Study on Properties and Statistical Inference of Natural Discrete New Polynomial Exponential Distribution

Papu Mahanata
Roll No: M.Sc-Sem (IV) Stat-06

Department of Statistics
Visva-Bharati, Santiniketan

June 2025

Roadmap

1. Introductory Remarks
2. Properties of NDNPED
3. NDxgamma: A Special Case
4. Estimation and Simulation
5. Goodness of Fit
6. Concluding Remarks

Introductory Remarks

- **Objective**: Propose the Natural Discrete New Polynomial Exponential Distribution (NDNPED) for modeling overdispersed count data.
- **Motivation**: Limitations of traditional discrete distributions (e.g., Poisson) in handling high variance, right skewness, or zero-inflated data.
- **Applications**: Biology, reliability analysis, insurance, and ecological data.
- NDNPED is a mixture of geometric and negative binomial distributions.

NDNPED Probability Mass Function

The PMF of NDNPED is:

$$p(w;\theta) = \frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}} \binom{w+k}{w} (1-\theta)^{w} \theta^{k+1}}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}}, \quad w = 0, 1, 2, \ldots$$

- $0 < \theta < 1$, $a_{k,\theta}$ are non-negative constants.
- Mixture of $(r+1)$ negative binomial distributions.

Moment and Shape Properties

- **Probability Generating Function**:

$$P_X(s) = \frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{(1-\bar{\theta}s)^{k+1}}}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}}, \quad |s| < \frac{1}{1-\theta}$$

- **Factorial Moment**:

$$\mu'_{(l)} = h(\theta) \left(\frac{\bar{\theta}}{\theta}\right)^{l} \sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(l+k+1)}{\theta^{k+1}}$$

- **Index of Dispersion (ID)**:

$$ID(X) = \frac{\bar{\theta}}{\theta} \left[\frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+3)}{\theta^{k+1}}}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+2)}{\theta^{k+1}}} - \frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+2)}{\theta^{k+1}}}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}} \right] + 1$$

- NDNPED is overdispersed ($ID \geqslant 1$).

Reliability Properties

- **Reliability Function**:

$$S(t) = \frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}} I_\theta(t, k+1)}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}}$$

- **Failure Rate**:

$$\lambda(t) = \frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}} \binom{t+k}{t} (1-\theta)^t \theta^{k+1}}{\sum_{k=0}^{r} a_k \frac{\Gamma(k+1)}{\theta^{k+1}} I_\theta(t, k+1)}$$

- **Mean Residual Lifetime (MRL)**:

$$MRL(t) = \frac{\sum_{x=t+1}^{\infty} \sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}} I_\theta(x, k+1)}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}} I_\theta(t, k+1)}$$

NDxgamma Distribution

Putting the value $r = 2,\ a_{0,\theta} = 1,\ a_{1,\theta} = 0,\ a_{2,\theta} = \frac{\theta}{2}$, we get PMF for NDxgamma distribution

Definition

NDxgamma is a mixture of $Geo(\theta)$ and $NB(3, \theta)$ with mixing probabilities $\frac{\theta}{1+\theta}$ and $\frac{1}{1+\theta}$.

PMF:

$$p(x, \theta) = \frac{\theta^2}{1+\theta}(1-\theta)^x \left(1 + (x+2)(x+1)\frac{\theta}{2}\right); \quad x = 0, 1, 2, .., \infty, \theta \in (0, 1)$$

CDF:

$$F(x, \theta) = 1 - \frac{(1-\theta)^{x+1}(\theta(x+2)(\theta+\theta x+2)+2)}{2(\theta+1)}$$

The PMF of NDxgamma for θ=0.1, 0.3, 0.5, 0.9

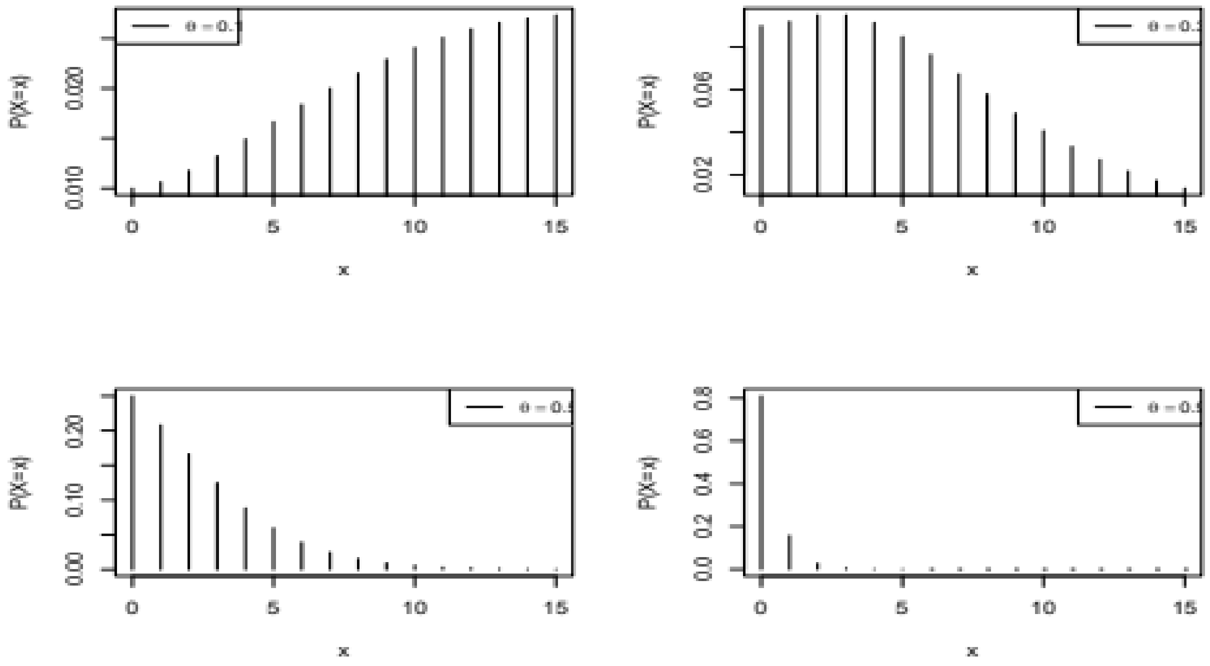

Figure: The PMF of NDxgamma for θ=0.1, 0.3, 0.5, 0.9

The CDF of NDxgamma for θ=0.1, 0.3, 0.5, 0.9

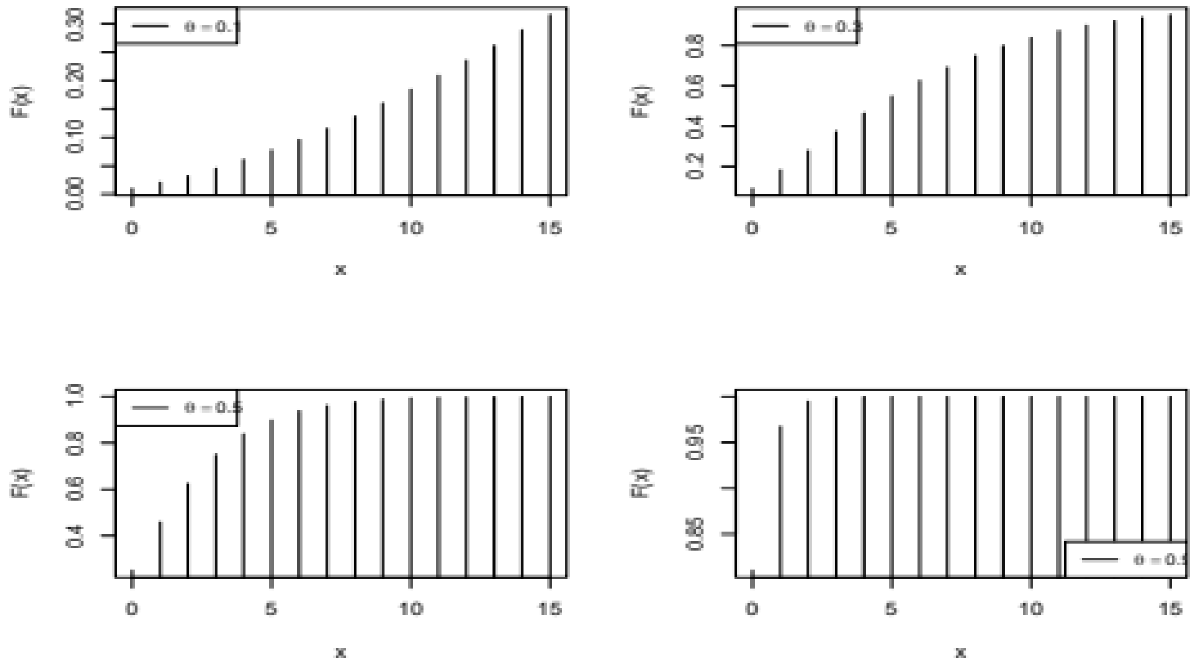

Figure: The CDF of NDxgamma for θ=0.1, 0.3, 0.5, 0.9

Mean, variance, skewness, and Kurtosis plots for different values of θ

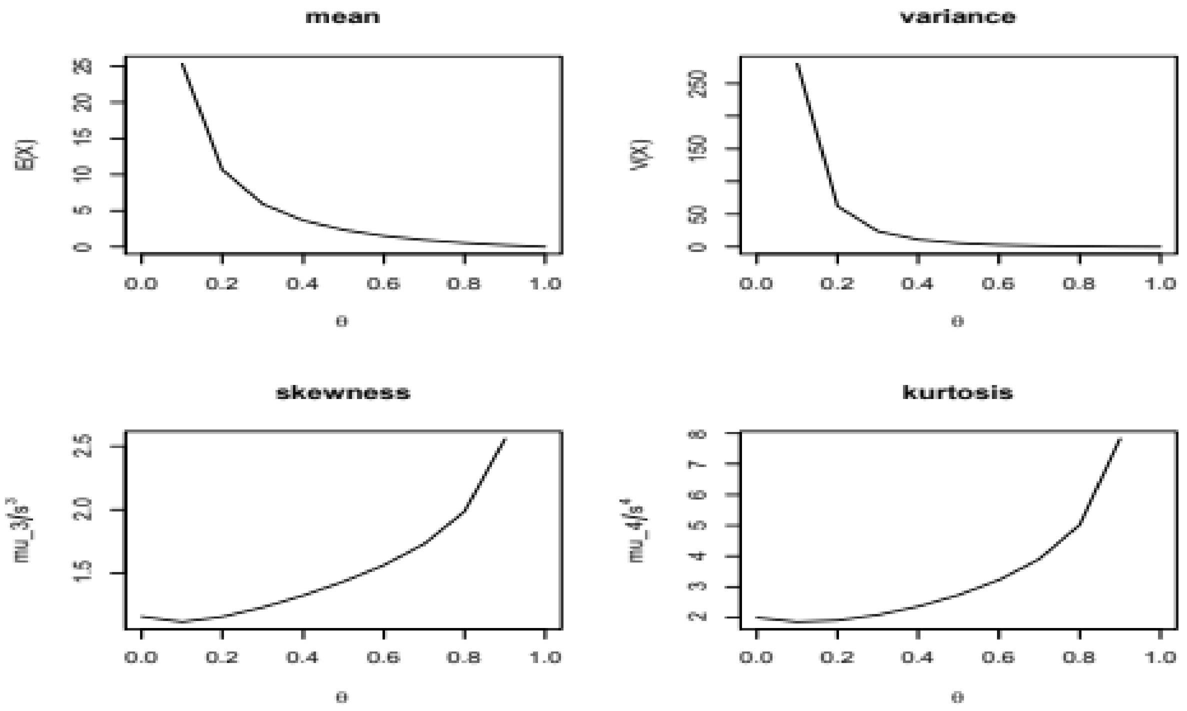

Figure: Mean, variance, skewness, and Kurtosis plots for different values of θ

NDxgamma Properties

Putting the value $r = 2,\ a_{0,\theta} = 1,\ a_{1,\theta} = 0,\ a_{2,\theta} = \frac{\theta}{2}$, we get m^{th} factorial moment for NDxgamma distribution

$$\begin{aligned} \mu'_{(m)} &= E\left(X(X-1)..(X-m+1)\right) \\ &= \left(\frac{1-\theta}{\theta}\right)^m \left(\frac{m!}{\theta+1}\right)\left(\theta + \frac{(m+1)(m+2)}{2}\right). \end{aligned} \tag{1}$$

Similarly, the m^{th} central moment of NDxgamma is given by

$$\mu_m = \frac{\theta^3(\Phi(1-\theta,-m-2,0) + 3\Phi(1-\theta,-m-1,0))}{2(\theta+1)} + \theta^2\Phi(1-\theta,-m,0), \tag{2}$$

where, $\Phi(z,s,a) = \sum_{k=0}^{\infty} \dfrac{z^k}{(a+k)^s}$ is the Lerchi transcendent.

Putting m=1 and m=2 the equations respectively, we get mean and variance as

- **Mean**:

$$\mu = \frac{-\theta^2 - 2\theta + 3}{\theta(\theta+1)}$$

- **Variance**:

$$\mu_2 = \frac{(\theta-3)(\theta-1)(3\theta+1)}{\theta^2(\theta+1)^2}$$

- **Index of Dispersion**:

$$ID = \frac{(\theta-3)(\theta-1)(3\theta+1)}{\theta(\theta+1)(3-\theta^2-2\theta)}$$

- Positively skewed and leptokurtic.
- Log-concave, implying Increasing Failure Rate (IFR) and Decreasing Mean Residual Lifetime (DMRL).

Parameter Estimation

- **Maximum Likelihood Estimation (MLE)**:

$$l(\theta) = n \log\left(\frac{\theta^2}{1+\theta}\right) + \sum_{i=1}^{n} \log(1-\theta) + \sum_{i=1}^{n} \left[1 + (1+x_i)(2+x_i)\frac{\theta}{2}\right]$$

Solve numerically:

$$\frac{dl}{d\theta} = \frac{(2+\theta)n}{\theta(1+\theta)} - \sum_{i=1}^{n} \log x_i + \sum_{i=1}^{n} \left[\frac{(1+x_i)(2+x_i)}{2+(1+x_i)(2+x_i)}\right] = 0$$

- **Method of Moments (MME)**:

$$\hat{\theta} = \frac{\sqrt{\bar{X}^2 + 16\bar{X} + 16} - \bar{X} - 2}{2(\bar{X}+1)}$$

Simulation Results

- Simulated NDxgamma with $\theta = 0.4$, 1,000 repetitions.
- Metrics:
  - Bias: $\frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)$
  - MSE: $\frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2$
- Results: Both Bias and MSE decrease as sample size n increases, confirming the precision of MLE.

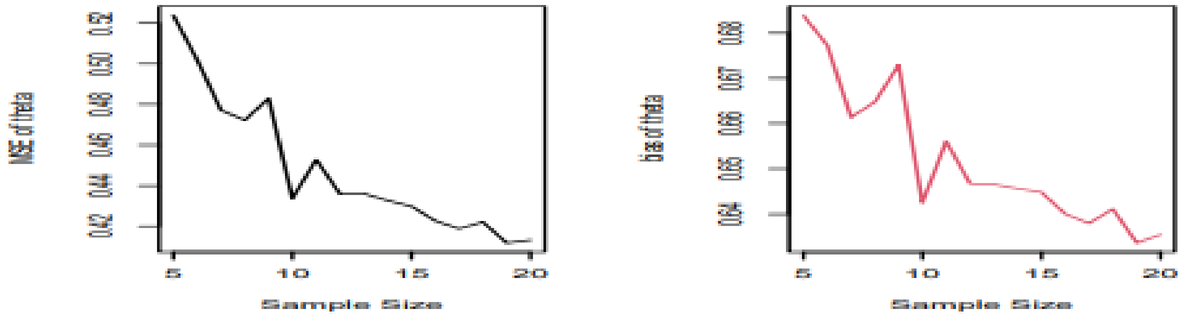

Figure: Bias and MSE for $\theta = 0.4$

Goodness of Fit: Real Datasets

- **Datasets**:
  1. Accidents of 647 women (Consul and Jain, 1973).
  2. European red mites on apple leaves (Garman, 1951).
  3. European corn borers (McGuire et al., 1957).
- Compared NDxgamma with Poisson, Zero-Inflated Poisson (ZIP), Dxgamma-I, and Dxgamma-II.
- Metrics: AIC, χ^2 statistic.

Real Datasets

X	Observed frequencies
0	447
1	132
2	42
3	21
4	3
≥ 5	2

Accident Data

X	Observed frequencies
0	70
1	38
2	17
3	10
4	9
5	3
6	2
7	1

Mite count Data

X	Observed frequencies
0	188
1	83
2	36
3	14
4	2
5	1

Corn Borer data

Goodness of Fit Results

Model Comparison for Accident Data

Model	AIC	Chi-square (χ^2)
Poisson	1236.372	110.31
ZIP	1190.544	6.42
Dxgamma-I	1201.249	23.81
Dxgamma-II	1201.757	24.50
NDxgamma	**1189.988**	**6.27**

Table: Model Comparison for Accident Data

Conclusion: The NDxgamma model provides the best fit with the lowest AIC and χ^2. It outperforms Poisson and ZIP, handling overdispersion effectively.

Goodness of Fit Results

Model Comparison for Mite Count Data

Model	AIC	Chi-square (χ^2)
Poisson	487.620	108.54
ZIP	456.866	14.44
Dxgamma-I	452.217	9.65
Dxgamma-II	452.235	9.60
NDxgamma	**448.009**	**4.82**

Table: Model Comparison for Mite Count Data

Conclusion: NDxgamma is most suitable for this overdispersed biological data, outperforming Poisson, ZIP, and Dxgamma models.

Goodness of Fit Results

Model Comparison for Corn Borer Data

Model	AIC	Chi-square (χ^2)
Poisson	726.490	18.66
ZIP	714.079	1.24
Dxgamma-I	713.928	2.70
Dxgamma-II	713.976	4.27
NDxgamma	**713.888**	**1.20**

Table: Model Comparison for Corn Borer Data

Conclusion: NDxgamma again provides the best fit, slightly outperforming ZIP and handling zero-inflation with overdispersion effectively.

Concluding Remarks

- NDNPED is a flexible discrete distribution for overdispersed count data.
- Key features:
  - Closed-form expressions for moments, reliability, and entropy.
  - Superior fit to biological and ecological datasets.
  - Efficient parameter estimation via MLE and MME.
- NDxgamma, a special case, balances simplicity and modeling power.
- Future work: Explore additional applications in insurance and reliability engineering.

Thank You!