Chapter 1

Introduction
Chapter 1

Introduction

Discrete probability distributions arise whenever we deal with a problem of random nature involving counting. It is one of the most important basic fields of study in the theory of statistics. Feller (1968) introduced discrete probability distributions of order $k$, when he extended the notion of success to a success run of length $k$. A *Run* is usually defined as an uninterrupted sequence of like symbols, (‘S’ or ‘F’). In a given sequence of independent Bernoulli trials, two possible types of success runs have been discussed:

(i) non-overlapping success runs of length $k$, in which each success contributes to success runs of length $k$ at most once.

(ii) overlapping success runs of length $k$, in which each success can contribute to success runs of length $k$ up to $k$ times (Ling, 1988)
Thus the distributions associated with runs of $k$ like outcomes are distributions of order $k$. The interest on this topic was further upsurged by Phillipou (1983), who initiated the topic of the exact distribution theory of the discrete distributions of order $k$ and introduced the Geometric and Negative Binomial distributions of order $k$, referred to as the Type I Geometric and Negative Binomial distributions of order $k$ by Ling (1988). Further he introduced the closely related Poisson distributions of order $k$. Since it is not easy to calculate the probabilities of the distribution of order $k$ from the definition, Aki et al. (1984) treated the Poisson distribution of order $k$ as a special case of the extended form of Adelson’s (1966) Stuttering Poisson distribution.

Various modifications have also been made on the underlying sequence, e.g., replacing the Bernoulli trials for other random sequences such as some urn models, a binary sequence of order $k$; (Aki (1985), Hirano and Aki (1987), Aki (1992), Aki and Hirano (1994, 1995), Dhar and Jiang (1995), Balakrishnan (1997) and Balakrishnan and Koutras (1998)), Markov dependent trials and higher order Markov dependent trials. Consequently, the class of discrete distributions of order $k$ has been extended and become fruitful. Moreover, some other kinds of extensions are to be examined; adoption of different ways of counting the number of
runs, constructing a multivariate version of the class of the distributions (e.g. Johnson, Kotz and Balakrishnan (1997, ch.42)), and introduction of various waiting time problems.

1.1 Review of Literature

1.1.1 Quasi Binomial Distributions

Consul (1974) has first introduced the notion of urn model with predetermined strategy with a two urn model and developed quasi binomial type I distribution (QBD I) with probability function

\[
\Pr(X = x) = \binom{n}{x} p(p + x\phi)^{x-1}(1 - p - x\phi)^{n-x}, \quad x = 0(1)n, \quad \frac{p}{n} \leq \phi \leq \frac{1-p}{n}
\]

(1.1.1)

where \( p, \phi \) and \( n \) denotes the shape parameters. Mishra and Sinha (1981) observed that (1.1.1) is a probability distribution even for negative values of \( \phi \).

The Quasi binomial type I is used to model Bernoulli trials. The parameter \( p \) denotes the initial probability of success, \( n \) denotes the number of Bernoulli trials, and \( \phi \) denotes how the probability of success increases or decreases with the number of successes. Specifically when \( \phi=0 \), the quasi binomial type I distribution reduces to classical binomial
distribution and when $\phi \neq 0$ the probability of success in the $x^{th}$ trial becomes $p + x\phi$. Das et al. (1998) have used QBD I in migration. Consul and others have studied QBD I extensively and given justification and mentioned applications of these distributions in various fields.

Consul and Mittal (1975) has obtained another type of quasi binomial distribution via urn model using four urn, the probability mass function (pmf) of this distribution is

$$\Pr(X = x) = \binom{n}{x} \frac{p(1-p-n\phi)}{(1-n\phi)} (p+x\phi)^{x-1} (1-p-x\phi)^{n-x-1}, \quad x = 0(1)n,$$

$$-\frac{p}{n} \leq \phi \leq \frac{1-p}{n}; \quad (1.1.2)$$

The pmf (1.1.2) is termed as the quasi binomial distribution of type II (QBD II).

Berg and Mutafchiev (1990) have shown applications of some QBDs and modified QBDs in random mapping problems. Consul (1990) studied properties of (1.1.1) with deductions of moments, inverse moments, maximum likelihood estimation and data fitting.

Using Abel's generalization of Binomial identities (Riordan, 1968)) has proposed a class of QBDs with the probability function (Das (1993, 1994))
Pr(X = x) = \binom{n}{x} \frac{(p + x\phi)^{rs}(q + (n - x)\phi)^{n-x\phi}}{B_n(p, q; s, t; \phi)}, x = 0(1)n \tag{1.1.3} 

where s and t are integers, \( p + q + n\phi = 1 \) and 

\[ B_n(p, q; s, t; \phi) = \sum_{x=0}^{n} \binom{n}{x} (p + x\phi)^{rs}(q + (n - x)\phi)^{n-x\phi} \tag{1.1.4} \]

Chakraborty (2001) and Chakraborty and Das (2006) have derived a class of weighted quasi-binomial distributions as weighted distribution of QBDs including the class of QBDs. Also the moments, inverse moments, recurrence relations among moments, bounds for mode, problem of estimation and fitting of data from real life situations using different methods and limiting distributions of the class have also been studied. Consul’s (Comm. Statist. Theory Methods 19(2) (1990) 477-504) results on QBD I have been found out as particular cases.

1.1.2 A Family of Abel Series Distributions

A shift invariant operator of the form \( Q = DT \), where \( Q \) is a unital operator, is said to be a delta operator. We write \( Q = Df(D) \).

To every delta operator \( Q \) there exists a unique sequence of polynomials \( p_0(x), p_1(x), \ldots \) such that \( Qp_n(x) = np_{n-1}(x) \) and \( p_n(0) = 0 \) for \( n > 0 \). Such a sequence is called basic sequence of the delta operator \( Q \). A
sequence of polynomials is said to be of *binomial type* whenever

\[ P_0(x) = 1, \quad P_1(x) = x \quad \text{and} \quad P_n(x + a) = \sum_{i} \binom{n}{i} P_i(x) P_{n-i}(a). \]

Every basic sequence is of binomial type. Conversely, every sequence of polynomials \( P_i(x) \) of binomial type such as \( P_0(x) = 1 \) and \( P_1(x) = x \) is the basic sequence for a unique delta operator. The most important sequence of binomial type is the sequence of Abel polynomials, namely, the sequence \( P_n(x) = x(x + na)^{n-1} \) for \( a \in \mathbb{Q} \). The sequence of Abel polynomial is the basic sequence associated to the operator \( De^{-aD} \). Because of the importance of this fact, we review the proof, which amounts to the following computation:

\[
De^{-aD} x(x + na)^{n-1} = (x + (n - 1)a)^{n-1} + (n - 1)(x - a)(x + (n - 1)a)^{n-2}
\]

\[
= (x + (n - 1)a + (n - 1)(x - a))(x + (n - 1)a)^{n-2}.
\]

\[
= nx(x + (n - 1)a)^{n-2}.
\]

It follows from the above verification that the sequence of Abel polynomials is a sequence of polynomials of binomial type. The identity stating that the sequence of polynomials \( P_n(x) = x(x + na)^{n-1} \) is of binomial type is due to Abel.

A polynomial \( A_n(x; a) \) given by the associated Sheffer
Chapter 1. Introduction

sequence with \( f(t) = te^{at} \) given by,

\[ A_n(x; a) = x(x - na)^{n-1} \]

Then the generating function is

\[ \sum_{k=0}^{\infty} \frac{A_k(x; a)}{k!} = e^{xW(a)/a} \]

where \( W(x) \) is the Lambert \( W \)-function. Then the associated binomial identity is

\[ (x + y)(x + y - an)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} xy(x - ak)^{k-1}[y - a(n - k)]^{n-k-1} \]

where \( \binom{n}{k} \) is a binomial coefficient, a formula originally due to Abel.

The first few Abel Polynomials are

\[
\begin{align*}
A_0(x; a) & = 1 \\
A_1(x; a) & = x \\
A_2(x; a) & = x(x - 2a) \\
A_3(x; a) & = x(x - 3a)^2 \\
A_4(x; a) & = x(x - 4a)^3
\end{align*}
\]

Nandi and Das (1994) have proposed a simple series distribution called as the Abel series distribution (ASD). They have used the expansion of suitable functions in power series to obtain various distributions from the Abel series distribution. Using suitable functions in
Abel series distribution, they have derived some new distributions such as the Quasi-logarithmic series distribution and a variant of quasi-negative Binomial distribution. From the variant they have obtained quasi-geometric distribution (QGD) as a particular case. Further they have obtained some well known distributions, viz., quasi-binomial distribution of type I (QBD I), quasi-binomial distribution of type II (QBD II), Generalized Poisson distribution (GPD), etc. from ASD. Some inferential properties as well as problem of estimation of parameters of underlying distributions were obtained.

By considering the Abel Polynomials introduced by Comtet (1974, p. 128) Nandi and Das (1994) defined the probability function of Abel series distribution family as

\[
p(k) = a(a-kz)^{k-1} \beta(k, z) / f(a)
\]  

(1.1.5)

where \( f(a) = \sum_{k \geq 0} a(a-kz)^{k-1} \beta(k, z) \)

and \( \beta(k, z) = \left( \frac{1}{k!} \right) \left[ D^k f(a) \right]_{a=zk} \)

In particular, if \( z = 0 \), then (1.1.5) reduces to the common power series distribution.
1.1.3 A Family of Multivariate Abel Series Distributions

On passing from univariate to multivariate distributions, some essentially new features require attention. These are connected with relationships among sets of random variables and include regression, correlation, and more generally conditional distributions.

Nandi and Das (1996), considering the multivariate Abel polynomials, have defined the probability function of the family of multivariate Abel series distribution as,

\[ p(k) = p(k; a, z) = \frac{\prod_{i=1}^{m} a_i (a_i - k, z)^{k_i} / k_i! \left[ D^{k}f(a) \right]_{a=k} }{f(a)} \]

where \( k = k_1 + k_2 + k_3 + \cdots + k_m \), each \( k_i \) being a non-negative integer, \( a_i (i = 1(1)m) \) and \( z \) are parameters and \( f(a) \) is given by

\[ f(a) = f(a, z) = \sum_k \left[ \prod_{i=1}^{m} a_i (a_i - k, z)^{k_i} / k_i! \left[ D^{k}f(a) \right]_{a=k} \right] \]

where \( \sum_k \) denotes the sum over all the partition of the number \( k \) such that \( k = k_1 + k_2 + k_3 + \cdots + k_m \), each \( k_i \) being a non-negative integer and the factor

\[ \beta(k, z) = \left[ D^{k}f(a) \right]_{a=k} \left( \prod_{i=1}^{m} k_i! \right) \]

is independent of \( a \), which is always greater than zero. The domain
a = (a₁, ..., aₙ) is a subspace of an m-dimensional parameter space subject to restrictions aᵢ ≥ 0, if z ≤ 0, and (aᵢ - k, z) ≥ 0, if z ≥ 0, z belonging to a suitable subset of real numbers.

From MASDs, a new distribution called quasi multivariate logarithmic series distribution was obtained by them, also a variant of quasi negative multinomial distribution was found out.

### 1.1.4 A Class of Inflated Generalized Poisson distribution

Nandi et al. (1999) has defined a class of generalized Poisson distribution having the pmf

\[
p(k; a, s, \lambda) = \frac{(a + k \lambda)^{k+s} e^{-k\lambda}}{k! K(a; s; \lambda)}, \quad a + k \lambda > 0, \quad k = 0, 1, 2, \ldots;
\]

\[
0 < \left(\frac{\lambda + a}{k + 1}\right) \left(1 + \frac{\lambda}{a + k \lambda}\right)^{k+s} e^{-k \lambda} < 1, \quad \text{for large } k
\]

\[
= 0, \quad \text{elsewhere}
\]

\[
(1.1.6)
\]

where

\[
K(a; s; \lambda) = \sum_{k \geq 0} \frac{(a + k \lambda)^{k+s} e^{-k\lambda}}{k!}.
\]

Chakraborty (2001) has found out the probability generating function (pgf) of a class of generalized Poisson distribution as

\[
G(t) = \frac{K(at; s; \lambda t)}{t^i K(a; s; \lambda)}
\]

\[
(1.1.7)
\]
where $t = u e^{\lambda(t-1)}$ and $K(a;s;\lambda)$ is defined in (1.1.6).

Dey and Das (2004) have defined a class of Inflated Generalized Poisson distribution whose probability function is as follows

$$\Pr(X = 0) = w + (1 - w) \frac{a^t}{K(a;s;\lambda)}$$

$$\Pr(X = x) = (1 - w) (a + x\lambda)^{x-1} \frac{e^{-x\lambda}}{x!K(a;s;\lambda)} , \ x \geq 1$$

(1.1.8)

where $K(a;s;\lambda)$ is defined in (1.1.6).

The probability generating function of a class of inflated generalized Poisson distribution is given by

$$H(u) = w + (1 - w) \frac{K(at;s;\lambda t)}{t^t K(a;s;\lambda)}$$

(1.1.9)

Putting different valued of $s$ in (1.1.8) and (1.1.9) they obtained different forms of inflated GPDs and their pgf's.

1.2 Genesis of the distributions of order $k$

Feller (1968) introduced discrete distributions of order $k$ when he extended the notion of success to a success run of length $k$. The distribution of the number of trials until the first occurrence of consecutive $k$ successes in Bernoulli trials with success probability $p$ is
called geometric distribution of order $k$. This definition is due to Philippou, Georghiou and Philippou (1983). Though the distribution seems to have been interested since De Moivre (1667-1754)'s works (Todhunter (1965), Feller (1968), Johnson, Kotz and Kemp (1992, pp. 426-432) and Balakrishnan and Koutras (2002)), Philippou et al. (1983)'s contribution in this field seems to be very important. Since Philippou et al. (1983) called the distribution a geometric distribution of order $k$ and defined a Poisson distribution of order $k$ and a negative binomial distribution of order $k$, attention has been paid to interrelationships among the so called discrete distribution of order $k$ and exact distribution theory has been developed extensively by many researchers.

From early 80's, reliability theory of the consecutive-$k$-out-of-$n$: $F$ systems began to be studied (Kontoleon (1980), Chiang and Niu (1981) and Derman, Liberman and Ross (1982)), and a large number of studies have been made as relationships between the reliability of the system and discrete distributions of order $k$ have been clear (e.g. Lambiris and Papastavridis (1985), Fu (1985, 1986a, 1986b), Fu and Beihua (1987), Aki (1985), Hirano (1986), Philippou (1986), Papastavridis (1987), Aki and Hirano (1988, 1989, 1996, 1997), Chrysaphinou and Papastavridis (1990c), Fu and Koutras (1994a, b), Godbole (1990a, 1993), Griffith
Various modifications have also been made on the underlying sequence, e.g., replacing the Bernoulli trials for other random sequences such as some urn models, a binary sequence of order \( k \) (Aki (1985), Hirano and Aki (1987), Aki (1992), Aki and Hirano (1994, 1995), Dhar and Jiang (1995), Balakrishnan (1997) and Balakrishnan and Koutras (1998)), Markov dependent trials and higher order Markov dependent trials. Consequently, the class of discrete distributions of order \( k \) has been extended and has become fruitful. Moreover, some other kinds of extensions are to be examined; adoption of different ways of counting the number of runs, constructing a multivariate version of the class of distributions (e.g. Johnson, Kotz and Balakrishnan (1997, Chapter 42)), and introduction of various waiting time problems.

### 1.2.1 Discrete distributions of order \( k \)

First, we will pay attention to the following relationships among the binomial distribution \( B(n, p) \), the negative binomial distribution \( NB(r, p) \), the geometric distribution \( G(p) \), the Poisson distribution \( P(\lambda) \), and the logarithmic series distribution \( LS(p) \). The binomial distribution
is related to the negative binomial distribution by the notion of inverse sampling, i.e., \( \sum_{x=1}^{\infty} NB(r, p; x) = \sum_{x=1}^{\infty} B(n, p; x) \).

By setting \( r = 1 \), in the negative binomial distribution \( NB(r, p) \), we get the geometric distribution \( G(p) \). The \( r \)-th convolution power of the geometric distribution \( G(p) \) becomes the negative binomial distribution \( NB(r, p) \). Setting \( p \to 1, r \to \infty \) and \( r(1 - p) \to \lambda \) in the negative binomial distribution \( NB(r, p) \) yields the Poisson distribution \( P(\lambda) \). By compounding a Poisson distribution, we can get a negative binomial distribution; the conditional distribution of \( X \) given \( Y = y \) is a Poisson distribution \( P(\lambda y) \), and \( Y \) follows the gamma distribution with density \( \frac{\alpha'}{\Gamma(r)} y^{r-1}e^{-\alpha y} \), then \( X \) is distributed as \( NB(r, p) \), where \( q = 1 - p = \frac{\lambda}{\lambda + \alpha} \). When \( X \) follows the negative binomial distribution \( NB(r, p) \), the conditional probability \( P(X = x \mid X > 0) \) converges to \( LS(p; x) \) as \( r \to 0 \). By generalizing the Poisson distribution \( P(-r \log p) \) with the generalizer \( LS(p) \), we obtain the negative binomial distribution \( NB(r, p) \).

It is to be noted here that the geometric \( G(p) \), the binomial \( B(n, p) \),
and the negative binomial $NB(r, p)$ distributions are based on independent Bernoulli trials with success probability $p$. By changing the role of success for that of a success run of length $k$, we can obtain naturally a geometric distribution of order $k$, $G_k(p)$, a binomial distribution of order $k$, $B_k(n, p)$, and a negative binomial distribution of order $k$, $NB_k(r, p)$. By defining a Poisson distribution of order $k$, $P_k(p)$ and a logarithmic series distribution of order $k$, $LS_k(p)$, appropriately, corresponding relationships to the above among distributions of order $k$ can be obtained (Philippou et al. (1983) and Aki, Kuboki and Hirano (1984)). Aki (1985) has defined a binary sequence of order $k$ as an extension of a sequence of Bernoulli trials. Based on the sequence, the extended geometric distribution of order $k$, $EG_k(p_1, p_2, ..., p_k)$, the extended binomial distribution of order $k$, $EB_k(n, p_1, p_2, ..., p_k)$, the extended negative binomial distribution of order $k$, $ENB_k(r, p_1, p_2, ..., p_k)$, the extended Poisson distribution of order $k$, $EP_k(\lambda_1, \lambda_2, ..., \lambda_k)$ and the extended logarithmic series distribution of order $k$, $ELS_k(p_1, p_2, ..., p_k)$ are defined and still retain the corresponding relationship to the above. These distributions are equivalent to the multiparameter distributions of order $k$ defined by Philippou (1988), such as $NB_k(r, q_1, q_2, ..., q_k)$, $P_k(\lambda_1, \lambda_2, ..., \lambda_k)$ and $LS_k(p_1, p_2, ..., p_k)$ in the sense that
they can be obtained by changing the parameters appropriately from the corresponding distributions. We also note that the cluster negative binomial distribution defined by Xekalaki and Panaretos (1989) based on an urn model with replacement is equivalent to the extended negative binomial distribution of order $k$; a genesis scheme of the multiparameter (or cluster) negative binomial distribution of order $k$ ($NB_k(r,q_1,q_2,\ldots,q_k)$) is the following: An urn contains balls bearing the letters $F_1,\ldots,F_k$ and $S(=S_0)$ with respective proportions $q_1,q_2,\ldots,q_k$ and $p$ ($0 < q_i < 1$ for $1 \leq i \leq k$, $q_1 + q_2 + \ldots + q_k < 1$ and $q_1 + q_2 + \ldots + q_k + p = 1$). Balls are drawn from the urn with replacement until $r$ balls ($r \geq 1$) bearing the letter $S$ appear. Then the distribution of the sum of indices of the letters on the balls drawn is $NB_k(r,q_1,q_2,\ldots,q_k)$. The sampling scheme called a cluster sampling can be seen as the $\{S,F\}$-trials until the $r$-th occurrence of consecutive $k$ successes by replacing the event $F_i$ for $S_1\ldots S F$, $i = 1,\ldots,k$, and $S_0$ for $S_1\ldots S$. Distributions of order $k$ and extended distributions of order $k$ are further extended to multivariate distributions by Philippou, Antzoulakos and Tripsiannis (1989), Philippou and Antzoulakos (1990), Philippou, Antzoulakos and Tripsiannis (1990), Antzoulakos and Philippou (1994)
1.2.2 Discrete distribution related to succession events

A geometric distribution of order $k$ is a distribution of waiting time (number of trials) until the first occurrence of consecutive $k$ successes in Bernoulli trials. A good deal of effort has been made on waiting times for the first occurrence of a given pattern or word in independent or dependent trials (e.g. Rajarshi (1974), Li (1980), Gerber and Li (1981), Chrysaphinon and Papastavridis (1990a, b), Chrysaphinon et al. (1994) and Mori (1991)). Ebneshahrashoob and Sobel (1990) considered two succession events, a success run of length $k$ and a failure run of length $r$ and derived the exact distribution of the sooner waiting time (the number of trials until the first occurrence of one of the two events) and that of the later waiting time (the number of trials until the first occurrence of the both events). Aki (1992) generalized this problem to the case of a sequence of independent nonnegative-integer-valued random variables and obtained the exact solution of the $n$-th waiting time problem. Aki and Hirano (1993) and Balasubramanium, Viveros and Balakrishnan (1993) solved the sooner and later problem in the sequence of a Markov chain. Ling (1992) and Sobel and Ebneshahrashoob (1992) developed the
sooner and later problem in the case of a frequency quota. Let $F_0$ be the event that $l$ runs of "0" of length $r$ occur and let $F_1$ be the event that $m$ runs of "1" of length $k$ occur in the sequence $X_1, X_2, \ldots$. Uchida and Aki (1995) obtained recurrence relations of the probability generating functions of the distributions of the waiting time for the sooner and later occurring events between $F_0$ and $F_1$. The waiting time to finish a volleyball game is a special example of the distributions. Aki, Balakrishnan and Mohanty (1996) studied the sooner and later waiting time problems based on higher order Markov dependent trials. Aki (1997) studied the distribution of the number of occurrences of the sooner event until the first occurrence of the later event based on a two-state Markov dependent trial and generalized the problem to a multivariate case. Koutras (1997a) investigated relationships between the number of appearances of a pattern and the waiting time of the $r$-th occurrence of the pattern in terms of their generating functions. Koutras (1997b) developed a general technique for the evaluation of the exact distribution in a wide class of waiting time problems. Koutras and Alexandrou (1997b) investigated the sooner waiting time problems in a sequence of trinary trials. Considering some enumeration schemes of the number of runs, we
can derive more fruitful results. Feller (1968, chapter XIII) defined a way of counting the number of runs exactly of length \( k \) as counting the number from scratch every time a run occurs. For example, the sequence $SSS \mid SFSSS \mid SSS \mid F$ contains 3 success runs of length 3. Though Feller’s enumeration scheme is convenient for applying renewal theory, some other enumeration schemes are also interested from the practical point of view. In the classical literature, a “success run of length \( k \)” meant an uninterrupted sequence of either exactly \( k \), or of at least \( k \) successes (Feller (1968)). In this way of counting the number of runs of length 3 (or more), the above example $SSSS \mid FSSSSSS \mid F$ contains 2 success runs of length 3 (or more). In Goldstein (1990) a Poisson approximation of the distribution of the number of success runs of length \( k \) or more until the \( n \)-th trial was proposed. Schwager (1983) and Ling (1988, 1989) studied the distributions on the number of overlapping runs of length \( k \). In this enumeration scheme, $SSSSFSSSSSSSF$ contains 6 overlapping success runs. Ling (1988) obtained a recurrence relation for the pgf of the number of overlapping success runs of length \( k \) until the \( n \)-th trial in independent trials with success probability \( p \). The distribution is called the type II binomial distribution of order \( k \) ($B_i^u(n,p)$). Hirano et al. (1991) obtained the exact pgf of $B_i^u(n,p)$ and its probability function. Ling (1989) and...
Hirano et al. (1991) studied the distribution of the number of trials until the \( r \)-th overlapping success run occurs. The distribution is called the type III negative binomial distribution of order \( k \) \( (NB'''_k(r, p)) \). Hirano and Aki (1993) derived the pgf's of the distributions of the number of overlapping success runs of length \( k \) and success runs of length \( k \) or more until the \( n \)-th trial in the sequence of \( \{0,1\} \)-valued Markov chain. We have to note here that the probability that a success run of length \( k \) does not occur coincides with each other for the above three enumeration schemes. When we investigate the distribution of runs by using a probability generating function, the distribution of the number of success runs of length \( k \) or more becomes surprisingly simple, and as a byproduct the elegant result of Lambiris and Papastavridis (1985) on the reliability of a consecutive-\( k \)-out-of-\( n:F \) system is derived simply and alternatively (without using enumerative combinatorics) (Hirano and Aki (1993)).

It has been observed that the reliability of a consecutive-\( k \)-out-of-\( n:F \) system can be obtained as an application of the distribution theory of runs (e.g. Hirano (1994) and Chao, Fu and Koutras (1995)). The results of Goldstein (1990) have application to analysis of DNA data.

Shmueli and Cohen (2000) applied exact probabilities of \( G_k(p) \),
Further we have the following important application called start-up demonstration tests; a start-up demonstration test is a mechanism by which a vendor demonstrates to a customer the reliability of an equipment with regard to its starting. The vendor repeats start-ups of the equipment, such as lawn mowers, chain saws, water pumps and outboard motors until a specified number of consecutive successful start-ups are observed. The pioneering work of the problem is Hahn and Gage (1983). Since Hahn and Gage (1983) regarded the sequence of start-ups as an independent trial with success probability $p$ for simplicity, the number of attempted start-ups until the item is accepted follows the geometric distribution of order $k$, where $k$ is the required number of consecutive successful start-ups to achieve acceptance. Viveros and Balakrishnan (1993) studied statistical inference from start-up demonstration test data and introduced a Markov dependence model. Balakrishnan, Balasubramaniam and Viveros (1995) introduced corrective action models and Balakrishnan, Mohanty and Aki (1997) gave the joint probability generating function of some random variables appearing in the Markov dependence model of the start-up demonstration test with corrective actions.
Though these results on consecutive systems and start-up demonstration tests have been developed independently from theory of distributions of order $k$, the results on statistical inference are related deeply to Philippou, Georghiou and Philippou (1983) and Aki and Hirano (1989). And the distribution of number of start-ups (or trials) based on Markov dependent trials is related to the results of Aki (1985), Aki and Hirano (1993), Balasubramanium, Viveros and Balakrisnan (1993), Aki and Hirano (1994, 1995), Mohanty (1994), Uchida and Aki (1995) and Aki, Balakrishnan and Mohanty (1996).

In addition to the above applications, we have to mention the moving window detection procedure for discrete data. A consecutive-$k$-out-of $-r$-from-$n$:F system is a sequence of $n$ linearly-ordered components such that the system fails if and only if there are $r$ consecutive components from which at least $k$ are failed. Since it includes a consecutive-$k$-out-of $-n$:F system (in the case of $r = k$) and $k$-out-of $-n$:F system (in the case of $n = r$), the system is very general one (Griffith (1986), Papastavridis and Sfakianakis (1991), Sfakianakis, Kounius and Hillaris (1992), Papastavridis and Koutras (1993) and Cai (1994)). To calculate the reliability of the system, it suffices to derive the distribution of the waiting time for the first observation of the window
(consecutive trials) of length \( r \) which has at least \( k \) failures in independent trials with success probability \( p \). The problem is called the moving window detection procedure and has been studied in the theory of radar detection, time sharing systems and quality control (Greenberg (1970), Saperstein (1973), Nelson (1978), Mirstik (1978) and Glaz (1983)). The moving window detection procedure has been extended in more general situations and they are treated and studied as the scan statistics (Glaz (1989), Glaz and Naus (1991), Koutras and Alexandron (1995) and Glaz and Balakrishnan (1998)). By putting together waiting time problems and the enumeration schemes of the number of runs, more interesting problems can be solved. For example the following results were obtained.

For simplicity, we consider only the case of Bernoulli trials. The distribution of the number of failures until the first occurrence of consecutive \( k \) successes is a Geometric distribution (of order 1), \( G(p^k) \).

The number of successes until the first occurrence of consecutive \( k \) successes is a shifted geometric distribution of order \((k-1)\), \( G_{k-1}(p,k) \).

Similarly, the number of overlapping success runs of length \( l \) until the first occurrence of consecutive \( k \) successes becomes a shifted geometric distribution of order \((k-1)\), \( G_{k-1}(p,k-l+1) \). Aki and Hirano (1994) studied the above results based on a Markov chain. Aki and Hirano
(1995) derived the joint distribution of the number of failures, successes and success runs of length less than $k$ until the first occurrence of consecutive $k$ successes in a two state Markov chain. Hirano, Aki and Uchida (1997) studied the problem based on the $m$-th order Markov dependent trials and showed that the distributions of the success runs of length $l$ in the cases of $m \leq l < k$ and $m > l$ are completely different.

There are not so many papers which treat statistical inference of parameters of the above distributions, since the probability functions are too complicated. It is not so easy to obtain maximum likelihood estimate of a parameter of the complicated distribution like above even by using computers. We can mention Douglas (1995) and Shumway and Gurland (1960a, 1960b) for estimating parameters of some special generalized Poisson distributions such as Naymann type A contagious distribution, Poisson binomial distribution and Poisson Pascal distribution. Philippou et al. (1983) proposed moment estimation of the parameters of $G_s(p)$, $NB_k(p)$ and $P_k(\lambda)$. Aki and Hirano (1989) discussed methods for obtaining the maximum likelihood estimates of the parameters of discrete distributions of order $k$ by using the Newton-Raphson method and recurrence relations of derivatives of probabilities of the distributions. After some numerical calculations, they also showed that the asymptotic
efficiencies of the moment estimators are very close to one for given parameters which belong to relatively wide ranges. Aki (1990) treated a method for obtaining the m.l.e. of a two parameter model; as an application simultaneous estimation of the parameters $r$ and $p$ in the negative binomial distribution of order $k$. Balakrishnan, Balasubramanium and Viveros (1995) discussed statistical inference for the probability $p$ of a successful start-up based on demonstration test data. Aki and Hirano (1996) gave the distribution of the lifetime of consecutive-$k$-out-of-$n$: $F$ system with i.i.d. components as a mixture of the distributions of order statistics of $n$ observations.

1.2.3 Some approaches to the distribution theory

Here we have studied some approaches of discrete distribution theory. It is seen that the standard method is enumerative combinatorics, that is to give all typical sequences and, by splitting them into subsequences which can be interpreted. This method is useful when trials are independent or nearly independent (e.g. when the sequence of trials forms a Markov chain), it becomes even difficult even to write down all the typical sequences (Aki, Balakrishnan and Mohanty (1996)). There are further two major approaches to solving recent complicated problems.
One is to give recurrence relations of the conditional probability generating functions of the desired random variable by considering the condition of one-step ahead from every condition. In some cases (e.g. when the system of recurrence relations is linear) the system of equations can be solved generally with respect to the conditional probability generating functions. Even if the system of equations can not be solved generally, in many cases the conditional probability generating functions can be written explicitly if the number of trials and the length of a run are specified (e.g., Ebneshahrashoob and Sobel (1990), Aki (1992), Aki and Hirano (1995), Uchida and Aki (1995), Aki, Balakrishnan and Mohanty (1996), Balakrishnan, Mohanty and Aki (1996), Hirano, Aki and Uchida (1997), Aki (1997), Han, Q. and Aki (1998, 2000a), Han, S. I. and Aki (2000), and Inoue and Aki (2001,2002a)). The other effective approach is so called a Markov chain imbedding method, which is to imbed the sequence of trials into a Markov chain with an appropriate state space. Then, the desired probability can be obtained by multiplying transition probability matrices (Fu and Koutras (1994b), Fu (1996), Koutras and Alexandrou (1995), Koutras (1997b), Koutras and Alexandrou (1997a,b), Han, Q. and Aki (1999, 2000b), Inoue and Aki (2002b) and Balakrishnan and Koutras (2002)).
Recently, Stefanov and Pakes (1997) presented a new approach for constructing joint generating function for quantities of interest associated with pattern formation in binary sequences. The method is based on exponential family of Markov chains.

Among the above approaches, the method of conditional generating functions seems to be useful for investigating more complicated situation in future. By developing the method, Aki (1999) derived the distribution of numbers of “1”-run of a specified length in a directed tree whose vertices are regarded to be \{0,1\}-valued random variables with a directed Markov distribution. The resulting system of equations of conditional probability generating functions is no longer linear if the directed tree is not a sequence. Hence, it may be difficult to apply the Markov chain imbedding method to the problem. Aki and Hirano (1999) have treated sooner and later waiting time problems for runs in Markov dependent bivariate trials by using the method of conditional probability generating functions.

1.2.4 Definitions of some distributions of order $k$ and related concepts

1.2.4.1 Geometric distribution of order $k$

*(Philippou et al., 1983)*
A random variable ‘X’ is said to have the Geometric distribution of order $k$ with parameter $p$, to be denoted by $G_k(p)$, if

$$P(X = x) = \sum_{x_1, x_2, \ldots, x_k} \left( \frac{x_1 + x_2 + \ldots + x_k}{x_1, x_2, \ldots, x_k} \right) p^{x_1} \left( \frac{q}{p} \right)^{x_2 + \ldots + x_k}; x \geq k$$

where the summation is over a non-negative integers $x_1, x_2, \ldots, x_k$, such that $x_1 + 2x_2 + \ldots + kx_k = x - k$ and $q = 1 - p$.

1.2.4.2 Negative binomial distribution of order $k$

(Philippou et al., 1983)

A random variable ‘X’ is said to have a Negative binomial distribution of order $k$, with parameters $r$ and $p$, denoted by $NB_k(r, p)$, if

$$P(X = x) = \sum_{x_1, x_2, \ldots, x_k} \left( \frac{x_1 + x_2 + \ldots + x_k + r - 1}{x_1, x_2, \ldots, x_k, r - 1} \right) p^{x_1} \left( \frac{q}{p} \right)^{x_2 + \ldots + x_k}; x \geq kr$$

where the summation is over all non-negative integers $x_1, x_2, \ldots, x_k$ such that $x_1 + 2x_2 + \ldots + kx_k = x - kr$

This is the distribution of sum of ‘$r$’ i.i.d. random variables distributed as $G_k(p)$. 
1.2.4.3 Poisson distribution of order $k$

*(Philippou et al., 1983)*

A random variable ‘$X$’ is said to have the Poisson distribution of order $k$ with parameter $\lambda$, denoted by $P_k(\lambda)$, if

$$P(X = x) = e^{-\lambda} \frac{\lambda^{x_1 + x_2 + \cdots + x_k}}{x_1! x_2! \cdots x_k!}; x = 0, 1, 2, ...$$

where the summation is over all the non-negative integers $x_1, x_2, ..., x_k$, such that $x_1 + 2x_2 + \cdots + kx_k = x$.

1.2.4.4 Logarithmic Series distribution of order $k$

*(Aki, Kuboki & Hirano, 1984)*

A random variable ‘$X$’ is said to have Logarithmic Series distribution of order $k$ with parameter ‘$p$’, denoted by $LS_k(p)$, if

$$P(X = x) = \sum \frac{(x_1 + x_2 + \cdots + x_k - 1)!}{(-k \log p) x_1! x_2! \cdots x_k!} \left(\frac{q}{p}\right)^{x_1 + x_2 + \cdots + x_k}; x = 1, 2, 3, ...$$

where the summation is over all the non-negative integers $x_1, x_2, ..., x_k$, such that $x_1 + 2x_2 + \cdots + kx_k = x$.

1.2.4.5 Binomial distribution of order $k$
1.2.4.5.1 Type-I Binomial distribution of order k
(Using non-overlapping counting)
(Hirano, 1984 and Philippou and Makri, 1986)

A random variable ‘X’ is said to have the Type-I Binomial
distribution of order k with parameters \((n, p)\), \(0<p<1\), denoted by
\(B_{k,I}(n,p)\), if it has the pmf

\[
P(X = x) = \sum_{i=0}^{\frac{k-1}{k}} \sum_{x_1,x_2,...,x_k} \binom{x_1 + x_2 + ... + x_k + x}{x_1,x_2,...,x_k,x} p^x q^{k-x} \; ; x = 0,1,...,\left\lfloor \frac{n}{k} \right\rfloor
\]

where the inner summation is over non-negative integers \(x_1,x_2,...,x_k\), such
that \(x_1 + 2x_2 + ... + kx_k = x\), \(q = 1-p\) and \([x]\) denotes the greatest the greatest
integer less than or equal to \(x\) for any real number \(x\).

1.2.4.5.2 Type-II Binomial distribution of order k
(Using greater than or equal counting)

A random variable ‘X’ is said to have the Type-II Binomial
distribution of order k, denoted by \(B_{k,II}(n,p)\), and the probability mass
function is given by

\[
P(X = x) = \sum_{i=0}^{\frac{k-1}{k}} \sum_{x_1,x_2,...,x_n} \binom{x_1 + x_2 + ... + x_n}{x_1,x_2,...,x_n} p^x q^{n-x} \; ; x = 0,1,...,\left\lfloor \frac{n}{k} \right\rfloor
\]

where \(\sum_{x}\) is over all non-negative integers \(x_1,x_2,...,x_n\) satisfying
\[ \sum_{j=1}^{n} jx_j = n-i \text{ and } \sum_{j=k+1}^{n} x_j = x; \]

and \( \sum \) is over all non-negative integers \( x_1, x_2, \ldots, x_n \) satisfying

\[ \sum_{j=1}^{n} jx_j = n-i \text{ and } \sum_{j=k+1}^{n} x_j = x-1 \]

### 1.2.4.5.3 Type-III Binomial distribution of order \( k \)

(Using overlapping counting)

A random variable ‘\( X \)’ is said to have the Type-III Binomial distribution of order \( k \), denoted by \( B_{k,iii}(n,p) \), and the probability mass function is

\[ P(X = x) = \sum_{i=1}^{x} \sum_{x_i, x_2, \ldots, x_n} \left( \begin{array}{c} x_i + x_2 + \cdots + x_n \\ x_1, x_2, \ldots, x_n \end{array} \right) p^n \left( \frac{q}{p} \right)^{y_{1}y_{2}\cdots y_{n}} \]

where \( \sum \) is over all non-negative integers \( x_1, x_2, \ldots, x_n \) satisfying

\[ \sum_{j=1}^{n} jx_j = n-i \text{ and } \sum_{j=k+1}^{n} (j-k)x_j = x - \max(0, i-k+1) \]

### 1.2.5 Concept of Run

A run is a specified sequence of outcomes that may occur at some point in the series of trials.

For example, (i) let \( R_i = (1, 1, 1, 1, 1) = 11111 \)
and \( R_2 = (1, 1, 0, 0, 1, 1) = 110011 \)

If '1' denotes a success and '0' denotes a failure, then \( R_1 \) will be called "success runs", whereas \( R_2 \) will be called as "success-failure runs".

(ii) Let \( n \) and \( k \) be fixed positive integers such that \( n \geq k \).

Now consider a sequence for \( n = 14 \) and \( k = 3 \)

The sequence contains

(i) 2 success runs one each of lengths 6, 4 and 1 [Mood (1940)]

\[
\underbrace{SSSSSS}\, FF \underbrace{SSSS\, S}_{6\, 4\, 1}
\]

(ii) 3 success runs of length 3 [Feller (1968)]

\[
\underbrace{SSS}\, SFS \underbrace{SSS}\, SFS
\]

(iii) 6 overlapping success runs of length 3 [Ling (1998)]

\[
\underbrace{SSSSSS}\, FF \underbrace{SSS}\, FS
\]

(iv) 2 success runs of length 3 or more [Goldstein (1990)]

\[
\underbrace{SSSSSS}\, FF \underbrace{SSSS}\, FS
\]
1.2.5.1 Non-overlapping and overlapping success runs

The number of non-overlapping success runs of length \( k \) (\( k \) is a positive integer) was first introduced by Feller (1968) and is usually denoted by \( N_{n,k} \). Another counting scheme was proposed by Ling (1988) gives rise to the number \( M_{n,k} \) of overlapping success runs of length \( k \).

Instead of giving the mathematical definition of the above mentioned variables, we mention the following illustrative example:

In a sequence of \( n = 12 \) trials, the outcomes are,

\[
SFSSSSFSSFS
\]

Let \( k = 2 \), then

\[
N_{n,k} = N_{12,2} = \text{number of non-overlapping success runs of length } k = 2, \text{ in a sequence of } n = 12, \text{ in a sequence of } n = 12 \text{ trials}
\]

\[
= 3
\]

and

\[
M_{n,k} = M_{12,2} = \text{number of overlapping success runs of length } k = 2, \text{ in a sequence of } n = 12 \text{ trials}
\]

\[
= 5
\]
1.3 Applications

1.3.1 Run Theory

The study of pattern and the Run in statistics have been found wise applications in Industry, non-parametric statistical inference, quality control, computer science, networking, cryptography, molecular biology, epidemiology, psychology, remote sensing, ecology etc. A few of the applications has been enlisted in the following.

1.3.2 Acceptance Sampling Plans [Ahn and Kuo (1994)]

In quality control, control charts are used as a standard tool in statistical process control. The use of sampling plans for inspecting the lots can be regarded as Bernoulli trials. For each sample under inspection from a lot, the probability of obtaining $x$ defectives in a random sample of size $n$, drawn from a lot of size $N$ and containing $X$ defective is

$$P(x,n;X,N) = \binom{X}{x} \binom{N-X}{n-x} \frac{N^n}{n^n} , \quad \max(n+X-N,0) < x < \min(X,n)$$

For large $N$, the probability function tends to

$$P(x,n) = \binom{n}{x} p^x (1-p)^{n-x} , \quad x = 0,1,2,...,n$$

where $p$ is the probability of success.
The acceptance of the lot can be viewed as success of a component in reliability system and the rejection is equivalent to failure of a system. The lot is accepted, if \( k \)-out-of-\( n \) consecutive trials result in success, i.e., a success run of length at most \( k \), the probability can be obtained from the pgf.

1.3.3 Reliability [Papastavridis(1990), Godbole (1993), Koutras (1996)]

The reliability system, called \( k \)-out-of-\( n \): \( F \) system are closely connected to length of the longest run of successes \( (L_n) \) in a sequence of Bernoulli trials. This system consists of a sequence of \( n \) components (labeled as 1, 2, ..., \( n \)) in which each component is either working or failed and each component works independently of each other. The entire system fails whenever at least \( k \) consecutive components fail. The reliability of this system is

\[
R_n = P(L_n < k)
\]

In other system of reliability, it is directly related to the number of non-overlapping runs of length \( k \) in \( n \) Bernoulli trials, is known as \( m \)-consecutive-\( k \)-out-of-\( n \): \( F \) system. This system fails whenever \( m \) overlapping runs of \( k \) consecutive failed in \( n \) observed components of the system.
1.3.4 **Scan Statistics** [Glaz and Naus (1990), Glaz and Balakrishnan (1999)]

The discrete scan statistics is used for testing the null hypothesis of uniformity against an alternative hypothesis of clustering which specifies an increased occurrence of events in a connected subsequence of the observations. The scan statistics have been found interesting applications in many areas of science including analysis of DNA and Protein sequence, sociology, radar detection, epidemiology, minefield detection, quality control etc.

1.3.5 **Ecology** [Pielon (1962, 1963a, 1963b; Knight (1974))]

Ecology studies are used to study the distribution of infected trees with respect to healthy ones in a forest, i.e., segregation of two classes of trees, viz., infected and healthy, segregation of species, etc. To perform test of segregation, the sampling is done by means of narrow belt transects. Such a transact, through a two species population, must always contains runs of individuals of the other. The frequency distributions of these run lengths will through light on the relationship between two species that are segregated.
1.4 Objective of the Thesis

1. Some aspects of a class of quasi binomial distribution of order $k$, various distributional properties, limiting distribution and properties of some of the members of this class in particular has been studied.

2. Abel series distribution of order $k$ and Abel series distribution family of order $k$ has been defined and from which many discrete distributions of order $k$ have been derived.

3. We have obtained a family of multivariate Abel series distributions of order $k$ from which a new distribution called quasi multivariate logarithmic series distribution has been derived and study some properties, generalizations and compounding.

4. Zero-modified distributions of order $k$ have been defined and also obtained different zero modified distributions of order $k$, studied their properties in general.

1.5 Chapterisation Scheme

An introduction of the work is given in the first chapter including review of literature and genesis of distributions of order $k$.

In chapter two the probability functions of quasi binomial distributions of type I (QBD I) of order $k$ (QBD I $(k)$) and quasi binomial
distributions of type II (QBD II) of order $k$ (QBD II ($k$)) has been defined and their particular cases of QBD I (Consul, 1974) and QBD II (Consul and Mittal, 1975). It has been shown that under certain underlying conditions, both the QBDs of order $k$ tend to GPD of order $k$ having the probability function. Then a class of QBDs of order $k$ has been derived by using Abel’s generalization of the binomial formula (Riordon, 1968) after extending it to of order $k$ and simultaneously some particular cases have been studied along with some recurrence relations and limiting distributions of QBDs of order $k$ have been defined and studied.

In the third chapter, first of all an effort has been made to develop a class of univariate discrete distributions of order $k$, i.e., the Abel Series distributions of order $k$ (ASD($k$)) generated by suitable functions of real valued parameters in the Abel polynomials. Then a new distribution called the quasi logarithmic series distribution of order $k$ (QLSD($k$)) is derived from ASD($k$) and many more distributions, viz. quasi binomial distributions of order $k$ (QBD($k$)), Generalized Poisson distribution of order $k$ (GPD($k$)) and quasi negative binomial distribution of order $k$ (QNBD ($k$)) have been derived as particular cases of ASDs of order $k$. Some inferential properties have also been discussed.

The fourth chapter begins with considering the multivariate Abel
series of order $k$, here we have defined multivariate Abel series distributions of order $k$ and from the MASDs of order $k$, a new distribution called Quasi Multivariate logarithmic series distribution of order $k$ has been obtained. Also a variant of quasi negative multinomial distribution of order $k$ is studied. Moreover, by using a new method of derivation some well known distributions, viz. quasi multinomial distribution of type-I of order $k$ (QMD-I $(k)$), quasi multinomial distribution of type-II of order $k$ (QMD-II $(k)$), multiple generalized Poisson distribution of order $k$ (MGPD $(k)$) etc. are obtained. A property, i.e., the limiting distribution of the QNMD of order $k$ has been found out.

In the last chapter a zero-modified distribution of order $k$ has been introduced and defined and an attempt is made to study different zero-modified distributions of order $k$ such as Binomial distributions of order $k$, Poisson distributions of order $k$, Geometric distributions of order $k$, Negative Binomial distributions of order $k$ and logarithmic series distributions of order $k$. Also a class of inflated generalized Poisson distribution of order $k$ has been defined and studied. Some of their inferential properties also have been discussed.

Finally, two appendices have been provided to list the various formulae and results related to the present work.