CHAPTER II

Polya and Negative Polya-Eggenberger Distributions

2.1 Introduction 48
2.2 Polya-Eggenberger Distribution (PED) 49
2.3 Size-Biased Polya-Eggenberger Distribution (SBPED) 53
2.3.1 Some Models of SBPED 54
2.3.2 Structural Properties of SBPED 55
2.3.3 Relation of SBPED with other Distributions 57
2.3.4 Estimation of the Parameters of SBPED 58
2.4 Negative Polya-Eggenberger Distribution (NPED) 61
2.4.1 Some Models of NPED 61
2.4.2 Structural Properties of NPED 64
2.4.3 Relation of NPED with other Distributions 70
2.4.4 Estimation of the parameters of NPED 72
2.5 Size-Biased Negative Polya-Eggenberger Distribution (SBNPED) 76
2.5.1 Some Models of SBNPED 76
2.5.2 Structural Properties of SBNPED 77
2.5.3 Relation of SBNPED with other Distributions 81
2.5.4 Estimation of the parameters of SBNPED 82
2.6 A New Contagious Distribution Analogous to Negative Polya-Eggenberger Distribution (CDANPED) 86
2.6.1 The proposed model (SBNPED) 86
2.6.2 Structural Properties of CDANPED 88
2.6.3 Relation of CDANPED with other Distributions 92
2.7 Goodness of Fit 95
Chapter II
Polya and Negative Polya-Eggenberger Distribution

2.1 Introduction

The contagion model was first given originally by Greenwood and Yule (1920) in a work related to the study of accidents where it was assumed that the occurrence of an accident by chance in some way has the effect of increasing or decreasing the probability of occurrence of additional accidents. Shortly afterwards Greenwood and Yule’s work, Polya and Eggenberger (1923) discovered Polya-Eggenberger distribution which they described as ‘truly contagious’ distribution.

Greenwood and Yule discussed two types of what may conveniently be called contagion: with one type there is true contagion in the sense of Polya and Eggenberger, where each “favorable” event increases (or decreases) the probability of succeeding favorable events; with the second type the events are independent and an apparent contagion is actually due to an inhomogeneity of the population. The two applications are very different in nature as well in practical implications. As pointed out by Feller (1943), it is therefore most remarkable that Greenwood and Yule found their distribution assuming an apparent contagion; in their opinion this distribution contradicts true contagion. On the contrary, as pointed out above, Polya and Eggenberger arrived at the same distribution assuming true contagion, while the possibility of an apparent contagion due to inhomogeneity seems not to have been noticed by them. Though, it may be mentioned here that both the proneness (or apparent contagion) and contagion models given by Greenwood and Yule lead to negative binomial distribution.

The original work of Greenwood and Yule has lead other authors to a consideration of the problem of contagion and to many additional distributions. During this process a huge amount of work on contagious distributions can be seen in literature because of the diversified applications of these distributions in real life
problem especially in the field of entomology, ecology, bacteriology and accidental statistics. The data arising from these fields can not be described by usual distribution functions but rather by some type of contagious distributions.

A class of contagious distributions is derived from a certain biological model which takes into account the fact that the distribution of larvae over the plots of a field depends upon the fact that the larvae are hatched from egg-masses which appear at random over the field has been derived by Neyman (1939), Evans (1953) and Beal and Rescia (1953). This class of distributions has been successful in accounting for the distribution of some insect populations (cf. Beal-1940). Feller (1943) proposed a general class of contagious distributions from which Neyman’s Types A, B, C contagious distributions can be derived. Beal and Rescia (1953) suggested another generalization of Neyman’s contagious distributions. Bliss and Fisher (1953) showed that the negative binomial distribution is useful as a possible underlying distribution for insect populations. Contagious distributions have also been used in the study of accident and medical statistics by Dubourdieu (1939), Greenwood and Yule (1920), Lundberg (1940) and Newbold (1927).

Besides above applications there is a numerous literature present on the contagious distributions e.g. Feller (1940), Bliss and fisher (1953), Thompson (1954), Duglas (1955), Gurland (1957), Archibald (1948), Barton (1957), Beall (1940), Bosch(1963), Chaddha (1963), Katti and Gurland (1962), Subrahmaniam (1966), Woodbury (1949), Engen (1974), Gordon (1989), Kanazawa (1999), Kwang-Su and Chi-Hyuck [(2000),(2002)] etc., see; Johnson and Kotz (1969, 1993), S and Kemp. A.W (1992) and Bibliography of this thesis for detailed references.

In this chapter, we have concentrated mainly on Polya-Eggenberger distribution and negative Polya-Eggenberger distribution which are discussed in detail in the succeeding sub-sections of this chapter. The generalizations of these distributions are discussed fully in the next chapter-III.

2.2 Polya-Eggenberger Distribution
The Polya-Eggenberger distribution (PED) is considered as a truly contagious distribution and was introduced by Polya and Eggenberger (1923) through urn model. Some further analysis was given by Polya in (1930). The genesis of this distribution is conveniently expressed in terms of random drawings of colored balls from an urn. Initially it is supposed that there are \(a\) white balls and \(b\) black balls in the urn. One ball is drawn at random, and then replaced, together with \(s\) balls of the same color. If this procedure is repeated \(n\) times, and \(x\) represents the total number of times a white ball is drawn, then the distribution of \(x\) is given by

\[
P(x = k) = \binom{n}{x} \frac{(a)(a+s)...(a+x-1s)(b+s)...(b+n-x-1s)}{(a+b)(a+b+s)...(a+b+n-1s)}
\]

(2.2.1)

Where \(x = 0, 1, 2, ..., n\); \(a, b, \) and \(s\) are parameters of the distribution. The distribution represented by (2.2.1) is known as Polya-Eggenberger distribution with parameters \((n, a, b, s)\).

Taking \(\alpha = (a/s); \gamma = (b/s)\), we get alternative form of (2.2.1) in ascending factorials as

\[
P(x = k) = \binom{n}{k} \frac{\alpha^x \gamma^{n-x}}{(\alpha + \gamma)^n}
\]

(2.2.2)

The distribution (2.2.2) is the most convenient form of Polya-Eggenberger distribution (PED).

Other way of representing (2.2.1) is

\[
P(X = x) = \binom{n}{x} \frac{(-a)^x (-b)^x}{(a+b)^x}
\]

(2.2.3)

An alternative form of (2.2.1) in terms of parameters \(n; P = a/(a+b); Q = 1-P = b/(a+b)\) and \(\alpha = s/(a+b)\) is

\[
P(x = x) = \binom{n}{x} \frac{P(P+\alpha)...(P+x-1\alpha)Q(Q+\alpha)...(Q+n-x-1\alpha)}{(1+\alpha)(1+2\alpha)...(1+n-1\alpha)}
\]
It is possible for $s$ (and so $\alpha$) to be negative. However $s$ must satisfy the inequality

$$(a + b) + s(n - 1) > 0$$

Srodka, T. (1964) gave the recurrence relation among the moments about zero of the PED (2.2.3) as

$$
\mu_{r+1} = (c + b + rs)^{-1} \sum_{j=0}^{r} \binom{r}{j} - (a - sn) \binom{r}{j+1} - s \binom{r}{j+2} \mu_{r-1}
$$

and the $r$th factorial moment is given by

$$
\mu_r = E[x^r]
$$

$$
= n^r \prod_{j=0}^{k-r} \frac{(P + j\alpha)}{l + j\alpha} \sum_{k-r}^{n-r} \binom{n-r}{k-r} \prod_{j=0}^{k-r-1} (P + j\alpha') \prod_{j=0}^{n-r-1} (Q + j\alpha') \prod_{j=0}^{k-r} (1 + j\alpha')
$$

Where $P' = (P + r\alpha)(1 + r\alpha)^{-1}$; $Q' = Q(1 + r\alpha)^{-1}$; $\alpha' = \alpha(1 + r\alpha)^{-1}$

In particular, the first four central moments of Polya-Eggenberger distribution are

$$
\mu_1 = nP
$$

$$
\mu_2 = n(n - 1)P(P + \alpha)(1 + \alpha)^{-1} + nP - (nP)^2
$$

$$
= nPQ(1 + n\alpha)(1 + \alpha)^{-1}
$$

$$
\mu_3 = nPQ(Q - P)(1 + n\alpha)(1 + 2n\alpha)(1 + \alpha)^{-1}(1 + 2\alpha)^{-1}
$$

$$
\mu_4 = \frac{nPQ(1 + n\alpha)(1 + 3n\alpha)(1 - 3PQ) + (n - 1)(\alpha + 3PQ(1 + n\alpha))}{(1 + \alpha)(1 + 2\alpha)(1 + 3\alpha)}
$$

From these formulas we obtain the moment ratios
Polya (1930) pointed out the following particular cases: if \( s \) is positive, then success and failure are both contagious; if \( s = 0 \), then events are independent and (2.2.1) reduces to a binomial distribution; when \( s = -1 \), the outcome is a classical hypergeometric distribution, whereas when \( s = 1 \), the outcome is a negative hypergeometric distribution. These relationships are apparent from the genesis of the distribution. Also, taking \( a = b = s \), we obtain a discrete rectangular distribution.

The negative binomial distribution is obtained as a limiting distribution as \( n \to \infty \), \( a(a+b)^{-1} \to 0 \) and \( s(a+b)^{-1} \to 0 \) in such a way that \( na(a+b)^{-1} \) and \( ns(a+b)^{-1} \) tend to finite non-zero values \( \theta \), \( p \), respectively. This limiting form of (2.2.1) is sometimes referred to as a “Polya” distribution [e.g. Gnedenko, B. V. (1961), Eisenhart and Zelen (1958), Arley and Buch (1950), Hald (1952)]. On the other hand Bosch (1963) called the general distribution (2.2.1) a “Polya” distribution and reports further that it is some times called a Skellam distribution [see also Skellam (1948)]. Patil and Joshi (1968) term the negative binomial a “Polya-Eggenberger” and the distribution (2.2.4) simply a “Polya” distribution. It is also the Type IIA generalized hypergeometric distribution of Kemp and Kemp (1956), where a method of estimating the parameters by the method of maximum likelihood has also been discussed. It is also known by beta-binomial distribution, a beta mixture of binomial distribution. The term binomial-beta was given by Ish and Hayakawa (1960). Sarkadi (1957) has shown that the Polya-Eggenberger distribution is also related to the distribution of the “number of exceedances,” i.e. of the number of random variables, among a set of size \( n \), the values of which are larger than the values of at least \( (N - m + 1) \) out of \( N \) other random variables, all \( (n + N) \) random variables being
independent and having identical continuous distributions. Gumbel and von Schelling (1950) showed that this distribution is defined by

\[ P(Y = x) = \binom{n}{m} m^m \left( \frac{n}{n+m} \right)^{n+m-1} \]

This is a Polya-Eggenberger distribution with parameters \( n \), \( p = \frac{m}{n+m} \) and \( \alpha = (N+1)^{-1} \).

2.3. The Size-Biased Polya-Eggenberger Distribution

Size-biased distributions arise naturally for some sampling plans in reliability, biometry, and survival analysis and in several contexts in forestry and ecology. Simple power relationships (e.g., basal area and diameter at breast height) between variables are one such area of interest arising from a modeling perspective. Another, probability proportional to size (PPS) sampling, is found in the most widely used methods for sampling standing or dead and fallen material in the forest. Often it is desirable or necessary to estimate a parametric probability density model based on size-biased data. Traditional equal probability methods may not be appropriate, or may be less efficient in such circumstances, and estimation is better conducted utilizing size-biased theory.

Size-biased distributions are a special case of the more general form known as weighted distributions. First introduced by Fisher (1934) to model ascertainment bias, weighted distributions were later formalized in a unifying theory by Rao (1965). Such distributions arise naturally in practice when observations from a sample are recorded with unequal probability, such as from probability proportional to size (PPS) designs. Briefly, if the random variable \( X \) has distribution \( f(x; \theta) \), with unknown parameters \( \theta \), then the corresponding weighted distribution is of the form

\[ f^w(x; \theta) = \frac{w(x) f(x; \theta)}{E[w(x)]} \]

(2.3.1)

Where \( w(x) \) is a non-negative weight function such that \( E[w(x)] \) exists.
A special case of interest arises when the weight function is of the form \( w(x) = x^\alpha \). Such distributions are known as size-biased distributions of order \( \alpha \) and are written as [Patil and Ord, 1976; Patil, 1981; Mahfoud and Patil, 1982]:

\[
f^{\alpha}(x, \theta) = \frac{\alpha f(x, \theta)}{\mu_\alpha}
\]  

(2.3.2)

Where \( \mu_\alpha = \sum x^\alpha f(x, \theta) \) is the \( \alpha \)th raw moment of \( f(x, \theta) \).

The size-biased version of the Polya-Eggenberger distribution has not been introduced so far. The purpose of this section is to introduce and study the size-biased Polya-Eggenberger distribution in detail.

2.3.1 Some Models of SBPED

If \( X \) is a Polya-Eggenberger variate with probability mass function (2.2.2), then its mean is given by

\[
\mu_\alpha = \frac{n \alpha}{(\alpha + \gamma)} 
\]  

(2.3.3)

Taking \( \alpha = 1 \) in (2.3.2), then \( \mu_1 = E(X) \) is the mean of the Polya-Eggenberger variate \( X \) given by (2.3.3) and the equation (2.3.2) reduces to

\[
p(x) = \frac{x}{n \alpha} \binom{n}{x} \frac{\gamma^{x} \mu^{n-x}}{(\alpha + \gamma)^{n}} 
\]  

(2.3.4)

Which can be expressed as

\[
p(x) = \binom{n-1}{x-1} \frac{\gamma^{x-1} \mu^{n-x}}{(\alpha + \gamma + 1)^{n-1}}, \quad x = 1, 2, \ldots, n 
\]  

The equation (2.3.4) gives the probability mass function of the size-biased Polya-Eggenberger distribution.

The Mixture Model

The size-biased Polya-Eggenberger distribution is obtained by compounding the size-biased binomial distribution (1.2.4) through the value of \( \rho \) with the beta distribution.
Thus, the mixture model of SBPED is obtained as

\[
P(x) = \left( \begin{array}{c} n-1 \\ x \end{array} \right) \frac{1}{\beta(\alpha,\gamma)} \int_0^1 p^{\gamma n-x-1} (1-p)^{\alpha x} dp \\
= \frac{\alpha x}{\gamma n-x-1} 
\]

Replacing \( \alpha \) with \( \alpha + 1 \), (2.3.6) coincides with the probability mass function of the size-biased Polya-Eggenberger distribution (2.3.4).

2.3.2 Structural Properties of SBPED

2.3.2.1 Recurrence relation between probabilities

Taking \( x = x + 1 \) in (2.3.4) and dividing the resulting equation by (2.3.4), we get

\[
\frac{P(x+1)}{p(x)} = \frac{(n-x)}{x} \frac{(\alpha x)}{\gamma + n-x-1}
\]

Which can be written as

\[
P(x+1) = \left[ \frac{(n-x)}{x} \frac{(\alpha x)}{\gamma + n-x-1} \right] p(x) \tag{2.3.6}
\]

Which is the required recurrence relation between probabilities of the proposed model (SBPED).

2.3.2.2 Recurrence relation between moments

Multiplying equation (2.3.6) by \( x^k \) and summing the resulting equation over \( x \), we get

\[
\mu_{k+1}(n,\alpha,\gamma) = \sum_{x=1}^{n} x^k \frac{(n-1)!}{(n-x-1)!(x-1)!} \frac{(\alpha+1)^x \gamma^{n-x}}{(\alpha + \gamma + 1)^{n-1}}
\]
\[
\sum_{x=1}^{n-1} x^k \frac{(n-1)!}{(n-1-x)! (x-1)!} \frac{(\alpha + 2)!}{(\alpha + \gamma + 2)!} y^{n-1-k} = \frac{(n-1)(\alpha + 1)!}{(\alpha + \gamma + 1)!} \sum_{x=1}^{n-1} x^k \frac{(n-1)!}{(n-1-x)! (x-1)!} \frac{(\alpha + 2)!}{(\alpha + \gamma + 2)!} y^{n-1-k} \]

Which subsequently reduces to

\[
\mu_{k+r}(n, \alpha, \gamma) = \frac{(n-1)(\alpha + 1)!}{(\alpha + \gamma + 1)!} \mu_{k+r}(n-1, \alpha + 1, \gamma) \quad (2.3.7)
\]

The equation (2.3.7) is the required recurrence relation between moments of the proposed model (2.3.4). In particular

\[
\mu_1 = \frac{(n-1)(\alpha + 1)!}{(\alpha + \gamma + 1)!} \quad (2.3.8)
\]

\[
\mu_2 = \frac{(n-1)(n-2)(\alpha + 1)! (\alpha + 2)!}{(\alpha + \gamma + 1)! (\alpha + \gamma + 2)!} 
\]

\[
\mu_3 = \frac{(n-1)(n-2)(n-3)(\alpha + 1)! (\alpha + 2)! (\alpha + 3)!}{(\alpha + \gamma + 1)! (\alpha + \gamma + 2)! (\alpha + \gamma + 3)!} 
\]

\[
\mu_r = \frac{(n-1)(n-2)(n-3)(n-4)(\alpha + 1)! (\alpha + 2)! (\alpha + 3)! (\alpha + 4)!}{(\alpha + \gamma + 1)! (\alpha + \gamma + 2)! (\alpha + \gamma + 3)! (n-3)! (\alpha + \gamma + 4)!} 
\]

2.3.2.3 Factorial moments

Suppose \( \mu'_{k+r}(n, \alpha, \gamma) \) denotes the \( k \)th factorial moments of the proposed model then by definition, we have

\[
\mu'_{k+r}(n, \alpha, \gamma) = \sum_{x=1}^{n-1} x^k \frac{(n-1)!}{(n-x)! (x-1)!} \frac{(\alpha + 1)! y^{n-1-k}}{(\alpha + \gamma + 1)!^{n-1-k}} 
\]

\[
= \sum_{x=k}^{n} x^k \frac{(n-1)!}{(n-x)! (x-k)!} \frac{(\alpha + 1)! y^{n-1-k}}{(\alpha + \gamma + 1)!^{n-1-k}} 
\]

Taking \( x = x + k - 1 \) in the equation above, we get

\[
\mu'_r = \sum_{x=1}^{n-k+1} (x + k - 1)^k \frac{(n-1)!}{(n-k-1)! (x-1)!} \frac{(\alpha + 1)! y^{n-1-k}}{(\alpha + \gamma + 1)!^{n-1-k}} 
\]

A simplification of the equation above gives
A comparison of the above with (2.3.9) gives

$$
\mu'_j = \frac{(n-1)^j(\alpha+1)^j}{(\alpha+\gamma+1)^{j-1}} \left[ \mu_j(n-1, \alpha + k - l, \gamma) + (k-1) \right]
$$

Using (2.3.8), we obtain

$$
\mu'_j = \frac{(n-1)^j(\alpha+1)^j}{(\alpha+\gamma+1)^{j-2}} \left[ \frac{(n-k)(\alpha+k)}{(\alpha+\gamma+k)} + (k-1) \right]
$$

Which gives the first four factorial moments as

$$
\mu'_1 = \frac{(n-1)(\alpha+1)}{\alpha+\gamma+1}
$$

$$
\mu'_2 = \frac{(n-1)(n-2)(\alpha+1)(\alpha+2)}{(\alpha+\gamma+1)(\alpha+\gamma+2)} + \frac{(n-1)(\alpha+1)}{\alpha+\gamma+1}
$$

$$
\mu'_3 = \frac{(n-1)(n-2)(n-3)(\alpha+1)(\alpha+2)(\alpha+3)}{(\alpha+\gamma+1)(\alpha+\gamma+2)(\alpha+\gamma+3)} + \frac{2(n-1)(n-2)(\alpha+1)(\alpha+2)}{(\alpha+\gamma+1)(\alpha+\gamma+2)}
$$

$$
\mu'_4 = \frac{(n-1)(n-2)(n-3)(n-4)(\alpha+1)(\alpha+2)(\alpha+3)(\alpha+4)}{(\alpha+\gamma+1)(\alpha+\gamma+2)(\alpha+\gamma+3)(\alpha+\gamma+4)} + \frac{3(n-1)(n-2)(n-3)(\alpha+1)(\alpha+2)(\alpha+3)}{(\alpha+\gamma+1)(\alpha+\gamma+2)(\alpha+\gamma+3)}
$$

### 2.3.3 Relation of SBPED with other Distributions

**Theorem 2.3.1** Let $X$ be a size-biased Polya-Eggenberger variate with parameters pmf (2.3.4). If $\gamma \to \infty$ such that $\alpha \gamma^{-1} = \theta$ and $n \to \infty$ such that $n \theta = \lambda$, show that $X$ tends to size-biased Poisson distribution (1.2.13) with parameter $\lambda$.

**Proof:** The pmf of the proposed model can be put as
\[ P(x) = \frac{(n-1)\ldots(n-x+1)}{(x-1)!} \frac{(\alpha+1)\ldots(\alpha+2)(\alpha+x-1)\ldots(\alpha+x-n-1)}{(\alpha+\gamma+1)\ldots(\alpha+\gamma+n-1)} \]

Taking limit \( \gamma \to \infty \) such that \( \frac{\alpha}{\gamma} = \theta \), we get

\[ P(x) = \frac{(n-1)\ldots(n-x+1)}{(x-1)!} \frac{\theta^{x-1}}{(1+\theta)^{x-1}} \]

Proceeding to limit \( n \to \infty \) such that \( n\theta = \lambda \), the equation above reduces to size-biased Poisson distribution (1.2.13) with parameters \( \lambda \).

### 2.3.4 Estimation of the parameters of SBPED

In this section, we discuss different methods of estimation of the proposed model as:

#### 2.3.4.1 Moment method:

The size-biased Polya-Eggenberger distribution (2.3.4) has three parameters \( (n, \alpha, \gamma) \). The parameter \( n \) is where as the other two parameters are to be estimated. Let \( m'_1, m'_2 \) be the sample moments (about origin) of size-biased Polya-Eggenberger distribution. The method of moments consists in comparing the sample moments with the population moments of the distribution. The two equations thus obtained are

\[ m'_1 = \frac{(n-1)(\alpha+1)}{(\alpha+\gamma+1)} \]  \hspace{1cm} (2.3.11)

\[ m'_2 = \frac{(n-1)(n-2)(\alpha+1)(\alpha+2)}{(\alpha+\gamma+1)(\alpha+\gamma+2)} \]  \hspace{1cm} (2.3.12)

Dividing (2.3.12) by (2.3.11), we obtain after few steps

\[ (t+1) \frac{m'_2}{m'_1} = n\alpha - 2\alpha + 2n - 4 \]  \hspace{1cm} (2.3.13)

Where \( \alpha + \gamma + 1 = t \)  \hspace{1cm} (2.3.14)

Also from equation (2.3.11), we have

\[ \alpha = \frac{m'_1+1-n}{n-1} \]  \hspace{1cm} (2.3.15)
Again, from (2.3.13), we obtain

\[ \alpha = \frac{(t + 1) \frac{m'_r}{m'_l} + 4 - 2n}{n - 2} \]  

(2.3.16)

Eliminating \( \alpha \) between (2.3.15) and (2.3.16), we get

\[ \frac{m'_l + l - n}{n - 1} = \frac{(t + 1) \frac{m'_r}{m'_l} + 4 - 2n}{n - 2} \]

Which gives on simplifications

\[ t = \frac{m'_r(n - 1) - n + 1}{m'_r(n - 2) - n(m'_r + m'_s) - 2m'_r + m'_s} \]

Substituting the value of \( t \) from the equation above into equation (2.3.15), the value of \( \alpha \) can be obtained. After substituting the value of \( t \) and \( \alpha \) into (2.3.14) the value of \( \gamma \) can be obtained.

2.3.4.2 Using mean and first two cell frequencies:

Taking \( x = 1, 2 \) in the size-biased Polya-Eggenberger distribution (2.3.4) and then equate these probabilities with their corresponding relative frequencies \( \frac{f_i}{N} \), we obtain

\[ \frac{\gamma^{[n-1]}}{(\alpha + \gamma + 1)^{[n-1]}} = \frac{f_i}{N} \]  

(2.3.17)

\[ \frac{(\alpha + 1)\gamma^{[n-1]}}{(\alpha + \gamma + 1)^{[n-1]}} = \frac{f_2}{N} \]  

(2.3.18)

Where \( N = \sum f_i \). Dividing (2.3.18) by (2.3.17), we get

\[ \frac{f_2}{f_i} = \frac{(n - 1)(\alpha + 1)}{(\gamma + n - 2)} \]  

(2.3.19)

Eliminating \( \alpha \) between (2.3.19) and (2.3.11), we get the value of \( \gamma \) as
Now, the value of $\alpha$ can be easily obtained from (2.3.11).

### 2.3.4.3 The method of Maximum-likelihood.

The log likelihood function of size-biased Polya-Eggenberger distribution (2.3.4) is given by

$$
\log L = N \log \Gamma(n) - \sum f_i \log \Gamma(n-x+1) - \sum f_i \log \Gamma(x+1)
+ \sum f_i \log \Gamma(\alpha + x) - N \log \Gamma(\alpha + 1) + \sum f_i \log \Gamma(\gamma + n - x)
- N \log \Gamma(\gamma) + N \log \Gamma(\alpha + \gamma + 1) - \sum f_i \log \Gamma(\alpha + \gamma + n)
$$

Where $f_i$ is the observed frequency for the variate value 'x' and $N = \sum f_i$.

The proposed model has two unknown parameters viz. $(\alpha, \gamma)$. The log likelihood equations for estimating $\alpha$ and $\gamma$ are

1. $\frac{\partial \log L}{\partial \alpha} = 0$ which gives

   $$
   -N \sum_{k=1}^{\gamma-l} \frac{l}{\alpha + \gamma + k} + \sum f_i \sum_{k=1}^{\gamma-l} \frac{l}{\alpha + k} = 0
   $$

   (2.3.20)

2. $\frac{\partial \log L}{\partial \gamma} = 0$ which gives

   $$
   \sum f_i \sum_{k=0}^{\gamma-l} \frac{l}{\gamma + k} + \sum f_i \sum_{k=1}^{\gamma-l} \frac{l}{\alpha + \gamma + k} = 0
   $$

   (2.3.21)

The above equations are not simple to provide direct solution and thus an iterative method of solution such as Newton-Rampson, Fisher's scoring method etc. are required to solve these equations. We may solve the following system of equations

$$
(\hat{\theta} - \theta_0) \left[ \frac{\partial^2 \log L}{\partial \theta^2} \right]_{\theta_0} = \left[ \frac{-\partial \log L}{\partial \theta} \right]_{\theta_0} - \theta_0
$$
Where \( \theta = (\alpha, \gamma) \) is a parameter vector, the ML estimate of \( \theta \) and \( \theta_0 \) is the trial value of \( \theta \), which may be first obtained by equating the theoretical frequencies with the observed frequencies.

2.4. Negative Polya-Eggenberger Distribution

The Polya-Eggenberger distribution and its inverse, the negative Polya-Eggenberger distribution was introduced by Polya and Eggenberger (1923) through an urn model. The negative Polya-Eggenberger distribution is known by different names in the literature. Janardan and Schaeffer (1977) have called this as inverse Markov-Polya distribution. Another name is the beta-negative binomial distribution which was obtained analogously to the beta-binomial distribution by Kemp and Kemp (1956a). The generalized Waring distribution introduced by Irwin (1968) as an accident proneness-liability model is another parameterization of negative Polya-Eggenberger distribution which was subsequently applied by Irwin (1968, 1975a, b, c.) to data on accidents. Xekalaki (1981, 1983a, b, c, d) also studied this distribution in considerable detail under the name generalized Waring distribution. The negative Polya-Eggenberger distribution also belongs to Kemp and Kemp (1956a) generalized hypergeometric distribution.

In this section, we explored some interesting properties of negative Polya-Eggenberger distribution (NPED) and obtained it by compounding negative binomial distribution with beta distribution of I-kind. The distribution generates a number of univariate contagious or compound (or mixture of) distributions as its particular cases. The distribution is unimode, overdispersed and all of its positive and negative integer moments exist. The difference equation of the proposed model shows that it is a member of the Ord's family of distribution. Further, we propose different methods of estimation besides application of NPED to some data, available in the literature, has been given and its goodness of fit demonstrated (see; section 2.7).

2.4.1 Some models of NPED

Urn model
The negative Polya-Eggenberger distribution is related to Polya-Eggenberger distribution in the same way as negative binomial distribution is related to binomial distribution. The negative Polya-Eggenberger distribution is obtained by inverse Polya-Eggenberger sampling scheme. Consider an urn containing \( a \) white balls and \( b \) black balls. A ball is drawn at random and replaced back, together with \( c \) additional balls of the same colour before the next draw is made. Let \( N \) be a random variable denoting the number of trials to be made till the \( k \)th white ball is drawn. We wish to find the probability \( P(N = x) \). \( x = k, k + 1, \ldots \) This probability is given by

\[
P(N = x) = P((k-1) \text{ white balls in first (x-1) trials}) \times P( k^{\text{th}} \text{ white ball drawn at x}^{\text{th}} \text{ trials})
\]

\[
\binom{x-1}{k-1} \frac{a(a+c)\ldots(a+k-2c)b(b+c)\ldots(b+x-k-1c)}{(a+b)(a+b+c)\ldots(a+b+x-2c)} \frac{a+(k-1)c}{a+b+(x-1)c}
\]

Let \( Y \) be another random variable denoting the number of black balls preceding the \( k \)th white ball then \( y = x - k \Rightarrow x = y + k \) and the equation above reduces to

\[
P(N = y+k) = \binom{y+k-1}{k-1} \frac{a(a+c)\ldots(a+k-1c)b(b+c)\ldots(b+y-lc)}{(a+b)(a+b+c)\ldots(a+b+y+k-1c)} \quad y = 0, 1, 2, \ldots
\]

Which is the probability mass function of negative Polya-Eggenberger distribution

**The Mixture Model**

A certain mixture distribution arises when all (or some) parameters of a distribution vary according to some probability distribution called the mixing distribution. A well-known example of discrete-type mixture distribution is the negative-binomial distribution which can be obtained as a poisson mixture with gamma distribution.

Let \( X \) has a conditional negative-binomial distribution with parameter \( p \), that is, \( X \) has a conditional probability mass function (pmf)

\[
P(X / p) = P(X = x / p) = \binom{n+x-1}{x} p^x (1-p)^n
\]

(2.4.1)
for \( r=0,1,2, \ldots; \quad 0 < p < 1, \quad n > 0 \)

Now, suppose \( p \) is a continuous random variable with probability density function given by (2.3.5)

Bhattacharya (1966) showed that the conditional pmf of \( X \) is given by

\[
f(x) = P(X = x) = \int f(x|p) g(p) dp
\]

The equation above together with (2.4.1) and (2.3.5) gives

\[
P(X = x) = \binom{n+x-1}{x} \frac{\alpha(\alpha + 1) \ldots (\alpha + x - 1) \gamma(\gamma + 1) \ldots (\gamma + n - 1)}{\alpha + \gamma)(\alpha + \gamma + 1) \ldots (\alpha + \gamma + n + x - 1)}
\]  
(2.4.2)

Taking \( \alpha = a/c, \gamma = b/c \), the equation above reduces to the negative Polya-Eggenberger distribution with pmf

\[
P(X = x) = \binom{n+x-1}{x} \frac{a(a + c) \ldots (a + x - 1)c b(b + c) \ldots (b + n - 1)c}{(a + b)(a + b + c) \ldots (a + b + n + x - 1)c}
\]  
(2.4.3)

The proposed model (2.4.3) can be put into different forms for the mathematical convenience and to study some of its properties. The model (2.4.3) in terms of ascending factorials can be put as

\[
P(X = x) = \binom{n+x-1}{x} \frac{a^{(x-1)} b^{(n-1)}}{(a + b)(a + b + c) \ldots (a + b + n + x - 1)c}, \quad x = 0, 1, 2, \ldots
\]  
(2.4.4)

Where \( a^{(x-1)} = a(a + c) \ldots (a + x - 1 c) \)

Another form of (2.4.3) can be

\[
P(X = x) = \binom{n+x-1}{x} \frac{\alpha^{(x-1)} \gamma^{(n-1)}}{(a + \gamma)(a + \gamma + 1) \ldots (a + \gamma + n + x - 1)}, \quad x = 0, 1, 2, \ldots
\]  
(2.4.5)

Where \( \alpha^{(x-1)} = \alpha(\alpha + 1) \ldots (\alpha + x - 1) \) and \( \alpha = a/c, \gamma = b/c \). The model represented by (2.4.5) is the most convenient model, used through this section, for the mathematical computations.
Another form of (2.4.3) in terms of \( n, p = \frac{a}{t(a+b)} \), \( Q = 1 - p = \frac{b}{t(a+b)} \) and
\[
\delta = \frac{c}{(a+b)}
\]
can be
\[
P(X = x) = \binom{n+x-1}{x} \frac{\prod_{j=0}^{x-j} (p+j\delta)}{\prod_{j=0}^{y-1} (Q+j\delta)} \frac{\prod_{j=0}^{y+j} (1+j\delta)}{\prod_{j=0}^{n-1} (1+j\delta)}, \quad x = 0, 1, 2, \ldots
\]
(2.4.6)

A number of special cases can be deduced from the proposed model (2.4.3) by assigning different set of values to its parameters. For \( c = 0 \), (2.4.3) reduces to negative-binomial distribution, beta-negative binomial distribution is obtained when \( c = 1 \), negative hyper geometric distribution when \( c = -1 \), geometric series distribution when \( (c = 0, n = 1) \), Bernoulli-delta distribution (geometric) when \( c = 0 \) and replacing \( x \) with \( x - n \).

2.4.2 Structural properties of NPED

In this section, some of the interesting properties of the proposed model (NPED) has been explored which are described as follows:

2.4.2.1 Recurrence relation between probabilities

Expressing the pmf of the proposed model as
\[
P(X = x) = \frac{(n+x-1)!}{(n-1)! x!} \frac{\alpha^x}{\gamma^n} \frac{\gamma^n}{(\alpha+\gamma)^n x!}
\]
(2.4.7)

Taking \( x = x + 1 \) in the equation above and dividing the resulting equation by (2.4.5), we get the recurrence relation
\[
P(X = x + 1) = \left[ \frac{(n+x)}{(x+1)} \frac{(\alpha+x)}{(\alpha+\gamma+n+x)} \right] P(X = x)
\]
(2.4.8)

Which gives the difference equation of the proposed model as
\[
\Delta P_{x-1} = \frac{n\alpha-n-\alpha}{\gamma+1} P_x
\]
where
\[
\frac{n\alpha-n-\alpha}{\gamma+1} + \frac{1}{\gamma+1} \left[ \frac{n\alpha+\gamma}{\gamma+1} - 1 \right] x + \frac{x(x+1)}{\gamma+1}
\]
The difference equation above exhibits that the proposed model is a member of the Ord's family of distribution.

### 2.4.2 Unimodality

The proposed model is a unimodal by the following result of Holgate (1970):  

**Lemma.** If the mixing distribution is non-negative, continuous and unimodal, then the resulting distribution is unimodal.

The proposed model is a unimodal since the mixing distribution is a beta distribution of 1-kind which is unimodal. To show the unimodality of the distribution we prove the following theorem.

**Theorem 2.4.1.** The proposed model is a unimodal for all values of \((n,\alpha,\gamma)\) and the mode is at \(x=0\) if \(n\alpha<l\) and for \(n\alpha>l\) the mode is at some other point \(x=M\) such that

\[
\frac{n(\alpha-1)-(\alpha+\gamma)}{(\gamma+1)} < M < \frac{(n-1)(\alpha+1)}{(\gamma+1)}
\]

(2.4.9)

**Proof.** The recurrence relation (2.4.8) gives the ratio

\[
\frac{P(x+1)}{P(x)} = \frac{(n+x)}{(x+1)} \frac{(\alpha+x)}{(\alpha+\gamma+n+x)}
\]

(2.4.10)

Which is less than one, that is,

\[
\frac{P(x+1)}{P(x)} < 1 \quad \text{if} \quad n\alpha<l \quad \forall (n,\alpha,\gamma)>0
\]

Hence, for \(n\alpha<l\), the ratio \(\frac{P(x+1)}{P(x)}\) is a non-increasing function, therefore, the mode of the proposed model exists at \(x=0\). Suppose for \(n\alpha>l\) the mode exists at \(x=M\), then the ratio defined by (2.4.10) gives the two inequalities

\[
\frac{P(M+1)}{P(M)} = \frac{(n+M)}{(M+1)} \frac{(\alpha+M)}{(\alpha+\gamma+n+M)} < 1
\]

(2.4.11)

and

\[
\frac{P(M)}{P(M-1)} = \frac{n+M-1}{M} \frac{(\alpha+M-1)}{(\alpha+\gamma+n+M-1)} > 1
\]

(2.4.12)
By the inequality (2.4.11), we have

\[
\frac{\alpha-1-(\alpha+\gamma)}{(\gamma+1)} \leq M
\]  
(2.4.13)

and the inequality (2.4.12) gives

\[
M < \frac{(n-1)(\alpha+1)}{(\gamma+1)}
\]  
(2.4.14)

On combining the inequalities (2.4.13) and (2.4.14), the result (2.4.9) follows.

2.4.3 Mean and variance

Mean and variance of the proposed model can be easily obtained by using the properties of conditional mean and conditional variance as follows:

**Mean:**

By the conditional mean, we have

\[
\text{Mean} = E(X) = E[E(X|p)]
\]  
(2.4.15)

Where \( E(X|p) \) is the conditional expectation of \( X \) given \( p \) and for given \( p \) the random variable \( X \) follows negative binomial distribution (2.4.1) with mean and variance given by

\[
E(X|p) = np(1-p) \quad \text{and} \quad V(X|p) = np^2(1-p)
\]  
(2.4.16)

The equation (2.4.15) together with (2.4.16) and (2.4.2) gives mean of the proposed model as

\[
E(X) = \frac{n\alpha}{(\gamma-1)} \quad \gamma > 1.
\]

**Variance:**

Similarly, by the conditional variance, we have

\[
V(X) = E[V(X|p)] + V[E(X|p)]
\]  
(2.4.17)

Using (2.4.16), we obtain

\[
V(X) = nE[p^{-2}(1-p)] + n^2E[p^{-2}(1-p)^2] - n^2E(p^{-1}(1-p))^2
\]
Since \( p \) is varying as beta distribution (4.2.2), this reduces equation above to

\[
V(X) = \frac{n^2}{\beta(\gamma, \alpha)} \int_0^1 p^{\alpha+1}(1-p)^{\beta-1} dp + \frac{n^2}{\beta(\gamma, \alpha)} \int_0^1 p^{\alpha+1}(1-p)^{\beta-1} dp - \left( \frac{n^2}{\beta(\gamma, \alpha)} \right)^2
\]

By an application of beta integral, the equation above gives variance as

\[
V(X) = \frac{n\alpha}{(\gamma-1)} + \frac{n(n+1)\alpha(\alpha+1)}{(\gamma-1)(\gamma-2)} - \left( \frac{n\alpha}{(\gamma-1)} \right)^2 \text{ for } \gamma > 2
\]

2.4.2.4 Recurrence relation between moments

The recurrence relation (2.4.8) gives

\[
(1+x)^{\gamma+1}P_{\gamma+1}(n, \alpha, \gamma) = (1+x)^{\gamma} \frac{\alpha+x}{\alpha+\gamma+n+x} P_{\gamma}(n, \alpha, \gamma+1)
\]

Which subsequently reduces to

\[
(1+x)^{\gamma+1}P_{\gamma+1}(n, \alpha, \gamma) = (1+x)^{\gamma} \frac{n\alpha}{\gamma-1} P_{\gamma}(n+1, \alpha+1, \gamma-1), \quad \gamma > 1 \quad (2.4.18)
\]

Where \( P_{\gamma}(n+1, \alpha+1, \gamma-1) \) denotes the pmf of the proposed model with parameters \((n+1, \alpha+1, \gamma-1)\). Summing both sides of (2.4.18) over the values of \( \gamma \), we obtain the recurrence relation

\[
\mu_{\gamma+1}(n, \alpha, \gamma) = \frac{n\alpha}{(\gamma-1)} \sum_{r=1}^{\gamma} \mu_{r}(n+1, \alpha+1, \gamma-1) \quad (2.4.19)
\]

Where \( \mu_{r}(n+1, \alpha+1, \gamma-1) \) denotes the \( j \)th moment about origin of the proposed model with parameters \((n+1, \alpha+1, \gamma-1)\). The recurrence relation (2.4.19) gives the first four moments about origin as

\[
\mu_{1} = \frac{n\alpha}{(\gamma-1)}, \quad \gamma > 1
\]

\[
\mu_{2} = \frac{n\alpha}{(\gamma-1)} + \frac{n(n+1)\alpha(\alpha+1)}{(\gamma-1)(\gamma-2)}, \quad \gamma > 2
\]

\[
\mu_{3} = \frac{n\alpha}{(\gamma-1)} + \frac{3n(n+1)\alpha(\alpha+1)}{(\gamma-1)(\gamma-2)} + \frac{n(n+1)(n+2)\alpha(\alpha+1)(\alpha+2)}{(\gamma-1)(\gamma-2)(\gamma-3)}, \quad \gamma > 3
\]
Now, the central moments can be easily obtained from the moments about origin of the proposed model which are given by

\[
\mu_j = \frac{n\alpha}{(\gamma - 1)} \left[ l - \frac{3n\alpha}{(\gamma - 1)} + 2\left(\frac{n\alpha}{(\gamma - 1)}\right)^2 \right] + \frac{3n(n+1)\alpha(\alpha + 1)}{(\gamma - 1)(\gamma - 2)} (y-1)
\]

\[
\times \left[ l - \frac{n\alpha}{(\gamma - 1)} \right] + \frac{n(n+1)(n+2)\alpha(\alpha + 1)(\alpha + 2)}{(\gamma - 1)(\gamma - 2)(\gamma - 3)} (y-1)
\]

\[
\mu_j = \frac{n\alpha}{(\gamma - 1)} \left[ l - \frac{4n\alpha}{(\gamma - 1)} + 6\left(\frac{n\alpha}{(\gamma - 1)}\right)^2 \right] - 3\left(\frac{n\alpha}{(\gamma - 1)}\right)^3
\]

\[
+ \frac{n(n+1)\alpha(\alpha + 1)}{(\gamma - 1)(\gamma - 2)} \left[ l - \frac{12n\alpha}{(\gamma - 1)} + 6\left(\frac{n\alpha}{(\gamma - 1)}\right)^2 \right] (y-1)
\]

\[
+ \frac{n(n+1)(n+2)\alpha(\alpha + 1)(\alpha + 2)}{(\gamma - 1)(\gamma - 2)(\gamma - 3)} \left[ l - \frac{4n\alpha}{(\gamma - 1)} \right] (y-1)
\]

\[
+ \frac{n(n+1)(n+2)(n+3)\alpha(\alpha + 1)(\alpha + 2)(\alpha + 3)}{(\gamma - 1)(\gamma - 2)(\gamma - 3)(\gamma - 4)} (y-1)
\]

\[
\mu_j = \frac{n\alpha}{(\gamma - 1)^2} \left[ l - \frac{5n\alpha}{(\gamma - 1)} + 10\left(\frac{n\alpha}{(\gamma - 1)}\right)^2 - 10\left(\frac{n\alpha}{(\gamma - 1)}\right)^3 \right] (y-1)
\]

\[
+ \frac{n(n+1)\alpha(\alpha + 1)}{(\gamma - 1)^2} \left[ l - \frac{15n\alpha}{(\gamma - 1)} + 20\left(\frac{n\alpha}{(\gamma - 1)}\right)^2 - 15\left(\frac{n\alpha}{(\gamma - 1)}\right)^3 \right] (y-1)
\]

\[
+ \frac{n(n+1)(n+2)\alpha(\alpha + 1)(\alpha + 2)}{(\gamma - 1)^2} \left[ l - \frac{20n\alpha}{(\gamma - 1)} + 30\left(\frac{n\alpha}{(\gamma - 1)}\right)^2 - 20\left(\frac{n\alpha}{(\gamma - 1)}\right)^3 \right] (y-1)
\]

\[
+ \frac{n(n+1)(n+2)(n+3)\alpha(\alpha + 1)(\alpha + 2)(\alpha + 3)}{(\gamma - 1)^2} \left[ l - \frac{30n\alpha}{(\gamma - 1)} + 60\left(\frac{n\alpha}{(\gamma - 1)}\right)^2 - 30\left(\frac{n\alpha}{(\gamma - 1)}\right)^3 \right] (y-1)
\]

\[
+ \frac{n(n+1)(n+2)(n+3)(n+4)\alpha(\alpha + 1)(\alpha + 2)(\alpha + 3)(\alpha + 4)}{(\gamma - 1)^2} \left[ l - \frac{40n\alpha}{(\gamma - 1)} + 120\left(\frac{n\alpha}{(\gamma - 1)}\right)^2 - 60\left(\frac{n\alpha}{(\gamma - 1)}\right)^3 \right] (y-1)
\]

The proposed model is over dispersed for \( \gamma > 2 \) and its coefficient of variation is given by

\[
CV = 1 + \frac{(n+1)(\alpha + 1)}{(\gamma - 2)} - \frac{n\alpha}{(\gamma - 1)} (y-1)
\]

\[\gamma > 2\]

2.4.2.5 Probability generating function

Suppose \( G_x(u) \) denotes the probability generating function of the proposed model, we have

\[
G_x(u) = E(u^x) = \sum_{x=0}^{n} u^x \frac{(n+x-1)!}{(n-1)!} \frac{\alpha\gamma^x}{(\alpha + \gamma)^{n+x}} (y-1)
\]
Which on simplifications yields the probability generating function of the proposed model as

\[
G_X(u) = \frac{\gamma^n}{(\alpha + \gamma)^{n+1}} \sum_{i=0}^{n} \binom{n}{i} \left( \frac{\gamma}{\alpha + \gamma} \right)^i 
\]

(2.4.20)

Where \( \sum_{i=0}^{n} \binom{n}{i} \left( \frac{\gamma}{\alpha + \gamma} \right)^i \) is a Gaussian hypergeometric function [see: section 1.5.1].

**Remarks:** If we replace \( u \) by \((1+i)\) or \((1-i)\), the equation above yields the descending or ascending factorial moment generating functions of the proposed model, respectively.

### 2.4.2.6 Factorial moments.

The rth factorial moment about origin \( \mu_{(r)} \) of the proposed model is defined as

\[
\mu_{(r)} = E(X^{(r)}) = \sum_{x=0}^{\infty} x^r \frac{(n+x-1)! \cdot \alpha^{(x)} \cdot \gamma^n}{(n-1)! \cdot x! \cdot (\alpha + \gamma)^n} 
\]

Which reduces to

\[
\mu_{(r)} = \sum_{x=r}^{\infty} \frac{(n+x-1)!}{(n-1)!} \cdot \frac{\alpha^{(x)} \cdot \gamma^n}{(\alpha + \gamma)^n} 
\]

Taking \( x = x+1 \), the equation above yields the rth factorial moment of the proposed model as

\[
\mu_{(r)} = \frac{n^{(r)} \cdot \alpha^{(r)} \cdot (\gamma-r-1)!}{(\gamma-1)!} \quad (2.4.21)
\]

In particular, the first four factorial moments are

\[
\begin{align*}
\mu_{(1)} &= \frac{n \cdot \alpha}{(\gamma-1)} \quad \gamma > 1 \\
\mu_{(2)} &= \frac{n(n+1) \cdot \alpha(\alpha+1)}{(\gamma-1)(\gamma-2)} \quad \gamma > 2 \\
\mu_{(3)} &= \frac{n(n+1)(n+2) \cdot \alpha(\alpha+1)(\alpha+2)}{(\gamma-1)(\gamma-2)(\gamma-3)} \quad \gamma > 3
\end{align*}
\]
2.4.2.7 Negative integer factorial moments

The negative moments are useful in many problems of applied statistics, especially in life testing and in survey sampling, where ratio estimates are used. In this section, we obtained the expression for the $r$th negative integer ascending factorial moment of the proposed model in terms of Gaussian hypergeometric function. Suppose $\varphi'_{[r]}$ denotes the $r$th negative integer ascending factorial moment of the proposed mode then we have

$$\varphi'_{[r]} = \mathbb{E}[(x + 1)^r]^{-1} = \sum_{x=0}^{\infty} \frac{1}{(x+1)^r} \frac{(n+x-1)!}{(n-1)!} \frac{(\alpha+r)!}{(\alpha+\gamma)!} \frac{\gamma^n}{r!}$$

Where

$$\frac{1}{(x+1)^r} = \frac{x!}{(x+r)!} \frac{j^{(r)}}{j^{(r+1)}}$$

Using the result above in (2.4.22), we obtain

$$\varphi'_{[r]} = \frac{\gamma^n}{r!(\alpha+\gamma)!} \times_{F_{2}} [n, \alpha, 1; r+1, \alpha + \gamma + n; 1]$$

Where $\times_{F_{2}} [n, \alpha, 1; r+1, \alpha + \gamma + n; 1]$ is known as the generalized hypergeometric function [see; section 1.5.3]. The relation (2.4.23) gives the first four negative integer ascending factorial moments as

$$\varphi'_{[1]} = \frac{\gamma^n}{(\alpha+\gamma)!} \times_{F_{2}} [n, \alpha, 1; 2, \alpha + \gamma + n; 1]$$
$$\varphi'_{[2]} = \frac{\gamma^n}{2(\alpha+\gamma)!} \times_{F_{2}} [n, \alpha, 1; 3, \alpha + \gamma + n; 1]$$
$$\varphi'_{[3]} = \frac{\gamma^n}{6(\alpha+\gamma)!} \times_{F_{2}} [n, \alpha, 1; 4, \alpha + \gamma + n; 1]$$
$$\varphi'_{[4]} = \frac{\gamma^n}{24(\alpha+\gamma)!} \times_{F_{2}} [n, \alpha, 1; 5, \alpha + \gamma + n; 1]$$

2.4.3 Relation of NPED with other distributions.
Theorem 2.4.2. Let \( X \) be a negative Polya-Eggenberger variate with parameters \((n, \alpha, \gamma)\). If \( \gamma \to \infty \) such that \( \alpha \gamma^{-1} = \lambda \) and \( \lambda = \theta n^{-1} \) as \( n \to \infty \) then show that \( X \) tends to a Poisson distribution (1.2.7) with parameter \( \theta \).

**Proof:** Expressing the pmf of the proposed model as

\[
P(X=x) = \frac{\alpha(\alpha + 1) \ldots (\alpha + x - 1) \gamma (\gamma + 1) \ldots (\gamma + n - 1)}{(\alpha + \gamma)(\alpha + \gamma + 1) \ldots (\alpha + \gamma + n + x - 1)}
\]

Taking limit \( \gamma \to \infty \) such that \( \alpha \gamma^{-1} = \lambda \), the equation above gives

\[
P_{\gamma \to \infty}(X=x) = \left(1 + \frac{x-1}{n}\right) \left(1 + \frac{x-2}{n}\right) \ldots \left(1 + \frac{1}{n}\right) \frac{(n\lambda)^x}{x!(1 + \lambda)^{n+x}}
\]

Substituting \( \lambda = \frac{\theta}{n} \) and taking limit \( n \to \infty \), the equation above reduces to the Poisson distribution (1.2.7) with parameter \( \theta \).

**Theorem 2.4.3.** Let \( X \) be a negative Polya-Eggenberger variate with parameters \((n, \alpha, \gamma)\). Show that zero-truncated negative Polya-Eggenberger distribution tends to logarithmic series distribution (1.2.29).

**Proof:** The pmf of the zero-truncated negative Polya-Eggenberger distribution is

\[
P(X=x) = \binom{n+x-1}{x} \frac{\alpha^x \gamma^x}{(\alpha + \gamma)^{x+n}} \frac{(\alpha + \gamma)^{n}}{(\alpha + \gamma)^{|n|} - \gamma^{|n|}}
\]

Substituting \( \alpha \gamma^{-1} = \lambda \) and proceeding to limit \( \gamma \to \infty \), we get

\[
P_{\gamma \to \infty}(X=x) = \frac{n\Gamma(n+x)}{\Gamma(n+1)\Gamma(x+1)} \frac{\lambda^x}{(1 + \lambda)^{n+x}} \frac{1}{1-(1+\lambda)^{-n}}, \quad x=1,2,\ldots
\]

Taking \( \frac{\lambda}{1+\lambda} = t \) in the equation above, we get

\[
P(X=x) = \frac{n\Gamma(n+x)}{\Gamma(n+1)\Gamma(x+1)} \frac{t^x(1-t)^n}{1-(1-t)^n}
\]

Proceeding to the limit \( n \to 0 \), the equation above reduces to the logarithmic series distribution (1.2.29) with parameter \( t \).
2.4.3 Estimation of the parameters of NPEI

In this section, we propose different methods of estimation of NPEI which are discussed in the subsequent sub-sections.

2.4.3.1 Moment method:

Let \( m'_1, m'_2, \ldots, m'_j \) be the sample moments about origin and \( \mu_1, \mu_2, \mu_3 \) be the population moments about origin of the proposed model. Comparing the sample moments with the population moments of the proposed model, we obtain

\[
x = \frac{n\alpha}{(\gamma - 1)} \quad (2.4.25)
\]

\[
s^2 + \bar{x}^2 = \frac{n\alpha}{(\gamma - 1)} + \frac{n(n+1)\alpha(\alpha+1)}{(\gamma - 1)(\gamma - 2)}, \quad (2.4.26)
\]

Where \( s^2 \) is the sample variance given by \( s^2 = m'_2 - \bar{x}^2 \) and mean \( \bar{x} = m'_1 \)

Using (2.4.25) in (2.4.26), we get

\[
l_i (\gamma - 2) - \bar{x}^2(\gamma - 1) = n + \alpha + 1, \quad (2.4.27)
\]

where \( l_i = s^2 + \bar{x}^2 - \bar{x} \)

By comparing the third sample moment with its corresponding population moment, we get

\[
m'_3 = \frac{n\alpha}{(\gamma - 1)} + \frac{3n(n+1)\alpha(\alpha+1)}{(\gamma - 1)(\gamma - 2)} + \frac{n(n+1)(n+2)\alpha(\alpha+1)(\alpha+2)}{(\gamma - 1)(\gamma - 2)(\gamma - 3)} \quad (2.4.28)
\]

The equation (2.4.28) together with (2.4.25) and (2.4.26) gives

\[
\frac{(\gamma - 3)(m'_3 - \bar{x} - 3l_i)}{l_i} = \frac{(\gamma - 2)l_i}{x} = n + \alpha + 3 \quad (2.4.29)
\]

Eliminating \( n \) and \( \alpha \) between (2.4.29) and (2.4.27), we get the estimate of \( \gamma \) as

\[
\gamma = \frac{\bar{x}(3m'_3 - 7l_i) - 4l'_1^2 + x^2(l_i - 3)}{x(m'_3 - 3l_i) - 2l'_1^2 + x^2(l_i - 1)}
\]
Substituting the value of $n$ from (2.4.27) into (2.4.25), we get a quadratic equation in $\alpha$ as

$$\alpha^2 \bar{x} - \alpha \{ l \gamma - 2 - \bar{x}'(\gamma - 1) - \bar{x} + \bar{x}(\gamma - 1) = 0$$

Which can be solved for $\alpha$. After estimating $\alpha$, the value of $n$ can be obtained either from (2.4.29) or (2.4.25).

2.4.3.2 Using mean and first three cell frequencies:

Equating the first three probabilities of the proposed model with their corresponding relative frequencies $\frac{f_0}{N}, \frac{f_1}{N}, \frac{f_2}{N}$, we get

\[
\frac{\gamma^{n+1}}{(\alpha + \gamma)^{n+1}} = \frac{f_0}{N} \tag{2.4.30}
\]

\[
\frac{n\alpha \gamma^{n+1}}{(\alpha + \gamma)^{n+1}} = \frac{f_1}{N} \tag{2.4.31}
\]

\[
\frac{n(n+1)\alpha(\alpha + 1)\gamma^{n+1}}{2(\alpha + \gamma)^{n+2}} = \frac{f_2}{N} \tag{2.4.32}
\]

Dividing (2.4.31) by (2.4.30) and then using (2.4.25) in the resulting equation, we get

\[
\gamma = \frac{tf_1 + \bar{x}f_0}{xf_0} \tag{2.4.33}
\]

Where $\bar{x}$ is the mean of the proposed mode (2.4.5) and $t$ is given by

\[
\alpha + \gamma + n = t \tag{2.4.34}
\]

Using (2.4.25) and (2.4.34) in the equation obtained by dividing (2.4.32) with (2.4.31), we get

\[
\gamma = \frac{2(t+1)f_2 - (t+1-x)f_1}{(x-1)f_1} \tag{2.4.35}
\]

Eliminating $\gamma$ between (2.4.33) and (2.4.35), we obtain the estimate of $t$ as

\[
\hat{t} = \frac{2xf_0f_2}{f_1^2(x-1) - 2xf_0f_2 + xf_0f_1}
\]
The equation (2.4.33) together with the result obtained above gives the estimate of \( \gamma \) as

\[
\gamma = \frac{t f_1 + 2x f_0}{xf_0}
\]

Substituting the value of \( n \) from (2.4.34) into (2.4.25), we get a quadratic equation in \( \alpha \) as

\[
\alpha^2 - \alpha(t^* - \gamma) + x(\gamma - 1) = 0
\]

Which can be used to estimate \( \alpha \). The estimate of \( n \) can be obtained from (2.4.34) or (2.4.25).

### 2.4.23 Maximum likelihood method

The log likelihood function of the proposed model (2.4.5) is given by

\[
\log L = N \left[ \log(\gamma)^{n-1} - \log(\alpha + \gamma)^{n-1} \right] + \sum f_x \log(n)^{x-1} + \sum f_x \log(\alpha)^{x-1} - \sum f_x \log(\alpha + \gamma + n)^{x-1} - \sum f_x \log(x!)
\]

Where \( f_x \) is the observed frequency for the variate value \( x \) and \( N = \sum f_x \)

The method of maximum likelihood method of estimation gives the three likelihood equations for three unknown parameters as

1) \[
\frac{\partial \log L}{\partial \alpha} = 0 = \sum f_x \sum_{k=0}^{n-1} \frac{l}{\alpha + \gamma + k} - \sum f_x \sum_{k=0}^{n-1} \frac{l}{\alpha + k}
\]

2) \[
\frac{\partial \log L}{\partial \gamma} = 0 = N \sum_{k=0}^{n-1} \frac{l}{\gamma + k} - \sum f_x \sum_{k=0}^{n-1} \frac{l}{\alpha + \gamma + k}
\]

3) \[
\frac{\partial \log L}{\partial n} = 0 = N \frac{\partial \log(\gamma)^{n-1}}{\partial n} - N \frac{\partial \log(\alpha + \gamma)^{n-1}}{\partial n} + \sum f_x \frac{\partial \log(n)^{x-1}}{\partial n} - \sum f_x \frac{\partial \log(\alpha + \gamma + n)^{x-1}}{\partial n} - \sum f_x \frac{\partial \log(x!)}{\partial n}
\]

Converting the likelihood equation above in terms of gamma functions, we get

\[
N \frac{\partial \log \Gamma(\gamma + n)}{\partial n} - N \frac{\partial \log \Gamma(\alpha + \gamma + n)}{\partial n} + \sum f_x \frac{\partial \log \Gamma(n + x)}{\partial n} - \sum f_x \frac{\partial \log \Gamma(n)}{\partial n}
\]
The differentiation of the equation above is not straightforward and can be solved with the help of the following recurrence relation: (see pages 6-8 Johnson, Kotz and Kemp 1993 for details)

\[ \psi(x+n) = \psi(x) + \sum_{j=1}^{n} (x+j-1)^{-1}, \quad n = 1, 2, \ldots \quad (2.4.37) \]

Where \( \psi(x) = \frac{d}{dx} \log \Gamma(x) / \Gamma(x) \) is called digamma function. A good approximation for \( \psi(x) \) is

\[ \psi(x) = \log \Gamma(x) - \frac{1}{2x} \]

The relation (2.4.37) together with the above result gives

\[ \psi(x+n) = \log \Gamma(x) - \frac{1}{2x} + \sum_{j=1}^{n} (x+j-1)^{-1}, \quad n = 1, 2, \ldots \quad (2.4.38) \]

By an application of (2.4.38), the equation (2.4.36) gives the third likelihood equation as

\[
N \left[ \log \Gamma(\gamma + n) + \sum_{k=1}^{\gamma} (n+k-1)^{-1} \right] - 2N \left[ \log \Gamma(\alpha + \gamma + n) + \sum_{k=1}^{\alpha + \gamma} (n+k-1)^{-1} \right] \\
+ \sum_{k=1}^{\alpha + \gamma} (n+k-1)^{-1} \right] + \sum_{k=1}^{\alpha + \gamma} (n+k-1)^{-1} \right] \\
- N \log \Gamma(n) + \sum_{k=1}^{\alpha + \gamma} (n+k-1)^{-1} \right] = 0
\]

The three likelihood equations are not simple to provide direct solution, however, different iterative procedures such as Fisher’s scoring method, Newton-Rampson method etc. can be employed to solve these equations. We may solve the following system of equations

\[
(\hat{\theta} - \theta_0) \left[ \frac{\partial^2 \log L}{\partial \theta^2} \right]_{\theta_0} = - \frac{\partial \log L}{\partial \theta} \bigg|_{\theta_0}
\]
Where $\theta = (n, \alpha, \gamma)$ is a parameter vector, the ML estimate of $\theta$ and $\theta_0$ is the trial value of $\theta$ which may be first obtained by equating the theoretical frequencies with the observed frequencies.

2.5 Size-biased negative Polya-Eggenberger distribution

Some times we come across the situations when a value with zero frequency either is ignored or is impossible to observe. This leads to truncation of the distribution at zero-variate value. The distribution thus truncated is not straightforward and is difficult to study, especially, in case of estimation and thus renders its common use. Here, the size-biased distributions play a very important role and are helpful because of their simplicity and gives best fit as compared to zero-truncated distributions.

In this section, we are concerned with size-biased negative Polya-Eggenberger distribution (SBNPED). As discussed in section 2.3, weighted distributions arise when the observations generated from a stochastic process are not given equal chance of being recorded; instead, they are recorded according to some weighted function. The size-biased negative Polya-Eggenberger distribution has not been introduced so far. In this section, we introduce and study the various aspects of this distribution where as its goodness of fit test has been demonstrated with the help of some data sets in section 2.7 of this chapter.

2.5.1 Some Models of SBNPED

If $X$ is a negative Polya-Eggenberger variate with probability mass function given by (2.4.53). A size-biased negative Polya-Eggenberger distribution is obtained when the weights of NPED (2.4.5) is taken proportional to the variate value $X$. This is equivalent to taking $\alpha = 1$ in (2.3.2) with $\mu = E(X)$, the mean of the Polya-Eggenberger variate $X$ given by (2.4.16), we obtain from (2.3.2),

$$p(x) = \frac{x^{(n+1)\alpha} (\gamma - 1)^{\alpha \gamma / n}}{n \alpha \gamma / (\gamma - 1)}$$

Which gives on simplifications
This is called the size-biased negative Polya-Eggenberger distribution (SBNPED) with parameters \((n, \alpha, \gamma)\).

The Mixture Model:

The size-biased negative Polya-Eggenberger distribution can also be obtained by compounding the negative binomial distribution (1.2.16) through the values of parameter \(p\) with beta distribution (2.3.5), we obtain the pmf of SBNPED as

\[
P(X = x) = \binom{n + x - 1}{x-1} \frac{1}{\beta(x, \gamma)} \left( \frac{1}{x!} \right) \frac{1}{(\alpha + \gamma)^{n + x}} \int_0^1 p^{x-2} (1-p)^{n+x-1} dp
\]

Which gives on simplification

\[
P(X = x) = \binom{n + x - 1}{x-1} \frac{\gamma^{x-1} \alpha^{n-x}}{\alpha + \gamma)^{n + x}} \quad x = 1, 2, \ldots
\]

Replacing \(a\) with \(a+1\) and \(y\) with \(y-1\), the equation above coincides with the size-biased negative Polya-Eggenberger distribution (2.5.1).

2.5.2 Structural properties of SBNPED

2.5.2.1 Recurrence relation between probabilities

The probability mass function of size-biased negative Polya-Eggenberger distribution (2.5.1) can be expressed as

\[
P(X = x) = \frac{(n+x-1)!}{n! (x-1)!} \frac{(\alpha+1)^{x-1} (\gamma-1)^{n-x}}{(\alpha+y)^{n+x}}
\]

Taking \(x = x+1\), we get

\[
P(X = x+1) = \frac{(n+x)!}{n! x!} \frac{(\alpha+1)^{x} (\gamma-1)^{n-x+1}}{(\alpha+y)^{n+x+1}}
\]

Divide (2.5.3) by (2.5.2), we obtain the recurrence relation

\[
P(X = x+1) = \left( \frac{(n+x)}{x} \frac{(\alpha+x)}{(\alpha+y+n+x)} \right) P(X = x)
\]
Since negative Polya-Eggenberger distribution belongs to Ord's family of distribution so is the size-biased negative Polya-Eggenberger distribution.

2.5.2.2 Recurrence relation between moments

Suppose \( \mu'_r(n, \alpha, \gamma) \) is the \( r \)-th moment about origin of the proposed model (2.5.1), we have

\[
\mu'_r(n, \alpha, \gamma) = \sum_{x=0}^{n} x^r \frac{(n+x-1)!}{n! (x-1)!} \frac{(\alpha+1)^{x-1} (\gamma-1)^{n-x}}{(\alpha+\gamma)^{n+x}}
\]  

(2.5.5)

Taking \( x = x+1 \), we get

\[
\mu'_r(n, \alpha, \gamma) = \sum_{x=0}^{n} \left(1+x\right)^r \frac{(n+x)!}{n! x!} \frac{(\alpha+1)^x (\gamma-1)^{n-x}}{(\alpha+\gamma)^{n+x+1}}
\]

\[
= \sum_{x=0}^{n} \left[ \sum_{j=0}^{r} \binom{r}{j} \frac{(n+1+x-1)!}{(n+1-x)!} \frac{(\alpha+1)^j (\gamma-1)^{n-x}}{(\alpha+\gamma)^{n+x+j}} \right]
\]

Using (2.5.5), we obtain the recurrence relation

\[
\mu'_r(n, \alpha, \gamma) = \sum_{j=0}^{r} \binom{r}{j} \mu'_j(n+1, \alpha+1, \gamma-1)
\]  

(2.5.5)

Where \( \mu'_j(n+1, \alpha+1, \gamma-1) \) is the \( j \)-th moment about origin of the negative Polya-Eggenberger distribution (2.4.5) with parameters \( (n+1, \alpha+1, \gamma-1) \). This recurrence relation (2.5.6) can be used to determine the moments of the proposed model when the moments of negative Polya-Eggenberger distribution are known. However, the moments of size-biased negative Polya-Eggenberger distribution can be obtained directly from its moment generating function (2.5.7) or factorial moment relation (2.5.8).

2.5.2.3 Moment generating function

Moment generating function of the proposed model (2.5.1) can be obtained as:

\[
M_X(t) = E(e^{\alpha t})
\]

\[
= \sum_{x=1}^{\infty} \frac{e^{\alpha t} \frac{(n+x-1)!}{n! (x-1)!} \frac{(\alpha+1)^{x-1} (\gamma-1)^{n-x}}{(\alpha+\gamma)^{n+x}}}{(\alpha+\gamma)^{n+x}}
\]
Taking $x = x + 1$, we get

$$M_x(t) = \frac{(\gamma-1)^{n+1}}{(\alpha + \gamma)^{n+1}} \sum_{i=0}^{n+1} \frac{(n+1)!}{(\alpha + \gamma + n + 1)!} \left( \frac{t^i}{i!} \right)$$

$$= \frac{(\gamma-1)^{n+1}}{(\alpha + \gamma)^{n+1}} \left[ F \{ n + 1, \alpha + 1; \alpha + \gamma + n + 1, t \} \right]$$

(2.5.7)

Which is the required moment generating function of the proposed model (2.5.1).

This gives the first four moments about origin of SBNPED as

$$\mu_1 = 1 + \frac{(n+1)(\alpha+1)}{(\gamma-2)}$$

$$\mu_2 = 1 + \frac{3(n+1)(\alpha+1)}{(\gamma-2)} + \frac{(\alpha+1)(\alpha+2)(n+1)(n+2)}{(\gamma-2)(\gamma-3)}$$

$$\mu_3 = 1 + \frac{7(n+1)(\alpha+1)}{(\gamma-2)} + \frac{6(n+1)(n+2)(\alpha+1)(\alpha+2)}{(\gamma-2)(\gamma-3)(\gamma-4)}$$

$$= \left[ \frac{(\alpha+1)(\alpha+2)(\alpha+3)(n+1)(n+2)(n+3)}{(\gamma-2)(\gamma-3)(\gamma-4)} \right]$$

$$\mu_4 = 1 + \frac{15(\alpha+1)(n+1)}{(\gamma-2)} + \frac{25(n+1)(n+2)(\alpha+1)(\alpha+2)}{(\gamma-2)(\gamma-3)(\gamma-4)}$$

$$+ \frac{10(n+1)(n+2)(n+3)(\alpha+1)(\alpha+2)(\alpha+3)}{(\gamma-2)(\gamma-3)(\gamma-4)(\gamma-5)}$$

After obtaining the moments about origin, central moments can be easily obtained from the moments about origin, thus variance and higher central moments are

$$\mu_2' = \frac{(n+1)(\alpha+1)}{(\gamma-2)} + \frac{(\alpha+1)(\alpha+2)(n+1)(n+2)}{(\gamma-2)(\gamma-3)} \left[ \frac{(n+1)(\alpha+1)}{(\gamma-2)} \right]^2$$

$$\mu_3' = \frac{(n+1)(\alpha+1)}{(\gamma-2)} + \frac{3(n+1)(n+2)(\alpha+1)(\alpha+2)}{(\gamma-2)(\gamma-3)}$$

$$+ \frac{(n+1)(n+2)(n+3)(\alpha+1)(\alpha+2)(\alpha+3)}{(\gamma-2)(\gamma-3)(\gamma-4)}$$

$$- 3 \left[ \frac{(n+1)(\alpha+1)}{(\gamma-2)} \right]^2 + \left[ \frac{(\alpha+2)(n+2)}{(\gamma-3)} \right] + \left[ \frac{(n+1)(\alpha+1)}{(\gamma-2)} \right]^3$$
\[
\mu_j = \frac{(n+1)(\alpha+1)}{(\gamma - 1)} + \frac{7(n+1)(n+2)(\alpha+1)(\alpha+2)}{(\gamma - 2)(\gamma - 3)} + \frac{6(n+1)(n+2)(n+3)(\alpha+1)(\alpha+2)(\alpha+3)}{(\gamma - 2)(\gamma - 3)(\gamma - 4)} + \frac{(n+1)(n+2)(n+3)(n+4)(\alpha+1)(\alpha+2)(\alpha+3)(\alpha+4)}{(\gamma - 2)(\gamma - 3)(\gamma - 4)(\gamma - 5)}
\]

\[
-4 \left\{ \frac{(n+1)(\alpha+1)}{(\gamma - 2)} \right\} \left\{ 3 + \frac{3(n+2)(\alpha+2)}{(\gamma - 3)} + \frac{(n+2)(n+3)(\alpha+2)(\alpha+3)}{(\gamma - 3)(\gamma - 4)} \right\} + 6 \left\{ \frac{(\alpha+2)(n+2)}{(\gamma - 3)} \right\} \left\{ \frac{(n+1)(\alpha+1)}{(\gamma - 2)} \right\} - 3 \left\{ \frac{(n+1)(\alpha+1)}{(\gamma - 2)} \right\}
\]

### 2.5.2.4 Factorial moments

The \(r\)th factorial moment \(\mu_{r}(n,\alpha,\gamma)\) of the proposed model \((2.5.1)\) is obtained as

\[
\mu_{r}(n,\alpha,\gamma) = E[X^r] = \frac{\Gamma(n+k)}{n!\Gamma(n+r)} (\alpha+1)^{n+k} \left( \frac{1}{\alpha+\gamma} \right)^{n+r}
\]

Taking \(x = x+1\), we get

\[
\mu_{r}(n,\alpha,\gamma) = \sum_{x=0}^{n+k} \frac{(n+x+k-1)! (\alpha+1)^{(n+k-1)} (\gamma-1)^{(n+x-1)} \Gamma(n+r)}{n! (x-1)! (\alpha+\gamma)^{(n+x)}}
\]

\[
= (n+1)^{(k)} \sum_{x=0}^{n+k} \frac{(n+k+x-1)! (\alpha+1)^{(n+k-1)} (\gamma-1)^{(n+x-1)} \Gamma(n+r)}{(n+k) (x-1)! (\alpha+\gamma)^{(n+x)}}
\]

\[
+ k(n+1)^{(k-1)} \sum_{x=0}^{n+k} \frac{(n+k+x-1)! (\alpha+1)^{(n+k-1)} (\gamma-1)^{(n+k-1)} \Gamma(n+r)}{(n+k-1)! x! (\alpha+\gamma)^{(n+k+x)}}
\]

Which gives on simplification the relation for the \(r\)th factorial moment as

\[
\mu_{r}(n,\alpha,\gamma) = \frac{(\alpha+1)^{(k-1)} (n+1)^{(k-1)} \Gamma(n+k) (\alpha+k)(n+k) + k(\gamma-k-1)}{(\gamma-1)^{(k)}} (2.5.8)
\]
We can use this expression to obtain the factorial moments of size-biased negative Polya-Eggenberger distribution. The first four factorial moments are obtained by taking $k = 1, 2, 3, 4$ in (2.5.8). Thus,

$$
\mu'_1 = 1 + \frac{(n+1)(\alpha+1)}{(\gamma-2)}
$$

$$
\mu'_2 = \frac{(\alpha+1)(\alpha+2)(n+1)(n+2)}{(\gamma-2)(\gamma-3)} + \frac{2(n+1)(\alpha+1)}{(\gamma-2)}
$$

$$
\mu'_3 = \frac{(\alpha+1)(\alpha+2)(\alpha+3)(n+1)(n+2)(n+3)}{(\gamma-2)(\gamma-3)(\gamma-4)} + \frac{3(n+1)(n+2)(\alpha+1)(\alpha+2)}{(\gamma-2)(\gamma-3)}
$$

$$
\mu'_4 = \frac{(\alpha+1)(\alpha+2)(\alpha+3)(\alpha+4)(n+1)(n+2)(n+3)(n+4)}{(\gamma-2)(\gamma-3)(\gamma-4)(\gamma-5)} + \frac{4(n+1)(n+2)(n+3)(\alpha+1)(\alpha+2)(\alpha+3)}{(\gamma-2)(\gamma-3)(\gamma-4)}
$$

2.5.3 Relation of SBNPED with other distribution.

Theorem 2.5.1. If $X$ is a size-biased negative Polya-Eggenberger distribution with parameters $(n+1, \alpha+1, \gamma-1)$. If $\gamma \to \infty$ such that $\alpha \gamma^{-1} = \lambda$ and $\lambda n = \theta$ as $n \to \infty$, show that $X$ approaches to size-biased Poisson distribution (1.2.13) with parameter $\theta$.

Proof. The pmf of the size-biased negative Polya-Eggenberger distribution can be written as

$$
P(X=x) = \frac{\frac{(n+x-1)(n+x-2)...(n+1)\alpha^{x-i}(\gamma-1)^{i}}{(x-1)!}}{(\alpha+\gamma)^{n+x}}
$$

(2.5.9)

Now, we can write

$$
\frac{(\alpha+1)^{l+1}(\gamma-1)^{l+1}}{(\alpha+\gamma)^{l+n+1}}
$$

$$
= \frac{(\alpha \gamma^{-1} + \gamma^{-1})...(\alpha \gamma^{-1} +(x-1)\gamma^{-1})(1-\gamma^{-1})(1+\gamma^{-1})...(1+(n-1)\gamma^{-1})}{(\alpha \gamma^{-1} + 1)(\alpha \gamma^{-1} + \alpha + 1 + \gamma^{-1})...(\alpha \gamma^{-1} + 1 + (n+x-1)\gamma^{-1})}
$$

Taking limit $\gamma \to \infty$ in such a way so that $\alpha \gamma^{-1} = \lambda$, we obtain
Substituting this value in (2.5.9), we get

\[ P(X = x) = \frac{(n\lambda)^{x-1}}{(x-1)!1^{x-1+n}} \left(1 + \frac{x-1}{n}, \frac{x-2}{n}, \ldots, \frac{x-n}{n}\right) \]

Substituting \( n\lambda = \theta \) and taking limit \( n \to \infty \), we obtain size-biased Poisson distribution (1.2.13) with parameter \( \theta \).

2.5.4 Estimation of the parameters of SBNPED

In this section, we obtained different methods of estimation of the proposed model which are discussed below.

2.5.4.1 Moment method

Equating the sample moments denoted by \( m'_1, m'_2, m'_3 \) with their corresponding population moments \( \mu'_1, \mu'_2, \mu'_3 \) of the size-biased negative Polya-Eggenberger distribution, we obtain

\[
\overline{x} - 1 = \frac{(n+1)(\alpha+1)}{\gamma - 2} \quad (2.5.10)
\]

\[
s^2 + \overline{x}^2 - 1 = \frac{3(n+1)(\alpha+1)}{(\gamma - 2)} + \frac{(n+1)(n+2)(\alpha+1)(\alpha+2)}{(\gamma - 2)(\gamma - 3)} \quad (2.5.11)
\]

where \( s^2 = m'_2 - \overline{x}^2 \) is the sample variance.

Using (2.5.10) in (2.5.11), we obtain

\[
\frac{l_1(\gamma - 3) - (\overline{x} - 1)^2(\gamma - 2)}{(\overline{x} - 1)} = n + \alpha + 3 \quad (2.5.12)
\]

where \( l_1 = s^2 + \overline{x}^2 - 3\overline{x} + 2 \)

\[
m'_3 = 1 + \frac{7(n+1)(\alpha+1)}{(\gamma - 2)} + \frac{6(n+1)(n+2)(\alpha+1)(\alpha+2)}{(\gamma - 2)(\gamma - 3)} + \frac{(n+1)(n+2)(n+3)(\alpha+1)(\alpha+2)(\alpha+3)}{(\gamma - 2)(\gamma - 3)(\gamma - 4)} \quad (2.5.13)
\]

The equation (2.5.13) together with (2.5.10) and (2.5.11) gives
Eliminating \( n \) and \( \alpha \) between (2.5.12) and (2.5.14), we get after few steps

\[
\gamma = \frac{2(x-1)[2(m_j - 7x - 6l_j + 6) - 6l_j + x]}{(x-1)[m_j - 7x - 6l_j + 6 + (x-1)l_j + 2l_j^2]} \quad (2.5.15)
\]

Now, eliminating \( n \) between (2.5.12) and (2.5.10) and in the resulting equation using (2.5.15), we get obtain a quadratic equation in \( \alpha \) as

\[
\alpha^2(x-1) - \alpha[l_j(\gamma - 3) - (x-1)^2(\gamma - 2) - 2(x-1)]
\]

\[-[l_j(\gamma - 3) - 2(x-1)^2(\gamma - 2) - 2(x-1)] = 0
\]

Which can be solved for \( \alpha \). Now, the estimate of \( n \) can be obtained either from (2.5.12) or (2.5.10).

2.5.4.2 Using mean and first three cell frequencies

Taking \( x=1.2.3 \) in the pmf of the size-biased negative Polya-Eggenberger distribution (2.5.1) and then equating these probabilities with their corresponding relative frequencies \( f_1, f_2, f_3 \), where \( N = \sum f_i \), we get the three equations as

\[
\frac{(\gamma - 1)^{m+1}}{(\alpha + \gamma)^{m+1}} = \frac{f_1}{N} \quad (2.5.16)
\]

\[
\frac{(n+1)(\alpha + 1)(\gamma - 1)^{m+1}}{(\alpha + \gamma)^{m+1}} = \frac{f_2}{N} \quad (2.5.17)
\]

\[
\frac{(n+1)(n+2)(\alpha + 1)(\alpha + 2)(\gamma - 1)^{m+1}}{2(\alpha + \gamma)^{m+1}} = \frac{f_3}{N} \quad (2.5.18)
\]

Divide (2.5.17) by (2.5.16) and then using (2.5.10), we get

\[
\gamma = \frac{(1+1)f_2 + 2(x-1)f_1}{(x-1)f_1} \quad (2.5.19)
\]

Where \( \alpha + \gamma + n = t \)
Now, dividing (2.5.18) by (2.5.17) and then use of (2.5.10) in the resulting equation gives

\[ \gamma = \frac{2(t+2)f_3 - (t+5-2t)f_2}{(x-2)f_2} \quad (2.5.21) \]

Eliminate \( \gamma \) between (2.5.19) and (2.5.21), we get

\[ \hat{t} = \frac{(x-1)f_1 f + (2x^2 - 7x + 5)f_1 f_2 - (x-2)f_2^2 - 2(x-1)f_1}{(x-2)f_2^2 - 2(x-1)f_1 f_3 + (x-1)f_1 f_2} \quad (2.5.22) \]

Substituting the value of \( \hat{t} \) in (2.5.19), we obtain

\[ \hat{\gamma} = \frac{(t+1)f_2 + 2(x-1)f_1}{(x-1)f_1} \quad (2.5.23) \]

Eliminate \( n \) between (2.5.20) and (2.5.10), we get a quadratic equation in \( \alpha \) as

\[ \alpha^2 - (1-\gamma + 1)n \alpha - \gamma - 2 + 1 - \hat{t} = 0 \]

Which can be solved for \( \alpha \). Finally, the value of \( n \) can be estimated from (2.5.20).

2.5.4.3 Maximum likelihood method

The log likelihood function of size-biased negative Polya-Eggenberger distribution (2.5.1) is given by

\[ \log L = N \left[ \log(y-1)^{n+1} - \log(\alpha + \gamma y)^n \right] + \sum f_x \log(n+1)^{x-1} \]

\[ + \sum f_x \log(\alpha + 1)^{x-1} - \sum f_x \log(\alpha + \gamma + n)^{x} \quad (2.5.24) \]

Since the proposed model has three unknown parameters viz \((\alpha, \gamma, n)\), therefore, the three likelihood equations are

1). \( \frac{\partial \log L}{\partial \alpha} = 0 \) gives

\[ -N \frac{\partial \log(\alpha + \gamma y)^n}{\partial \alpha} + \sum f_x \frac{\partial \log(\alpha + n+1)^{x-1}}{\partial \alpha} - \sum f_x \frac{\partial \log(\alpha + \gamma + n)^{x}}{\partial \alpha} = 0 \]

Which on simplification gives the likelihood equation

\[ \sum f_x \sum_{k=0}^{n+y-1} \frac{1}{\alpha + \gamma + k} - \sum f_x \sum_{k=1}^{y-1} \frac{1}{\alpha + k} = 0 \quad (2.5.25) \]
This on simplification gives the second likelihood equation

\[ N \sum_{k=1}^{n} \frac{1}{\gamma + k} - \sum f_i \sum_{k=1}^{\gamma + n} \frac{1}{\alpha + \gamma + k} = 0 \]  

(2.5.26)

3) \[ \frac{\partial \log L}{\partial \alpha} = 0 \] gives

\[ N \frac{\partial \log (\gamma - 1)^{(n+1)}}{\partial \alpha} = -N \frac{\partial \log (\alpha + \gamma)^{n+1}}{\partial \alpha} \]

\[ + \sum f_i \frac{\partial \log (n+1)^{(n+1)}}{\partial \alpha} - \sum f_i \frac{\partial \log (\alpha + \gamma + n)^{(n+1)}}{\partial \alpha} = 0 \]

Using \( \log (\gamma - 1)^{(n+1)} = \log \Gamma (\gamma + n) - \log \Gamma (\gamma - 1) \), we obtain

\[ N \frac{\partial \log \Gamma (n+\gamma)}{\partial \alpha} - N \frac{\partial \log \Gamma (n+\alpha + \gamma)}{\partial \alpha} + \sum f_i \frac{\partial \log \Gamma (n+x)}{\partial \alpha} - \sum f_i \frac{\partial \log \Gamma (n+1+x)}{\partial \alpha} \]

\[ - \sum f_i \frac{\partial \log \Gamma (n+\alpha + \gamma + x)}{\partial \alpha} + \sum f_i \frac{\partial \log \Gamma (n+\alpha + \gamma)}{\partial \alpha} = 0 \]  

(2.5.27)

The above equation is not straightforward for differentiation; however, by using (2.4.38), we obtain the third likelihood equation

\[ N \sum_{j=1}^{x} (n+j-1)^{-1} \sum j \sum_{j=1}^{x} (n+j-1)^{-1} - \sum f_i \sum_{j=1}^{x} (n+j-1)^{-1} - \sum f_i \sum_{j=1}^{x} (n+j-1)^{-1} - N \frac{N}{n} = 0 \]  

(2.5.28)

Where \( f \) is the observed frequency for the variate value \( x \) and \( N = \sum f_i \).

The equations given above are not simple to provide direct solution and thus an iterative method of solution such as Fisher's scoring method are required to solve these equations. We may solve the following system of equations

\[ (\hat{\theta} - \theta_0) \left[ \frac{\partial^2 \log L}{\partial \theta^2} \right]_{\theta_0}^{-1} = -\frac{\partial \log L}{\partial \theta} \]

Where \( \theta = (n, \alpha, \gamma) \) is a parameter vector, the ML estimate of \( \theta \) and \( \theta_0 \) is the trial value of \( \theta \) which may be first obtained by equating the theoretical frequencies with observed frequencies.
2.6. A New Contagious Distribution Analogous to Negative Polya-Eggenberger Distribution (CDANPED)

In this section, we propose a new contagious distribution analogous to negative Polya-Eggenberger distribution (CDANPED) that has an application in queuing theory, accidental statistics, industry etc. The first application has been described in section 2.6.1 where as its other applications have been demonstrated with the help of three data sets. (see; section 2.7). The proposed model is unimode and all of its positive integer moments exist. These properties along with some other interesting properties have been explored in this section. The relation of the proposed model (CDANPED) with some other distributions has also been discussed.

2.6.1 The proposed model (CDANPED)

In the theory of queuing, suppose we have a single queue beginning with $r$ customers. First, If we assume that the random arrival time of a customer is at a constant rate $\lambda$ with a constant time $\beta$ occupied in serving each customer, then it has been shown by Borel (1942) and Tanner (1953) that the probability distribution of total number of customers served before the queue vanishes is

$$P(X = x) = \frac{r}{(x-r)!} \alpha^{x-r} e^{-\lambda \beta}, \quad x = r, r+1, \ldots \quad (2.6.1)$$

This is known as Borel-Tanner distribution. If $\beta$, the time occupied in serving each customer, is not a constant but varies as negative exponential distribution with poisson random arrival time of a customer at a constant rate $\lambda$ then it has been shown by Frank A. Haight (1961) that the probability distribution of total number of customers served before the queue vanishes is a distribution analogous to Borel-Tanner distribution with probability mass function given by

$$P(X = x) = \frac{\alpha^{x-r}}{x \binom{x-r-1}{x-1}} \frac{\alpha^{x-r}}{(1+\alpha)^{x-r}}, \quad x = r, r+1, \ldots \quad (2.6.2)$$

Where $\alpha$ represents the traffic intensity and $r = \frac{\alpha}{(1+\alpha)}$, a constant, represents the probability of the arrival of the customers, $0 < r < 1$. The assumption that the probability
\( t \) to be a constant does not seem to be realistic. In fact, there may be cases when both the arrival time as well as the service time are not at a constant rate but may vary. In (2.6.2), if we assume that \( t = \frac{\mu}{1 + \alpha} \) varies then the probability of total number of customers served before the queue vanishes is given by

\[
P(X = x) = E \left[ \frac{r}{X} \left( \frac{2^r - 1}{2^r} \right) \right]
\]

(2.6.3)

Where expectation is to be taken over \( t \). Suppose \( t \) varies as beta distribution (2.3.5) with parameters \( (\alpha, \beta) \), then (2.6.3) reduces to

\[
P(X = x) = \frac{r}{x} \left( \frac{2^r - 1}{2^r} \right) \beta(a, b) \int_0^\infty \frac{1}{\beta(a, b)} e^{-(1+\alpha)x} \frac{b^r e^{-r t}}{(1-t)^{\alpha + \beta - 1}} dt
\]

Taking \( x = x + r \) i.e. starting with an idle queue, we obtain

\[
P(X = x) = \frac{r}{x + r} \left( \frac{2^r - 1}{2^r} \right) \beta(a, b) \frac{b^r e^{-r x}}{(x + r)^{\alpha + \beta - 1}}
\]

Which reduces to

\[
P(X = x) = \frac{r}{2x + r} \left( \frac{2^{r+1}}{2^r} \right) \frac{b^{r+1}}{(a + b)^{r+1}}
\]

(2.6.4)

The distribution represented by (2.6.4) is a new contagious distribution. Hence the distribution of total number of customers served before the queue vanishes assuming that we start with an idle queue wherein the random arrival time of customers is a poisson process and the time occupied in serving each customer follows negative exponential distribution and the probability distribution of arrival of customers follows a beta distribution is a contagious distribution (2.6.4).

**Remarks:** The proposed distribution is a particular case of Hassan and Bilal’s (2006) generalized negative Polya-Eggenberger distribution when \( \beta = 2 \) and Sen and Mishra’s (1996) generalized Polya-Eggenberger model (an urn model) when \( \mu = 1 \). Therefore, (2.6.4) represents a true probability distribution.
2.6.2 Structural properties of CDANPEID

2.6.2.1 Probability generating function.

The probability generating function $G_X(u)$ of the proposed model (2.6.4) is obtained as

$$G_X(u) = E(u^X) = \sum_{x=0}^{\infty} u^x \frac{r(2x+r-1)!}{(x+r-1)!} \frac{a^{(x)} b^{(x+r)}}{(a+b)^{2x+r}}$$

Which reduces to

$$G_X(u) = \frac{r b^{(r)}}{(a+b)^{r}} \sum_{x=0}^{\infty} \frac{r^{(2x)}}{r^{(x)} (a+b)^{r+x}} \frac{a^{(x)} (b+r)^{(x)}}{x!} u^x$$

A use of Gauss multiplication theorem $a^{(mn)} = k^{mn} \prod_{s=1}^{n} \left( \frac{x+s-1}{a} \right)$ gives the probability generating function as

$$G_X(u) = \frac{r b^{(r)}}{(a+b)^{r}} \sum_{x=0}^{\infty} \frac{a^{(x)} (b+r)^{(x)}}{(a+b+r)^{x}} \frac{r^{(x)}}{x!} \left( \frac{r+1}{2} \right)^{x} u^x$$

Converting the series on the right hand side of the equation above into hypergeometric function, we get

$$G_X(u) = \frac{r b^{(r)}}{(a+b)^{r}} F_{x} \left[ \frac{a+b+r+r+1}{2}, \frac{a+b+r}{2}, \frac{a+b+r+1}{2}, \frac{a+b+r+1}{2}, u \right]$$

(2.6.5)

Where $F_x \left[ a+b+r+r+1, \frac{a+b+r+1}{2}, \frac{a+b+r+1}{2}, \frac{a+b+r+1}{2}, u \right]$ is a generalized hypergeometric function [see; section 1.5]. The result (2.6.5) can be used to obtain the moments of the proposed model (2.6.4) but they turn out to be in terms of the generalized hypergeometric function which are not easy to be worked with.

2.6.2.2 Unimodality

The proposed model (2.6.4) is a unimode by the following result of Holgate (1970).

Lemma. If the mixing distribution is non-negative, continuous and unimode then the resulting distribution is unimodal.
The proposed model (2.6.4) is a unimodal since the mixing distribution is beta distribution which is unimodal for $a > 1$ and $b > 1$.

2.6.2.3 Recurrence relation

Taking $x = x + 1$ in the proposed model (2.6.4) and dividing the resulting equation by (2.6.4), we get the recurrence relation for probabilities of the proposed model as

$$P(X = x + 1) = \frac{(r + x)(r + 2x + 1)(a + x)(b + r + x)}{(r + x + 1)(x + 1)(a + b + r + 2x + 2)} P_t(Y = x)$$

(2.6.5)

Multiplying (2.6.6) by $(x + 1)^r$ and summing the resulting equation over the values of $x$, we get

$$\mu'(r, a, b) = \frac{r a}{(r + 1)(a + b)} \sum_{x=0}^{\infty} (x + 1)^{r+1} P(x, r + 1, a + 1, b)$$

Where $\mu'(r, a, b)$ represents the $k$th moment about origin of the proposed model with parameters $(r, a, b)$. A use of binomial theorem in the equation above and then rearranging the terms in the resulting equation, we get

$$\mu'(r, a, b) = \frac{r a}{(r + 1)(a + b)} \sum_{j=0}^{k} \sum_{x=0}^{\infty} x^j (r + 2x + 1) P_t(x, r + 1, a + 1, b)$$

Which subsequently gives the recurrence relation

$$\mu'(r, a, b) = \frac{r a}{(a + b)} \left[ \sum_{j=0}^{k} \mu_j'(r + 1, a + 1, b) + \frac{2}{(r + 1)} \mu'_t(r + 1, a + 1, b) \right]$$

(2.6.7)

In particular, $k = 1$ gives

$$\mu'_1(r, a, b) = \frac{r a}{(a + b)} \left[ 1 + \frac{2}{(r + 1)} \mu'_t(r + 1, a + 1, b) \right]$$

A repeated use of (2.6.8) on the function $\mu'_t(.)$ gives

$$\mu'_t(r, a, b) = \frac{r a}{(a + b)} \left[ 1 + \frac{2(a + 1)}{(a + b + 1)} + \frac{2(a + 1)(a + 2)}{(a + b + 1)(a + b + 2)} \right]$$
Using hypergeometric function, we obtain

\[
\mu_2(r,a,b) = \frac{ra}{(a+b)} \cdot {}_2F_1[a+1,1; a+b+2] \tag{2.6.8}
\]

Taking \( k = 3 \) in (2.6.7), we get

\[
\mu_3(r,a,b) = \frac{ra}{(a+b)} \left[ \frac{1}{(r+1)} \mu_2(r+1,a+1,b) + \frac{2}{(r+1)} \mu_2(r,a+1,b) \right] + \cdots \tag{2.6.9}
\]

Using (2.6.9) repeatedly on the function \( \mu_3(r,a,b) \) and then applying (2.6.8) in the resulting equation, we obtain

\[
\mu_3(r,a,b) = \frac{ra}{(a+b)} \left[ \frac{1}{(r+1)} \mu_2(r+1,a+1,b) + \frac{2}{(r+1)} \mu_2(r,a+1,b) \right]
\]

\[
\times \left[ (r+3) \cdot \frac{(a+1)}{(a+b+1)} \cdot {}_2F_1[a+2,1; a+b+2] + ... \right] + \frac{ra}{(a+b)} \]

\[
\times \left[ \frac{2(a+1)(a+2)}{(a+b+1)(a+b+2)} \cdot {}_2F_1[a+3,1; a+b+3,2] + ... \right]
\]

\[
\times \left[ \frac{2(a+1)(a+2)(a+3)}{(a+b+1)(a+b+2)(a+b+3)} \cdot {}_2F_1[a+4,1; a+b+4,2] + ... \right]
\]

Which gives on simplifications the second moment as

\[
\mu_3(r,a,b) = \frac{ra}{(a+b)} \cdot {}_2F_1[a+1,1; a+b+1,2] + \frac{(r+3)ra(a+1)}{(a+b)(a+b+1)} \cdot {}_2F_1[a+2,2; a+b+2,2] + \cdots
\]

\[
\times \frac{2ra(a+1)(a+2)}{(a+b)(a+b+1)(a+b+2)} \cdot {}_2F_1[a+3,3; a+b+3,3] + ...
\]

Taking \( k = 3 \) in (2.6.7), the third moment about origin of the proposed model is obtained as
\[
\mu'_1(r,a,b) = \frac{ra}{(a+b)^2} \left[ 1 + \frac{2r}{(r+1)} \mu'_1(r+1,a+1,b) + \frac{2^2(a+1)(a+2)}{(r+1)(a+b+1)} \mu'_1(r+2,a+2,b) + \frac{2^3(a+1)}{(r+1)(a+b+1)} \right]
\]

A repeated use of (2.6.10) on the function \(\mu'_1(.)\) together with the equation (2.6.9) gives

\[
\mu'_2(r,a,b) = \frac{ra}{(a+b)^2} \left[ 1 + \sum_{n=1}^{\infty} \frac{2^n(a+1)(a+2)}{(r+n)(a+b+1)} \right]
\]

Now, reducing the function \(\mu'_2(.)\) into the function \(\mu'_1(.)\) by making repeated use of (2.6.9) and then substituting the values of \(\mu'_1(.)\) from the equation (2.6.8) into the resulting equation, we obtain

\[
\mu'_2(r,a,b) = \frac{ra}{(a+b)^2} F_1[a+1,1;a+b+1,2] + \frac{ra(a+1)}{(a+b)(a+b+1)}
\]
Similarly, the expression for the fourth moment about origin of the proposed model can be obtained from (2.6.7). The required expression, after lengthy calculations, is

\[
\mu_4(r,a,b) = \frac{ra}{(a+b)^2} \left[ \begin{array}{c} F_i[a+1,1; a+b+1,2] + \frac{(a+1)}{(a+b+1)} \times \\
F_i[a+2,2; a+b+2,2] + \frac{(a+2)}{(a+b+2)} \times \\
F_i[a+3,3; a+b+3,2] + \frac{(a+3)}{(a+b+3)} \times \\
F_i[a+4,4; a+b+4,2] + \frac{(a+4)}{(a+b+4)} \times \\
\end{array} \right]
\]

It is not easy to obtain the central moments from the moments about origin as they are coming in messy forms and are not shown here.

2.6.3 Relation of CDANPED with other distributions
Theorem 2.6.1. Let \( X \) be a random variable following the proposed model (2.6.4) with parameters \((r,a,b)\), if \( b \to \infty \) such that \( ab^{l-r} = \lambda \) and \( \lambda_r = r \alpha \) as \( r \to \infty \), then show that \( X \) tends to a Poisson distribution (1.2.7) with parameters \( \lambda_r \).

Proof: Expressing the pmf of the proposed model as

\[
P(X = x) = \frac{r(r+2x-1)\cdots(r+2x-2)\cdots(r+x+1)}{x!}
\]

Which gives on simplifications

\[
P(X = x) = \frac{r(r+2x-1)\cdots(r+2x-2)\cdots(r+x+1)}{x!}
\]

Taking limit \( b \to \infty \) in such a way so that \( ab^{l-r} = \lambda \), the equation above reduces to

\[
\lim_{b \to \infty} P(X = x) = \frac{(r+2x-1)\cdots(r+x+1)}{x!}
\]

Substituting \( \lambda = \frac{\lambda_r}{r} \) in the equation above and proceeding to limit \( r \to \infty \), we get Poisson distribution (1.2.7) with parameter \( \lambda \).

Theorem 2.6.2. Let \( X \) be a random variable following the proposed model (2.6.4) with parameters \((r,a,b)\). If \( a \to \infty \) such that \( b = a \theta \) then show that \( X \) tends to generalized negative binomial distribution.

Proof: The pmf of the proposed model can be written as

\[
P(x) = \frac{r}{2x+r} \binom{2x+r}{x} \frac{a(a+1)\cdots(a+x-1)}{(a+b)(a+b+1)\cdots(a+b+2x+r-1)}
\]

Dividing numerator and denominator by \( a^{2x+r} \) and in the resulting equation taking \( \theta = \frac{b}{a} \) and proceeding to limit \( a \to \infty \), we get

\[
\lim_{a \to \infty} \frac{r}{2x+r} \binom{2x+r}{x} \left( \frac{1}{1+\theta} \right)^x \left( 1 - \frac{1}{1+\theta} \right)^{2x+r-x}
\]
Which is a particular case of Jain and Consul's (1971) generalized negative binomial distribution with parameters \( r, (1+\theta)^{-1} \).

**Remarks:** Taking \( \frac{a}{b} = \theta \) in (2.6.1) and proceeding to limit \( b \to \infty \) we get a new generalization of Haight's distribution with pmf

\[
P(x) = \frac{r}{2x+r} \binom{2x+r}{x} \left( \frac{\theta}{1+\theta} \right)^x, \quad x=0,1,2,... \quad 0 < \theta < 1
\]

as it reduces to Haight's distribution (1961) when \( r = 1 \).

**Theorem 2.6.3.** Show that the proposed model truncated at zero tends to generalized Logarithmic series distribution.

**Proof:** The pmf of the zero-truncated proposed model is defined as

\[
P(X=x) = \frac{r}{2x+r} \binom{2x+r}{x} \frac{(a+b)^r}{(a+b)^{2x+r}} \frac{(a+b)^r}{(a+b)^{2x+r} - b^r}, \quad x=1,2,...
\]

Substituting \( \frac{b}{a} = \theta \) in the equation above and proceeding to limit \( a \to \infty \), we get

\[
\lim_{a \to \infty} P(X=x) = \frac{r}{2x+r} \binom{2x+r}{x} \left( \frac{\theta}{1+\theta} \right)^x \frac{1}{1-\theta'(1+\theta)^{-r}}
\]

(2.6.12)

Now, we can write

\[
1-\theta'(1+\theta)^{-r} = 1 - \left( \frac{1}{1+\theta} \right)^r
\]

A use of binomial theorem in the equation above gives

\[
1-\theta'(1+\theta)^{-r} = -r \left( \frac{1}{1+\theta} + \frac{(r-1)}{2!} \left( \frac{1}{1+\theta} \right)^2 + \frac{1}{r!} \left( \frac{1}{1+\theta} \right)^r \right)
\]

Substituting this value in (2.6.12) and then proceeding to limit \( r \to \theta \), we get a logarithmic series distribution with pmf

\[
P(X=x) = \frac{1}{x!} \frac{1}{1-\theta'} \left( \frac{1}{1+\theta} \right)^x \left( 1 - \frac{1}{1+\theta} \right)^x, \quad x=1,2,...
\]
Which is a particular case of Jam and Gupta's (1973) generalized logarithmic series distribution with parameter \(\frac{1}{l+i}\).

**Theorem 2.6.4.** Let \(X\) is a random variable following the proposed model (2.6.4) with parameters \((r,a,b)\). If \(a \to l\) and \(r \to l\) then show that \(X\) tends to generalized Yule's distribution.

**Proof:** Proceeding to limits as \(r \to l\), \(\alpha \to l\) in (2.6.11), we get

\[
P(x) = \frac{b\Gamma(2x+1)\Gamma(b+x+1)}{\Gamma(x+2)\Gamma(b+2x+2)}, \quad x=0,1,2,...
\]

Which is a particular case of the generalized Yule's distribution with zero's defined by Kenwar Sen (2004).

### 2.7. Goodness of fit

In this section, we present three data sets [tables (2.7.1)-(2.7.3)] to examine the fitting of the proposed models NPED and CDANPED and comparing these with the negative binomial distribution and Consul and Jain's (1971) GNBD.

Due to complicated likelihood function, the maximum likelihood estimate of the parameters of all the proposed models are not straightforward and need some iterative procedure such as Fisher's scoring method, Newton-Rampson method etc. for their solution. R-software provides one among such solutions. In R-software there is a function called "nlm" (non-linear minimization) that minimizes the negative log-likelihood function or equivalently maximizes the log likelihood function for estimating the parameters of the distribution by adopting Newton-Rampson iterative procedure. A random start procedure is employed, that is, for a set of random starting points the function "nlm" searches recursively until global maxima is reached. In order to verify that the global maximum has been reached the gradient should be equal to zero. The closer the value of the random starting points to the ML estimate the lesser number of iterations are required to obtain the global maximum.
As mentioned above, the maximum likelihood equations of all the proposed models discussed in this chapter and in succeeding chapters are not straightforward to provide the maximum likelihood (ML) estimates of the parameters of these models. Therefore, a computer program in R-software has been used to estimate the parameters and has been shown in the bottom of their respective tables.

Table 2.7.1

Absenteeism among shift-workers in steel industry; data of Arbous and Sichel, 1954

<table>
<thead>
<tr>
<th>Count</th>
<th>Obs. freq.</th>
<th>Expected frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NBD</td>
</tr>
<tr>
<td>0</td>
<td>7</td>
<td>12.02</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>16.16</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>17.77</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>18.03</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>17.65</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>16.80</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>15.72</td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>14.52</td>
</tr>
<tr>
<td>8</td>
<td>09</td>
<td>13.28</td>
</tr>
<tr>
<td>9</td>
<td>09</td>
<td>12.06</td>
</tr>
<tr>
<td>10</td>
<td>08</td>
<td>10.89</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>09.78</td>
</tr>
<tr>
<td>12</td>
<td>08</td>
<td>08.75</td>
</tr>
<tr>
<td>13</td>
<td>07</td>
<td>07.80</td>
</tr>
<tr>
<td>14</td>
<td>02</td>
<td>06.93</td>
</tr>
<tr>
<td>15</td>
<td>12</td>
<td>06.14</td>
</tr>
<tr>
<td>16</td>
<td>03</td>
<td>05.43</td>
</tr>
<tr>
<td>17</td>
<td>05</td>
<td>04.79</td>
</tr>
<tr>
<td>18</td>
<td>04</td>
<td>04.22</td>
</tr>
<tr>
<td>Count</td>
<td>Obs. freq.</td>
<td>Expected frequencies</td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
<td>----------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NBD</td>
</tr>
<tr>
<td>0</td>
<td>206</td>
<td>198.85</td>
</tr>
<tr>
<td>1</td>
<td>143</td>
<td>166.90</td>
</tr>
<tr>
<td>2</td>
<td>128</td>
<td>124.33</td>
</tr>
<tr>
<td>3</td>
<td>107</td>
<td>88.70</td>
</tr>
<tr>
<td>4</td>
<td>71</td>
<td>61.89</td>
</tr>
<tr>
<td>5</td>
<td>36</td>
<td>42.60</td>
</tr>
<tr>
<td>6</td>
<td>32</td>
<td>29.05</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>19.68</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>13.27</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>8.91</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>5.96</td>
</tr>
</tbody>
</table>

**Table 2.7.2**

The data has been taken from Beall-Ruscia Table VII.
Table 2.731

Accidents to 647 women working on H.E. shells during 5 weeks

<table>
<thead>
<tr>
<th>Count</th>
<th>Obs. Freq.</th>
<th>Expected frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NBD</td>
</tr>
<tr>
<td>0</td>
<td>447</td>
<td>445.89</td>
</tr>
<tr>
<td>1</td>
<td>132</td>
<td>134.90</td>
</tr>
<tr>
<td>2</td>
<td>42</td>
<td>44.59</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>14.09</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>4.96</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2.57</td>
</tr>
<tr>
<td>TOTAL</td>
<td>647</td>
<td>647</td>
</tr>
</tbody>
</table>

ML Estimate

<table>
<thead>
<tr>
<th>ML Estimate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>0.86512</td>
</tr>
<tr>
<td>p</td>
<td>0.34969</td>
</tr>
<tr>
<td>\beta</td>
<td>8.26367</td>
</tr>
<tr>
<td>\alpha</td>
<td>17.78681</td>
</tr>
<tr>
<td>\gamma</td>
<td>7.727532</td>
</tr>
<tr>
<td>r</td>
<td>7.272532</td>
</tr>
<tr>
<td>a</td>
<td>2.213498</td>
</tr>
<tr>
<td>b</td>
<td>42.140665</td>
</tr>
</tbody>
</table>

n = 647, p = 0.34969, \beta = 8.26367, \alpha = 17.78681, \gamma = 7.727532, r = 7.272532, a = 2.213498, b = 42.140665
From all the tables (2.7.1)-(2.7.2) it is clear that the CANPED gives a very close fit as compared to other distributions. Thus, the CANPED provides a better alternative to explain the data than the compared distributions.

Now, we present two data sets [tables (2.7.4)-(2.7.5)] to examine the fitting of SBNPED and comparing that with the zero-truncated NPED. The parameters have been estimated with the help of a computer programme in R-soft.

Table 2.7.4

Observed and expected number of households of size group 16 and above having at least one migrant.

<table>
<thead>
<tr>
<th>No. of migrants</th>
<th>Obs. No. of households</th>
<th>Expected number of households</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SBNPED</td>
</tr>
<tr>
<td>1</td>
<td>101</td>
<td>101.38</td>
</tr>
<tr>
<td>2</td>
<td>70</td>
<td>68.87</td>
</tr>
<tr>
<td>3</td>
<td>34</td>
<td>34.09</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>15.03</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>6.29</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2.57</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1.05</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0.72</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>230</strong></td>
<td><strong>230</strong></td>
</tr>
</tbody>
</table>

**ML estimate**

\[ n = 15.977758 \]
\[ \alpha = 1.510667 \]
\[ \gamma = 44.257112 \]

\[ n = 2.202049e-04 \]
\[ \alpha = 8.313891e-07 \]
\[ \gamma = 9.875006e-01 \]

\[ \chi^2 = 0.471 \]
\[ \text{d.f} = 1 \]

\[ \chi^2 = 81.302 \]
\[ \text{d.f} = 1 \]
Table 2.7.5

Observed and expected number of households of size group 16 and above having at least one migrant.

<table>
<thead>
<tr>
<th>No. of migrants</th>
<th>Obs. No. of households</th>
<th>Expected number of households</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SBNPED</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>18.27</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>21.15</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>14.89</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>8.33</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>4.10</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1.87</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.81</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0.58</td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>70</td>
</tr>
</tbody>
</table>

| ML estimate     |                        | n = 8.062025 | n = 9.863161e-04 |
|                 |                        | α = 8.062025 | α = 2.345760e-07 |
|                 |                        | γ = 53.789650| γ = 9.899945e-01 |
| χ²              | 0.392                  |              | 66.972         |
| d.f             | 1                      |              | 1              |

From both the tables (2.7.4)-(2.7.5) it is clear that size-biased negative Polya-Eggenberger distribution gives a very close fit as compared to zero-truncated negative Polya-Eggenberger distribution that is why it is preferred to use size-biased negative Polya-Eggenberger distribution because of it’s greater ability and better power to explain the data. Thus, size-biased negative Polya-Eggenberger distribution provides a better alternative to zero-truncated negative Polya-Eggenberger distribution.