A Generalization to the Family of Discrete Distributions

Abdullah Albalawi[1], Mian Arif Shams Adnan[2], Tareq F Khan[3]
[1]Ball State University, Muncie, IN 47304, USA
[2]Jahangirnagar University, Dhaka 1342, Bangladesh

Abstract
An alternative approaches for a couple of discrete distributions like Binomial and Multinomial, Poisson, etc having more general form of sampling method (more than one outcome in one trial) compared to tradition sampling heuristics have been suggested and termed as Generalized Binomial, Generalized Multinomial, Generalized Poisson, Generalized Geometric respectively. It is evident that the traditional existing distributions are the special cases of the proposed generalized distributions. The basic distributional properties of the proposed distributions have also been examined including the limiting form. Real life examples are cited for the respective distributions.

Key Words: Concentration ; Limiting Property ; Progression.

1. Introduction

The discrete distributions are widely used in the diversified field and among them the distribution like Binomial, Multinomial, Poisson and Geometric are the most commonly used discrete distributions. The other discrete distributions include uniform or rectangular, hypergeometric, negative binomial, power series etc. The truncated and censored forms of the different discrete distributions are used as probability distribution in statistics literature and real life. The binomial distribution was first studied in connection with the games of pure chance but it is not limited within narrow area, where the Multinomial distribution is considered as the generalization of the binomial distribution. The number of mutually exclusive outcomes from a single trial are k in multinomial distribution compared to two outcomes namely success or failure of Binomial distribution.

The usual binomial distribution is the discrete probability distribution of the number of successes 0 to n resulted from n independent Bernoulli trails each of which yields success with probability p and failure with probability q. One of the important assumptions regarding binomial variate representing the number of success is that it can take only values in the sequence of $0,1,2,\cdots,n$. But in real world, the successes of binomial may not occur in the usual way rather it may occur in a different sequence such as (i) $0,2,4,\cdots,2n$, (ii) $2,4,\cdots,2n$, (iii) $0,3,6\cdots,3n$, (iv) $3,6,9,\cdots,3n$ and so on.

In cases (ii) and (iv), truncated distribution is the better option to find the probability of number of success. In truncated distribution, it is assumed that the truncated values of the random variable have certain probabilities. If it is considered that there is no existence of the truncated values, the truncated distribution cannot provide the probability due to the mathematical cumbersome. In the remaining cases, Binomial and truncated Binomial are completely helpless.

In this context, we have suggested an alternative approach of Binomial and Multinomial distribution having generalized sequence of the values of the random variables. For convenience the distributions are defined as relatively generalized distribution. The number of successes of the proposed distributions is represented by the arithmetic progression $a + nd$, where, a is non-negative integer and termed as minimum number of success, d is positive integer representing the concentration of success and n is a non-negative integer indicating the total number of trails.

To justify the sequence with real life situation, let us consider an example of number of defective shoes. It is well known that the shoes are produced pair wise. That is, if we make n attempts to identify the number of defective shoes, then it is usual that the number of defective shoes would occur pair wise. In this context, the number success in the form $0,2,4,\cdots,2n$ is justified. If it is known that the minimum number of defective items in n attempts are a, then these occurrences are not chance outcome and may be regarded as constant and then we are interested to find the probability of chance outcome taking form $a, a+2, a+4, \cdots, 2n$, where $a > 0$ and n is non-negative integer. That is, number of defective shoes larger than a. The other possible sequences are also justifiable in this way by real life examples.

One may confuse the proposed form of the distribution with the usual truncated distribution. The major difference is that in our form the probability exists only for the possible number of success shown in the sequence. In the example discussed above, the defective number of shoes take the values $0,2,4,\cdots,2n$. This indicates that there is no existence of the number of defective item(s) of the form $1,3,5,\cdots$ and thus no probabilities. On the other hand, in truncated distribution, there is existence of the number of success which is truncated and also they have the probabilities.

A number of authors published their work under the heading of generalized binomial distribution but they were different in terms of the key concept of our present work. Altham (1978) showed two generalizations of the binomial distribution when the random variables are identically distributed but not independent and assumed to have symmetric joint distribution with no second or higher order "interactions". Two generalizations are obtained depending on whether the "interaction" for discrete variables is "multiplicative" or "additive". The distribution has a new parameter $\theta > 0$ which controls the shape of the distribution and flexible to allow for both over or under-disperse than traditional Binomial distribution. Whereas, the beta-binomial distribution allows only for over-disperse distribution than the corresponding Binomial distribution (Johnson, Kemp and Kotz 2005). Dwass (1979) have provided a unified approach to a family of discrete distributions that includes the hypergeometric, Binomial, and Polya distributions by considering the simple sample scheme where after each drawing there is a "replacement" whose magnitude is a fixed real number. Paul (1985) derived a new three parameter distribution, a generalization of the binomial, the beta-binomial and correlated beta-binomial distribution. Further a modification on beta-correlated binomial distribution was proposed by Raul (1987). In the generalization of the probability distribution, Panaretos and Xekalaki (1986) developed cluster binomial and multinomial model and their probability distributions. In his study, Madsen (1993) discussed that in many cases binomial distribution fails to apply because of more variability in the data than that can be explained by the distribution. He pointed out a characterization of sequences of exchangeable Bernoulli random variables which can be used to develop models which is more fluxion than the traditional binomial distribution. His study exhibited sufficient conditions which will yield such models and show how existing models can be combined to generate further models.

A generalization of the binomial distribution is introduced by Drezner and Farnum (1993) that allow the dependence between trials, non-constant probabilities of success from trial to trial and which contains usual binomial distribution as special case. A new departure in the generalization was carried out by Fu and Sproule (1995) by adopting the assumption that the underlying Bernoulli trials take on the values α or β where $\alpha < \beta$, rather than the conventional values 0 or 1. This rendered a four parameter binomial distribution of the form $B(n, p, \alpha, \beta)$. In a recent work, Altham and Hankin (2012) introduced two generalizations of multinomial and Binomial distributions which arise from the multiplicative Binomial distribution of Altham (1978). The forms of the generalized distributions are of exponential family form and termed as "multivariate multinomial distribution" and "multivariate multiplicative binomial distribution". Like the Altham's generalized distribution, both the distribution has an additional shape parameter θ which corresponds usual distribution if it takes value 1 and over and under-disperse for greater and less than 1 respectively.

The Poisson and Geometric distributions and their applications in statistical modeling and in many other scientific fields are well recognized. The traditional Poisson variate corresponds to the number of occurrences of the rare event in a fixed interval of time or space or other intervals and assumed to lie in discrete order between 0 to ∞. On the other hand, the usual Geometric variate is used to present the number of failure preceding the first success having the same range of the values of Poisson variate. But, in real world the event may occur in different fashions. Consider the sampling scheme where the number of occurrences possess the sequence such as (i) $0,2,4,\cdots,\infty$, (ii) $2,4,\cdots,\infty$, (iii) $3,6,9,\cdots,\infty$, (iv) $1,4,7,\cdots,\infty$ and so on. For instance, consider the total number of births those who born as twins at a particular hospital during a specified time interval. Let us define the number of new births as success and thus the number of success possess values $0,2,4,\cdots,\infty$. The usual Poisson distribution is completely helpless to deals with this particular sequence of number of success along with others mentioned above. Again, consider the example of twin births in geometric sense. In the sampling scheme, let us define the event as success if both the twins are male and stopped the sampling and number of births as twins in either combination of girl and boy or both girls are considered as failure. With these sequences of the values of the random variable, the traditional Geometric distribution cannot be applied to find a certain probability of a particular event.

In order to deal with the problems where traditional distributions are unaided, our study have suggested new generalization of the traditional Poisson and Geometric distributions where the random variable for each distribution is expressed by an arithmetic progression $a + nd$, where a is an integer representing the minimum values of the random variable, n is a pre-assigned non-negative integer indicating the number of trails and d is also a positive integer representing the concentration of the occurrences.

A series of studies have carried out under the heading generalization of Poisson distribution where the key concept our work is completely different. Consul and Jain (1973) first suggested generalization of Poisson distribution having two parameters λ_1 and λ_2 which is obtained as a limiting form of the generalized negative Binomial distribution. In usual Poisson distribution, the mean and variance are same, while the variance of the suggested generalized distribution is greater than, equal to or smaller than the mean depending on whether the value of the parameter λ_2 is positive, zero or negative. Later Consul (1989) studied more extensively the distribution to cover the diversity of the observed number of occurrences for various factors. He also mentioned that proving the sum of all of the probabilities to unity is very difficult. In this context, Lerner et al. (1997) provide a more direct proof using the analytic functions. Some remarks on generalized negative binomial and Poisson distributions were made by Nelson (1975). Paul (1978) proposed a generalized compound Poisson model for panel data analysis on consumer purchase. Lin (2004), in his study discussed about the generalized Poisson models and their applications in insurance and finance sector. On the basis of gamma function and digamma function a new two-parameter count distribution is derived by Hagmark (2012). He unveiled that the derived distribution can attain any degree of over/under dispersion or zero-inflation/deflation where the usual Poisson model has no dispersion flexibility.

Several authors have worked on generalization of the Geometric distribution which are also differ from the concept of our current work. Mishra (1982) proposed generalized geometric series distribution (GGSD) and shown that traditional geometric and Jain and Consul's (1971) generalized negative binomial distribution are the special cases of the proposed distribution. A generalized Geometric distribution and its properties was introduced and studied by Philippou et al. (1983) from the motivation of the work by Philippou and Muwafi (1982). The distribution was defined under the heading of Geometric distribution of order k where the distribution turned into the traditional form if order $k = 1$. Considering the length-biased version of the generalized log-series distributions by Kempton (1975) and Tripathi and Gupta (1985), Tripathi et al. (1987) derived two version of two parameter generalized Geometric distribution. Another generalization of the Geometric distribution having two parameters was obtained by Gomez-Deniz (2010) obtained. He showed the generalization can be obtained either by using Marshal and Olkin (1997) scheme and adding a parameter to the Geometric distribution or from generalized exponential distribution presented in the same paper by Marshal and Olkin (1997). Nassar and Nada (2013) discovered a five parameter new generalization of the Pareto-geometric distribution by compounding Pareto and Geometric distribution and generalized it by logit of the beta random variable.

2.1 New Approach of Binomial Distribution

In the traditional Binomial sampling each time we draw a sample having either a success or a failure, we continue this method up to n trials and we may have number of successes starting from 0 and may ended up to n, but in real world, the success of Binomial distribution may not occur in the usual sequences rather it may happened as

i) $0,2,4,\cdots,2n$
ii) $2,4,\cdots,2n$
iii) $0,3,6\cdots,3n$
iv) $3,6,9,\cdots,3n$

and so on.

These indicate that the number of success may follow an arithmetic progression $a+nd$, where, a is a non-negative integer representing the minimum number of success, d is a positive integer representing the concentration of success occurring and n is a non-negative integer indicating the total number of trails.

Definition 2.1.1: A random variable *X* is said to have a relatively general binomial distribution if it has the following probability mass function

$$P(x;a,n,d,p)=\frac{\binom{a+nd}{x}p^x q^{a+nd-x}}{\sum_{x=a}^{a+nd}\binom{a+nd}{x}p^x q^{a+nd-x}};\ x=a,a+d,a+2d,\cdots,a+nd$$
$$=\frac{1}{k}\binom{a+nd}{x}p^x q^{a+nd-x}$$

where, $a\geq 0$ is the minimum number of success, $d>0$ is the concentration of occurrence of success, n is a predefined finite number of non-negative integer represents the number of trials and p is the probability of success such that $p+q=1$ and $k=\sum_{x=a}^{a+nd}\binom{a+nd}{x}p^x q^{a+nd-x}$ is a constant. The probability mass function $P(x;a,n,d,p)$ stands for probability of getting x success out of maximum of $a+nd$ successes in n trials.

Theorem 2.1.1: *For the generalized binomial distribution with parameter $a\geq 0, d>0, n\geq 0$ and p and usual binomial distribution with parameter $n\geq 0$ and , the probability mass function of generalized binomial distribution reduces to the probability mass function of usual binomial distribution when $a=0$.*

Theorem 2.1.2: *The moment generating function of generalized binomial distribution is*

$$M_X(t)=\frac{\sum_{x=a}^{a+nd}\binom{a+nd}{x}(pe^t)^x q^{a+nd-x}}{\sum_{x=a}^{a+nd}\binom{a+nd}{x}p^x q^{a+nd-x}}.$$

Theorem 2.1.3: *The mean and variance of generalized binomial distribution are $(a+nd)p$ and $(a+nd)pq$ respectively.*

Theorem 2.1.4: *If X follows the generalized binomial distribution then find the 3rd and 4th raw and central moments of X respectively are*

$$\mu_3' = (a+nd)^3p^3 - 3(a+nd)^2p^3 + 2(a+nd)p^3 + 3(a+nd)^2p^2 - 3(a+nd)p^2 + (a+nd)p$$
$$\mu_4' = (a+nd)^4p^4 - 6(a+nd)^3p^4 + 11(a+nd)^2p^4 - 6(a+nd)p^4 + 6(a+nd)^3p^3 - 18(a+nd)^2p^3 + 12(a+nd)p^3 + 7(a+nd)^2p^2 - 7(a+nd)p^2 + (a+nd)p$$
$$\mu_3 = (a+nd)pq(1-2p)$$
$$\mu_4 = (a+nd)pq\big[1 + 3\big((a+nd)-2\big)pq\big]$$
The special case for 3rd and 4th raw and central moments holds for $a = 0$ and $d = 1$ and turned to the form of usual Binomial distributions.

Theorem 2.1.5: *The shape charactersitics of generalized binomial distribution are:*
$$Measures\ of\ Skewness\ (\beta_1) = \frac{(1-2p)^2}{(a+nd)pq}, Coefficient\ of\ Skewness\ (\gamma_1) = \sqrt{\beta_1} = \frac{1-2p}{\sqrt{(a+nd)pq}}$$
From coefficient of skewness, the following conclusion can be drawn and still surprising that the nature of skew depends only on p only and which is similar to the usual binomial distribution as:

- *i) The distribution is positively skewed if $p < \frac{1}{2}$.*
- *ii) On the other hand, the distribution is negatively skewed if $p > \frac{1}{2}$.*
- *iii) And the distribution is symmetric if $p = \frac{1}{2}$.*

$$Measures\ of\ kurtosis\ (\beta_2) = 3 + \frac{(1-6pq)}{(a+nd)pq}, Coefficient\ of\ kurtosis\ (\gamma_2) = \beta_2 - 3 = \frac{(1-6pq)}{(a+nd)pq}$$
These equations tell us that the generalized distribution is

- *i) Mesokurtic if $= \frac{1}{6}$.*
- *ii) Platykurtic $pq > \frac{1}{6}$ and*
- *iii) Leptokurtic if $pq < \frac{1}{6}$.*

Theorem 2.1.6: *The maximum likelihood estimator of the parameter p is $\frac{x}{a+nd}$, where x is the total number of success from maximum of $a + nd$ success in n trials.*

Theorem 2.1.7: *Normal distribution is a limiting form of generalized binomial distribution.*

2.2 New Approach of Multinomial Distribution

Under the sampling scheme described in Section 2.1, consider the situation where one of the k mutually exclusive outcomes is possible from a single trial other than only success or failure. More specifically, if the k outcomes are denoted by $e_1, e_2, e_2, \cdots, e_k$ and the number of occurrences of the respective outcomes are denoted by $x_1, x_2, \cdots, x_k$ such that $\sum_{i=1}^{k} x_i = a + nd$, where a is non-negative integer and termed as minimum number of success, d is positive integer representing the concentration of success and n is a non-negative integer indicating the total number of trails, then our suggested general form of Binomial as well as traditional Multinomial distribution cannot provide the probability that the event e_1 occurred x_1 times, the event e_2 occurred x_2 times and so on the event e_k occurred x_k times. To overcome this situation, we have suggested the new approach of Multinomial distribution and termed as relatively more general Multinomial or generalized Multinomial distribution. In this section we would present only the definition of the suggested distribution and statement of the theorems that we have derived for our present work to keep the paper size standard and other reason is that the derivations are very much similar to the suggested Binomial distribution.

Definition 2.2.1: k discrete random variable $\mathrm{X}_1, \mathrm{X}_2, \cdots, \mathrm{X}_\mathrm{k}$ is said to have a generalized multinomial distribution if it has the following probability mass function

$$\mathrm{P}(\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_\mathrm{k}; \mathrm{a}, \mathrm{n}, \mathrm{d}, \mathrm{p}_1, \mathrm{p}_2, \cdots, \mathrm{p}_\mathrm{k}) = \frac{\frac{1}{\mathrm{x}_1!\,\mathrm{x}_2!\cdots\mathrm{x}_\mathrm{k}!}\mathrm{p}_1^{\mathrm{x}_1}\mathrm{p}_2^{\mathrm{x}_2}\cdots\mathrm{p}_\mathrm{k}^{\mathrm{x}_\mathrm{k}}}{\sum_{\mathrm{x}_1,\mathrm{x}_2,\cdots,\mathrm{x}_\mathrm{k}=\mathrm{a}}^{\mathrm{a}+\mathrm{nd}}\frac{1}{\mathrm{x}_1!\,\mathrm{x}_2!\cdots\mathrm{x}_\mathrm{k}!}\mathrm{p}_1^{\mathrm{x}_1}\mathrm{p}_2^{\mathrm{x}_2}\cdots\mathrm{p}_\mathrm{k}^{\mathrm{x}_\mathrm{k}}}$$
$$= \frac{\frac{(\mathrm{a}+\mathrm{nd})!}{\mathrm{x}_1!\,\mathrm{x}_2!\cdots\mathrm{x}_\mathrm{k}!}\mathrm{p}_1^{\mathrm{x}_1}\mathrm{p}_2^{\mathrm{x}_2}\cdots\mathrm{p}_\mathrm{k}^{\mathrm{x}_\mathrm{k}}}{\sum_{\mathrm{x}_1,\mathrm{x}_2,\cdots,\mathrm{x}_\mathrm{k}=\mathrm{a}}^{\mathrm{a}+\mathrm{nd}}\frac{(\mathrm{a}+\mathrm{nd})!}{\mathrm{x}_1!\,\mathrm{x}_2!\cdots\mathrm{x}_\mathrm{k}!}\mathrm{p}_1^{\mathrm{x}_1}\mathrm{p}_2^{\mathrm{x}_2}\cdots\mathrm{p}_\mathrm{k}^{\mathrm{x}_\mathrm{k}}}$$

where, $\mathrm{a} \geq 0, \mathrm{d} > 0, \mathrm{n} \geq 0$ and $\mathrm{p}_1, \mathrm{p}_2, \cdots, \mathrm{p}_\mathrm{k}$ such that $\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{p}_\mathrm{i} = 1$, are the parameters of the distribution. And $\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{x}_\mathrm{i} = \mathrm{a} + \mathrm{nd}$.

The defined function is a probability mass function as it holds the conditions of the probability function and easily reduces to traditional Multinomial distribution if $a = 0$ and $d = 1$ and thus may be said that the traditional Multinomial distribution is a special case of the suggested Multinomial distribution.

Theorem 2.2.1: *Generalized Binomial distribution is a special case of generalized Multinomial distribution.*

Theorem 2.2.2: *The moment generating function of generalized Multinomial distribution is*

$$M_{X_1,X_2,\cdots,X_k}(t_1, t_2, \cdots, t_k) = \frac{\sum_{x_1,x_2,\cdots,x_k=a}^{a+nd}\frac{(a+nd)!}{\prod_{i=1}^{k} x_i!}\prod_{i=1}^{k}(p_i e^{t_i})^{x_i}}{\sum_{x_1,x_2,\cdots,x_k=a}^{a+nd}\frac{(a+nd)!}{\prod_{i=1}^{k} x_i!}\prod_{i=1}^{k} {p_i}^{x_i}}$$

and the mean and the variance of X_i $(i = 1,2,\cdots,k)$ *respectively are* $E(X_i) = \mu'_{1_i} = (a+nd)p_i$, $V(X_i) = (a+nd)p_i(1-p_i)$.

Theorem 2.2.3: *The 3rd and 4th raw and central moments of generalized Multinomial distribution are*

$$\mu'_{3_i} = (a+nd)^3{p_i}^3 - 3(a+nd)^2{p_i}^3 + 2(a+nd){p_i}^3 + 3(a+nd)^2{p_i}^2 - 3(a+nd){p_i}^2 + (a+nd)p_i$$

$$\mu'_{4_i} = (a+nd)^4{p_i}^4 - 6(a+nd)^3{p_i}^4 + 11(a+nd)^2{p_i}^4 - 6(a+nd){p_i}^4 + 6(a+nd)^3{p_i}^3 - 18(a+nd)^2{p_i}^3 + 12(a+nd){p_i}^3 + 7(a+nd)^2{p_i}^2 - 7(a+nd){p_i}^2 + (a+nd)p_i$$

$$\mu_{3_i} = (a+nd)p_i(1-p_i)(1-2p_i)$$

and $\mu_4 = (a+nd)p_i(1-p_i)\big[1 + 3\big((a+nd)-2\big)p_i(1-p_i)\big]$

For $a = 0$ and $d = 1$, the moments reduces to those of traditional Multinomial distribution.

Theorem 2.2.4: *Shape characteristics of generalized multinomial distribution are*

Measure of Skewness $\beta_{1_i} = \frac{(1-2p_i)^2}{(a+nd)p_i(1-p_i)}$, *Coefficient of Skewness* $\gamma_{1_i} = \sqrt{\beta_{1_i}} = \frac{1-2p_i}{\sqrt{(a+nd)p_i(1-p_i)}}$

Measure of Kurtosis $\beta_{2_i} = 3 + \frac{\{1-6p_i(1-p_i)\}}{(a+nd)p_i(1-p_i)}$, *Coefficient of Kurtosis* $\gamma_{2_i} = \beta_{2_i} - 3 = \frac{\{1-6p_i(1-p_i)\}}{(a+nd)p_i(1-p_i)}$

Theorem 2.2.5: *The maximum likelihood estimator of the parameters of generalized multinomial distribution is* $\widehat{p}_i = \frac{x_i}{(a+nd)}$, *where* x_i *is the number of success comprising* $(a+nd-x_i)$ *is the total number of failure subject to condition that maximum number of success is* $a+nd$.

2.3 Proposed Generalized Poisson Distribution

In the count data model, the traditional Poisson variate representing the number of occurrences takes the value ranges from 0 to ∞. Consider the following sequences for the values of the Poisson variate indicating the number of occurrence (i) $0,2,4,\cdots,\infty$; (ii) $2,4,\cdots,\infty$; (iii) $3,6,9,\cdots,\infty$; (iv) $1,4,7,\cdots,\infty$ and so on. The arithmetic progression $a+nd$ can be used to represent the number of occurrences where, a is a non-negative integer representing the minimum number of occurrence, d is a positive integer representing the concentration of occurrence and n is a non-negative integer indicating the total number of trails. To tackle with the situation where the number of occurrences follows an arithmetic progression, we formulate a probability function and defined as generalized Poisson distribution. The sequence can easily take the traditional sequence $0,1,2,\cdots,\infty$ of Poisson distribution for the values $a=0$ and $d=1$ and then our proposed distribution turned into the usual Poisson distribution. Thus, a series of probabilistic problems can be solved by the proposed distribution including the problems solved by the traditional one. In this section, we define the proposed distribution and provide some of its important properties.

Definition 2.3.1: A random variable X is said to have a Poisson distribution with parameter λ, a, n and d if it has the following probability mass function

$$\mathrm{P}(\mathrm{x};\lambda,\mathrm{a},\mathrm{n},\mathrm{d})=\frac{\lambda^{\mathrm{x}}}{\mathrm{x}!\sum_{\mathrm{n}=0}^{\infty}\frac{\lambda^{\mathrm{a+nd}}}{(\mathrm{a}+\mathrm{nd})!}}\qquad;\mathrm{x}=\mathrm{a},\mathrm{a}+\mathrm{d},\mathrm{a}+2\mathrm{d},\cdots,\mathrm{a}+\mathrm{nd}$$

where, $\mathrm{a}\geq 0$ is the minimum number of occurrence, $\mathrm{d}>0$ is the concentration of occurrence, n is a pre-assigned non-negative integer such that $\mathrm{a}=0$ if $\mathrm{n}=0$ and $\lambda\geq 0$ is the mean number of occurrences.

It can be clearly shown that $P(x;\lambda,a,n,d)\geq 0$ and $\sum_{x=a}^{a+nd}P(x;\lambda,a,n,d)=1$ for different values of X in terms of $a\geq 0, d>0$ and $n>0$. Hence, the suggested function of generalized Poisson distribution is a probability mass function.

The generalized Poisson distribution tends to the traditional Poisson distribution when minimum number of occurrence $a=0$ and concentration of occurrence $d=1$. Thus, we may conclude that the traditional Poisson distribution is a special case of the proposed generalized Poisson distribution. The advantage is to give the solution of finding probabilities of the count data where the values of the random variable can take infinite number of sequence including the traditional $0,1,2,\cdots,\infty$.

Theorem 2.3.1: *Generalized Binomial distribution tends to Generalized Binomial Distribution as sample size tens to infinity and probability of success tends to zero.*

Theorem 2.3.2: *The moment generating function of the generalized Poisson distribution and its first four raw and central moments.The moment generating function of the generalized Poisson distribution is*

$$M_X(t)=\frac{1}{\sum_{n=0}^{\infty}\frac{\lambda^{a+nd}}{(a+nd)!}}\sum_{n=0}^{\infty}\frac{(\lambda e^{t})^{a+nd}}{(a+nd)!}$$

Differentiating it with respect to t in first to 4th order and equating $t=0$, we obtain the first four raw moments respectively and which are as follows:

$$\mu_1'=\frac{\sum_{n=0}^{\infty}(a+nd)\frac{\lambda^{a+nd}}{(a+nd)!}}{\sum_{n=0}^{\infty}\frac{\lambda^{a+nd}}{(a+nd)!}},\ \mu_2'=\frac{\sum_{n=0}^{\infty}\{a+nd\}^2\frac{\lambda^{a+nd}}{(a+nd)!}}{\sum_{n=0}^{\infty}\frac{\lambda^{a+nd}}{(a+nd)!}},\ \mu_3'=\frac{\sum_{n=0}^{\infty}(a+nd)^3\frac{\lambda^{a+nd}}{(a+nd)!}}{\sum_{n=0}^{\infty}\frac{\lambda^{a+nd}}{(a+nd)!}},\ \mu_4'=\frac{\sum_{n=0}^{\infty}(a+nd)^4\frac{\lambda^{a+nd}}{(a+nd)!}}{\sum_{n=0}^{\infty}\frac{\lambda^{a+nd}}{(a+nd)!}}.$$

The corresponding central moments can be obtained by using the relationship between the raw and central moments. Consequently, we may find the measures of skewness and kurtosis. The form of the central moments and thus measures of skewness and kurtosis are seem to long and complicated equations and not in a concrete form, but very much simple to calculate the mentioned properties for specific values of *a* and

d. All of the raw moments, central moments and measures of skewness and kurtosis derived from the generalized Poisson distribution tend to the form of traditional Poisson distribution when $a = 0$ and $d = 1$, which again justify that property that the usual Poisson distribution is the special case of the generalized Poisson distribution.

2.4 Proposed Generalized Geometric Distribution

In this section, we discuss about the generalized Geometric variate having relatively more general form of the values represented by an arithmetic progression which is mentioned above in the generalized Poisson distribution case and its distribution and the properties. We have proposed the distribution for the random variable taking values in the form $a + nd$ where a is a non-negative integer representing minimum number of failure, d is a positive integer indicating how the failures are occur, that is concentration of occurrence of failure per trail and $n > 0$ is an integer representing the total number of trails. Clearly, the range of the random variable is similar to that of usual Geometric distribution if $a = 0$ and $d = 1$. As we proposed generalization of traditional Geometric distribution, it is the necessary condition that it reduces to traditional distribution if the random variable posses the range 0 to ∞.

Definition 2.4.1: A random variable X is said to have a generalized geometric distribution if it has the following probability mass function

$$P(x; a, n, d, p) = q^x \frac{(1 - q^d)}{q^a};\ x = a, a + d, a + 2d, \cdots, a + nd$$

where, *a is a non-negative integer representing minimum number of failure, d is a positive integer indicating concentration of failure per trail and $n > 0$ is an integer representing the total number of trails such that $a = 0$ if $n = 0$ and p is the probability of success such that $p + q = 1$ are the parameters of the distribution. The probability function $q^x \frac{(1-q^d)}{q^a}$ provides the probability of getting the d successes following maximum of $\{a + nd\}$ failures in n trials.*

The function in the above Equation is a probability function as it satisfies the following properties of probability function for several values of the parameters.

i) $P(x; a, n, d, p) \geq 0$

ii) $\sum_{x=a}^{a+nd} P(x; a, n, d, p) = 1$

The probability function reduces to pq^x, where $x = 0, 1, 2, \cdots, \infty$ if $a = 0$ and $d = 1$. Thus, the name generalized Geometric and traditional one is considered as the special case of the suggested distribution. The other distributional properties are examined below:

Theorem 2.4.1: *Prove that the moment generating function of generalized Geometric distribution is*

$$M_X(t) = \frac{(1 - q^d)e^{ta}}{[1 - (qe^t)^d]}$$

Theorem 2.4.2: *The first four raw moments of the generalized Geometric distribution are respectively*

$\mu_1' = a + \frac{dq^d}{(1-q^d)}$, $\mu_2' = a^2 + \frac{ad}{(1-q^d)} + \frac{d(a+d)q^d}{(1-q^d)} + \frac{2(dq^d)^2}{(1-q^d)^2}$, $\mu_3' = a^3 + \frac{2a^2dq^d}{(1-q^d)} + \frac{a^2d}{(1-q^d)} + \frac{ad^2}{(1-q^d)} + \frac{2ad^2q^d}{(1-q^d)^2} + \frac{2ad^2q^d}{(1-q^d)} + \frac{d^3q^d}{(1-q^d)} + \frac{4ad^2q^{2d}}{(1-q^d)^2} + \frac{6d^3q^{2d}}{(1-q^d)^2} + \frac{6d^3q^{3d}}{(1-q^d)^3}$, $\mu_4' = a^4 + \frac{a^4dq^d}{(1-q^d)} + \frac{a^3dq^d}{(1-q^d)} + \frac{a^2d^2q^d}{(1-q^d)} + \frac{2a^2d^2q^{2d}}{(1-q^d)^2} + \frac{a^3d}{(1-q^d)} + \frac{2a^2d^2}{(1-q^d)} + \frac{ad^3}{(1-q^d)} + \frac{2a^2d^2q^d}{(1-q^d)^2} + \frac{2ad^3q^d}{(1-q^d)^2} + \frac{a^3dq^d}{(1-q^d)} + \frac{3a^2d^2q^d}{(1-q^d)} + \frac{3ad^3q^d}{(1-q^d)} + \frac{d^4q^d}{(1-q^d)} + \frac{2a^2d^2q^{2d}}{(1-q^d)^2} + \frac{4ad^3q^{2d}}{(1-q^d)^2} + \frac{2d^4q^{2d}}{(1-q^d)^2} + \frac{2a^2d^2q^d}{(1-q^d)^2} + \frac{4ad^3q^d}{(1-q^d)^2} + \frac{6ad^3q^{2d}}{(1-q^d)^3} + \frac{4a^2d^2q^{2d}}{(1-q^d)^2} + \frac{14ad^3q^{2d}}{(1-q^d)^2} + \frac{12d^4q^{2d}}{(1-q^d)^2} + \frac{12ad^3q^{3d}}{(1-q^d)^3} + \frac{18d^4q^{3d}}{(1-q^d)^3} + \frac{6ad^3q^{3d}}{(1-q^d)^3} + \frac{18d^4q^{3d}}{(1-q^d)^3} + \frac{24d^4q^{4d}}{(1-q^d)^4}$

Theorem 2.4.3: *The first four central moments of the generalized Geometric distribution are respectively* $\mu_1 = \mu_1' = a + \frac{dq^d}{(1-q^d)}$, $V(X) = \mu_2 = ad + \frac{d^2q^d}{(1-q^d)^2}$, $\mu_3 = -2a^2d + \frac{ad^2}{(1-q^d)}(1-2q^d) + \frac{d^3q^d}{(1-q^d)^3}(1+q^d)$, $\mu_4 = \frac{a^4dq^d}{(1-q^d)} - \frac{4d^2q^{2d}}{(1-q^d)^2} + \frac{a^3d}{(1-q^d)}(3-4q^d) - \frac{2a^2d^2}{(1-q^d)^2}(1-4q^d+q^{2d}) + \frac{ad^3}{(1-q^d)^2}(1+4q^d-q^{2d}) + \frac{d^4q^d}{(1-q^d)^4}(1+7q^d+q^{2d})$

Theorem 2.4.4: *Find the MLE estimator of the probability of success after x failure of generalized geometric distribution will be* $\hat{p} = 1 - \hat{q} = 1 - \left(\frac{a-x}{a-d-x}\right)^{\frac{1}{d}}$.

Discussion and Conclusion

We have suggested relatively more general form of two discrete distributions such as Binomial and Multinomial for the different sampling scheme which is described above and termed as generalized distribution. It is evident from the generalized distribution that if sampling is drawn in the usual manner, then our suggested distributions reduces to the traditional form and thus it may conclude that the traditional Binomial and Multinomial distribution are the special cases of our proposed generalized Binomial and Multinomial distribution. Like the traditional distributions, all of the distributional properties including limiting theorems of the suggested distributions have derived. The truncated cases of the traditional distribution can be address more accurately by our new approach of the distributions. In general, the new approach of the distributions are providing more access and broaden the scope from the theoretical point of view as well as from the standpoint of real world problem solving. Generalized sequence of the number of success of the proposed distributions may be considered as an added advantage in the distribution theory.

In probability theory and statistics Poisson distribution and Geometric distribution have great importance. In both the distribution, the value of the random variable ranges from 0 to ∞ with a constant increment of 1 for each trail. We proposed the generalized version of both the distribution where the number of occurrences can be expressed by an arithmetic progression $a + nd$. It is shown that the traditional forms of the distributions are the special case of the proposed distributions. Thus, both the generalized Poisson and generalized Geometric distribution can be applied in the cases where their traditional distribution is the only way. In parallel, the proposed distributions facilitates to solving the probability of the certain value of the random variable having infinitely many sequences other than traditional sequence. In this context, the scope of the proposed distribution is much wider than their traditional form. In addition, some of the distributional properties are derived and examined for both of the suggested distributions. Overall, the generalized Poisson and generalized Geometric distributions may play a critical and vital role in the distribution theory and thus in the complicated real life problems.

Reference

Altham, P. M. E. (1978). Two Generalizations of the Binomial Distribution. *Journal of the Royal Statistical Society C,* 27 (2), 162-167.

Altham, P. M. E. and Hankin, R. K. S. (2012). Multivariate Generalizations of the Multiplicative Binomial Distribution: Introducing the MM Package. *Journal of Statistical Software*, 46(12), 1-23.

Consul, P. C. (1989). *Generalized Poisson Distribution: Properties and Applications (Statistics.*

Text Book and Monograms). New York: Marcel Dekker Inc.
Consul, P. C. and Jain, G. C. (1973). A Generalization of the Poisson Distribution. *Technometrics* 15: 791-799.
Drezner, Z. and Farnum, N. (1993). A generalized binomial distribution. *Communications in Statistics - Theory and Methods*, 22(11), 3051-3063.
Dwass, M. (1979). A Generalized Binomial Distribution. *The American Statistician*, 33(2), 86-87.
Fu, J. and Sproule, R. (1995). A generalization of the binomial distribution. *Communications in Statistics - Theory and Methods*, 24(10), 2645-2658.
Gomez-Deniz, E. (2010). Another generalization of the geometric distribution. *Test* 19: 399-415.
Hagmark, P. (2012). An Exceptional Generalization of the Poisson Distribution. *Open Journal of Statistics* 2: 313-318.
Jain, G. C. and Consul, P. C. (1971). A generalized negative binomial distribution. *SIAM Journal of Applied Mathematics* 21: 501-513.
Johnson, N. L., Kemp, A. W. and Kotz, S. (2005). *Univariate Discrete Distributions*. Hoboken: John Wiley & Sons.
Johnson, N. L., Kemp, A. W. and Kotz, S. (2005). *Univariate Discrete Distributions*. 3rd ed. John Wiley & Sons, Hoboken.
Kempton, R. A. (1975). A generalized form of Fisher's logarithmic series. *Biometrika* 62: 28-38.
Lerner, B., Lone, A., and Rao, M. (1997). On generalized Poisson distributions. *Probability and Mathematical Statistics* 17: 377–385.
Lin, X. S. (2004). Generalized Poisson Models and Their Applications in Insurance and Finance. *Journal of the American Statistical Association* 99: 563.
Madsen, R. W. (1993). Generalized binomial distributions. *Communications in Statistics - Theory and Methods*, 22(11), 3065-3086.
Marshall, A. N, and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with applications to the exponential and Weibull families. *Biometrica* 84: 641-652.
Nassar, M. and Nada, M. (2013). A new generalization of the Pareto-geometric distribution. *Journal of the Egyptian Mathematical Society* 21: 148-155.
Nelson, D. L. (1975). Some Remarks on Generalizations of the Negative Binomial and Poisson Distributions. *Technometrics* 17: 135-136.
Panaretos, J. and Xekalaki, E. (1986). On Generalized Binomial and Multinomial Distributions and Their Relation to Generalized Poisson Distribution. *Annals of the Institute of Statistical Mathematics*, 38(2), 223-231.
Paul, A. E. (1978). A Generalized Compound Poisson Model for Consumer Purchase Panel Data Analysis. *Journal of the American Statistical Association* 73: 706-713.
Paul, S. R. (1985). A Three-Parameter Generalization of the Binomial Distribution. *Communications in Statistics - Theory and Methods*, 14(6), 1497-1506.
Philippou, A. N. and Muwafi, A. A. (1982). "Waiting for the kth consecutive success and the Fibonacci sequence of order k. *The Fibonacci Quarterly* 20:28-32.
Philippou, A. N., Georghiou, C. and Philippou, G. N. (1982). A Generalized Geometric Distribution and some of its Properties. *Statistics and Probability Letters* 1: 171-175.
Raul, S. R. (1987). On the beta-correlated binomial (bcb) distribution-a three parameter generalization of the binomial distribution. *Communications in Statistics - Theory and Methods*, 16(5), 1473-1478.
Tripathi, R. C. and Gupta, R. C. (1985). A generalization of the log-series distribution.

Communications in Statistics, *Series A* 14: 1779-1799.
Tripathi, R. C., Gupta, R. C. and White, T. G. (1987). Some Generalizations of the Geometric Distribution. *Sankhya: The Indian Journal of Statistics*, *Series* B 49: 218-223.

Appendix

Proof 2.1.1: We know, the probability mass function of the suggested generalized binomial distribution is

$$P(x;a,n,d,p)=\frac{\binom{a+nd}{x}p^xq^{a+nd-x}}{\sum_{x=a}^{a+nd}\binom{a+nd}{x}p^xq^{a+nd-x}}\ \ ;\ x=a,a+d,a+2d,\cdots,a+nd$$

$P(x;a,n,d,p)\geq 0$ for all values of *X* with different values of a, d, n and p. Again, $\sum_{x=a}^{a+nd}P(x;a,n,d,p)$ $=\sum_{x=a}^{a+nd}\frac{\binom{a+nd}{x}p^xq^{a+nd-x}}{\sum_{x=a}^{a+nd}\binom{a+nd}{x}p^xq^{a+nd-x}}=\frac{1}{\sum_{x=a}^{a+nd}\binom{a+nd}{x}p^xq^{a+nd-x}}\sum_{x=a}^{a+nd}\binom{a+nd}{x}p^xq^{a+nd-x}=1.$

As $P(x;a,n,d,p)\geq 0$ and $\sum_{x=a}^{a+nd}P(x;a,n,d,p)=1$, so we may conclude that $P(x;a,n,d,p)$ is a probability function.

Proof 2.2.1: The form of the probability mass function of generalized Multinomial distribution is

$$P(x_1,x_2,\cdots,x_k;a,n,d,p_1,p_2,\cdots,p_k)=\frac{\frac{1}{x_1!\,x_2!\cdots x_k!}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}}{\sum_{x_1,x_2,\cdots,x_k=a}^{a+nd}\frac{1}{x_1!\,x_2!\cdots x_k!}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}}$$

Considering $k=2$ such that $x_1+x_2=a+nd$ and $p_1+p_2=1$, we obtain

$$P(x_1,x_2;a,n,d,p_1,p_2)=\frac{\frac{(a+nd)!}{x_1!\,x_2!}p_1^{x_1}p_2^{x_2}}{\sum_{x_1,x_2=a}^{a+nd}\frac{(a+nd)!}{x_1!\,x_2!}p_1^{x_1}p_2^{x_2}}$$

$$=>P(x_1;a,n,d,p_1,p_2)=\frac{\frac{(a+nd)!}{x_1!\,(a+nd-x_1)!}p_1^{x_1}p_2^{a+nd-x_1}}{\sum_{x_1=a}^{a+nd}\frac{(a+nd)!}{x_1!\,(a+nd-x_1)!}p_1^{x_1}p_2^{a+nd-x_1}}$$

Letting $x_1=x$, $p_1=p$ and $p_2=1-p_1=1-p=q$, we have

$$P(x;a,n,d,p)=\frac{\frac{(a+nd)!}{x!\,(a+nd-x)!}p^xq^{a+nd-x}}{\sum_{x=a}^{a+nd}\frac{(a+nd)!}{x!\,(a+nd-x)!}p^xq^{a+nd-x}}$$

$$=>P(x;a,n,d,p)=\frac{\binom{a+nd}{x}p^xq^{a+nd-x}}{\sum_{x=a}^{a+nd}\binom{a+nd}{x}p^xq^{a+nd-x}}$$

Hence, $P(x;a,n,d,p)=\frac{\binom{a+nd}{x}p^xq^{a+nd-x}}{\sum_{x=a}^{a+nd}\binom{a+nd}{x}p^xq^{a+nd-x}}\ \ ;\ x=a,a+d,a+2d,\cdots,a+nd$

which is the probability mass function of generalized Binomial distribution.

*Proof 2.3.1***:** The probability mass function of generalized Binomial variate *X* with parameter a, n, d, and p is $P(x;a,n,d,p)=\frac{\binom{a+nd}{x}p^xq^{a+nd-x}}{\sum_{x=a}^{a+nd}\binom{a+nd}{x}p^xq^{a+nd-x}}\quad ;\ x=a,a+d,a+2d,\cdots,a+nd$

Under the following assumptions, the generalized Poisson distribution can be derived from generalized Binomial distribution.

i) p, the probability of success in a Bernoulli trail is very small. i.e. $p \to 0$.
ii) n, the number of trails is very large. i.e. $n \to \infty$.
iii) $(a+nd)p = \lambda$ is finite constant, that is average number of success is finite. Under this condition, we have $(a+nd)p = \lambda \quad \therefore p = \frac{\lambda}{(a+nd)}$ and $q = 1 - \frac{\lambda}{(a+nd)}$.

Proof ***2.4.4:*** The likelihood function of generalized geometric distribution itself is a probability mass function and which is $L = q^x \frac{(1-q^d)}{q^a}$

Taking logarithm in both sides of the equation, we have

$lnL = xlnq + ln(1-q^d) - alnq$

Differentiating equation (2.2.12) with respect to q and equating to zero, we have

$$\frac{\delta}{\delta q}(lnL) = \frac{\delta}{\delta q}\left[xlnq + ln(1-q^d) - alnq\right] = 0$$

$$=> \frac{x}{\hat{q}} - \frac{d\hat{q}^{d-1}}{(1-\hat{q}^d)} - \frac{a}{\hat{q}} = 0$$

$$=> \frac{x(1-\hat{q}^d) - d\hat{q}^d - a(1-\hat{q}^d)}{\hat{q}(1-\hat{q}^d)} = 0$$

$$=> x - x\hat{q}^d - d\hat{q}^d - a + a\hat{q}^d = 0$$

$$=> \hat{q}^d(a-d-x) = a - x$$

$$=> \hat{q}^d = \frac{a-x}{a-d-x}$$

$=> \hat{q} = \left(\frac{a-x}{a-d-x}\right)^{\frac{1}{d}}$ which is the MLE estimator of probability of failure before first success.

Hence, the MLE estimator of probability of success after x failure will be $\hat{p} = 1 - \hat{q} = 1 - \left(\frac{a-x}{a-d-x}\right)^{\frac{1}{d}}$