CHAPTER – 1

Introduction and Summary:

1.1. Introduction:

The research work in the area of modeling and fitting of distributions to data sets has been receiving greater attention in the recent years and seems to have explosive growth. Usefulness of such work has been realized in almost all fields including science, technology, engineering, management, business, health and social sciences. In view of this, proposed new models and their applications / analysis will be appreciated. These will contribute for the overall development of all branches of studies and applications.

A model is a mathematical expression whose structure and hence behavior corresponds in some sense to a particular reality or phenomenon. In other words statistical model is a probability distribution constructed to enable inferences to be drawn or decision to be made on the basis of data. When the data collected or the experiment conducted is under ideal conditions, a well-known model may give a good fit for the data set or the outcomes of the experiment. These fitted distributions describe the random behavior of the variable involved and can be used for decision making. When the situation is not ideal due to presence of some distorting factor, standard models may not
give a good fit. One possible technique to take an account for disturbing factors is to modify or extend the existing model. The main objective of the work reported in this thesis is to introduce new statistical models, essentially derived from well-known families of distributions and carry out their analysis.

A new class of distributions can be generated by using a distribution or class of distributions. Some well-known methods to generate a class of distributions are, transformation of variables, use of conditional distributions, introducing of additional parameters or the use of copulas. For example if X has exponential distribution with mean 1 then $Y = X^\alpha$ has Weibull distribution. If X and Y are i.i.d. continuous random variables having density $f$, symmetric about 0 then the density corresponding to the conditional density of X given $Y < m X$ ($-\infty < m < \infty$) is

$$\frac{f(x) F(mx)}{\int_{-\infty}^{\infty} f(x) F(mx) \, dx}, \quad -\infty < x < \infty.$$ 

In particular if $m = 0$, then $F(mx) = 1/2$ for all $x \in \mathbb{R}$ and the above reduces to $f(x)$. If $X$ has density $g(x)$ then the density corresponding to the truncated variable with support $A$ is

$$f(x) = \begin{cases} 
\frac{g(x)}{P(A)}, & \text{if } x \in A \\
0, & \text{otherwise}
\end{cases}$$

where $P(A) = \int_A g(x) \, dx$.

By introducing an additional parameter $\varepsilon$, Mudholkar and Hutson (2000) have proposed the epsilon–skew-normal distribution. Copulas can be used for constructing families of bivariate distributions. Copula C is a function on
[0, 1]^2 to [0, 1], itself is a distribution function on [0, 1]^2, so that for given marginals F and G the function H(x, y) = C (F(x), G(y)) is a distribution function. For example if \( \theta \in [-1, 1] \), then the function

\[
C_{\theta} (u, v) = uv + \theta uv(1-u)(1-v)
\]

is a one parameter family of copulas. A generalization of this family can be found in Quesada and Rodriguez (1995). Another popular technique of generating distributions is to use the technique of mixing two distributions. If \( H_1(x) \) and \( H_2(x) \) are two distribution functions then \( H(x, \theta) = \theta H_1(x) + (1-\theta) H_2(x) \), \( 0 \leq \theta \leq 1 \) is a distribution function.

In this thesis we introduce Contour-Transformation, a technique based on geometric approach and use it to generate new classes of distributions. Contour Transformation (CT) technique is different from the above mentioned methods in the sense that it may not be possible to obtain the density function obtained by a CT of a given density by using any of the above well-known methods. Also we introduce ‘Selfininvertible densities’ a typical class of densities on the support \((0,\infty)\). Further using length (Lebesgue measure) of contours of a density \( f \), we defined a density \( h \) being referred as the \( f \)-contour density. Multivariate versions of these are also considered whenever possible and certain properties have been proved.

### 1.2. Review of Literature:

In the literature some methods to modify a model to give a better fit for the data set have been reported. Marshal and Olkin (1997) have extended the class of exponential and Weibull family by introducing an additional parameter \( \alpha \). For a given distribution function \( F \), let \( \bar{F}(x) = 1 - F(x) \). The new model proposed by Marshal-Olkin is given by
\[
\tilde{G} = \frac{\alpha \tilde{F}(x)}{1 - (1 - \alpha) \tilde{F}(x)}, -\infty < x < \infty, \quad 0 < \alpha < \infty.
\]

\[
1 - \frac{F(x)}{\alpha + (1 - \alpha)F(x)}
\]

Here we note that \( G(x) = \frac{F(x)}{\alpha + (1 - \alpha)F(x)} \). For \( \alpha > 0 \), the function \( 1/G(x) = (1 - \alpha) + \alpha /F(x) \) is non-increasing function and \( G(-\infty) = 0, G(\infty) = 1 \), the function \( G \) is a distribution function.

Clearly when \( \alpha = 1 \) we get the original distribution function. Corresponding density function is given by

\[
g(x; \alpha) = \frac{\alpha f(x)}{\{1 - (1 - \alpha) \tilde{F}(x)\}^2}, -\infty < x < \infty, \quad 0 < \alpha < \infty
\]

and the hazard rate function is given by

\[
h(x; \alpha) = \frac{h_F(x)}{\{1 - (1 - \alpha) \tilde{F}(x)\}}, -\infty < x < \infty, \quad 0 < \alpha < \infty
\]

where \( h_F(x) \) denotes the hazard rate function of original model with distribution function \( F \). Marshal-Olkin Weibull distribution is useful to measure the mortality rate due to the spreading of diseases and unexpected natural calamities. Marshal-Olkin semi Weibull distribution and generalized Weibull distribution are studied by Jose and Alice (2005). Ghitany, et al. (2005) gives a detailed study of Marshal-Olkin Weibull distribution that can be obtained as a compound distribution mixing with exponential distribution.

In recent days there has been growing trend to propose flexible parametric families of asymmetric distributions to incorporate possible asymmetry being exhibited in the data sets. Skewed distributions play an important role in the analysis of the data relating to reliability, survival studies, life span, strength and hardness of the materials. They are also useful
as model for distribution of sales, income, insurance premiums and claims. It is well known that symmetric distributions are not suitable for modeling all types of data. Though there are many techniques to generate skew models it is desirable to use those leading to skew models with nice properties.

Many researchers have developed different methods to construct skewed distributions. A simple departure from the normal distribution has been proposed by Azzalini (1985) who defined the skew-normal distribution with p.d.f.

\[
2 \phi(x; \mu, \sigma) \Phi(\alpha (x - \mu))
\]

where \( \phi(x; \mu, \sigma) \) is the p.d.f. of a normal random variable with mean \( \mu \) and variance \( \sigma^2 \), and \( \alpha \) is a shape parameter controlling skewness. For \( \alpha = 0 \), the p.d.f. (1) reduces to the normal one, whereas for \( \alpha > 0 \) or \( \alpha < 0 \) the p.d.f. is skewed to the right or to the left, respectively. An extension of (1) to the multivariate setting was proposed by Azzalini and DallaValle (1996), defining the p.d.f.

\[
2 \phi_p(x; \mu, \Sigma) \Phi(\alpha^T (x - \mu)), \quad X \in \mathbb{R}^p,
\]

where, \( \phi_p(x; \mu, \Sigma) \) is the \( p \)-dimensional normal p.d.f. with mean \( \mu \) and correlation matrix \( \Sigma \), \( \Phi(.) \) is the c.d.f. of standard normal variate and \( \alpha \) is a \( p \)-dimensional shape parameter. A \( p \)-dimensional random vector \( X \) with a multivariate skew-normal distribution is denoted by \( X \sim SNp(\mu, \Sigma, \alpha) \). Its expectation and variance are given by \( E(X) = \mu + \sqrt{2} \delta/\pi \) and \( \text{Var}(X) = \Sigma^{-}(2/\pi) \delta \delta^T \) where, \( \delta = \Sigma \alpha / \sqrt{1 + \alpha^T \Sigma \alpha} \). Corresponding multivariate extensions were developed in Azzalini (1986), Azzalini and Capitano (1999) and Azzalini (2005). For historical development and various standard statistical models and analysis of well-known statistical models one may refer to Johnson, et al. (1994, 1995, 1997), Rao (1965) and Davison (2003) and

\[ f(z) = K_m^{-1} \phi(z | \mu, \Sigma) Q_m(z), \]

where \( Q_m \) is interpreted as skewing function and \( K_m = E[Q_m(Z)] \). Azzalini (1986) handled skewness and heavy tails simultaneously by proposing skew exponential power distribution which is more flexible and incorporate a wide range of models near to the normal distribution. Fernandez, et al. (1995) have proposed \( \nu \)-spherical multivariate densities and studied their robustness properties. Mudholkar and Hutson (2000) have introduced ‘The epsilon-skew normal distribution for analyzing near normal data’. It is a unimodal distribution with mode at 0 and probability mass \((1+\varepsilon)/2\) below the mode. Its p.d.f. is of the form

\[
f_\varepsilon(x) = \begin{cases} 
\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2(1+\varepsilon)^2}\right) & \text{if } x < 0 \\
\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2(1-\varepsilon)^2}\right) & \text{if } x \geq 0 
\end{cases}
\]

where \(-1 < \varepsilon < 1\).

It gives better fit for data sets of heights of 219 volcanoes. Deshmukh and Nabar (2006) proposed the epsilon skew exponential power distribution. Salinas et al. (2007) gave an extension of the skew exponential power distribution. The general family of asymmetric probability density function is introduced by Arellano-Valle, et al. (2004). Fernandez and Steel (1998) have proposed a technique of generating unimodal skewed distributions from a symmetric unimodal distributions by using a scalar parameter, known as the skew parameter \( k \). The model is given by
where, \( k > 0 \). Dey and Liu (2005) have proposed a new class of skew-elliptical distribution which is based on linear constraints and linear combinations. Wang and Genton (2006) obtained multivariate skew-slash distribution which combines skewness with heavy tails. Arnold, et al. (2008) have obtained a general family of multivariate densities with given contours and these include circular and elliptical densities not necessarily having the same centers, but have not used transformation of the contours to generate densities. Asymmetric Weibull distribution plays an important role in reliability analysis and is used to model the breaking strength of materials in quality control analysis, for more details one can refer to Juric and Kozubowski (2004). By using copulas, a class of multivariate distribution can be obtained with specified marginals, for further details one may refer to a review article by Lai (2004). Rattihalli and Basugade (2008) have generated a class of multivariate densities by using contour transformation. For further details about asymmetric distributions and their relevance to practical problems one may refer to Kotz, et al. (2001). Nelder and Wedderburn (1972) were the first to show, by introducing the class of generalized linear models that a large variety of non-normal data may be analyzed by simple technique. Generalized Linear Models were originally developed for exponential family of distributions, but these ideas can be extended to wider class of models called dispersion models, for the further details one may refer to Jorgensen (1997). Karian and Dudewicz (2000) described fitting of Generalized Lambda Distributions (GLD) through different methods. The family of Generalized Lambda Distributions with parameters \( \lambda_1, \lambda_2, \lambda_3, \lambda_4 \) is specified in terms of its percentile function as
\[ Q(y; \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{\lambda_2}{\lambda_3 y^{\lambda_3 - 1} + \lambda_4 (1 - y)^{\lambda_4 - 1}} \text{ at } x = Q(y). \]

For further details one can refer to Karian and Dudewiez (2000).

When \( X \) has unbounded support, by conditioning on a bounded set, the natural parameter space can be enlarged yielding a new family of distributions. For example, when the distribution of \( X \) is exponential with rate \( \lambda > 0 \), by conditioning on \( X \) to lie in \([0, k]\), the p.d.f. is

\[
f(x) = \begin{cases} 
\frac{\lambda}{1 - \exp(-\lambda k)} \exp(-\lambda x), & 0 < x < k \\
0, & \text{otherwise}
\end{cases}
\]

Since \( \frac{\lambda}{1 - \exp(-\lambda k)} > 0 \) and \( \exp(-\lambda x) > 0, 0 < x < k \) for \( \lambda \neq 0 \) and \( \int_0^k f(x)dx = 1 \), \( f(x) \) is a p.d.f.. The density function \( f \) is decreasing (increasing) on its support according as \( \lambda > 0 (\lambda < 0) \). For \( \lambda = 0 \) as a natural extension \( \lambda \to 0 \) we define

\[
f(x) = \begin{cases} 
\frac{1}{k}, & 0 < x < k \\
0, & \text{otherwise}
\end{cases}
\]

Hence the natural parameter space of the conditional distributions so obtained is \((-\infty, \infty)\). A family \( \mathcal{F} \) of distributions can be extended to a family \( \mathcal{G} \) by considering

\[
\mathcal{G} = \{G(x, \alpha) = H(F(x), \alpha), F \in \mathcal{F} \text{ and } H \in \mathcal{H} \}
\]

where, \( \mathcal{H} = \{H(x, \alpha): 0 \leq x \leq 1, \alpha > 0 ; H(0^+, \alpha) = 0, H(1^-, \alpha) = 1, H(x, 1) = x, H(x, \alpha) \text{ increasing in } x\} \).
For example i) \( H = \{ x^\alpha, \alpha > 0 \} \) generates exponentiated family of distributions. ii) \( H (x, \alpha) = \frac{x}{\{ \alpha + (1- \alpha) x \}}, \ 0 < \alpha \leq 1 \), is a function used by Marshal & Olkin (1997) to extend the class of exponential distributions. For further details one may refer to Rattihalli (2004). Also for multivariate models using conditional specifications one may refer to Arnold, et al. (2002).

1.3. Motivation of the work:

Sometimes a proposed well-known statistical model may not give a very satisfactory fit for a collected data set. This might be due to typical inherent behavior of the phenomenon that gives rise to the specific data in hand and the model attempted being incapable of capturing such a behavior. For example in a completely automatised production process uncontrolled error could be symmetric, but when a human or man power is also involved in the process the typical attitude of the person will have an impact on the behavior of the errors, because of which errors would no more be symmetric. Another illustration is that of random position of a particle under perfect ideal conditions being symmetric about a known point. But when ideal conditions are disturbed say (by a force), random behavior of the position of the particle would no more be symmetric. In such situations, to account for the impacts like the attitude or the force, suitable models have to be proposed.

The salient features of the research work presented are as follows.

i) To introduce new families of distributions using contour transformations.

ii) To study their distributional properties and obtain permissible limits for the parameters used in the contour transformations.
iii) To study the performance of the proposed estimators may be based on simulation.

iv) To generate multivariate densities with different shapes of the contours, may be having different centers (using contour transformations).

iv) To identify new class of densities based on some characterizing property and study their distributional properties.

v) Generation of densities based on the size of contours of a density function.

1.4. Basic concept and Preliminaries:

Let \( f(x), x \in \mathbb{R}^k \) be a p. d. f. and with modal value zero.

For \( 0 \leq u \leq f(0) \), the set \( \{ x : f(x) = u \} \) is called as a contour of \( f \), or an \( u \)-level set of \( f \).

However in the following for convenience the set

\[
C_f(u) = \{ x : f(x) \geq u \}
\]

be called the \( u \)-contour of \( f \). The support of the density \( f \) is

\[
C_f(0) = \lim_{k \to \infty} \left\{ c_f(u_k) \right\},
\]

where \( u_k \) is any sequence decreasing to zero.

Let \( f(x), x \in \mathbb{R}^k \) be a p. d. f. and \( \mathbb{C}_f = \{ C_f(u) : 0 \leq u \leq f(0) \} \) be the class of all contours of \( f \). Let \( \mathbb{A}_k \) and \( \mathbb{A} \) be the Lebesgue measures on \( \mathbb{R}^k \) and \( \mathbb{R} \) respectively.
**Definition 1: Contour Transformation:**

A Contour Transformation (CT) transforms each member \( C_f(u) \) of the Class \( \mathcal{C}_f \) to \( C^*(u) \), so that \( \land_k(C_f(u)) = \land_k(C^*(u)) \) for \( 0 \leq u \leq f(0) \) and empty set is transformed to empty set itself, so that \( C^*(u) = C_{f^*}(u) \), for some density \( f^* \). The p. d. f. \( f^* \) is said to be obtained by a CT of the p. d. f. \( f \) and is denoted by CT (\( f \)).

**Definition 2: Basic Contour Transformation:**

The CT that transforms \( C_f(u) \) to \( (0, \land(C_f(u))) \), for \( u \geq 0 \) is called the Basic Contour Transform. The p.d.f. corresponding to the contours \( (0, \land(C_f(u))) \) be called as the Basic Contour Transformed density of \( f \) and be denoted by \( B(f) \). The function \( g(.) \) defined in (2.2.7) is \( B(f) \).

**Unimodality in Multivariate distributions:**

A function \( f(x): \mathbb{R}^k \to [0, \infty) \) is said to be unimodal if the set \( \{ x \mid f(x) \geq u \} \) is convex for all \( u \geq 0 \).

Let \( f(x) \) be a unimodal p.d.f. and \( C_f(u) = \{ x \in \mathbb{R}^k : f(x) \geq u \} \). It is clear that \( \{ C_f(u), 0 < u \leq f(0) \} \) is decreasing class of convex contours. It is to be noted that if a set \( A \) is convex then \( MA + \theta \) is also convex and \( \land(MA + \theta) = |
\begin{vmatrix} M \end{vmatrix} \land(A) \), where, \( | M | \) is the determinant of \( M \) and \( \land(A) \) is the Lebesgue measure (volume) of the set \( A \).

If \( C^*(u) = MC(u) + \theta \) with \( | M | = 1 \) then \( \{ C^*(u), 0 < u \leq f(0) \} \) satisfies the conditions

(i) \( C^*(u) \) is non increasing and

(ii) \( \land_k(C(u)) = \land_k(C^*(u)) \)

and corresponding density \( f^*(x) \) with \( C^*(u) = C_{f^*}(u) \) is given by
\[ f'(x) = \sup\{u : f(x) \geq u\}. \]

Thus contour transformation \( C^*(u) = MC(u) + \theta \) corresponds to the density of the random vector \( MX + \theta \), when \( f \) is the density of random vector \( X \). We note that if \( M \) is not orthogonal it does not corresponds to just a rotation. For example, if \( f(x) \) corresponds to \( N_p(\mu, \Sigma) \) distribution, then

\[
C(u) = \{x: (x-\mu)'\Sigma^{-1}(x-\mu) \leq -2 \log\left(\frac{1}{(2\pi)^{p/2}}\right)\}. 
\]

**Gauss Hypergeometric Functions:** The Gauss hypergeometric function \( _2F_1(a, b; c; z) \) is defined as

\[
_2F_1(a, b; c; z) = \sum_{t=0}^{\infty} \frac{(a)_t (b)_t (z)^t}{(c)_t t!} 
\]

where \( a, b \) and \( c \) are the parameters independent of \( z \). Here \( (a)_t = a(a+1)(a+2) \ldots (a+t-1) \). For details one may refer to Bateman (1953).

**Acceptance Rejection Method:**

If \( X \) is a random variable with continuous distribution function \( F \), a realization on \( X \) can be obtained by noting the fact that \( F(X) \) has \( U(0,1) \) distribution. That is a realization is obtained by solving the equation \( F(x) = u \) where \( u \) is realization on Uniform random variable over \( (0, 1) \). That is a realization \( x \) of \( X \) is given by \( x = F^{-1}(u) \).

If it is difficult to obtain \( F \) and/or difficult to obtain \( F^{-1} \), one can use the following method, known as acceptance rejection method.

In the following we describe this method to generate observation with density \( f \), by using a value generated from the density \( g \).

Let \( X \) has the density \( f \) and \( Y \) has the density \( g \). The density \( g \) is suitably selected so that an observation from it can be easily generated and \( f(x) \leq cg(x), c \geq 1 \). The constant \( c \) is to be taken as small as possible (since \((c-1)/c\) is...
the proportions of rejections). The acceptance / rejection method is described below.

Step 1. Generate $Y$, (say $y$) a realization from $g$.

Step 2. Generate independently the uniform random variable $V$ (say $v$) over $(0, cg(y))$.

Step 3. If $v \leq f(y)$ then accept this value of $y$ as a realization on $X$, otherwise reject (discard) the value and go to step 1.

**Remark 1:** Let $(W, V)$ be a bivariate random vector with density $h(w, v) = 1$ for $-\infty < w < \infty$, $0 < v < \infty$, where $f$ is a density function. Then marginal density of $W$ is $f$ and conditional density of $V$ given $W=w$ is uniform over $(0, f(w))$.

**Remark 2:** Let $(W, V)$ have uniform bivariate density on a sub set $A$ of $\mathbb{R}^2$. The conditional distribution of $(W, V)$ given that $(W, V) \in B$, a sub set of $A$ is uniform over the set $B$. In the above acceptance rejection method $f(w, v) = c^{-1}$, $A = \{(w, v); (-\infty < w < \infty), 0 < v < cg(w)\}$ and $B = \{(w, v); (-\infty < w < \infty) 0 < v < f(w)\}$. Since $g$ and $f$ are density functions the lebesgue measure $\Lambda (\Lambda_2)$ of the set $A$ and $B$ are $c$ and $1$ respectively. Let random vector $(W, V)$ have density $h(w, v) = 1$ for $(-\infty < w < \infty)$, $0 < v < f(w)$, where $f$ is a density function. Then marginal density of $W$ is $f$ and the conditional density of $V$ given $W=w$ is uniform over $(0, f(w))$. We note that $f(w)$ can be infinity only on a countable set $C$ and $P(W \in C) = 0$. For further details one may refer to LucDevroye (1986).

1.5. Organization of the Study:

Chapter 1 is introductory in nature. It includes introduction and motivation of the topic of the study. In addition a literature review and related
details are described in brief.

In Chapter -2, we describe some notations and define contour transformation. Further we introduce Basic contour transform, a typical CT, leading to a new decreasing density on $(0,\infty)$. The ‘contour transform equivalence’ is defined and based on it equivalent classes of densities are defined. Also we establish stochastic ordering between two univariate random variables with the densities $f$ and $f^*$ when $f$ is unimodal and $f^*=\text{CT}(f)$. Further, a method of generating a slanted density by using a CT depending on an additional parameter $\beta$ (to be referred as slant parameter) is described and range for values of the parameter $\beta$ is obtained, so that the conditions required for the CT are satisfied. Estimators of $\beta$ for slanted standard Laplace model are obtained by using (i) sample moment and (ii) sample quantile. The performance of the proposed estimators is studied using simulation. Generation of asymmetric densities from symmetric ones by using CT that depends on a parameter $\epsilon$ (to be referred as skewness parameter) is discussed and it is shown that an $\epsilon$-skew Laplace model gives a better fit than that of Laplace for the data used by Azzalini and Dalla Valle (1996). C-program to estimate $\theta$, $\sigma$, and $\epsilon$ is given in Appendix A. The Part of this work has been accepted for publications in the Journal of Indian Statistical Association.

Chapter -3 deal with the development of multivariate models using Contour Transformation. This also includes the CT yielding densities with change in the form or shape of the contours. Some simple properties of generated densities are discussed. Also we consider a p-variate C-contoured unimodal probability density function (p.d.f.) $f$ with modal value $\theta$ and by using contour transformation a new family of $C_\delta^*$- contoured density functions $\{f^*(x, \delta): \delta \in \Delta, \text{a suitable set of parameters}\}$ is obtained. Some properties of $f^*(x, \delta)$ are studied. The part of this work is published in Statistical Science and Interdisciplinary Research -Vol.4, Rattihalli and Basugade (2008).
In Chapter -4, we propose a class of densities on \((0,\infty)\) such that the inverse of the density function is the density function itself and study their properties. Also we obtain maximum likelihood estimator and moment estimator of the parameter \(\alpha\) and based on simulation the performance of the proposed estimators are studied. The MATLAB program to obtain Maximum likelihood estimate and moment estimate are given in Appendix B and C. The Part of this work has been accepted for possible publications in International Review of Pure and Applied Mathematics, Basugade and Rattihalli (2009).

Other methods of generating densities are discussed in chapter-5. If \(f\) is a p.d.f. with model value 0 then \(h(u) = \wedge(C_f(u)), 0 < u \leq f(0)\) is a p. d. f.. Such types of densities are obtained when \(f\) is symmetric and in discrete cases also. Some of their distributional properties are discussed.