Chapter 1

Introduction

1.1 General Introduction

The theory of discrete probability distribution is considered as a useful branch in statistics, having its applications in a wide variety of disciplines. Its origin began with the work of Bernoulli [8] and Poisson [83]. In recent years, the mixture of distribution has drawn continuous attention since the traditional models such as Poisson, Negative Binomial, Geometric etc. which can be formulated on the basis of simple models, have found to be inadequate in situations which occur in number of phenomena. Hence, the univariate mixture distributions came into existence and have become an immensely useful branch of statistics. The mixture distributions combine two or more of the elementary distributions through the process of combining or generalizing. A detailed study on these discrete mixture distributions and their properties can be found in the works of Neyman [75], Haight [48, 49], Khatri [64], Katti [60], Johnson and Kotz [57], Johnson et al. [56] etc.

1.2 Preliminaries

In distributional theory, a mixture of distribution may be defined as being made up of two or more component distributions is a mechanism to construct new distributions from the given ones for which empirical justification must be sought. It is an important class of distribution in distributional theory which provides much flexibility than the traditional models. The notion of mixing often has a simple and direct interpretation in terms of the physical situation under investigation where the random variable concerned may be the result of actual mixing of a number of different populations.

A large number of mixture distributions were derived as it has been observed that in many situations the simple basic distribution fails to describe a set of data for which some researcher's belief that the model underlying the distribution has some of the particular characteristics of the mixture model for which further research was made to see if any simpler mixture distribution will describe the data to any degree of satisfaction.

Medgyessy [71] define mixture of distributions as a superimposition of distributions with different functional forms or different parameters in specified proportion. If $F_j(x_1, x_{2,} \ldots, x_n), (j = 0, 1, \ldots, m)$ represents different cumulative distribution functions (c.d.f) and $w_j \geq 0$ and $\sum_{j=0}^{m} w_j = 1$ then,

$$F(x_1, x_{2,} \ldots, x_n) = \sum_{j=0}^{m} w_j F_j(x_1, x_{2,} \ldots, x_n)$$

is a proper cumulative distribution function. The mixture distribution $\{F_j\}$ is finite or infinite or infinite according as m is finite or infinite. Thus, there are two important categories of mixture distributions viz. finite mixture and countable and continuous mixture of discrete distributions.

(i) Finite mixture of discrete distributions

The concept of finite mixture distribution was introduced by Pearson (1915). A finite mixture distribution is formed from k different component distributions with cumulative distribution functions (cdf) $F_1(x), F_2(x), \ldots, F_k(x)$ and mixing weights $w_1, w_{2,} \ldots, w_k$, where $w_j > 0$, $\sum_{j=1}^{k} w_j = 1$, by taking the weighted average $F(x) = \sum_{j=1}^{k} w_j F_j(x)$ as the cdf of new mixture distribution.

If the component distributions are defined on non-negative integers with

$$P_j(x) = F_j(x) - F_j(x-1),$$

then the mixture distribution is a discrete distribution with pmf

$$P[X = x] = \sum_{j=1}^{k} w_j P_j(x).$$

The problem of finite mixture distribution arises due to unavailability of data for each conditional distribution separately but is available only for the overall mixture distribution. Such situation often arises as it is impossible to observe some underlying variables which split the observations into groups. In this case interest often lies on estimating the mixing proportions and on estimating the parameters in the conditional distributions. Zero modified or inflated distribution is an example of finite mixture of distribution. Now, we may define inflated distribution as the random variable whose probability mass function at the point $X = x$ is given as

$$P[X = x] = \begin{cases} w + (1 - w)P_x, & x = 0 \\ (1 - w)P_x, & x = 1, 2, 3, \dots \end{cases}$$

where $P_x, (x = 0, 1, 2, \dots)$ is the pmf of the original distribution without inflation. It is also possible to take $w < 0$, provided $w + (1 - w)P_x \geq 0$.

The probability generating function (pgf) of inflated distribution is

$$g(t) = w + (1 - w)G(t),$$

where, $G(t)$ is the pgf of original distribution without inflation.

(ii) Countable and Continuous Mixture of Discrete Distributions

A mixture distribution also arises when the cumulative distribution function of a random variable depends on the parameter $\theta_1, \theta_2, \dots, \theta_m$ and some (or all) of those parameters may vary. A mixture of this type is represented by

$$\mathcal{F}_A \bigwedge_{\Theta} \mathcal{F}_B$$

where $\mathcal{F}_A$ represents the original distribution and $\mathcal{F}_B$ represents the mixing distribution. The new distribution has the cumulative distribution function

$$E[F(X|\theta_1, \theta_2, \dots, \theta_m)],$$

where the expectation is with respect to the joint distribution of the k parameters that vary. This includes situations where the source of a random variable is unknowable. When Θ has a discrete distribution with probabilities $p_i, i = 0,1, \dots$ we call the outcome a countable mixture of discrete distribution. The pmf of the mixture is

$$P[X = x] = \sum_{i \geq 0} p_i P_i(x),$$

where, $$P_j(x) = F_j(x) - F_j(x-1).$$

When the points of increase of the mixing distribution are continuous, the outcome is called as a continuous mixture. A continuous mixture of discrete distribution arises when a parameter corresponding to some features of a model for a discrete distribution can be regarded as a random variable taking continuous values. The theory of countable and continuous mixture of discrete distribution was first studied by Greenwood and Yule [46] and Lundberg [67].

Our study comprises of finite mixture of discrete distributions i.e. observations are available from a population which is known to be a mixture of some sub-population. Each sub-population that is considered will be assumed to have the same type of distribution (or different type of distribution) but with different parameter values.

A wide class of mixture distribution in univariate case has been constructed by the process of compounding and generalizations.

The term compounding has often been used in terms of 'mixing'. A compound distribution may be defined as a distribution that results from allowing the parameter of a distribution to vary. Let X_1 be a random variable having the distribution function $F(x|\theta)$, the parameter θ being treated as a random variable, $X_2$with distribution function $H(\theta)$. The distribution which is obtained by summing X_1 using the mixing parameter θ has the distribution function

$G(x) = \sum_{\Theta} F(x|\theta)H(\theta)$ where Θ is the parameter space, integration is used when X_2 has a continuous distribution. Symbolically, it may be written as

$$X_1 \wedge_{\Theta} X_2$$

where X_1 represents the original distribution, θ the varying parameter and X_2 be the compounding distribution.

Generalization is a process which generates a variety of distributions. Feller [35] defines generalization as the process in which the new distribution results from the combination of two independent distributions in a particular way. Gurland [47] reinforced the use of the term generalization.

Let Y_1 and Y_2 be two random variables having the distribution functions F_1 and F_2 and the probability generating functions $g_1(t)$ and $g_2(t)$ respectively. Then, the distribution with probability generating function

$$g(t) = g_1\{g_2(t)\}$$

is called an Y_1 distribution generalized by the generalizing Y_2 distribution. Symbolically, it is represented as,

$$Y_1 \vee Y_2$$

Mixture of distributions have received an increasing amount of attention in statistical literature due to an increased interest in mathematics involved in dealing with mixtures and largely due to an increasing number of specific problems encountered in certain applications.

Mixtures of discrete distributions are also applicable in the area of life testing and acceptance testing along with biological applications. In recent years' different authors have discussed various applications of different mixture distributions obtained by them in the areas such as accident data, error data, modeling and waiting survival times etc.

1.3 Brief historical overview

The theory of discrete probability distribution was developed during the works of James Bernoulli and Poisson. James Bernoulli derived the binomial distribution and published it in the year 1713 whereas; Poisson distribution was derived by French mathematician Simeon D. Poisson in the year 1837 derived Poisson distribution as a limiting form of binomial distribution. Also, the negative binomial distribution was obtained by Greenwood and Yule [46] as a consequence of certain assumptions in accident proneness models.

In the recent past, it has been observed that the traditional models like binomial, Poisson, negative binomial, logarithmic series etc. fails to describe count data as it exhibits the properties of over-dispersion and long tail behavior. So, there arises the demand to further generalize the traditional models in encounters such problems. One such method that is used to generalize these models is mixture of the distributions. A large number of discrete mixture distributions are proposed by different researchers. These distributions have drawn its attention in the field of Insurance, economics, social sciences, engineering and so on.

Everitt and Hand [33], Consul [25], and Johnson et al. [56] in their books give some detailed information of the vast area of the discrete mixture distribution, their properties and applications. Everitt and Hand [33] in their book Finite Mixture Distributions reviewed the literature of Karl Pearson [81] "estimation of five parameters in a mixture of two normal distributions" and in addition to indicate the practical details of fitting such distributions to sample data. Titterington [100] used finite mixture distribution in speech recognition and in image analysis.

Modified distribution which is also known as inflated distribution is an example of finite mixture of discrete distribution. Singh [96] introduced inflated distribution and studied inflated Poisson distribution to serve the probabilistic description of an experiment with slight inflation at a point say zero. The generalized inflated Poisson distribution was studied by Pandey [76]. Heilborn [51] discussed on mechanisms producing zero-modified distribution in the content of generalized linear model. Zero-inflated Poisson distribution was proposed by Lambert [65]. Holgate [52] studied zero-modified geometric distribution as a model for the length of residence of animals in a specified habitat. Heilbron [51] in the content of generalized linear model discussed on mechanisms producing zero modified distribution. Bohing [12] reviewed the related literature of Zero-inflated Poisson (ZIP) distribution Proposed by Lambert [65].

The works on countable and continuous mixture of discrete distributions by "accident proneness" theory and to actual risk theory of Greenwood and Yule [46] and Lundberg [67].

Different mixtures of Poisson distributions are discussed in the book by Johnson et al. [56] where the mixing distributions are continuous or countable.

A Poisson mixture of Poisson distributions known as Neyman type A distribution has been used to describe plant distributions, especially when the reproduction of the species produces cluster. Evan [32] used this distribution and found that it gives good result for plant production. Cresswell and Froggatt [27] derived the Neyman Type A distribution in context of bus driver accidents.

A Poisson mixture of negative binomial distribution called as the Poisson–Pascal distribution was introduced by Skellam [97] in the context of spatial distribution of plants. Katti and Gurland [47] studied its properties and estimation and derived it from an entomological model. The Hermite distribution which is a Poisson mixture of binomial distribution is studied by Kemp and Kemp [62]. A new distribution known as Gegenbauer distribution is introduced by Plunkett and Jain [82] by mixing the Hermite distribution with gamma distribution. Borah [13] studied the probability and moment properties of Gegenbauer distribution. Medhi and Borah [70] investigated the four parameter generalized Gegengauer distribution and estimated the parameters by the method of moment and ratio of first two frequencies and $\bar{x}$ and s^2. Also the generalized Gegenbauer distribution obtained by Medhi and Borah [70]

The Lindley distribution is proposed by Lindley [66] in the context of Bayesian statistics, as a counter example to fiducial statistics. Sankaran [85] introduced Poisson Lindley distribution while modeling count data, assuming the parameter of Poisson distribution to follow Lindley distribution. Poisson Lindley distribution is a special case of Bhattacharya's [9] more complicated mixed Poisson distribution. Borah and Deka Nath [16] studied inflated Poisson Lindley distribution with inflation at point zero and applied it to problem of biological data. Borah and Deka Nath [15] also studied on some mixtures of Poisson Lindley distribution derived by using Gurland's generalization. Borah and Begum [14] studied on some properties of Poisson Lindley distribution and its derived distribution. Also, the two forms of geometrically infinite divisible two parameter Poisson-Lindley distribution has been studied by Borah and Begum and has been fitted to some biological and ecological data for comparison.

Ghitany and Al Mutairi [38] studied the size biased version of Poisson Lindley distribution and discussed various properties and applications. Ghitany et al. [40] introduced the zero-truncated Poisson Lindley distribution and applications. Ghitany and Al Mutairi [39] also studied on estimation method of Poisson Lindley distribution and showed that the method of moment and method of likelihood estimators of the parameter are consistent and asymptotically normal. Zakerzadeh and Dolati [103] obtained an extended version of Poisson Lindley distribution named as generalized Poisson-Lindley distribution by compounding the Poisson distribution with the generalized Lindley distribution. Various properties and applications have been studied and also showed that the cumulative distribution of the generalized Poisson Lindley distribution can be expressed in terms of incomplete beta function ratio and hence as a sum of two binomial terms. Shanker and Mishra [92] obtained size biased quasi Poisson Lindley distribution.

A two parameter Poisson Lindley distribution which is a special case of Sankaran's [85] one parameter Poisson Lindley distribution is introduced by Shanker et al. [90]. Ghrine and Zeghdoudi [42] studied on Poisson quasi Lindley distribution. Generalized Poisson Lindley distribution has been introduced by Bhati et al. [10] which is another form of two parameters Poisson Lindley distribution. Some statistical properties have been studied and knowledge about the parameters is obtained through method of moments, maximum likelihood method and EM algorithm. Wongrin and Bodhisuwan [101] obtained a three parameter Poisson mixture of generalized Lindley distribution known as the Poisson generalized Lindley distribution with applications. Zero truncated two parameters Poisson-Lindley distribution has been studied by Shanker and Shukla [93] and studied various statistical properties.

1.4 Objectives:

The main objectives of the present study are as follows

- To derive certain discrete mixture distributions from the continuous distributions.
- To study certain properties of generalized derived distributions.
- To study the flexibility of the distribution
- To estimate the parameters of the proposed distributions using different method of estimation.
- To fit the distributions to some reported data sets

1.5 Organization

The thesis entitled "A Study on certain Univariate Discrete Mixture Distributions" comprises of seven chapters.

The first chapter is an introductory one which gives an account of the relevant works done earlier by different authors in the theory of univariate discrete probability distributions.

The second chapter is based on new quasi Poisson Lindley which is obtained by compounding Poisson distribution with that of Lindley distribution. Different properties of the distribution have been studied including the probability recurrence relation and factorial moment recurrence relation. The parameters are estimated by considering the method of moment.

In the third chapter we have obtained the size biased version of new quasi Poisson Lindley distribution. The shape of the probability function has been shown graphically. The probability generating function and factorial moment generating function has also been obtained. Parameters are estimated by considering the first two moments and the method of maximum likelihood

The fourth chapter is based on generalized two parameter Poisson-Lindley distribution which was obtained by Shanker and Mishra [93]. Additional statistical properties like the shape of the probability function, skewness, kurtosis and Index of dispersion are obtained. The expressions for recurrence relation for probabilities and factorial moment recurrence relation have been obtained. The parameters by obtained by the method of moment.

In the fifth chapter, we have made a comparative study on zero-truncated Poisson-Lindley, new generalized Poisson-Lindley and generalized two-parameter Poisson-Lindley distributions. Different statistical properties of this distribution have been studied. The parameters are obtained by the method of maximum likelihood.

A new discrete mixture distribution named Poisson-Sushila distribution which is obtained by compounding Poisson distribution with that of Lindley distribution has been obtained in chapter 6. Properties like unimodality, recurrence relations, and measures of dispersion have also been studied. The parameters are obtained by the method of moments.

The seventh chapter is based on some properties of Poisson size biased new quasi Lindley (PSBNQL) distribution which is obtained by compounding Poisson distribution with the size biased new quasi Lindley distribution. Graphical representation of PSBNQL distribution been obtained. Properties such as recurrence relations, moments, index of dispersion, and estimation of parameters have been discussed.