**Revista Colombiana de Estadística - Theoretical Statistics**
January 2025, volume 48, issue 1, pp. 1–21
http://doi.org/10.15446/rce.v48n1.113025

# A Flexible Discrete Probability Model for Partly Cloudy Days

## Un modelo de probabilidad discreto flexible para días parcialmente nublados

Tassaddaq Hussain[1,a], Hassan S. Bakouch[2,3,b], Zahid Ur Rehman[4,c], Mohammad Shakil[5,d], Qingsong Shan[4,e], Qianning Liu[4,f]

[1]Department of Statistics, Faculty of Natural and Applied Science, Mirpur University of Science and Technology (MUST), Mirpur, Pakistan

[2]Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt

[3]Department of Mathematics, College of Science, Qassim University, Buraydah, Saudi Arabia

[4]Department of Statistics, Faculty of Natural Sciences, Jiangxi University of Finance and Economics, Nanchang, China

[5]Department of Liberal Arts and Sciences (Mathematics/Statistics), Miami Dade College city,

### Abstract

In this article, a discrete-valued probability model is proposed, its mathematical properties and formulation are studied under the nabla structure, which include discrete Laplace transformation, moments, recurrence relation between moments, index of dispersion, and asymptotic distribution of extremes. Furthermore, application of model with rereference to the partly cloudy days is discussed. Moreover, the model compatibility is checked by chi-square, Anderson–Darling, Cramér–von Mises, information criterion and Vuong statistics, and it is found that the proposed model is the best strategy for such data analysis.

***Key words*:** Statistical model; Failure Rate; Recurrence relation of moments; Estimation.

[a]Ph.D. E-mail: tafkho2000@gmail.com
[b]Ph.D. E-mail: hassan.bakouch@science.tanta.edu.eg
[c]Ph.D. E-mail: zahid.ms@must.edu.pk
[d]Ph.D. E-mail: mshakil@mdc.edu
[e]Ph.D. E-mail: mathhuman@outlook.com
[f]Ph.D. E-mail: mathhuman@outlook.com

**Resumen**

En este artículo se propone un modelo de probabilidad en tiempo discreto, se estudian sus propiedades matemáticas y su formulación bajo la estructura nabla que incluye transformación discreta de Laplace, momentos, relación de recurrencia entre momentos, índice de dispersión y distribución asintótica de extremos. Además, se discute la aplicación del modelo con referencia a los días parcialmente nublados. Además, la compatibilidad del modelo se verifica mediante chi-cuadrado, Anderson-Darling, Cramér-von Mises, criterio de información y estadísticas de Vuong y se encontró que el modelo propuesto es la mejor estrategia para dicho análisis de datos.

***Palabras clave*:** Modelo estadístico; Tasa de fallas; Relación de recurrencia de momentos; Estimación.

# 1. Introduction

Modeling of the discrete stochastic processes like, the number of fog days per month, the number of rain droplets per mm, the number of precipitation days, the number of times cloud collisions occur to produce rain, the number partly cloudy days produce rain, the number of space mission carried out in cloudy days, the number of cloudy hours necessary to decrease the temperature, the number of days when the temperature is less than 32°, the number of partially cloudy days per month, the number of clear days per months and the number of thunder storms per year, etc., often occur in real life scenario see (Carter, 1972; Lloyd-Smith, 2007; Banik & Kibria, 2009). For determining the likelihood of above phenomenon, many discrete probability models are often used, which include Poisson, binomial, negative binomial, zero-truncated and zero-inflated models.

However, such models often exhibit three dispersion pattern types: under-, over- and equi-dispersions, which usually affect the goodness-of-fit statistics if these are not adjusted, which may result in un-realistic conclusion. In these dispersion patterns, over-dispersion is a crucial notion in the analysis of count data. Many times, data acknowledge more unpredictability than anticipated under the supposed distribution. Nevertheless, the likelihood for over-dispersion occurs because the generally used distributions identify certain bonds between the variance and the mean. Also, in the modern complex era the discrete processes often display an over-dispersion. In this regard, the negative binomial, generalized Poisson and Poisson-Lindley models may be utilized for modeling the over- and equi-dispersion as well as under-dispersion patterns as these handle the over-dispersion with an extra parameter to exhibit the variance of a variable but factual unable to handle the under-dispersion cases as efficiently as observed in over-dispersion cases. Moreover, the under-dispersion can communicate to forms of manifestation which are much more consistent than the arbitrariness linked as seen in the Poisson process. It is also theoretically possible but rare in practice. However, the Poisson models usually estimate exceptional-event processes (e.g., accident incidents, failures in engineering or processing, etc.). These processes usually depend on equi-dispersion restrictions that render it useless as a real-life pattern of dispersion resulting in the incapability of fitting the Poisson model, see Nory et al. (2022).

In view of the above discussion and on the motivational aspect, we decided to model the number of cloudy days pattern by a discretized model, which will be quite effective in under-dispersion conditions. The proposed model is a discretized version of generalized Lindley distribution, which is obtained via discrete fractional calculus discussed by Ganji & Gharari (2018*b*). The model possesses the same hazard function behavior as that of the continuous version and is quite helpful in all dispersion patterns quite effectively. Secondly, the proposed strategy is based on the nabla structure, which has so far not been search out for modeling the environmental indicators. Thirdly, the model is a simple structure and have the ability to exhibit a bathtub, increasing and decreasing failure rate behavior. Fourthly, it is not only possessing a unique mode but also bimodal in nature and genrates a family of discrete distributions like discrete chi-square, discrete Lindley, two-parameter truncated Lindley and discrete gamma distributions. Fifthly it is a mixture model and have ability to model over-dispersed data sets too.

The rest of the article is systematized as listed. Discretization procedures, along with the derivation of the proposed model, are described in Section 2. In Section 3, we shall deal with the mathematical chattels of the purported model, which include probability generating function, moments, index of dispersion, skewness, kurtosis and distribution of extreme order statistics. However, Section 4 is reserved for parameter estimation. Model compatibility and application as well as evaluation statistics along with four practical data application on partly cloudy days are studied in Section 5. Finally, conclusion is drawn in Section 6.

# 2. Discretiztion Procedures and Proposed Distribution

An essential and dynamic field of mathematical research is the count data modeling via the discretization of continuous probability models. In this regard, researchers of the current era have innovated and proposed a number of discretized distributions, among them being Weibull, Lindley, gamma, normal, Rayleigh, inverse Weibull, inverse Rayleigh. For this purpose, they have adopted various methodologies. In this regard Chakraborty (2015), Chakraborty (2015) and Ganji & Gharari (2018*a*) have presented the discretizing procedures in an effective and pleasant manner. These procedures are listed below.

## 2.1. Procedure-1

In this procedure, researchers usually sustain the probabilities of a continuous random variable $X$ at integer points only and define the probability mass function (PMF) of discretized random variable $Y$, stated by Chakraborty (2015) and defined as

$$\mathcal{P}_m = \frac{f(x)}{\sum_{m=-\infty}^{\infty} f(m)}, \quad m = 0, \pm 1, \pm 2, \ldots$$

### 2.2. Procedure-2

In this procedure, researchers usually try to preserve the survival function (SF) of a continuous random variable $X$, i.e., $\mathcal{S}_X(x)$ and define a discretized random variable $Y$ as $Y = (X) =$ the greatest integer less than or equal to $X$ with corresponding PMF as given by Chakraborty (2015) and Roy (2004),

$$\begin{aligned}\mathcal{P}_m &= \mathbb{P}(m \leq X < m+1) = \mathcal{F}_X(m+1) - \mathcal{F}_X(m) \\ &= \mathcal{S}_X(m) - \mathcal{S}_X(m+1). \quad m = 0, 1, 2, 3, \ldots\end{aligned}$$

### 2.3. Procedure-3

This procedure sustains the hazard function (HRF) of the continuous variable $X$. However, the SF of the discretized random variable $Y$ is given by

$$\mathbb{P}(Y \geq m) = (1 - \mathcal{H}_X(1))(1 - \mathcal{H}_X(2)) \cdots (1 - \mathcal{H}_X(m-1)), \; m = 1, 2, \ldots, k.$$

By Chakraborty (2015), the PMF is given as

$$\begin{aligned}\mathcal{P}_m &= (1 - \mathcal{H}_X(1))(1 - \mathcal{H}_X(2)) \cdots (1 - \mathcal{H}_X(m-1)) \left(1 - (1 - \mathcal{H}_X(k))\right), \\ &= (1 - \mathcal{H}_X(1))(1 - \mathcal{H}_X(2)) \cdots (1 - \mathcal{H}_X(m-1))\mathcal{H}_X(m),\end{aligned}$$

$$\mathcal{P}_m = \left\{ \begin{array}{l} \mathcal{H}_X(0), \quad m = 0 \\ (1 - \mathcal{H}_X(1))(1 - \mathcal{H}_X(2)) \cdots (1 - \mathcal{H}_X(m-1))\mathcal{H}_X(m), \; m = 1, 2, 3 \ldots, k \\ 0, \;\; \text{elsewhere} \end{array} \right\}$$

However, the value of $k$ is governed so as to characterize the HRF: $0 \leq \mathcal{H}_X(x) < 1$ and its PMF $\mathcal{P}_m$, see Chakraborty (2015).

### 2.4. Procedure-4

This procedure was developed by Hagmark (2008) in which they discretized the continuous stochastic variate $X$ with CDF $\mathcal{F}_X(x)$ and defined its count counterpart $Y$ with CDF as, see Chakraborty (2015),

$$\mathcal{F}_Y(y) = \mathbb{P}(Y \leq y) = \int_y^{y+1} \mathcal{F}_X(y) dx.$$

### 2.5. Procedure-5

Bohner & Peterson (2001, 2003) and Ganji & Gharari (2018$a$) have discretized a continuous variable with the help of the monomial Taylor and exponential function by using discrete fractional calculus.

## 2.6. Discretized Lindley Family of Distributions

In view of the above definitions, we propose a flexible family of nabla discrete Lindley distributions by considering the procedure-5 as stated by Bohner & Peterson (2001, 2003), Ganji & Gharari (2018*a*,*b*), Gharari & Ganji (2021) and Gharari et al. (2023). Zakerzadeh & Dolati (2009) defined the probability density function (PDF) of the generalized Lindley (GL) distribution as

$$f(x;\theta,\alpha,\beta)=\frac{(\alpha+\beta x)e^{-\theta x}(\theta x)^{\alpha-1}\theta^2}{\Gamma(\alpha+1)(\theta+\beta)}.x>0,\theta>0,\beta>0\,\alpha>0. \tag{1}$$

On the basis of (1) we shall define a generalized nabla discrete Lindley (GNDL) distribution as follows.

**Proposition 1.** *A random variable $Y$ has a GNDL distribution with $p,\alpha$ and $\beta$ parameters if its PMF is given by*

$$\mathcal{P}_x=\frac{(1-p)^{\alpha+1}p^{x-1}(\alpha+(\alpha+x-1)\beta)(x)_{\alpha-1}}{(1-p+\beta)\Gamma(\alpha+1)}, \tag{2}$$

*for* $x=1,2,3,\ldots,0<p=1-\theta<1,\alpha>0,\beta\geq 0.$

***Proof.*** Let $Y$ be a discrete random used to represent the number of partly cloudy days at certain locality. Then PMF of discretized version of Equation (1) can be obtained as

$$\mathcal{P}_x=\mathbb{P}(Y=x)=\frac{\theta^\alpha\theta}{\theta+\beta}(\frac{\mathfrak{h}_{\overline{\alpha-1}}(x)}{e_\theta(\rho(x),0)})+\frac{\beta}{\theta+\beta}(\frac{\mathfrak{h}_{\overline{\alpha}}(x)\theta^{\alpha+1}}{e_\theta(\rho(x),0)}). \tag{3}$$

Based on Procedure-5 mentioned above and by incorporating $\rho(x)=x-1$, $e_\theta(\rho(x),0)=(1-\theta)^{-(x-1)}$ and $\mathfrak{h}_{\overline{\alpha-1}}(x)=\frac{x^{\overline{\alpha-1}}}{\Gamma(\alpha)}$ in Equation (3) we get

$$\mathcal{P}_x=\frac{\alpha\theta^{\alpha+1}}{\theta+\beta}(\frac{x^{\overline{\alpha-1}}(1-\theta)^{x-1}}{\Gamma(\alpha+1)})+\frac{\beta}{\theta+\beta}(\frac{x^{\overline{\alpha+1-1}}\theta^{\alpha+1}(1-\theta)^{x-1}}{\Gamma(\alpha+1)}),$$

where $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$ and $x^{\overline{\alpha}}=x(x+1)(x+2)...(\alpha+x-1)$. Hence, we have the expression

$$\mathcal{P}_x=\frac{\theta^{\alpha+1}}{(\theta+\beta)\Gamma(\alpha+1)}(\alpha+(\alpha+x-1)\beta)(1-\theta)^{x-1}(x)^{\overline{\alpha-1}},x=1,2,3,\ldots$$

On incorporating $1-\theta=p$ in the expression above, we get

$$\mathcal{P}_x=\frac{(1-p)^{\alpha+1}p^{x-1}(\alpha+(\alpha+x-1)\beta)(x)^{\overline{\alpha-1}}}{(1-p+\beta)\Gamma(\alpha+1)},$$

which after simplification yields Equation (2). Now on using the property of nabla ascending factorial moment we can exhibit that $\sum_{x=1}^{\infty}\mathcal{P}_x=1$. □

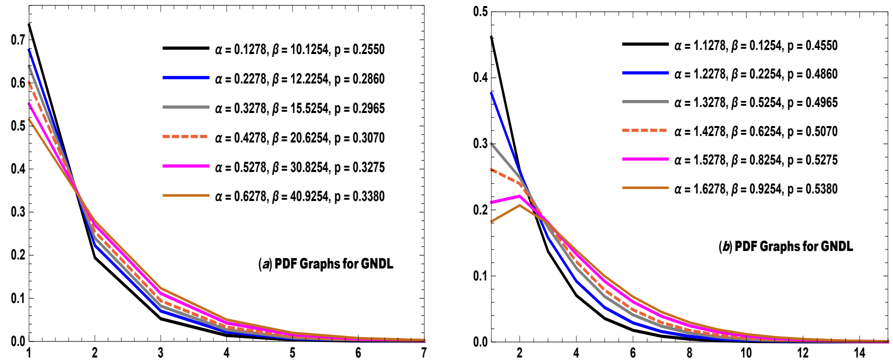


Figure 1: PDF plots of GNDL.

Table 1: A set of submodels of the GNDL.

| Distribution | Parameters' values | Refrence |
|---|---|---|
| Discrete Gamma ($\nabla^{DG}_{\alpha,\theta}$) | $x \in \mathbb{N}$, $\beta = 0$ and $\theta = 1 - p$ | see Ganji and Gharari (2018). |
| Discrete Lindley ($\nabla^{DL}_{\theta}$) | $x \in \mathbb{N}$, $\beta = 1, \alpha = 1$ and $1 - p = \theta$ | see Bakouch et. al., (2022) |
| Geometeric ($\nabla^{DE}_{\theta}$) | $x \in \mathbb{N}$, $\alpha = 1, \beta = 0$ and $\theta = 1 - p$ | see Ganji and Gharari (2018) |
| Discrete Chi square ($\nabla^{DC}_{\frac{\alpha}{2},\frac{1}{2}}$) | $x \in \mathbb{N}$, $\alpha = \frac{\alpha}{2}, p = \frac{1}{2}, \beta = 0$ and $1 - \theta = p$ | see Ganji and Gharari (2018) |
| Truncated Lindley ($\nabla^{DG}_{p,\beta}$) | $x \in \mathbb{N}$, $\alpha = 1$ and $\theta = 1 - p$ | see Kiani (2020) |

The GNDL contains the submodels given in Table 1 with natural numbers support.

The proposed PMF can adopt various shapes ranging from reverse J, and symmetric curve. Usually, for $\alpha < 1$ and $p \to 0$ and for all $\beta$ it generates a reverse J shape and for $\alpha > 1$ and $p \to 1$ it changes into positively skewed one. However, for large $\alpha$, i.e., $\alpha \to \infty$, it exhibits a symmetrical curve behavior, but the peakedness from decreases leptokurtic, mesokurtic to platykurtic as both $\alpha$ and $p$ goes on increasing, see Figure 1.

Further, GNDL is a weighted version of $\nabla^{\mathrm{DG}}_{\alpha,\theta}$ with weight $\frac{\theta}{\beta+\theta}$ and $\nabla^{\mathrm{DG}}_{\alpha+1,\theta}$ with weight $\frac{\beta}{\beta+\theta}$, respectively. Moreover, it has the probability recurrence relation

$$x\mathcal{P}_{x+1} = ((\beta+1)\alpha + (x-1)\beta) = p(x+\alpha-1)\mathcal{P}_x(\beta x + (\beta+1)\alpha)).$$

The failure (hazard) rate function (HRF) describes the immediate future failure in such a way that unit has not failed at time $x$, and for the GNDL it is defined as

$$\mathcal{H}(x) = \frac{(1-p)(\alpha + (\alpha + x - 1)\beta))}{((x-1)\beta + \alpha(\beta + 1 - p)_2F_1(1, -1 + x + \alpha; x; p))}. \tag{4}$$

From Equation (4) it is evident that $\mathcal{H}(x) = \frac{(\alpha+(\alpha+x-1)\beta))}{(x\beta-\beta+\alpha(1+\beta))}$ as $p \to 0$; $\mathcal{H}(x) = \frac{(1+x)(1-p)^2}{x-p+1}$ as $\beta \to 1, \alpha \to 1$ and $\mathcal{H}(x) = \frac{1}{{}_2F_1(1,-1+x+\alpha;x;p)}$ as $\beta \to 0$. However, $\mathcal{H}(x) = 1-p$ as $x \to \infty$, hence, the $\mathcal{H}(x)$ is constrained above. From Figure 2, it is obvious that $\mathcal{H}(x)$ can adopt bathtub when $\alpha < 1$, $\beta > 0$ and for all $p$, increasing for $\alpha > 1$ for all $\beta$ and $p$, and decreasing failure rate shapes for $\alpha \le 1, \beta \to 0$, respectively.

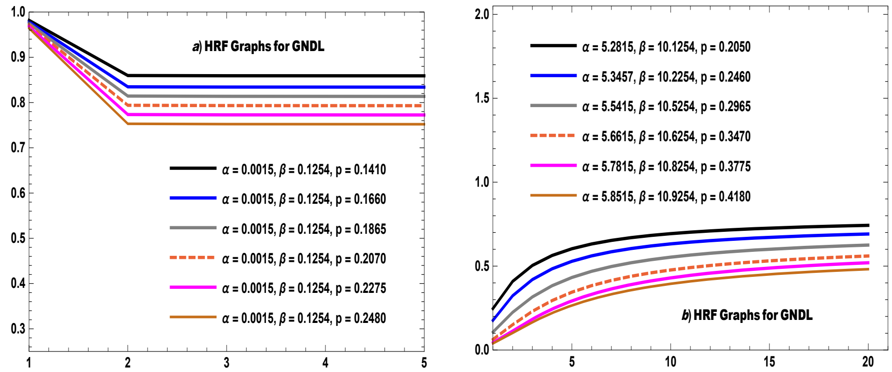


FIGURE 2: HRF plots of GNDL.

## 3. Mathematical Properties of the GNDL Distribution

The Laplace transform, has been viewed as a compelling device for explaining both ordinary and partial differential equations. Equivalent to the Laplace transform, the discrete Laplace transform (DLT) is used in signal processing, and in the theory of analytic functions. Moreover, Z-transform is utilized to solve linear systems of difference equation, see Ameen et al. (2019) and yields the generating function of the values of any function at non-negative integers. Such generating functions are the only source to recognize the PMF of any discrete distribution.

**Definition 1.** The $\mathfrak{L}_d$-transform of a sequence $\{x\}_{k=1}^{\infty}$ is function $\mathfrak{L}(s)$ of complex variable defined by

$$\mathfrak{L}_d(s) = \mathfrak{L}\{a_k\}(s) = \sum_{k=1}^{\infty} \frac{a_k}{(1+s)^{k+1}}, \tag{5}$$

for all values of $s$ for which the series converges.

The coming proposition provides the DLT of the GNDL.

**Proposition 2.** *The DLT ($\mathfrak{L}_d(s)$) of the GNDL is expressed as*

$$\mathfrak{L}_{X_d}(s) = \frac{(1-p)^{\alpha+1}(1+\beta-\frac{p}{1+s})}{(1+s)(1-\frac{p}{1+s})^{\alpha+1}(1-p+\beta)}. \tag{6}$$

***Proof*.** By definition, $\mathfrak{L}_{X_d}(s)$ can be expressed as

$$\mathfrak{L}_{X_d}(s) = \mathfrak{L}_d\{\mathcal{P}_x\}(s) = \sum_{x=1}^{\infty} \left(\frac{1}{1+s}\right)^x \mathcal{P}_x, \tag{7}$$

by incorporating Equation (2) in the above expression we get

$$\mathfrak{L}_{X_d}(s) = \frac{(1-p)^{\alpha+1}}{p(1-p+\beta)\Gamma(\alpha+1)} \sum_{k=1}^{\infty} k^{\overline{\alpha-1}}(\alpha(1+\beta)+(k-1)\beta)(\frac{p}{1+s})^k,$$

hence

$$\mathfrak{L}_{X_d}(s) = \frac{(1-p)^{\alpha+1}}{p(1-p+\beta)\Gamma(\alpha+1)}\{(\alpha-\beta+\alpha\beta)\sum_{k=1}^{\infty} k^{\overline{\alpha-1}}(\frac{p}{1+s})^k + \beta\sum_{k=1}^{\infty} kk^{\overline{\alpha-1}}(\frac{p}{1+s})^k\}. \tag{8}$$

The DLT of $\mathfrak{L}_d\{p^k k^{\overline{\alpha-1}}\}(s)$ is

$$\sum_{k=1}^{\infty} k^{\overline{\alpha-1}}(\frac{p}{1+s})^k = \frac{p\Gamma(\alpha)}{(1+s)(1+s-p)^{\alpha}} \tag{9}$$

and $\mathfrak{L}_d\{kp^k k^{\overline{\alpha-1}}\}(s)$ is

$$\sum_{k=1}^{\infty} kk^{\overline{\alpha-1}}(\frac{p}{1+s})^k = \frac{p(1+\frac{p(-1+\alpha)}{1+s})\Gamma(\alpha)}{(1-\frac{p}{1+s})^{\alpha+1}}. \tag{10}$$

Incorporating Equation (9) and (10) into Equation (8) we get

$$\mathfrak{L}_{Xd}(s) = \frac{(1-p)^{\alpha+1}(p-(1+s)(1+\beta))}{(1+s)^2(1-p/(1+s))^{\alpha+1}(1-p+\beta)},$$

which after simplification yields (6). □

**Corollary 1.** *On putting* $\frac{1}{1+s} = e^t$ *in Equation* (6)*, the moment generating function (MGF) of the GNDL is*

$$\mathcal{M}_X(t) = \frac{\exp(t)(1-p)^{\alpha+1}(1+\beta-p\exp(t))}{(1-p\exp(t))^{\alpha+1}(1-p+\beta)}. \tag{11}$$

**Corollary 2.** *The MGF of* $X$ *is the convolution of three independent random variables, namely* $X_{Degenrate}(1), X_{NB}(p, \alpha+1)$ *and* $X_{Bernoulli}(\frac{p}{1+\beta-p})$*, i.e.,* $X = X_1 + X_2 + X_3$

$$\mathcal{M}_Y(t) = \mathcal{M}_{X_1}(t)\mathcal{M}_{X_2}(t)\mathcal{M}_{X_3}(t). \tag{12}$$

*where* $\mathcal{M}_{X_1}(t) = e^t$, $\mathcal{M}_{X_2}(t) = \frac{(1-p)^{\alpha+1}}{(1-pe^t)^{\alpha+1}}$ *and* $\mathcal{M}_{X_3}(t) = \frac{1-pe^t+\beta}{1-p+\beta}$.

**Corollary 3.** *The probability generating function (PGF) of GNDL can be obtained by substituting* $\frac{1}{1+s} = t$ *in Equation* (6) *as*

$$\mathcal{G}_X(t) = \frac{t(1-p)^{\alpha+1}(1+\beta-pt)}{(1-pt)^{\alpha+1}(1-p+\beta)}. \tag{13}$$

*This completes the proof.*

On differentiating Equation (12) $i$th respect to (w.r.t) $t$, then we get the raw moment by the expression $\mu'_r = \frac{d^r\mathcal{M}_Y(t)}{dt^r}|_{t=0}$. In this regard, we have computed its mean, variance, index of dispersion ($\mathbb{ID}$), skewness and kurtosis, which are given below.

$$\mu'_1 = 1 + p(\frac{1+\alpha}{1-p} - \frac{1}{1-p+\beta}),$$

$$\mu_2' = \frac{(1-p+\beta)(1+p+3p\alpha+(p\alpha)^2)-(1-p)p(3+p(2\alpha-1))}{(1-p)^2(1-p+\beta)}.$$

Now the variance of $Y$ is defined as $\mathbb{V}ar(Y) = \mu_2' - \mu_1'^2$ and for the GNDL it is expressed as

$$\mathbb{V}ar(Y) \quad = \quad \frac{p\alpha}{(1-p)^2} + \frac{p\beta(1-p^2+\beta)}{(1-p+\beta)^2(1-p)^2}.$$

After defining the mean and variance, we define another characteristic that helps us to identify the dispersion known as dispersion index, abbreviated as $\mathbb{ID}$ and defined as $\frac{\text{Variance}}{\text{Mean}}$. It may assume three different values i) over-dispersion, i.e. ($\mathbb{ID} > 1$), ii) under-dispersion ($\mathbb{ID} < 1$) and iii) equi-dispersion ($\mathbb{ID} = 1$). $\mathbb{ID}$ behavior of the GNDL can be described as; for $p \leq 0.5$ $\forall\beta$ and $\alpha$ the distribution is under-dispersion and $p > 0.5$ or $p \to 1$ $\forall\beta$ and $\alpha$ the distribution portrays, an over-dispersion characteristic, $0 < p < 0.2$ and $\forall, \beta$ and $\alpha \to \infty$ the GNDL exhibits an equi-dispersion behavior while as $p \to 1$ and the distribution becomes over-dispersed.

$$\mathbb{ID} \quad = \quad \frac{p(\alpha(\beta-p+1)^2+\beta(\beta-p^2+1))}{(-p+1)(\beta-p+1)(1+\beta+p(-2+p+\alpha(\beta-p+1)))},$$

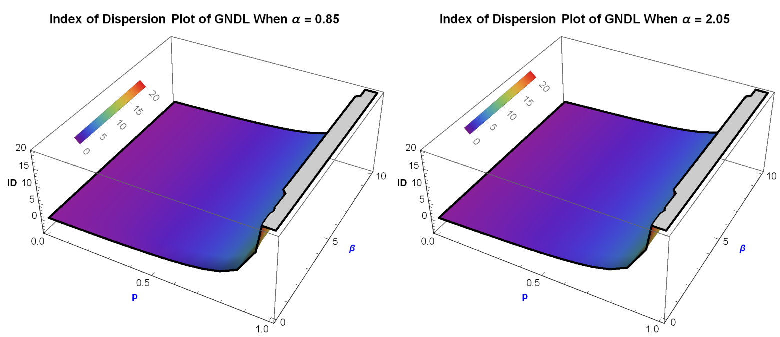


FIGURE 3: ID plots of GNDL.

Similarly, the skewness ($\frac{\mu_3^2}{\mu_2^3}$) and the kurtosis ($\frac{\mu_4}{\mu_2^2}$) help to determine shape and peakedness of the curve. In this regard, the GNDL behaved as positively skewed and leptokurtic in nature.

**Proposition 3.** *Let $r^{th}$ and $(r+1)^{th}$ are $r$ and $r+1$ order moments of the GNDL about the origin, then a recursive relation between them can be established as*

$$\mu_{r+1}' = p\frac{d\mu_r'}{dp} + \frac{((1-p)p-(1+p\alpha)+(1-p+\beta))\mu_r'}{(1-p)(1-p+\beta)},$$

*where* $0 < p < 1$, $\beta \geq 0$ *and* $\alpha > 0$.

***Proof*.** The $r^{th}$ moment about origin can be stated as

$$\mu_r' = \frac{(1-p)^{\alpha+1}}{(1-p+\beta)\Gamma(\alpha+1)} \sum_{x=1}^{\infty} x^r p^{x-1} (x)_{\alpha-1} (\alpha + (\alpha+x-1)\beta), \tag{14}$$

the derivative of Equation (14) w.r.t $p$ yields

$$\begin{aligned} \frac{d\mu_r'}{dp} &= \frac{-(\alpha+1)(-p+1)^{\alpha}}{(1-p+\beta)\Gamma(\alpha+1)} \sum_{x=1}^{\infty} x^r p^{x-1} (x)_{\alpha-1} (\alpha + (\alpha+x-1)\beta) \\ &+ \frac{(-p+1)^{\alpha+1}}{(1-p+\beta)\Gamma(\alpha+1)} \sum_{x=1}^{\infty} (x-1) x^r p^{x-2} (x)_{\alpha-1} (\alpha + (\alpha+x-1)\beta) \\ &+ \frac{(1-p)^{\alpha+1}}{(1-p+\beta)^2\Gamma(\alpha+1)} \sum_{x=1}^{\infty} x^r p^{x-1} (x)_{\alpha-1} (\alpha + (\alpha+x-1)\beta), \end{aligned}$$

which after simplification yields

$$\frac{d\mu_r'}{dp} = \frac{\mu_{r+1}'}{p} - \frac{((1-p)p - (1+p\alpha) + (1-p+\beta))\mu_r'}{p(1-p)(1-p+\beta)},$$

$$r = 0, 1, 2, 3, \ldots, \mu_0' = 1, 0 < p < 1, \alpha > 0 \text{ and } \beta \geq 0.$$

□

Likewise, the $r^{th}$ moment about mean can be stated as

$$\mu_r = \frac{(1-p)^{\alpha+1}}{(1-p+\beta)\Gamma(\alpha+1)} \sum_{x=1}^{\infty} (x-\mu_1')^r p^{x-1} (x)_{\alpha-1} (\alpha + (\alpha+x-1)\beta),$$

where $\mu_1' = 1 + p(\frac{1+\alpha}{1-p} - \frac{1}{1-p+\beta})$. On differentiating the above equation w.r.t $p$ we get the next corollary.

**Corollary 4.**

$$\begin{aligned} \mu_{r+1} + p\mu_1'\mu_r &= p\frac{d\mu_r}{dp} + p\left(\frac{1}{p} + \frac{1+\alpha}{1-p} - \frac{1}{\beta+1-p}\right)\mu_r \\ &- \left(\frac{rp}{(\beta+1-p)^2} + \frac{r}{\beta+1-p} - \frac{r(\alpha+1)}{(1-p)^2}\right)p\mu_{r-1}. \end{aligned}$$

$$r = 1, 2, 3, \ldots, \mu_0' = 1, \mu_1 = 0, 0 < p < 1, \alpha > 0 \textit{ and } \beta \geq 0.$$

## 3.1. Asymptotic Distribution of Extreme Values

Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample of size $n$, drawn identically and independently from the GNDL with CDF $\mathcal{F}(x) = 1 - \mathcal{SF}(x)$. Let $Y_{1:n} < Y_{2:n} < \ldots < Y_{n:n}$ are order statistics of size $n$. Suppose $\mathcal{M}_n = y_{n:n}$ and $\mathcal{N}_n = y_{1:n}$ that are the greatest and smallest reflection, correspondingly, the aim is discussing the asymptotic distributions of these extremes. Now on utilizing the Theorem (1.6.2) see Leadbetter et al., (1987), the distribution of maxima, after using L'Hôpital's rule, is

$$\lim_{s\to\infty} \frac{1-\mathfrak{F}(s+xg(s))}{1-\mathfrak{F}(s)} = \lim_{s\to\infty} \frac{\mathbb{P}(Y=x+s)}{\mathbb{P}(Y=s)},$$

$$\lim_{s\to\infty} \frac{(x+s)_{\alpha-1}(\alpha+(\alpha+x+s-1)\beta)p^{x+s-1}}{s_{\alpha-1}(\alpha+(\alpha+s-1)\beta)p^{s-1}} = p^x,$$

and its standardized form is Gumbel type distribution, i.e. $p^x$. So, according to Theorem (1.6.3) of Leadbetter et al. (1987), the norming constants $q_n > 0, s_n > 0, u_n > 0$ and $v_n > 0$, we obtain

$$\mathbb{P}\left\{q_n(\mathcal{M}_n - s_n) \le x\right\} \to e^{-p^x},$$

For the minima of distribution, we apply the Theorem (1.6.2), i.e.

$$\lim_{s\to 0} \frac{1-\mathfrak{F}(x_F - sx)}{1-\mathfrak{F}(x_F - s)} = \lim_{s\to 0} \frac{x\mathbb{P}(Y=sx)}{\mathbb{P}(Y=s)} = x,$$

where $\lim_{x_F\to\infty} \mathfrak{F}(x_F) = 1$, hence

$$\mathbb{P}\left\{u_n(\mathcal{N}_n - v_n) \le x\right\} \to e^{-x},$$

as $n \to \infty$. Thus, the sample maxima and sample minimum domain of attractions are the domain of attraction of Gumbel and Fréchet distributions, respectively.

## 3.2. Order Statistics with Asymptotic Distributions

Let $Y_i = Y_{i:n}$ be the $i^{th}$ order statistics, then the PMF of $Y_i$ is defined as

$$\mathbb{P}_r(Y_{(i)} = x) = \mathbb{P}_r(Y_{(i)} \le x+1) - \mathbb{P}_r(Y_{(i)} \le x), \tag{15}$$

and

$$\mathbb{P}_r(Y_{(i)} \le x) = \mathbb{P}_r(\text{at least i's of Y's are} \le x),$$

$$\mathbb{P}_r(Y_{(i)} \le x) = \sum_{j=i}^{n} (\mathbb{P}_r(Y_{(1)} \le x))^j (1 - \mathbb{P}_r(Y_{(1)} \le x))^{n-j},$$

since $Y_1, Y_2, \ldots, Y_n$ are i.i.d and

$$\begin{aligned}\sum_{j=i}^{n} (\mathbb{P}_r(Y_{(1)} \le x))^j (1 - \mathbb{P}_r(Y_{(1)} \le x))^{n-j} &= \int_0^{\mathcal{F}(x)} \frac{1}{\beta(i, n-i+1)} u^i (1-u)^{n-i} du \\ &= \mathbf{I}_{\mathcal{F}(x)}(i, n-i+1),\end{aligned}$$

where $\mathbf{I}_{\mathcal{F}(x)}(i, n-i+1)$ is the incomplete beta function and $\mathcal{F}(x)$ is the CDF of GNDL. Thus the PMF of the $i^{th}$ order statistics is

$$\mathbb{P}_r(Y_{(i)} = x) = \int_0^{\mathcal{F}(x+1)} \frac{1}{\beta(i, n-i+1)} u^i (1-u)^{n-i} du - \int_0^{\mathcal{F}(x)} \frac{1}{\beta(i, n-i+1)} u^i (1-u)^{n-i} du,$$

$$\mathbb{P}_r(Y_{(i)}=x)=\mathbf{I}_{\mathcal{F}(x+1),\mathcal{F}(x)}(i,n-i+1),$$

where $\mathcal{F}(x)$ is expressed as

$$\begin{aligned}\mathcal{F}(x) &= 1-\frac{(1-p)^{\alpha+1}p^{x}\Gamma(\alpha+x)}{(1-p+\beta)\Gamma(\alpha+1)}\{(\alpha+(\alpha+x-1)\beta)_2F_1(1,\alpha+x,1+x,p)\\ &+ \beta_2F_1(2,x+\alpha+1,1+x,p)\}.\end{aligned}$$

It is also recorded that

$$\mathbb{P}_r(Y_{(i)}=x)=\frac{n!}{i!(n-i)!}\sum_{j=0}^{n-i}\binom{n-i}{j}\frac{(-1)^j}{(i+j)}((\mathcal{F}(x+1))^{i+j}-(\mathcal{F}(x))^{i+j}),$$

as

$$\sum_{j=0}^{n-i}\binom{n-i}{j}\frac{(-1)^j}{(i+j)}(\mathcal{F}(x))^{i+j}=\frac{(\mathcal{F}(x))^i}{i}{}_2F_1(-n+i,i;i+1;\mathcal{F}(x)),$$

$$\sum_{j=0}^{n-i}\binom{n-i}{j}\frac{(-1)^j}{(i+j)}(\mathcal{F}(x+1))^{i+j}=\frac{(\mathcal{F}(x+1))^i}{i}{}_2F_1(-n+i,i;i+1;\mathcal{F}(x+1)).$$

Therefore the PMF of $i^{th}$ order statistics is

$$\begin{aligned}\mathbb{P}_r(Y_{(i)}=x) &= \frac{1}{i}\binom{n}{i}\{(\mathcal{F}(x+1))^i_2F_1(-n+i,i;i+1;\mathcal{F}(x+1))\\ &- (\mathcal{F}(x))^i_2F_1(-n+i,i;i+1;\mathcal{F}(x))\}.\end{aligned}$$

# 4. Parameter Estimation with Inference

If $x_1,x_2,\ldots,x_n$ are $n$ random values of a sample drawn identically independently from the GNDL and the form of a joint probability function is expressed in the log-likelihood function as

$$\begin{aligned}\ell(\Theta)=\mathfrak{L}(p;\alpha,\beta)=n(\alpha+1)\ln(1-p)+\sum_{i=1}^{n}x_i\ln(p)\\ +\sum_{i=1}^{n}\ln(\alpha+(\alpha+x_i-1)\beta)+\sum_{i=1}^{n}\ln((x_i)_{\alpha-1})\\ -n\ln(1-p+\beta)-n\ln(\Gamma(\alpha+1)).\end{aligned}\tag{16}$$

The partial derivative of Equation (16) w.r.t $p,\beta$ and $\alpha$ and equating them to zero yields MLEs of $p$, $\beta$ and $\alpha$, respectively, as

$$-\frac{n(1+\alpha)}{1-p}+\frac{n}{1-p+\beta}+\sum_{i=1}^{n}\frac{x_i-1}{p}=0\tag{17}$$

$$-\frac{n}{1-p+\beta}+\sum_{i=1}^{n}\frac{\alpha+x_i-1}{\alpha+\beta(\alpha+x_i-1)}=0,\tag{18}$$

and

$$-n\ln(1-p)-n\Psi(0,\alpha+1)+\sum_{i=1}^{n}\Psi(0,\alpha+x_i-1)+\sum_{i=1}^{n}\frac{\beta+1}{\alpha+\beta(\alpha+x_i-1)}=0, \quad (19)$$

where $\Psi^n(z)$ is the polygamma function and defined as $\Psi^n(z) = \frac{d^n\Psi(z)}{dz^n}$. Since the closed form solutions of the above equations are not possible, then the MLEs are computed via computational package Mathematica [12.0]. The MLEs of $p, \beta$ and $\alpha$ can be obtained by solving the non-linear normal equations appearing in Equations (17), (18), and (19). Also, the second order derivative of the earlier equations helps to determine the information matrix, which is essential to find the variance-covariance matrix and confidence interval of the estimators. In this regard, the information matrix has the form

$$\mathcal{I}((p,\alpha,\beta)|_{p=\hat{p},\alpha=\hat{\alpha},\beta=\hat{\beta}}) = \begin{bmatrix} \mathcal{E}(\frac{-\partial^2\ell(\Theta)}{\partial p^2}) & \mathcal{E}(\frac{-\partial^2\ell(\Theta)}{\partial p\partial\beta}) & \mathcal{E}(\frac{-\partial^2\ell(\Theta)}{\partial p\partial\alpha}) \\ \mathcal{E}(\frac{-\partial^2\ell(\Theta)}{\partial\beta^2}) & \mathcal{E}(\frac{-\partial^2\ell(\Theta)}{\partial p\partial\beta}) & \mathcal{E}(\frac{-\partial^2\ell(\Theta)}{\partial\beta\partial\alpha}) \\ \mathcal{E}(\frac{-\partial^2\ell(\Theta)}{\partial\alpha^2}) & \mathcal{E}(\frac{-\partial^2\ell(\Theta)}{\partial\alpha\partial\beta}) & \mathcal{E}(\frac{-\partial^2\ell(\Theta)}{\partial\alpha^2}) \end{bmatrix}.$$

However, in view of the regularity conditions, the GNDL $(p, \alpha, \beta)$ model fulfills the regularity conditions as declared by Rohatgi & Saleh (2015, pp. 419). Thus, the confidence interval belt MLE vector of $\hat{\Theta} = (\hat{p},\ \hat{\alpha},\ \hat{\beta})$ is consistent and asymptotically normal family, i.e., $\sqrt{n}[\hat{\Theta}^T - \Theta^T] \sim TVN[0, \mathcal{I}^{-1}]$, where $\mathcal{I}^{-1}$ is the inverse of the expected Fisher information matrix that generates a covariance matrix, which is based on the expectation of second order loglikelihood derivatives.

# 5. Application and Model's Compatibility

Here, we shall concentrate the consideration on model assortment and endorsement in modeling development. But, the model assortment is a stimulating chore, and the chief of an appropriate model, and it is made on the basis of whatever perception is simple with the use of well-deemed reasoning. An inadequate fitting, either logical or graphical, may result due to i) the model is erroneously stated, ii) the model description is right, but regrettably brings a massive bias. In general, endorsement requires extra data, other evidence and additional checking as well as vigilant measurement of the results.

## 5.1. Measures of Goodness-of-Fit

In order to test the hypothesis that $\mathrm{H}_0 : \mathcal{F}_n(x) = \mathcal{F}_o(x)$, where $\mathcal{F}_o(x)$ is the CDF from a specified distribution. In this regard, researchers usually adopt these measures both for discrete and continuous set up like.

- $\chi^2$(Chi Square)-test, for a sample of $n-$ values due to Karl Pearson defined as
$$\chi^2 = \sum_{i=1}^{k} \frac{(f_o(i) - f_e(i))^2}{f_e(i)},$$
where $f_o(i)$ and $f_e(i)$ are the observed and expected frequencies, respectively and $k$ denotes the number of possible classes.

- Kolmogorov Smirnov (KS) test defined as
$$\text{KS} = \max_{1\leq i\leq k}\{\frac{i}{k} - z_i, z_i - \frac{i-1}{k}\},$$
where $z_i$ is CDF of the distribution.

- Anderson–Darling ($\text{AD}_0^*$)-test usually attaches more mass to the tails, which is defined as
$$A_0^* = \left(\frac{2.25}{k^2} + \frac{0.75}{k} + 1\right)\left\{-k - \frac{1}{k}\sum_{i=1}^{k}(2i-1)\ln(z_i(1 - z_{k-i+1}))\right\}.$$

- Cramér–von Mises ($\text{C}VM_0^*$)-test derived version of KS test defined as
$$W_0^* = \sum_{i=1}^{K}\left(z_i - \frac{2i-1}{2k}\right)^2 + \frac{1}{12k}.$$

- Akaike information criterion (AIC) defined as $AIC = 2m - 2\ell(\hat{\Theta})$, where $m$ denote the number of parameters.

- Corrected Akaike information criterion (AICc) expressed as $AICc = AIC + \frac{2m(m+1)}{n-m-1}$.

- Bayesian information criterion (BIC), which is defined as $BIC = m\ln(n) - 2\ell(\hat{\Theta})$.

- Hannan-Quinn information criterion (HQIC) expressed as $HQIC = -2\ell(\hat{\Theta}) + 2m\ln(\ln(m))$.

- Consistent Akaike information criterion (CAIC) given as $CAIC = -2\ell(\hat{\Theta}) + m(\ln(n) + 1)$.

- Vuong test proposed by Vuong, (see Vuong, 1989) is also used for model selection purposes.

For comprehensive details about these measures readers are referred to Hussain et al. (2019), Murthy et al. (2004) and Vuong (1989), respectively.

## 5.2. Working Methodology and Competing Models

In order to model partly cloudy days of various regions in USA and assist the space scientists around the globe, we first search out the relevant data sets, which can be accessed from `http://www.ncdc.noaa.gov/data-access/quick`. Next, for selecting the application areas, we have to search out the possible competitors of the proposed model. Since the proposed model is based on the nabla structure with a domain belonging to $\mathbb{N}$, so we have decided to choose those competitors that not only work very well in over-, under- and equi-dispersed data sets very well but also should be defined on positive integers. These benchmark indicators have paved the path to reach the truth in an elegant manner. So the selection of competitor models includes the zero truncated generalized Poisson (ZTGP), see Consul & Famoye (1989), zero truncated negative binomial (ZTNB), see Sampford (1955) and zero truncated Poisson Lindley (ZTPL) distributions, see Aderoju et al. (2017).

## 5.3. Descriptive Summary of Data Sets

Four data sets are used for analysis purposes, which deal with the partly cloudy days of 40,34, 40, and 38 stations. Data exhibits the average number of days per category of cloudiness that denotes 4/10 to 7/10 mean sky cover. From Table 2 it is obvious that the data sets I, II, and IV are negatively skewed, while the III data set is positively skewed. Moreover, non stationarity is observed in cloudy data sets behaviour, which is probably due to different region of USA. Such statistics are computed online by using the Wald-Wolfowitz test available at `https://rdrr.io/cran/trend/man/ww.test.html`. Furthermore, Figure 4 indicates non existence of outliers in the selected data sets.

## 5.4. Clouds and Their Impacts

Since, the key incentive for writing this article is to model the environmental effects, for this purpose, we have decided to model the environmental cloudy days, data sets. Clouds are potent causes of universal climate shift. These are the vital element shaping local weather and the Earth's climate structure. By monitoring clouds, we can get evidence about moisture, temperature, and wind situations at different altitudes in the atmosphere. This evidence supports forecasting meteorological conditions. For this purpose NASA and other space agencies have a number of satellites orbiting the Earth and collecting data about clouds and the Earth's energy.

### 5.4.1. Partly Cloudy Days

The following data sets behavior, demonstrate the mean number of partly cloudy days. Partly cloudy involves 4/10 to 7/10 average sky cover by clouds whereas cloudy involves 8/10 to 10/10 average sky cover by clouds. From July 1996 sky coverage is described to eighths (oktas). Clear implies 0-2 oktas, partly cloudy suggests 3-6 oktas and cloudy implies 7-8 oktas. The cloudiness data are unaffected from year to year. Revises to the cloudiness data concluded with the conversion to Automated Surface Observing Stations (ASOS) at National Weather Service (NWS) sites. In the 1990s cloud dimensions carried by trained witnesses commenced to be phased out in approval of computerized radars that assessed clouds lower than 12,000 feet. In this regard, the first, second, third and fourth data sets observe the number of partially cloudy days of February 2015:6, 6, 7, 6, 4, 4, 6, 4, 4, 7, 5, 6, 4, 4, 3, 5, 5, 3, 4, 4, 6, 4, 3, 3, 3, 6, 7, 7, 7, 7, 6, 6, 8, 8, 5, 6, 8, 7, 6, 7; March 2015: 8, 8, 8, 8, 5, 5, 7, 5, 5, 8, 6, 7, 6, 6, 3, 6, 7, 6, 6, 5, 7, 4, 6, 4, 4, 8, 8, 7, 9, 7, 7, 7, 9, 9; April 2015: 8, 7, 9, 8, 6, 6, 7, 6, 10, 9, 4, 8, 10, 6, 4, 7, 6, 7, 7, 6, 6, 5, 5, 6, 5, 9, 7, 7, 9, 6, 7, 7, 9, 9, 7, 8, 8, 10, 9, 10 and November 2015: 7, 7, 7, 6, 5, 3, 3, 5, 6, 6, 6, 5, 5, 5, 2, 5, 7, 4, 4, 4, 6, 3, 5, 4, 3, 7, 6, 6, 6, 6, 6, 6, 8, 8, 5, 7, 7, 8, respectively. Wald-Wolfowitz test for randomness is also used to check the independence assumption, which indicates that data set I, II and IV do not follow the independece at 5% level of significance with Z-score $Z_I = 3.6362, Z_I = 2.2922$ and $Z_{IV} = 2.8261$, respectively, while the third data set III follows the assumption at 5% level of significance with Z-score $Z_{III} = 1.65$. Moreover, the empirical summary of these data sets is given below

Table 2: Descriptive Summary of the Data Sets.

| Data Set | Sample Size | Mean | Median | S.D | $\mathbb{ID}$ | SK | KU |
|---|---|---|---|---|---|---|---|
| I | 40 | 5.43 | 6.0 | 1.5340 | 0.4338 | -0.0941 | 1.8639 |
| II | 34 | 6.50 | 7.0 | 1.5618 | 0.3752 | -0.2906 | 2.3406 |
| III | 40 | 7.25 | 7.0 | 1.6447 | 0.3731 | 0.0131 | 2.2211 |
| IV | 38 | 5.50 | 6.0 | 1.5201 | 0.4202 | -0.3743 | 2.4971 |

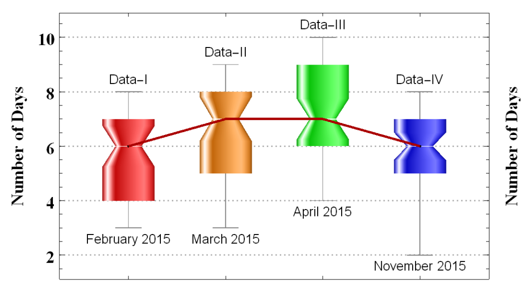


Figure 4: Box and Whisker plots of Data Sets.

### 5.4.2. Analysis of Data Sets

Goodness-of-fit statistics of the first data, as portrayed in Table 3 indicate that GNDL is a good strategy for the analysis of data set-I. In this table we observed that although the proposed model yields a competing statistics for all the parameteric and non-parameteric tests, yet $\chi^2$ goodness-of-fit test advocates the GNDL by generating the highest $p$-value as compared to the competing models. In addition, Table 4 also indicates that GNDL is a good choice by showing minimum values of all information criteria. Moreover, the Vuong statistics yield values of

137.5935, 137.5926 and 137.6067 (for comprehensive details readers are refered to Hussain et al., 2019) when comparing GNDL with ZTGP, ZTNB and ZTPL distributions, respectively. Thus indicating that GNDL is the right choice for this data set.

TABLE 3: MLEs and measures of fit for Data-I

| Distribution | $\hat{p}$ | $\hat{\alpha}$ | $\hat{\beta}$ | $\chi^2$ | $p$-value | KS | $A_0^*$ | $W_0^*$ |
|---|---|---|---|---|---|---|---|---|
| GNDL$(p;\alpha;\beta)$ | 0.00007 | 60733.716 | 10.6369 | 0.5959(2) | 0.7423 | 0.1508 | 0.0321 | 0.2491 |
| ZTGP$(p;\beta)$ | $2.90\times10^{-9}$ | - | 5.4005 | 0.7498(2) | 0.6873 | 0.1885 | 0.0321 | 0.2719 |
| ZTNB$(p;\beta)$ | 0.00003 | - | 196700.19 | 0.7498(2) | 0.6874 | 0.1885 | 0.0321 | 0.2711 |
| ZTPL$(p;\beta)$ | $1.93\times10^{6}$ | - | 355406.547 | 0.7498(2) | 0.6874 | 0.1884 | 0.0321 | 0.2721 |

TABLE 4: Criteria based on log-likelihood for Data-I.

| Distribution | $-l$ | AIC | AICC | BIC | HQIC |
|---|---|---|---|---|---|
| GNDL | 76.9293 | 159.859 | 160.525 | 164.925 | 161.69 |
| ZTGP | 78.9685 | 161.937 | 162.261 | 165.315 | 163.158 |
| ZTNB | 78.9688 | 161.938 | 162.262 | 165.315 | 163.159 |
| ZTPL | 78.9685 | 161.937 | 162.261 | 165.315 | 163.158 |

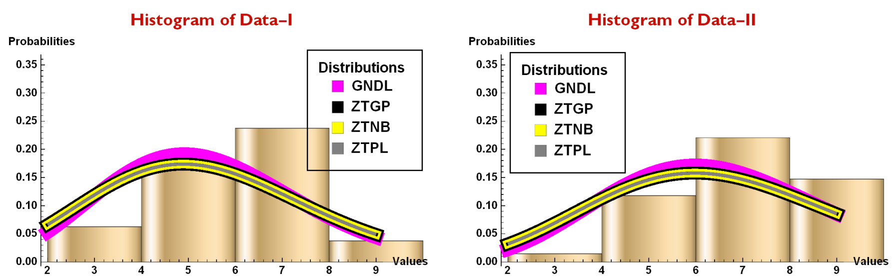


FIGURE 5: Histogram of data set-I and -II and estimated PMF.

Similalry, in modeling the second data set, we have observed that the GNDL yields the lowest values of all the goodness-of-fit test statistics and the corresponding highest $p$-value, which are portrayed in Table 5. In addition, the information criteria as portrayed in Table 6 are also minimum, except for the BIC, which arises due to over fitting issue of the GNDL, i.e., all goodness-of-fit statistics yield smart values. Moreover, the Vuong statistics, when compare the GNDL with ZTGP, ZTNB and ZTPL gives values: 70.1713, 69.8878 and 70.1658, respectively, which clearly indicate the superiority of the GNDL for such a data set. Furthermore, the histograms in Figure 5 of the first and second data sets also advocate the GNDL for its suitability of fit.

TABLE 5: MLEs and measures of fit for Data-II.

| Distribution | $\hat{p}$ | $\hat{\alpha}$ | $\hat{\beta}$ | $\chi^2$ | $p$-value | KS | $A_0^*$ | $W_0^*$ |
|---|---|---|---|---|---|---|---|---|
| GNDL$(p;\alpha;\beta)$ | 0.00004 | 127932.544 | 10.3022 | 3.2110(3) | 0.3602 | 0.1322 | 0.0195 | 0.1783 |
| ZTGP$(p;\beta)$ | $1.5124\times10^{-9}$ | - | 6.4901 | 4.0873(2) | 0.2522 | 0.1639 | 0.0257 | 0.2260 |
| ZTNB$(p;\beta)$ | 0.00008 | - | 82591.7697 | 3.8532(3) | 0.2778 | 0.1638 | 0.0256 | 0.2259 |
| ZTPL$(p;\beta)$ | $1.9372\times10^{6}$ | - | 298497.9241 | 3.8496(2) | 0.1459 | 0.1639 | 0.0257 | 0.2260 |

Table 6: Criteria based on log-likelihood for Data-II.

| Distribution | $-l$ | AIC | AICC | BIC | HQIC |
|---|---|---|---|---|---|
| GNDL | 67.8076 | 141.615 | 142.415 | 146.194 | 143.177 |
| ZTGP | 69.4374 | 142.875 | 143.262 | 145.928 | 143.916 |
| ZTNB | 69.4383 | 142.877 | 143.264 | 145.929 | 143.918 |
| ZTPL | 69.4375 | 142.875 | 143.262 | 145.928 | 143.916 |

The other two data sets also deal with the number of partly cloudy days of different cities of USA. In modeling the third data set, we observed from Table 2 that it is positively skewed one and from Table 7 one can infer that the GNDL yields minimum values of all the goodness-of-fit test statistics and the highest $p$-value. However, from Table 8 we have observed that the information criteria are minimum, except BIC, which is a little bit higher due over parameterization. In addition, on comparing the Vuong statistics 176.7445, 176.7844 and 176.7618 of the GNDL with ZTGP, ZTNB and ZTPL distributions, respectively, we found that GNDL is the most suitable choice for modeling the third data set.

Table 7: MLEs and measures of fit for Data-III.

| Distribution | $\hat{p}$ | $\hat{\alpha}$ | $\hat{\beta}$ | $\chi^2$ | $p$-value | KS | $A_0^*$ | $W_0^*$ |
|---|---|---|---|---|---|---|---|---|
| GNDL$(p;\alpha;\beta)$ | 0.00006 | 105767 | 10.7972 | 6.1898(4) | 0.1854 | 0.1538 | 0.0210 | 0.1951 |
| ZTGP$(p;\beta)$ | $1.49\times10^{-9}$ | - | 7.2448 | 7.4732(4) | 0.1129 | 0.1861 | 0.0304 | 0.2618 |
| ZTNB$(p;\beta)$ | 0.00004 | - | 182458.34 | 7.4729(4) | 0.1129 | 0.1861 | 0.0304 | 0.2619 |
| ZTPL$(p;\beta)$ | $1.97\times10^{6}$ | - | 271433.76 | 7.4730(4) | 0.1129 | 0.1861 | 0.0304 | 0.2619 |

Table 8: Criteria based on log-likelihood for Data-III.

| Distribution | $-l$ | AIC | AICc | BIC | HQIC |
|---|---|---|---|---|---|
| GNDL | 81.9183 | 169.837 | 170.503 | 174.903 | 171.668 |
| ZTGP | 83.7181 | 171.436 | 171.76 | 174.814 | 172.657 |
| ZTNB | 83.7185 | 171.437 | 171.761 | 174.815 | 172.658 |
| ZTPL | 83.718 | 171.436 | 171.76 | 174.814 | 172.657 |

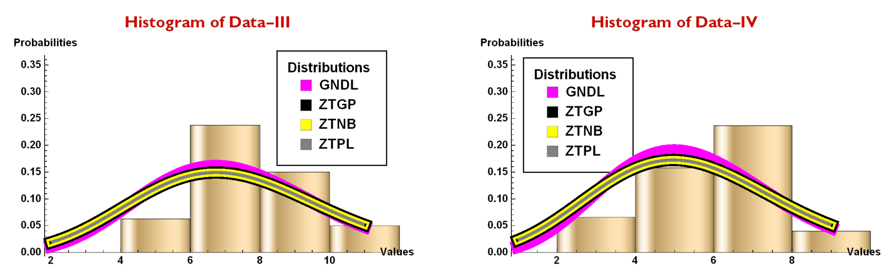


Figure 6: Histogram of data set-III and -IV and estimated PMF.

While modeling the fourth data set, we have observed that the $\chi^2$ and KS goodness-of-fit statistics of the GNDL least while Cramér-von Misses and Anderson Darling are not only minimum for ZTGP, ZTNB and ZTPL but also equally well in behavior. These goodness-of-fit statistics are portrayed in Table 9. Thus indicating the GNDL is the best strategy for modeling the fourth data set too. Moreover, the histogram of data set-III and-IV as portrayed in Figure 6, also advocates the GNDL for its best fitting behavior.

TABLE 9: MLEs and measures of fit for Data-IV.

| Distribution | $\hat{p}$ | $\hat{\alpha}$ | $\hat{\beta}$ | $\chi^2$ | $p$-value | KS | $A_0^*$ | $W_0^*$ |
|---|---|---|---|---|---|---|---|---|
| GNDL$(p;\alpha;\beta)$ | 0.00003 | 174051.85 | 0.0 | 4.0496(3) | 0.2562 | 0.1232 | 0.0210 | 0.1841 |
| ZTGP$(p;\beta)$ | $0.10\times10^{-9}$ | - | 5.4769 | 5.2572(3) | 0.1539 | 0.1271 | 0.0211 | 0.1705 |
| ZTNB$(p;\beta)$ | 0.000023 | - | 235858.92 | 5.2574(3) | 0.1539 | 0.1271 | 0.0211 | 0.1705 |
| ZTPL$(p;\beta)$ | $1.92\times10^{6}$ | - | 350593.81 | 5.2572(3) | 0.1539 | 0.1271 | 0.0211 | 0.1705 |

Moreover, Vuong statistics of GNDL-ZTGP, GNDL-ZTNB and GNDL-ZTPL distributions are 27.0677, 27.0665, and 27.0675, indicating that the GNDL is a reasonably good choice for such data analysis. Furthermore, Table 10 indicates that the proposed model is the least loss of information model with minimum values of model selection benchmarks.

TABLE 10: Criteria based on log-likelihood for Data-IV.

| Distribution | $-l$ | AIC | AICC | BIC | HQIC |
|---|---|---|---|---|---|
| GNDL | 73.4597 | 152.919 | 153.625 | 157.832 | 154.667 |
| ZTGP | 75.2843 | 154.569 | 154.911 | 157.844 | 155.734 |
| ZTNB | 75.2845 | 154.569 | 154.912 | 157.844 | 155.734 |
| ZTPL | 75.2843 | 154.569 | 154.911 | 157.844 | 155.734 |

# 6. Conclusion

A discretized version of the generalized Lindley distribution by using discrete fractional calculus under the name GNDL, is suggested and its several chattels are observed. These properties includes discrete Laplace transformation, moment and probability generating functions, $r^{th}$ moments recurrence relations and recurrence relation between probabilities. The GNDL is log-concave, IFR, unimodal and DMRL, log-convex, DFR and IMRL and BTS in behavior. It is also observed that the GNDL is not only pliable but also mathematically amenable for over-, under- and equi-dispersed count data sets. Moreover, the purported model is strongly recommended for modeling the discrete environmental indicators defined over the domain $\mathbb{N}$, which can help the scientists in planning the space mission in a realistic sense.

This strategy thus will be quite helpful for NASA and other space agencies for technically handling the hazards in launching space missions and predicting the cloudy days in a well mannered way. Moreover, the GNDL is also capable to work efficiently in over- and equi-dispersion cases for modeling the number of earth quake of certain magnitudes, which hopefully be addressed in future communication.

[Received: February 2024 — Accepted: August 2024]

## References

Aderoju, S. A., Jolayemi, E. T. & Ibrahim, A. O. (2017), 'A comparative study on zero-truncated generalized poisson-lindley and zero-truncated poisson-lindley distributions', *International Journal of Mathematical Archive* **8**(3), 92–98.

Ameen, R., Hasan, K. O. S. E. & Jarad, F. (2019), 'On the discrete laplace transform', *Results in Nonlinear Analysis* **2**(2), 61–70.

Banik, S. & Kibria, B. M. (2009), 'On some discrete distributions and their applications with real life data', *Journal of Modern Applied Statistical Methods* **8**(2), 8.

Bohner, M. & Peterson, A. (2001), *Dynamic Equations on Time Scales*, Birkhäuser, Boston.

Bohner, M. & Peterson, A. (2003), *Advances in Dynamic Equations on Time Scales*, Birkhäuser, Boston.

Carter, M. C. (1972), 'A model for thunderstorm activity: Use of the compound negative binomial-positive binomial distribution', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **21**(2), 196–201.

Chakraborty, S. (2015), 'Generating discrete analogues of continuous probability distributions-a survey of methods and constructions', *Journal of Statistical Distributions and Applications* **2**, 1–30.

Consul, P. C. & Famoye, F. (1989), 'The truncated generalized poisson distribution and its estimation', *Communications in Statistics-Theory and Methods* **18**(10), 3635–3648.

Ganji, M. & Gharari, F. (2018*a*), 'The discrete delta and nabla mittag-leffler distributions', *Communications in Statistics-Theory and Methods* **47**(18), 4568–4589.

Ganji, M. & Gharari, F. (2018*b*), 'A new method for generating discrete analogues of continuous distributions', *Journal of Statistical Theory and Applications* **17**(1), 39–58.

Gharari, F., Bakouch, H. & Karakaya, K. (2023), 'A pliant model to count data: Nabla poisson-lindley distribution with a practical data example', *Bulletin of the Iranian Mathematical Society* **49**(3), 32.

Gharari, F. & Ganji, M. (2021), 'A new stochastic order based on discrete laplace transform and some ordering results of the order statistics', *Communications in Statistics-Theory and Methods* **52**(6), 1963–1980.

Hagmark, P. E. (2008), 'On construction and simulation of count data models', *Mathematics and Computers in Simulation* **77**, 72–80.

Hussain, T., Bakouch, H. S. & Chesneau, C. (2019), 'A new probability model with application to heavy-tailed hydrological data', *Environmental and Ecological Statistics* **26**, 127–151.

Leadbetter, M. R., Lindgren, G. & Rootzen, H. (1987), *Extremes and Related Properties of Random Sequences and Processes*, Springer Verlag, New York.

Lloyd-Smith, J. O. (2007), 'Maximum likelihood estimation of the negative binomial dispersion parameter for highly over dispersed data, with applications to infectious diseases', *PloS one* **2**(2), e180.

Murthy, D. N. P., Xie, M. & Jiang, R. (2004), *Weibull Models*, John Wiley and Sons.

Nory, M. K. M., Abd, T. & Algamal, Z. Y. (2022), 'Count data modeling under over dispersion issue: A comparative study', *Mathematical Statistician and Engineering Applications* **71**(3), 1398–1406.

Rohatgi, V. K. & Saleh, A. M. E. (2015), *An introduction to probability and statistics*, John Wiley and Sons.

Roy, D. (2004), 'Discrete rayleigh distribution', *IEEE Transactions on Reliability* **53**(2), 255–260.

Sampford, M. R. (1955), 'The truncated negative binomial distribution', *Biometrika* **42**(1/2), 58–69.

Vuong, Q. H. (1989), 'Likelihood ratio tests for model selection and non-nested hypotheses', *Econometrica* **57**(2), 307–333.

Zakerzadeh, H. & Dolati, A. (2009), 'Generalized lindley distribution', *Journal of Mathematical Extension* **3**(2), 1–17.