1.1 Introduction

Normal distribution is one of the most celebrated distributions in statistical literature. The history of normal distribution started when Laplace in his book *Analytical Theory of Probabilities*, used the normal curve to describe the distribution of errors. Subsequently, Carl Friedrich Gauss (1809) used the normal curve to analyze astronomical data. The importance of normal distribution rests partially on because of Central Limit Theorem (CLT), one of the fundamental theorems that form a bridge between Probability and Statistics. Use of the Normal distribution as a statistical model for both theoretical as well as data analysis is wide spread. It is a fact that even though normal distribution is more mathematically complex than many other distributions, it occurs more frequently.
The probability density function of a normal distribution with location parameter $\theta$ and scale parameter $\sigma$ is usually indexed by the symbol $\phi$ and is given by

$$\phi(x; \theta, \sigma) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty, \quad \sigma > 0.$$ 

The importance of normal distribution is not only due to central limit theorem but also to the availability of estimates of the parameters with desirable properties. Even though normality assumption is usual in statistical data analysis, unfortunately, departure from normality also seems to be more frequent than expectations. If there is a shift in the assumption of underlying distributions, many of the celebrated estimates will underperform in terms of their efficiencies and are not robust. In fact Huber (1981), by means of an example, showed that if there is a very small contamination in the normality how it will affect the relative efficiencies of the estimates. Many examples of departure from normality and its implications in the estimates are reported in Huber (1981), Stigler (1977).

A large number of alternative distributions were suggested in place of normal probability density functions. In robustness studies one considers various symmetric non-normal error distributions. For example, the double exponential due to Laplace and the parabolic distribution with probability density function $f(x) = (3/4r^3)(r^2-x^2), \quad |x| \leq r$, due to Euler, were the early competitors of the normal model, and many symmetric distributions with finite variance were studied as error distribution in the past. Since under the normal error model optimal estimating and testing procedures available, the question as to whether errors are in fact normally distributed or not assumed some significance and was tackled by residual analysis and probability plotting method along with some tests for normality. When the data seem to be heavier tailed than normal, Laplace distribution can be used instead of normal. When the information is available in the form
of first few moments we use normality assumption when the skewness $\beta_1 = 0$ and kurtosis $\beta_2 = 3$ where, $\beta_1 = \mu_3^2/\mu_2^3$ and $\beta_2 = \mu_4/\mu_2^2$, with $\mu_i$, being the $i^{th}$ central moment of the population. The basic idea here is that the Gram-Charlier expansion or Edgworth expansion for all probability density functions reduces to the normal when skewness is zero and kurtosis is equal to three, when higher order terms are neglected. If the normality assumption is ruled out Edgworth series expansion can be used to approximate a probability density function of the population. The Edgworth series expansion of the standard form of the probability density function $f(x)$ with skewness measure $\beta_1$ and kurtosis measure $\beta_2$ is given by

$$f(x) = \phi(x)[1 + \frac{\sqrt{\beta_1}}{6}H_3(x) + \frac{\beta_2 - 3}{24}H_4(x) + \frac{\beta_1}{72}H_6(x)],$$

(1.1.1)

where, $\phi$ is the normal probability density function and $H_r(x)$ denotes the $r^{th}$ degree Hermite polynomial given by

$$\frac{d^r(e^{-x^2})}{dx^r} = (-1)^rH_r(x)e^{-x^2}, \quad r = 0, 1, 2, \ldots$$

(1.1.2)

However a serious drawback of these expansions, is that the density is non-negative only for restricted values of $\beta_1$ and $\beta_2$. For example, see, Barton and Dennis (1952) and Draper and Tierney (1972).

The normal distribution possesses some interesting properties that make it a favourite model for Mathematicians, Statisticians and Data Analysts. For example it is the only distribution for which the sample mean $\bar{X}$ and sample variance $S^2 = \sum (X_i - \bar{X})^2/n$ are independently distributed, where $X_1, X_2, \ldots, X_n$ is a random sample taken from the population. In the simplest case of the linear regression model $X_i = \theta + \epsilon_i$, where $\epsilon_i$’s are independent random variables with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$, $i = 1, 2, \ldots, n$, it is well known that the least squares estimators of $\theta$ and $\sigma^2$ are the sample mean $\bar{X}$ and the sample variance $S^2$ respectively. If $\epsilon_i$’s are assumed to be normal, $N(0, \sigma^2)$
then $\bar{X}$ and $S^2$ are independent and are the maximum likelihood estimators of $\theta$ and $\sigma^2$ respectively. Godambe and Thompson (1989) extended this idea to more general semi-parametric models in which $E(X) = \theta$, $Var(X) = \sigma^2$, and skewness $\beta_1$ and kurtosis $\beta_2$ of $X$ are known. They showed that the optimal estimating equations for $\theta$ and $\sigma^2$, when $\beta_1 = 0$ and $\beta_2 = 3$, coincide with the likelihood equations of the normal probability density function with mean $\theta$ and variance $\sigma^2$. It is well known that, within the Pearson family of frequency distributions, the normal probability density function is the only one with $\beta_1 = 0$ and $\beta_2 = 3$. Usually in most of the statistical inferential problems related to regression analysis and robustness studies, we assume that the error distribution is continuous and symmetric. Hence $\beta_1 = 0$ is a natural restriction. A question that naturally arises is, whether there exist symmetric probability density functions other than normal, with kurtosis equal to 3. While Hildebrand (1971) density given in chapter 2, was an answer to this question, Kale and Sebastian (1996) showed that there exist a wide class of symmetric distributions with Pearson’s measure of kurtosis $\beta_2 = 3$. A member of this class can be obtained by considering a mixture of two symmetric non-normal densities, with centers of symmetry being the same, say zero, the kurtosis of one component strictly less than 3 and that of the other component strictly greater than 3. They showed that the probability density functions in this class can have a variety of shapes and can be very much different from the normal density. The following are some other examples of a non-normal symmetric distribution with kurtosis three. Consider the probability density functions

$$m_p(x) = \frac{(p+1)}{2p} \{1 - |x|^p\}, |x| < 1, p > 0, \quad (1.1.3)$$

and

$$g_q(x) = \frac{(q-1)}{2} |x|^{-q}, |x| > 1, q > 5. \quad (1.1.4)$$

In the case of (1.1.3), $\beta_2(m_p) = 3$ for $p = \sqrt{10} - 3$ and in the case of (1.1.4), $\beta_2(g_q) = 3$.
for \( q = 3 + \sqrt{6} \). Now, it is easy to show that if we have two densities symmetric about zero with \( \beta_2 = 3 \), any mixture of the standardized versions of these probability density functions would be symmetric around zero with \( \beta_2 = 3 \). For more examples and method of construction of such densities we refer to Kale and Sebastian (1996).

Motivated from the example of Kale and Sebastian (1996) and Hildebrand (1971), in this thesis we try to identify and characterize a fairly large class of symmetric mesokurtic distributions in which normal distribution is a special case. We also identify a subclass of this family of densities with close resemblance to the normal model. The performance of the estimators in regression model and time series analysis in which, we replace the normal error by one of the member of this class is also considered.

### 1.2 Summary

The thesis is divided into five chapters. Chapter 1 provides an introduction and summary of the thesis. In Chapter 2, we consider the general symmetric mesokurtic family and prove some characterization results. As an extension of Kagan et.al. (1973), we prove that if \( X_1, X_2, \ldots, X_n \) are \( n \) independent random variables from a symmetric mesokurtic family of probability density functions, then any linear combination of \( X_i \)'s will be closed under symmetric mesokurtic property. We also prove that if the kurtosis value \( \gamma_2 = \beta_2 - 3 \) of any linear combination of a set of independent random variables is unchanged if and only if the random variables are from a symmetric mesokurtic family.

While all the densities considered by Hildebrand (1971) and Kale and Sebastian (1996) were a mixture of two symmetric probability density functions, we give an example of a class of symmetric probability density functions with \( \beta_2 = 3 \), which cannot be considered as a mixture of density. This class contains normal distribution as a particular case and is obtained by considering one more term in the usual Edgworth series approximations for a density with given moments. This class contains both unimodal and multimodal densities. We show using examples that a symmetric mesokurtic family...
contains distributions for which the sample median is asymptotically a better estimate than sample mean for estimating population mean. A simulation study is conducted to assess the performance of the estimates of location and scale parameters when the random variable is from the subfamily of symmetric mesokurtic family newly introduced in this chapter.

In Chapter 3, we consider a comparative study between normal distribution and the class of non-normal symmetric mesokurtic distributions. We have shown that $\beta_2$ does not measure the peakedness of distribution at the center, even for unimodal densities. It is also shown using examples that $\beta_2$ is not a measure of tail weight, tail length or 'lack of shoulders'. Any test based on $\beta_2$ will have power zero to detect shifts from normality within the class of symmetric mesokurtic distributions. It can be seen that even though P-P plots and Q-Q plots are not sensitive to the shifts from normality within this particular class, kurtosis measure due to Horn (1983) is a good option to detect such type of shifts. We also discuss concepts like spread function and Van Zwet's (1964) ordering and try to use them for identifying shifts from normality to other symmetric distributions with $\beta_2 = 3$. A moment based ordering is introduced in this chapter followed by its application in ordering probability distributions.

Chapter 4 deals with estimation of parameters from the symmetric mesokurtic family introduced in chapter 2. The most important and conventional method of estimation, that is Maximum Likelihood (ML) method of estimation cannot be easily applied here, as the terms in the likelihood equation are intractable. As a result, we cannot have any closed form expressions for maximum likelihood estimates. Solving the likelihood equations using iterative procedure is also not feasible due to multiple roots, slow convergence or even divergence. To overcome these difficulties, we use Modified Maximum Likelihood (MML) method of estimation to find the estimates of location and scale parameters in this family. This can be done by first expressing the ML equations in terms of order statistics and then replacing the intractable terms by their linear approxima-
tions. An overview of MML estimation is available in Tiku and Suresh (1992) and Tiku (1998) and the references cited therein. The expressions for the estimates are given and a simulation study is conducted to assess the accuracy of the estimates. Simulation study as well as theoretical results shows that the estimates are asymptotically unbiased and fully efficient. In general they have less variance than least square estimates and are equally good as maximum likelihood estimates.

The last chapter deals with some applications of the newly introduced symmetric mesokurtic family in regression analysis and time series modeling. The class of symmetric mesokurtic distributions can be used as error distribution in linear regression models, instead of Gaussian error models. Considering this class of distributions as error models, we have derived estimates for the unknown parameters. Here again the philosophy of modified maximum likelihood estimation is used, as the likelihood equations are intractable. Asymptotic properties are derived and a simulation study is conducted by generating random samples from the model. Results of the study are presented and computer programs are given in the appendix. The last section of chapter 5 considers the time series models of the form

\[ Y_t = \sum_{j=1}^{p} a_j Y_{t-j} + \epsilon_t, \quad t = 1, 2, ..., n, \]

where, \( \epsilon_t \) are independently and identically distributed with mean zero and variance \( \sigma^2 \).

Estimation of parameters from such an Auto Regressive (AR) model has received a great deal of attention in statistical literature. In such studies, error models are broadly classified into two as, Gaussian and Non-Gaussian. See for example; Durbin (1960) has used Gaussian error models, while Huber (1981), Tiku et. al. (1986, 1999) have used different Non-Gaussian error models. Tiku et. al. (2000) in their paper considered Auto Regressive models with the underlying distribution as a class of non-normal symmetric innovations. The well-known assumption regarding the distribution of \( \epsilon_t \) that it is distributed independently and identically as normal with mean zero and variance \( \sigma^2 \). But apart from the restrictiveness of normality assumption many authors have illustrated non-normality in real life situations. For example, see, Ganadesikan (1977).
Non robustness of the estimators under the normality assumption is noticed by authors like Huber (1981), Martin and Yohai (1985) and Tiku et.al. (1986). In the same line as Tiku et. al. (2000), we use modified maximum likelihood estimation procedure to obtain the estimators in AR models when the innovations are a class of symmetric mesokurtic distributions. A simulation study is conducted using Monte Carlo runs generated from this model. A matlab program that is used for Monte Carlo simulation study is given in the appendix.
References


