HARRIS FAMILY OF DISCRETE DISTRIBUTIONS AND PROCESSES

Thesis submitted to the
University of Calicut
for the award of the degree of

**DOCTOR OF PHILOSOPHY
IN STATISTICS**

By

SHERLY SEBASTIAN

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673635
INDIA

MARCH 2007

DEDICATED TO MY BELOVED DAUGHTER RYTHU

DECLARATION

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree, diploma, associateship, fellowship, membership etc., of any university or institute and, to the best of my knowledge and belief, it contains no material previously published by any other person, except where due reference made in the text of the thesis.

Calicut University Campus

30th March 2007

Sherly Sebastian

CERTIFICATE

This is to certify that the thesis entitled HARRIS FAMILY OF DISCRETE DISTRIBUTIONS AND PROCESSES that is being submitted by Smt. Sherly Sebastian for the award of Doctor of Philosophy, to the University of Calicut, is based on the bonafide research work carried out by her under our supervision and guidance. The work reported herein does not form part of any other thesis or dissertation submitted previously for the award of any degree or diploma of any other university or institution.

Dr. N. Raju,
Reader,
Department of Statistics
University of Calicut.
Kerala – 673 635
India

Dr. E. Sandhya,
Lecturer (Selection Grade)
Department of Statistics
Prajyothi Niketan College, Pudukkad
Kerala – 680 301
India

ACKNOWLEDGEMENT

First and foremost I praise and thank God, the Almighty for giving me the strength and power to under take and complete this task successfully.

I am deeply indebted to my supervisors Dr. N. Raju, Reader, Department of Statistics, University of Calicut and Dr. E. Sandhya, Selection Grade Lecturer in Statistics, Prajyothi Niketan College, Pudukkad, Thrissur for their inspiring guidance, constant encouragement and help throughout the preparation of this thesis. I thank them profusely for the interest, care and the immeasurable support that they extended to my research work.

I am also indebted to Dr. C. Chandran, Head of the Department, Dr. M. Manoharan (Former Head) and Dr. K. Jayakumar, for providing me the necessary facilities, support and help.

My gratitude knew no bounds while remembering the precious help and guidance extended to me by Dr. S. Satheesh, BSNL, Thrissur, in my research work during the entire span of my research. I express my heartfelt thanks to Mr. N.K. Narayananunny, DE, BSNL, Thrissur for his kind help.

It is a pleasure for me to record my indebtedness to Rev. Sr. Rose Dheera, former Principal, Vimala College, Thrissur and Rev. Sr. Lekha, Principal, Vimala College, Thrissur for the encouragement in all my endeavors. My thanks are also to my colleagues and my friends for the help in various ways.

The co-operation by the office staff, Librarian, research scholars and students of the Department of Statistics, University of Calicut is here remembered and acknowledged.

I am thankful to Dr. T. M. Jacob, Department of Statistics, Nirmala College, Moovattupuzha, for the help in doing the simulation study.

The financial support received from the UGC, X th Plan through FIP fellowship is gratefully acknowledged.

Words are inadequate to express my sincere thanks to my parents, brother, sisters and relatives for their moral support and prayers throughout my research work.

I am greatly indebted to my Husband Mr. Jos M. K., Selection grade Lecturer, Department of Statistics, St. Thomas College, Thrissur, for his continuous support and help in all my endeavors. I thank him specially for being the source of inspiration in taking up the research work initially. My thanks are also to my son Fenil and daughter Dony for their co-operation and patience in completing the work successfully.

I take this opportunity to express my sincere thanks to all those who helped me in my research work.

Sherly Sebastian

CONTENTS

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 Probability Models

A model takes into account the essentials of a phenomenon or a system ignoring minor details. A model may be iconic giving shape to an abstraction, symbolic, expressing the phenomenon in symbols or logical, involving mathematical relationships or statements underlying the various aspects of the phenomenon under observation. As it provides an approximation to the real-world situation, the conclusions drawn from the model are also approximations. If the deductions following the model agree with the actual observation, the model is satisfactory and it is used for further study of the phenomenon.

Mathematical or deterministic models are used extensively in every day life and they hold true under ideal conditions. But the ideal situation envisaged in a deterministic model hardly exists in every day life. Most of the real-life phenomena are governed by chance-laws. Any realistic model of a real-world phenomenon must take into account the possibility of randomness. That is, the quantities we are interested in will not be predictable in advance, but will exhibit an inherent variation that should be taken into account by the model. This is usually accomplished by allowing the model to be probabilistic in nature. Regularity in the occurrence of chance phenomenon enables us to introduce random or chance variables in the model, which will improve the model and make it closer to reality. A model involving a random variable (r.v) or chance factor is called a stochastic or probability model.

Practical and useful probability models may refer to non-observable worlds. For example, in insurance, probability theory is used to calculate the probability of ruin; that is, the theory is used to avoid certain undesirable situations, and consequently it applies to situations that are not actually observed.

The first step in the construction of a model is to identify the essential characteristics of the phenomenon and their inter-relationships. Based on experience, logical foundations and preliminary observations, a family of models is proposed. This may involve specification of the family of probability distributions of the r.vs and relations between them. We are concerned with theoretical discrete distributions in which variables are distributed according to some definite probability law which can be expressed mathematically. We can fit a mathematical model or a function of the form $y = f(x)$ to the observed data.

Lattice distributions

Most of the discrete distributions used in statistics belong to a much narrower class, the lattice distributions. In these distributions the intervals between the values of any one r.v for which there are nonzero probabilities are all integral multiples of one quantity (which depends on the r.v). Points with these coordinates form a lattice. By an appropriate linear transformation it can be arranged that all variables take values that are integers. For most of the discrete distributions the support will be a set of non-negative integers and for such non-negative lattice variables certain special methods of analysis are available.

There are a number of ways of classifying non-negative lattice distributions. For example, they can be classified on the basis of the form of the probability mass function (p.m.f) $f(x)$, a function of x. One such class is generalized power series distributions which is discussed in chapter 3. Some of the important distributions belonging to this class are the binomial, Poisson, negative binomial and logarithmic distributions. The description of many of these distributions starts with the concept of Bernoulli trials.

Bernoulli trials

Repeated independent trials are called Bernoulli trials if there are only two possible outcomes, say, "success", S and "failure", F for each trial and their probabilities remain the same throughout the trials. The probabilities of success and failure are usually denoted by p and $q = 1 - p$ respectively. Clearly, p and q must be non-negative, and $p + q = 1$.

The binomial distribution

Let X denote the total number of successes produced in a succession of n Bernoulli trials with success probability p. Then X is said to have a binomial distribution with parameters (n, p) if its p.m.f is given by

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2 \ldots n$$

where $0 < p < 1$ and n is a positive integer.

When $n = 1$, the distribution is known as the Bernoulli distribution.

The binomial distribution is a power series distribution with finite support. It arises whenever the underlying events have two possible outcomes, the chances of which remain constant. The importance of the distribution has extended from its original application in gaming to many other areas. Its importance in model building is evidenced by Khatri and Patel (1961), Katti (1966) and Douglas (1980). Johnson and Kotz (1977) have given many instances of urn models with Bernoulli trials.

The Poisson distribution

In many applications we deal with Bernoulli trials where, comparatively speaking, n is large and p is small so that the product $\lambda = np$ is of moderate magnitude. In such cases it is convenient to use an approximation to binomial distribution which is due to Poisson.

A r.v X is said to have a Poisson distribution with parameter λ, if for some $\lambda > 0$,

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad x = 0, 1, 2, \ldots$$

The distribution is a power series distribution with infinite non-negative integer support. It belongs to the exponential family of distributions also. Poisson distribution arises in the case of rare events. That is, it can be considered in cases where, in addition to the requirements of independence of trials and consistency of probability from trial to trial, the number of trials, n is very large and the probability of occurrence of the outcome under observation, p is small. "Student" (W.S. Gosset) (1907) used the Poisson distribution to represent the number of particles falling in a small area A when a large number of such areas are spread at random over a surface large in comparison with A. It is used in quality control for the number of defective items per batch. Walsh (1955), van der Waerden (1960) are examples.

In the empirical treatment of count data the Poisson distribution is often used as a yardstick to assess the degree and nature of randomness. It is the counting distribution for a Poisson process, which has great importance concerning counts of events per unit of time, particularly in queueing theory. Certain very common distributions such as the negative binomial are both Poisson-stopped-sum distributions and mixed Poisson distributions. There are also others that belong to only one of these two families of distributions.

The negative binomial distribution

The negative binomial distribution is a frequently encountered standard discrete distribution belonging to a two-parameter family. Many different models give rise to this distribution and often it appears in different forms because of different parameterizations. One very common parameterization employs two real-valued parameters p and r with $0< p<1$ and $r > 0$. Under this parameterization, the p.m.f of a r.v with a negative binomial distribution ($NB(p, r)$) takes the following form:

$$f(x) = \binom{r+x-1}{x} p^r q^x, \quad x = 0, 1, 2, \ldots; \quad q = 1 - p.$$

The $NB(p, r)$ is the probability distribution of the number of failures before the rth success in a Bernoulli process with success probability p. A common derivation of this distribution is through an inverse sampling model. That is, if one goes on sampling with replacement until he gets a fixed number of defectives, then the number of items sampled follows a negative binomial distribution. More specifically, the negative binomial r.v X represents the number of independent Bernoulli trials required to obtain r occurrences of an event that has a constant probability of occurrence p at each trial. Further, the negative binomial distribution arises as a continuous mixture of Poisson distribution where the mixing distribution of the Poisson rate is a gamma distribution.

The equality of mean and variance is an important characteristic of the Poisson distribution, whereas for the binomial distribution mean is always greater than the variance. There are observable phenomena which give rise to empirical discrete distributions which show a variance larger than the mean. It has been shown by different investigators that in such cases the negative binomial distribution provides an excellent model as this distribution has a variance larger than the mean. It can be used as a robust alternative to the Poisson distribution and it is especially useful for discrete data over an unbounded positive range whose sample variance exceeds the sample mean, a feature called 'overdispersion'. If a Poisson distribution is used to model such a data, the model mean and variance are equal. In that case, the observations are overdispersed with respect to the Poisson model. Since the negative binomial distribution has one more parameter than the Poisson, the second parameter can be used to adjust the variance independently of the mean.

Many researchers, including Arbous and Kerrich (1951), Greenwood and Yule (1920) and Kemp (1970), have applied it to accident statistics. Furry (1937) and Kendall (1949) have shown its applicability in birth-and-death processes. Medical and military applications have been described by Chew (1964) and by Bennett and Birch (1964). Bliss and Fisher (1953) successfully fitted the negative binomial distribution to a large number of biometrical data sets. When Martin and Katti (1965) fitted the negative binomial and certain other distributions to ecological data sets, they found that the negative binomial distribution has very wide applicability.

Wilson and Room (1983) and Perry (1984) have used the negative binomial for modeling entomological data. The distribution has been used to model family size by Rao *et al.* (1973). Physical applications involving queueing theory and other stochastic processes were described by Taylor and Karlin (1984). Autoregressive moving-average processes with geometric and negative binomial marginal distributions were discussed by McKenzie (1986). Recently Wang *et al.* (2001) introduced a UMPU test for testing the equality of means for independent samples of count data from two negative binomial populations in the presence of common dispersion parameter. Johnson *et al.* (1992) provides a detailed discussion on its genesis, historical remarks and its properties as well as its applications.

An important point to be noted here is that the parameter r need not be an integer. For the special case where r is an integer, the negative binomial distribution is known as the Pascal distribution. Other names for the negative binomial distribution are the binomial waiting time distribution and the Polya distribution.

A further specialization occurs when $r = 1$. In this case it is known by the name geometric distribution.

The geometric distribution

Suppose that Bernoulli trials with probability of success p are performed until a success occurs. If X denotes the number of failures before the first success, then X is said to follow a geometric distribution with parameter p if its p.m.f is given by

$$f(x) = q^x p, \quad x = 0, 1, 2, \ldots$$

where $0 < p < 1$, $q = 1 - p$ and $p + q = 1$.

If X denotes the number of trials required to get the first success then we get a modified (decapitated) geometric distribution with p.m.f,

$$f(x) = q^{x-1} p, \quad x = 1, 2, \ldots$$

where $0 < p < 1$, $q = 1 - p$ and $p + q = 1$.

The geometric distribution is the discrete analogue of the exponential distribution. Further, it is shown that the sum of geometric r.vs is a negative binomial variable.

Applications of the geometric distribution in queueing theory and applied stochastic models are discussed by Taylor and Karlin (1984). Mann *et al.* (1974) treated applications in reliability theory. Johnson *et al.* (1992) give a detailed discussion on its properties, various characterizations and applications.

The current study is based on a lattice distribution which has close relation with negative binomial and geometric. We have named it Harris distribution after Harris (1948) who considered its probability generating function (p.g.f) defined by

$$P(s)=\frac{s}{(m-(m-1)s^k)^{1/k}}, \quad k>0 \text{ integer and } m>1 \tag{1.1.1}$$

He introduced this p.g.f while considering a simple mathematical model for a branching stochastic process. The model represents the generation-by-generation growth of a family where the fundamental r.v Z_n is the number of individuals in the nth generation. Under certain conditions this model describes the size of a family at a sequence of points in time. It is proved that in the case of a family with non-zero chance of indefinite survival, the r.vs $Z_n/E(Z_n)$ converge in probability to a r.v W with distribution function (d.f) $G(w)$. He has discussed some special cases in which the moment generating function (m.g.f) $\varphi(s)$ and the d.f $G(w)$ of W can be found explicitly. By considering generating functions of the form given in (1.1.1) he has obtained the asymptotic distribution of $Z_n/E(Z_n)$ as *Gamma* $(1/k, 1/k)$ where *Gamma* (α, β) is defined by its probability density function (p.d.f)

$$f(x)=\frac{\beta^\alpha}{\Gamma\alpha}e^{-\beta x}x^{\alpha-1}, \quad x>0$$

where $\alpha>0$, $\beta>0$ are the shape and the scale parameters respectively and $\Gamma\alpha$ is the gamma function defined by

$$\Gamma\alpha = \int_0^\infty e^{-x} x^{\alpha-1} dx, \quad \alpha > 0 .$$

Continuous models that might be considered in certain fission processes, giving the number of individuals existing at a given time are also discussed under the assumption that the probability of fission is independent of age. He suggests that there should exit a family of functions $f_n(s)$ defined for all positive n such that $f_{n_1}\left(f_{n_2}(s)\right) = f_{n_1+n_2}(s)$; such that for each positive n, $f_n(s)$ is a p.g.f and such that for n = 0, 1, 2,... the functions $f_n(s)$ coincide with the iterates s, $f(s)$, $f(f(s))$, ...

The functional equation $\varphi(ms) = f(\varphi(s))$ shows that $f(s) = \varphi(m\varphi^{-1}(s))$, whence $f_n(s) = \varphi(m^n\varphi^{-1}(s))$ for integral n. This has led him to consider $f(s) = \dfrac{s}{m-(m-1)s}$. Then the iterates $f_n(s) = \dfrac{s}{m^n - (m^n - 1)s}$ and they are generating functions for all positive n, satisfying the required relation $f_{n_1}(f_{n_2}) = f_{n_1+n_2}$. If $g(s)$ is some function such that the function $f(s) = g^{-1}\left(\dfrac{g(s)}{m-(m-1)g(s)}\right)$ is a generating function for all $m > 1$, with $g(1) = 1$, then $f_n(s) = g^{-1}\left(\dfrac{g(s)}{m^n - (m^n - 1)g(s)}\right)$ and they are generating functions for all $n > 0$ as $f(s)$ is a generating function for all $m > 1$. The simplest function $g(s)$ which satisfies the requirements is $g(s) = s^k$, where k is any positive integer. In this case $f(s)$ has the form considered in (1.1.1) and $f_n(s) = s\left(m^n - (m^n - 1)s^k\right)^{-1/k}$. As $n \to 0$, $f_n(s) = (1 - \frac{n}{k}\log m)s + \dfrac{n\log m}{k} s^{k+1} + 0(n^2)$, which may be interpreted as follows

A particle in existence at a given time may, in a short time interval Δt, either split into $(k+1)$ particles, with probability $\frac{\Delta t \log m}{k}$; or it may remain unaltered, with probability $1-\frac{\Delta t \log m}{k}$. If it splits, each particle produced has the same chances for splitting as its parent, etc. If we begin with a single particle at time $t = 0$, the asymptotic probability distribution of Z_t / m^t where Z_t is the number of particles at time t and $E(Z_t) = m^t$ is *Gamma* $(1/k, 1/k)$.

From the p.g.f (1.1.1) one can readily see that this is a generalization of the geometric distribution on $\{1, 2, 3, \ldots\}$ to which it reduces when $k = 1$ and its atoms (probability carrying integers) are k integers apart or the probabilities are concentrated on the points $1, 1+k, 1+2k, \ldots$. It is also true that these probabilities coincide with that of the negative binomial distribution on $\{0, 1, 2, \ldots\}$ with parameters $1/m$ and $1/k$. But not much study concerning the properties of this distribution is seen in the literature.

The distribution corresponding to the p.g.f (1.1.1) is denoted by $H_1(m,k,1/k)$. In the notation the suffix 1 suggests that the support of the distribution starts from unity, m determines the probabilities, k implies that the atoms of the distribution are k integers apart and $1/k$ is the exponent.

The main interest that led us to a close examination of this distribution is its role played in schemes with random (N) sample sizes (random-sums or N-sums and random-extremes or N-extremes) in general and in branching processes and time series models in particular where N is a non-negative integer-valued r.v. Satheesh *et al.* (2002) and Satheesh and Nair (2002*a*) have been discussed this p.g.f in the context of N-sums and N-extremes.

Some basic concepts and definitions, which are needed to understand the results connected with random sums, extremes, stability etc. are given below. Most of the results have been taken from Feller (1966, 1968), Johnson *et al.* (1992) and Lukacs (1970).

1.2 Preliminaries

Definition 1.1 (Probability generating function) Suppose that X is a r.v that assumes the values 0, 1, 2, ... and that $P(X = n) = p_n$, n = 0, 1, 2, ... with $\sum p_n = 1$, then the generating function $P(s) = \sum_{n=0}^{\infty} p_n s^n$ of the sequence of probabilities $\{p_n\}$ is known as the probability generating function of the r.v X.

As $P(1) = 1$, the series $P(s)$ converges absolutely at least for $-1 \leq s \leq 1$. Also the p.g.f defines a distribution uniquely. Again,

$$P(s) = \sum_{n=0}^{\infty} P(X = n)s^n = E(s^X)$$

where $E(s^X)$ is the expectation of the function of the r.v X.

Definition 1.2 (Laplace transform) The Laplace transform (LT) of a positive real valued r.v X with p.d.f $f(x)$ is defined as

$$\varphi(\lambda) = E(e^{-\lambda X}) = \int_0^{\infty} e^{-\lambda x} f(x)dx, \quad \lambda > 0$$

Definition 1.3 (Self-decomposability) If for every $\alpha \in (0,1)$ there exists a r.v X_α independent of X such that $X \stackrel{d}{=} \alpha X + X_\alpha$, then the r.v X (or its distribution) is said to be self-decomposable.

(The symbol $\stackrel{d}{=}$ expresses the equality of distributions).

In terms of characteristic function (c.f) it may be written as:

Definition 1.4 A probability distribution on R is said to be self-decomposable (or, of class L) if its c.f satisfies

$$\phi(t) = \phi(\alpha t)\phi_{\alpha}(t), \qquad t \in R; \quad \alpha \in (0,1),$$

with ϕ_{α} a c.f.

Definition 1.5 (Infinite divisibility) A r.v X is called infinitely divisible (i.d), if for each $n \geq 1$, there exits independent and identically distributed (i.i.d) r.vs $X_1, X_2, \dots X_n$ such that

$$X \stackrel{d}{=} X_1 + X_2 + \dots + X_n$$

In terms of characteristic functions it may be defined as follows:

Definition 1.6 A r.v is said to be i.d if and only if it has a c.f $\phi(t)$ that can be represented for every positive integer n as the nth power of some c.f $\phi_n(t)$, i.e.

$$\phi(t) = (\phi_n(t))^n, \qquad t \in R$$

Definition 1.7 (N – compounding) Consider a r.v S_N which can be written as

$$S_N = X_1 + X_2 + \dots + X_N \tag{1.2.1}$$

where $X_1, X_2, \dots, X_N$ are i.i.d r.vs and N is some discrete r.v independent of $X_1, X_2, \dots, X_N$. If S_N has the representation (1.2.1) we say that S_N is closed under N-compounding. In particular, when N is geometric, we say that S_N is closed under geometric compounding.

Definition 1.8 (N – sum stability) Let $X, X_1, X_2, \dots$ be non-degenerate i.i.d r.vs with a common c.f $\phi(t)$ and N a positive integer valued r.v independent of X with p.g.f $P(s)$. Set the N-sum of X_i's, $S_N = X_1 + X_2 + \dots + X_N$. When cS_N and X are identically distributed for some $c > 0$ (S_N and X are of the same type), or equivalently

$$P(\phi(ct)) = \phi(t), \text{ for all } t \in R, \tag{1.2.2}$$

we say that the distribution of X is stable under the operation of summation with respect to (w.r.t) the r.v N (X is N-sum stable). Notice that for N-sum stability we need relation (1.2.2) to be satisfied for some $c>0$ only. Often N is described as the compounding r.v.

Definition 1.9 (*N* - extreme stability) Let $F(x)$, $x \in R$ be the d.f of a continuous r.v X and $P(s)$ be the p.g.f of a positive integer valued r.v N which is independent of X. Then, $F(x)$ is N-max stable if

$$P[F(x)] = F(a+bx),$$

and N-min stable if

$$P[\overline{F}(a+bx)] = \overline{F}(x),$$

for all $x \in R$ and some $a \in R$ and $b>0$, where $\overline{F}(x) = 1-F(x)$ is the survival function. The distributions of N-max and N-min of $F(x)$ should be of the same type as that of $F(x)$.

1.3 Harris Distribution in Literature

In literature the p.g.f (1.1.1) of Harris distribution has been widely used in summation schemes. As already stated, $H_1(m,k,1/k)$ defined on $\{1, 1+k, 1+2k, \ldots\}$ is a generalization of geometric distribution supported by the set $\{1, 2, \ldots\}$. The geometric summation schemes are studied by Sandhya (1991*a*, *b*), Pillai and Sandhya (1996), Cinlar and Agnew (1968), Yannaros (1989) etc. Sandhya (1991*a*) observes that if (1.2.1) is true with $N \sim H_1(m,k,1/k)$ law, then X has a gamma distribution with LT $\varphi(\lambda) = \left(1+(k\lambda/m)\right)^{-1/k}$, the same connection that is observed between $H_1(m,k,1/k)$ and gamma distribution by Harris (1948).

Satheesh *et al.* (2002) have considered a generalization of stability of geometric sums by studying distributions that are stable under summation w.r.t $H_1(m,k,1/k)$ law. They discussed the stability of N – sums of r.vs when N is Harris (with the p.g.f (1.1.1)).

They considered the following class of continuous functions in the sequel.

$$\psi(u) = m\,\psi(bu) \text{ for all } u\in R \text{ and some } m,\ b > 0 \text{ with } \psi(0) = 0 \tag{1.3.1}$$

It has been shown that $0<b<1<m$ alone is possible and there exist a unique $\alpha>0$ such that $mb^{\alpha}=1$. Also when (1.3.1) is satisfied for two values of b, say b_1 and b_2 such that $\ln b_1/\ln b_2$ is irrational then $\psi(u) = \eta\,|u|^{\alpha}$, for some $\eta > 0$.

In the context of branching processes, when the p.g.f of N is (1.1.1) Harris (1948) showed that the N-sum of *Gamma*$(1/k,\ 1)$ law with LT $1/[1+\lambda]^{1/k}$ is of the type as *Gamma*$(1/k,\ 1)$. In fact the stability holds here for every $m>1$ as shown by Sandhya (1991*a*).

While discussing the stability w.r.t $H_1(m,k,1/k)$ law they have proved the following results:

Theorem 1.1

A c.f $\phi(t)$ is $H_1(m,k,1/k)$-sum stable for some $m>1$, if and only if $\phi(t) = 1/(1+\psi(t))^{1/k}$, where $\psi(t)$ satisfies (1.3.1) with $\alpha\in(0,2]$, $b = c = m^{-1/\alpha}$, 'm' and 'k' being that in $H_1(m,k,1/k)$.

Theorem 1.2

If $\phi(t)$ is N-sum stable w.r.t $H_1(m_1,k,1/k)$ and $H_1(m_2,k,1/k)$ such that $\ln m_1/\ln m_2$ is irrational, then $\phi(t) = 1/[1+\eta|t|^{\alpha}]^{1/k}$.

Theorem 1.3

A LT $\varphi(\lambda)$ is N-sum stable w.r.t $H_1(m,k,1/k)$ law for some $m>1$, if and only if $\varphi(\lambda) = [1+\psi(\lambda)]^{-1/k}$, where $\psi(\lambda)$ satisfies (1.3.1) with $\alpha\in(0,1]$, $b = c$, 'm' and 'k' being that in $H_1(m,k,1/k)$. Further, the LT $1/(1+\lambda^{\alpha})^{1/k}$ is $H_1(m,k,1/k)$-sum stable for every $m>1$.

When $\varphi(\lambda) = [1+\psi(\lambda)]^{-\beta}$, $\beta>0$, it is noticed that $\varphi(\lambda)$ is not N-sum stable w.r.t $H_1(m,k,1/k)$ law.

They have used the following lemma in developing (i) p.g.fs representing the discrete analogue of the LTs of continuous laws on $[0,\infty)$, (ii) stability of integer valued r.vs in the N-sum scheme.

Lemma 1.1

If $\varphi(\lambda)$ is a LT, then $P(s) = \varphi(1\text{-}s)$, $0<s<1$ is a p.g.f.

In the discrete set up corresponding to equation (1.2.2), they have described the stability of N-sums with $Q(s)$, the p.g.f of X as

$$P[Q(1-c+cs)] = Q(s) \text{ for all } s\in(0,1) \text{ and some } 0<c<1. \tag{1.3.2}$$

Invoking Lemma 1.1 and (1.3.2) they have the following result on the stability of a discrete r.v w.r.t $H_1(m,k,1/k)$.

Theorem 1.4

A p.g.f $Q(s)$ is $H_1(m,k,1/k)$-sum stable for some $m>1$, if and only if $Q(s) = 1/[1+\psi(1\text{-}s)]^{1/k}$, where $\psi(1\text{-}s)$ satisfies (1.3.1), $b = c = m^{-1/\alpha}$, 'm' and 'k' being that in $H_1(m,k,1/k)$.

While discussing a method for identifying the compounding r.v N for N-sum stability they have proved the following general result.

Theorem 1.5

The generalized Semi Mittag-Leffler ($SML(a,b,\beta)$) law with LT $\varphi(\lambda) = [1+\psi(\lambda)]^{-\beta}$, $\beta>0$ is N-sum stable if and only if N is $H_1(m,k,1/k)$, $\beta = 1/k$, $c = b$, 'm' and 'k' being same in both.

Setting $\psi(\lambda) = \lambda^{\alpha}$, they have showed that a positive generalized Linnik$(\alpha,0,\beta)$ law (generalized ML) is N-sum stable if and only if N is $H_1(m,k,1/k)$, $\beta k = 1$ and $mc^{\alpha} = 1$. Also by setting $\psi(\lambda) = \lambda$, they have stated that $Gamma(\beta, 1)$ law is N-sum stable if and only if N is $H_1(m,k,1/k)$, $\beta = 1/k$ and $c = 1/m$.

In the discrete set up, they have the following result.

Theorem 1.6

The p.g.f $Q(s) = [1+\psi(1-s)]^{-\beta}$, where $\psi(1-s)$ satisfies (1.3.1) is N-sum stable if and only if N is $H_1(m,k,1/k)$, $\beta = 1/k$, $k>0$ integer, $c = b$, 'm' and 'k' being same in both.

Also, it is given that the geometric law on $\{0,1,2, \ldots\}$ and discrete analogues of SML and ML laws are stable only w.r.t a geometric-sum, and the discrete analogue of $Gamma(\beta, 1)$ is stable only w.r.t a Harris-sum and further $\beta=1/k$ for a positive integer k.

The role of Harris distribution on the stability of extremes has been discussed in Satheesh and Nair (2004). They start with the definition on N- max stability and N- min stability. The geometric law on $\{1,2,3, \ldots\}$ with expectation $1/p$ is denoted by $Geo(p)$.

Marshall and Olkin (1997) introduced a parameterization scheme for a survival function $\overline{F}(x)$, $x \in R$ that is similar in structure to the *Geo(p)*-minimums by defining another survival function

$$\overline{G}(x,a) = \frac{a\overline{F}(x)}{1-(1-a)\overline{F}(x)}, \quad x \in R, \quad a>0$$

and showed that this family is *Geo(p)*-extreme (that is, both *Geo(p)*-max and *Geo(p)*-min) stable. They attributed this property partially to the fact that *Geo(p)* laws are closed under their own compounding. They also concluded that the N-min stability could not be expected, if the *Geo(p)* distribution is replaced by another distribution. Satheesh and Nair (2004) have shown that a Harris-sum of Harris distributions is again Harris. They have proved the following results and generalized the Marshall-Olkin parameterization scheme from the geometric to the Harris distribution.

Theorem 1.7

N-max stability of $F(x)$ implies N-min stability of $\overline{F}(x)$ (and vice-versa) if and only if $P_u(s)$ satisfies

$$1-P_u(1-s) = P_u^{-1}(s), \quad \forall\ 0<s<1. \tag{1.3.3}$$

where $P_u(s)$ denotes the p.g.f of N with parameter $u>0$.

Theorem 1.8

If the p.g.f $P_u(s)$ of N satisfies (1.3.3), then the survival functions of N-max and N-min of $F(x)$ have the same algebraic structure in terms of $\overline{F}(x)$, if and only if

$$P_u^{-1}(s) = P_\eta(s), \text{ for some } \eta>0. \tag{1.3.4}$$

Theorem 1.9

Suppose the p.g.f $P_u(s)$ of N satisfies (1.3.3) and (1.3.4). If $P_u(s)$ also satisfies

$$P_u[P_v(s)] = P_{uv}(s) \quad \text{for all } |s|<1 \text{ and } u, v>0, \tag{1.3.5}$$

then the survival functions of N-min of N-maxs and N-max of N-mins of $F(x)$ have the same algebraic structure as that of N-min of $F(x)$.

They have also proved the following lemma which shows that (1.3.5) implies (1.3.4).

Lemma 1.2

If a one-to-one function $P_u(s)$, $u>0$ satisfies (1.3.5) then it satisfies (1.3.4) with $\eta = 1/u$ and $P_1(s) = s$ for all s.

In order to examine whether the conditions (1.3.3) and (1.3.5) characterize the p.g.f of the $Geo(p)$ law, they used the p.g.f (1.1.1) of the Harris law, which satisfies (1.3.5). They have recorded the following lemma which verify that the $H_1(m,k,1/k)$ laws are closed under their own compounding.

Lemma 1.3

Let $P_u(s) = \dfrac{s}{[u-(u-1)s^k]^{1/k}}$ and $Q_v(s) = \dfrac{s}{[v-(v-1)s^k]^{1/k}}$. Then

$$P_u(Q_v(s)) = \frac{s}{[uv-(uv-1)s^k]^{1/k}}. \quad \text{Also } P_u^{-1}(s) = P\eta(s),\ \eta = 1/u.$$

The parameter k is the same for both the p.g.fs and it is the same in the compound also. Thus it also satisfies (1.3.4) with $\eta = 1/u$, but it is not a solution of (1.3.3). Using this result they have stated that a p.g.f $P_u(s)$, $u>0$ satisfies (1.3.3) and (1.3.5) (and hence (1.3.4)) if and only if it is the p.g.f of the *Geo* (u) law with mean $1/u$.

Marshall and Olkin (1997) parameterization scheme is generalized on the following lines.

They have taken $P_u(s)$ and $Q_v(s)$ in lemma 1.3 as the p.g.fs of the r.vs N and M respectively and $\{X_i\}$ independent copies of a r.v X with d.f $F(x)$, $x \in R$. Assuming the mutual independence of N, M and X and putting $U = Min(X_1, \ldots, X_N)$ and $V = Max(X_1, \ldots, X_N)$ it is shown that,

$$P\{U > x\} = \frac{\overline{F}(x)}{[u-(u-1)[\overline{F}(x)]^k]^{1/k}} \quad \text{and} \tag{1.3.6}$$

$$P\{V < x\} = \frac{F(x)}{\{u-(u-1)[F(x)]^k\}^{1/k}}, \tag{1.3.7}$$

where $x \in R$, $k>0$ integer and $u>1$.

They have proved that the family of distributions of the form (1.3.6) is M-min stable and that of the form (1.3.7) is M-max stable and hence they have N-min and N-max stability of $F(x)$ w.r.t a non-geometric (Harris) r.v N. They have remarked that though Harris laws are closed under their own compounding the families (1.3.6) and (1.3.7) are not Harris-extreme stable, because the p.g.f of the Harris law is not a solution of (1.3.3).

Sandhya (1996) had used Harris distribution to demonstrate the notion of random infinite divisibility w.r.t non-negative integer-valued r.vs and Satheesh (2004*a*) has used it to demonstrate that an integer-valued infinitely divisible r.v X with $P(X=1)>0$ can have gaps in its support. It is also worth mentioning here that in the context of N-maximums a reparameterization of this distribution under the name 'extended geometric distribution' had been considered by Voorn (1987). Distributional and divisibility properties, characterizations, simulation and estimation problems of $H_1(m,k,1/k)$ law have been given in Sandhya, *et al.* (2006*a*, *b*). Also Satheesh *et al.* (2005) have developed a time series model that has an inherent N-sum structure where N is Harris distributed. Sherly *et al.* (2006) have obtained a stochastic model where the variable under consideration follows a Harris distribution.

As already mentioned $H_1(m,k,1/k)$ denotes the Harris distribution on $\{1, 1+k, 1+2k, \ldots\}$. Other member of the Harris family of distributions considered here is that on $\{0, k, 2k, \ldots\}$ with p.g.f, $P(s) = (m-(m-1)s^k)^{-1/k}$, $k>0$ integer and $m>1$ and we denote this by $H_0(m,k,1/k)$.

1.4 Organization of Work

This thesis consists of seven chapters, including the present introductory chapter. A brief summary of the other chapters are given below.

In chapter 2, we discuss the genesis and the distributional properties of the Harris family of distributions and a characterization of it. We obtain Harris distribution as a gamma mixture of Poisson distribution. Moments, cumulants and their recurrence relations are derived. It is shown that this distribution has a single mode at $x = 1$ and is positively skewed and leptokurtic. The relations between the Harris, geometric, negative binomial and gamma distributions are revealed in this chapter. Also the divisibility properties of the family are studied. It is proved that $H_0(m,k,1/k)$ and $H_1(m,k,1/k)$ r.vs are i.d. Further, $H_0(m,k,1/k)$ r.vs are self-decomposable while $H_1(m,k,1/k)$ r.vs are not selfdecomposable. Again, $H_1(m,k,1/k)$ r.v induces N-sum stability in the generalized Linnik distributions.

The maximum likelihood and moment estimators for both the parameters of the Harris family of distributions are derived and they are evaluated by the simulated observations. Also the minimum variance unbiased estimator (MVUE) of m is obtained. It can be seen that all these estimators coincide. Further, the Bayes estimator of the parameter m is considered in this chapter. Probability plots of Harris distributions, which give an idea about their shapes, are exhibited.

In chapter 3, Harris distribution is shown to be a member of exponential family. It has also been proved that this distribution belongs to the family of generalized power series distributions. Using the properties of power series distributions the properties of Harris distribution are studied and verified.

For a fixed k, a sufficient estimator for $\theta = (1\text{-}(1/m))^{1/k}$ is derived. Further, we develop the UMVUE of $(m\text{-}1)/m$. Again, $T = \sum X_i$ is a completely sufficient statistic, where $X_1, X_2, \ldots, X_n$ are i.i.d according to the Harris distribution $H_1(m,k,1/k)$, which belongs to the power series family. The UMP and UMPU tests of the population mean as well as the sampling distribution of the test statistic are discussed. Some numerical examples illustrating the UMP tests of the population mean and the corresponding power curves are also included in this chapter.

Chapter 4 mainly contains characterizations of certain distributions which are related to Harris distribution. Proceeding as in Johnson *et al.* (1992), we have introduced an extended geometric distribution and it is characterized in terms of convolution equations. A mixture of this distribution is characterized using the property of having completely monotone probability sequence (CMPS). Also we show that CMPSs are log-convex and hence they are i.d. The concept of fractional success is developed and explained and some real life examples are mentioned. It is shown that a mixed binomial process with gaps is renewal if and only if it is binomial with gaps. Further, a characterization which reveals the relation between extended geometric and Harris distribution is obtained. Characterizations of Harris and negative binomial distributions using the p.g.fs and a characterization of gamma distribution in terms of LT are also provided.

It is well known that a geometric random sum of exponential r.vs is exponential and a Poisson random sum of logarithmic r.vs is negative binomial. Some results about random sums are developed, provided with converses and presented in the form of characterizations of various distributions.

In chapter 5, we have developed two stochastic models where the variables under consideration follow Harris distribution. In the first model, Harris distribution is presented as an appropriate marketing distribution for a specific manufacturing unit. We have considered the policy of a manufacturing unit producing costly articles. The company provides incentives to the marketing agents in order to increase the selling volume. This model is a typical example illustrating the concept of fractional success. Here selling of each item is a fractional success and obtaining an incentive is a success and success occurs only when the k fractional successes happen *i.e.* when k items are sold. Here the rate λ of the process is considered as a function of n, the number of occurrences of the event at the instant. We have developed a stochastic process where the variable follows Harris law. We named this process 'Harris process'.

Using the above model we can evaluate the expected number of articles sold as well as the average number of incentives obtained. Further, the total number of failures preceding a success follows a geometric distribution, which can be used to evaluate the efficiency of a marketing executive and the demand for the product.

In the second model an alternate approach is used to obtain Harris process. Here the intensity parameter λ of the Poisson process follows a gamma distribution and Harris distribution is obtained as a mixture by considering a linear function of Poisson variable. Generalized mixed Poisson process is defined and Harris process is viewed as a special case of it. Further Yule – Furry process (see Bhat, 2000, section 8.1.3, for details) is shown to be a particular case of Harris process. The mean, variance, auto-correlation function and the distribution of the first occurrence time of the processes are derived. The processes are non-stationary and time homogeneous.

In chapter 6, we develop a first order autoregressive model $\{X_n\}$ that is marginally stationary where X_n is the sum / extreme of k i.i.d observations. We prove that stationary solutions to these models are also either semi-self-decomposable / extreme-semi-self-decomposable or, sum / extreme stable with respect to Harris distribution.

In the last chapter, as an illustration of fit to data, we have used a telephone data from Bharat Sanchar Nigam Limited (BSNL). The data is classified into three sets corresponding to k = 30, 60 and 180. We notice that the data is overdispersed in all the three cases. As we know Harris distribution, generalized Poisson distribution and extended geometric distribution are overdispersed distributions, we tested the goodness of fit corresponding to these three distributions. We found that Harris distribution is a good fit to the data while the generalized Poisson model and extended geometric model are not.

Again, Harris distribution is fitted using two estimates of m, namely unbiased estimate and Bayes estimate. Comparing the p-values of the chi-square for the goodness of fit test, we conclude that for small values of k, the unbiased estimate is better than the Bayes estimate and for large values of k Bayes estimate is better. Testing of hypotheses concerning the population mean of the telephone data, corresponding power functions and power curves and the probability plots of the test statistic are also given.

Further a generalized Harris distribution is developed in this chapter. A brief discussion of its distributional properties and a characterization of this distribution is given. Also by modifying the model 1 of chapter 5, we develop a stochastic model where the variable under consideration follows a generalized Harris distribution.

HARRIS FAMILY OF DISCRETE DISTRIBUTIONS

Sherly Sebastian “Harris family of discrete distributions and processes” Thesis. Department of Statistics , University of Calicut, 2007

CHAPTER 2

HARRIS FAMILY OF DISCRETE DISTRIBUTIONS

2.1 Introduction

In this chapter, Harris distribution is studied in more detail. Section 2.2 gives a brief view of the preliminary concepts and results. In section 2.3, the genesis and the distributional properties of Harris family of distributions is studied. It is observed that a r.v $X \sim H_1(m,k,1/k)$ if and only if $Y=(X-1)/k \sim NB\,(1/m,1/k)$, a relationship which is very useful in deriving the moments and distributional properties of the distribution. We present Harris distribution as a gamma mixture of distribution of a linear function of Poisson r.v. Also we obtain a characterization of Harris distribution which is similar to the characterization of a negative binomial distribution by Sathe and Ravi (1997). The divisibility properties of the distribution are studied in section 2.4. It is shown that the family of Harris distributions is i.d. Further, $H_0(m,k,1/k)$ r.vs are self-decomposable while $H_1(m,k,1/k)$ r.vs are not self-decomposable. Again, $H_1(m,k,1/k)$ r.v induces N-sum stability in the generalized Linnik distributions. Section.2.5 deals with estimation and simulation. We derive the moment and maximum likelihood estimators for both the parameters and verify them by the simulated observations. The minimum variance unbiased estimator and Bayes estimator of the parameter m are also obtained.

The contents of this chapter has been accepted for publication in the ***Proceedings of the National Seminar*** *conducted by Kerala Statistical Association and the Department of Statistics, University of Kerala, in March 2006.*

2. 2 Preliminaries

Let $P(s)$ be the p.g.f of a r.v X. Then

Result 2.1

The nth factorial moment of X is given by

$$E\big(X(X-1)\ldots(X-n+1)\big)=\frac{d^n}{ds^n}P(s)|\,s=1 \quad \text{for } n=1,2,\ldots$$

Result 2.2

The mean and variance of X in terms of the p.g.f of X are given by

$E(X)=P'(1)$ and

$V(X)=P''(1)+P'(1)-\big(P'(1)\big)^2$.

Also, $P(e^t)$ is the moment generating function and $P(1+t)$ is the factorial moment generating function.

Result 2.3

The sequence of probabilities $\{p_n\}$ is given by

$p_n = P(X=n)$ = coefficient of s^n in the expansion of $P(s)$ as a power series in s.

Randomization and mixtures

Let F be a distribution function depending on a parameter θ, and u a probability density. Then

$$W(x)=\int_{-\infty}^{+\infty} F(x,\theta)u(\theta)d\theta$$

is a monotonic function of x increasing from 0 to 1 and hence a distribution function. If F has a continuous density f, then W has a density w given by

$$w(x)=\int_{-\infty}^{+\infty} f(x,\theta)u(\theta)d\theta$$

The parameter θ is treated as r.v. The process is called randomization and the new probability density $w(x)$ is called a mixture.

Gamma and beta functions

When n is not a positive integer, meaning can be given to $n!$ by defining $(n\text{-}1)! = \Gamma(n)$ where $\Gamma(n)$ is the gamma function defined in section 1.1.

Definition 2.1 The beta function $B(a,b)$ is defined by

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx, \ a > 0, \ b > 0.$$

Also, $B(a,b) = B(b,a)$ and $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

Definition 2.2 The incomplete beta function $B_p(a,b)$ is defined by

$$B_p(a,b) = \int_0^p x^{a-1}(1-x)^{b-1}dx, \quad 0 < p < 1,$$

and the incomplete beta function ratio (incomplete beta function)

$$I_p(a,b) = \frac{B_p(a,b)}{B(a,b)}.$$

Further, $I_p(a,b) = 1 - I_{1-p}(b,a)$

Gaussian hypergeometric functions

The hypergeometric function, or more precisely the Gaussian hypergeometric function, has the form

$${}_2F_1[a,b;c;x] = 1 + \frac{ab}{c1!}x + \frac{a(a+1)b(b+1)}{c(c+1)2!}x^2 + \ldots, \qquad c \neq 0, \text{-}1, \text{-}2, \ldots \tag{2.2.1}$$

The Euler transformation

$${}_2F_1[a,b;c;x] = (1-x)^{c-a-b}\,{}_2F_1[c-a, c-b; c; x] \tag{2.2.2}$$

Again, the incomplete beta function

$$B_p(a,b) = a^{-1} p^a {}_2F_1[a, 1-b; a+1; p] \tag{2.2.3}$$

For details, see Johnson *et al.* (1992).

2.3 Genesis and Distributional Properties of Harris Family

The p.g.f $P(s)$ given in (1.1.1) can also be written as

$$P(s) = \frac{s}{m^{1/k}}\left(1-\left(1-\frac{1}{m}\right)s^k\right)^{-1/k}$$

$$= \frac{s}{m^{1/k}}\left[1+\frac{1}{k}\left(1-\frac{1}{m}\right)s^k + \frac{1}{2!}\frac{1}{k}\left(\frac{1}{k}+1\right)\left(1-\frac{1}{m}\right)^2 s^{2k} + \ldots\right.$$

$$\left. + \frac{1}{r!}\frac{1}{k}\left(\frac{1}{k}+1\right)\left(\frac{1}{k}+2\right)\ldots\left(\frac{1}{k}+r-1\right)\left(1-\frac{1}{m}\right)^r s^{rk} + \ldots\right]$$

2.3.1 Probability Distribution

Let $X \sim H_1(m,k,1/k)$. Then the coefficient of s^{1+nk} in the above expansion gives

$$P(X = 1+nk) = \binom{(1/k)+n-1}{n}\left(\frac{1}{m}\right)^{1/k}\left(1-\frac{1}{m}\right)^n \quad \text{or}$$

$$= \binom{-(1/k)}{n} p^{1/k}(-q)^n, \qquad n = 0,1,2,\ \ldots\ ; \quad k>0 \text{ integer}, \ m>1$$

and $\binom{(1/k)+n-1}{n} = \dfrac{\Gamma((1/k)+n)}{n!\,\Gamma(1/k)}$.

This is the $(n+1)^{\text{th}}$ term in the expansion of $((1/p)-(q/p))^{-1/k}$ where $p=1/m$, $q=1-p$. These probabilities are attached to the points 1, $1+k$, $1+2k$, ... which are k integers apart.

Definition 2.3 A discrete r.v X defined on $\{1, 1+k, 1+2k, \ldots\}$ follows Harris distribution with parameters m and k if its probability mass function is given by

$$P(X=1+nk) = \binom{(1/k)+n-1}{n}\left(\frac{1}{m}\right)^{1/k}\left(1-\frac{1}{m}\right)^{n}, \qquad n=0,1,2,\ldots \tag{2.3.1}$$

where $\binom{(1/k)+n-1}{n} = \dfrac{\Gamma((1/k)+n)}{n!\,\Gamma(1/k)}$, $k>0$ integer, $m>1$.

Now, take $x = 1+nk$ in (2.3.1), then the p.m.f of $H_1(m,k,1/k)$ is

$$P(X=x) = \binom{\frac{x}{k}-1}{\frac{x-1}{k}}\left(\frac{1}{m}\right)^{\frac{1}{k}}\left(1-\frac{1}{m}\right)^{\frac{x-1}{k}}, \quad x=1,1+k,1+2k,\ldots \tag{2.3.2}$$

The recurrence relation for these probabilities is;

$$P\left(X=1+(n+1)k\right) = \frac{1+nk}{1+n}\,\frac{(1-(1/m))}{k}\,P\left(X=1+nk\right), \quad n=0,1,2,\ldots \tag{2.3.3}$$

Now, we have the following properties of Harris law.

(*a*) Let X be a r.v degenerate at $k>0$ integer and Y be a negative binomial r.v with parameters $1/m$ and $1/k$ ($NB(1/m,1/k)$) having p.g.f $P(s) = (m-(m-1)s)^{-1/k}$ and let $N = 1+Y$. Then the N-sum of independent observations on X is $H_1(m,k,1/k)$ distributed.

(*b*) Consider $\varphi_\theta(\lambda) = (1+\theta\lambda)^{-1/k}$, $k>0$ integer, $0<\theta<1$, the LT of $Gamma(1/k,1/\theta)$ and let $\varphi(\lambda) = (1+\lambda)^{-1/k}$. Then $\varphi(\varphi_\theta^{-1}(\lambda)) = \lambda/(\frac{1}{\theta}-(\frac{1}{\theta}-1)\lambda^k)^{1/k}$, which is the p.g.f of $H_1(1/\theta,k,1/k)$.

(*c*) Here we obtain the Harris distribution as a gamma mixture of Poisson distribution. Consider the r.v $X = 1+kY$ where Y has a conditional Poisson distribution with parameter μ so that

$$P(Y=n\mid\mu)=\frac{e^{-\mu}\mu^{n}}{n!},\quad n=0,1,2,\ldots$$

where μ itself is assumed to have a gamma distribution with density

$$g(\mu)=\frac{(1/(m-1))^{1/k}}{\Gamma(1/k)}e^{-\frac{\mu}{m-1}}\mu^{(1/k)-1},\quad k>0\text{ integer},\ m>1,\ \mu>0.$$

The r.v X has support $1,\ 1+k,\ 1+2k,\ \ldots$ and the conditional distribution is;

$$P(X=1+nk\mid\mu)=P(Y=n\mid\mu)=\frac{e^{-\mu}\mu^{n}}{n!},\quad n=0,1,2,\ldots$$

The marginal distribution of X is given by

$$P(X=1+nk)=\int_0^\infty P(X=1+nk\mid\mu)g(\mu)d\mu$$

$$=\int_0^\infty\frac{e^{-\mu}\mu^{n}}{n!}\frac{(1/(m-1))^{1/k}}{\Gamma(1/k)}e^{-\mu/(m-1)}\mu^{(1/k)-1}d\mu$$

$$=\frac{(1/(m-1))^{1/k}}{n!\,\Gamma(1/k)}\int_0^\infty e^{-((1/(m-1))+1)\mu}\mu^{(1/k)+n-1}d\mu$$

$$=\frac{(1/(m-1))^{1/k}}{n!\,\Gamma(1/k)}\frac{\Gamma((1/k)+n)}{(m/(m-1))^{(1/k)+n}}$$

$$=\binom{(1/k)+n-1}{n}\left(\frac{1}{m}\right)^{1/k}\left(1-\frac{1}{m}\right)^{n},\quad n=0,1,2,\ldots$$

by using the gamma integral. The marginal distribution of X is $H_1(m,k,1/k)$.

The relationship between the Harris and the negative binomial distribution is stated in the following theorem. The proof follows from the corresponding p.g.fs.

Theorem 2.1

A r.v $X\sim H_1(m,k,1/k)$ if and only if $Y=(X-1)/k\sim NB(1/m,1/k)$.

$X \sim H_0(m,k,1/k)$ if and only if $Y = X/k \sim NB(1/m,1/k)$.

Notice that the parameter $1/k$ in $NB(1/m,1/k)$ need not be an integer. See Johnson, *et al.* (1992, p.199-200).

Remark 2.1

A point worth noticing here is that usually integer-valued r.vs are not closed under division. However, here an integer-valued r.v is obtained by dividing another integer-valued r.v by a constant.

Certain relations between the Harris family, geometric, negative binomial and gamma distributions are given below;

(i) When $k = 1$ in the p.g.f (1.1.1) we get the p.g.f of the geometric distribution on $\{1, 2, \ldots\}$ as $P(s) = s/(m-(m-1)s) = ps/(1-qs)$ where $p=1/m$, $q=1-p$.

(ii) The p.g.f, $s/(m-(m-1)s^k) = ps/(1-qs^k)$ is the p.g.f of the geometric distribution on $\{1, 1+k, 1+2k, \ldots\}$.

(iii) The p.g.f, $1/(m-(m-1)s^k) = p/(1-qs^k)$ is the p.g.f of the geometric distribution on $\{0, k, 2k, \ldots\}$.

(iv) The p.g.f (1.1.1) can also be written as $P(s) = \{s^k/(m-(m-1)s^k)\}^{1/k} = (G(s))^{1/k}$ where $G(s) = s^k/(m-(m-1)s^k) = ps^k/(1-qs^k)$ is the p.g.f of the geometric distribution on $\{k, 2k, 3k\ldots\}$. Hence the kth root of the p.g.f of the geometric distribution on $\{k, 2k, 3k\ldots\}$ is the p.g.f of $H_1(m,k,1/k)$.

(v) The p.g.f (1.1.1) is the special case of $\theta=1$ in the p.g.f $P^*(s) = s^\theta/(m-(m-1)s^k)^{1/k}$, $\theta>0$ integer, which is the p.g.f of the linear function $Z=\theta+kY$ of Y where $Y \sim NB(1/m,1/k)$.

(vi) The following relation between the p.g.f $P(s) = (m-(m-1)s)^{-\beta}$ of a $NB(1/m,\beta)$ distribution and the LT $\varphi_\theta(\lambda) = (1+\theta\ \lambda)^{-\beta}$ of a $Gamma(\beta,\ 1/\theta\ )$ is also worth recording here. $P(\mathrm{s}) = \{(1-(1-p)s)/p\}^{-\beta} = (1+\theta(1-s))^{-\beta} = \varphi_\theta(1-s)$, where $p = 1/m$, $q = 1\text{-}p$ and $\theta = q/p$.

2.3.2 Distribution Function

Let $X \sim H_1(m,k,1/k)$ with p.m.f given in (2.3.1). Then the d.f is given by;

$$F(1+nk) = P(X \le 1+nk) = \sum_{j=0}^{n} \binom{(1/k)+j-1}{j} \left(\frac{1}{m}\right)^{1/k} \left(1-\frac{1}{m}\right)^{j}, \quad n = 0, 1, 2, \ldots$$

This sum can be expressed in terms of an incomplete beta function ratio by adopting an approach similar to the one given in Johnson *et al.* (1992, p.210).

By theorem 2.1, $Y = (X-1)/k \sim NB\,(1/m, 1/k)$. Let $p = 1/m$ and $q = 1\text{-}\ (1/m)$. Then we have

$$P(X > 1+(n-1)k) = P(Y > n-1)$$

$$= P(Y = n) + P(Y = n+1) + P(Y = n+2) + \ldots$$

$$= \binom{(1/k)+n-1}{n} p^{1/k} q^{n} + \binom{(1/k)+n}{n+1} p^{1/k} q^{n+1} + \binom{(1/k)+n+1}{n+2} p^{1/k} q^{n+2} + \ldots$$

$$= \frac{((1/k)+n-1)!}{n!((1/k)-1)!} p^{1/k} q^{n} \left(1 + \frac{1((1/k)+n)}{(n+1)} \frac{q}{1!} + \frac{1.2((1/k)+n)((1/k)+n+1)}{(n+1)(n+2)} \frac{q^2}{2!} + \ldots\right)$$

$$= \frac{((1/k)+n-1)!}{n!((1/k)-1)!} p^{1/k} q^{n}\ {}_2F_1[1,\ (1/k)+n;\ n+1;\ q] \qquad \text{[by using (2.2.1)]}$$

$$= \frac{\Gamma((1/k)+n)}{n\Gamma(n)\Gamma(1/k)} p^{1/k} q^{n} (1-q)^{(n+1)-1-((1/k)+n)}\ {}_2F_1[(n+1)-1, (n+1)-((1/k)+n); n+1; q]$$

[by using (2.2.2)]

$$= \frac{1}{nB(n,1/k)} q^n {}_2F_1[n, 1-(1/k); n+1; q] \qquad \text{since } B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$$

$$= \frac{B_q(n,1/k)}{B(n,1/k)} \qquad \text{[by using (2.2.3)]}$$

$$= I_q(n,1/k) \qquad \text{[by using the definition 2.2]}$$

$$P(X > 1+nk) = I_q(n+1, 1/k) = I_{1-p}(n+1, 1/k)$$

Hence the d.f is

$$F(1+nk) = P(X \le 1+nk) = 1 - P(X > 1+nk) = 1 - I_{1-p}(n+1, 1/k)$$

$$= I_p(1/k, n+1) \quad \text{since } I_p(a,b) = 1 - I_{1-p}(b,a)$$

$$= I_{1/m}(1/k, n+1), \quad n = 0, 1, 2, \cdots$$

where $I_p(a,b) = B_p(a,b)/B(a,b)$.

Theorem 2.2

Let X and Y be i.i.d $H_1(m,k,1/k)$ r.vs. Then the conditional distribution of X given $X+Y$ is;

$$P(X = 1+nk / X+Y = 2+tk) = \frac{\binom{(1/k)+n-1}{n}\binom{(1/k)+t-n-1}{t-n}}{\binom{(2/k)+t-1}{t}}$$

where $n = 0,1,2,\ldots, t$ and $t = 0,1,2,\ldots$.

In particular, if $k = 1$, the conditional distribution is uniform.

Proof.

$$P(X = 1+nk) = P(Y = 1+nk) = \binom{(1/k)+n-1}{n}\left(\frac{1}{m}\right)^{\frac{1}{k}}\left(1-\frac{1}{m}\right)^n$$

$$P(X=1+nk/X+Y=2+tk)=\frac{P(X=1+nk,Y=1+(t-n)k)}{P(X+Y=2+tk)}$$

$$=\frac{P(X=1+nk)P(Y=1+(t-n)k)}{\sum_{n=0}^{t}P(X=1+nk)P(Y=1+(t-n)k)}$$

$$=\frac{\binom{(1/k)+n-1}{n}\binom{(1/k)+t-n-1}{t-n}}{\sum_{n=0}^{t}\binom{(1/k)+n-1}{n}\binom{(1/k)+t-n-1}{t-n}}$$

$$=\frac{\binom{(1/k)+n-1}{n}\binom{(1/k)+t-n-1}{t-n}}{\binom{(2/k)+t-1}{t}}$$

If $k=1$, that is, if X and Y are i.i.d geometric r.vs, then

$$P(X=1+n/X+Y=2+t)=1/(1+t),\quad n=0,1,2,\ldots t;\quad t=0,1,2,\cdots$$

and hence the conditional distribution is uniform.

2.3.3 Moments

Let $\mu_{(r)}$, μ_r' and μ_r denote respectively the r^{th} factorial moment, r^{th} moment about zero and r^{th} central moment of a distribution. We now evaluate the first four of them for the $H_1(m,k,1/k)$ distribution. Using the result 2.1 we can determine the various factorial moments. Now using the relationship between the moments we can find the raw moments and central moments. From this one can identify the mean and variance of the $H_1(m,k,1/k)$ distribution. Also the moments of Harris distribution can be easily determined via the moments of $NB(1/m,1/k)$ distribution. We have

$$
\begin{aligned}
\mu_{(1)} &= m \\
\mu_{(2)} &= m\,(m-1)\,(k+1) \\
\mu_{(3)} &= m\,(m-1)\,(k+1)\,(k-m+2\,(k+1)\,(m-1)\,) \\
\mu_{(4)} &= m\,(m-1)\,(k+1)\left(6m^2k^2+5m^2k+m^2-6mk^2-13mk-5m+k^2+5k+6\right)
\end{aligned}
$$

$$
\begin{aligned}
\mu_1' &= m \\
\mu_2' &= m^2 + m\,(m-1)\,k \\
\mu_3' &= m^3 + m\,(m-1)\,k\,(2mk+3m-k\,) \\
\mu_4' &= m^4 + m\,(m-1)\,k\left(6m^2k^2+11m^2k+6m^2-6mk^2-7mk+k^2\,\right)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mu_1 &= 0 \\
\mu_2 &= m\,(m-1)\,k \\
\mu_3 &= m\,(m-1)\,(2m-1)\,k^2 \\
\mu_4 &= m\,(m-1)\left(k\,(6m^2-6m+1)+3\,m\,(m-1)\right)k^2
\end{aligned}
$$

Harris distribution is positively skewed since;

$$\mu_3 = m\,(m-1)\,(2m-1)k^2 > 0.$$

$$\beta_1 = \frac{(2m-1)^2\,k}{m\,(m-1)}$$

$$\gamma_1 = (2m-1)\sqrt{\frac{k}{m\,(m-1)}}$$

Harris distribution is leptokurtic as $\beta_2 > 3$;

$$\beta_2 = 3 + 6k + \frac{k}{m\,(m-1)}$$

$$\gamma_2 = 6k + \frac{k}{m\,(m-1)}$$

It's coefficient of variation is; $\mathrm{C.V} = \sqrt{(1-(1/m))k}$.

From the p.g.f of $H_1(m,k,1/k)$ it follows that the moment generating function is

$$M_X(t) = \frac{e^t}{(m-(m-1)\,e^{tk})^{1/k}}$$

and so the cumulant generating function is

$$K_X(t) = t - \frac{1}{k}\log(m) - \frac{1}{k}\log\left(1-\left(\frac{m-1}{m}\right)e^{tk}\right)$$

The r^{th} cumulant $k_r = \left[\dfrac{d^r}{dt^r}K_X(t)\right]_{t=0}$ for $r = 1,2,\ldots$

Hence the first four cumulants are;

$$k_1 = m$$

$$k_2 = m\,(m-1)\,k$$

$$k_3 = m\,(m-1)\,(2m-1)\,k^2$$

$$k_4 = m\,(m-1)\,(6m^2-6m+1)\,k^3$$

We have the following recurrence relations for the moments.

$$\mu_r' = E(X^r) = \sum_{n=0}^{\infty}(1+nk)^r P(X=1+nk)$$

$$= \sum_{n=0}^{\infty}(1+nk)^r \binom{(1/k)+n-1}{n}\left(\frac{1}{m}\right)^{1/k}\left(1-\frac{1}{m}\right)^n$$

Differentiation w.r.t m and simplification gives

$$\mu_{r+1}' = m\left((m-1)k\frac{d\mu_r'}{dm}+\mu_r'\right), \qquad r = 0,1,2,\ \ldots$$

Now, $\mu_r = E(X-m)^r = \sum_{n=0}^{\infty}(1+nk-m)^r P(X=1+nk)$ gives

$$\mu_{r+1} = m\,(m-1)k\left(\frac{d\mu_r}{dm}+r\mu_{r-1}\right), \qquad r = 1,2,\ \ldots$$

Since the r^{th} cumulant $k_r = \left[\frac{d^r}{dt^r} K_X(t)\right]_{t=0}$ and $\frac{dk_r}{dm} = \left[\frac{d^r}{dt^r}\frac{d}{dm} K_X(t)\right]_{t=0}$

the cumulants satisfy the relationship

$$k_{r+1} = m\,(m-1)\,k\frac{dk_r}{dm}, \qquad r = 1, 2, \ldots$$

From the relation (2.3.3) it can be seen that;

$$P(X = 1 + nk) > P(X = 1 + (n+1)k), \qquad n = 0, 1, 2, \ldots$$

and hence there is a single mode at $X = 1$.

2.3.4 Additive Property

If $X_1, X_2, \ldots, X_n$ are n i.i.d random variables following Harris distribution $H_1(m,k,1/k)$

then, $T = \sum_{i=1}^{n} X_i$ follows $H_n(m,k,n/k)$.

Probability generating function of X_i is given by $P_{X_i}(s) = \frac{s}{(m-(m-1)s^k)^{1/k}}$. Then

the p.g.f of $T = \sum_{i=1}^{n} X_i$ is

$$P_{X_1+X_2+\ldots+X_n}(s) = P_{X_1}(s)P_{X_2(s)}\ldots P_{X_n}(s) = \frac{s^n}{(m-(m-1)s^k)^{n/k}}$$

This is the p.g.f of Harris distribution $H_n(m,k,n/k)$ and hence the additive property.

Now we present a characterization of the Harris distribution motivated by a characterization of the negative binomial distribution by Sathe and Ravi (1997).

Theorem 2.3

Let X be a r.v with $P\{X = nk+1\} = a_{nk+1}$, $n = 0, 1, 2, \ldots$ and let $E(X) = \mu$ be finite.

Then X has Harris distribution if and only if

$$\frac{dP(X > nk+1)}{d\mu} = \frac{nk+1}{\mu k} a_{nk+1}, \qquad n = 0,1,2, \ldots, \quad k>0 \text{ an integer.} \tag{2.3.4}$$

Proof.

Let $X \sim H_1(\mu,k,1/k)$. Then,

$$P(X > nk+1) = 1 - P(X \le nk+1) = 1 - \sum_{r=0}^{n} \binom{(1/k)+r-1}{r} \left(\frac{1}{\mu}\right)^{1/k} \left(1-\frac{1}{\mu}\right)^r \quad \text{and}$$

$$\frac{d}{d\mu} P(X > nk+1) = -\frac{d}{d\mu} \sum_{r=0}^{n} \binom{(1/k)+r-1}{r} \left(\frac{1}{\mu}\right)^{1/k} \left(1-\frac{1}{\mu}\right)^r$$

$$= -\sum_{r=0}^{n} \binom{(1/k)+r-1}{r} \frac{d}{d\mu} \left(\frac{1}{\mu}\right)^{1/k} \left(1-\frac{1}{\mu}\right)^r$$

Now differentiation and simplification gives

$$\frac{dP(X > nk+1)}{d\mu} = \frac{1}{k}\left(\frac{1}{\mu}\right)^{\frac{1}{k}+1} \sum_{r=0}^{n} \binom{(1/k)+r-1}{r} \left(1-\frac{1}{\mu}\right)^r - \left(\frac{1}{\mu}\right)^{\frac{1}{k}+2} \sum_{r=0}^{n} r \binom{(1/k)+r-1}{r} \left(1-\frac{1}{\mu}\right)^{r-1}$$

$$= \frac{1}{k}\left(\frac{1}{\mu}\right)^{\frac{1}{k}+1} \sum_{r=0}^{n} \binom{(1/k)+r-1}{r} \left(1-\frac{1}{\mu}\right)^r - \left(\frac{1}{\mu}\right)^{\frac{1}{k}+2} \frac{1}{k} \sum_{r=1}^{n} \binom{(1/k)+r-1}{r-1} \left(1-\frac{1}{\mu}\right)^{r-1}$$

$$= \frac{1}{k}\left(\frac{1}{\mu}\right)^{\frac{1}{k}+1} \sum_{r=0}^{n} \binom{(1/k)+r-1}{r} \left(1-\frac{1}{\mu}\right)^r - \left(\frac{1}{\mu}\right)^{\frac{1}{k}+2} \frac{1}{k} \sum_{r=0}^{n-1} \binom{(1/k)+r}{r} \left(1-\frac{1}{\mu}\right)^{r}$$

$$= \frac{1}{k}\left(\frac{1}{\mu}\right)^{\frac{1}{k}+1} \left[\sum_{r=0}^{n} \binom{(1/k)+r-1}{r} \left(1-\frac{1}{\mu}\right)^r - \frac{1}{\mu} \sum_{r=0}^{n-1} \binom{(1/k)+r}{r} \left(1-\frac{1}{\mu}\right)^{r} \right]$$

$$= \frac{1}{k}\left(\frac{1}{\mu}\right)^{\frac{1}{k}+1} \left[\sum_{r=0}^{n} \binom{(1/k)+r-1}{r} \left(1-\frac{1}{\mu}\right)^r - \left(1-\left(1-\frac{1}{\mu}\right)\right) \sum_{r=0}^{n-1} \binom{(1/k)+r}{r} \left(1-\frac{1}{\mu}\right)^{r} \right]$$

$$= \frac{1}{k}\left(\frac{1}{\mu}\right)^{\frac{1}{k}+1} \left[\sum_{r=0}^{n} \binom{(1/k)+r-1}{r} \left(1-\frac{1}{\mu}\right)^r - \sum_{r=0}^{n-1} \binom{(1/k)+r}{r} \left(1-\frac{1}{\mu}\right)^{r} + \sum_{r=0}^{n-1} \binom{(1/k)+r}{r} \left(1-\frac{1}{\mu}\right)^{r+1} \right]$$

$$=\frac{1}{k}\left(\frac{1}{\mu}\right)^{\frac{1}{k}+1}\left[\sum_{r=0}^{n}\binom{(1/k)+r-1}{r}\left(1-\frac{1}{\mu}\right)^{r}-\sum_{r=0}^{n-1}\left\{\binom{(1/k)+r-1}{r}+\binom{(1/k)+r-1}{r-1}\right\}\left(1-\frac{1}{\mu}\right)^{r}\right.$$

$$\left.+\sum_{r=0}^{n-1}\binom{(1/k)+r}{r}\left(1-\frac{1}{\mu}\right)^{r+1}\right]$$

$$=\frac{1}{k}\left(\frac{1}{\mu}\right)^{\frac{1}{k}+1}\left[\sum_{r=0}^{n}\binom{(1/k)+r-1}{r}\left(1-\frac{1}{\mu}\right)^{r}-\sum_{r=0}^{n-1}\binom{(1/k)+r-1}{r}\left(1-\frac{1}{\mu}\right)^{r}\right.$$

$$\left.-\sum_{r=1}^{n-1}\binom{(1/k)+r-1}{r-1}\left(1-\frac{1}{\mu}\right)^{r}+\sum_{r=0}^{n-1}\binom{(1/k)+r}{r}\left(1-\frac{1}{\mu}\right)^{r+1}\right]$$

$$=\frac{1}{k}\left(\frac{1}{\mu}\right)^{\frac{1}{k}+1}\left[\binom{(1/k)+n-1}{n}\left(1-\frac{1}{\mu}\right)^{n}-\sum_{r=0}^{n-2}\binom{(1/k)+r}{r}\left(1-\frac{1}{\mu}\right)^{r+1}+\sum_{r=0}^{n-1}\binom{(1/k)+r}{r}\left(1-\frac{1}{\mu}\right)^{r+1}\right]$$

$$=\frac{1}{k}\left(\frac{1}{\mu}\right)^{\frac{1}{k}+1}\left[\binom{(1/k)+n-1}{n}\left(1-\frac{1}{\mu}\right)^{n}+\binom{(1/k)+n-1}{n-1}\left(1-\frac{1}{\mu}\right)^{n}\right]$$

$$=\frac{1}{k}\left(\frac{1}{\mu}\right)^{\frac{1}{k}+1}\binom{(1/k)+n-1}{n}\left(1-\frac{1}{\mu}\right)^{n}(1+nk)$$

$=\dfrac{nk+1}{\mu k}a_{nk+1}$, $n=0,1,2,\ldots$, which is condition (2.3.4).

Now, assuming the condition (2.3.4) and using $P(X>nk+1)=1-P(X\le nk+1)$ we have

$$-\frac{d}{d\mu}\sum_{r=0}^{n}P(X=rk+1)=\frac{nk+1}{\mu k}a_{nk+1},\quad n=0,1,2,\ldots$$

$$-\sum_{r=0}^{n}\frac{da_{rk+1}}{d\mu}=\frac{nk+1}{\mu k}a_{nk+1},\quad n=0,1,2,\ \ldots$$

Consequently, we get

$$-\frac{da_1}{d\mu}=\frac{a_1}{\mu k} \qquad \text{and}$$

$$-\frac{da_{nk+1}}{d\mu} = \frac{nk+1}{\mu k}a_{nk+1}-\frac{(n-1)k+1}{\mu k}\,a_{(n-1)k+1}, \quad n\geq 1 \tag{2.3.5}$$

Defining $P(\mu,s)=\sum_{n=0}^{\infty}a_{nk+1}\,s^{nk+1}$, $0<s\leq 1$, we get

$$\frac{\partial P}{\partial s}=\sum_{n=0}^{\infty}(nk+1)a_{nk+1}\,s^{nk} \qquad \text{and} \qquad \frac{\partial P}{\partial \mu}=\frac{\partial a_1}{\partial \mu}\,s+\sum_{n=1}^{\infty}\frac{\partial a_{nk+1}}{\partial \mu}\,s^{nk+1}$$

Thus from (2.3.5) we get

$$\frac{\partial P}{\partial \mu}=-\frac{a_1 s}{\mu k}-\sum_{n=1}^{\infty}\left[\frac{nk+1}{\mu k}a_{nk+1}-\frac{(n-1)k+1}{\mu k}a_{(n-1)k+1}\right]s^{nk+1}$$

$$-\mu k\frac{\partial P}{\partial \mu}=a_1 s+\sum_{n=1}^{\infty}(nk+1)\;a_{nk+1}\,s^{nk+1}-\sum_{n=1}^{\infty}\big((n-1)k+1\big)\;a_{(n-1)k+1}\,s^{nk+1}$$

$$=a_1 s+s\sum_{n=1}^{\infty}(nk+1)\;a_{nk+1}\,s^{nk}-s^{k+1}\sum_{n=1}^{\infty}\big((n-1)k+1\big)a_{(n-1)k+1}\,s^{(n-1)k}$$

$$=s\sum_{n=0}^{\infty}(nk+1)a_{nk+1}\,s^{nk}-s^{k+1}\sum_{n=1}^{\infty}\big((n-1)k+1\big)a_{(n-1)k+1}\,s^{(n-1)k}$$

$$=s\frac{\partial P}{\partial s}-s^{k+1}\frac{\partial P}{\partial s}$$

$$\mu k\frac{\partial P}{\partial \mu}+s\left(1-s^{k}\right)\frac{\partial P}{\partial s}=0 \tag{2.3.6}$$

This is of the form $W(x,y,z)\frac{\partial z}{\partial x}+Q(x,y,z)\frac{\partial z}{\partial y}=R(x,y,z)$ where x, y are independent variables and z is a dependent variable. The general solution of this equation is $F(u,v)=0$ where F is an arbitrary function of u and v and $u(x,y,z)=c_1$ and

$v(x,y,z)=c_2$ are the solutions of the system $\dfrac{dx}{W(x,y,z)}=\dfrac{dy}{Q(x,y,z)}=\dfrac{dz}{R(x,y,z)}$. See section 3 in Sneddon (1957, p.10). From (2.3.6) we get the system of differential equations

$$\frac{d\mu}{\mu k}=\frac{ds}{s(1-s^k)}=\frac{dP}{0}$$

which is similar to the example 1.4.4 in Amaranath (1997, p.16).

Integrating, $\dfrac{d\mu}{\mu k}=\dfrac{ds}{s(1-s^k)}$ gives

$$\frac{1}{k}\log\mu=\int\frac{ds}{s(1-s^k)}+c, \text{ where } c \text{ is a constant}$$

$$=\frac{1}{k}\log\left(\frac{s^k}{1-s^k}\right)+c$$

or $\quad \mu(1-s^k)/s^k = c$

Then we have, $u=c_1=\mu(1-s^k)/s^k$

Now, $\dfrac{d\mu}{\mu k}=\dfrac{dP}{0}\Rightarrow\dfrac{dP}{d\mu}=0\Rightarrow P$ is a constant $=c_2=v$. Notice that $v=P(\mu,s)$.

The general solution of (2.3.6) is $F(u,v)=0$. *i.e,* $v=f(u)$

$$P(\mu,s)=f\left(\mu(1-s^k)/s^k\right) \tag{2.3.7}$$

where f is determined by the initial conditions $\mu=1$ and $P(1,s)=s$.

Putting $\mu=1$ in (2.3.7) we get

$$s=f\left((1-s^k)/s^k\right) \tag{2.3.8}$$

Now, taking $(1-s^k)/s^k=y$ or $s=\{1/(y+1)\}^{1/k}$ in (2.3.8) we have

$$f(y)=\{1/(y+1)\}^{1/k}$$

And hence from (2.3.7),

$$P(\mu,s) = \left(\frac{\mu(1-s^k)}{s^k}+1\right)^{-1/k} = \left(\frac{s^k}{\mu-\mu s^k+s^k}\right)^{1/k}$$

which is the p.g.f of $H_1(\mu,k,1/k)$. This completes the proof.

2.4 Divisibility Properties of Harris Family

The p.g.f of the $H_0(m,k,1/k)$ distribution is $P(s) = \{m-(m-1)s^k\}^{-1/k}$ which corresponds to a r.v $U = kV$, where V is $NB(1/m,1/k)$ with p.g.f $\{m-(m-1)s\}^{-1/k}$. We know that a negative binomial distribution is i.d. Since infinite divisibility is not affected by change of origin and scale, it follows that U is i.d. Hence;

Theorem 2.4

$H_0(m,k,1/k)$ and $H_1(m,k,1/k)$ r.vs are i.d .

Sreehari (1979) has shown that *Gamma*(β,c) distributions are self-decomposable. We now give an alternate proof that *Gamma*($1/k$, $1/c$) distributions are self-decomposable using its Harris–sum stability. As a consequence of this we get that $H_0(m,k,1/k)$ distributions are self-decomposable. Satheesh *et al.* (2002) showed that *Gamma*($1/k$, $1/c$) law is $H_1(m,k,1/k)$-sum stable in the sense:

$$\frac{(1+cs)^{-1/k}}{[m-(m-1)/(1+cs)]^{1/k}} = \frac{(1+cs)^{-1/k}}{[(1+mcs)/(1+cs)]^{1/k}} = (1+mcs)^{-1/k}.$$

From the above when $m = 1/c$ we have

$$\left(1+s\right)^{-1/k} = \left(1+cs\right)^{-1/k}\left(m-(m-1)(1+cs)^{-1}\right)^{-1/k}.$$

Here the second factor on the right hand side is the LT of a $H_0(m,k,1/k)$-sum of *Gamma*$(1/k,\ 1/c)$ variables. The relation is also true for each $0<c<1$ since we may choose the $H_0(m,k,1/k)$ with $m=1/c$ accordingly. Hence *Gamma*$(1/k,\ 1/c)$ r.v is self-decomposable. Consequently the corresponding negative binomial r.vs are also self-decomposable by corollary 2.1 in Satheesh and Nair (2002*a*). Since self-decomposability is not affected by change of scale we have proved;

Theorem 2.5

$H_0(m,k,1/k)$ r.vs are self-decomposable.

Remark 2.2

$H_1(m,k,1/k)$ r.vs are not self-decomposable as it has no probability at zero, a necessary condition for a discrete distribution to be self-decomposable, (Satheesh 2004*a*).

Kozubowski and Panorska (1998) have developed ν- stable laws and according to them the ν-stable law with c.f $\{1+|t|^{\alpha}\}^{-\beta}$ is obtained as the weak limit of negative binomial sums. The above c.f is that of the generalized Linnik distribution. But negative binomial r.v does not induce N-sum stability in the generalized Linnik distributions. It is the $H_1(m,k,1/k)$ r.v that induces N-sum stability in generalized Linnik distributions and that too only when $\beta = 1/k$, see Satheesh *et al.* (2002). Here we show that the above c.f can be obtained as the weak limit of Harris-sums also. Let $N_m \sim H_1(m,k,1/k)$. The next lemma follows from Satheesh (2004*b*).

Lemma 2.1

$$\frac{N_m}{m} \xrightarrow{d} kU \text{ as } m \to \infty \text{ where } U \text{ is } Gamma(1/k,\ 1).$$

Lemma 2.2

$$N_m \xrightarrow{p} \infty \text{ as } m \to \infty.$$

Proof.

Let $p_{m,k} = P\{N_m = k\}$, $k = 1, 2, \ldots$ be the probability distribution corresponding to the r.vs $\{N_m, m>1\}$. Then $N_m \xrightarrow{p} \infty$ as $m \to \infty$ is equivalent to $p_{m,k} \to 0$ as $m \to \infty$ for every $k = 1, 2, \ldots$. In terms of p.g.fs this is equivalent to $\underset{m\to\infty}{Lt}\ P_m(s) = 0$ for all $s \in (0,1)$. Since gamma distribution with LT φ is absolutely continuous, $\varphi(\infty) = 0$ and since the p.g.f of N_m is derived from the LT φ (see (vi) on page.7), $\underset{m\to\infty}{Lt}\ P_m(s) = 0$. That completes the proof of the lemma.

This lemma enables one to invoke the Szasz (1972) transfer theorem where the conditions to be satisfied by the indexing r.v N_m are, $\frac{N_m}{m} \xrightarrow{d} v$ and $N_m \xrightarrow{p} \infty$ as $m \to \infty$. Here v is the r.v w.r.t which v- stable laws are described. In our case it is the *Gamma*$(1/k,k)$ r.v. The following result now follows by invoking Szasz's (1972) theorem.

Theorem 2.6

Let $S_n = X_1+X_2+ \ldots + X_n$ where $X_1, X_2, \ldots$are i.i.d r.vs and $S_n \xrightarrow{d} U$ so that U is i.d with c.f $\exp\{-\psi(t)\}$. Then $S_{N_m} \xrightarrow{d} V$ where the c.f of V is $\{1+\psi(t)\}^{-1/k}$. The converse is also true.

Notice that irrespective of whether $N_m \sim H_0(m,k,1/k)$ or $H_1(m,k,1/k)$ we get the above result since the conclusion in lemma 2.1 & 2.2 are true for both, see Satheesh (2004*b*). Notice also that instead of $N_m \sim H_1(m,k,1/k)$ we may consider the general case of $N_m \sim H_1(m,k,\beta)$ and see that the above arguments hold good. This result extends ν-stable distributions of Kozubowski and Panorska (1998) when ν is gamma where the discussion is limited to negative binomial sums. Also, this is a stronger description of φ-stable distributions of Satheesh (2004*b*) when φ is the LT of a gamma distribution since there the description is based only on lemma 2.1.

2.5 Estimation and Simulation

Usually in Harris distribution the parameter k will be known from the data or from the mechanism generating it, since k is the gap in the support of the distribution. Hence we need estimate the parameter m only. But in clinical trials and biological experiments the trials are to be carried out with the pre-estimated gap. Hence we assume that both the parameters are unknown. We develop maximum likelihood estimators and moment estimators for both parameters m and k of the Harris family of distributions and then evaluate their performances by simulation of the distribution. The obtained estimate of k may not be integer valued and hence the integral part is taken as the safer estimate of k.

2.5.1 Estimator of *m* When *k* is Known

(a) Method of Maximum Likelihood

The p.m.f of $H_1(m,k,1/k)$ is given by

$$f(x)=\binom{(1/k)+r-1}{r}\left(\frac{1}{m}\right)^{1/k}\left(1-\frac{1}{m}\right)^{r}, \quad \text{where } \ x=1+rk, \quad r=0,1,2,\ldots$$

Consider a sample $X=(X_1, X_2, \ldots, Xn)$ from this distribution. Let $x =(x_1, x_2, \ldots, x_i, \ldots, x_n)$ denotes the observed values of X. Then the likelihood function is given by

$$L=\prod_{i=1}^{n} f(x_i) \quad \text{where } \ x_i = 1+r_i k, \quad i=1,2,\ldots n; \quad r_i = 0,1,2,\ldots$$

$$=\prod_{i=1}^{n}\binom{(1/k)+r_i-1}{r_i}\left(\frac{1}{m}\right)^{1/k}\left(1-\frac{1}{m}\right)^{r_i}$$

$$=\left(\frac{1}{m}\right)^{\frac{n}{k}}\left(1-\frac{1}{m}\right)^{\sum_{i=1}^{n} r_i}\prod_{i=1}^{n}\binom{(1/k)+r_i-1}{r_i}, \quad \text{where } \sum_{i=1}^{n} r_i=\sum_{i=1}^{n}\left(\frac{x_i-1}{k}\right)=\frac{n}{k}(\bar{x}-1)$$

$$\log L=\sum_{i=1}^{n}\log\binom{(1/k)+r_i-1}{r_i}+\frac{n}{k}\log\left(\frac{1}{m}\right)+\frac{n}{k}(\bar{x}-1)\log\left(1-\frac{1}{m}\right)$$

$$\frac{\partial \log L}{\partial m}=-\frac{n}{km}+\frac{n(\bar{x}-1)}{km(m-1)}=\frac{n}{m(m-1)k}(\bar{x}-m)$$

Solution of $\dfrac{\partial \log L}{\partial m}=0$ gives the maximum likelihood estimator of m,

$$\hat{m}=\bar{x}.$$

(b) Method of Moments

Using the method of moments, that is, by equating the sample mean $\bar{x}$ to the population mean, we have the following estimate for m.

$$\hat{m}=\bar{x}.$$

2.5.2 Estimators When Both the Parameters are Unknown

Consider the situation where both parameters m and k are unknown. In this case we develop the maximum likelihood and moment estimators.

(a) Method of Maximum Likelihood

Putting $K=1/k$ and $p=1/m$ in the p.m.f of $H_1(m,k,1/k)$ we have the likelihood function

$$L=\prod_{i=1}^{n}\binom{K+r_i-1}{r_i}p^{nK}(1-p)^{n\bar{r}}, \qquad \bar{x}=1+\bar{r}k=1+(\bar{r}/K)$$

Now, following the development in Simon (1961) we have;

$$\log L=\sum_{i=1}^{n}\log\binom{K+r_i-1}{r_i}+nK\log p+n(\bar{x}-1)K\log(1-p)$$

$$\frac{\partial\log L}{\partial p}=\frac{nK}{p}-\frac{n(\bar{x}-1)K}{1-p}=0 \quad \text{gives} \quad \hat{p}=\frac{1}{\bar{x}} \quad \text{and hence} \quad \hat{m}=\bar{x}$$

The maximum likelihood equation corresponding to K is

$$\sum_{i=1}^{n}\left(\frac{1}{K}+\frac{1}{K+1}+\cdots+\frac{1}{K+r_i-2}+\frac{1}{K+r_i-1}\right)+n\log p+n(\bar{x}-1)\log(1-p)=0$$

Substituting the value of p, we have

$$\sum_{i=1}^{n}\left(\frac{1}{K}+\frac{1}{K+1}+\cdots+\frac{1}{Kx_i-2}+\frac{1}{Kx_i-1}\right)+n\log\left(\frac{1}{\bar{x}}\right)+n(\bar{x}-1)\log\left(1-\frac{1}{\bar{x}}\right)=0$$

This equation has only one unknown K and, therefore, can be solved by trial and error method. Then $\hat{k}=1/\hat{K}$.

(b) Method of Moments

Using the method of moments, that is, by equating the sample mean $\bar{x}$ and the sample variance s^2 to the corresponding population values we have the following estimates.

$$\hat{m}=\bar{x} \qquad \text{and} \qquad \hat{k}=\frac{s^2}{\bar{x}(\bar{x}-1)}.$$

The maximum likelihood estimate and the moment estimate of m coincide. The results based on simulation are presented in the tables appended.

2.5.3 Unbiased Estimator of '*m*'

Given a sample of *n* observations $x_1, x_2, \ldots, x_n$ from a Harris distribution $H_1(m,k,1/k)$, we have

$$E(\bar{x}) = E\left(\frac{\sum x_i}{n}\right) = \frac{1}{n}\sum E(x_i) = \frac{1}{n}\sum m = m$$

$\bar{x}$ is an unbiased estimator of m.

2.5.4 Minimum Variance Unbiased Estimator of '*m*'

Proceeding as given in section 2.5.1, we have

$$\frac{\partial \log L}{\partial m} = -\frac{n}{km} + \frac{n(\bar{x}-1)}{km(m-1)} = \frac{n}{m(m-1)k}(\bar{x}-m)$$

Clearly, by the method of minimum variance, the minimum variance unbiased estimator (MVUE) for m is

$$\hat{m} = \bar{x}$$

with minimum variance, $V(\bar{x}) = m(m\text{-}1)k/n$.

2.5.5 Bayes Estimate for *m*

The Bayes estimate for *m* is obtained by proceeding as given in Rohatgi (1990).

Let $x_1, x_2, \ldots, x_i, \ldots, x_n$ be *n* observations taken from the Harris distribution $H_1(m,k,1/k)$. Take $p = 1/m$ in the p.m.f of the distribution.

Let the a priori distribution of p be uniform with p.d.f $\pi(p) = 1$, $0 < p < 1$. Also, let the loss function $L(p, d) = [p - d(X)]^2$.

If $X = x$ is observed, then the joint p.m.f of X's and p is given by

$$u(x,p) = \pi(p)u(x/p)$$

$$= \pi(p)\prod_{i=1}^{n} f(x_i/p)$$

$$= p^{n/k}(1-p)^z \prod_{i=1}^{n} \binom{(1/k)+r_i-1}{r_i}, \quad \text{where } z = \sum_{i=1}^{n} r_i = \sum_{i=1}^{n}\left(\frac{x_i-1}{k}\right) = \frac{n}{k}(\bar{x}-1)$$

and the joint marginal p.m.f of X is given by

$$g(x) = \int_0^1 u(x,p)dp = \int_0^1 \left[\prod_{i=1}^{n} \binom{(1/k)+r_i-1}{r_i}\right] p^{n/k}(1-p)^z dp.$$

It follows that the posterior distribution of p given x is

$$h(p/x) = \frac{u(x,p)}{g(x)}$$

$$= \frac{p^{n/k}(1-p)^z \prod_{i=1}^{n} \binom{(1/k)+r_i-1}{r_i}}{\int_0^1 p^{n/k}(1-p)^z \left[\prod_{i=1}^{n} \binom{(1/k)+r_i-1}{r_i}\right] dp}$$

$$= \frac{p^{n/k}(1-p)^z}{\int_0^1 p^{n/k}(1-p)^z dp}$$

$$= \frac{p^{n/k}(1-p)^z}{\beta((n/k)+1, z+1)},$$

and the Bayes estimate is

$$d^*(x) = E(p/X=x)$$

$$= \int_p p\, h(p/x) dp$$

$$= \int_0^1 p \frac{p^{n/k}(1-p)^z}{\beta((n/k)+1, z+1)} dp$$

$$= \frac{\beta((n/k)+2, z+1)}{\beta((n/k)+1, z+1)}$$

Substituting the value of z and simplifying we get, $d^*(x) = \dfrac{n+k}{n\bar{x}+2k}$.

This is the Bayes estimate of p w.r.t π, the a priori distribution of p and hence the Bayes estimate of the Harris parameter m is

$$\tilde{m}(n, \bar{x}) = \frac{n\bar{x}+2k}{n+k} .$$

The following figures give us an idea about the shape of the Harris distribution. They correspond to the probability distribution given below.

TABLE 2.1

Harris probability distribution

H_1(2,2,1/2)		H_1(4,5,1/5)		H_1(50,5,1/5)	
x	$P(x)$	x	$P(x)$	x	$P(x)$
1	0.707107	1	0.757858	1	0.457305
3	0.176777	6	0.113679	6	0.089632
5	0.066291	11	0.051155	11	0.052704
7	0.027621	16	0.028136	16	0.037876
9	0.012084	21	0.016881	21	0.029695
11	0.005438	26	0.010635	26	0.024445
13	0.002492	31	0.006913	31	0.020762
15	0.001157	36	0.004592	36	0.018021
17	0.000542	41	0.003100	41	0.015895
19	0.000256	46	0.002118	46	0.014192
21	0.000122	51	0.001462	51	0.012796
23	5.81E-05	56	0.001016	56	0.011628
25	2.78E-05	61	0.000711	61	0.010636
27	1.34E-05	66	0.000501	66	0.009781
		71	0.000354	71	0.009038
		76	0.000251	76	0.008385
		81	0.000179	81	0.007806
		86	0.000128	86	0.007290
		91	9.18E-05	91	0.006827

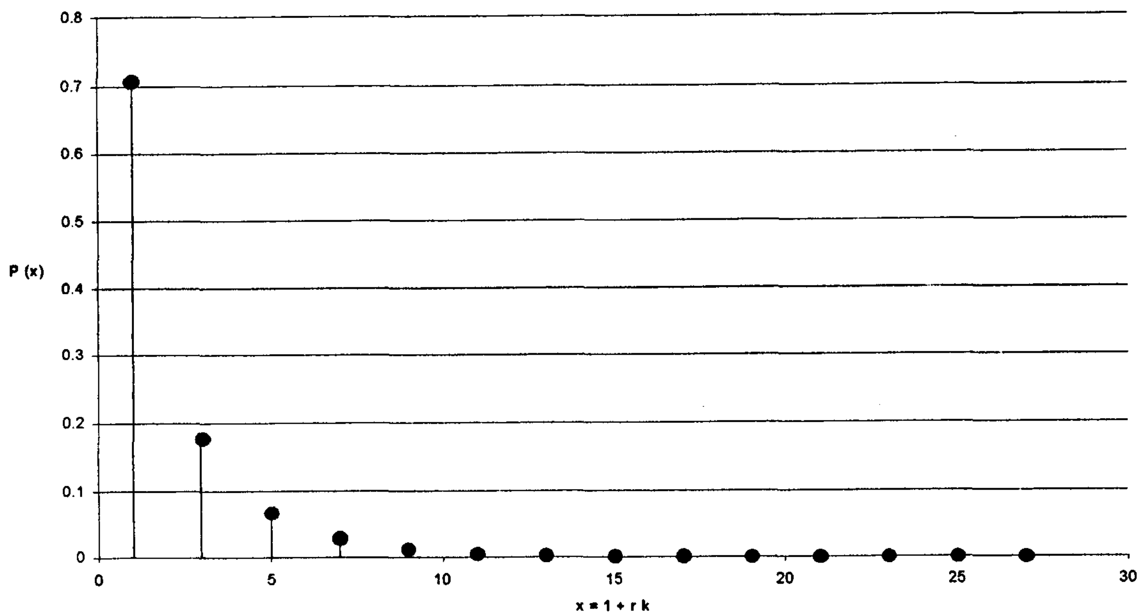


FIGURE 2.1 – Probability plot of $H_1(2, 2, \frac{1}{2})$

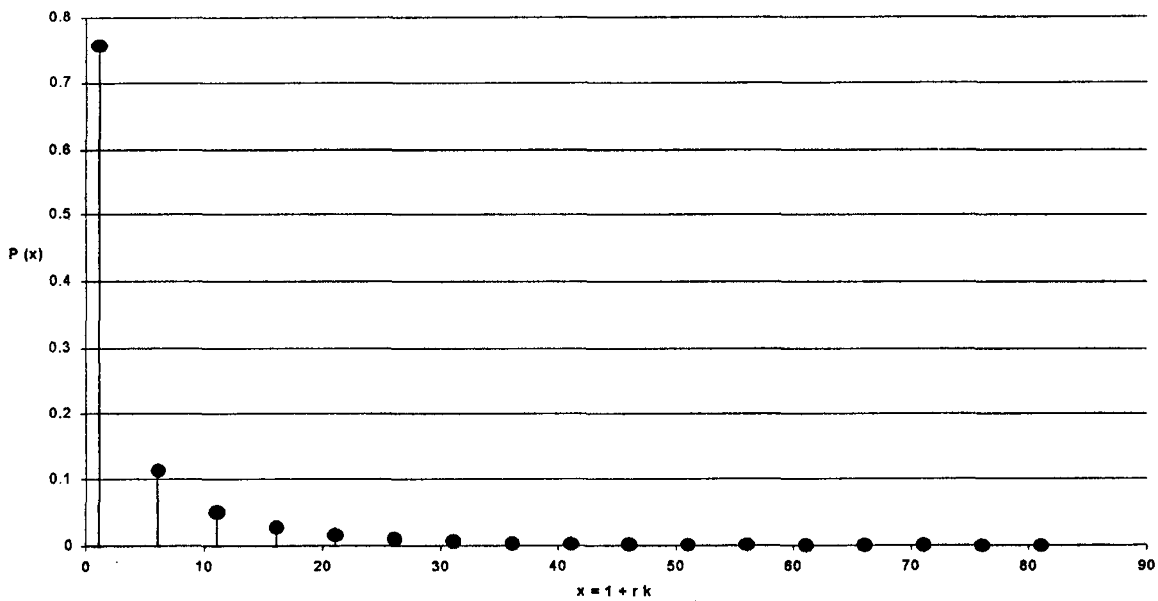


FIGURE 2.2 – Probability plot of $H_1(4, 5, 1/5)$

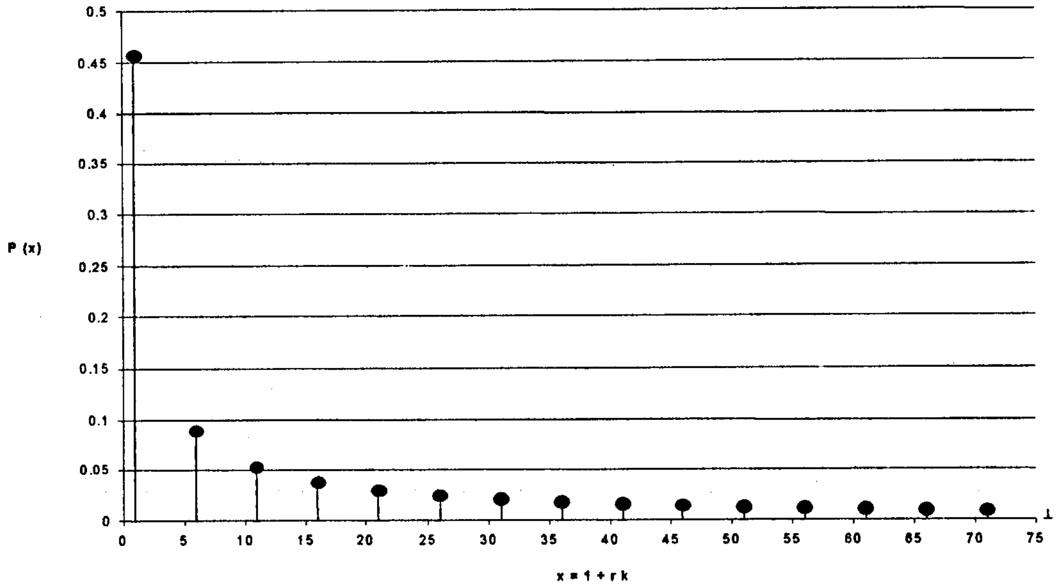


FIGURE 2.3 - Probability plot of $H_1(50,5,1/5)$

TABLE 2.2

Maximum likelihood estimates of m and k using simulated samples of different sizes n

n	k=2, m=10 $\hat{k}$	$\hat{m}$	k=2, m=2 $\hat{k}$	$\hat{m}$	k=4, m=2 $\hat{k}$	$\hat{m}$
100	2.44	8.56	2.00	2.06	4.95	1.72
200	2.53	10.8	1.91	1.94	5.00	1.78
300	2.49	10.3	1.92	1.95	4.81	1.91
400	2.49	9.77	1.98	1.98	4.83	1.89
500	2.48	9.83	2.00	2.00	4.95	1.85

TABLE 2.3

Moment estimates of m and k using simulated sample of size 200 and number of repetitions 50

	k	2		4		10		20		30		50	
m		$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$
1.25	Estimate	1.25	2.00	1.27	3.88	1.24	9.51	1.27		1.23		1.25	
	SE	0.01	0.05	0.01	0.09	0.02	0.30	0.03		0.03		0.04	
1.5	Estimate	1.52	1.97	1.52	3.90	1.52	9.72	1.51	18.9	1.47		1.50	
	SE	0.01	0.05	0.02	0.15	0.03	0.41	0.04	0.85	0.04		0.06	
2	Estimate	2.01	1.97	2.04	3.74	1.95	9.13	2.03	17.4	1.88		1.98	
	SE	0.02	0.05	0.03	0.07	0.04	0.29	0.06	0.69	0.08		0.11	
10	Estimate	9.79	1.95	9.99	4.01	10.5	9.25	10.3	18.1	9.70	42.6	10.1	28.0
	SE	0.13	0.05	0.20	0.10	0.33	0.34	0.48	1.30	0.62	2.80	0.59	1.45
50	Estimate	50.7	2.03	50.1	4.02	49.1	9.24	50.8	18.9	50.1	27.3	46.1	42.0
	SE	0.67	0.06	0.98	0.09	1.24	0.36	2.08	0.86	2.83	1.43	2.80	2.52

TABLE 2.4

Moment estimates of m and k using simulated sample of size 100 and number of repetitions 50

	k	2		4		10		20		30		50	
m		$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$
1.25	Estimate	1.26	1.89	1.24	3.73	1.27		1.24		1.26		1.22	
	SE	0.01	0.06	0.01	0.11	0.03		0.04		0.04		0.05	
1.5	Estimate	1.53	1.95	1.50	3.89	1.48	8.55	1.50		1.47		1.44	
	SE	0.02	0.06	0.02	0.15	0.04	0.40	0.05		0.06		0.08	
2	Estimate	1.97	1.81	2.02	3.80	2.00	8.86	1.82		1.98		1.97	
	SE	0.03	0.04	0.04	0.14	0.07	0.36	0.09		0.09		0.13	
10	Estimate	9.67	1.87	9.66	4.16	9.67	8.52	10.2	18.5	9.96	19.9	10.6	35.3
	SE	0.15	0.06	0.27	0.23	0.40	0.35	0.76	1.49	0.69	1.05	1.03	2.23
50	Estimate	49.2	1.98	49.1	3.64	45.5	8.73	51.0	16.6	45.9	22.2	50.9	32.1
	SE	0.90	0.07	1.25	0.13	1.90	0.39	3.59	0.91	3.32	1.23	5.15	1.99

TABLE 2.5

Moment estimates of m and k using simulated sample of size 50 and number of repetitions 100

m	k	2		4		10		20		30		50	
		$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$
1.25	Estimate	1.25	1.91	1.24		1.23		1.31		1.22		1.27	
	SE	0.01	0.05	0.02		0.02		0.04		0.04		0.05	
1.5	Estimate	1.52	1.87	1.49	3.86	1.57		1.46		1.46		1.53	
	SE	0.02	0.05	0.02	0.11	0.04		0.05		0.06		0.08	
2	Estimate	2.05	1.86	2.08	3.72	2.06		2.08		1.91		2.07	
	SE	0.03	0.05	0.04	0.12	0.06		0.09		0.10		0.15	
10	Estimate	10.4	1.87	10.4	3.42	10.5	8.62	9.81	14.7	10.2	17.4	8.97	35.3
	SE	0.24	0.06	0.25	0.12	0.39	0.45	0.52	0.63	0.76	0.75	1.19	2.23
50	Estimate	50.7	1.95	51.6	3.83	49.4	7.94	48.5	13.7	54.0	17.9	53.2	32.1
	SE	0.92	0.06	1.36	0.17	2.34	0.28	3.22	0.58	3.80	0.73	4.98	1.99

TABLE 2.6

Moment estimates of m and k using simulated sample of size 500 and number of repetitions 50

m	k	2		4		10		20		30		50	
		$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$	$\hat{m}$	$\hat{k}$
1.25	Estimate	1.25	1.99	1.25	4.01	1.26	9.58	1.23	19.2				
	SE	0.05	0.03	0.01	0.07	0.01	0.21	0.01	0.50				
1.5	Estimate	1.49	2.01	1.48	3.92	1.50	9.72	1.50	18.8	1.43	29.1	1.57	48.6
	SE	0.01	0.03	0.01	0.08	0.02	0.22	0.02	0.54	0.03	1.70	0.04	2.88
2	Estimate	1.98	1.99	2.00	3.89	2.00	9.59	2.00	20.6	1.98	25.7	1.99	44.2
	SE	0.01	0.03	0.02	0.07	0.02	0.27	0.04	0.88	0.05	0.95	0.06	1.93
10	Estimate	10.0	2.00	10.1	4.10	9.87	9.44	10.4	19.5	9.68	29.3	9.46	44.8
	SE	0.08	0.03	0.11	0.08	0.18	0.28	0.24	0.61	0.32	1.23	0.38	2.06
50	Estimate	50.2	1.99	49.4	4.07	51.0	9.47	50.1	20.1	53.0	29.3	50.5	44.6
	SE	0.52	0.02	0.56	0.07	1.15	0.29	1.47	0.71	1.62	1.15	2.17	2.07

HARRIS DISTRIBUTION AS A MEMBER OF GENERALIZED POWER SERIES FAMILY

Sherly Sebastian "Harris family of discrete distributions and processes" Thesis. Department of Statistics , University of Calicut, 2007

CHAPTER 3

HARRIS DISTRIBUTION AS A MEMBER OF GENERALIZED POWER SERIES FAMILY

3.1 Introduction

Noack (1950) defined a r.v Z taking non-negative integral values z with probabilities

$$P(Z=z)=\frac{a(z)\theta^z}{g(\theta)}, \qquad z=0,1,\ldots,$$

where $a(z)\geq 0$; $\theta\geq 0$ and $g(\theta)=\sum_{z=0}^{\infty}a(z)\theta^z$.

The distribution of Z is called a power series distribution (*PSD*). θ is the *power parameter* of the distribution, and $g(.)$ is the *series function.*

Patil (1961,1962) allowed the set of values that the variate can take to be any nonempty enumerable set S of nonnegative integers and considered the generating function

$$C(\theta)=\sum_{x\in S}a(x)\theta^x$$

with $a(x)\geq 0$; $\theta\geq 0$ so that $C(\theta)$ is positive, finite and differentiable.

He defined a r.v X taking non-negative integral values in S with probabilities

$$P(X=x)=\frac{a(x)\theta^x}{C(\theta)}, \qquad x\in S \tag{3.1.1}$$

Patil called this distribution a *generalized power series distribution* (*GPSD*). The properties that hold for a *PSD* also hold for *GPSD*. Some of the distributions belonging to this extended class are the binomial, Poisson, negative binomial and logarithmic distributions, and their related multivariate distributions. Furthermore, the sum of n mutually independent r.vs each having the same *GPSD*, has a distribution of the same class, with series function $[C(\theta)]^n$.

In this chapter Harris distribution is shown to be a member of exponential family. It has also been proved in section 3.2 that this distribution belongs to the family of *GPSD*. It can be easily seen that proper choice of S and $C(\theta)$ reduces the *GPSD* in particular, to the Harris distribution. Also the properties of the distribution are reviewed by using the properties of *PSD* given in Johnson *et al.* (1992). Estimators of natural parameter and estimates of some of its functions are given in section 3.3. The UMP and UMPU tests of the population mean and the sampling distribution of the test statistic are discussed in section 3.4. Further, some UMP tests with power curves are also included in this section.

3.2 Harris Distribution, a Member of *GPSD*

Taking $a(x) = \binom{(x/k)-1}{(x-1)/k}$ for $x \in S = \{1, 1+k, 1+2k, \ldots\}$, $k>0$ integer and $a(x) = 0$ otherwise;

$\theta = \left((m-1)/m\right)^{1/k}$ and $C(\theta) = \theta(1-\theta^k)^{-1/k}$ in (3.1.1) we get

$$P(X = x) = \binom{(x/k)-1}{(x-1)/k} \frac{\theta^x}{\theta(1-\theta^k)^{-1/k}}$$

$$= \binom{(x/k)-1}{(x-1)/k} (1-\theta^k)^{1/k} (\theta^k)^{(x-1)/k}$$

$$= \binom{(x/k)-1}{(x-1)/k} \left(\frac{1}{m}\right)^{1/k} \left(1-\frac{1}{m}\right)^{(x-1)/k}, \quad x = 1,\ 1+k,\ 1+2k,$$

As this is the p.m.f of $H_1(m,k,1/k)$ given in (2.3.2), it belongs to the family of *GPSD*.

Note that

$$C(\theta) = \theta(1-\theta^k)^{-1/k}$$

$$= \theta + \binom{1/k}{1}\theta^{1+k} + \binom{(1/k)+1}{2}\theta^{1+2k} + \ldots + \binom{(1/k)+r-1}{r}\theta^{1+rk} + \ldots$$

$$= \sum_{r=0}^{\infty}\binom{(1/k)+r-1}{r}\theta^{1+rk} \quad = \sum_{x\in S}\binom{(x/k)-1}{(x-1)/k}\theta^{x} \quad = \sum_{x\in S} a(x)\theta^{x}$$

where $a(x)$ is the coefficient of θ^x in the expansion of $C(\theta)$.

Now, using the properties of *PSD*, we have the following results.

The p.g.f is

$$P(s) = C(\theta s)/C(\theta)$$

$$= \frac{\theta s(1-\theta^k s^k)^{-1/k}}{\theta(1-\theta^k)^{-1/k}}$$

$$= s\left(\frac{1-\theta^k}{1-\theta^k s^k}\right)^{1/k}$$

$$= \frac{s}{(m-(m-1)s^k)^{1/k}}$$

The moment generating function

$$P(e^t) = C(\theta e^t)/C(\theta) = \frac{e^t}{(m-(m-1)e^{tk})^{1/k}}$$

The mean and variance are

$$E(X) = \mu = \theta C'(\theta)/C(\theta) = \theta\frac{\frac{d}{d\theta}\theta(1-\theta^k)^{-1/k}}{\theta(1-\theta^k)^{-1/k}} = 1/(1-\theta)^k = m$$

$$V(X) = \theta\frac{d\mu}{d\theta} = \frac{k\theta^k}{(1-\theta^k)^2} = m(m-1)k$$

as has already been proved.

Factorial moment generating function is

$$P(1+t) = C(\theta(1+t))/C(\theta) = (1+t)/(m-(m-1)(1+t)^k)^{1/k},$$

Cumulant and factorial cumulant generating functions are

$$K(t) = \text{In}\, P(e^t) = \text{In}\,[C(\theta e^t)/C(\theta)\,] = t - (1/k)\text{In}(m\text{-}(m\text{-}1)e^{tk}) \text{ and}$$

$$\text{In}P(1+t) = \text{In}(1+t) - (1/k)\text{In}(m\text{-}(m\text{-}1)(1+t)^k)$$

3.3 Estimation of Natural Parameter and Some of its Functions

Unbiased estimator for $\theta^k = (m\text{-}1)/m$

Proceeding as given in Lehmann (1983), since $a(x)>0$ for all $x \in S$, θ^{rk} is U-estimable, for any positive integer r, its unique unbiased estimator $T_r(x)$ is obtained by solving the equations

$$E(T_r(x)) = \theta^{rk} \quad \text{for all } \theta \in \Omega.$$

$$\sum_{x\in S} T_r(x)a(x)\theta^x = \theta^{rk}C(\theta) = \sum_{x\in S} a(x)\theta^{x+rk}$$

since $C(\theta) = \sum_{x\in S} a(x)\theta^x$ · Comparison of the coefficients of θ^x yields

$$T_r(x) = \begin{cases} 0 & \text{if } \; x < 1+rk \\ a(x\text{-}\,rk)\,/\,a(x) & \text{if } \; x \geq 1+rk \end{cases}$$

Again, substituting for $a(x)$, we have,

$$T_r(x) = \binom{\frac{x-rk}{k}-1}{\frac{x-rk-1}{k}} \Bigg/ \binom{\frac{x}{k}-1}{\frac{x-1}{k}}$$

$$= \frac{\Gamma\left(\frac{x}{k}-r\right)\Gamma\left(\frac{x-1}{k}+1\right)}{\Gamma\left(\frac{x}{k}\right)\Gamma\left(\frac{x-1}{k}+1-r\right)}$$

$$= \frac{(x-1)(x-1-k)\ldots((x-1)-(r-1)k)}{(x-k)(x-2k)\ldots(x-rk)}, \qquad \text{for } x \geq 1+rk$$

In particular, $E(T_1(x)) = \theta^k = (m-1)/m = 1-(1/m)$ and

$$T_1(x) = \begin{cases} 0 & \text{if } x < 1+k \\ a(x-k)\,/\,a(x) & \text{if } x \geq 1+k \end{cases} \qquad \text{or}$$

$$T_1(x) = \begin{cases} 0 & \text{if } x = 1 \\ (x-1)\,/(x-k) & \text{if } x > 1 \end{cases}$$

$T_1(x)$ is an unbiased estimator of $1-(1/m)$ and $1- T_1(x)$ is an unbiased estimator of $1/m$.

Sufficient estimator for $\theta = ((m\text{-}1)/m)^{1/k}$

Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d according to $H_1(m,k,1/k)$, which belongs to the generalized power series family (3.1.1). Let

$$T = X_1 + X_2 + \ldots + X_n$$

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \frac{\theta^{\sum x_i} \prod a(x_i)}{(C(\theta))^n}$$

where $a(x_i) = \binom{(x_i/k)-1}{(x_i-1)/k}$ for $x_i \in S = \{1, 1+k, 1+2k, \ldots\}$, $k > 0$ integer and $a(x_i) = 0$

otherwise; $\theta = ((m-1)/m))^{1/k}$ and $C(\theta) = \theta(1-\theta^k)^{-1/k}$.

By factorization theorem, $T = \sum X_i$ is sufficient for $\theta = ((m\text{-}1)/m)^{1/k}$, when k is known.

According to the lemma 3.1 given in Lehmann (1983, p.96) the distribution of $T = \sum X_i$ is the power series family

$$P(T = t) = \frac{A(t,n)\theta^t}{(C(\theta))^n}, \qquad t = \Sigma x_i = n, n+k, n+2k, \ldots$$

where $A(t, n)$is the coefficient of θ^t in the expansion of $(C(\theta))^n$.

$A(t, n) = \sum_t a(x_1)a(x_2)\cdots a(x_n)$ where $\sum_t$ indicates that the summation extends over all n-tuples of integers $(x_1, x_2, \ldots, x_n)$ with $x_1 + x_2 + \ldots + x_n = t$.

Let $\psi(T)$ be some measurable function of T.

$$E[\psi(T)] = \sum_{t\in N} \psi(t)P(T = t) \text{ where } N = \{n, n+k, n+2k, \ldots\}$$

$$= \sum_{t\in N} \psi(t) \frac{A(t,n)\theta^t}{(C(\theta))^n}$$

$$E[\psi(T)] = 0 \Rightarrow \psi(t) = 0$$

T is complete.

Uniformly minimum variance unbiased estimator (*UMVUE*) of $(m$-1$)/m$

To find *UMVUE* of $(m-1)/m$ we make use of the following theorem [See Rohatji (1976, p.356)].

Theorem 3.1 (Lehmann-Scheffe Theorem)

If T is a complete sufficient statistic and there exits an unbiased estimate h of θ, there exits a unique *UMVUE* of θ, which is given by $E(h/T)$.

For any positive integer r, define $U_r(t)$ as follows.

$$U_r(t) = \begin{cases} 0 & \text{if } t < n + rk \\ A(t-rk, n)/A(t, n) & \text{if } t \geq n + rk \end{cases}$$

Then, $E(U_r(t)) = \theta^{rk}$

In particular, $E(U_1(t)) = \theta^k = (m-1)/m$ where

$$U_1(t) = \begin{cases} 0 & \text{if } t = n \\ A(t-k, n) / A(t, n) & \text{if } t > n \end{cases} \tag{3.3.1}$$

Since $U_1(t)$ is an unbiased estimator of θ^k and $T = \sum X_i$ is a completely sufficient statistic, by Lehmann-Scheffe theorem $U_1(t)$ is the unique *UMVUE* of $\theta^k = (m-1)/m$ on the basis of a sample of size n. Now,

$$\begin{aligned} V(U_1(t)) &= E(U_1(t))^2 - \{E(U_1(t))\}^2 \\ &= E(U_1(t))^2 - \theta^{2k} \\ &= E(U_1(t))^2 - E(U_2(t)) \\ &= E\{(U_1(t))^2 - U_2(t)\} \end{aligned}$$

$(U_1(t))^2 - U_2(t)$ is the *UMVUE* of $V(U_1(t))$.

It is given in Johnson *et al.* (1992, p.73) that a *MVUE* exists if and only if the *PSD* has support $\{a_1+a_2x\}$ where $x = 0,1,2,\ldots$, and a_1 and a_2 are nonnegative integers. Harris distribution belongs to the class of G*PSD* and has support $\{a_1+a_2x\}$ where $x = 0,1,2,\ldots$; $a_1 = 1$ and $a_2 = k > 0$, an integer. So *MVUE* $\hat{\theta}$ of $\theta = ((m-1)/m)^{1/k}$ exits and is

$$\hat{\theta} = \begin{cases} 0 & \text{if } t = n \\ A(t-1, n) / A(t, n) & \text{if } t > n \end{cases}$$

where $A(t, n)$ is the coefficient of θ^t in the expansion of $(C(\theta))^n$. This is (3.3.1) with $k = 1$ and $t = \Sigma x_i$.

Maximum likelihood estimator

Suppose $x_1, x_2, \ldots, x_n$ are n independent observations from $H_1(m,k,1/k)$. The likelihood function is then

$$L(\theta) = \prod_{i=1}^{n} \frac{a(x_i)\theta^{x_i}}{C(\theta)}$$

$$= \theta^t (C(\theta))^{-n} \prod_{i=1}^{n} a(x_i)$$

where $t = \sum_{i=1}^{n} x_i$, $\quad a(x_i) = \binom{(x_i/k)-1}{(x_i-1)/k}$ for $x_i \in S = \{1, 1+k, 1+2k, \ldots\}$, $k > 0$ an integer;

$\theta = ((m-1)/m)^{1/k}$ and $C(\theta) = \theta(1-\theta^k)^{-1/k}$

$$\log L = t \log\theta - n \log C(\theta) + \sum_{i=1}^{n} \log a(x_i)$$

The maximum likelihood estimator $\hat{\theta}$ of θ satisfies the equation

$$(\hat{\theta})^{-1} t - \{nC'(\hat{\theta})/C(\hat{\theta})\} = 0$$

$$\hat{\theta}C'(\hat{\theta})/C(\hat{\theta}) = t/n$$

$$\hat{\mu} = \bar{x}$$

where the mean μ of $H_1(m,k,1/k)$ is m and $\bar{x}$ is the sample mean.

So, as has already been proved in chapter 2, maximum likelihood estimate of m is

$$\hat{m} = \bar{x}$$

3.4 Testing of Population Mean

For a fixed k, Harris distribution belongs to the one parameter exponential family of distributions.

The p.m.f of $H_1(m,k,1/k)$ can be written as

$$P(X=x) = \binom{(x/k)-1}{(x-1)/k} \left(\frac{1}{m}\right)^{1/k} \left(1-\frac{1}{m}\right)^{(x-1)/k}, \qquad x = 1, 1+k, 1+2k, \ldots$$

$$= \binom{(x/k)-1}{(x-1)/k}\left(\frac{1}{m-1}\right)^{1/k}\left(\frac{m-1}{m}\right)^{x/k}$$

$$= \binom{(x/k)-1}{(x-1)/k} e^{-\frac{1}{k}\log(m-1)} e^{x\frac{1}{k}\log(\frac{m-1}{m})}$$

$$= Q(m)e^{\eta(m)T(x)}a(x)$$

where $Q(m) = e^{-\frac{1}{k}\log(m-1)}$, $\eta(m) = \frac{1}{k}\log(\frac{m-1}{m}), T(x) = x$ and $a(x) = \binom{(x/k)-1}{(x-1)/k}$; Q and η are real-valued functions of parameters and T is a real- valued statistic. This is the density form of the distributions belonging to the one-parameter exponential family.

If X_i, $i = 1,2,\ldots,n$ are independent Harris variables with $E(X_i) = m$, then their joint distribution is

$$P_m(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

$$= \prod_{i=1}^{n} \binom{(x_i/k)-1}{(x_i-1)/k}\left(\frac{1}{m}\right)^{\frac{1}{k}}\left(1-\frac{1}{m}\right)^{\frac{x_i-1}{k}}$$

$$= \left(\frac{1}{m}\right)^{\frac{n}{k}}\left(1-\frac{1}{m}\right)^{\sum_{i=1}^{n}\left(\frac{x_i-1}{k}\right)} \prod_{i=1}^{n} \binom{(x_i/k)-1}{(x_i-1)/k}$$

$$= h(x)\exp\left(\frac{n}{k}\log\left(\frac{1}{m}\right)\right)\exp\left(\left(\frac{\sum x_i}{k} - \frac{n}{k}\right)\log\left(\frac{m-1}{m}\right)\right)$$

$$= h(x)\exp\left(-\frac{n}{k}\log(m-1)\right)\exp\left(\frac{\sum x_i}{k}\log\left(\frac{m-1}{m}\right)\right)$$

$$= h(x)Q(m)\exp(\eta(m)T(x))$$

where $Q(m) = \exp\left(-\frac{n}{k}\log(m-1)\right)$, $\eta(m) = \frac{1}{k}\log(\frac{m-1}{m})$, $T(x) = \sum_{i=1}^{n} x_i$,

$h(x)=\prod_{i=1}^{n}\binom{(x_i/k)-1}{(x_i-1)/k}$; Q and η are real-valued functions of the parameters and T is a real-valued statistic.

This is the density form of the distributions belonging to the one-parameter exponential family. Further, $\eta(m)$ is strictly increasing.

Sampling distribution of T(x)

The sampling distribution of $T(x)=\sum_{i=1}^{n} x_i \sim H_n(m,k,n/k)$. The probability distributions of T for different sample size n with fixed m=2 and k=3 are given in Table 3.1 and the corresponding probability plots are given in Figures 3.1 – 3.5. The asymptotic normality of T is evident from the Figures.

Next we consider the following tests of population mean m when k is known.

One sided tests

Consider testing the hypotheses

$H_0 : m \leq m_0$ against $H_1 : m > m_0$.

Using corollary 2 of Lehmann (1986, p.80), for testing H_0 against H_1 there exists a UMP test ϕ, which is given by,

$$\phi(x) = \begin{cases} 1 & \text{when } \sum x_i > c \\ \gamma & \text{when } \sum x_i = c \\ 0 & \text{when } \sum x_i < c \end{cases}$$

where γ and c are determined by $E_{m_0}\phi(X)=\alpha$.

$$P_{m_0}\left(T(x)>c\right)+\gamma P_{m_0}\left(T(x)=c\right)=\alpha$$

$$P_{m_0}\left(T(x) \le c\right) - \gamma P_{m_0}\left(T(x) = c\right) = 1 - \alpha \tag{3.4.1}$$

where $T(x) = \sum_{i=1}^{n} x_i \sim H_n(m, k, n/k)$ with p.m.f

$$P(T(x) = t) = \frac{\Gamma(t/k)}{\Gamma(n/k)\Gamma(1 + ((t-n)/k))}\left(\frac{1}{m}\right)^{n/k}\left(1 - \frac{1}{m}\right)^{(t-n)/k}, \quad t = n,\ n+k,\ n+2k, \ldots \tag{3.4.2}$$

It follows from (3.4.1) and (3.4.2) that

$$\sum_{t=n}^{c} \frac{\Gamma(t/k)}{\Gamma(n/k)\Gamma(1 + ((t-n)/k))}\left(\frac{1}{m_0}\right)^{n/k}\left(1 - \frac{1}{m_0}\right)^{(t-n)/k}$$

$$-\gamma \frac{\Gamma(c/k)}{\Gamma(n/k)\Gamma(1 + ((c-n)/k))}\left(\frac{1}{m_0}\right)^{n/k}\left(1 - \frac{1}{m_0}\right)^{(c-n)/k} = 1 - \alpha \tag{3.4.3}$$

The power function is given by

$$\begin{aligned}
\beta(m) &= E_m \phi(X) \\
&= P_m\left(T(x) \le c\right) - \gamma P_m\left(T(x) = c\right) \\
&= \sum_{t=n}^{c} \frac{\Gamma(t/k)}{\Gamma(n/k)\Gamma(1 + ((t-n)/k))}\left(\frac{1}{m}\right)^{n/k}\left(1 - \frac{1}{m}\right)^{(t-n)/k} \\
&\qquad -\gamma \frac{\Gamma(c/k)}{\Gamma(n/k)\Gamma(1 + ((c-n)/k))}\left(\frac{1}{m}\right)^{n/k}\left(1 - \frac{1}{m}\right)^{(c-n)/k}
\end{aligned} \tag{3.4.4}$$

Similarly, a UMP test ϕ for testing $H_0 : m \ge m_0$ against $H_1 : m < m_0$ has the form

$$\phi(x) = \begin{cases} 1 & \text{when } \sum x_i < c \\ \gamma & \text{when } \sum x_i = c \\ 0 & \text{when } \sum x_i > c \end{cases}$$

where γ and c are determined such that

$$E_{m_0} \phi(X) = \alpha . \tag{3.4.5}$$

It follows from (3.4.2) that condition (3.4.5) becomes

$$\sum_{t=n}^{c}\frac{\Gamma(t/k)}{\Gamma(n/k)\Gamma(1+((t-n)/k))}\left(\frac{1}{m_0}\right)^{n/k}\left(1-\frac{1}{m_0}\right)^{(t-n)/k}$$

$$+\gamma\frac{\Gamma(c/k)}{\Gamma(n/k)\Gamma(1+((c-n)/k))}\left(\frac{1}{m_0}\right)^{n/k}\left(1-\frac{1}{m_0}\right)^{(c-n)/k}=\alpha \qquad (3.4.6)$$

The power function is given by

$$\beta(m)=E_m\phi(X)$$

$$=\sum_{t=n}^{c}\frac{\Gamma(t/k)}{\Gamma(n/k)\Gamma(1+((t-n)/k))}\left(\frac{1}{m}\right)^{n/k}\left(1-\frac{1}{m}\right)^{(t-n)/k}$$

$$+\gamma\frac{\Gamma(c/k)}{\Gamma(n/k)\Gamma(1+((c-n)/k))}\left(\frac{1}{m}\right)^{n/k}\left(1-\frac{1}{m}\right)^{(c-n)/k} \qquad (3.4.7)$$

Two sided tests

Now consider two- sided hypotheses of the form

H_0 : $m \le m_1$ or $m \ge m_2$ $(m_1 < m_2)$ against $H_1 : m_1 < m < m_2$

Using theorem 6 of Lehmann (1986, p.101), for testing H_0 against H_1 there exists a UMP test ϕ given by

$$\phi(x) = \begin{cases} 1 & \text{when } c_1 < \sum x_i < c_2 \quad (c_1 < c_2) \\ \gamma_j & \text{when } \sum x_i = c_j, \quad j = 1,2 \\ 0 & \text{when } \sum x_i < c_1 \text{ or } > c_2 \end{cases}$$

where the γ's and c's are determined by

$$E_{m_1}\phi(X)=E_{m_2}\phi(X)=\alpha\,. \qquad (3.4.8)$$

It follows from (3.4.2) and (3.4.8) that

$$\sum_{t=c_1+k}^{c_2-k} \frac{\Gamma(t/k)}{\Gamma(n/k)\Gamma(1+((t-n)/k))}\left(\frac{1}{m_1}\right)^{n/k}\left(1-\frac{1}{m_1}\right)^{(t-n)/k}$$

$$+\sum_{j=1}^{2}\gamma_j \frac{\Gamma(c_j/k)}{\Gamma(n/k)\Gamma(1+((c_j-n)/k))}\left(\frac{1}{m_1}\right)^{n/k}\left(1-\frac{1}{m_1}\right)^{(c_j-n)/k} = \alpha \tag{3.4.9}$$

and

$$\sum_{t=c_1+k}^{c_2-k} \frac{\Gamma(t/k)}{\Gamma(n/k)\Gamma(1+((t-n)/k))}\left(\frac{1}{m_2}\right)^{n/k}\left(1-\frac{1}{m_2}\right)^{(t-n)/k}$$

$$+\sum_{j=1}^{2}\gamma_j \frac{\Gamma(c_j/k)}{\Gamma(n/k)\Gamma(1+((c_j-n)/k))}\left(\frac{1}{m_2}\right)^{n/k}\left(1-\frac{1}{m_2}\right)^{(c_j-n)/k} = \alpha \tag{3.4.10}$$

The power function is given by

$$\beta(m) = E_m\phi(X)$$

$$= \sum_{t=c_1+k}^{c_2-k} \frac{\Gamma(t/k)}{\Gamma(n/k)\Gamma(1+((t-n)/k))}\left(\frac{1}{m}\right)^{n/k}\left(1-\frac{1}{m}\right)^{(t-n)/k}$$

$$+\sum_{j=1}^{2}\gamma_j \frac{\Gamma(c_j/k)}{\Gamma(n/k)\Gamma(1+((c_j-n)/k))}\left(\frac{1}{m}\right)^{n/k}\left(1-\frac{1}{m}\right)^{(c_j-n)/k} \tag{3.4.11}$$

Now consider testing the hypotheses

$H_0 : m_1 \le m \le m_2$ against $H_1 : m < m_1$ or $m > m_2$.

Using results in section 2 of Lehmann (1986, p.135), UMP test doesn't exit for testing H_0 against H_1. In this case there exit a UMP unbiased test ϕ given by

$$\phi(x) = \begin{cases} 1 & \text{when } \sum x_i < c_1 \text{ or } > c_2 \\ \gamma_j & \text{when } \sum x_i = c_j, \quad j = 1,2 \\ 0 & \text{when } c_1 < \sum x_i < c_2 \end{cases} \tag{3.4.12}$$

where γ's and c's are determined by

$$E_{m_1}\phi(X) = E_{m_2}\phi(X) = \alpha\,. \tag{3.4.13}$$

It follows from (3.4.2) and condition (3.4.13) that

$$\sum_{t=c_1}^{c_2} \frac{\Gamma(t/k)}{\Gamma(n/k)\Gamma(1+((t-n)/k))}\left(\frac{1}{m_1}\right)^{n/k}\left(1-\frac{1}{m_1}\right)^{(t-n)/k}$$

$$-\sum_{j=1}^{2}\gamma_j \frac{\Gamma(c_j/k)}{\Gamma(n/k)\Gamma(1+((c_j-n)/k))}\left(\frac{1}{m_1}\right)^{n/k}\left(1-\frac{1}{m_1}\right)^{(c_j-n)/k} = 1-\alpha \tag{3.4.14}$$

and

$$\sum_{t=c_1}^{c_2} \frac{\Gamma(t/k)}{\Gamma(n/k)\Gamma(1+((t-n)/k))}\left(\frac{1}{m_2}\right)^{n/k}\left(1-\frac{1}{m_2}\right)^{(t-n)/k}$$

$$-\sum_{j=1}^{2}\gamma_j \frac{\Gamma(c_j/k)}{\Gamma(n/k)\Gamma(1+((c_j-n)/k))}\left(\frac{1}{m_2}\right)^{n/k}\left(1-\frac{1}{m_2}\right)^{(c_j-n)/k} = 1-\alpha \tag{3.4.15}$$

The power function is given by

$$\beta(m) = E_m\phi(X)$$

$$= \sum_{t=c_1}^{c_2} \frac{\Gamma(t/k)}{\Gamma(n/k)\Gamma(1+((t-n)/k))}\left(\frac{1}{m}\right)^{n/k}\left(1-\frac{1}{m}\right)^{(t-n)/k}$$

$$-\sum_{j=1}^{2}\gamma_j \frac{\Gamma(c_j/k)}{\Gamma(n/k)\Gamma(1+((c_j-n)/k))}\left(\frac{1}{m}\right)^{n/k}\left(1-\frac{1}{m}\right)^{(c_j-n)/k} \tag{3.4.16}$$

Now consider testing the hypothesis

$H_0 : m = m_0$ against the alternatives $H_1 : m \neq m_0$

For testing H_0 against H_1 there exists a UMP unbiased test ϕ given by (3.4.12) but the constants γ's and c's are determined by

$E_{m_0}\phi(X) = \alpha$ and

$$E_{m_0}\{T(X)\phi(X)\} = E_{m_0}\{T(X)\}\alpha = nm_0\alpha. \tag{3.4.17}$$

It follows from (3.4.2) that conditions (3.4.17) becomes

$$\sum_{t=c_1}^{c_2} \frac{\Gamma(t/k)}{\Gamma(n/k)\Gamma(1+((t-n)/k))}\left(\frac{1}{m_0}\right)^{n/k}\left(1-\frac{1}{m_0}\right)^{(t-n)/k}$$

$$-\sum_{j=1}^{2} \gamma_j \frac{\Gamma(c_j/k)}{\Gamma(n/k)\Gamma(1+((c_j-n)/k))}\left(\frac{1}{m_0}\right)^{n/k}\left(1-\frac{1}{m_0}\right)^{(c_j-n)/k} = 1-\alpha \qquad (3.4.18)$$

The power function is given by

$$\beta(m) = E_m \phi(X)$$

$$= \sum_{t=c_1}^{c_2} \frac{\Gamma(t/k)}{\Gamma(n/k)\Gamma(1+((t-n)/k))}\left(\frac{1}{m}\right)^{n/k}\left(1-\frac{1}{m}\right)^{(t-n)/k}$$

$$-\sum_{j=1}^{2} \gamma_j \frac{\Gamma(c_j/k)}{\Gamma(n/k)\Gamma(1+((c_j-n)/k))}\left(\frac{1}{m}\right)^{n/k}\left(1-\frac{1}{m}\right)^{(c_j-n)/k} \qquad (3.4.19)$$

Numerical illustrations:

(1) Consider the problem of testing $H_0 : m \leq m_0$ against $H_1 : m > m_0$.

(A) Let $m_0 = 2$, $k = 3$, $n = 10$ and $\alpha = .05$

The test function is given by

$$\phi(x) = \begin{cases} 1 & \text{when } \sum x_i > 34 \\ 0.219 & \text{when } \sum x_i = 34 \\ 0 & \text{when } \sum x_i < 34 \end{cases}$$

It follows from (3.4.4) that the power function,

$$\mathrm{A}(m) = \sum_{t=10}^{34} \frac{\Gamma(t/3)}{\Gamma(10/3)\Gamma(1+((t-10)/3))}\left(\frac{1}{m}\right)^{10/3}\left(1-\frac{1}{m}\right)^{(t-10)/3}$$

$$-0.219 \frac{\Gamma(34/3)}{\Gamma(10/3)\Gamma(1+((34-10)/3))}\left(\frac{1}{m}\right)^{10/3}\left(1-\frac{1}{m}\right)^{(34-10)/3}$$

(B) Let $m_0 = 3$, $k = 3$, $n = 10$ and $\alpha = .05$

The test function is given by

$$\phi(x) = \begin{cases} 1 & \text{when} \quad \sum x_i > 55 \\ 0.326 & \text{when} \quad \sum x_i = 55 \\ 0 & \text{when} \quad \sum x_i < 55 \end{cases}$$

The power function,

$$\mathrm{B}(m) = \sum_{t=10}^{55} \frac{\Gamma(t/3)}{\Gamma(10/3)\Gamma(1+((t-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(t-10)/3}$$

$$-0.326 \frac{\Gamma(55/3)}{\Gamma(10/3)\Gamma(1+((55-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(55-10)/3}$$

(C) Let $m_0 = 4$, $k = 3$, $n = 10$ and $\alpha = .05$

The test function is given by

$$\phi(x) = \begin{cases} 1 & \text{when} \quad \sum x_i > 76 \\ 0.495 & \text{when} \quad \sum x_i = 76 \\ 0 & \text{when} \quad \sum x_i < 76 \end{cases}$$

The power function,

$$\mathrm{C}(m) = \sum_{t=10}^{76} \frac{\Gamma(t/3)}{\Gamma(10/3)\Gamma(1+((t-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(t-10)/3}$$

$$-0.495 \frac{\Gamma(76/3)}{\Gamma(10/3)\Gamma(1+((76-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(76-10)/3}$$

The values of A(m), B(m) and C(m) for various values of m are given in Table 3.2 and the corresponding power curves are given in Figure 3.6.

(2) Consider the problem of testing $H_0 : m \geq m_0$ against $H_1 : m < m_0$

(A) Let $m_0 = 8$, $k = 3$, $n = 10$ and $\alpha = .05$

The test function is given by

$$\phi(x) = \begin{cases} 1 & \text{when} \quad \sum x_i < 28 \\ 0.32 & \text{when} \quad \sum x_i = 28 \\ 0 & \text{when} \quad \sum x_i > 28 \end{cases}$$

It follows from (3.4.7) that the power function,

$$\mathrm{A}(m) = \sum_{t=10}^{28} \frac{\Gamma(t/3)}{\Gamma(10/3)\Gamma(1+((t-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(t-10)/3}$$

$$+0.32 \frac{\Gamma(28/3)}{\Gamma(10/3)\Gamma(1+((28-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(28-10)/3}$$

(B) Let $m_0 = 10$, $k=3$, $n=10$ and $\alpha = .05$

The test function is given by

$$\phi(x) = \begin{cases} 1 & \text{when} \quad \sum x_i < 34 \\ 0.298 & \text{when} \quad \sum x_i = 34 \\ 0 & \text{when} \quad \sum x_i > 34 \end{cases}$$

The power function,

$$\mathrm{B}(m) = \sum_{t=10}^{34} \frac{\Gamma(t/3)}{\Gamma(10/3)\Gamma(1+((t-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(t-10)/3}$$

$$+0.298 \frac{\Gamma(34/3)}{\Gamma(10/3)\Gamma(1+((34-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(34-10)/3}$$

(C) Let $m_0 = 12$, $k=3$, $n=10$ and $\alpha = .05$

The test function is given by

$$\phi(x) = \begin{cases} 1 & \text{when} \quad \sum x_i < 40 \\ 0.279 & \text{when} \quad \sum x_i = 40 \\ 0 & \text{when} \quad \sum x_i > 40 \end{cases}$$

The power function,

$$C(m) = \sum_{t=10}^{40} \frac{\Gamma(t/3)}{\Gamma(10/3)\Gamma(1+((t-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(t-10)/3}$$

$$+0.279 \frac{\Gamma(40/3)}{\Gamma(10/3)\Gamma(1+((40-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(40-10)/3}$$

The values of A(m), B(m) and C(m) for various values of m are given in Table 3.3 and the corresponding power curves are given in Figure 3.7.

(3) Consider testing of two- sided hypotheses of the form

$H_0 : m \le m_1$ or $m \ge m_2$ $(m_1 < m_2)$ against $H_1 : m_1 < m < m_2$

(A) Let $m_1 = 2$, $m_2 = 12$, $k = 3$, $n = 10$ and $\alpha = .05$

The test function is given by

$$\phi(x) = \begin{cases} 1 & \text{when } 34 < \sum x_i < 46 \\ 0.449 & \text{when } \sum x_i = 34 \\ 0.759 & \text{when } \sum x_i = 46 \\ 0 & \text{when } \sum x_i < 34 \text{ or } > 46 \end{cases}$$

It follows from (3.4.11) that the power function,

$$A(m) = \sum_{t=37}^{43} \frac{\Gamma(t/3)}{\Gamma(10/3)\Gamma(1+((t-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(t-10)/3}$$

$$+0.449 \frac{\Gamma(34/3)}{\Gamma(10/3)\Gamma(1+((34-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(34-10)/3}$$

$$+0.759 \frac{\Gamma(46/3)}{\Gamma(10/3)\Gamma(1+((46-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(46-10)/3}.$$

(B) Let $m_1 = 2$, $m_2 = 10$, $k = 3$, $n = 10$ and $\alpha = .05$

The test function is given by

$$\phi(x) = \begin{cases} 1 & \text{when} \quad 34 < \sum x_i < 43 \\ 0.724 & \text{when} \quad \sum x_i = 34 \\ 0.287 & \text{when} \quad \sum x_i = 43 \\ 0 & \text{when} \quad \sum x_i < 34 \text{ or } > 43 \end{cases}$$

The power function,

$$\mathrm{B}(m) = \sum_{t=37}^{40} \frac{\Gamma(t/3)}{\Gamma(10/3)\Gamma(1+((t-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(t-10)/3}$$

$$+0.724 \frac{\Gamma(34/3)}{\Gamma(10/3)\Gamma(1+((34-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(34-10)/3}$$

$$+0.287 \frac{\Gamma(43/3)}{\Gamma(10/3)\Gamma(1+((43-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(43-10)/3}.$$

(C) Let $m_1 = 2,\ m_2 = 8,\ k = 3,\ n = 10$ and $\alpha = .05$

The test function is given by

$$\phi(x) = \begin{cases} 1 & \text{when} \quad 31 < \sum x_i < 37 \\ 0.179 & \text{when} \quad \sum x_i = 31 \\ 0.845 & \text{when} \quad \sum x_i = 37 \\ 0 & \text{when} \quad \sum x_i < 31 \text{ or } > 37 \end{cases}$$

The power function,

$$\mathrm{C}(m) = \frac{\Gamma(34/3)}{\Gamma(10/3)\Gamma(1+((34-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(34-10)/3}$$

$$+0.179 \frac{\Gamma(31/3)}{\Gamma(10/3)\Gamma(1+((31-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(31-10)/3}$$

$$+0.845 \frac{\Gamma(37/3)}{\Gamma(10/3)\Gamma(1+((37-10)/3))} \left(\frac{1}{m}\right)^{10/3} \left(1-\frac{1}{m}\right)^{(37-10)/3}.$$

The values of $\mathrm{A}(m)$,$\mathrm{B}(m)$ and $\mathrm{C}(m)$ for various values of m are given in Table 3.4 and the corresponding power curves are given in Figure 3.8.

TABLE 3.1

Sampling distribution of T

t	P(t)	t	P(t)	t	P(t)	t	P(t)	t	P(t)
1	0.794000	5	0.315000	10	0.099000	25	0.003100	45	3.05E-05
4	0.132000	8	0.262000	13	0.165000	28	0.013000	48	0.000229
7	0.044000	11	0.175000	16	0.179000	31	0.030000	51	0.000916
10	0.017000	14	0.107000	19	0.159000	34	0.052000	54	0.002594
13	0.007145	17	0.062000	22	0.126000	37	0.074000	57	0.005836
16	0.003096	20	0.035000	25	0.092000	40	0.091000	60	0.011000
19	0.001376	23	0.020000	28	0.064000	43	0.101000	63	0.018000
22	0.000623	26	0.011000	31	0.043000	46	0.103000	66	0.028000
25	0.000285	29	0.005825	34	0.028000	49	0.099000	69	0.038000
		32	0.003128	37	0.017000	52	0.090000	72	0.049000
		35	0.001668	40	0.011000	55	0.078000	75	0.058000
		38	0.000885	43	0.006504	58	0.065000	78	0.066000
		41	0.000467	46	0.003884	61	0.052000	81	0.072000
		44	0.000246	49	0.002291	64	0.041000	84	0.075000
				52	0.001336	67	0.031000	87	0.075000
				55	0.000772	70	0.023000	90	0.072000
				58	0.000442	73	0.017000	93	0.068000
				61	0.000252	76	0.012000	96	0.062000
						79	0.008503	99	0.055000
						82	0.005893	102	0.048000
						85	0.004027	105	0.041000
						88	0.002716	108	0.034000
						91	0.001811	111	0.028000
						94	0.001194	114	0.022000
						97	0.000780	117	0.018000
						100	0.000504	120	0.014000
						103	0.000323	123	0.011000
						106	0.000206	126	0.008013
								129	0.006010
								132	0.004455
								135	0.003267
								138	0.002371
								141	0.001704
								144	0.001214
								147	0.000857
								150	0.000600
								153	0.000417
								156	0.000287
								159	0.000196

TABLE 3.2
Values of power functions given in illustration (1)

m	A(m)	B(m)	C(m)
1.1	.000	.000	.000
1.5	.003	.000	.000
2.0	.050	.001	.000
2.5	.163	.014	.001
3.0	.301	.050	.006
4.0	.539	.185	.050
5.0	.697	.345	.140
8.0	.898	.688	.466
10	.943	.803	.625
12	.966	.871	.734
15	.982	.926	.835
18	.989	.954	.892
20	.992	.966	.917
25	.996	.982	.954
30	.998	.989	.972
35	.999	.993	.982
40	.999	.995	.988
45	.999	.997	.991
50	1.00	.998	.994
55	1.00	.998	.995
60	1.00	.999	.996
65	1.00	.999	.997
70	1.00	.999	.998
80	1.00	.999	.998
90	1.00	1.00	.999
100	1.00	1.00	.999
120	1.00	1.00	1.00

TABLE 3.3
Values of power functions given in illustration (2)

m	A(m)	B(m)	C(m)
1.1	1.00	1.00	1.00
1.3	.997	1.00	1.00
1.4	.991	.999	1.00
1.6	.961	.991	.998
1.8	.909	.971	.991
2.0	.842	.937	.976
3.0	.497	.664	.788
4.0	.284	.427	.560
5.0	.170	.276	.387
6.0	.108	.184	.271
7.0	.072	.127	.194
8.0	.050	.091	.142
9.0	.036	.067	.106
10	.027	.050	.081
12	.016	.030	.050
14	.010	.019	.033
16	.007	.013	.022
18	.005	.009	.016
20	.003	.007	.012
25	.002	.003	.006
30	.001	.002	.003
35	.001	.001	.002
40	.000	.001	.001
45	.000	.001	.001
50	.000	.000	.001
55	.000	.000	.001
60	.000	.000	.000

TABLE 3.4
Values of power functions given in illustration (3)

m	A(m)	B(m)	C(m)
1.1	.000	.000	.000
1.4	.001	.001	.002
1.5	.003	.004	.005
1.8	.024	.025	.027
2.0	.050	.050	.050
2.4	.118	.108	.096
3.0	.205	.172	.137
3.5	.243	.194	.145
4.0	.253	.195	.139
6.0	.190	.134	.087
8.0	.120	.081	.050
10	.076	.050	.030
12	.050	.032	.019
12.5	.045	.029	.017
13	.041	.026	.016
14	.034	.022	.013
15	.029	.018	.011
17	.021	.013	.007
20	.013	.008	.005
25	.007	.004	.002
30	.004	.003	.001
35	.003	.002	.001
40	.002	.001	.001
45	.001	.001	.000
50	.001	.001	.000
55	.001	.000	.000
60	.000	.000	.000

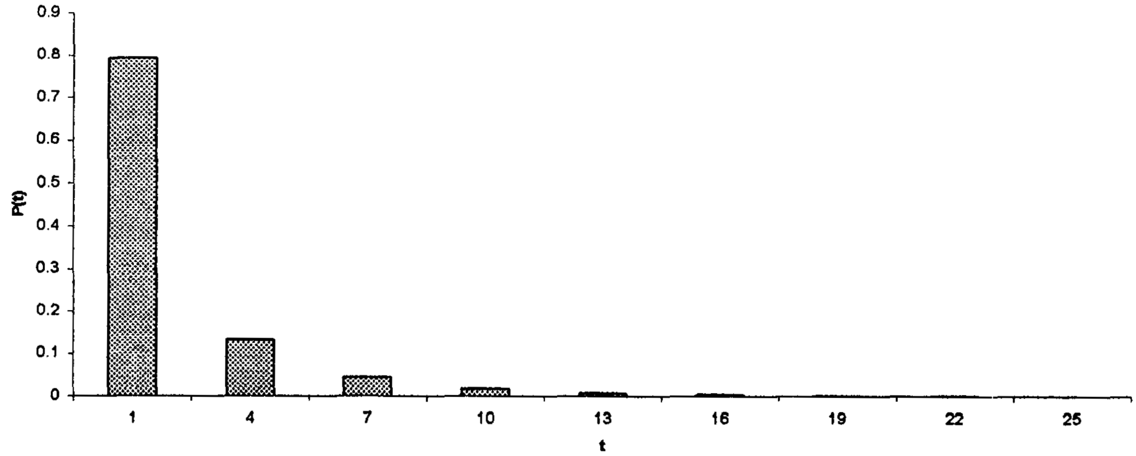


FIGURE 3.1 – Probability plot of T for $m = 2$, $k = 3$ and $n = 1$

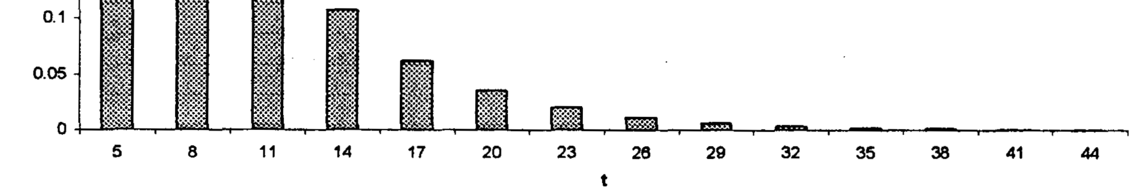


FIGURE 3.2 – Probability plot of T for $m = 2$, $k = 3$ and $n = 5$

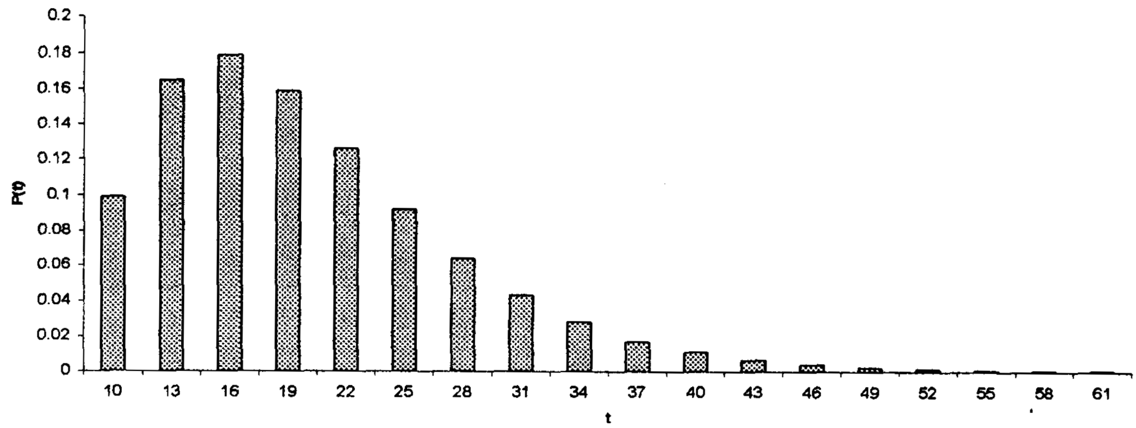


FIGURE 3.3 – Probability plot of T for $m = 2$, $k = 3$ and $n = 10$

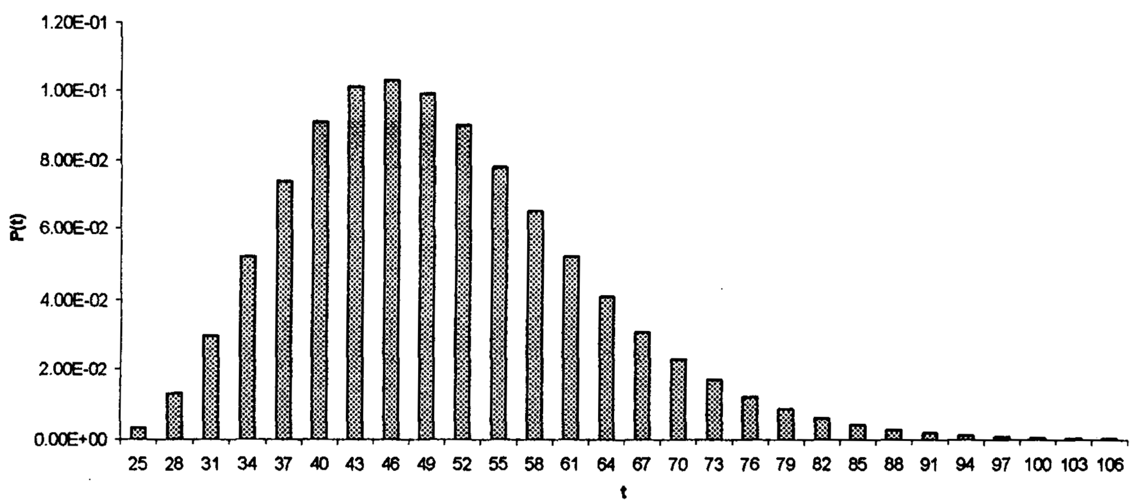


FIGURE 3.4 – Probability plot of T for $m = 2$, $k = 3$ and $n = 25$

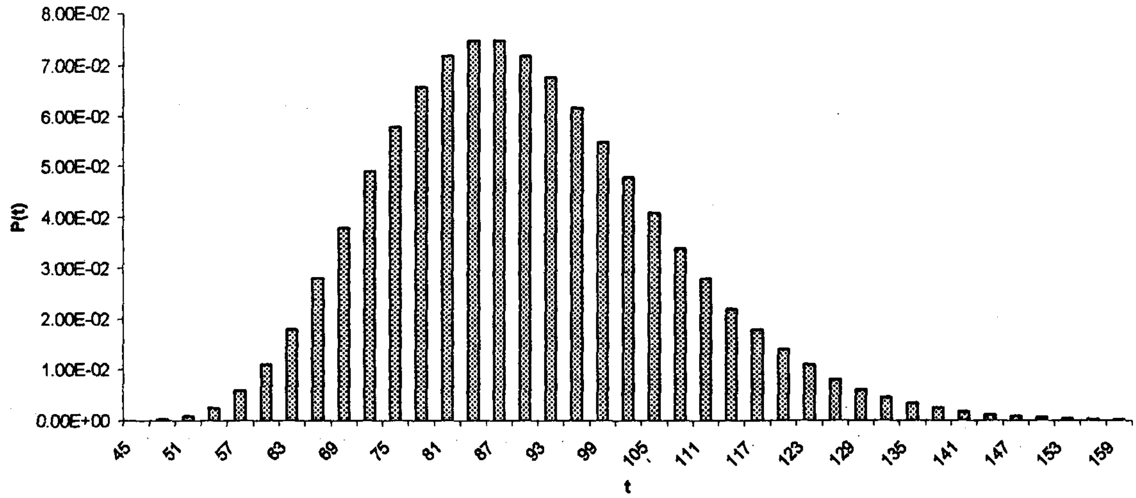


FIGURE 3.5 – Probability plot of T for $m = 2$, $k = 3$ and $n = 45$

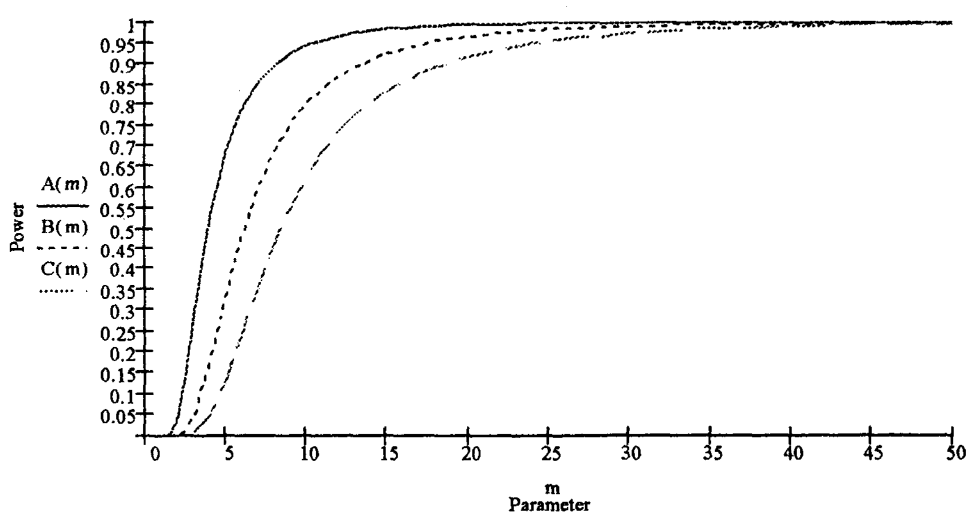


FIGURE 3.6 – Power curves of illustration (1).

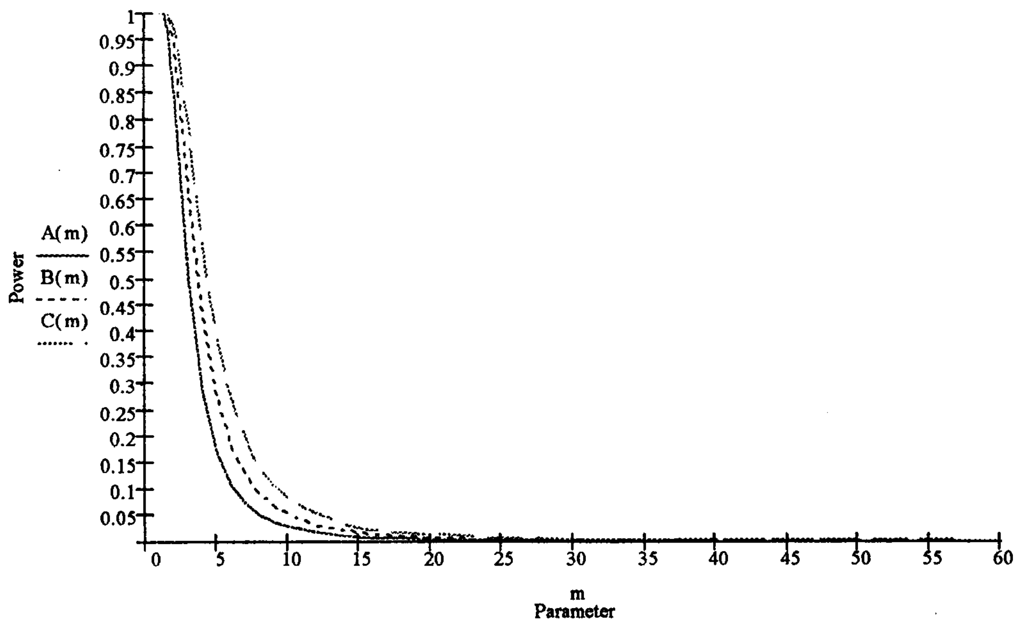


FIGURE 3.7 – Power curves of illustration (2).

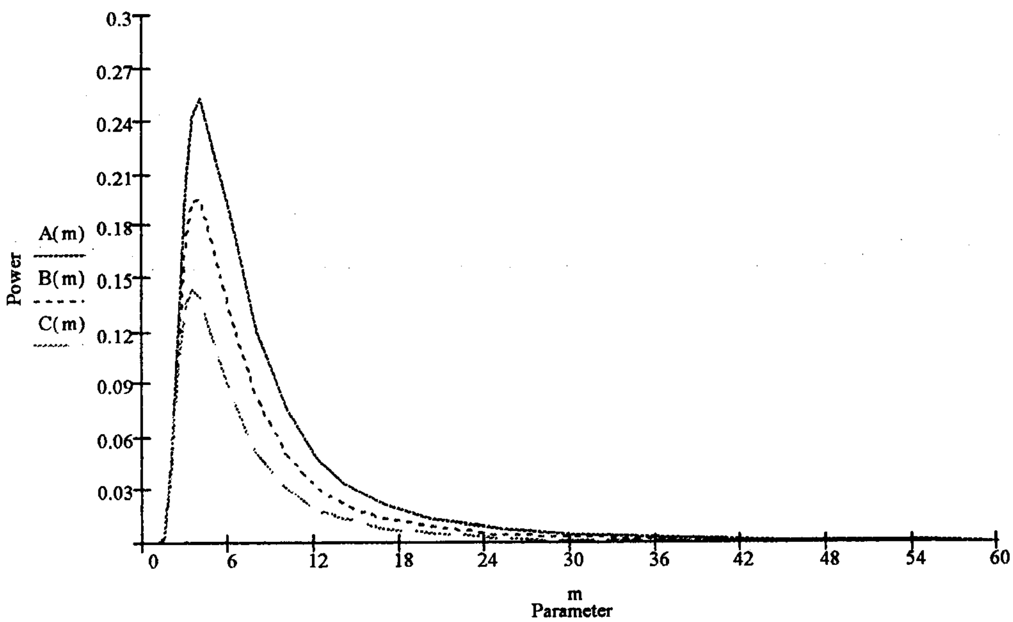


FIGURE 3.8 – Power curves of illustration (3).

CHARACTERIZATIONS OF EXTENDED GEOMETRIC, HARRIS, NEGATIVE BINOMIAL AND GAMMA DISTRIBUTIONS

Sherly Sebastian “Harris family of discrete distributions and processes ” Thesis. Department of Statistics , University of Calicut, 2007

CHAPTER 4

CHARACTERIZATIONS OF EXTENDED GEOMETRIC, HARRIS, NEGATIVE BINOMIAL AND GAMMA DISTRIBUTIONS

4.1 Introduction

We have already seen that Harris distribution is related to many discrete distributions like geometric, negative binomial etc. and continuous distribution gamma. Here we derive some characterizations of Harris and the related distributions mentioned above. We start with defining an extended geometric distribution. A geometric distribution may be extended to cover the case of a variable taking values a, $a + k$, $a + 2k$, ... ($k>0$). See, Johnson *et al.* (1992, p.201). Here we consider a geometric distribution with parameter p defined on $\{a,\ a + k,\ a + 2k,\ \dots\}$ where $a\geq 0$ and $k>0$ are integers. We call the corresponding distribution as extended geometric distribution and denote it by $Geo_a(p, k)$. In the notation the suffix a suggests that the support of the distribution starts from a and k implies that the atoms of the distribution are k integers apart.

Let the r.v Z have a geometric distribution on $\{0, 1, 2, \dots\}$ with parameter p ($Z \sim Geo(p)$). The p.m.f of Z is

$$P(Z = n) = q^n p,\ \ n = 0,1,2,\dots,\ \ 0 < p < 1,\ \ p + q = 1.$$

Then the r.v $Y = a + kZ$ is $Geo_a(p, k)$ with p.m.f

$$P(Y = a + nk) = q^n p,\ \ n = 0,1,2,\dots \tag{4.1.1}$$

The contents of this chapter has been accepted for publication in ***STARS. Int. Journal.***

where $a \geq 0$ and $k > 0$ are integers, $0 < p < 1$ and $p + q = 1$. *i.e.*, $Y \sim Geo_a(p, k)$.

The moments of $Geo_a(p, k)$ are easily available from the following simple but an important characterization property of the distribution.

A r.v $Z \sim Geo(p)$ if and only if the r.v $Y \sim Geo_a(p, k)$ where $Y = a + k\,Z$. Then,

$$E(Y) = a + k\,E(Z)$$

$$V(Y) = k^2\,V(Z)$$

Also the p.g.f of $Geo_a(p, k)$ is of the form

$$P(s) = \frac{ps^a}{1 - qs^k}$$

The standard results used in this chapter are given in section 4.2. Satheesh and Sandhya (1997) have given a characterization for mixtures of geometric distributions using the property of having completely monotone probability sequence (CMPS). Also the geometric distribution is characterized in terms of convolution equations. Similar characterizations developed in the case of extended geometric distributions are presented in section 4.3. Based on this convolution equation it is shown that a counting process is binomial with gaps (of interval k) in its support if and only if the inter-arrival times are extended geometric. Also it follows from the above result that a mixed binomial process with gaps is renewal if and only if it is binomial with gaps. Further it is shown that CMPSs on $\{a, a + k, a + 2k, \ldots\}$ are log-convex and hence they are i.d.

Relations among the Harris, geometric, negative binomial and gamma distributions are given in the second chapter. It is proved that a r.v $X \sim H_1(m, k, 1/k)$ if and only if the r.v $W = (X\text{-}1)/k$ follows a negative binomial distribution $NB(1/m, 1/k)$ with parameters $1/m$ and $1/k$. A negative binomial distribution with index $1/k$ is given in

Cadigan and Chen (2001). Characterizations of Harris and negative binomial distributions based on the p.g.fs are given in section 4.4. Also some relations between extended geometric distribution and Harris distribution are discussed in this section.

The connection between a negative binomial r.v and a geometric r.v as given in theorem 4.14 may be better explained by considering the concept of fractional successes. There are situations where a success can occur only when k (>1, an integer) fractional successes happen. For example, suppose a sales representative receives an incentive only when he sells k identical items. Here selling of each item is a fractional success and obtaining an incentive is a success and success occurs only when the k fractional successes happen *i.e.* when k items are sold. Such an example is considered in chapter 5. Also, in product control, an item is passed a quality test only when k identical quality tests are passed. This is another real life situation in which a success is constituted by fractional successes.

In chapter 1 connection between Harris and gamma r.vs in the context of branching process has been pointed out as established by Harris (1948). In chapter 2 we have shown the Harris distribution as a gamma mixture by considering a linear function of Poisson r.v with gamma distributed Poisson parameter. Again, Satheesh *et al.* (2002) showed that $Gamma(a,1/k)$ law is $H_1(a,k,1/k)$ sum stable. A characterization of gamma distribution is also given in this section and some random sum results yielding characterizations are presented in section 4.5. Another look at Harris sum stability of gamma variable is given in theorem 4.16.

4.2 Standard Results

Completely monotone sequence

Definition 4.1 A sequence $\{\mu_n\}$ such that $(-1)^r \Delta^r \mu_n \geq 0$ for all combinations r, n is called completely monotone. Here Δ is the differencing operator.

Theorem 4.1 (Hausdorff's Theorem), Feller (1966, p. 223)

A sequence of numbers μ_0, μ_1, ... represents the moments, $\mu_n = E(X^n)$, of some probability distribution F concentrated on $[0,1]$ if and only if $(-1)^r \Delta^r \mu_n \geq 0$, $\mu_0 = 1$.

Convolutions of sequences

Definition 4.2 Let $\{a_n\}$ and $\{b_n\}$ be any two numerical sequences. The new sequence $\{c_n\}$ defined by $c_n = a_0 b_n + a_1 b_{n-1} + a_2 b_{n-2} + \ldots + a_{n-1} b_1 + a_n b_0$ is called the convolution of $\{a_n\}$ and $\{b_n\}$ and will be denoted by

$$\{c_n\} = \{a_n\} * \{b_n\}$$

Generating function of survival sequence

Consider a r.v X assuming the values 0,1,2,.... Let $P(X=n) = p_n$ and $P(X>n) = q_n$. Then

$$q_n = p_{n+1} + p_{n+2} + \ldots, \qquad n = 0,1,2,\ldots,$$

The generating function of the sequence $\{q_n\}$ is given by

$$Q(s) = q_0 + q_1 s + q_2 s^2 + \ldots$$

Theorem 4.2

$$Q(s) = \frac{1 - P(s)}{1 - s}, \quad -1 < s < 1$$

where $P(s) = \sum_{n=0}^{\infty} p_n s^n$ is the p.g.f of the r.v X.

Theorem 4.3

If $\{a_n\}$ and $\{b_n\}$ are sequences with generating functions $A(s)$ and $B(s)$, and $\{c_n\}$ is their convolution, then the generating function $C(s) = \sum c_n s^n$ is the product

$$C(s) = A(s)B(s)$$

4.3 Characterizations of Mixtures of Extended Geometric Distribution

Pillai and Sandhya (1990) have shown that a distribution function on $(0,\infty)$ has complete monotone derivative or the density is completely monotone (CM) if and only if it is a mixture of exponential distributions. Naturally its discrete analogue will be geometric distribution and Satheesh and Sandhya (1997) characterize a discrete distribution with CMPS as a mixture of geometric distribution. Just like a mixture of geometric r.v on $\{0,1,2,\dots\}$ characterizes a CMPS $\{f(n)\}$, $n = 0,1,2,\dots$ we have the following result on extended geometric r.v characterizing a CMPS on $\{a+nk\}$, $n = 0,1,2,\dots$, $a\geq 0$, $k > 0$ integers.

Theorem 4.4

A probability sequence (PS) $\{f(a+nk)\}$, $n = 0,1,2,\dots$; $a\geq 0$ and $k>0$ are integers is completely monotone if and only if it is a mixture of $Geo_a(p, k)$.

Proof.

Consider an extended geometric r.v with PS given by (4.1.1). Then we have

$$P(Y < a+ nk) = 1 - P(Y \geq a+ nk) = 1 - q^n$$

Randomizing q with a distribution G concentrated on $(0,1)$, we have the d.f F of the mixture as

$$F(a+nk) = P(X < a+nk)$$

$$= \int_0^1 (1-q^n)dG(q)$$

$$= 1 - E(q^n), \quad n = 0,1,2,\ldots$$

$$F(a+nk) = 1 - m(n), \quad n = 0,1,2,\ldots$$

where $\{m(n)\}$, the moment sequence of G is CM by using Hausdorff's theorem (theorem 4.1) and definition 4.1. If $\{f(a+nk)\}$ denote the PS corresponding to this mixture, then,

$$\Delta F(a+nk) = F(a+(n+1)k) - F(a+nk) = f(a+nk), \quad n = 0,1,2,\ldots$$

Again,

$$\Delta F(a+nk) = \Delta(1-m(n)) = -\Delta m(n), \quad n = 0,1,2,\ldots$$

so that

$$\Delta F(a+nk) = f(a+nk) = -\Delta m(n), \quad n = 0,1,2,\ldots$$

and hence $\{f(a+nk)\}$ is CM, Δ being the differencing operator.

Conversely, let us start with a CMPS $\{f(a+nk)\}$. The corresponding survival sequence is given by $S(a+nk) = P(X \geq a+nk)$.

Now in view of definition 4.1 we have,

$$(-1)^r \Delta^r f(a+nk) \geq 0$$

$$(-1)^r \Delta^r (\Delta F(a+nk)) \geq 0$$

$$(-1)^r \Delta^{r+1} (1 - S(a+nk)) \geq 0$$

$$(-1)^{r+1} \Delta^{r+1} S(a+nk) \geq 0$$

Also, when $n = 0$, $S(a+nk) = S(a) = P(X \geq a) = 1$. Hence by Hausdorff's theorem $\{S(a+nk)\}$ represents the moments of some probability distribution G concentrated on $(0,1)$. Now,

$$F(a+nk) = 1 - S(a+nk)$$

$$= 1 - \int_0^1 q^n dG(q)$$

$$= \int_0^1 (1-q^n) dG(q), \quad n = 0,1,2,\ldots$$

and hence $\{f(a+nk)\}$ corresponds to a mixture of $Geo_a(p, k)$. This completes the proof.

Now consider a non-negative integer-valued r.v U. Then the r.v $V = kU$, $k>0$ an integer, is with PS $\{g(nk)\}$, where $g(nk) = P(V=nk)$, $n = 0,1,2,\ldots$. The survival sequence is taken as

$$S(nk) = P(V>nk) = g((n+1)k) + g((n+2)k) + \ldots$$

Sandhya and Satheesh (1997) characterize an exponential distribution based on a convolution equation connecting its distribution function, survival function and density function. Again Satheesh and Sandhya (1997) characterize the discrete analogue of exponential distribution, viz, geometric distribution on similar lines and use this result to show that a mixed binomial process is renewal if and only if it is binomial. Here we characterize extended geometric distribution in a similar fashion based on its p.m.f and survival function.

Theorem 4.5

A PS $\{g(nk)\}$, $k > 0$ an integer satisfies the convolution equation

$$\{g(nk)\} * \{S(nk)\} = \{ng(nk)\}, \quad n = 0,1,2,\ldots \text{if and only if}$$

$$g(nk) = q^{n-1}p, \quad n = 1,2,\ldots$$

$$= 0 \quad \text{otherwise}, \quad 0 < p < 1, \ p + q = 1.$$

Proof.

The survival sequence corresponding to the PS $g(nk) = q^{n-1}p$, $n = 1, 2, \ldots$ is

$$S(nk) = q^n p + q^{n+1}p + \ldots = q^n, \qquad n = 0, 1, 2, \ldots$$

The convolution of the sequences $\{g(nk)\}$ and $\{S(nk)\}$ is the sequence $\{C(nk)\}$ where

$$C(nk) = g(0)S(nk) + g(k)S((n-1)k) + \ldots + g((n-1)k)S(k) + g(nk)S(0)$$

$$= 0.q^n + p.q^{n-1} + qp.q^{n-2} + \ldots + q^{n-1}p.1$$

$$= nq^{n-1}p$$

$$= ng(nk)$$

Hence we have the convolution equation

$$\{g(nk)\} * \{S(nk)\} = \{ng(nk)\}, \qquad n = 0,1,2, \ldots$$

Now suppose that the convolution equation is true and let $P(s)$, $Q(s)$ and $C(s)$ be the generating functions of the sequences $\{g(nk)\}, \{S(nk)\}$ and $\{C(nk)\}$ respectively. Then

$$P(s) = \sum_{n=0}^{\infty} g(nk)s^{nk}$$

From theorem 4.2 we have

$$Q(s) = \sum_{n=0}^{\infty} S(nk)s^{nk} = \frac{1-P(s)}{1-s^k}$$

which can be proved as follows.

$$(1- s^k)\, Q(s) = (1- s^k)\, \{\, S(0) + S(k)\, s^k + S(2k)\, s^{2k} + \ldots + S(nk)\, s^{nk} + \ldots \,\} \tag{4.3.1}$$

$$1- P(s) = 1 - g(k)\, s^k - g(2k)\, s^{2k} - \ldots - g(nk)\, s^{nk} - \ldots \tag{4.3.2}$$

For $n \geq 1$, coefficient of s^{nk} in (4.3.1) is $S(nk) - S((n\text{-}1)k) = - g(nk)$, which is the coefficient of s^{nk} in (4.3.2).

For $n = 0$, coefficient of s^0 in (4.3.1) is $S(0) = 1$, which is coefficient of s^0 in (4.3.2).

This proves the expression for $Q(s)$.

Now,

$$C(s) = \sum_{n=0}^{\infty} ng(nk)s^{nk} = \frac{s}{k}\sum_{n=1}^{\infty} nkg(nk)s^{nk-1} = \frac{s}{k}P'(s)$$

By using theorem 4.3, we have

$$P(s)Q(s) = C(s)$$

$$P(s)\,\frac{1-P(s)}{1-s^k} = \frac{s}{k}P'(s)$$

$$kP(s)(1-P(s)) = s\left(1-s^k\right)\frac{dP(s)}{ds}$$

Now, integration gives the solution of this differential equation for a p.g.f. Hence we have,

$$\frac{1}{k}\int \frac{dP(s)}{P(s)(1-P(s))} = \int \frac{ds}{s(1-s^k)}$$

$$\frac{1}{k}\log\left(\frac{P(s)}{1-P(s)}\right) = \frac{1}{k}\log\left(\frac{s^k}{1-s^k}\right) + \frac{1}{k}\log c,$$ where $(1/k)\log c$ is a constant.

$$\frac{P(s)}{1-P(s)} = \frac{cs^k}{1-s^k}$$

$$P(s) = \frac{cs^k}{1-s^k+cs^k} = \frac{cs^k}{(1-c+c)-(1-c)s^k}$$

$$P(s) = \frac{as^k}{a+b-bs^k} = \frac{ps^k}{1-qs^k}, \quad p = \frac{a}{a+b}, \quad q = 1-p$$

This is the p.g.f of a $Geo_k(p, k)$.

From an example in Feller (1968, p.313), Satheesh and Sandhya (1997) stressed that a counting process is binomial if and only if the waiting times are geometric. Here we consider a process having $Geo_k(p,k)$ distribution as inter-arrival times. Consider any situation where one trial is a bunch of k Bernoulli trials and success happens only when all the k Bernoulli trials in the bunch turns out as successes (i.e. a success occurs only when k fractional successes happen in a particular trial). Let N_n stand for the number of successes in nk Bernoulli trials with success probability p=1− q>0. Let Z_{nk} stand for the number of fractional successes from all the successes occurred. Then N_n takes values 0,1,2,…., n and Z_{nk} takes values 0,k,2k, …., nk. Also,

$$P\{Z_{nk} = rk\} = P\{N_n=r\} = \binom{n}{r} q^{n-r}(1-q)^r,\ r = 0,1,2, \ldots n.$$

The process $\{Z_{nk}, n = 1,2,3, \ldots\}$ is termed as a binomial process with gaps of interval k. Let $f(nk)$ = P{success occurring for the first time in n^{th} trial} = $q^{n-1}p$, n = 1,2, …. , where $f(0)$=0 and the p.g.f of $f(nk)$ is $P(s) = \sum_{n=0}^{\infty} f(nk)s^{nk} = \dfrac{ps^k}{1-qs^k}$. That is, the waiting times follow $Geo_k(p,k)$ distribution. A similar situation where the inter-arrival times follow $Geo_1(q,k)$ distribution is given in Feller (1968, p.315) associated with a counter model. Now we have proved;

Theorem 4.6

A counting process is binomial with gaps of interval k if and only if the inter-arrival times are $Geo_k(p,k)$ distributed.

This leads to the following theorem.

Theorem 4.7

A mixed binomial process with gaps is renewal if and only if it is binomial with gaps.

Proof.

Randomizing the parameter q in the process $\{Z_{nk}\}$ with a distribution G concentrated on (0,1) we have a mixed binomial process $\{Z_{nk}\}$ with gaps and

$$P(Z_{nk} = rk) = \binom{n}{r}\int_0^1 q^{n-r}(1-q)^r dG(q)$$

$$= \binom{n}{r}(-1)^r \Delta^r m(n-r)$$

where $m(.)$ is the moment sequence of G. Now if $\{Z_{nk}\}$ is renewal with waiting time distribution function F, then

$$P(Z_{nk} = rk) = \binom{n}{r}(-1)^r \Delta^r m(n-r) = F^{r*}(nk) - F^{(r+1)*}(nk).$$

For $r = 0$, $P(Z_{nk} = 0) = m(n) = 1 - F(nk)$

For $r = 1$, $P(Z_{nk} = k) = -n\Delta m(n-1) = F(nk) - F^{2*}(nk)$.

Hence, $\quad n\Delta F((n-1)k) = F(nk) * [1 - F(nk)]$

or, $\quad \{f(nk)\} * \{S(nk)\} = \{nf(nk)\}$

Now, invoking theorems 4.5 and 4.6 we have the result.

Now we characterize a $Geo_0(p,k)$ distribution as in theorem 4.5.

Theorem 4.8

A PS $\{g(nk)\}$, $k > 0$ an integer, satisfies the convolution equation

$\{g(nk)\} * \{S(nk)\} = \{(n+1)g((n+1)k)\}$, $n = 0,1,2,\ldots$ if and only if

$g(nk) = q^n p$, $n = 0,1,2,\ldots$

$= 0$ otherwise, $0 < p < 1$, $p + q = 1$.

Proof.

Corresponding to the PS $g(nk) = q^n p$, $n = 0,1,2,\ldots$ we have the survival sequence $S(nk) = q^{n+1}$, $n = 0,1,2,\ldots$. Then it can be shown that

$$\{g(nk)\} * \{S(nk)\} = \{(n+1)g((n+1)k)\}, \quad n = 0,1,2,\ldots$$

Now, assume that the convolution equation is true. Then proceeding as in theorem 4.5 we have

$$P(s)\ \frac{1-P(s)}{1-s^k} = \frac{P'(s)}{k\,s^{k-1}}$$

where $P(s)$ is the p.g.f of $\{g(nk)\}$. The solution of this differential equation for a p.g.f is

$$P(s) = \frac{c}{1+c-s^k} = \frac{p}{1-qs^k}, \quad p = \frac{c}{1+c}, \ q = 1-p.$$

This is the p.g.f of a $Geo_0(p, k)$.

Satheesh and Sandhya (1997) have shown that distributions having CMPS are log-convex. Here we establish the log-convexity of CMPS on $\{a, a+k, a+2k, \dots\}$.

Theorem 4.9

Distributions with CMPSs on $\{a, a+k, a+2k, \dots\}$ are log-convex.

Proof.

If $\{f(a+nk)\}$ is a CMPS, then by Hausdorff's theorem, $f(a+nk) = \int_0^1 x^n dM(x)$ for a finite measure M and by Schwarz's inequality for arbitrary measures we have

$$\int_0^1 x^{n+1} dM(x) \int_0^1 x^{n-1} dM(x) \geq \left(\int_0^1 x^n dM(x) \right)^2 \quad \text{and hence}$$

$$f(a+(n+1)k) f(a+(n-1)k) \geq f^2(a+nk) \tag{4.3.3}$$

This completes the theorem.

Theorem 4.10

Mixtures of extended geometric distributions are i.d.

Proof.

Putting $Y = (X\text{-}a)/k$ in (4.3.3), we get, $f(n+1)f(n-1) \geq f^2(n) \Rightarrow Y$ is log-convex and hence it is i.d. Since infinite divisibility is not affected by change of origin and scale, X is also i.d and hence the theorem.

4.4 Characterizations of Harris, Negative Binomial and Gamma Distributions

In this section we characterize the Harris distribution $H_1(m,k,1/k)$. Its p.g.f $P(s)$ given in (1.1.1) on differentiation with respect to s gives,

$$P'(s) = m/\left(m-(m-1)s^k\right)^{1+(1/k)}, \quad \text{where} \quad m = P'(1) \tag{4.4.1}$$

Now we have the following theorems.

Theorem 4.11

A r.v $X \sim H_1(m,k,1/k)$ if and only if its p.g.f $P(s)$ satisfies the equation

$$s^{1+k}P'(s) = m(P(s))^{1+k}$$

Proof.

Let $X \sim H_1(m,k,1/k)$. Then we have (1.1.1) and (4.4.1). Now

$$sP'(s) = mP(s)/\left(m-(m-1)s^k\right) \tag{4.4.2}$$

Again, from (1.1.1) we have

$$m - (m\text{-}1)\, s^k = (s / P(s))^k \tag{4.4.3}$$

From (4.4.2) and (4.4.3) we get

$$s^{1+k}P'(s) = m(P(s))^{1+k} \tag{4.4.4}$$

Conversely, starting with (4.4.4) we have,

$$\int dP(s)/(P(s))^{1+k} = m\int ds/s^{1+k} \quad \text{and hence} \quad 1/(\mathrm{P(s)})^k = (m/s^k) - c$$

When $s = 1$, $P(s) = 1$ and hence $c = m$-1. So $P(s) = s/\left(m-(m-1)s^k\right)^{1/k}$, which is the p.g.f of $H_1(m,k,1/k)$.

Theorem 4.12

$H_1(m,k,1/k)$ is characterized by the equation $s\,(1\text{-}s^k)\,P'(s) = P(s)\{1\text{-}\,(P(s))^k\}$.

Proof.

Let $X \sim H_1(m,k,1/k)$. Then we have the equations given in theorem 4.11. Now from (4.4.3) we get

$$m = (s/P(s))^k\left\{\left(1-(P(s))^k\right)/\left(1-s^k\right)\right\}.$$

Now substituting this m in (4.4.4) we get

$$s\,(1\text{-}s^k)\,P'(s) = P(s)\{1\text{-}\,(P(s))^k\} \tag{4.4.5}$$

Conversely, starting with a p.g.f satisfying (4.4.5) we have

$$\int\frac{dP(s)}{P(s)\{1-(P(s))^k\}} = \int\frac{ds}{s(1-s^k)}$$

Now, this on integration and simplification gives

$$\frac{(P(s))^k}{1-(P(s))^k} = \frac{cs^k}{1-s^k}$$

Then the solution $P(s) = s/\left(m-(m-1)s^k\right)^{1/k}$, m=1/c, is the p.g.f of $H_1(m,k,1/k)$.

Corollary 4.1

Value of k in $H_1(m,k,1/k)$ can be evaluated by

$$k = \frac{sP(s)P''(s) - s(P'(s))^2 + P(s)P'(s)}{P'(s)\left(sP'(s) - P(s)\right)}$$

where $P'(s)$ and $P''(s)$ are the first and second order derivatives of $P(s)$ with respect to s.

Proceeding as given in theorem 4.12 we have,

$s\,(1\text{-}s^k).P'(s) = P(s)\{1\text{-}\,(P(s))^k\}.$

Now from (4.4.4) we get, $P'(s) = m\,(P(s)/s)^{1+k}$ which on differentiation with respect to s gives

$$P''(s) = m(1+k)(P(s))^k\{sP'(s) - P(s)\}/s^{k+2} \tag{4.4.6}$$

Now dividing (4.4.6) by (4.4.4) we get

$$\frac{P''(s)}{P'(s)} = \frac{(1+k)\{sP'(s) - P(s)\}}{sP(s)}, \quad \text{which gives}$$

$$k = \frac{sP(s)P''(s) - s(P'(s))^2 + P(s)P'(s)}{P'(s)\left(sP'(s) - P(s)\right)}.$$

Next, we have a theorem which shows the relationship between a Harris distribution and an extended geometric distribution.

Theorem 4.13

$X_1, X_2, \ldots, X_k$ are i.i.d $H_1(m,k,1/k)$ r.vs if and only if $Y = \sum_{i=1}^{k} X_i$ is $Geo_k(1/m,k)$.

Proof.

Let $X_1, X_2, \ldots, X_k$ be k i.i.d $H_1(m,k,1/k)$ r.vs with p.g.f

$$P_{X_i}(s) = \frac{s}{\left(m-(m-1)s^k\right)^{1/k}}, \quad i = 1, 2, \ldots, k$$

The p.g.f of $Y = \sum_{i=1}^{k} X_i$ is then $P_Y(s) = \left(P_{X_i}(s)\right)^k = \dfrac{s^k}{m-(m-1)s^k}$, $m>1$.

This is the p.g.f of $Geo_k(1/m, k)$ distribution. Hence $Y = \sum_{i=1}^{k} X_i \sim Geo_k(1/m,k)$.

Conversely, let $Y = \sum_{i=1}^{k} X_i \sim Geo_k(1/m,k)$ where X_i's are i.i.d r.vs.

Now, the p.g.f of Y is given by

$$P_Y(s) = \frac{s^k}{m-(m-1)s^k} = \left(\frac{s}{\left(m-(m-1)s^k\right)^{1/k}}\right)^k = \left(P_{X_i}(s)\right)^k.$$

where $P_{X_i}(s)$ is the p.g.f of $H_1(m,k,1/k)$ and hence X_1, X_2, …, X_k are k i.i.d $H_1(m,k,1/k)$ r.vs.

Following theorem shows the relationship between $NB(p,1/k)$ and $Geo(p)$.

Theorem 4.14

$X_1, X_2, \ldots, X_k$ are k i.i.d $NB(p,1/k)$ r.vs if and only if $Y = \sum_{i=1}^{k} X_i$ is a $Geo(p)$.

Proof.

Let $X \sim NB(p,1/k)$ with p.m.f

$$f(x) = \binom{(1/k)+x-1}{x} p^{1/k} q^x, \qquad x = 0,1,2,\ldots$$

Then the p.g.f, $P_X(s) = E(s^X) = \sum_{x=0}^{\infty} s^x \binom{(1/k)+x-1}{x} p^{1/k} q^x = \left(\frac{p}{1-qs}\right)^{1/k}$

Let $X_1, X_2, \ldots, X_k$ be k i.i.d $NB(p,1/k)$ variables with p.g.f

$$P_{X_i}(s) = \left(\frac{p}{1-qs}\right)^{1/k} \qquad \text{for} \quad i = 1, 2, \ldots k$$

and let $Y = \sum_{i=1}^{k} X_i$. Now, the p.g.f of Y is given by

$$P_Y(s) = \left(P_{X_i}(s)\right)^k = \frac{p}{1-qs}$$

This is the p.g.f of $Geo(p)$.

Conversely, let $Y = \sum_{i=1}^{k} X_i$, the sum of k i.i.d r.vs be $Geo(p)$. Now, the p.g.f of Y

is given by

$$P_Y(s) = \frac{p}{1-qs}$$

Since $X_1, X_2, \ldots, X_k$ are i.i.d r.vs we have $P_Y(s) = (P_{X_i}(s))^k$ and hence

$$P_{X_i}(s) = \left(\frac{p}{1-qs}\right)^{1/k} \quad \text{for} \quad i = 1, 2, \ldots k$$

where $P_{X_i}(s)$ is the p.g.f of $NB(p,1/k)$ r.v and hence $X_1, X_2, \ldots, X_k$ are k i.i.d $NB(p,1/k)$ variables.

Next we present a characterization of gamma distribution, which is the continuous analogue of negative binomial distribution.

Theorem 4.15

The LT $\varphi(\lambda)$ of a probability distribution satisfies the differential equation $\varphi'(\lambda) = \varphi(\lambda)\log\varphi(\lambda)/(1+\lambda)\log(1+\lambda)$ if and only if the distribution is $Gamma(\alpha,1)$ with density function $f(x) = e^{-x}x^{\alpha-1}/\Gamma\alpha$, $x>0$, $\alpha>0$.

Proof.

Let X be a $Gamma(\alpha,1)$ r.v with LT $\varphi(\lambda) = (1+\lambda)^{-\alpha}$. Then

$\alpha = -\log\varphi(\lambda)/\log(1+\lambda)$.

Differentiation of $\varphi(\lambda)$ with respect to λ gives

$$\varphi'(\lambda) = -\alpha\varphi(\lambda)/(1+\lambda)$$

Substituting the value of α we get

$$\varphi'(\lambda) = \varphi(\lambda)\log\varphi(\lambda)/(1+\lambda)\log(1+\lambda) \tag{4.4.7}$$

Now, suppose that equation (4.4.7) is true. Then we have,

$$\int d\varphi(\lambda)/\varphi(\lambda)\log\varphi(\lambda) = \int d\lambda/(1+\lambda)\log(1+\lambda)$$

Now, $u = \log\varphi(\lambda)$ and $t = \log(1+\lambda)$ gives $u = ct$, *i.e.*

$$\varphi(\lambda) = (1+\lambda)^c \tag{4.4.8}$$

which gives $c = \log\varphi(\lambda) / \log(1+\lambda)$ and hence from (4.4.7)

$$\varphi'(\lambda) = c\varphi(\lambda)/(1+\lambda) \tag{4.4.9}$$

When $\lambda = 0$, $\varphi(0) = 1$ and $\varphi'(0) = -E(X) = -\alpha$ (say) and hence from (4.4.9) we have $c = -\alpha$. Hence from (4.4.8), $\varphi(\lambda) = (1+\lambda)^{-\alpha}$ which is the LT of a gamma distribution with parameter α.

4.5 Random Sum Characterizations

Random sums, that is, sums of a random number of i.i.d r.vs with the random number independent of the summands, often produce useful results. It is well-known that a geometric random sum of exponential r.vs is exponential and a Poisson random sum of logarithmic r.vs is negative binomial. In this section the results about the random sums are presented in the form of characterizations of various distributions.

Suppose X and $X_1, X_2, \ldots$ are i.i.d non-negative r.vs independent of N, a non-negative integer-valued r.v. A random sum S is defined by

$$S = X_1 + X_2 + \ldots + X_N \qquad (N \geq 1). \tag{4.5.1}$$

Let $P_N(s)$ denote the p.g.f of N and $\varphi_X(\lambda)$ the LT of the distribution of the r.v X. Then the LT for the random sum S is given by

$$\varphi_S(\lambda) = P_N(\varphi_X(\lambda)) \tag{4.5.2}$$

When X is discrete, the p.g.f of the random sum is

$$P_S(s) = P_N(P_X(s)) \tag{4.5.3}$$

Theorem 4.16

Suppose that S, given by (4.5.1), is a random sum. Then any two of the following three conditions imply the third.

(a) The individual summands are $Gamma(\alpha, 1/k)$ distributed with p.d.f

$$f_X(x) = \frac{\alpha^{1/k}}{\Gamma(1/k)} e^{-\alpha x} x^{(1/k)-1}, \qquad 0 < x < \infty$$

where $\alpha>0$ and $k>0$ is an integer.

(b) The number of summands has a $H_1(m,k,1/k)$ distribution with p.m.f

$$f(x) = \binom{(1/k)+n-1}{n}\left(\frac{1}{m}\right)^{1/k}\left(1-\frac{1}{m}\right)^n, \quad x = 1+nk, \quad n = 0,1,2,\ldots$$

where $m>1$ and $k>0$ an integer.

(c) The random sum S has a gamma distribution $Gamma(\beta, 1/k)$ where $\beta=(\alpha/m)>0$ and $k>0$ is an integer.

Proof.

The proof is straight forward. We use (4.5.2), together with the LTs and p.g.f for the above distributions

$$\varphi_X(\lambda) = \left(1+\frac{\lambda}{\alpha}\right)^{-1/k} \tag{4.5.4}$$

$$P_N(s) = \frac{s}{(m-(m-1)s^k)^{1/k}} \tag{4.5.5}$$

$$\varphi_S(\lambda) = \left(1+\frac{\lambda}{\beta}\right)^{-1/k}, \ \beta = \alpha/m \tag{4.5.6}$$

and their uniqueness.

(a) and (b) $\Rightarrow$ (c)

Inserting (4.5.4) and (4.5.5) in (4.5.2) gives

$$\varphi_S(\lambda) = \frac{\left(1+\dfrac{\lambda}{\alpha}\right)^{-1/k}}{\left(m-(m-1)\left(1+\dfrac{\lambda}{\alpha}\right)^{-1}\right)^{1/k}}$$

which gives (4.5.6).

(b) and (c) $\Rightarrow$ (a)

By inserting (4.5.5) and (4.5.6) in (4.5.2) we get

$$\left(1+\frac{m\lambda}{\alpha}\right)^{-1/k} = \frac{\varphi_X(\lambda)}{\left(m-(m-1)(\varphi_X(\lambda))^k\right)^{1/k}}$$

which gives (4.5.4).

(c) and (a) $\Rightarrow$ (b)

Inserting (4.5.4) and (4.5.6) in (4.5.2) yields

$$\left(1+\frac{m\lambda}{\alpha}\right)^{-1/k} = P_N\left(\left(1+\frac{\lambda}{\alpha}\right)^{-1/k}\right) \tag{4.5.7}$$

Taking $s=\left(1+\dfrac{\lambda}{\alpha}\right)^{-1/k}$ in (4.5.7) and solving gives (4.5.5).

Remark 4.1

When k=1 in theorem 4.16, the number of summands has a geometric distribution. Then the random sum is exponentially distributed if and only if the individual summands are exponentially distributed.

Theorem 4.16 can be considered as the compact expression of three characterization results, in the sense that any one of them is expressible as an 'if and only if' statement; the above remark is an example.

Now we present a random sum characterization motivated by Milne and Yeo (1989). The p.g.f of a modified logarithmic distribution given therein is

$$P(s) = 1 - \alpha + \frac{\alpha \ln(1-\theta s)}{\ln(1-\theta)}, \quad 0<\alpha \leq 1 \text{ and } 0<\theta<1.$$

The log-convexity and hence the infinite divisibility of the distribution are given by Hansen and Willekens (1990). But here we consider a related p.g.f

$$P_X(s) = \left(1 - \frac{\alpha}{k} + \frac{\alpha \ln(1-\theta s^k)}{k \ln(1-\theta)}\right)^{1/k}$$

The fact that the function $P_X(s)$ is a p.g.f follows from infinite divisibility of $P(s)$, since if $P(s)$ is i.d $\left(P(s^k)\right)^{1/k}$ is also a p.g.f, $k>0$ integer. The argument s^k in any p.g.f implies that the distribution has gaps of interval k. Hence the function $P_N(s) = e^{\lambda(s^k-1)}$ corresponds to the p.g.f of a Poisson distribution with gaps of interval k.

Theorem 4.17

Suppose that S is given by (4.5.1). Then, any two of the following three conditions imply the third.

(a) The individual summands have a distribution with p.g.f

$$P_X(s) = \left(1 - \frac{\alpha}{k} + \frac{\alpha \ln(1-\theta s^k)}{k \ln(1-\theta)}\right)^{1/k} \tag{4.5.8}$$

where $k>0$ an integer, $0<\alpha\leq 1$ and $0<\theta<1$.

(b) The number of summands follows a generalized Poisson law with p.g.f

$$P_N(s) = e^{\lambda(s^k-1)} \tag{4.5.9}$$

where $k>0$ an integer and $\lambda>0$.

(c) The random sum S has $H_0(1/(1-\theta),k,1/k)$ distribution with p.g.f

$$P_S(s) = \left(\frac{1-\theta}{1-\theta s^k} \right)^{1/k} \tag{4.5.10}$$

where $k>0$ an integer and $0<\theta<1$.

Proof.

The result follows from the relation (4.5.3), together with the above p.g.fs and their uniqueness.

(a) and (b) $\Rightarrow$ (c): Inserting (4.5.8) and (4.5.9) in (4.5.3) and solving gives (4.5.10) with $\lambda = -[\ln(1-\theta)]/\alpha$.

(b) and (c) $\Rightarrow$ (a): Inserting (4.5.9) and (4.5.10) in (4.5.3) and solving results in (4.5.8) with $\lambda = -[\ln(1-\theta)]/\alpha$.

(c) and (a) $\Rightarrow$ (b): Inserting (4.5.10) and (4.5.8) in (4.5.3) and substituting $t = P_X(s)$ yields (4.5.9) with $\lambda = -[\ln(1-\theta)]/\alpha$.

The next theorem gives a similar characterization of extended geometric distribution. With the same notations as in theorem 4.17 we have,

Theorem 4.18

Any two of the following three conditions imply the third.

(a) $P_X(s) = \left(1-\alpha + \frac{\alpha \ln(1-\theta s^k)}{\ln(1-\theta)} \right)^{1/k}$

(b) $P_N(s) = e^{\lambda(s^k-1)}$

(c) $P_S(s) = \frac{1-\theta}{1-\theta s^k}$.

Proof.

Follows as the proof of theorem 4.17.

HARRIS PROCESSES

Sherly Sebastian “Harris family of discrete distributions and processes ” Thesis. Department of Statistics , University of Calicut, 2007

CHAPTER 5

HARRIS PROCESSES

5.1 Introduction

In this chapter we present the Harris distribution as an appropriate marketing distribution for a specific manufacturing unit. Here we develop two models. In the first model the rate λ of the process is considered as a function of n, the number of occurrences of the event at the instant. Then the r.v under consideration follows a Harris distribution and the corresponding stochastic process is termed as 'Harris process'. In the second model, the Poisson rate λ follows a gamma distribution and Harris distribution is obtained as a mixture. The stochastic process considered in this case is also a Harris process. Again, it is proved that Yule – Furry process is a particular case of Harris process. It may be noted that Greenwood and Yule (1920) derived the relationship between the Poisson and negative binomial distribution by considering the intensity parameter λ of the Poisson process as a gamma distributed r.v. Also, Barndorff-Nielsen and Yeo (1969) defined a negative binomial process as a conditional Poisson process whose intensity function $\lambda(t)$ is a gamma process.

This chapter is organized as follows: In section 5.2, the preliminary concepts and results used in the construction of the two models are given. Section 5.3 contains the description and analysis of model 1. Model 2 is presented and analyzed in section 5.4, which provides an alternate approach for obtaining Harris process. In section 5.5

The contents of this chapter has been presented during the 25th Conference of ISPS, held on 28-30 December 2005, at Dept. of Statistics, Bangalore University.

generalized mixed Poisson distribution is defined. Further, generalized mixed Poisson process is developed and Harris process discussed in model 2 is obtained as a special case of generalized mixed Poisson process.

5. 2 Preliminaries

Stochastic Process

Families of random variables which are functions of time, say t, are known as stochastic processes or random processes. Let $X(t)$ be a random variable which is a function of t. The family of random variables $\{X(t), t \in T\}$ is a stochastic process.

State Space

The set of possible values of a single random variable X_n of a stochastic process $\{X_n, n \geq 1\}$ is known as its state space. The state space is discrete if it contains a finite or a denumerable infinity of points; otherwise, it is continuous.

Stationary Process

If for arbitrary $t_1, t_2, \ldots t_n$, the joint distribution of the vector random variables

$$(X(t_1), X(t_2), \ldots, X(t_n)) \quad \text{and} \quad (X(t_1+h), X(t_2+h), \ldots, X(t_n+h))$$

are the same for all $h > 0$, then the stochastic process $\{X(t), t \in T\}$ is said to be stationary of order n. It is strictly stationary if it is stationary of order n for any integer n.

Evolutionary Process

A process which is not stationary is said to be evolutionary.

Poisson Process

The process $\{X(t), t \in [0, \infty)\}$ where the probability mass function of $X(t)$ is

$$P(X(t)=n)=\frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad \lambda>0, \quad n=0,1,2,\ldots\forall t$$

is said to be a Poisson Process. This process is evolutionary as the distribution of $X(t)$ depends on t.

5. 3 Model 1

Consider a company producing costly articles like car, television, refrigerator etc. Increased competition has forced many companies to appoint efficient marketing executives to improve their sales. One of the main problems in a manufacturing unit is the uncertainty in the selling volume which depends on the selling rate λ. In order to enhance the selling of products the company provides incentives to its marketing executives for their performance. Suppose that the company has decided to market its product by taking a policy as follows:

A person is allowed to become a sales executive of the company when he sells one article and he will be given one incentive initially. Thereafter, for the selling of each k articles, (k>0 integer) one additional incentive will be given. Let $I(t)$ denote the number of additional incentives obtained in the interval of duration t starting from an initial epoch t = 0. The family of r.vs $\{I(t),\ t \geq 0\}$ is a stochastic process in continuous time with discrete state space $\{0,1,2,\ldots\}$. Now, let $N(t)$ denote the minimum number of articles sold in order to have $(n+1)$ incentives in the time interval $(0, t]$, where $n = 0,1,2, \ldots$. The family of r.vs $\{N(t),\ t\geq 0\}$ is a stochastic process with $N(0) = 1$. Here the time t is continuous and the state space of $N(t)$, $\{1,\ 1+k,\ 1+2k,\ \ldots\}$ is discrete and integer-valued.

Thus the whole system is a two-dimensional continuous time stochastic process $\{N(t), I(t), t \geq 0\}$ in the state space $\{(1+nk, n);\ \ n \geq 0,\ \ k > 0 \text{ integers}\}$.

Let $P_{nk+1}(t)$ be the probability that the r.v $N(t)$ assumes the value $nk+1$.

$P_{nk+1}(t) = P\{N(t) = nk+1\}$, $n = 0, 1, 2, \ldots$, $k > 0$ integer.

$P_{nk+1}(t)$ is a function of the time t and $\sum_{n=0}^{\infty} P_{nk+1}(t) = 1$. $\{P_{nk+1}(t)\}$ represents the probability distribution of the r.v $N(t)$ for every value of t.

Let the probability of getting r incentives in $(t, t+\Delta t)$ given that n incentives were obtained by epoch t be given by

$$
\begin{aligned}
P\big(I(\Delta t) = r / I(t) = n\big) &= \lambda_n \Delta t + 0(\Delta t), && r = 1 \\
&= 0(\Delta t) && r \geq 2 \\
&= 1 - \lambda_n \Delta t + 0(\Delta t) && r = 0
\end{aligned}
$$

where $\lambda_n = (nk+1)\lambda$ is a linear function of n and the initial selling rate $\lambda > 0$ is a parameter. Similar models with different linear structures are used in birth and death processes. For examples, see section 6.3 of Ross (2003).

Theorem 5.1

Under the above conditions $N(t)$ follows a Harris distribution, *i.e.* $P_{nk+1}(t)$ is given by the Harris law:

$$
P_{nk+1}(t) = \binom{(1/k)+n-1}{n} (e^{-t\lambda k})^{1/k} (1 - e^{-t\lambda k})^n , \quad n = 0, 1, 2, \ldots
$$

Proof.

Consider $P_{nk+1}(t+\Delta t)$ for $n \geq 0$. For $n \geq 1$, the probability of selling $(nk+1)$ items in order to have $(n+1)$ incentives by time $(t+\Delta t)$ can be written as

$$P_{nk+1}(t+\Delta t) = P_{nk+1}(t)\big(1-(nk+1)\lambda\,\Delta t\big) + P_{(n-1)k+1}(t)\big((n-1)k+1\big)\lambda\,\Delta t + 0(\Delta t) \quad (5.3.1)$$

where as for $n = 0$,

$$P_1(t+\Delta t) = P_1(t)\big(1-\lambda\,\Delta t\big) + 0(\Delta t) \qquad (5.3.2)$$

Combining (5.3.1) and (5.3.2), we get

$$P_{nk+1}(t+\Delta t) = P_{nk+1}(t)\big(1-(nk+1)\lambda\,\Delta t\big) + P_{(n-1)k+1}(t)\big((n-1)k+1\big)\lambda\,\Delta t + 0(\Delta t) \quad (5.3.3)$$

for $n \geq 0$ with $P_{1-k}(t) = 0$. Hence we have

$$\frac{P_{nk+1}(t+\Delta t) - P_{nk+1}(t)}{\Delta t} = \big((n-1)k+1\big)\lambda\ P_{(n-1)k+1}(t) - (nk+1)\lambda\ P_{nk+1}(t) + \frac{0(\Delta t)}{\Delta t}$$

Taking the limit as $\Delta t \rightarrow 0$, we get

$$\frac{d}{dt}P_{nk+1}(t) = \big((n-1)k+1\big)\lambda P_{(n-1)k+1}(t) - (nk+1)\lambda P_{nk+1}(t) \qquad (5.3.4)$$

Since the process starts by the selling of the first item $N(0) = 1$. To the system (5.3.3), we add the initial condition

$$P_1(0) = 1; \quad P_r(0) = 0 \qquad \text{for} \quad r \neq 1. \qquad (5.3.5)$$

The probability generating function $P(s, t)$ for the probability distribution $P_{nk+1}(t)$ is

$$P(s,t) = \sum_{n=0}^{\infty} P_{nk+1}(t)\, s^{nk+1}$$

where $P(s, t)$ is a function of s and t. Then,

$$\frac{\partial P}{\partial s} = \sum_{n=0}^{\infty}(nk+1)P_{nk+1}(t)\, s^{nk} \qquad \text{and} \qquad \frac{\partial P}{\partial t} = \sum_{n=0}^{\infty}\frac{\partial}{\partial t}P_{nk+1}(t)\, s^{nk+1}$$

Now, the initial condition (5.3.5) can be expressed as

$$P(s,0) = s$$

Multiply (5.3.4) by s^{nk+1} and sum over all values of n. This gives

$$\sum_{n=0}^{\infty}\frac{\partial}{\partial t}P_{nk+1}(t)\,s^{nk+1}=\lambda\sum_{n=0}^{\infty}\big((n-1)k+1\big)P_{(n-1)k+1}(t)\,s^{nk+1}\;-\lambda\sum_{n=0}^{\infty}(nk+1)P_{nk+1}(t)\;s^{nk+1}$$

$$\frac{\partial P(s,t)}{\partial t}=\lambda s^{k+1}\sum_{n=1}^{\infty}\big((n-1)k+1\big)P_{(n-1)k+1}(t)\;s^{(n-1)k}\;-\lambda s\sum_{n=0}^{\infty}(nk+1)P_{nk+1}(t)\;s^{nk}$$

$\dfrac{\partial P(s,t)}{\partial t}=\lambda\, s^{k+1}\,\dfrac{\partial P(s,t)}{\partial s}-\lambda\, s\dfrac{\partial P(s,t)}{\partial s}$. Thus we have

$$\frac{\partial P(s,t)}{\partial t}+\lambda\, s\,(1-s^k)\,\frac{\partial P(s,t)}{\partial s}=0 \tag{5.3.6}$$

Auxiliary equations are

$$\frac{dt}{1}=\frac{ds}{\lambda\, s(1-s^k)}=\frac{dP}{0},$$

which is similar to the example 1.4.4 in Amaranath (1997, p.16).

Integrating $\dfrac{dt}{1}=\dfrac{ds}{\lambda\, s(1-s^k)}$ we get

$$t=\frac{1}{\lambda k}\log\left(\frac{s^k}{1-s^k}\right)+\log c_1$$

or $\quad e^t\left(\dfrac{1-s^k}{s^k}\right)^{1/\lambda k}=c_1$, a constant.

Now, integrating $\dfrac{dt}{1}=\dfrac{dP}{0}$ we get $P(s,t)=c_2$, a constant.

If $u=c_1$ and $v=c_2$ then general solution of (5.3.6) can be written as $v=f(u)$, *i.e.*

$$P(s,t)=f\left(e^t\left(\frac{1-s^k}{s^k}\right)^{1/\lambda k}\right) \tag{5.3.7}$$

where f is an arbitrary function. Initial conditions $t=0$ and $P(s,0)=s$ in (5.3.7) gives

$$s = f\left(\left(\frac{1-s^k}{s^k}\right)^{1/\lambda k}\right) = f(y) \quad \text{(say) where}$$

$$y = \left(\frac{1-s^k}{s^k}\right)^{1/\lambda k} \qquad \text{or} \qquad s = \left(\frac{1}{1+y^{\lambda k}}\right)^{1/k}$$

$$f(y) = \left(\frac{1}{1+y^{\lambda k}}\right)^{1/k}$$

Therefore from (5.3.7) we have $P(s,t) = \left(\frac{1}{1+e^{t\lambda k}\left((1-s^k)/s^k\right)}\right)^{1/k}$. i.e.

$$P(s,t) = \frac{s}{(e^{t\lambda k} - (e^{t\lambda k}-1)s^k)^{1/k}} \text{, where} \quad e^{t\lambda k} > 1 \tag{5.3.8}$$

This is the p.g.f of the Harris distribution $H_1(e^{t\lambda k}, k, 1/k)$. Hence $N(t)$ follows a Harris distribution. Picking out the coefficient of s^{nk+1} on the right side of (5.3.8), we have

$$P_{nk+1}(t) = \binom{(1/k)+n-1}{n}(e^{-t\lambda k})^{1/k}(1-e^{-t\lambda k})^n, \quad n = 0, 1, 2, \ldots \tag{5.3.9}$$

This completes the theorem.

Corollary 5.1

The mean and variance of the process are

$E\{N(t)\} = e^{t\lambda k}$ and

$V\{N(t)\} = e^{t\lambda k}(e^{t\lambda k} - 1)k$

Remarks 5.1

(i) The mean number of items sold in order to have $(n+1)$ incentives in an interval of length t is $e^{t\lambda k}$, so that the mean number of items sold for $(n+1)$ incentives per unit time ($t = 1$), i.e. in an interval of unit length is $e^{\lambda k}$.

(ii) The mean and the variance of $N(t)$ are functions of t and the distribution of $N(t)$ functionally dependent on t. As such the process $\{N(t), t \geq 0\}$ is not stationary – it is evolutionary.

(iii) We have $N(t) = 1 + k\,I(t)$

$P(I(t) = n) = P(N(t) = nk + 1)$

$$= \binom{(1/k)+n-1}{n} (e^{-t\lambda k})^{1/k} (1 - e^{-t\lambda k})^n \ , \quad n = 0, 1, 2, \ldots$$

which is the p.m.f of the negative binomial distribution $NB(e^{-t\lambda k}, 1/k)$ defined on $\{0, 1, 2, \ldots\}$. The mean and variance of $I(t)$ are

$E\{I(t)\} = (e^{t\lambda k} - 1) / k$ and

$V\{I(t)\} = e^{t\lambda k} (e^{t\lambda k} - 1) / k$

(iv) In the above model we consider a situation where a success can occur only when k (> 1, an integer) fractional successes happen. For a marketing executive, selling of each item is a fractional success and obtaining an incentive is a success and success occurs only when the k fractional successes happen *i.e.* when k items are sold. Let X_i denote the number of failures preceding the i^{th} fractional success, $i = 1, 2, \ldots, k$. Then $X_i \sim NB(p, 1/k)$ where the probability of success p is assumed to be a constant for each Bernoulli trial. Then their sum $Y = \sum_{i=1}^{k} X_i$, the total number of failures preceding a success follows a $Geo(p)$, a geometric distribution with parameter p. *i.e.* the sum of k i.i.d negative binomial variables following $NB(p, 1/k)$ is $Geo(p)$. This distribution can be used to evaluate the efficiency of a marketing executive and the demand for the product.

$\{N(t), t \geq 0\}$ is a stochastic process which we call 'Harris process'.

The auto-correlation coefficient between $N(t)$ and $N(t+s)$

If $\{N(t)\}$ is a Harris process then the (correlation) auto-correlation coefficient between $N(t)$ and $N(t+s)$ is

$$\rho(t, t+s) = \left(\frac{e^{t\lambda k} - 1}{e^{s\lambda k}\left(e^{(t+s)\lambda k} - 1\right)} \right)^{1/2}$$

Proof.

$$E\{N(T)\} = e^{t\lambda k} \quad V\{N(T)\} = e^{t\lambda k}(e^{t\lambda k} - 1)k$$

and $E(N^2(T) = ke^{2t\lambda k} + e^{2t\lambda k} - ke^{t\lambda k}$ for $T = t$ and $t + s$.

Since $N(t)$ and $\{N(t+s) - N(t)\}$ are independent

$$E\{N(t)\, N(t+s)\} = E[N(t)\{N(t+s) - N(t) + N(t)\}]$$

$$= E\{N^2(t)\} + E\{N(t)\}E\{N(t+s) - N(t)\}$$

$$= ke^{2t\lambda k} + e^{2t\lambda k} - ke^{t\lambda k} + e^{t\lambda k}\{e^{(t+s)\lambda k} - e^{t\lambda k}\}$$

$$= e^{\lambda k(2t+s)} + ke^{t\lambda k}(e^{t\lambda k} - 1)$$

The auto covariance between $N(t)$ and $N(t+s)$ is given by

$$C(t, t+s) = E\{N(t)N(t+s)\} - E\{N(t)\}E\{N(t+s)\}$$

$$= e^{\lambda k(2t+s)} + ke^{t\lambda k}(e^{t\lambda k} - 1) - e^{t\lambda k}e^{(t+s)\lambda k}$$

$$= e^{t\lambda k}(e^{t\lambda k} - 1)k$$

Hence the auto correlation function

$$\rho(t, t+s) = \frac{C(t, t+s)}{[V\{N(t)\}V\{N(t+s)\}]^{1/2}}$$

$$=\left(\frac{e^{t\lambda k}-1}{e^{s\lambda k}\left(e^{(t+s)\lambda k}-1\right)}\right)^{1/2}$$

Distribution of occurrence time of the first additional incentive

The distribution of the occurrence time of the first additional incentive is exponential.

Proof.

Taking $n = 0$ in (5.3.9) we get

$$P_1(t) = P\{N(t) = 1\} = P\{I(t) = 0\} = e^{-\lambda t}$$

which is the probability that at time t no additional incentive is obtained. *i.e.* it is the probability that the first additional incentive obtains at some instant greater than t.

Now if U denotes the time of occurrence of the first additional incentive then

$$P(U > u) = e^{-\lambda u}$$

$$P(U \leq u) = 1 - e^{-\lambda u}$$

which is the d.f of the occurrence time of the first additional incentive. The corresponding frequency function is

$$f(u) = \lambda e^{-\lambda u},\ 0 \leq u < \infty .$$

Occurrence time of the first additional incentive is *Exponential* (λ).

Yule-Furry process as a particular case of Harris process

When $k = 1$, the distribution of $N(t)$ is given by

$$P_n(t) = P(N(t) = n) = (e^{-t\lambda})(1 - e^{-t\lambda})^{n-1}\ ,\quad n = 1, 2, \ldots$$

which is the p.m.f of the modified geometric or decapitated geometric with parameter $e^{-\lambda t}$ and its p.g.f is given by

$$P(s,t) = \frac{e^{-\lambda t}s}{1-(1-e^{-\lambda t})s}$$

The mean and variance of the corresponding process are

$E\{N(t)\} = e^{t\lambda}$ and

$V\{N(t)\} = e^{t\lambda}(e^{t\lambda} - 1)$

So when $k = 1$, the stochastic process $\{N(t), t \geq 0\}$ is a Yule-Furry process.

5.4 Model 2

In this section we have developed a stochastic model where the variable under consideration follows a Harris distribution. We start with a Poisson process having rate λ distributed as a gamma r.v. Then a linear function of the Poisson variable leads to a Harris process.

The Poisson distribution arises naturally in the study of count data. Count data refers to just about anything measured by counts and count models are used in a wide range of disciplines. For a homogeneous Poisson process$\{X(t), t \geq 0\}$ having rate λ, the number of events in any interval of length t is Poisson distributed with mean λt. That is, for all $t \geq 0$

$$P(X(t) = n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \qquad n = 0, 1, 2, \ldots$$

See Karlin and Taylor (1975), for details.

Consider a linear function of $X(t)$, say, $Z(t)=k\,X(t)+1$ where $k > 0$ integer. The family of r.vs $\{Z(t), t \geq 0\}$ is a stochastic process with state space $\{1, 1+k, 1+2k, \ldots\}$.

Let $P'_{nk+1}(t) = P(Z(t) = nk+1)$

Theorem 5.2

Under the above conditions $Z(t)$ follows a Harris distribution, provided the mixing distribution is $Gamma(1/k, a)$. *i.e.* $P'_{nk+1}(t)$ is given by the Harris law:

$$P'_{nk+1}(t) = \binom{(1/k)+n-1}{n}\left(\frac{a}{a+t}\right)^{1/k}\left(\frac{t}{a+t}\right)^{n}, \quad n = 0, 1, 2, \ldots$$

Proof.

The conditional distribution of $Z(t)$ given λ is

$$P(Z(t) = nk+1 \mid \lambda) = P(X(t) = n \mid \lambda) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \ldots$$

Assume that the parameter λ is a gamma r.v with p.d.f

$$f(\lambda) = \frac{a^{1/k}}{\Gamma(1/k)} e^{-a\lambda}\lambda^{(1/k)-1}, \quad k > 0 \text{ integer}, a > 0, \lambda > 0.$$

The mean and variance of λ are $\mu_\lambda = 1/ak$ and $\sigma_\lambda^2 = 1/a^2k$ respectively.

The unconditional distribution of $Z(t)$ is given by

$$\begin{aligned}
P'_{nk+1}(t) &= P(Z(t) = nk+1) \\
&= \int P(Z(t) = nk+1 \mid \lambda) f(\lambda) d\lambda \\
&= \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^n}{n!} \frac{a^{1/k}}{\Gamma(1/k)} e^{-a\lambda}\lambda^{(1/k)-1} d\lambda \\
&= \binom{(1/k)+n-1}{n}\left(\frac{a}{a+t}\right)^{1/k}\left(\frac{t}{a+t}\right)^{n} \\
&= \binom{(1/k)+n-1}{n}\left(\frac{1}{m}\right)^{1/k}\left(1-\frac{1}{m}\right)^{n}, \quad n = 0, 1, 2, \ldots
\end{aligned}$$

where $k > 0$ is an integer and $m = (a+t)/a > 1$. This is the p.m.f of the Harris distribution with index $1/k$. *i.e.* $Z(t) \sim H_1(m, k, 1/k)$.

Corollary 5.2

The mean and variance of the process are

$E\{Z(t)\} = (a+t)/a$ and

$V\{Z(t)\} = (a+t)tk/a^2$

Remarks 5.2

(1) The mean and variance of $Z(t)$ are functions of t. Since the distribution of $Z(t)$ is functionally dependent on t, the process $\{ Z(t), t \geq 0 \}$ is non- stationary.

(2) When $k = 1$, the distribution of $Z(t)$ is given by

$$P_n'(t) = P(Z(t) = n) = \left(\frac{a}{a+t}\right)\left(\frac{t}{a+t}\right)^{n-1} \quad n = 1, 2, \ldots$$

which is the p.m.f of the modified geometric with parameter $a/(a+t)$. So when $k = 1$, the stochastic process $\{Z(t), t \geq 0\}$ is Yule-Furry process.

The auto-correlation coefficient between $Z(t)$ and $Z(t+s)$

If $\{Z(t)\}$ is a Harris process then the auto-correlation coefficient between $Z(t)$ and $Z(t+s)$ is

$$\rho(t, t+s) = \left(\frac{(a+t)t}{(a+t+s)(t+s)}\right)^{1/2}$$

Proof.

$E\{Z(T)\} = (a+T)/a$, $V\{Z(T)\} = (a+T)Tk/a^2$

and $E(Z^2(T)) = (a+T)(a+T+Tk)/a^2$ for $T = t$ and $t+s$.

Since $Z(t)$ and $\{Z(t+s) - Z(t)\}$ are independent

$$E\{Z(t)\,Z(t+s)\} = E[Z(t)\{Z(t+s) - Z(t) + Z(t)\}]$$

$$= E\{Z^2(t)\} + E\{Z(t)\}E\{Z(t+s) - Z(t)\}$$

$$= (a+t)(a+t+s+tk)/a^2$$

The auto covariance between $Z(t)$ and $Z(t+s)$ is given by

$$C(t,t+s) = E\{Z(t)Z(t+s)\} - E\{Z(t)\}E\{Z(t+s)\}$$

$$= (a+t)tk/a^2$$

Hence the auto correlation function

$$\rho(t,t+s) = \left(\frac{(a+t)t}{(a+t+s)(t+s)}\right)^{1/2}$$

Distribution of the time of first occurrence

The interval of time upto the first occurrence has a Pearson's Type V1 distribution.

Proof.

Putting $n = 0$ in $P'_{nk+1}(t)$ we have,

$$P_1'(t) = P(Z(t) = 1) = P(X(t) = 0) = \left(a/(a+t)\right)^{1/k}$$

This is the probability that the first occurrence takes place at some instant greater than t.

If U denotes the time of the first occurrence, then

$$P(U > u) = \left(a/(a+u)\right)^{1/k}$$

Then the d.f of U is given by

$$P(U \le u) = 1 - \left(a/(a+u)\right)^{1/k}$$

The corresponding p.d.f

$$f(u) = \frac{1}{k}\left(\frac{a}{a+u}\right)^{1/k}\left(\frac{1}{a+u}\right), \qquad 0 < u < \infty$$

is the p.d.f of a Pearson's Type V1 distribution.

Note: If $W = U/a$ then the p.d.f of W is given by

$$g(w) = \frac{1}{k(1+w)^{(1/k)+1}}, \qquad 0 < w < \infty$$

This is the p.d.f. of $\beta_2(1, 1/k)$, a beta distribution of the second kind.

The Harris process $\{Z(t), t \geq 0\}$ is not renewal.

By Sandhya and Satheesh (1997) a mixed Poisson process is renewal if and only if the mixing distribution is degenerate. In model 2, $Z(t)$ is obtained as a mixed Poisson process with mixing distribution gamma and hence the process cannot be renewal.

Numerical illustration revealing the relationship between the mixing distribution *Gamma*$(1/k, a)$ and $Z(t)$ is given in the Table 5.1. Also their graphs are presented together for the purpose of comparison. We have chosen the relatively large value $t = 10$, in order to see how $P'_{nk+1}(t)$ behaves for small values of n. Figures 5.1-5.3 exhibit the relation between the distributions of λ and $Z(t)$ in the cases given in Table 5.1. It can be seen that the variability pattern of the distributions of λ and $Z(t)$ are alike.

5.5 Harris Process as a Special Case of Generalized Mixed Poisson Process

As mentioned in Gnedenko (1978), generalizing the concept of Poisson's law we will say that a r.v Y is Poisson distributed if it can assume only $nb+c$ where b and c are real constants and

$$P(Y = nb+c) = \frac{e^{-\lambda}\lambda^n}{n!}, \quad n = 0, 1, 2, \ldots \tag{5.5.1}$$

Let λ be a non negative r.v with p.d.f $f(\lambda)$, then, (5.5.1) gives the conditional distribution of Y given λ. The unconditional distribution of Y, in this case is

$$P\{Y = nb+c\} = \int_0^\infty P(Y = nb+c \mid \lambda) f(\lambda) d\lambda, \quad n = 0, 1, 2, \ldots$$

and this distribution is a generalized mixed Poisson distribution.

Similarly we can define a generalized mixed Poisson process. Consider a Poisson process $\{X(t),\ t \geq 0\}$ with intensity parameter λ and let $Y(t) = bX(t)+c$ where b and c are real constants. The family of r.vs $\{Y(t),\ t \geq 0\}$ is a stochastic process in continuous time with discrete state space $\{c,\ b+c,\ 2b+c,\ \ldots,\ nb+c,\ \ldots\}$and for all $t \geq 0$,

$$P\{Y(t) = nb+c\} = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \ldots \tag{5.5.2}$$

When λ is a non negative r.v, (5.5.2) gives the conditional distribution of $Y(t)$ given λ. The unconditional distribution of $Y(t)$ is

$$P\{Y(t) = nb+c\} = \int_0^\infty P(Y(t) = nb+c \mid \lambda) f(\lambda) d\lambda$$

$$= \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^n}{n!} f(\lambda) d\lambda, \quad n = 0,1,2,\ldots$$

The r.v $Y(t)$ follows generalized mixed Poisson law and the stochastic process $\{Y(t),\ t \geq 0\}$ in this case is called generalized mixed Poisson process.

Proposition

Let $Y(t)$ follows generalized mixed Poisson law, where the mixing distribution of λ is *Gamma*$(1/k,\ a)$ with mean $\mu_\lambda = 1/ak$ and variance $\sigma_\lambda^2 = 1/a^2k$. Then

(1) $E\{Y(t)\} = \dfrac{bt}{ak} + c$

(2) $V\{Y(t)\} = \dfrac{b^2 t}{ak} + \dfrac{b^2 t^2}{a^2 k}$

(3) The generating function $P_{Y(t)}(s)$ of $Y(t)$ is given by

$$P_{Y(t)}(s) = s^c \left(\frac{a}{a + t(1 - s^b)} \right)^{1/k}, \quad s \leq 1;$$

Proof.

We have

$$E\{Y(t)\} = E[E\{Y(t) \mid \lambda\}] = \frac{bt}{ak} + c \quad \text{and}$$

$$V\{Y(t)\} = E[V\{Y(t) \mid \lambda\}] + V[E\{Y(t) \mid \lambda\}] = \frac{b^2 t}{ak} + \frac{b^2 t^2}{a^2 k},$$

which proves (1) and (2).

Now the p.g.f

$$P_{Y(t)}(s) = \sum_{n=0}^{\infty} s^{nb+c}\, P\{Y(t) = nb + c\}$$

$$= s^c \int_0^{\infty} e^{\lambda t s^b} e^{-\lambda t} \frac{a^{1/k}}{\Gamma(1/k)} e^{-a\lambda} \lambda^{(1/k)-1} d\lambda$$

$$= s^c \varphi\{t(1 - s^b)\}, \quad s \leq 1$$

where $\varphi(s) = \int_0^{\infty} e^{-\lambda s} f(\lambda) d\lambda$ is the LT of λ. Hence we have

$$P_{Y(t)}(s) = s^c \left(\frac{a}{a + t(1 - s^b)} \right)^{1/k}, \quad s \leq 1;$$

which proves (3).

Remark 5.3

When $b = k$ and $c = 1$,

$E\{Y(t)\} = (a+t)/a$

$V\{Y(t)\} = (a+t)tk/a^2$ and

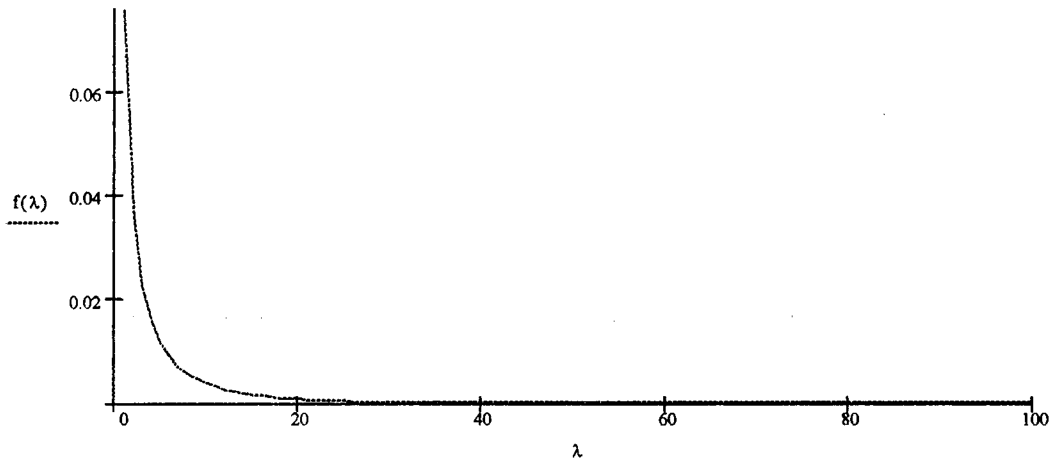
$$P_{Y(t)}(s) = \left(\frac{a}{a+t(1-s^k)}\right)^{1/k}, \quad s \le 1.$$

Here $Y(t) \sim H_1(m,k,1/k)$ with mean, $m = (a+t)/a > 1$. The stochastic process $\{Y(t),\ t \ge 0\}$ in this case is a Harris process, which is given in model 2. This shows that Harris process is a special case of generalized mixed Poisson process.

TABLE 5.1

Distributions illustrated in Figures 5.1 – 5.3

	a	k	t	μ_λ	σ_λ	$\sigma_\lambda / \mu_\lambda$	E(Z)	S.D(Z)	S.D(Z)/E(Z)
Fig. 5.1	0.1	10	10	1.0	3.162	3.162	101	317.805	3.147
Fig. 5.2	0.5	2	10	1.0	1.414	1.414	21	28.983	1.38
Fig. 5.3	2.0	5	10	0.1	0.224	2.240	6	12.247	2.041



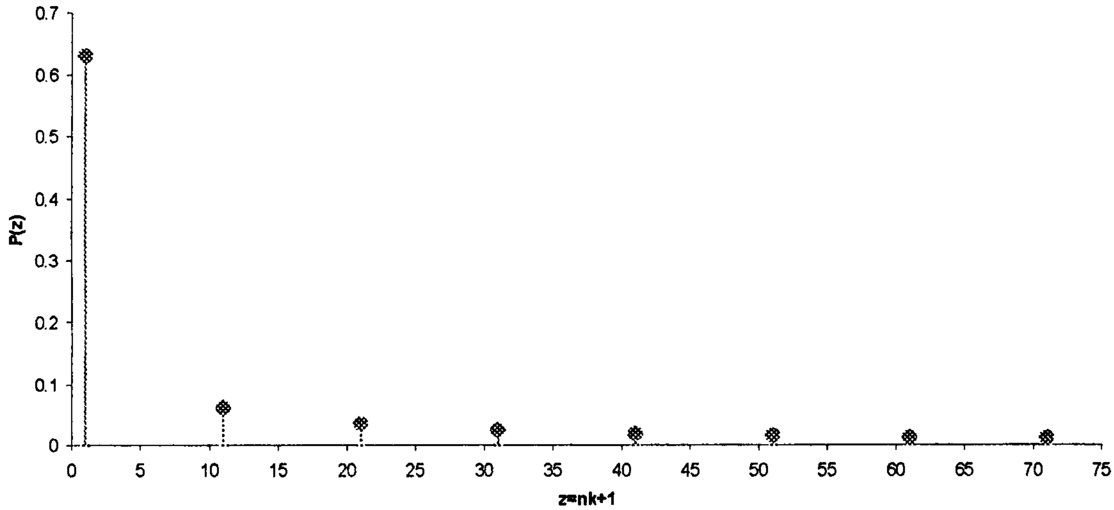
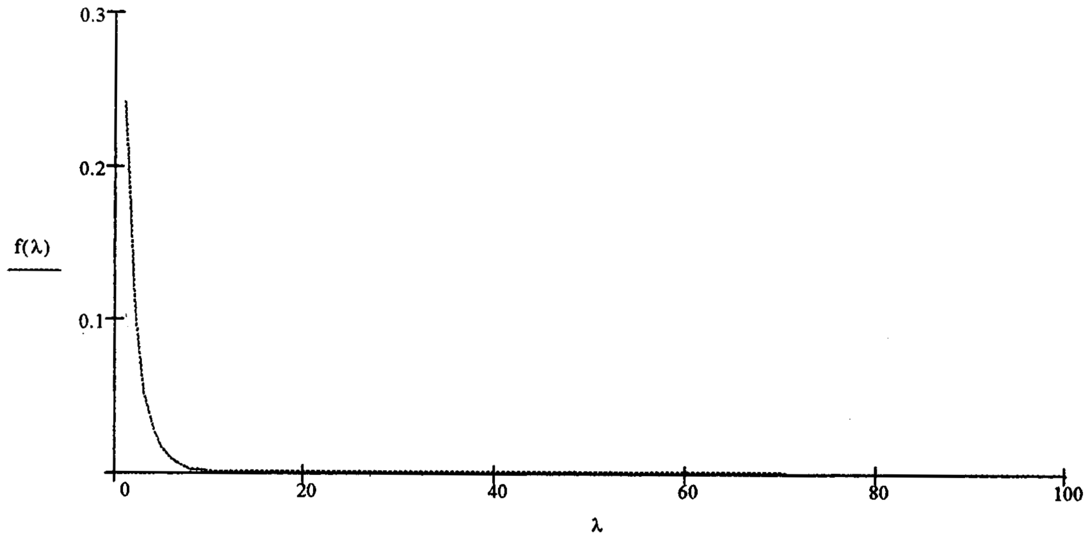


FIGURE 5.1 - Distributions of λ and $Z(t)$ for $a = 0.1$, $k = 10$ and $t = 10$.



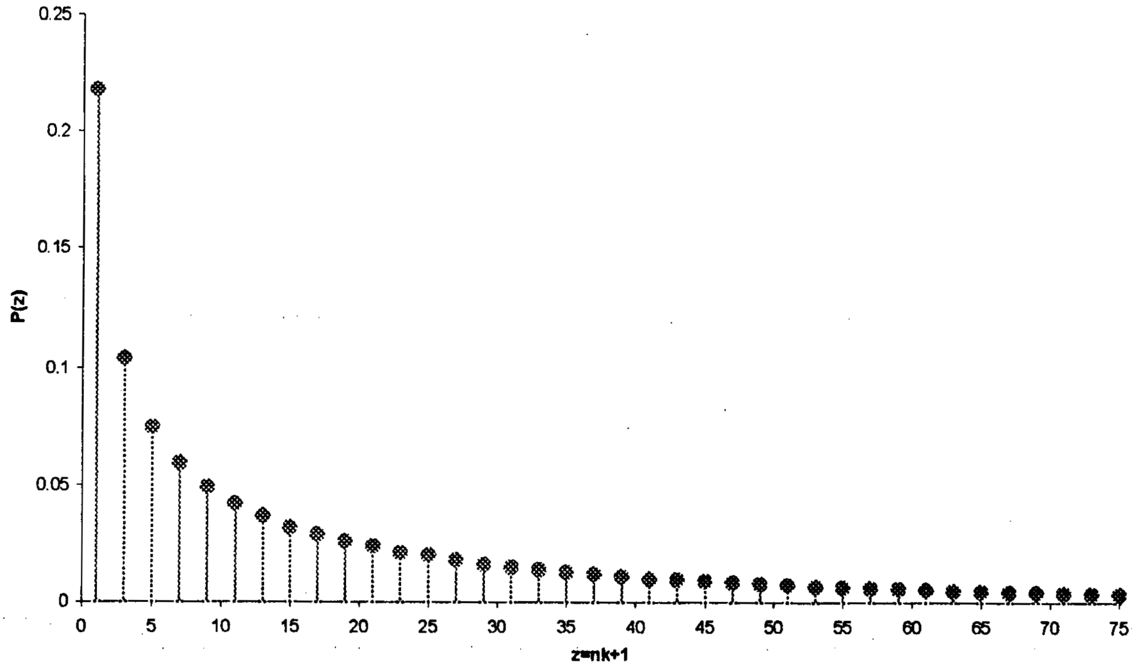
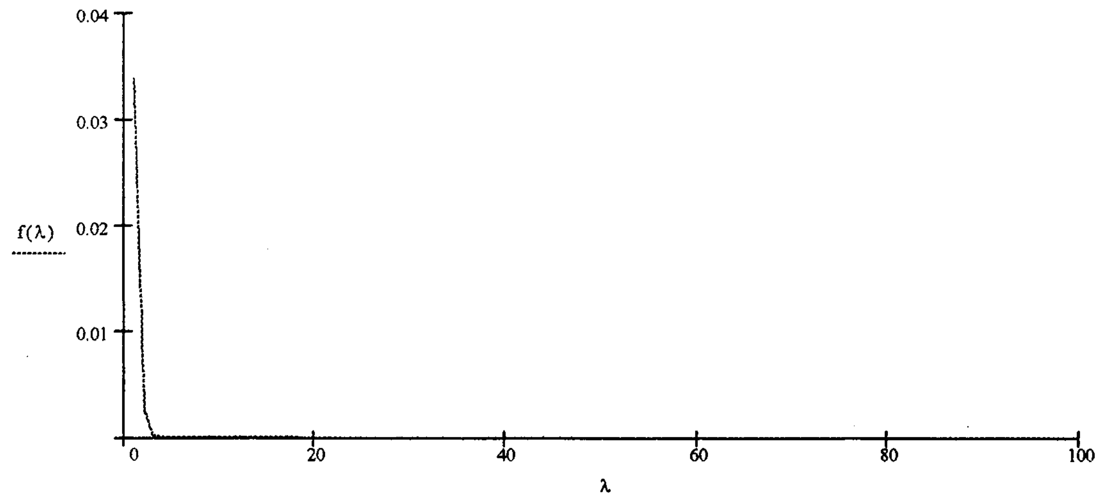


FIGURE 5.2 - Distributions of λ and $Z(t)$ for $a = 0.5$, $k = 2$ and $t = 10$.



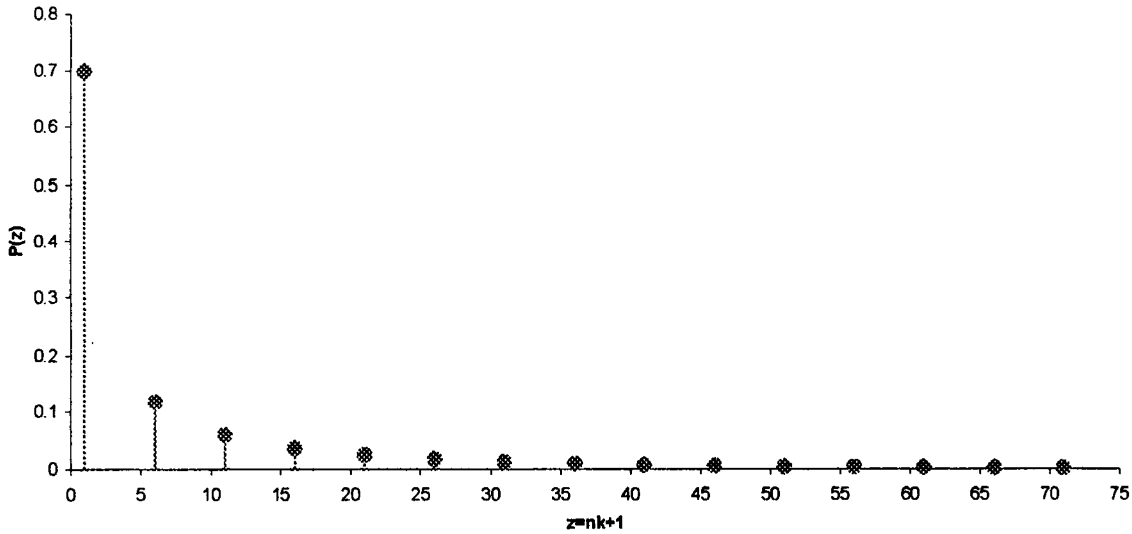


FIGURE 5.3 - Distributions of λ and $Z(t)$ for $a = 2$, $k = 5$ and $t = 10$.

A GENERALIZATION OF STATIONARY AR(1) SCHEMES

Sherly Sebastian "Harris family of discrete distributions and processes " Thesis. Department of Statistics , University of Calicut, 2007

CHAPTER 6

A GENERALIZATION OF STATIONARY AR(1) SCHEMES

6.1 Introduction

A sequence $\{X_n\}$ of r.vs describes the additive first order autoregressive (AR(1)) scheme considered here if there exists an innovation sequence $\{\varepsilon_n\}$ of i.i.d r.vs satisfying

$$X_n = bX_{n-1} + \varepsilon_n, \quad \forall\ n>0 \text{ integer and some } 0<b<1. \tag{6.1.1}$$

$\{X_n\}$ is marginally stationary if $X_n \overset{d}{=} X_{n-1}$ $\forall\ n>0$ integer.

The AR(1) sequence $\{X_n\}$ may be composed of k independent AR(1) sequences $\{Y_{i,n}\}$, $i = 1,2, \ldots, k$, and where for each $n>0$ integer $Y_{i,n}$, $i = 1,2, \ldots, k$ are identically distributed. That is, for each n, $X_n \overset{d}{=} \sum_{i=1}^{k} Y_{i,n}$ and $Y_{i,n}$, $i = 1,2, \ldots, k$ are i.i.d, k being a fixed positive integer. For example, the variable X_n could be the quantity of water flowing through a river, or the number of patients in a hospital, or the sales of a particular item by an agency, or the number of items produced in a factory having more than one plant for the production of the same. In all these cases the resultant observation X_n is either the sum of the quantities $Y_{i,n}$, $i = 1,2, \ldots, k$ of water flowing through k tributaries of the river, or the sum of the number of patients $Y_{i,n}$, $i = 1,2, \ldots, k$ in k different specialities in the hospital, or the sum of the sales $Y_{i,n}$, $i = 1,2, \ldots, k$ by the agency through their k different retail outlets or the sum of the quantities produced $Y_{i,n}$, $i = 1,2, \ldots, k$ at the k different plants in the factory. Thus we generalize the AR(1) model in (6.1.1) where the

---

The contents of this chapter has been accepted for publication in the journal, ***Statistical Methods.***

observations X_n are the sum of k i.i.d r.vs, and then extend the discussion to the maximum and minimum schemes.

The definitions and some preliminary results are given in section 6.2. In section 6.3 we establish the connection between a summation scheme introduced in (6.1.1), which is described by (6.3.1) and semi-self decomposability. A maximum scheme of model (6.3.1) is introduced in (6.3.2) and it is proved that a sequence $\{Y_{i,n}\}$ satisfies (6.3.2) if $Y_{i,n}$ is max-semi-self decomposable. Also a necessary and sufficient condition is obtained in order that $\{Y_{i,n}\}$ satisfy the scheme (6.3.3). Again we consider a max. version of (6.3.2) given by (6.3.4) and consequently a necessary and sufficient condition for $\{Y_{i,n}\}$ to satisfy (6.3.4) is given.

In section 6.4, we consider a discrete version of (6.3.3) given by (6.3.3a). Here we show that a sequence $\{Y_{i,n}\}$ of non-negative integer-valued r.vs satisfies AR(1) scheme (6.3.3a), if and only if $Y_{i,n}$ is discrete generalized semi-*ML*. Discrete semi- self decomposability is introduced and it is proved that a sequence $\{X_n\}$ of integer-valued r.vs defines a marginally stationary AR(1) scheme on $\{0, h, 2h, 3h, \ldots\}$, if X_n is discrete semi-self decomposable. In section 6.5 we define min-semi-self decomposable laws. It is shown that a sequence $\{Y_{i,n}\}$ is marginally stationary min-AR(1) scheme given by (6.5.1) if $Y_{i,n}$ is min-semi-self decomposable. Also a necessary and a sufficient condition for a sequence $\{Y_{i,n}\}$ of non-negative r.vs defines a marginally stationary min-AR(1) scheme given by (6.5.2) with $Y_{i,0} \overset{d}{=} \varepsilon_{i,1}$ is given in theorem 6.14.

6.2 Definitions and Preliminary Results

Definition 6.1 (Pillai, 1985). A c.f ϕ is semi-α-Laplace(a,b) if $\forall t \in R$ and for some $0<b<1<a$,

$$\phi(t) = 1/\{1+\psi(t)\}, \text{ where } \psi(t) = a\psi(bt), \; ab^{\alpha}=1, \; \alpha \in (0,2].$$

Satheesh *et al.* (2002) has considered a generalized semi-α-Laplace(a,b,k) law having c.f $\{1+\psi(t)\}^{-1/k}$ where $\psi(t)$ is as above. The generalized semi Mittag-Leffler(a,b,k) (ML) laws with LT $\varphi(\lambda) = \{1+\psi(\lambda)\}^{-1/k}$, $\lambda>0$, $\psi(\lambda) = a\psi(b\lambda)$, $0<b<1<a$, $ab^{\alpha}=1$, $\alpha \in (0,1]$ is its non-negative analogue and the generalized discrete semi-$ML(a,b,k)$ laws having p.g.f $P(s) = \{1+\psi(1\text{-}s)\}^{-1/k}$, $0<s<1$, where $\psi(1\text{-}s) = a\psi(b(1\text{-}s))$, $0<b<1<a$, $ab^{\alpha}=1$, $\alpha \in (0,1]$ is its non-negative integer-valued analogue, see Satheesh *et al.* (2002). In these families only the generalized Laplace law with c.f $\{1+\eta t^2\}^{-1/k}$ has finite variance. Similarly, in the non-negative case only the $Gamma(1/k, \eta)$ law has finite mean and in the discrete case only the negative binomial $(\eta, 1/k)$ $[NB(\eta, 1/k)]$ law has finite mean. Also if $\psi(t) = a\psi(bt)$ for two values of b, say b_1 and b_2 such that $ln(b_1)/ln(b_2)$ is irrational then $\psi(t) = \eta\,|t|^{\alpha}$, $\eta>0$.

Definition 6.2 (Maejima and Naito, 1998). A c.f ϕ is semi-selfdecomposable(b) (semi-selfdec(b)) if for some $0<b<1$ there exists a c.f ϕ_0 that is i.d such that

$$\phi(t) = \phi(bt)\phi_0(t), \; \forall t \in R.$$

If this relation holds for every $0<b<1$ then ϕ is selfdecomposable.

Definition 6.3 (Megyesi, 2002). A d.f F is max-semi-stable(a,c) if either,

$$F(x) = exp\{-x^{-\alpha}h(ln(x))\}, \; x>0, \; \alpha>0,$$

where $h(x)$ is a positive bounded periodic function with period $ln(c)$, $c>1$, and there exists an $a>1$ such that $ac^{-\alpha}= 1$, or,

$$F(x) = exp\{-|x|^{\alpha} h(ln(|x|))\}, x<0, \alpha >0,$$

where $h(x)$ is as above with period $|ln(c)|$, $c<1$, and there is an $a>1$ such that $ac^{\alpha}= 1$.

Remark 6.1

The first d.f in definition 6.3 can be represented in the form $exp\{-\psi(x)\}$, where $\psi(x)$ satisfies $\psi(x) = a\psi(cx)$, $x>0$ for some $a>1$, $c>1$, and $\alpha>0$ satisfying $ac^{-\alpha} =1$. Similarly, the second one also, where $\psi(x) = a\psi(cx)$, $x<0$, for some $a>1$, $c<1$, and $\alpha>0$ satisfying $ac^{\alpha}= 1$, see Satheesh and Sandhya (2006*b*).

Definition 6.4 (Becker-Kern, 2001). A non-degenerate d.f F is max-semi-selfdecomposable(c) if for some $c>1$ and $v\in R$ there is a non-degenerate d.f H such that

$$F(x) = F(c^{v}x +\beta)\, H(x), x\in R,$$

where $\beta = 0$ if $v \neq 0$ and $\beta = ln(c)$ if $v = 0$. Here we will consider the case of $\beta = 0$, ie. $v \neq 0$ only so that the above relation becomes $F(x) = F(cx)\, H(x)$, $\forall\ x\in R$, and some $c\in(0,1)\cup(1,\infty)$. If the relation holds for every $c\in(0,1)\cup(1,\infty)$, then F is max-selfdecomposable.

Satheesh and Sandhya (2005) showed that the sequence $\{X_n\}$ of r.vs generates a marginally stationary AR(1) scheme (6.1.1) if and only if X_n is semi-selfdecomposable(b) and also discussed the integer-valued analogue of it. Extending the discussion to the maximum scheme Satheesh and Sandhya (2006*b*) showed that $\{X_n\}$ generates a marginally stationary max-AR(1) scheme if and only if X_n is max-semi-selfdecomposable(c). The structure here is:

$$X_n = bX_{n-1} \vee \varepsilon_n,\ \forall\ n>0 \text{ integer and some } b>0, \text{ and } c=1/b. \tag{6.2.1}$$

They then modified the max-AR(1) scheme to:

$$\left.\begin{aligned} X_n &= bX_{n-1}, && \text{with probability } p \\ &= bX_{n-1} \vee \varepsilon_n, && \text{with probability } (1-p). \end{aligned}\right\} \tag{6.2.2}$$

Subsequently Satheesh and Sandhya (2006b) showed, assuming $X_0 \stackrel{d}{=} \varepsilon_1$, that a sequence $\{X_n\}$ of r.vs generates the marginally stationary max-AR(1) scheme (6.2.2), if and only if X_n is exponential max-semi-stable(a,c), a=1/p, c=1/b having d.f of the form:

$$F(x) = \frac{1}{1+\psi(x)} = \frac{1}{1+(1/p)\psi(cx)},\ \forall x \in R.$$

where $\psi(x)$ is as in remark 6.1 with $c>1$ if $X_n>0$ and $c<1$ if $X_n<0\ \forall n$.

In the additive AR(1) and minimum AR(1) schemes analogous to (6.2.2), Jayakumar and Pillai(1992), Pillai (1991), Balakrishna and Jayakumar (1997) have similar results.

Whether additive, minimum or maximum; the structure in (6.2.2) essentially captures a geometric sum or a geometric extreme scheme, the geometric law being supported by the set {1, 2, 3, ...}. A generalization of the geometric law is the $H_1(m,k,1/k)$ law on {1, 1+k, 1+2k, ...}, that is described by its p.g.f

$$P(s) = \frac{s}{(m-(m-1)s^k)^{1/k}},\quad k>0 \text{ integer and } m>1.$$

The stability properties of the $H_1(m,k,1/k)$ law in the summation scheme were studied by Satheesh *et al.* (2002) while Satheesh and Nair (2002*a*, 2004) studied them for the minimum and maximum. (See section 1.3 in chapter 1). More distributional and divisibility properties, simulation and estimation problems of $H_1(m,k,1/k)$ law are given in chapter 2. The stationary solution to the generalized AR(1) model that we discuss here has the property of Harris-sum/ extreme stability.

Thus in the next section we consider a generalization of (6.1.1) where for each $n>0$, $X_n \stackrel{d}{=} \sum_{i=1}^{k} Y_{i,n}$ and $Y_{i,n}$, $i=1,2, \ldots, k$ are i.i.d, k being a fixed positive integer. The case of the maximum scheme instead of addition is also considered along with this. The discussion is then extended to the addition scheme for integer-valued X_n in section 6.3. An off-shoot of the development here is the description of semi-selfdecomposable laws on $\{0, h, 2h, 3h, \ldots\}$, $h>0$ integer. In section 6.4 we discuss the case for the minimum scheme.

6.3 Generalizations of the Additive and Maximum AR(1) Schemes

We now modify the model (6.1.1) as follows. We have k independent AR(1) sequences $\{Y_{i,n}\}$, $i = 1,2, \ldots, k$, where $Y_{i,n} = bY_{i,n-1} + \varepsilon_{i,n}$, $\forall\, n>0$ integer and some $0<b<1$, where $Y_{i,n}$, $i = 1,2, \ldots, k$ are identically distributed. Here, $\varepsilon_{i,n}$, $i = 1,2, \ldots, k$, $n>0$ integer are i.i.d. Hence (6.1.1) is described in terms of $\{Y_{i,n}\}$ as;

$$\sum_{i=1}^{k} Y_{i,n} = b\sum_{i=1}^{k} Y_{i,n-1} + \sum_{i=1}^{k} \varepsilon_{i,n} \text{ , } \forall\, n>0 \text{ integer and some } 0<b<1. \tag{6.3.1}$$

Assuming $\{Y_{i,n}\}$ to be marginally stationary (that is, $Y_{i,n} \overset{d}{=} Y_{i,n-1}\ \forall\ n$) their c.f s satisfy;

$$\phi_y^k(t) = \phi_y^k(bt)\ \phi_\varepsilon^k(t),\ \text{for some } 0<b<1. \text{ Or}$$

$$= \{\phi_y(bt)\ \phi_\varepsilon(t)\}^k. \text{ Hence;}$$

Theorem 6.1

A sequence $\{Y_{i,n}\}$ of r.vs defines the marginally stationary AR(1) scheme (6.3.1) if $Y_{i,n}$ is semi-selfdecomposable(b).

Corollary 6.1

A sequence $\{Y_{i,n}\}$ describes the marginally stationary AR(1) scheme (6.3.1) for all $0<b<1$ if $Y_{i,n}$ is selfdecomposable.

In the maximum scheme the model (6.3.1) reads as:

$$\bigvee_{i=1}^{k} Y_{i,n} = b\{\bigvee_{i=1}^{k} Y_{i,n-1}\} \vee \{\bigvee_{i=1}^{k} \varepsilon_{i,n}\},\ \forall\ n>0 \text{ integer and some } b>0. \tag{6.3.2}$$

Assuming $\{Y_{i,n}\}$ to be marginally stationary, in terms of their d.f s this equation reads:

$$F^k(x) = F^k(cx)\ G^k(x),\ c=1/b. \text{ Or}$$

$$= \{F(cx)\ G(x)\}^k, \text{ for some } c>0. \text{ Hence;}$$

Theorem 6.2

A sequence $\{Y_{i,n}\}$ of r.vs defines the marginally stationary max-AR(1) scheme (6.3.2) if $Y_{i,n}$ is max-semi-selfdecomposable(c), $c=1/b$. $\{Y_{i,n}\}$ describes (6.3.2) $\forall\ b>0$ if $Y_{i,n}$ is max-selfdecomposable.

Now modifying (6.3.1) we may have, $\forall$ $n>0$ integer and some $0<b<1$:

$$\left.\begin{aligned} \sum_{i=1}^{k} Y_{i,n} &= b\sum_{i=1}^{k} Y_{i,n-1} && \text{with probability } p \\ \sum_{i=1}^{k} Y_{i,n} &= b\sum_{i=1}^{k} Y_{i,n-1} + \sum_{i=1}^{k} \varepsilon_{i,n} && \text{with probability } (1\text{-}p) \end{aligned}\right\} \tag{6.3.3}$$

Assuming $Y_{i,0} \stackrel{d}{=} \varepsilon_{i,1}$ and marginal stationarity of $\{Y_{i,n}\}$ their c.f s satisfy; for n=1:

$\phi^k(t) = p\phi^k(bt) + (1\text{-}p)\phi^k(bt)\,\phi^k(t)$. That is;

$\phi^k(t) = p\phi^k(bt)/\{1\text{-}(1\text{-}p)\phi^k(bt)\}$. Or;

$\phi^k(t) = \phi^k(bt)/\{m\text{-}(m\text{-}1)\phi^k(bt)\}$, where $m = 1/p$. Hence;

$$\phi(t) = [\phi^k(bt)/\{m\text{-}(m\text{-}1)\phi^k(bt)\}]^{1/k}. \tag{*}$$

Equation (*) means that $Y_{i,1}$ is $H_1(m,k,1/k)$-sum stable. Hence by the characterization of generalized semi-α-Laplace laws, Satheesh *et al.* (2002, theorem 2.1) (theorem 1.1 in chapter 1), $Y_{i,1}$ is generalized semi-α-Laplace(m,b,k) with c.f $\phi(t)=\{1+\psi(t)\}^{-1/k}$ where $\psi(t)$ satisfies $\psi(t) = m\psi(bt)$, m and k being that in $H_1(m,k,1/k)$ and $mb^{\alpha}=1$ for some $\alpha\in(0,2]$, $m=1/p$ in the model (6.3.3). That is, $Y_{i,1}$, $Y_{i,o}$ and $\varepsilon_{i,1}$ are generalized semi-α-Laplace(m,b,k). Now, by the marginal stationarity of $\{Y_{i,n}\}$ we get $Y_{i,2}$ is generalized semi-α-Laplace(m,b,k) and recursively $Y_{i,3}$, $Y_{i,4}$, ... also follow generalized semi-α-Laplace(m,b,k). Such an induction argument over the index n results in:

Theorem 6.3

Under the assumption $Y_{i,0} \stackrel{d}{=} \varepsilon_{i,1}$, a sequence $\{Y_{i,n}\}$ of r.vs defines the marginally stationary AR(1) scheme (6.3.3) if and only if $Y_{i,n}$ is generalized semi-α-Laplace($1/p,b,k$).

The max-analogue of the characterization of generalized semi-α-Laplace laws mentioned above is the following. The proof follows on the same lines as that of theorem 2.1 in Satheesh *et al.* (2002).

Theorem 6.4

The i.i.d r.vs X_i are $H_1(m,k,1/k)$-max stable if and only if its d.f is $F(x) = \{1+\psi(x)\}^{-1/k}$, where $\psi(x)$ satisfies $\psi(x) = m\psi(cx)$, m and k being that in $H_1(m,k,1/k)$ and $mc^{\alpha}=1$ for some $c>0$ and $\alpha>0$.

Satheesh and Sandhya (2006*a*) had discussed φ-max-semi-stable laws for a LT φ. In this terminology the above d.f is called a gamma-max-semi-stable($m,c,1/k$) law. When $k=1$ we have the exponential max-semi-stable(m,c) model that characterized (6.2.2).

Now, modifying (6.3.2) further we have, $\forall n>0$ integer and some $b>0$:

$$\left.\begin{array}{ll} \bigvee_{i=1}^{k} Y_{i,n} = b\{\bigvee_{i=1}^{k} Y_{i,n-1}\} & \text{with probability } p \\ \bigvee_{i=1}^{k} Y_{i,n} = b\{\bigvee_{i=1}^{k} Y_{i,n-1}\} \vee \{\bigvee_{i=1}^{k} \varepsilon_{i,n}\} & \text{with probability } (1-p) \end{array}\right\} \qquad (6.3.4)$$

Proceeding as in the additive scheme we have the max-analogue of theorem 6.3.

Theorem 6.5

Under the assumption $Y_{i,0} \stackrel{d}{=} \varepsilon_{i,1}$, a sequence $\{Y_{i,n}\}$ of r.vs defines the marginally stationary max-AR(1) scheme (6.3.4), if and only if $Y_{i,n}$ is gamma-max-semi-stable($1/p,1/b,1/k$).

Suppose we require (6.3.3) to be satisfied for all $b\in(0,1)$, then:

Corollary 6.2

Assuming $Y_{i,0}\overset{d}{=}\varepsilon_{i,1}$, a sequence $\{Y_{i,n}\}$ of r.vs defines the marginally stationary AR(1) scheme (6.3.3) for all $b\in(0,1)$, if and only if $Y_{i,n}$ is generalized Linnik (generalized α-Laplace) with c.f $\{1+\eta|t|^{\alpha}\}^{-1/k}$, $k>0$ integer, $\alpha\in(0,2]$ and $\eta>0$. Suppose we further demand $Y_{i,n}$ to have finite variance, then (6.3.3) characterizes the generalized Laplace law with c.f $\{1+\eta\, t^2\}^{-1/k}$, $k>0$ integer and $\eta>0$.

Suppose we need $Y_{i,n}>0$, then by the Harris-sum stability characterization of generalized semi-*ML* laws, Satheesh *et al.*(2002, corollary 2.1) (theorem 1.3 in chapter 1), we have:

Theorem 6.6

Under the assumption $Y_{i,0}\overset{d}{=}\varepsilon_{i,1}$, a sequence $\{Y_{i,n}>0\}$ of r.vs defines the marginally stationary AR(1) scheme (6.3.3), if and only if $Y_{i,n}$ is generalized semi-$ML(1/p,b,k)$.

Under the additional assumptions as in corollary 6.2 we have:

Corollary 6.3

Assuming $Y_{i,0}\overset{d}{=}\varepsilon_{i,1}$, a sequence $\{Y_{i,n}>0\}$ of r.vs defines the marginally stationary AR(1) scheme (6.3.3) for all $b\in(0,1)$, if and only if $Y_{i,n}$ is generalized *ML* with LT $\{1+\eta\,\lambda^{\alpha}\}^{-1/k}$, $k>0$ integer, $\alpha\in(0,1]$ and $\eta>0$. Further demanding $Y_{i,n}$ to have finite mean the *Gamma*$(1/k,\ 1/\eta)$ law with LT $\{1+\eta\lambda\}^{-1/k}$, $k>0$ integer and $\eta>0$ is characterized by (6.3.3).

6.4 Generalization of the Discrete AR(1) Scheme and Semi-selfdecomposable Laws with Gaps

To discuss (6.3.3) for discrete r.vs we need the description of integer-valued r.vs of the same type in Satheesh and Nair (2002*b*) that we present here as a remark.

Remark 6.2

If $\varphi(s)$ is a LT, then $P(s) = \varphi(1\text{-}s)$, $0<s<1$ is a p.g.f. If φ_1 and φ_2 are LTs, then the p.g.fs $P_1(s) = \varphi_1(1\text{-}s)$ and $P_2(s) = \varphi_2(1\text{-}s)$ are of the same type if and only if $\varphi_1(1\text{-}s) = \varphi_2(c(1\text{-}s))$, for all $0<s<1$ and some $c>0$. Two p.g.fs $P_1(s)$ and $P_2(s)$ are of the same type if and only if P_1 is a P_2 compounded Bernoulli law. Thus in the set-up of integer-valued r.vs the equivalent of r.vs of the same type is obtained by replacing bX by $b\circ X = \sum_{i=1}^{X} Z_i$, where $\{Z_i\}$ are i.i.d Bernoulli(b) r.vs independent of X with $P\{Z_i{=}0\} = 1{-}b$.

Consequently (6.3.3) becomes:

$$\left.\begin{aligned} \sum_{i=1}^{k} Y_{i,n} &= b\circ\sum_{i=1}^{k} Y_{i,n-1} && \text{with probability } p \\ \sum_{i=1}^{k} Y_{i,n} &= b\circ\sum_{i=1}^{k} Y_{i,n-1} + \sum_{i=1}^{k}\varepsilon_{i,n} && \text{with probability } (1\text{-}p) \end{aligned}\right\} \qquad (6.3.3a)$$

Discrete generalized semi-$ML(a,b,k)$ laws with p.g.f $P(s) = \{1+\psi(1\text{-}s)\}^{-1/k}$, $0<s<1$, where $\psi(1\text{-}s)$ satisfies $\psi(1\text{-}s) = a\psi(b(1\text{-}s))$, $ab^{\alpha} = 1$ for some $\alpha\in(0,1]$, were characterized in theorem 3.3 of Satheesh *et al.* (2002) (theorem 1.4 in chapter 1) by the property of Harris-sum stability of distributions on $\{0, 1, 2, \ldots\}$. Hence when $\{Y_{i,n}\}$ are integer-valued we have:

Theorem 6.7

Under the assumption $Y_{i,0} \overset{d}{=} \varepsilon_{i,1}$, a sequence $\{Y_{i,n}\}$ of non-negative integer-valued r.vs defines the marginally stationary AR(1) scheme (6.3.3a), if and only if $Y_{i,n}$ is discrete generalized semi-$ML(1/p,b,k)$.

Now we characterize discrete generalized ML and negative binomial laws respectively by additional conditions as in corollary 6.2 on the scheme.

Corollary 6.4

Assuming $Y_{i,0} \overset{d}{=} \varepsilon_{i,1}$, a sequence $\{Y_{i,n}\}$ of non-negative integer-valued r.vs defines the marginally stationary AR(1) scheme (6.3.3a) for all $b \in (0,1)$, if and only if $Y_{i,n}$ is discrete generalized ML with p.g.f $\{1+ \eta(1-s)^{\alpha}\}^{-1/k}$, $k>0$ integer, $\alpha \in (0,1]$ and $\eta>0$. Further demanding $Y_{i,n}$ to have finite mean the $NB(\eta, 1/k)$ law with p.g.f $\{1+ \eta(1-s)\}^{-1/k}$, $k>0$ integer and $\eta>0$ is characterized by (6.3.3a).

We next characterize distributions that have gaps in its support by Harris-sum stability as follows. Certain implications of having and not having gaps in the support were investigated in Satheesh (2004*a*).

Theorem 6.8

A distribution on $\{0, h, 2h, 3h, \ldots\}$, $h>0$ integer, is $H_1(m,k,1/k)$-sum stable if and only if its p.g.f is $P(s) = \{1+\psi(1-s^h)\}^{-1/k}$, $\psi(1-s^h)$ satisfying $\psi(1-s^h) = m\,\psi(b(1-s^h))$, $mb^{\alpha} = 1$ for some $\alpha \in (0,1]$.

Proof.

The assertion follows from lemma 4.1 and theorem 3.3 in Satheesh *et al.* (2002)

Now in terms of AR(1) models we have:

Theorem 6.9

Under the assumption $Y_{i,0} \overset{d}{=} \varepsilon_{i,1}$, a sequence $\{Y_{i,n}\}$ of r.vs on $\{0, h, 2h, 3h, \ldots\}$, $h>0$ integer, defines the marginally stationary AR(1) scheme (6.3.3a), if and only if the p.g.f of $Y_{i,n}$ is $P(s) = \{1+\psi(1-s^h)\}^{-1/k}$, $k>0$ integer, $0<s<1$, $p\psi(1-s^h) = \psi(b(1-s^h))$, $b^{\alpha} = p$, $\alpha\in(0,1]$.

Under the additional conditions on the scheme as in corollary 6.2 we have:

Corollary 6.5

Assuming $Y_{i,0} \overset{d}{=} \varepsilon_{i,1}$, a sequence $\{Y_{i,n}\}$ of r.vs on $\{0, h, 2h, 3h, \ldots\}$, $h>0$ integer, defines the marginally stationary AR(1) scheme (6.3.3a) $\forall b\in(0,1)$, if and only if the p.g.f of $Y_{i,n}$ is $\{1+\eta(1-s^h)^{\alpha}\}^{-1/k}$, $k>0$ integer, $\alpha\in(0,1]$ and $\eta>0$. Further demanding $Y_{i,n}$ to have finite mean the model (6.3.3a) characterizes the p.g.f of $Y_{i,n}$ as $\{1+\eta(1-s^h)\}^{-1/k}$, $k>0$ integer and $\eta>0$.

Remark 6.3

Recently Satheesh and Sandhya (2005) have described semi-selfdecomposable p.g.fs. Incidentally the above discussion suggests possible extension of the notion of selfdecomposable and semi-selfdecomposable laws to distributions on$\{0, h, 2h, 3h, \ldots.\}$, $h>0$ integer.

Definition 6.5 An integer-valued distribution on $\{0, h, 2h, 3h, \ldots\}$, $h>0$ integer, with p.g.f $P(s^h)$ is discrete semi-selfdecomposable (b) if for some $0<b<1$, there exists another p.g.f $P_o(s^h)$ that is i.d such that

$$P(s^h) = P(1-b+b\,s^h)\,P_o(s^h),\ \forall\ s\in(0,1).$$

The distribution is selfdecomposable if the above relation holds good for all $0<b<1$. If the p.g.fs are described in terms of LTs as in remark 6.2 above, then equivalently we have:

A p.g.f $P(s^h) = \varphi(1-s^h)$ is discrete semi-selfdecomposable (b) if for some $0<b<1$, there exists another p.g.f $P_o(s^h) = \varphi_o(1-s^h)$ such that

$$\varphi(1-s^h) = \varphi[b(1-s^h)]\ \varphi_o(1-s^h).$$

Example 6.1

By the $H_1(m,k,1/k)$-sum stability of the p.g.f $P(s) = \{1+\psi(1-s^h)\}^{-1/k}$ (see theorem 6.8) we get;

$$\frac{1}{\{1+\psi(1-s^h)\}^{1/k}} = \frac{1}{[1+\psi\{b(1-s^h)\}]^{1/k}}\ \frac{1}{[m-(m-1)/\{1+\psi[b(1-s^h)]\}]^{1/k}}.$$

Here the second factor itself is a p.g.f being the Harris-sum of the distribution with p.g.f $[1+\psi\{b(1-s^h)\}]^{-1/k}$, this Harris law being supported on $\{0, h, 2h, 3h, \ldots\}$. Further it is i.d since $H_0(m,k,1/k)$ is i.d. Hence the p.g.f $\{1+\psi(1-s^h)\}^{-1/k}$ is discrete semi-selfdecomposable (b) on $\{0, h, 2h, 3h, \ldots\}$, $h>0$ integer.

Theorem 6.10

A sequence $\{X_n\}$ of integer-valued r.vs defines a marginally stationary AR(1) scheme on $\{0, h, 2h, 3h, \ldots\}$, $h>0$ integer, with $0<b<1$ if X_n is discrete semi-selfdecomposable (b) on $\{0, h, 2h, 3h, \ldots\}$, $h>0$ integer.

Proof.

The proof follows on lines similar to that of theorem1 in Satheesh and Sandhya (2005).

6.5 Generalizations of the Minimum AR(1) Schemes

To extend the model to the minimum scheme we need describe min- semi-self dec(c) laws. Here we do not attempt a detailed study of this class and also restrict our discussion to the support $[0,\infty)$. More on this class is in Satheesh and Sandhya (2006*c*).

Definition 6.6 A non-degenerate d.f F with survival function (s.f) S is min-semi-selfdecomposable(c) if for some $0<c<1$ there is another s.f T such that

$$S(x) = S(cx)\, T(x),\ x>0.$$

If this is true for every $0<c<1$, then F is min-selfdecomposable.

The following results also give certain examples in these classes.

Theorem 6.11

Generalized semi-Pareto$(p,\alpha,1/k)$ law with s.f $[1+\psi(x)]^{-1/k}$, $x>0$, $p\psi(x) = \psi(p^{1/\alpha}x)$, $\forall x>0$, some $0<p<1$, and $\alpha>0$ is min-semi-selfdecomposable$(p^{1/\alpha})$.

Proof.

By the $H_1(m,k,1/k)$-min stability of the generalized semi-Pareto$(p,\alpha,1/k)$ law, (Satheesh and Nair (2002*a*), quoted below as theorem 6.13), it follows that we can write:

$$\{1+\psi(x)\}^{-1/k} = \{1+\psi(bx)\}^{-1/k}/\{m-(m-1)\{1+\psi(bx)\}^{-1}\}^{1/k},\quad p=1/m,\quad b=p^{1/\alpha}.$$

Now the assertion follows as done in example 6.1.

A more general approach to proving that a distribution is semi-selfdecomposable or max/ min semi-selfdecomposable is from the angle of mixtures as done in Satheesh and Sandhya (2005, 2006*b*, *c*).

Corollary 6.6

Semi-Pareto(p,α) laws of Pillai (1991) are min-semi-selfdecomposable$(p^{1/\alpha})$, which follows by its geometric-min stability, the geometric law being on $\{1, 2, \dots\}$.

Corollary 6.7

Pareto laws with s.f $1/\{1+x\}^{\alpha}$, $\alpha>0$, are min-selfdecomposable.

Now we modify the scheme (6.3.2) to the minimum structure as:

$$\bigwedge_{i=1}^{k} Y_{i,n} = b\{\bigwedge_{i=1}^{k} Y_{i,n-1}\} \wedge \{\bigwedge_{i=1}^{k} \varepsilon_{i,n}\}\quad \forall n>0 \text{ integer and some } b>1. \tag{6.5.1}$$

Assuming $\{Y_{i,n}\}$ to be marginally stationary, in terms of the s.f s S of $Y_{i,n}$ and T of $\varepsilon_{i,n}$ this equation reads:

$$S^k(x) = S^k(cx)\,T^k(x),\ c=1/b. \text{ Or}$$

$$= \{S(cx)\,T(x)\}^k, \text{ for some } 0<c<1. \text{ Hence;}$$

Theorem 6.12

A sequence $\{Y_{i,n}\}$ of non-negative r.vs defines a marginally stationary min-AR(1) scheme as in (6.5.1) if $Y_{i,n}$ is min-semi-selfdecomposable(c), c=1/b. $\{Y_{i,n}\}$ describes the structure (6.5.1) $\forall b>1$ if $Y_{i,n}$ is min-selfdecomposable.

The following is a restatement of proposition 3 in Satheesh and Nair (2002a).

Theorem 6.13

The i.i.d r.vs X_i are $H_1(m,k,1/k)$-min stable if and only if it is generalized semi-Pareto($p,\alpha,1/k$) with d.f $F(x) = 1-\{1+\psi(x)\}^{-1/k}$, $x>0$, $p\psi(x) = \psi(p^{1/\alpha}x)$, $\forall x>0$, some $0<p<1$, $p = 1/m$, and $\alpha>0$.

Now, modifying (6.5.1) further we have, $\forall n>0$ integer and some $b>1$:

$$\left.\begin{aligned} \bigwedge_{i=1}^{k} Y_{i,n} &= b\{\bigwedge_{i=1}^{k} Y_{i,n-1}\} && \text{with probability } p \\ \bigwedge_{i=1}^{k} Y_{i,n} &= b\{\bigwedge_{i=1}^{k} Y_{i,n-1}\} \wedge \{\bigwedge_{i=1}^{k} \varepsilon_{i,n}\} && \text{with probability } (1\text{-}p) \end{aligned}\right\} \tag{6.5.2}$$

Assuming marginal stationarity of $Y_{i,n}$ and $Y_{i,0} \overset{d}{=} \varepsilon_{i,1}$, and proceeding as in the additive scheme we have the following min-analogue of theorem 6.3.

Theorem 6.14

Under the assumption $Y_{i,0} \overset{d}{=} \varepsilon_{i,1}$, a sequence$\{Y_{i,n}\}$of non-negative r.vs defines the marginally stationary min-AR(1) scheme (6.5.2), if and only if $Y_{i,n}$ is generalized semi-Pareto($p,\alpha,1/k$) and $b = p^{-1/\alpha}$.

Notice that the schemes (6.3.2) and (6.3.4) include $b<1$ and the explosive case $b>1$ as well. This is also in tune with the range of the parameter c in max-semi-selfdecomposable and max-semi-stable laws. Thus we have potentially useful generalized AR(1) schemes in the additive, maximum and minimum structures.

APPLICATION AND GENERALIZATION OF HARRIS DISTRIBUTION

Sherly Sebastian “Harris family of discrete distributions and processes ” Thesis. Department of Statistics , University of Calicut, 2007

CHAPTER 7

APPLICATION AND GENERALIZATION OF HARRIS DISTRIBUTION

7.1 Introduction

A number of distributions have been devised for count data in which the variance is significantly larger than the mean (over-dispersed data). Harris distribution, generalized Poisson distribution and extended geometric distribution are three such distributions. In section 7.2 we consider the fitting of these three distributions to a telephone data. We found that Harris distribution is a good fit to the data while the generalized Poisson model and extended geometric model are not. Also, Harris distribution is fitted using two estimates of m, namely unbiased estimate and Bayes estimate.

Section 7.3 contains testing of hypotheses concerning the population mean of the observed telephone data. Values of the power functions and the corresponding power curves are provided. The probability plots of the test statistic are also given.

In section 7.4 a generalized Harris distribution is introduced. This section contains a brief discussion of its distributional properties and a characterization of this distribution. Also by modifying model 1 of chapter 5, we develop a stochastic model where the variable under consideration follows a generalized Harris distribution.

7.2 Fitting of Harris Distribution to an Observed Data

In this section as an illustration of fit to data, we have used a telephone data from Bharat

Sanchar Nigam Limited (BSNL). We study and compare models of approximation to the frequency data given in the columns headed "Observed" of the following tables.

When the link is obtained one call is recorded that can be of duration up to k seconds and subsequently additional calls are recorded for each k seconds. We assume that the minimum time taken for a call is one second. Let X denote the minimum time in seconds taken by a telephone caller for $r + 1$ calls; $r = 0, 1, 2, \ldots$ are the additional calls. If the duration of the call is between k and $2k+1$ seconds the count is two and in general, if the duration of the call is between $(n-1)k$ and $nk+1$ seconds the count is n. Thus the r.v X we are considering here takes values $1, 1+k, 1+2k, \ldots$. The criteria for fixing k is the air distance and the nature of telephone circles between the caller and the called party. The policy of BSNL is that k=30 for distance more than 500 kilometers, k= 60 for distances from 100 to 500 kilometers and k=180 for distance less than 100 kilometers.

The data is classified into three sets corresponding to $k = 30$, 60 and 180. Denote by $\bar{x}$ and s^2 the mean and variance of a sample of size n. The mean and variance of the three sets are (5.167, 387.639), (4.864, 585.072) and (11.286, 8707.004) respectively. Clearly the data is over-dispersed in all the three cases. As we know Harris distribution, generalized Poisson distribution and extended geometric distribution are all over-dispersed distributions, we tested the goodness of fit corresponding to these three distributions. Again, Harris distribution is fitted using two estimates of m, namely unbiased estimate and Bayes estimate. The columns labeled "$H_1(m,k,1/k)$" of the Tables 7.1, 7.2, 7.3 contain the corresponding frequencies calculated by using the Harris distribution with parameter $m = \bar{x}$, an unbiased estimate. The columns labeled

"$H_1^*(m,k,1/k)$" contain the expected frequencies calculated by using the Harris distribution with parameter $m = \dfrac{n\bar{x}+2k}{n+k}$, the Bayes estimate. The columns labeled "$GP(\lambda,k)$" contain the expected frequencies calculated by using the generalized Poisson distribution with parameter $\lambda=\bar{x}$ and the columns labeled "$EG(p, k)$" contain the expected frequencies calculated by using the extended geometric distribution with parameter $p=1/(1+\bar{x})$.

A comparison of the expected values present in the columns $H_1(m,k,1/k)$, $H_1^*(m,k,1/k)$, $GP(\lambda,k)$ and $EG(p, k)$ shows that $H_1(m,k,1/k)$ model fits much better the observed frequencies than the generalized Poisson model and the extended geometric model. In Table 7.1 the value of χ^2 for $H_1(m,k,1/k)$, for instance, is 0.502, which gives the p-value of 0.479 for the tested model. Thus, approximation of the observed data by the $H_1(m,k,1/k)$ model is apparently acceptable, while the generalized Poisson model and the extended geometric model are not. Therefore, Harris model may offer better goodness of fit to statistical data than generalized Poisson model and the extended geometric model.

Again, in Tables 7.1 and 7.2 the p-values of the chi-square for the goodness of fit test for the Harris model $H_1(m,k,1/k)$ with the unbiased estimate are greater than the p-values of the Harris model $H_1^*(m,k,1/k)$ with the Bayes estimate. But in Table 7.3 the p-value corresponding to Harris model with Bayes estimate is greater than that of the Harris model with unbiased estimate. Hence we conclude that for small values of k, the unbiased estimate is better than the Bayes estimate and for large values of k Bayes estimate turns to be better.

7.3 Testing of Population Mean of the Observed Data

It is very important to test whether the mean of the Harris population is greater than some specified value. We formed the null hypothesis of the form H_0: $m \leq m_0$ against H_1 : $m > m_0$. Following the UMP test developed in chapter 3, we carried out testing of H_0 against H_1 in three cases with gaps of $k = 30$, 60 and 180.

(A) In the case of phone data with gap, $k = 30$, we set up the hypothesis

$$H_0 : m \leq 3.75 \text{ against } H_1 : m > 3.75$$

Here $m_0 = 3.75$, $k = 30$, $n = 360$ and let the size of the test, $\alpha = .05$

The UMP size α test developed in this case is

$$\phi(x) = \begin{cases} 1 & \text{when } \sum x_i > 1950 \\ 0.537 & \text{when } \sum x_i = 1950 \\ 0 & \text{when } \sum x_i < 1950 \end{cases}$$

where $T = \sum_{i=1}^{n} x_i \sim H_{360}(3.75, 30, 360/30)$ under H_0.

Since $\sum x_i = 1860 < 1950$, we accept H_0 with size, $\alpha = 0.05$. Hence we conclude that the phone data with gap $k = 30$ is taken from a Harris population with mean, $m \leq 3.75$.

Now from (3.4.4) it follows that the power function

$$A(m) = \sum_{t=360}^{1950} \frac{\Gamma(t/30)}{\Gamma(360/30)\Gamma(1+((t-360)/30))}\left(\frac{1}{m}\right)^{360/30}\left(1-\frac{1}{m}\right)^{(t-360)/30}$$

$$-0.537\frac{\Gamma(1950/30)}{\Gamma(360/30)\Gamma(1+((1950-360)/30))}\left(\frac{1}{m}\right)^{360/30}\left(1-\frac{1}{m}\right)^{(1950-360)/30}$$

(B) Consider the problem of testing $H_0 : m \leq 3.5$ against $H_1 : m > 3.5$ in the case of phone data with gap, $k = 60$.

Here $m_0 = 3.5$, $k = 60$, $n = 528$ and $\alpha = .05$

The UMP size α test developed in this case is

$$\phi(x) = \begin{cases} 1 & \text{when } \sum x_i > 2808 \\ 0.557 & \text{when } \sum x_i = 2808 \\ 0 & \text{when } \sum x_i < 2808 \end{cases}$$

where $T = \sum_{i=1}^{n} x_i \sim H_{528}(3.5, 60, 528/60)$ under H_0.

Since $\sum x_i = 2568 < 2808$, we accept H_0 with size, $\alpha = 0.05$. Hence we conclude that the phone data with gap $k = 60$ is taken from a Harris population with mean, $m \leq 3.5$.

The power function is given by

$$B(m) = \sum_{t=528}^{2808} \frac{\Gamma(t/60)}{\Gamma(528/60)\Gamma(1+((t-528)/60))}\left(\frac{1}{m}\right)^{528/60}\left(1-\frac{1}{m}\right)^{(t-528)/60}$$

$$-0.557\frac{\Gamma(2808/60)}{\Gamma(528/60)\Gamma(1+((2808-528)/60))}\left(\frac{1}{m}\right)^{528/60}\left(1-\frac{1}{m}\right)^{(2808-528)/60}$$

(C) In the case of phone data with gap, $k = 180$, we set up the hypothesis

$H_0 : m \leq 6.5$ against $H_1 : m > 6.5$

Here $m_0 = 6.5$, $k = 180$, $n = 875$ and let $\alpha = .05$

The UMP size α test developed in this case is

$$\phi(x) = \begin{cases} 1 & \text{when } \sum x_i > 10060 \\ 0.231 & \text{when } \sum x_i = 10060 \\ 0 & \text{when } \sum x_i < 10060 \end{cases}$$

where $T = \sum_{i=1}^{n} x_i \sim H_{875}(6.5, 180, 875/180)$ under H_0.

Since $\sum x_i = 9875 < 10060$, we accept H_0 with size, $\alpha = 0.05$. Hence we conclude that the phone data with gap $k = 180$ is taken from a Harris population having mean, $m \leq 6.5$.

Now, the power function

$$C(m) = \sum_{t=875}^{10060} \frac{\Gamma(t/180)}{\Gamma(875/180)\Gamma(1+((t-875)/180))} \left(\frac{1}{m}\right)^{875/180} \left(1-\frac{1}{m}\right)^{(t-875)/180}$$

$$-0.231 \frac{\Gamma(10060/180)}{\Gamma(875/180)\Gamma(1+((10060-875)/180))} \left(\frac{1}{m}\right)^{875/180} \left(1-\frac{1}{m}\right)^{(10060-875)/180}$$

The values of $A(m)$, $B(m)$ and $C(m)$ for various values of m are given in Table 7.4, Table 7.5 and Table 7.6 and the corresponding power curves are given in Figure 7.1, Figure 7.2 and Figure 7.3.

Probability plots of sampling distribution of $T = \sum_{i=1}^{n} x_i$ corresponding to tests (A), (B) and (C) are given in Figures 7.4, 7.5 and 7.6 respectively.

7.4 Generalization of Harris Distribution

We have generalized the Harris distribution as follows.

Let $\{P_{a+nk}\}$, $n = 0, 1, 2, \ldots$ represent the p.m.f of a r.v X taking on values $a,\ a+k,\ a+2k, \ldots$. We say X has a generalized Harris distribution (GHD) if

$$P_{a+nk} = P(X = a+nk) = \binom{\beta+n-1}{n} \left(\frac{1}{m}\right)^{\beta} \left(1-\frac{1}{m}\right)^{n}, \quad n = 0, 1, 2, \ldots \tag{7.4.1}$$

where $a \geq 0$, $k > 0$ are integers, $\beta > 0$ and $m > 1$.

We write $X \sim H_a(m,k,\beta)$ for X following GHD with parameters a, k, m and β given by the p.m.f (7.4.1). In the notation the suffix a indicates that the support of the distribution starts from a, k implies that the atoms of the distribution are k integers apart, m determines the probabilities and β is the power parameter. The probabilities are concentrated on the points a, $a+k$, $a+2k$, … . Also these probabilities coincide with the probabilities of the negative binomial distribution $NB(1/m, \beta)$) defined on $\{0, 1, 2, \dots\}$.

Now we have the following cases.

If $a = 1$ and $\beta = 1/k$, then a GHD, $H_a(m,k,\beta)$ reduces to the HD, $H_1(m,k,1/k)$.

If $a = 0$ and $k = 1$, then a GHD, $H_a(m,k,\beta)$ reduces to the NBD, $NB(1/m, \beta)$ with support $\{0, 1, 2, \dots\}$.

If $a = 0$, $k = 1$ and $\beta = 1$, then the GHD, $H_a(m,k,\beta)$ reduces to the geometric distribution $Geo(1/m)$ with the parameter $1/m$ and support $\{0, 1, 2, \dots\}$.

If $a = 1$, $k = 1$ and $\beta = 1$, then the GHD, $H_a(m,k,\beta)$ reduces to the decapitated geometric distribution with the parameter $1/m$ and support $\{1, 2, \dots\}$.

Distributional properties of generalized Harris distribution

Proceeding as given in chapter 2, the following distributional properties of GHD is obtained. The mean, variance, p.g.f, etc. of the GHD can be easily determined.

Mean, $E(X) = a + (m-1)k\beta$

Variance, $V(X) = m(m-1)k^2\beta$ and

p.g.f, $$P_X(s) = \frac{s^a}{\left(m-(m-1)s^k\right)^\beta}$$

The recurrence relation for probabilities is given by

$$P(X = a+(n+1)k) = \frac{\beta+n}{n+1}\, q\, P(X = a+nk), \quad n = 0, 1, 2, \dots$$

Distribution function

The distribution function of a GHD, $H_a(m,k,\beta)$ with p.m.f as in (7.4.1) is given by;

$$F(a+nk) = I_{(1/m)}(\beta, n+1), \quad n = 0, 1, 2, \ldots$$

where $I_p(a,b) = B_p(a,b)/B(a,b)$ is the incomplete beta function ratio.

Additive property

If $X_1, X_2, \ldots, X_n$ are n independent r.vs following GHD, $H_{a_i}(m,k,\beta_i)$, $i = 1, 2, \ldots, n$,

then, $T = \sum_{i=1}^{n} X_i$ follows $H_A(m,k,B)$ where $A = \sum_{i=1}^{n} a_i$ and $B = \sum_{i=1}^{n} \beta_i$.

The p.g.f of X_i, $\quad P_{X_i}(s) = \dfrac{s^{a_i}}{(m-(m-1)s^k)^{\beta_i}}, \quad i = 1, 2, \ldots, n.$

Then the p.g.f of $T = \sum_{i=1}^{n} X_i$,

$$P_T(s) = \frac{s^A}{(m-(m-1)s^k)^B}, \quad A = \sum_{i=1}^{n} a_i \text{ and } B = \sum_{i=1}^{n} \beta_i$$

proves the additive property.

Theorem 7.1

Let X and Y be i.i.d $H_a(m,k,\beta)$ r.vs. Then the conditional distribution of X given $X+Y$ is;

$$P(X = a+nk / X+Y = 2a+tk) = \frac{\binom{\beta+n-1}{n}\binom{\beta+t-n-1}{t-n}}{\binom{2\beta+t-1}{t}}, \quad n = 0,1,2,\ldots, t; \quad t = 0,1,2,\ldots$$

Proof.

Proof is similar to the proof given in theorem 2.2.

Remark 7.1 If $\beta = 1$, then

$$P(X = a + nk / X + Y = 2a + tk) = 1/(1+t), \quad n = 0, 1, 2, \ldots t; \quad t = 0, 1, 2, \cdots$$

and hence the conditional distribution is uniform.

Remark 7.2 If $a = k = \beta = 1$, that is, if X and Y are i.i.d geometric r.vs, then

$$P(X = n+1 / X + Y = 2+t) = 1/(1+t), \quad n = 0, 1, 2, \ldots t; \quad t = 0, 1, 2, \cdots$$

which coincides with the result given in Rohatgi (1976, p.189)

Result 7.1 GHD, $H_a(p,k,\beta)$ tends to the distribution of a linear function of a Poisson variable when $p \to 1$ and $\beta \to \infty$ such that $\beta(1-p) = \lambda$ remains fixed.

Let $U \sim NB(p, \beta)$ and $V \sim P(\lambda)$. Then the p.g.f s of U and V are respectively

$$P_U(s) = (p/(1-qs))^\beta \qquad \text{and} \qquad P_V(s) = e^{-\lambda(1-s)}.$$

Rohatgi (1976, p. 201) proves that a NBD, $NB(p, \beta)$ with parameters p and β tends to the Poisson distribution $P(\lambda)$ with parameter λ as $p \to 1$ and $\beta \to \infty$ in such a way that $\beta(1-p) = \lambda$ remains fixed.

Under the above conditions $P_U(s) \to P_V(s)$. *i.e.* $(p/(1-qs))^\beta \to e^{-\lambda(1-s)}$ and hence $(p/(1-qs^k))^\beta s^a \to e^{-\lambda(1-s^k)} s^a$ as $p \to 1$ and $\beta \to \infty$ such that $\beta(1-p) = \lambda$ remains fixed.

Here $(p/(1-qs^k))^\beta s^a$ is the p.g.f of the GHD, $H_a(p,k,\beta)$ and $e^{-\lambda(1-s^k)} s^a$ is the p.g.f of the distribution of a linear function of a Poisson variable with parameter λ. Hence GHD tends to the Poisson distribution $P(\lambda)$ defined on $\{a, a+k, a+2k, \ldots\}$ as $p \to 1$ and $\beta \to \infty$ in such a way that $\beta(1-p) = \lambda$ remains fixed.

Now, parallel to the characterization of Harris distribution given in theorem 2.3 of chapter 2 we obtain the following characterization of GHD.

Theorem 7.2 Let X be a r.v with $P(X = a + nk) = b_{a+nk}$, $n = 0, 1, 2, \ldots$; $a \geq 0$ and $k > 0$ are integers and let $E(X) = \mu$ be finite. Then X has GHD if and only if

$$\frac{d}{d\mu}P(X > a+nk) = \frac{n+\beta}{k\beta+\mu-a}b_{a+nk}, \quad n = 0,1,2,\ldots; \; a \geq 0 \text{ and } k>0 \text{ are integers and } \beta > 0.$$

Proof.

Follows as the proof of theorem 2.3.

Generalized Harris distribution as a mixture

The GHD arises as a gamma mixture of generalized Poisson distribution.

Suppose that Y has a conditional Poisson distribution with parameter λ so that

$$P(Y = n \mid \lambda) = \frac{e^{-\lambda}\lambda^n}{n!}, \quad n = 0,1,2,\ldots$$

where λ itself is assumed to have a gamma distribution with density

$$g(\lambda) = \frac{b^\beta}{\Gamma\beta}e^{-b\lambda}\lambda^{\beta-1}, \quad b>0, \; \beta>0 \text{ and } \lambda>0$$

Now consider a r.v $X=a+kY$ where $a \geq 0$, $k>0$ are integers. The conditional distribution of X given λ is

$$P(X = a+nk \mid \lambda) = \frac{e^{-\lambda}\lambda^n}{n!}, \quad n = 0,1,2,\ldots$$

Then the unconditional distribution of X is given by

$$P(X = a+nk) = \int_0^\infty \frac{e^{-\lambda}\lambda^n}{n!}\frac{b^\beta}{\Gamma\beta}e^{-b\lambda}\lambda^{\beta-1}d\lambda$$

$$= \binom{\beta+n-1}{n}\left(\frac{1}{m}\right)^\beta\left(1-\frac{1}{m}\right)^n, \quad n = 0,1,2,\ldots$$

where $a\geq 0$ and $k>0$ are integers, $m = (1+b)/b > 1$ and $\beta > 0$. This is the p.m.f of the GHD.

Stability

Now, $H_0(m,k,\beta)$ p.g.f,

$$\frac{1}{\left(m-(m-1)s^k\right)^{\beta}}=\frac{1}{\left(\left(m-(m-1)s^k\right)^{\beta/n}\right)^{n}}$$

Each component has a $H_0(m,k,\beta/n)$ law. Hence $H_0(m,k,\beta)$ is stable in the summation scheme.

Connection with power series distribution

GHD belongs to the family of generalized power series distributions (GPSD) and using this result its properties can be reviewed.

Putting $x=a+nk$ in (7.4.1), the p.m.f of GHD can be written as

$$P(X=x)=\binom{(k\beta+x-a-k)/k}{(x-a)/k}\left(\frac{1}{m}\right)^{\beta}\left(1-\frac{1}{m}\right)^{\frac{x-a}{k}}, \quad x=a,\ a+k,\ a+2k,\ \ldots$$

where $a\geq 0$, $k>0$ are integers, $\beta>0$ and $m>1$.

Taking $a(x)=\binom{(k\beta+x-a-k)/k}{(x-a)/k}$ for $x \in S=\{a,\ a+k,\ a+2k,\ldots\}$ and $a(x)=0$ otherwise;

$\theta=\left(1-\frac{1}{m}\right)^{1/k}>0$ and $C(\theta)=\theta^a(1-\theta^k)^{-\beta}$ in (3.1.1), the p.m.f of GPSD we get

$$P(X=x)=\binom{(k\beta+x-a-k)/k}{(x-a)/k}\left(\frac{1}{m}\right)^{\beta}\left(1-\frac{1}{m}\right)^{\frac{x-a}{k}}, \quad x=a,\ a+k,\ a+2k,\ \ldots;\quad m=1/(1-\theta^k).$$

As this is the p.m.f of the GHD $H_a(m,k,\beta)$, it belongs to the family of GPSD. Note that

$$C(\theta)=\theta^a(1-\theta^k)^{-\beta}=\sum_{n=0}^{\infty}\binom{\beta+n-1}{n}\theta^{a+nk}=\sum_{x\in S}\binom{(k\beta+x-a-k)/k}{(x-a)/k}\theta^x=\sum_{x\in S}a(x)\theta^x$$

where $a(x)$ is the coefficient of θ^x in the expansion of $C(\theta)$.

Next we consider a modification of stochastic model 1 given in chapter 5.

Suppose when a person is appointed as a sales executive of the company, he will be given an initial amount 'a'. Thereafter, for the selling of each article, an additional amount k will be given, where a, $k>0$ are integers. Let $Y(t)$ denote the number of articles sold in an interval of duration t starting from an initial epoch $t = 0$. The family of r.vs$\{Y(t),\ t \geq 0\}$ is a stochastic process in continuous time with discrete state space $\{0,1,2,\ldots\}$. Now, let $X(t)$ denote the amount received by the sales executive in the time interval $(0, t]$. The family of r.vs$\{X(t),\ t\geq 0\}$ is a stochastic process with $X(0) = a$. Here the time t is continuous and the state space of $X(t)$, $\{a,\ a+k,\ a+2k,\ \ldots\ \}$ is discrete and integer-valued. Thus the whole system is a two-dimensional continuous time stochastic process $\{X(t),\ Y(t),\ t \geq 0\}$ in the state space $\{(a+nk,\ n);\ n \geq 0,\ a,\ k > 0 \text{ integers}\}$.

Let $P_{a+nk}(t)$ be the probability that the r.v $X(t)$ assumes the value $a+nk$.

$$P_{a+nk}(t) = P\{X(t) = a+nk\}\ ,\quad n = 0,1,2,\ldots\ ,\quad a,\ k > 0\quad \text{integers.}$$

$P_{a+nk}(t)$ is a function of the time t and $\sum_{n=0}^{\infty} P_{a+nk}(t) = 1$. $\{P_{a+nk}(t)\}$ represents the probability distribution of the r.v $X(t)$ for every value of t.

Let the probability of selling r articles in $(t,\ t+\ \Delta t)$ given that n articles were sold by epoch t be given by

$$\begin{aligned} P(Y(\Delta t) = r / Y(t) = n) &= \lambda_n \Delta t + 0(\Delta t), && r = 1 \\ &= 0(\Delta t) && r \geq 2 \\ &= 1 - \lambda_n \Delta t + 0(\Delta t) && r = 0 \end{aligned}$$

where $\lambda_n=(a+nk)\lambda$ is a linear function of n and the initial selling rate $\lambda > 0$ is a parameter.

Theorem 7.3

Under the above conditions $X(t)$ follows a generalized Harris distribution, *i.e.*

$P_{a+nk}(t)$ isgiven by

$$P_{a+nk}(t) = \binom{(a/k)+n-1}{n} (e^{-t\lambda k})^{a/k} (1 - e^{-t\lambda k})^n\ ,\quad n = 0,\ 1,\ 2,\ \ldots$$

Proof.

Similar to the proof given in theorem 5.1.

Corollary: The mean and variance of the process are

$E\{X(t)\} = ae^{t\lambda k}$ and

$\mathrm{Var}\{X(t)\} = ake^{t\lambda k}(e^{t\lambda k} - 1)$

Remarks:

(i) The expected gain of the sales executive in an interval of length t is $ae^{t\lambda k}$, so that his expected gain in an interval of unit length is $ae^{\lambda k}$.

(ii) The mean and variance of $X(t)$ are functions of t. Since the distribution of $X(t)$ is functionally dependent on t, the process $\{ X(t), t \geq 0 \}$ is not stationary.

(iii) We have $X(t) = a + k\,Y(t)$

$P(Y(t) = n) = P(X(t) = a + nk)$

$$= \binom{(a/k)+n-1}{n}(e^{-t\lambda k})^{a/k}(1 - e^{-t\lambda k})^n \ , \quad n = 0,1,2,\ldots$$

which is the p.m.f of the negative binomial distribution $NB(a/k, e^{-t\lambda k})$ defined on $\{0, 1, 2, \ldots\}$. The mean and variance of $Y(t)$ are

$E\{Y(t)\} = (e^{t\lambda k} - 1)a/k$ and

$\mathrm{Var}\{I(t)\} = e^{t\lambda k}(e^{t\lambda k} - 1)a/k$

Particular case

When $a = k = 1$, the distribution of $X(t)$ is given by

$$P_n(t) = P(X(t) = n) = (e^{-t\lambda})(1 - e^{-t\lambda})^{n-1} \ , \quad n = 1, 2, \ldots$$

which is the p.m.f of the decapitated geometric with parameter $e^{-\lambda t}$.

So when $a = k = 1$, the stochastic process $\{X(t), t \geq 0\}$ is the Yule-Furry process.

TABLE 7.1

Model fits to the countable time in seconds for an effective call measure of 30 seconds.

x	Observed	$H_1(m,30,1/30)$	$H_1^*(m,30,1/30)$	$GP(\lambda,30)$	$EG(p,30)$
1	338	340.92	341.28	313.317	316.098
31	10	9.000	9.000	43.516	38.549
61	4	3.960	3.600	3.022	4.701
91	3	2.088	2.016	0.140	0.573
121	3	1.277	1.219	0.005	0.070
151	1	0.831	0.783	0.000	0.009
≥ 181	1	1.924	2.102	0.000	0.000
Chi-square value		0.502	0.677	52.394	30.910
p-value		0.479	0.411	4.54E-13	2.70E-08

TABLE 7.2

Model fits to the countable time in seconds for an effective call measure of 60 seconds.

x	Observed	$H_1(m,60,1/60)$	$H_1^*(m,60,1/60)$	$GP(\lambda,60)$	$EG(p,60)$
1	510	514.272	514.8	495.072	496.057
61	9	6.864	6.864	31.880	30.011
121	5	2.749	2.662	1.026	1.816
181	2	1.468	1.398	0.022	0.000
241	1	0.880	0.824	0.000	0.000
≥ 301	1	1.767	1.452	0.000	0.000
Chi-square value		1.365	1.830	77.209	40.959
p-value		0.243	0.176	1.54E-18	1.554E-10

TABLE 7.3

Model fits to the countable time in seconds for an effective call measure of 180 seconds.

x	Observed	$H_1(m,180,1/180)$	$H_1^*(m,180,1/180)$	$GP(\lambda,180)$	$EG(p,180)$
1	859	863.625	863.625	826.402	827.703
181	6	4.372	4.305	47.223	44.741
361	2	2.003	1.942	1.349	2.418
541	2	1.221	1.164	0.026	0.131
721	2	0.836	0.785	0.000	0.007
901	1	0.610	0.564	0.000	0.000
1081	1	0.464	0.422	0.000	0.000
1261	1	0.363	0.325	0.000	0.000
≥ 1441	1	1.506	1.868	0.000	0.000
Chi-square value		2.239	2.125	91.373	56.409
p-value		0.135	0.145	1.19E-21	5.89E-14

TABLE 7.4
Values of power function-Test (A)

m	$A(m)$
1.1	.000
2.0	.000
2.5	.000
3.0	.003
3.5	.024
3.75	.050
3.8	.058
3.9	.072
4.0	.089
4.5	.210
5.0	.342
5.5	.485
6.0	.613
7.0	.797
8.0	.898
9.0	.950
10	.975
12	.993
14	.998
15	.999
16	.999
17	1.00
18	1.00

TABLE 7.5
Values of power function-Test (B)

m	$B(m)$
1.1	.000
2.0	.000
2.5	.001
2.7	.002
2.8	.004
3.2	.021
3.3	.029
3.4	.039
3.5	.050
3.6	.063
3.8	.095
4.0	.132
5.0	.377
6.0	.608
7.0	.769
8.0	.867
9.0	.923
10	.955
12	.984
15	.996
18	.999
20	.999
21	1.00

TABLE 7.6
Values of power function-Test (C)

m	$C(m)$
2.0	.000
4.0	.001
5.0	.007
6.0	.030
7.0	.076
13	.545
15	.662
20	.838
25	.918
30	.956
35	.974
40	.985
45	.990
46	.991
50	.994
55	.996
60	.997
65	.998
70	.998
80	.999
85	.999
91	.999
92	1.00

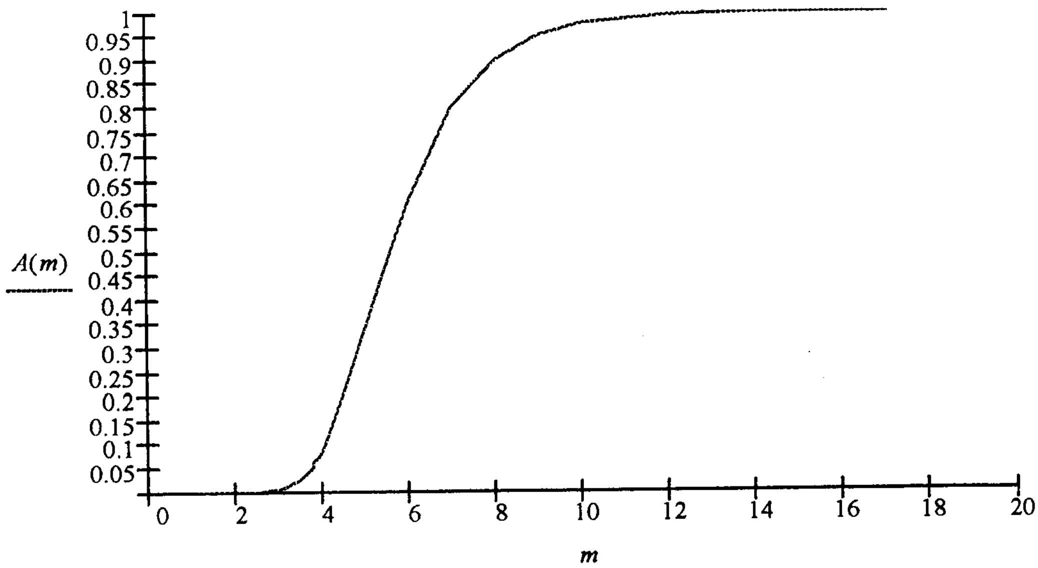


FIGURE 7.1 – Power curve of Test (A)

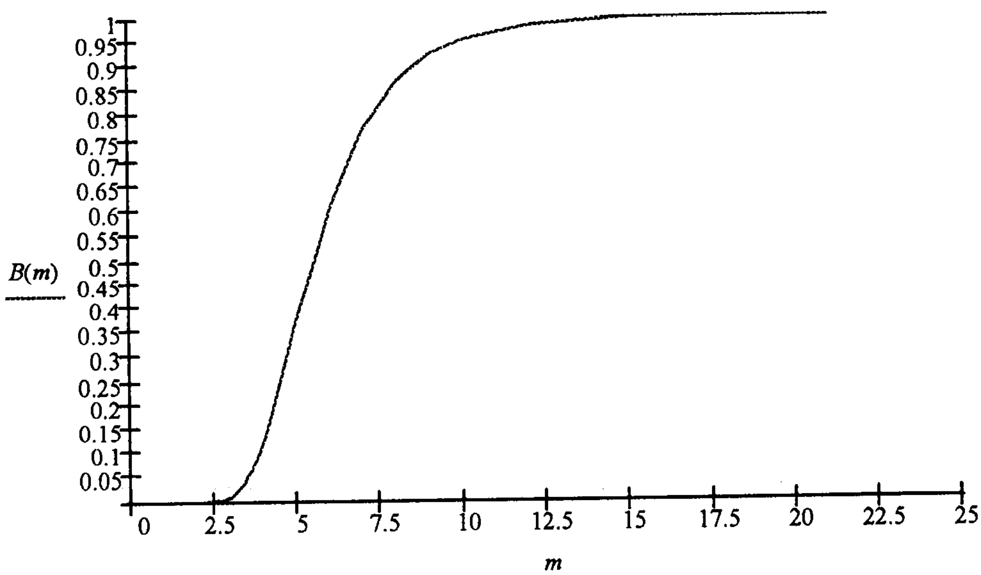


FIGURE 7.2 – Power curve of Test (B)

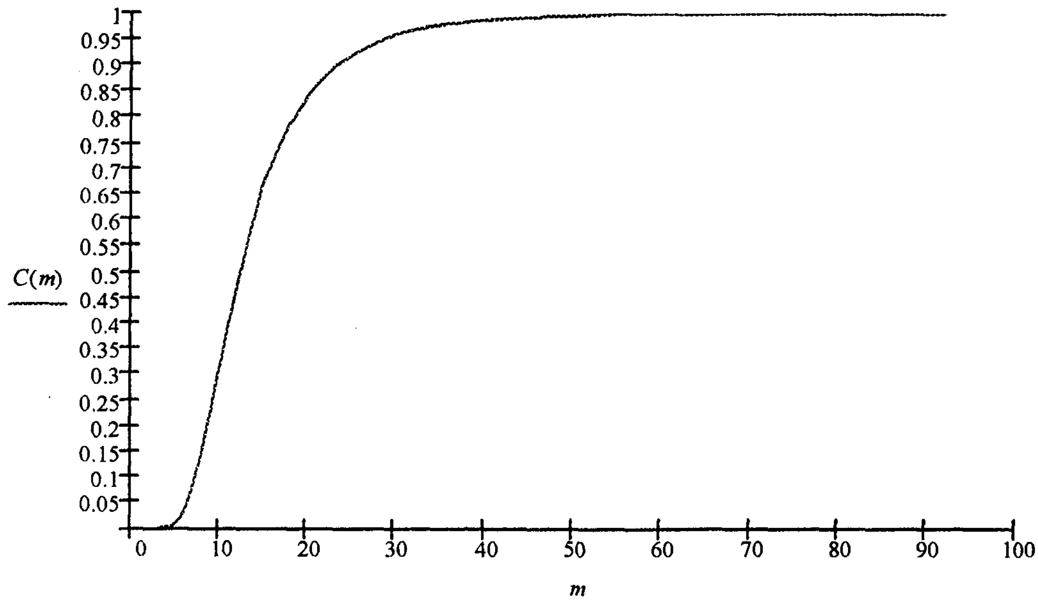


FIGURE 7.3 – Power curve of Test (C)

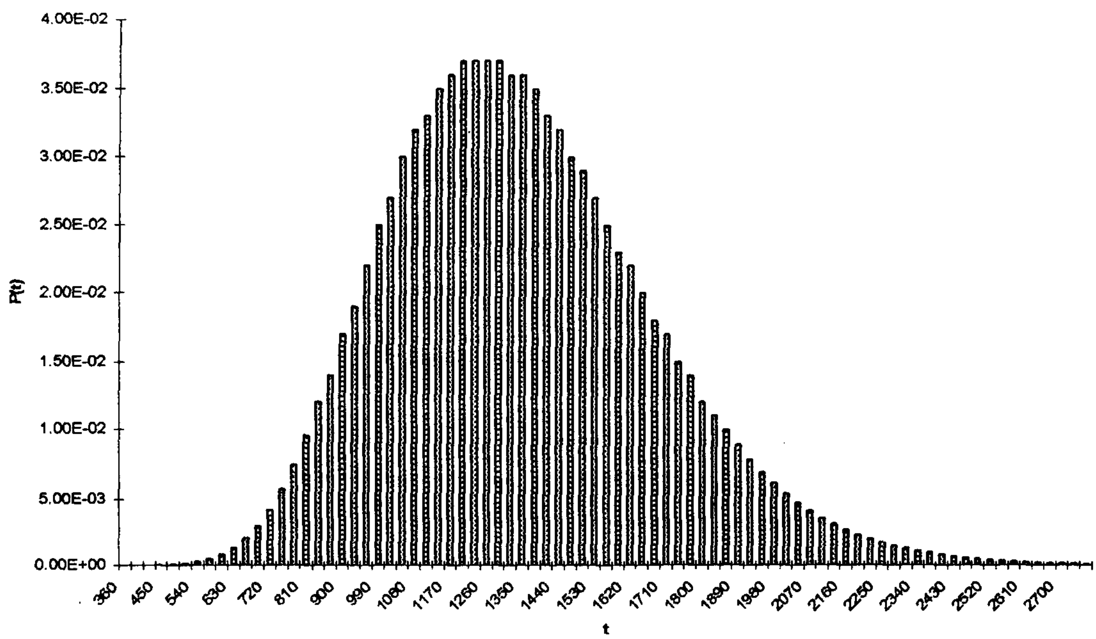


FIGURE 7.4 – Probability plot of sampling distribution of T of Test (A)

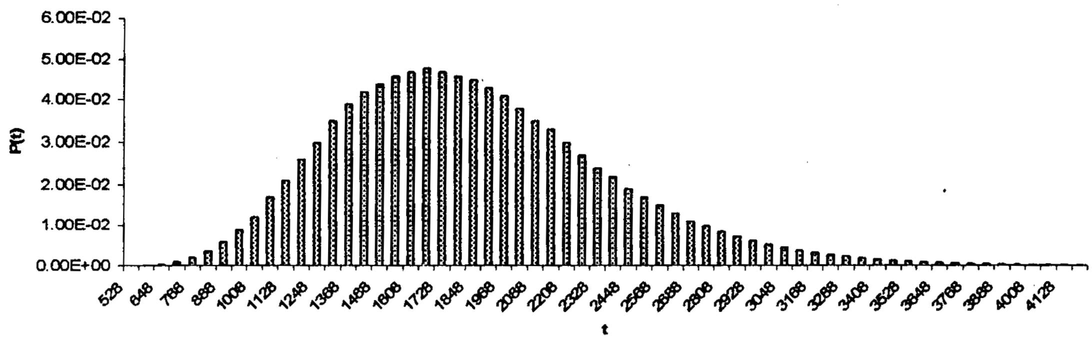


FIGURE 7.5 – Probability plot of sampling distribution of T of Test (B)

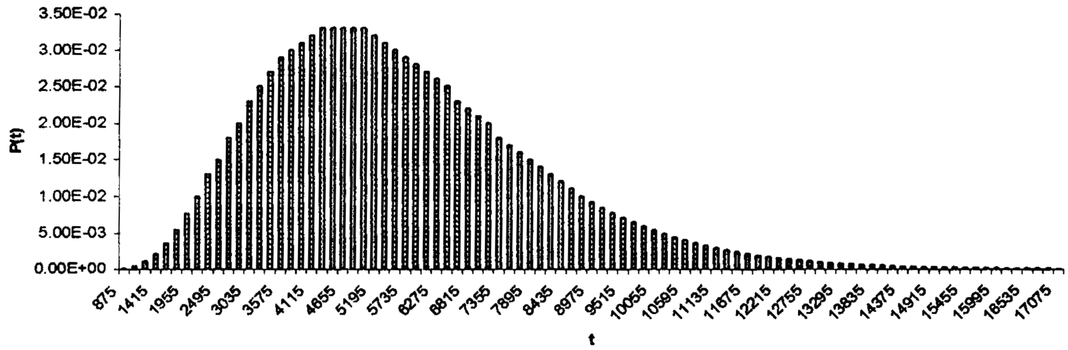


FIGURE 7.6 – Probability plot of sampling distribution of T of Test (C)

159

REFERENCES

[1] AMARANATH, T. (1997), *An Elementary Course in Partial Differential Equations*, Narosa Publishing House, New Delhi.

[2] ARBOUS, A.G. and KERRICH, J.E. (1951), Accident statistics and the concept of accident proneness, *Biometrics*, 7, 340-432.

[3] BALAKRISHNA, N. and JAYAKUMAR, K. (1997), Bivariate semi-Pareto distributions and processes, *Statistical Papers*, 38, 149 – 165.

[4] BARNDORFF-NIELSEN, O. and YEO, G.F. (1969), Negative binomial processes, *Journal of Applied Probability*, 6, 633-647.

[5] BECKER-KERN, P. (2001), Max-stable hemigroups: structure, domains of attraction and limit theorems with random sample size, *Probability and Mathematical Statistics*, 21, 441-465.

[6] BENNETT, B.M. and BIRCH, B. (1964), Sampling inspection tables for comparison of two groups using the negative binomial model, *Trabajos de Estadistica*, 15, 1-12.

[7] BHAT, B.R. (2000), *Stochastic Models - Analysis and Applications*, New Age International (P) Ltd., Publishers, New Delhi.

[8] BLISS, C.I. and FISHER, R.A. (1953), Fitting the negative binomial distribution to biological data and note on the efficient fitting of the negative binomial, *Biometrics*, 9, 176-200.

[9] CADIGAN, N.G. and CHEN, J. (2001), Properties of robust M–estimators for Poisson and negative binomial data, *Journal of Statistical Computation and Simulation*, 70, 273-288.

[10] CHEW, V. (1964), Applications of the negative binomial distribution with probability of misclassification, *Virginia Journal of Science* (*New Series*), 15, 34-40.

[11] CINLAR, E. and AGNEW, R.A. (1968), On the superposition of point processes, *Journal of the Royal Statistical Society*, B, 30, 576-581.

[12] DOUGLAS, J.B. (1980), *Analysis with standard contagious distributions*, Burtonsville, MD, International Co-operative Publishing House.

[13] FELLER, W. (1966), *An Introduction to Probability Theory and its Applications*, Vol.2, John Wiley and Sons, New York.

[14] FELLER, W. (1968), *An Introduction to Probability Theory and its Applications*, Vol.1, John Wiley and Sons, New York.

[15] FURRY, W.H. (1937), On fluctuation phenomena in the passage of high energy electrons through lead, *Physical Review*, 52, 569-581.

[16] GNEDENKO, B.V. (1978), *The Theory of Probability*, Mir Publishers Moscow.

[17] GREENWOOD, M and YULE, G. (1920), An inquiry into the nature of frequency distributions of multiple happenings etc., *Journal of the Royal Statistical Society, Series A*, 83, 255-279.

[18] HANSEN, B.G. and WILLEKENS, E. (1990), The generalized logarithmic series distribution, *Statistics and Probability Letters*, 9, 311-316.

[19] HARRIS, T.E. (1948), Branching Processes, *Annals of Mathematical Statistics*, 19, 474-494.

[20] JAYAKUMAR, K. and PILLAI, R. N. (1992), On class-L distributions, *Journal of the Indian Statistical Association*, 30, 103-108.

[21] JOHNSON, N. L. and KOTZ , S. (1977), *Urn models and their applications*, New York, Wiley.

[22] JOHNSON, N.L., KOTZ, S. and KEMP, A.W. (1992), *Univariate Discrete Distributions*, 2nd Edition, John Wiley and Sons, New York.

[23] KARLIN, S and TAYLOR, H.M. (1975), *A First Course in Stochastic Processes,* Academic Press, New York, 2nd ed.

[24] KATTI, S.K. (1966), Interrelations among generalized distributions and their components, *Biometrics*, 22, 44-52.

[25] KEMP, C.D. (1970), Accident proneness and discrete distribution theory, *Random Counts in Scientific Work*, 2: *Random Counts in Biomedical and Social Sciences*, 41-65. University Park, Pennsylvania State University Press.

[26] KENDALL, D.G. (1949), Stochastic processes and population growth, *Journal of the Royal Statistical Society, Series B*, 11, 230-282.

[27] KHATRI, C.G. and PATEL, I.R. (1961), Three classes of univariate discrete distributions, *Biometrics*, 17, 567-575.

[28] KOZUBOWSKI T.J. and PANORSKA A.K. (1998), Weak limits for multivariate random sums, *Journal of Multivariate Analysis*, 67, 398 - 413.

[29] LEHMANN E.L. (1983), *Theory of Point Estimation*, John Wiley and Sons.

[30] LEHMANN E.L. (1986), *Testing Statistical Hypotheses*, 2nd Edition, John Wiley and Sons.

[31] LUKACS, E. (1970), *Characteristic Functions*, 2nd Edition, Charles Griffin and Company Limited, London.

[32] MAEJIMA, M. and NAITO, Y. (1998). Semi-selfdecomposable distributions and a new class of limit theorems, *Probability Theory and Related Fields*, 112, 13-31.

[33] MANN, N.M., SCHAFER, R.E. and SINGPURWALLA, N.D. (1974), *Methods for Statistical Analysis of Reliability and Life Data*, New York, Wiley.

[34] MARSHALL, A.W. and OLKIN, I. (1997), On adding a parameter to a distribution with special reference to exponential and Weibull models, *Biometrika*, 84, 641-652.

[35] MARTIN, D.C. and KATTI, S.K. (1965), Fitting of some contagious distributions to some available data by the maximum likelihood method, *Biometrics*, 21, 34-48.

[36] McKENZIE, E. (1986), Autoregressive moving-average processes with negative binomial and geometric marginal distributions, *Advances in Applied Probability*, 18, 679-705.

[37] MEGYESI, Z. (2002). Domains of geometric partial attraction of max-semistable laws: structure, merge and almost sure limit theorems, *Journal of Theoretical Probability*, 15, 973-1005.

[38] MILNE, R.K. and YEO, G.F. (1989), Random sum characterizations, *Mathematical Scientist*, Vol.14, 120-126.

[39] NOACK, A. (1950), A class of random variables with discrete distributions, *Annals of Mathematical Statistics*, 21, 127-132.

[40] PATIL, G.P. (1961), *Contributions to Estimation in a Class of Discrete Distributions*, Ph.D. thesis, Ann Arbor, MI: University of Michigan.

[41] PATIL, G.P. (1962) Certain properties of the generalized power series distribution, *Annals of the Institute of Statistical Mathematics, Tokyo*, 14, 179-182.

[42] PERRY, J.N. (1984), Negative binomial model for mosquitoes, *Biometrics*, 40, 863.

[43] PILLAI, R. N. (1985), Semi-α-Laplace distributions. *Communications in Statistics-Theory and Methods*, 14, 991–1000.

[44] PILLAI, R. N. (1991), Semi-Pareto processes, *Journal of Applied Probability*, 28, 461 - 465.

[45] PILLAI, R.N. and SANDHYA, E. (1990), Distributions with complete monotone derivative and geometric infinite divisibility, Advances in Applied Probability, 22, 751-754.

[46] PILLAI, R.N. and SANDHYA, E. (1996), Geometric sums and Pareto law in reliability theory, I.A.P.Q.R. Trans., 21, 137-142.

[47] RAO, B.R., MAZUMDAR, S., WALLER, J.H. and LI, C.C. (1973), Correlation between the numbers of two types of children in a family, *Biometrics*, 29, 271-279.

[48] ROHATGI, V.K. (1976), *An Introduction to Probability Theory and Mathematical Statistics*, John Wiley and Sons.

[49] ROSS, S.M. (2003), *Introduction to Probability Models.* Academic Press, USA, 8th ed.

[50] SANDHYA, E. (1991a), Geometric Infinite Divisibility and Applications, Ph.D. Thesis (unpublished), University of Kerala, January 1991.

[51] SANDHYA, E. (1991b), On geometric infinite divisibility, *p*-thinning and Cox processes, *Journal of the Kerala Statistical Association*, 7, 1-10.

[52] SANDHYA, E. (1996), On a generalization of geometric infinite divisibility, *Proceedings of the 8th Kerala Science Congress, January*-1996, 355–357.

[53] SANDHYA, E. and SATHEESH, S. (1997), On exponential mixtures, mixed Poisson processes and generalized Weibull and Pareto models, *Journal of Indian Statistical Association*, 35, 45-50.

[54] SANDHYA, E., SHERLY, S. and RAJU, N. (2006*a*), Harris family of discrete distributions. Accepted for publication in the *Proceedings of the National Seminar* conducted by Kerala Statistical Association and the Department of Statistics, University of Kerala, March 2006.

[55] SANDHYA, E., SHERLY, S., JOS, M.K. and RAJU, N. (2006*b*), Characterizations of the extended geometric, Harris, negative binomial and gamma distributions. Accepted for publication in *STARS. Int. Journal.*

[56] SATHE, Y.S. and RAVI, S. (1997), A characterization of negative binomial distribution, *Journal of the Indian Statistical Association*, 35, 31-32.

[57] SATHEESH, S. (2004a), Why there are no gaps in the support of infinitely divisible discrete laws?, *Prob Stat Models*, 3, 1-7.

[58] SATHEESH, S. (2004*b*), Another look at random infinite divisibility, *Statistical Methods*, 6, 123-144.

[59] SATHEESH, S., NAIR, N.U. (2002*a*), A Note on maximum and minimum stability of certain distributions, *Calcutta Statistical Association Bulletin*, 53, 249-252.

[60] SATHEESH, S., NAIR, N.U. (2002*b*), Some classes of distributions on the non-negative lattice, *Journal of the Indian Statistical Association*, 40, 41–58.

[61] SATHEESH, S., NAIR, N.U. (2004), On the stability of geometric extremes, *Journal of the Indian Statistical Association*, 42, 99–109.

[62] SATHEESH, S., NAIR, N.U. and SANDHYA, E. (2002), Stability of random sums, *Stochastic Modeling and Applications*, 5, 17–26.

[63] SATHEESH, S. and SANDHYA, E. (1997), Distributions with completely monotone probability sequences, *Far East Journal of Theoretical Statistics*, 1,1, 69-75.

[64] SATHEESH, S. and SANDHYA, E. (2005). Semi-selfdecomposable laws and related processes, *Journal of the Indian Statistical Association*, 43, 157-166.

[65] SATHEESH, S. and SANDHYA, E. (2006*a*). Semi-stability of sums and maximums in samples of random size, in *Focus on Probability Theory*, Editor L. R. Velle, 43-72, Nova Science Publishers, New York.

[66] SATHEESH, S. and SANDHYA, E. (2006*b*). Max-Semi-selfdecomposable laws and related processes, *Statistics and Probability Letters*, 76, 1435-1440.

[67] SATHEESH, S. and SANDHYA, E. (2006*c*), Semi-selfdecomposable laws in the minimum scheme, submitted (presented at the National Conference, Department of Statistics, University of Kerala, Trivandrum, India, 16-18 of March 2006).

[68] SATHEESH, S., SANDHYA, E. and SHERLY, S. (2005), A Generalization of stationary AR (1) schemes, Accepted for publication in *Statistical Methods*.

[69] SHERLY, S., JOS, M.K., SANDHYA, E. and RAJU, N. (2006), Harris Processes, Submitted.

[70] SIMON, L.J. (1961), Fitting negative binomial distributions by the method of maximum likelihood, *Proceedings of the Casualty Actuarial Society*, XLVIII, 45 – 53.

[71] SNEDDON, I.N. (1957), *Elements of Partial Differential Equations*, McGraw-Hill International Editions.

[72] SREEHARI, M. (1979), An extension of Goldie's result and further results in infinite divisibility, *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 47, 19-25.

[73] STUDENT (1907), On the error of counting with a haemocytometer, *Biometrika*, 5, 351-360.

[74] SZASZ, D. (1972), On classes of limit distributions for sums of a random number of identically distributed independent random variables, *Theory of Probability and Applications*, 17, 401-415.

[75] TAYLOR, H.M. and KARLIN, S. (1984), *An Introduction to Stochastic Modeling*, Orlando, FL: Academic Press.

[76] VOORN, W.J. (1987), Characterizations of the logistic and log-logistic distributions by extreme value related stability with random sample size, *Journal of Applied Probability*, 24, 838-851.

[77] WAERDEN B.L. VAN DER (1960), Sampling inspection as a minimum loss problem, *Annals of Mathematical Statistics*, 31, 369-384.

[78] WALSH, J.E. (1955), The Poisson distribution as a limit for dependent binomial events with unequal probabilities, *Operations Research*, 3, 198-209.

[79] WANG YUHUA, LINDA J. YOUNG and DALLAS E. JOHNSON (2001), A UMPU test for comparing means of two negative binomial distributions, *Commu. Statist-Simula.*, 30(4), 1053-1075.

[80] WILSON, L.T. AND ROOM,P.M. (1983), Clumping patterns of fruit and arthropods in cotton, with implications for binomial sampling, *Environmental Entomology*, 12, 50-54.

[81] YANNAROS, N. (1989), On Cox and renewal processes, *Statistics and Probability Letters*, 7, 431-433.