Chapter 1
Introduction
Chapter 1

Introduction

1.1 General Introduction

A significant role is played by the distribution functions in modelling naturally occurring phenomena with all their properties and interrelationships. Large number of distribution functions have been proposed and defined in the literature, which are found to be applicable to many events in real life. Various methods exist in defining statistical distributions. Most of them have arisen from the need to model naturally occurring events. One such distribution is the extreme value distribution. Extreme value theory has been used to understand the tail properties of probability distributions. Applications of this theory has been extensive in the areas of telecommunication, finance and environmental data. This theory has originated mainly from the needs of astronomers in utilizing or rejecting outlying observations (Johnson et al. (1995) [92]). There are three types of extreme value distributions Type 1 (Gumbel-type distribution), Type 2 (Frechet-type distribution) and Type 3 (Weibull-type distribution). The three types of distributions may all be represented as members of a single family of generalized extreme value distribution (Johnson et al. (1995) [92]). The Type 2 and Type 3 extreme value distributions is confined to discussion in this thesis because of their wide applicability. Both the Type 2 and Type 3 extreme value distributions (Frechet and the Weibull distribution) are power transformations of an exponentially distributed variate. The reciprocal of a Weibull variate is a Frechet variate.

1.2 Importance of the study

Extreme value distribution has been widely used in the last 50 years to model environmental data, extreme levels of a river in hydrology, and the largest claim in actuarial analysis (Johnson et al. (1995) [92]). In all these cases observational study is involve. It is known that in observational studies with equal probability scientists, often cannot select sampling units. Well define sampling frames often do not exist for human, wildlife, insect, plant, or fish populations. Recorded observations on individuals in these populations are biased and will not have the original distribution unless every observation is given an equal chance of being recorded. The theory of weighted distributions provides a
unifying approach for correction of biases that exist in unequally weighted sample data. Also, the theory provides a means of fitting models to the unknown weighting function when samples can be taken both from the original distribution and the resulting bias distribution (McDonald (2010) [116]). A realization $x$ (the probability of recording the observation) of $X$ (the original observation) enters into the investigators record with probability proportional to a weight function $w(x)$ when the problem on weighted distribution is considered. The recorded $x$ is then not an observation of $X$, but rather an observation on a weighted random variable. Thus, the focus of this thesis is weighted extreme value distributions considering various weight functions.

The next problem is based on obtaining the order statistics of the weighted distributions. The theory of order statistics aims at deriving the distribution of order statistics for the said distributions. Order statistics arise naturally in many real-life applications involving data relating to life testing studies. In non-parametric statistics and inference, order statistics are among the most fundamental tool along with rank statistic. Minimum and maximum values of a sample are the special cases of the order statistics.

The study of percentage points of the newly developed probability distributions have been considered to find out the value of a random variable specified for a given probability, for which the random variable will be at, or below, with that probability.

Fillus and Fillus (2006) [51] introduced a new class of probability distributions as a linear combination of the available random variables. The new class of distributions has strong application in stochastic processes, finance and situations where application of actual distribution cannot be done (Shahbaz and Shahbaz (2009) [149]). This new class of distribution has been named by Fillus and Fillus (2006) [51] as the Pseudo-distributions. In milieu to concomitants of order statistics, the bivariate pseudo distribution has been extensively studied. The use of concomitant of order statistics arises in selection procedures when $k(< n)$ individuals are chosen on the basis of their $X$ values. Then the corresponding $Y$ values represent performance of an associated characteristic, i.e. concomitants are of interest in selection and prediction problems. To study a variable associated with another, distribution of the concomitants of order statistics are usually decisive. Hence the study of bivariate and trivariate Pseudo Frechet distribution and Pseudo weighted Weibull distribution have been considered along with the study of concomitants of order statistics for these distributions.

1.3 Literature Review

This section reviews different literatures to substantiate the findings of the current study. The review of various literatures has been included with a view to improve the research design on the basis of the outcomes of other researcher and to facilitate the interpretation of the results in a logical manner. The present review however, does not consider being exhaustive but endeavors have been made to refer most important and relevant literature on the study.
1.3. Literature Review

1.3.1 Review on Weighted Distribution

The distribution of the type $f_w(x) = \frac{w(x)f(x)}{W}$ where $W = \int w(x)f(x)dx$ with an arbitrary non-negative function $w(x)$ which may exceed unity has been introduced by Rao (1965) [142]. He gave practical examples where $w(x) = x$ or $x^\alpha$ were appropriate. He called distributions with arbitrary weight $w(x)$ weighted distributions. Fisher (1934) [52] introduced weighted distributions to model ascertainment bias which were later formalized in a unifying theory by Rao (1965)[142]. When the weight function is of the form $w(x) = x^\alpha$, such distributions are known as size-biased distributions of order $\alpha$. The most common cases of size-biased distributions occur when $\alpha=1$ or $2$; in the context of sampling, these special cases may be termed length and area-biased respectively. Patil and Taillie (1987) [137] calculated the Fisher information for certain exponential families, focusing primarily on $w(x) = x$ for nonnegative random variables. The weighted distribution concept of Patil et al. (1988) [136] can be traced to the study of the effect of methods of ascertainment estimation of frequencies by Fisher (1934) [52]. Gupta and Tripathi (1990) [68] studied the error made if an ordinary distribution is used instead of the length-biased version. Gupta and Tripathi (1996) [71] studied the weighted version of the bivariate three-parameter logarithmic series distribution, which has applications in many fields such as, ecology, social and behavioral sciences and species abundance studies. Relationships in the context of reliability were treated by several authors such as Gupta and Keating (1986) [69], Patil et al. (1986) [135], Jain et al. (1989) [90], Gupta and Kirmani (1990) [70], Oluyede and George (2002) [131] and several others. Results on size-biased distributions pertaining to parameter estimation in forestry, with special emphasis on the Weibull family has been reviewed by Gove (2003) [62]. Mir and Ahmad (2009) [119] discussed some size-biased probability distributions with their generalizations.

Weighted distributions have numerous applications. In ecology, Dennis and Patil (1984) [38] used stochastic differential equations to arrive at a weighted gamma distribution as the stationary probability density function for a stochastic population model with predation effects. Van Deusen (1986) [40] arrived at a size-biased distribution theory independently and applied it to fitting distributions of diameter at breast height (DBH) data arising from the horizontal point sampling (HPS) (Grosenbaugh (1958) [63]) inventories. Subsequently, Lappi and Bailey (1987) [111] used weighted distributions to analyze HPS diameter increment data. In fisheries, Taillie et al. (1995) [161] modeled populations of fish stocks using weighted distributions. Weighted distributions has been used by Magnusen et al. (1999) [114] to recover the distribution of canopy heights from airborne laser scanner measurements.

To introduce the concept of a weighted distribution, suppose $x$ is a non-negative random variable (rv) with its natural probability density function (pdf) $f(x)$. Let the weight function be $w(x)$ which is a non-negative function. Then the weighted density function is obtained as

$$g(x) = \frac{w(x)f(x)}{E[w(x)]} \quad x > 0$$

assuming that $E[w(x)] < \infty$ i.e the first moment of $w(x)$ exists.
1.3.2 Review on Weibull, Generalized Rayleigh and Frechet Distributions

The Weibull distribution is named after the Swedish physicist, Waloddi Weibull (1939 a, b) [169], [168]. In the Russian statistical literature this distribution is often referred to as the Weibull-Gnedenko distribution, since it is one of the three types of limit distributions of the sample maximum established rigorously by Gnedenko (1943) [61]. The name Frechet distribution is also used sometimes due to the fact that it was Frechet (1927) [56] who first identified this distribution to be an extermal distribution [Fisher and Tippett (1928) [54] later on showed to be one of three possible solutions]. The Weibull distribution includes the exponential and the Rayleigh distributions as special cases. The use of the distribution in reliability and quality control work was advocated by many authors following Kao (1958, 1959) [99],[100] and Berrettcm (1964) [18]. Situations where the Weibull distribution will likely arise has been mentioned by Gittus (1967) [59]. Malik (1975) [115] and Franck (1988) [55] has assigned some simple physical meanings and interpretations for the Weibull distribution, thus providing natural applications of this distribution in reliability problems particularly dealing with wearing styles. Hallinan (1993) [78] has provided an excellent review of the Weibull distribution by presenting historical facts, and the many different forms of this distribution as used by practitioners and possible confusions and errors that arise due to this non-uniqueness. Mudholkar et al. (1995) [120] introduced the exponentiated Weibull (EW) distribution, Xie and Lai (1995) [173] presented the additive Weibull distribution, Lai et al. (2003) [110] proposed the modified Weibull (MW) distribution and Carrasco et al. (2008) [25] defined the generalized modified Weibull (GMW) distribution. Based on Weibull distribution, various generalizations have been studied (Pham and Lai, 2007) [141].

The probability density function of the Weibull random variable \( X \) is given by (Johnson et al. (1994) [91])

\[
f(x) = c \cdot \left( \frac{x - \xi}{\alpha} \right)^{c-1} \cdot e^{-(\frac{x - \xi}{\alpha})^c}, \quad x > \xi, \quad c, \alpha > 0. \tag{1.3.2}
\]

Putting \( \xi = 0 \) yields the two-parameter Weibull distribution

\[
f(x) = \frac{c}{\alpha} \left( \frac{x}{\alpha} \right)^{c-1} e^{-\left(\frac{x}{\alpha}\right)^c}, \quad x > 0, \quad c, \alpha > 0. \tag{1.3.3}
\]

For the cases \( c = 1 \) and \( c = 2 \), Weibull distribution becomes the two parameter Exponential and Rayleigh distribution.

1.3. Literature Review


The Rayleigh distribution was originally derived by Lord Rayleigh (J. W. Strutt) in connection with a problem in the field of acoustics (Johnson et al. (1994) [91]). Miller (1964) [118] derived the Rayleigh distribution as the probability distribution of the distance from the origin to a point \((Y_1, Y_2, \ldots, Y_n)\) in \(N\)-dimensional Euclidean space, where the \(Y_i\)'s are independently and identically distributed \(N(0, \sigma^2)\) variables. The Rayleigh distribution is an important distribution in statistics and operations research. It is applied in several areas such as health, agriculture, biology, and other sciences. Substituting \(c = 2\) in equation (1.3.3) yields the Rayleigh distribution. Using \(\nu - 2\) in the pdf of the chi-distribution yields the Rayleigh density function (Johnson et al. (1994) [91]). If \(x_i, 1 \leq i \leq n\), are Gaussian variates and \(r^2\) is the sum of their squares, then we call \(r\) a generalized Rayleigh random variable. The probability distribution of the distance from the origin to a point \((Y_1, Y_2, \ldots, Y_n)\) in \(N\)-dimensional Euclidean space, where the \(Y_i\)'s are normal variables gives rise to the generalized Rayleigh distribution (Johnson et al. (1994) [91] p.452). Specifically, when \(Y_i\)'s are independent and identically distributed \(N(0, \sigma^2)\) variables, the probability density function of \(X = \sqrt{\sum_{i=1}^{n} Y_i^2}\) is given by

\[
  f(x) = \frac{2}{(2\sigma^2)^{\frac{N}{2}}} x^{N-1} e^{-\frac{x^2}{2\sigma^2}}, \quad x > 0, \ \sigma > 0. \tag{1.3.4}
\]

this generalized form of the Rayleigh distribution is also referred in the literature as the chi distribution with \(N\) degrees of freedom and scale parameter \(\sigma\).

The two-parameter generalized Rayleigh distribution is a particular member of the generalized Weibull distribution, originally proposed by Mudholkar and Srivastava (1993) [121], see also Mudholkar, Srivastava and Freimer (1995) [120]. Surles and Padgett (2001) [157] (Surles and Padgett; 2005 [158]) introduced a two-parameter Burr Type X distribution and correctly named as the generalized Rayleigh distribution.

The Frechet (extreme value type II) distribution is one of the probability distributions used to model extreme events. Frechet distribution is a special case of the generalized extreme value distribution. Frechet distribution was named after Maurice Frechet who wrote a related paper in 1927, further work was done by Fisher and Tippett [54] in 1928 and by Gumbel [66] in 1958. Frechet (1927) [56] had identified one possible limit distribution for the largest order statistics. Fisher and Tippett (1928) [54] showed that extreme limit distributions can only be one of three types. There are several applications of Frechet distribution ranging from accelerated life testing through to earthquakes, floods, rainfall, queues in supermarkets, sea currents, wind speeds and track race records.
The generalization of the standard Frechet distribution has been introduced by Nadarajah and Kotz (2003) [122]. Gupta and Nadarajah (2004) [67] introduced the beta Frechet (BF) distribution, which is a generalization of the exponentiated Frechet (EF) and Frechet distributions.

1.3.3 Review on Order Statistics

Order statistics play a very important role in statistical theory and methodology. Many authors have studied order statistics and associated inference, for example, David (1981) [32], Balakrishnan and Cohen (1991) [8], Arnold et al. (1992) [7], and David and Nagaraja (2003) [37]. Khatri (1962) [104] examined the probability function and the density function of a single order statistics, the joint probability function and density function of any two order statistics and joint density function of any three order statistics of independent and identically distributed (i.i.d.) random variables from a discrete parent. Vaughan and Venables (1972) [165] denoted the joint probability density function and marginal probability density function of order statistics of independent and not necessarily identically distributed (i.n.i.d.) random variables by means of permanents. David (1981) [32] considered the fundamental distribution theory of order statistics. Guilbaud (1982) [65] expressed probability of the functions of independent and not necessarily identically distributed random vectors as a linear combination of probabilities of the functions of independent and identically distributed random vectors and thus also for order statistics of random variables. Reiss (1989) [144] considered the joint probability density function, marginal probability density function and density function of any order statistics of i.i.d. random variables under a continuous density function and discontinuous density function. He also considered a probability density function of bivariate order statistics by marginal ordering of bivariate i.i.d. random vectors with a continuous density function by means of multinomial probabilities of appropriate cell frequency vectors, defining multivariate order statistics by marginal ordering of i.i.d. random vectors with a continuous density function. Balasubramanian et al. (1994) [13] established identities satisfied by distributions of order statistics from non-independent non-identical variables through operator methods based on difference and differential operators. Gan and Bain (1995) [57] obtained the joint probability function of any order statistics and also conditional distributions of discrete order statistics from a general discrete parent by tie-runs.

Tippett (1925) [163] tabulated, for certain sample sizes ranging from 3 to 1000, the cumulative distribution function of the largest order statistic in a sample from a normal population having zero mean and unit variance. Craig (1932) [30] gave general expressions for the exact distribution functions of the median, quartiles, and range of a sample size of \( n \). Daniels (1945) [31] has made an interesting application of the sampling theory of order statistics to develop the probability theory of breaking strength bundles of threads. Hastings et al. (1947) [81] has given the means, variances, covariances and correlations of order statistics in samples of ten or less from normal populations. Gupta (1960) [72] derived a recurrence relation for the expected value of the \( r \)-th order statistics for different values of \( m \) \( (E(X_{m}^{r}) \) for integer values of the shape parameter \( \alpha \). Gupta (1960) [72] used this relation to tabulate values of \( E(X_{m}^{r}) \) for various combinations of \( m \), \( n \) and \( \alpha \). Gupta (1960) [72] also discussed some illustrative applications to life testing and relia-
bility problems. Krishnaiah and Rizvi (1967) [107] extended the work of Gupta (1960) [72] with any positive shape-parameter. Breiter and Krishnaiah (1968) [23] tabulated the values of $m = 1, 2, 3, 4$ for various values of $\alpha$ obtained by using the recurrence relations by Krishnaiah and Rizvi (1967) [107]. Joshi (1969) [95] discussed methods for obtaining approximations and bounds for the moments of order statistics from a continuous parent distribution. Joshi (1969) [95] also showed that for the Cauchy distribution bounds and approximations of all finite moments can be obtained. Simple moment relations for the order statistics of independent standardized gamma variables and the order statistics of the symmetrical inverse multinomial distribution was presented by Young (1971) [176]. Recurrence relations between the moments of order statistics from the exponential and right truncated exponential distributions has been obtained by Joshi (1978) [96]. Several recurrence relations and identities for product moments of order statistics in a random sample of size $n$ from an arbitrary continuous distribution has been obtained by Joshi and Balakrishnan (1982) [94]. Balakrishnan and Malik (1986) [10] established some recurrence relations of order statistics from the linear-exponential distribution. Several recurrence relations and identities for the single and product moments of order statistics from some specific distributions have been reviewed by Balakrishnan et al. (1988) [12]. Sobel and Wells (1990) [155] showed that $E(X_{(m)}^n)$ can be expressed in terms of Dirichlet integrals (integrals involving gamma functions) and provided a table for reading Dirichlet integrals. Beg (1991) [17] generalized the results of Joshi and Balakrishnan (1982) [94] when the variables are independent but not assumed to be identically distributed using permanents. Recurrence relations for the single and product moments of order statistics from the doubly truncated parabolic and skewed distribution and linear-exponential distribution has been presented by Mohie El Din et al. (1997, 1991) [47], [46]. Hendi et al. (2006) [83] developed recurrence relations for the single and product moments of order statistics from doubly truncated Gompertz distribution. Explicit closed form expressions are derived for moments of order statistics from the chi square distribution by Nadarajah (2008) [123]. Order statistics from an exponentiated gamma (EG) distribution has been obtained by Shawky and Babokan (2009) [154]. They also derived exact expression for the single and double moments of order statistics from EG distribution.

Let $(X_1, X_2, \ldots, X_n)$ be $n$ independently and identically distributed variates. Let the variables be arranged in non-decreasing order and let $X_{(i)}$ denote the $i$-th order statistics. The smallest of the $X_i$s is denoted by $X_{1:n}$, the second smallest is denoted by $X_{2:n}$, and, finally, the largest is denoted by $X_{n:n}$. Thus $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$. The subject of order statistic deals with the properties and applications of these order statistics and functions involving them.

If the population is absolutely continuous, then the probability density function of $X_{(i:n)}$ is defined as (Balakrishnan and Rao (1998) [11])

$$f_{(i:n)}(x) = \frac{n!}{(i-1)!(n-i)!} [F(x)]^{i-1} [1 - F(x)]^{n-i} f(x), \quad -\infty < x < \infty$$

(1.3.5)

The pdf's of the smallest and largest order statistics are

$$f_{1:n}(x) = n [1 - F(x)]^{n-1} f(x), \quad -\infty < x < \infty,$$

$$f_{n:n}(x) = n [F(x)]^{n-1} f(x), \quad -\infty < x < \infty$$

(1.3.6)
The cumulative distribution function (cdf) of $X_{i:n}$ is given by (Arnold et al. (1992) [7])

$$F_{i:n}(x) = I_F(x; i, n - i + 1), \quad -\infty < x < \infty$$  \hfill (1.3.7)

which is just Pearson’s (1934) [140] incomplete beta function. The cdf’s of the smallest and largest order statistics are

$$F_{1:n}(x) = 1 - [1 - F(x)]^n, \quad -\infty < x < \infty,$$

$$F_{n:n}(x) = [F(x)]^n, \quad -\infty < x < \infty.$$  \hfill (1.3.8)

If the population is absolutely continuous, then the joint pdf of $X_{i:n}$ and $X_{j:n}$ ($1 \leq i < j \leq n$) is given by (Balakrishnan and Rao (1998) [11])

$$f_{i,j:n}(x_i, x_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \left[ F(x_j) - F(x_i) \right]^{j-i-1} \frac{f(x_j)}{1 - F(x_i)}, \quad x_i < x_j$$

The joint density function of the smallest and largest order statistics are

$$f_{1:n:n}(x_1, x_n) = n(n - 1) [F(x_n) - F(x_1)]^{n-2} f(x_1)f(x_n), \quad x_1 < x_n$$  \hfill (1.3.10)

The conditional density function of $X_{j:n}$ given that $X_{i:n} = x_i$ is given by (Arnold et al. (1992) [7])

$$f_{j:n}(x_j \mid X_{i:n} = x_i) = \frac{(n-i)!}{(j-i-1)!(n-j)!} \left[ \frac{F(x_j) - F(x_i)}{1 - F(x_i)} \right]^{j-i-1} \frac{f(x_j)}{1 - F(x_i)}, \quad 1 < j \leq n, \quad x_i \leq x_j < \infty$$  \hfill (1.3.11)

Considering the sample size $n$ to be odd, the pdf of the sample median $X_{n} = X_{(n+1)/2}$ is given by (Arnold et al. (1992) [7])

$$f_{X_{n}}(x) = \frac{n!}{[(n-1)/2]^2} \left[ F(x)(1 - F(x)) \right]^{n-1} f(x), \quad -\infty < x < \infty$$  \hfill (1.3.12)

The sample range $W_{n}$ may be defined as $W_{n} = X_{n:n} - X_{1:n}$. The pdf of the sample range may be written in the form (Arnold et al. (1992) [7])

$$f_{W_{n}}(w) = n(n - 1) \int_{-\infty}^{\infty} [F(x_1 + w) - F(x_1)]^{n-2} f(x_1)f(x_1 + w)dx_1, \quad 0 < w < \infty$$  \hfill (1.3.13)

The sample range can be generalized to the spacing $W_{i,j:n} = X_{j:n} - X_{i:n}$, $1 \leq i < j \leq n$. The $i$-th quasirange $W_{i:n} = X_{n+1-i:n} - X_{i:n}$ is a special case of $W_{i,j:n}$, and hence a spacing is sometimes called a generalized quasirange. The pdf of $W_{i,j:n}$ is given by (Arnold et al. (1992) [7])

$$f_{W_{i,j:n}}(w) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \int_{-\infty}^{\infty} \{F(x_i + w) - F(x_i)\}^{i-1} \{1 - F(x_i + w)\}^{n-j} f(x_i)f(x_i + w)dx_i, \quad 0 < w < \infty.$$  \hfill (1.3.14)
1.3. Literature Review

1.3.4 Review on Percentage Points

Different procedures for obtaining the percentage points of a distribution have been proposed in the literature. Pearson (1934) [140] has provided extensive tables of the incomplete Beta distribution. Cochran (1940) [28] has given an approximate formula for the percentage points of the incomplete Beta distribution, for large values of its parameters. Thompson (1941) [162] tabulated the percentage points of the incomplete Beta distribution accurate to five significant figures for \( v_1 = 1(1)10, 12, 15, 20, 24, 30, 40, 60, 120; v_2 = 1(1)30, 40, 60, 120, \infty \); \( \alpha = 0.005, 0.01, 0.025, 0.05, 0.10, 0.25, 0.50 \) where \( v_1 \) and \( v_2 \) denote the parameters of the Beta distribution and \( \alpha \) denotes the percentage. Eisenhart et al. (1948) [45] tabulated accurate to five significant figures, the percentile points of the median in samples from the uniform distribution of sizes \( N = 3(2)15(10)95 \) and for \( \alpha = 0.001, 0.005, 0.01, 0.025, 0.05, 0.10, 0.20, 0.25 \). The National Bureau of Standards (1953)[128] has published tables of \( F(X) - F(-X) = 2P - 1 \) to 15 decimal places for \( X=0(0.0001)1(0.001)7.8 \). The National Bureau of Standards (1964) [129] also has published tables of \( P = F(X) \) to 15 decimal places for \( X = 0(0.02)3(0.05)5 \). Interpolation in the National Bureau of Standards tables is time consuming and complicated. Leone et al. (1960) [112] tabulated the percentile points of the Binomial distribution for sample sizes \( N = 10(5)100 \) and \( \alpha = 0.0025, 0.005, 0.01, 0.025, 0.05, 0.1, 0.9, 0.95, 0.975, 0.99, 0.995, 0.9975 \). Tables of percentage points of the maximum signed value of the multivariate Student \( t \) distribution have been given in an unpublished report by Krishnaiah and Armitage (1966) [106] for all combinations of \( k = 1(1)10, \nu = 5(1)35, \rho = 0.05(0.05)0.9 \) and \( \gamma = 0.90, 0.95, 0.975 \) and 0.99. These tabulations have been published in the literature for \( \gamma = 0.95 \) and \( \gamma = 0.99 \) (Krishnaiah & Armitage (1966) [106]). Hahn (1969) [75] provided tabulations of factors for constructing one-sided and two-sided simultaneous prediction intervals given the values of a past sample of \( n \) observations from the same normal distribution. Chew (1968) [27] suggested two approximate procedures for obtaining such intervals. White (1970) [170] have given two 20 decimal place tables of percentile points of the normal distribution. A table of 100\( \gamma \) percentage points of the maximum absolute value \( |t| \) of the \( k \)-variate Student \( t \) distribution with \( \nu \) degrees of freedom and common correlation \( \rho \) was given for various values of \( k, \nu, \rho \) and \( \gamma \) by Hahn and Hendrickson (1971) [76]. The application of these tables to problems dealing with the construction of simultaneous confidence and prediction intervals was also briefly described by them.

Normal approximation to the percentage points of the \( \chi^2 \) distribution has been presented by Haldane (1938) [77]. Sixteen formulae for approximating percentage points of \( \chi^2 \) were examined at each of 15 significance levels by Zar (1978) [177]. Fisher & Cornish (1960) [53] suggested an approach based on expansion of the kurtosis, \( b_2 \), in normal sampling equals for a percentage point in terms of the cumulants and the corresponding normal deviate. Eeden (1961) [43] dealt with some approximations to the percentage points of the noncentral \( t \)-distribution. Most of these approximations are known in literature; some of them however were only known for the special case of the central \( t \)-distribution (Student's distribution), but could be generalized for the non-central \( t \)-distribution. Pearson (1965) [139] has given the type IV percentage point approximants to the distribution of the kurtosis, \( b_2 \), in normal sampling. Bowman and Shenton (1979) [22] have given approximate formulae for a set of percentage points of the Pearson system in terms of the skewness
and kurtosis.

Krutchkoff (1986) [108] addressed the issue of computer storage and time by introducing an algorithm that employed a probabilistic technique based on the normal approximation to the binomial distribution. Studies of the simulation literature conducted by Hoaglin and Andrews (1975) [86] and Hauck and Anderson (1984) [82] have shown a general absence of documentation of precision and justification for the number of iterations. Whether to use a single large-sample estimator for a percentile point or to combine a few smaller-sample estimators was considered by Juritz et al. (1983) [97] and Zelterman (1986) [179]. Dunn (1991) [42] presented a percentile point simulation algorithm. The algorithm is useful when computer storage and time considerations are at a premium. The algorithm employs various times and storage-saving ideas, including a pinching mechanism that reduces the proportion of simulated values stored as the number of iterations is increased. Hudak and Tiryakioglu (2009) [87] used geometric interpolation method to develop the distributions of percentiles when the shape and scale parameters differ from one. Further, since in practice, the shape and scale parameters that would be used in this interpolation are themselves estimates from data, Hudak and Tiryakioglu (2009) [87] used a step-by-step procedure for determining the distribution for the true percentiles.

Tables of percentage points for the distribution of order statistics has been prepared by Gupta (1960) [72], (1961) [73] one each for the standard normal and the standard gamma distribution and one prepared by Gupta and Shah (1965) [74] for the standard logistic distribution. Guenther (1977) [64] showed that the percentage points of order statistics associated with sampling from any standard distributions can be found easily and quickly making tables and special computer programs unnecessary.

1.3.5 Review on Pseudo Distribution and Concomitants of Order Statistics

A new class of continuous distribution has been defined by Filus and Filus (2000) [49] as a linear transformation on the Euclidean space \( R^n \rightarrow R^n \) for independent normally distributed random variables. The output of such a transformation is a random vector such that its density turns out to be a generalization of the multivariate Gaussian distribution and is called Pseudo Gaussian distribution. Filus and Filus (2001) [50] had further given a description of the distributions. A new class of multivariate distributions that contain Pseudo Weibull and Pseudo Gamma distributions have also been introduced by Filus and Filus (2006) [51]. Shahbaz et al. (2009) [152] introduced the pseudo Exponential distribution by compounding two exponential random variables. A new class of probability distributions known as bivariate Pseudo Rayleigh distribution has been defined by Shahbaz and Shahbaz (2009) [149]. The common properties of the distribution has been discussed by them. The trivariate pseudo Rayleigh distribution as a compound distribution of the three random variables was introduced by Shahbaz and Shahbaz (2011) [151].

In the framework of concomitants of order statistics, the bivariate pseudo distributions have been extensively studied. Spruill and Gastwirth (1982) [156] have used concomitants to estimate the correlation coefficient between two sensitive variables, data on which are kept separate, and merge of the data is not possible due to confidentiality
considerations. To allow for the selection based on more than one characteristic, Egorov and Nezvorov (1984) [44], Reiss (1989) [144], and Kaufmann and Reiss (1992) [101] made a further generalization by considering the ordering of more than a single $X$. The concomitants of order statistics for the situation in which the random vectors are independent but otherwise arbitrarily distributed was considered by Eryilmaz (2005) [48]. Yeo and David (1984) [175] considered the problem of choosing the best $k$ objects out of $n$ candidates on the basis of auxiliary measurements $X$, while the measurements of primary interest $Y$ are not available. The distribution of the $r$-th concomitant and joint distribution of $r$-th and $s$-th concomitants of order statistics for bivariate Pseudo-Weibull distribution has been obtained by Shahbaz and Ahmad (2009) [148]. The concomitants of order statistics find application in dealing with the estimation of parameters for multivariate data sets that are subject to some form of type II censoring; examples include Harrell and Sen (1979) [80], and Gill et al. (1990) [58]. The concomitants of order statistics also find application in the ranked-set sampling, which was, first introduced by McIntyre (1952) [117]. A comprehensive review of rank set sampling can be found in Wolfe (2004) [172], and in Chen et al. (2004) [26]. Abo-Eleneen (2001) [1] and Abo-Eleneen and Nagaraja (2002) [2] studied Fisher Information in pairs and collections of order statistics and their concomitants from bivariate samples. An excellent review of work on concomitants of order statistics is available in David and Nagaraja (1998) [36].

The finite-sample distribution theory for concomitants of order statistics has been investigated by several authors, for example by David (1973) [35], David et al. (1977) [34], Yang (1977) [174], Balasubramanian and Beg (1998) [14], Eryilmaz (2005) [48]. David and Galambos (1974) [33] showed that for any fixed $k$ and any choice $1 \leq r_1 < \ldots < r_k \leq n$, $Y_{[r_1;n]}, \ldots, Y_{[r_k;n]}$ are asymptotically independent provided that $\text{Var} [E (Y_{[r_i;n]} | X_{[r_i;n]})]$ approaches 0 as $n \to \infty$ for all $i = 1, 2, \ldots, k$. For the Gumbel's bivariate Exponential distribution, Balasubramanian and Beg (1998) [14] found the distribution of the concomitant of the $r$-th order statistic of one of the components. Begum and Khan (1999) [6] dealt with concomitants of order statistics from bivariate Pareto distribution. The probability density function of the $r$-th and joint pdf of the $r$-th and $s$-th concomitants of order statistics were derived for bivariate Inverse Rayleigh distribution by Aleem (2006) [5]. Tahmasebi and Behboodian (2010) [160] investigated the interesting properties of Shannon entropy for concomitants of order statistics in Generalized Morgenstern sub-family. Tahmasebi and Behboodian (2010) [160] also obtained a general expression of entropy for concomitants of order statistics in bivariate Pseudo-Weibull distribution. Shahbaz et al. (2010) [150] obtained the distribution of vector of two concomitants and joint distribution of order statistics for pseudc Exponential distribution. Expression for $p$-th moment of vector concomitants and product moment for the pair of vectors of concomitants also has been obtained by Shahbaz et al. (2010) [150]. The distribution of a pair of concomitants of order statistics for the trivariate pseudo Rayleigh distribution has been derived by Shahbaz and Shahbaz (2011) [151].

If $X_i, Y_i$ ($i = 1, 2, \ldots, n$) are $n$ independent random variables from some bivariate distribution function $F(x, y)$ and if the random variables are ordered by the $X$-variates as $x_{[1:n]} \leq x_{[2:n]} \leq \ldots \leq x_{[n:n]}$, then the $y$-variates associated with $X_{[k:n]}$, denoted by $y_{[1:n]}, y_{[2:n]}, \ldots, y_{[n:n]}$, are called the concomitant of the order statistic (David (1973)[35]) or induced order statistics (Bhattacharya (1974)[20]). The distribution of concomitants of
order statistics for a random sample of size \( n \); drawn from a bivariate distribution function \( F(x, y) \); has been defined by David and Nagaraja (2003) [37] as

\[
f_{[r:n]}(y) = \int_{-\infty}^{\infty} f(y \mid x)f_{r:n}(x)dx
\]  
(1.3.15)

\( f(y \mid x) \) is the conditional distribution of \( Y \) given \( X \). \( f_{r:n}(x) \) is the distribution of \( r \)-th order statistics. The distribution of \( r \)-th order statistics has been defined by David and Nagaraja (2003) [37] as

\[
f_{r:n}(x) = \frac{n!}{(r-1)!(n-r)!} f(x)[F'(x)]^{r-1}[1 - F(x)]^{n-r}
\]  
(1.3.16)

The distribution (1.3.15) was extended by Shahbaz and Shahbaz (2010) [150] when a random sample is available from trivariate distribution function \( F(x, y_1, y_2) \). The joint distribution of \( r \)-th and \( s \)-th order statistics is given as

\[
f_{r,s:n}(x_1, x_2) = C_{r,s:n} f(x_1)f(x_2)[F(x_1)]^{r-1}[F(x_2) - F(x_1)]^{s-r-1}
\]  
(1.3.17)

\( C_{r,s:n} = \frac{n!}{(r-1)!(s-r)!(n-s)!} \)

The joint distribution of \( r \)-th and \( s \)-th concomitant is given as

\[
f_{[r,s:n]}(y_1, y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{x_2} f(y_1 \mid x_1)f(y_2 \mid x_2)f_{r,s:n}(x_1, x_2)dx_1dx_2
\]  
(1.3.18)

Shahbaz and Shahbaz (2010) [150] have argued that the joint distribution of two bivariate concomitant can be derived by using

\[
f_{[r,s:n]}(y_1, y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{x_2} f(y_1, y_2 \mid x)f_{r,s:n}(x_1, x_2)dx_1dx_2
\]  
(1.3.19)

where \( y_1 = [y_{1r}, y_{2r}]', \ y_2 = [y_{1s}, y_{2s}]' \).
1.4 Objectives of the study

Keeping in view the importance of the study, the present research has been undertaken with the objectives given below.

- To develop new probability distributions based on weighted concept and study the various properties of the distributions. To determine the behavior of the hazard function and study the characterization of the developed distributions. In order to justify the suitability of the developed distributions, comparisons to be made between the developed weighted distributions with the original distributions using real life data.

- To study the order statistics of the weighted distribution.

- To study the percentage points of the weighted distributions along with the percentage points of the smallest and the largest order statistics for the weighted distributions.

- To develop bivariate and trivariate Pseudo Frechet distribution and Pseudo weighted Weibull distribution as a compound distribution of two and three random variables.

- To study the concomitant of order statistics for bivariate and trivariate Pseudo Frechet distribution and Pseudo weighted Weibull distribution.

1.5 Outline and Organization of the study

For clarity and better insight into the subject matter, this thesis has been divided into seven chapters each consisting of several subsections.

Chapter 1 Introduction

Chapter 2 Weighted Weibull Distribution

Chapter 3 Weighted Generalized Rayleigh Distribution

Chapter 4 Order Statistics of Weighted Distributions

Chapter 5 Percentage Points of Weighted Distributions

Chapter 6 Concomitants of Order Statistics for Pseudo Extreme Value Distributions

Chapter 7 Summary and Conclusion
The subject matter contained in different chapters is briefed below.

Chapter 1, which is introductory in nature, presents a brief conception of extreme value theory and concept of weighted distribution, some literature review, importance as well as the objectives of the study.

Chapters 2 and 3 deals with the development of new probability models based on Rao's (1965) [142] weighted concept. The probability density functions, cumulative distribution functions, reliability or survival functions, hazard functions of weighted Weibull distribution (WWD), length-biased weighted Weibull distribution (LBWWD), weighted Generalized Rayleigh distribution (WGRD) and length-biased weighted Generalized Rayleigh distribution (LBWGRD) have been studied. The moments of these distributions have been derived which has been used to compute the mean, variance, standard deviation and coefficients of variation, skewness and kurtosis. We have used Lingappaiah's (1988) [113] result for the characterization of the WWD and WGRD. Differential entropy has been obtained for these distributions. The suitability of the stated distributions have been justified with the help of real-life data.

Chapter 4 has been devoted to the study of the order statistics for the distributions developed in chapters 2 and 3 for independently and identically distributed cases only. The pdf of single order statistics $X_{i:n}$, cdf of $X_{i:n}$, joint pdf of $X_{i:n}$ and $X_{j:n}$, modal equations, single and product moments of the $i$-th order statistics, conditional density function of order statistics for weighted Weibull distribution, LBWWD, WGRD and LBWGRD have been derived. For the weighted Weibull distribution, LBWWD, WGRD and LBWGRD, we have obtained the pdf of sample median when the sample size is odd, the range and quasirange.

Chapter 5 deals on obtaining the percentage points of the weighted Weibull distribution and weighted Generalized Rayleigh distribution. For the weighted Weibull distribution and weighted Generalized Rayleigh distribution, we have computed the percentage points of the smallest and the largest order statistics.

Chapter 6 deals with the development of new classes of probability distributions namely bivariate Pseudo Frechet distribution and bivariate Pseudo weighted Weibull distribution based on the compound distribution technique. For the bivariate Pseudo Frechet distribution and bivariate Pseudo weighted Weibull distribution, we have derived the joint moments, ratio moments, moment generating function (mgf), the distribution and moment of the $r$-th concomitant and expression for the joint distribution of the $r$-th and $s$-th concomitant by using various transformations. The bivariate Pseudo Frechet distribution and bivariate Pseudo weighted Weibull distribution has been extended to trivariate Pseudo Frechet distribution and trivariate Pseudo weighted Weibull distribution using the same methodology. For the bivariate and trivariate Pseudo Frechet distribution, we have derived the joint moments, mgf, conditional moments, the distribution and moment of the $r$-th concomitant and expression for the joint distribution of the $r$-th and $s$-th concomitant. Expressions for the mgf of $r$-th concomitant and product moments of $r$-th and $s$-th concomitant for the bivariate and trivariate Pseudo Frechet distribution have also been obtained.

Chapter 7 summarizes and concludes the thesis pointing out the main findings and possible future work. A detailed bibliography of the relevant literature is given at the end.