \frac{(\pi \alpha^2 + 2\alpha + 6)}{\alpha(2 + \alpha + \pi \alpha^2)(1 - \theta)^2} + \frac{(2\pi \alpha^2 + 6\alpha + 24)}{\alpha^2(2 + \alpha + \pi \alpha^2)(1 - \theta)^2} \]  

(7)

The third moment about the origin is obtained as

\[ \mu''(3) = \frac{\alpha^3}{(2 + \alpha + \pi \alpha^2)} \int_0^\infty \left( \sum_{y=0}^\infty \frac{y^3 \lambda (\lambda + \theta y)^{-1} e^{-(\lambda + \theta y)}}{y!} \right) (\pi + \lambda + \lambda^2) e^{-\alpha y} d\lambda \]

\[ = \frac{\alpha^3}{(2 + \alpha + \pi \alpha^2)(1 - \theta)^3} \left[ \lambda (1 + 2\theta) \right] + \frac{3\lambda^2}{(1 - \theta)^3} + \frac{\lambda^3}{(1 - \theta)^3} \left[ (\pi + \lambda + \lambda^2) e^{-\alpha y} d\lambda \right] \]

\[ = \frac{\alpha^3}{(2 + \alpha + \pi \alpha^2)(1 - \theta)^3} \left[ (1 + 2\theta) \right] \left[ \frac{\pi}{\alpha^2} + \frac{2}{\alpha^2} + \frac{6}{\alpha^2} \right] + \frac{3}{(1 - \theta)^3} \left[ \frac{2\pi}{\alpha^2} + \frac{6}{\alpha^2} + \frac{24}{\alpha^2} \right] + \frac{1}{(1 - \theta)^3} \left[ \frac{6\pi}{\alpha^2} + \frac{24}{\alpha^2} + \frac{120}{\alpha^2} \right] \]

\[ = \frac{(1 + 2\theta)(\pi \alpha^2 + 2\alpha + 6) + 3(2\pi \alpha^2 + 6\alpha + 24) + (6\pi \alpha^2 + 24\alpha + 120)}{(1 - \theta)^3 \alpha(2 + \alpha + \pi \alpha^2)} + \frac{3(2\pi \alpha^2 + 6\alpha + 24)}{(1 - \theta)^4 \alpha^2(2 + \alpha + \pi \alpha^2)} + \frac{(6\pi \alpha^2 + 24\alpha + 120)}{(1 - \theta)^3 \alpha^2(2 + \alpha + \pi \alpha^2)} \]  

(8)

The fourth moment about origin has been obtained as
\[ \mu'_i = \frac{\alpha^2}{(2 + \alpha + \pi \alpha^2)} \int_0^\infty \sum_{n=0}^{\infty} \frac{y^n \lambda(\lambda + 2\theta y)^{n-1} e^{-(\lambda + 2\theta y)}}{y!} (\pi + \lambda + \lambda^2) e^{-\alpha y} d\lambda \]
\[ = \frac{\alpha^2}{(2 + \alpha + \pi \alpha^2)} \int_0^\infty \frac{\lambda(1 + 8\theta + 6\theta^2)}{(1 - \theta)^6} \frac{\lambda^2(7 + 8\theta)}{(1 - \theta)^6} + \frac{6\lambda^3}{(1 - \theta)^6} + \frac{\lambda^4}{(1 - \theta)^6} \left(\pi + \lambda + \lambda^2\right) e^{-\alpha y} d\lambda \]
\[ = \frac{\alpha^2}{(2 + \alpha + \pi \alpha^2)} \times \left[ \frac{(1 + 8\theta + 6\theta^2)}{(1 - \theta)^6} \left(\frac{\pi}{\alpha^2} + \frac{2}{\alpha^2} + \frac{6}{\alpha^4}\right) + \frac{(7 + 8\theta)}{(1 - \theta)^6} \left(\frac{2\pi}{\alpha^2} + \frac{6}{\alpha^2} + \frac{24}{\alpha^4}\right) \right] \]
\[ = \frac{(1 + 8\theta + 6\theta^2)(\pi \alpha^2 + 2\alpha + 6) + (7 + 8\theta)(2\pi \alpha^2 + 6\alpha + 24) + 6(6\pi \alpha^2 + 24\alpha + 120)}{(1 - \theta)^6 \alpha^2(2 + \alpha + \pi \alpha^2)} + \frac{(2\pi \alpha^2 + 120\alpha + 720)}{(1 - \theta)^6 \alpha^2(2 + \alpha + \pi \alpha^2)} \]
\[ = \frac{(1 + 8\theta + 6\theta^2)(\pi \alpha^2 + 2\alpha + 6)(2 + \alpha + \pi \alpha^2) + (1 - \theta)(2 + \alpha + \pi \alpha^2)(2\pi \alpha^2 + 6\alpha + 24) - (1 - \theta)(\pi \alpha^2 + 2\alpha + 6)^2}{(1 - \theta)^6 \alpha^2(\alpha(\pi \alpha^2 + \alpha + 2)^2)} \]
\[ = \frac{\alpha(\pi \alpha^2 + 2\alpha + 6)(2 + \alpha + \pi \alpha^2) + (1 - \theta)(2 + \alpha + \pi \alpha^2)(2\pi \alpha^2 + 6\alpha + 24) - (1 - \theta)(\pi \alpha^2 + 2\alpha + 6)^2}{(1 - \theta)^6 \alpha^2(\alpha(\pi \alpha^2 + \alpha + 2)^2)} \]
\[ \text{Theorem (1): GPMMD is an over-dispersed distribution.} \]

Proof: A distribution is said to be over-dispersed if Variance > Mean
\[ \alpha(\pi \alpha^2 + 2\alpha + 6)(2 + \alpha + \pi \alpha^2) + (1 - \theta)(2 + \alpha + \pi \alpha^2)(2\pi \alpha^2 + 6\alpha + 24) \]
\[ \text{Or,} \]
\[ -(1 - \theta)(\pi \alpha^2 + 2\alpha + 6)^2 \]
\[ \frac{(1 - \theta)((\alpha(\pi \alpha^2 + \alpha + 2)^2)}{\alpha(2 + \alpha + \pi \alpha^2)(1 - \theta)} > \frac{(\pi \alpha^2 + 2\alpha + 6)}{\alpha(2 + \alpha + \pi \alpha^2)(1 - \theta)} \]

Or,
\[ \alpha(\pi \alpha^2 + 2\alpha + 6)(2 + \alpha + \pi \alpha^2) + (1 - \theta)(2 + \alpha + \pi \alpha^2)(2\pi \alpha^2 + 6\alpha + 24) \]
\[ -(1 - \theta)(\pi \alpha^2 + 2\alpha + 6)^2 > (1 - \theta)^2 \alpha(\pi \alpha^2 + 2\alpha + 6)(2 + \alpha + \pi \alpha^2) \]

Or,
\[ -(2\theta - \theta^2)\alpha(\pi \alpha^2 + 2\alpha + 6)(2 + \alpha + \pi \alpha^2) + (1 - \theta)[(2 + \alpha + \pi \alpha^2)(2\pi \alpha^2 + 6\alpha + 24) - (\pi \alpha^2 + 2\alpha + 6)^2] > 0 \]
\[ \text{(11)} \]

Which is true because \( \alpha > 0, |\theta| < 1 \) and \( \pi = 22/7 \). Hence, GPMMD is an over-dispersed.

The third central moment of GPMMD can be obtained as follows
\[ \mu_3 = \mu'_3 - 3\mu'_2 \mu'_1 + 2(\mu'_1)^3 \]
The fourth central moment of GPMMD has been obtained as
\[ \mu_4 = \mu_1^3 - 4\mu_1\mu_2 + 6\mu_3(\mu_1^3) + 3(\mu_1^4) \]

(12)

(13)

**Methods of Estimation of the Parameter of GPMMD:**

Under this sub-section, the value of the parameter of this distribution has been estimated

(a) By using \( \mu_1 \) and \( p(Y = 0) \):

\[ p(Y = 0) = \frac{\alpha^3}{(2 + \alpha + \pi^2)} \left\{ \pi(1 + \alpha)^2 + (1 + \alpha) + 2 \right\} \]

(14)

Solving the expression (14), we get an estimated value of \( \alpha \) and putting it in

\[ \mu_1' = \frac{6(2 + \alpha + \pi^2)}{\alpha(2 + \alpha + \pi^2)(1 - \theta)} \]

We get an estimated value of \( \theta \).

(b) By using \( \mu_1' \) and \( \mu_2' \): From the mean

\[ (1 - \theta) = \frac{6(2 + \alpha + \pi^2)}{\alpha(2 + \alpha + \pi^2)\mu_1} \]

(15)

Putting the value of \( (1 - \theta) \) in the expression of \( \mu_1' \), we get
Solving this equation, we get an estimated value of $\alpha$. Putting the value of $\alpha$ in the expression (16), we get an estimated value of $\theta$.

- **Goodness of Fit and Applications of GPMMD:**

In order to check the validity of the theoretical work of this distribution, it seems necessary to describe the application and goodness of fit of this distribution with the support of secondary over-dispersed count data like number of errors per group, number of insects per leaf, number of death due to accident and etcetera used by other researchers. To test goodness of fit the following data are used.

Example (1): Distribution of mistakes in copying groups of random digits, Kemp and Kemp (1965).

<table>
<thead>
<tr>
<th>Number of errors per group</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed Frequency</td>
<td>35</td>
<td>11</td>
<td>8</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Example (2): Distribution of Pyrausta nabilalis in 1937, Beall (1940).

<table>
<thead>
<tr>
<th>Number of insects per leaf</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed Frequency</td>
<td>33</td>
<td>12</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Example (3): Distribution of mammalian cytogenic dosimetry lesions in rabbit lymphoblast included by Streptonigrin [NSC-45383], Exposure-70($\mu g/kg$).

<table>
<thead>
<tr>
<th>Class / Exposure ($\mu g/kg$)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed Frequency</td>
<td>200</td>
<td>57</td>
<td>30</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

**Table I: Observed Verses Expected Frequency of Example (1)**

<table>
<thead>
<tr>
<th>Number of errors per group</th>
<th>Observed frequency</th>
<th>Expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLD</td>
<td>PMD</td>
</tr>
<tr>
<td>Number of insects per leaf</td>
<td>Observed frequency</td>
<td>PLD</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------------</td>
<td>-----</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>1.0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>56.0</td>
</tr>
</tbody>
</table>

Table II: Observed Versus Expected Frequency of Example (2)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu'_1$</td>
<td>0.78333333</td>
</tr>
<tr>
<td>$\mu'_2$</td>
<td>1.8500</td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>1.7434</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>-</td>
</tr>
<tr>
<td>d.f.</td>
<td>2</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>1.78</td>
</tr>
<tr>
<td>P-value</td>
<td>0.61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu'_{12}$</td>
<td>1.8571</td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>1.8081</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>2.234</td>
</tr>
<tr>
<td>d.f.</td>
<td>2</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>0.53</td>
</tr>
<tr>
<td>P-value</td>
<td>0.625</td>
</tr>
</tbody>
</table>
The first example is related to number of errors per page (Kemp and Kemp, 1965). The second example is based on number of insects per leaf (Beall, 1940). The last example is related to class per exposer (Catcheside et al, 1946). Goodness of fit has been applied in all the examples mentioned. Theoretical frequencies due PLD and PMD have been tabulated to compare with the theoretical frequency of GPMMD. All these examples have already been mentioned in the Doctoral thesis (Sah, 2013).
Conclusion:
Comparison of GPMMD with PLD and PMD makes easy and simple with the use of following table.

Table- IV

<table>
<thead>
<tr>
<th>Table</th>
<th>PLD</th>
<th>PMD</th>
<th>GPMMD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>d.f.</td>
<td>P-Value</td>
<td>d.f.</td>
</tr>
<tr>
<td>I</td>
<td>2</td>
<td>1.78</td>
<td>0.61</td>
</tr>
<tr>
<td>II</td>
<td>2</td>
<td>0.53</td>
<td>0.83</td>
</tr>
<tr>
<td>III</td>
<td>3</td>
<td>3.91</td>
<td>0.43</td>
</tr>
</tbody>
</table>

In all these tables, the P-value of GPMMD is greater than that of PLD and PMD, so it is more useful for statistical modelling of such type of over dispersed data than PLD and PMD. It has also been observed that it gives a better fit than PMMD for example three.

CONFLICT OF INTEREST

The authors of this paper have written this paper selflessly with the aim of contributing only to continuous mixtures of generalised Poisson distribution. The authors do not intend to prejudice or offend anyone while writing this paper.

ACKNOWLEDGEMENT

Thank you to everyone who has given help or advice to improve quality of this paper.

REFERENCES


11. DOI: https://doi.org/10.22214/ijraset.2021.32839


17. DOI: 10.17605/OSF.IO/K5A7M