# Distribution of Income in Pakistan using L-moments

TH/15363

By

Muhammad Alam

Department of Mathematics and Statistics

Faculty of Basic and Applied Sciences

International Islamic University, Islamabad

Pakistan

2015

MS
519.24
MUD
K
M.Mu

# Distribution of Income in Pakistan using L-moments

By

**Muhammad Alam**

Supervised by

***Dr. Ishfaq Ahmad***

**Department of Mathematics and Statistics**

**Faculty of Basic and Applied Sciences**

**International Islamic University, Islamabad**

**Pakistan**

**2015**

# Distribution of Income in Pakistan using L-moments

By

Muhammad Alam

*A Dissertation*

*Submitted in the partial Fulfillment of the*

*Requirements for the degree of*

***MASTER OF SCIENCE***

***IN***

***STATISTICS***

Supervised by

***Dr. Ishfaq Ahmad***

Department of Mathematics and Statistics

Faculty of Basic and Applied Sciences

International Islamic University, Islamabad

Pakistan

2015

# Certificate

## Distribution of Income in Pakistan using L-moments

By

*Muhammad Alam*

A DISSERTATION SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF THE *MS IN STATISTICS*

We accept this dissertation as conforming to the required standard.

1. ______________
**Dr. M. Yousaf Shad**
(External Examiner)

2. ______________
**Dr. Zahid Iqbal**
(Internal Examiner)

3. ______________
**Dr. Ishfaq Ahmed**
(Supervisor)

4. ______________
**Dr. Rahmat Ellahi** 2/12/15
(Chairman)

**Department of Mathematics & Statistics**
**Faculty of Basic and Applied Sciences**
**International Islamic University, Islamabad**
**Pakistan**

**2015**

## Forwarding Sheet by Research Supervisor

The thesis entitled **"Distribution of income in Pakistan using L-moments"** submitted by **Muhammad Alam** (Registration # 01-BAS/MSST/F12) in partial fulfillment of M.S degree in Statistics has been completed under my guidance and supervision. I am satisfied with the quality of his research work and allow him to submit this thesis for further process to graduate with Master of Science degree from Department of Mathematics and Statistics, as per IIU Islamabad rules and regulations.

Dated:____________________

**Dr. Ishfaq Ahmad,**
Assistant Professor
Department of Maths & Statistics,
International Islamic University,
Islamabad.

## Acknowledgements

I bow my head to **Almighty Allah** in gratitude who blessed me the courage to complete this thesis. May each and every moment of my life be devoted to his praise. All respects and regards to the Holy Prophet **Muhammad (P.B.U.H)**, who came as a light of knowledge for all seekers. I feel highly obliged in my deepest and sincerest gratitude to my respected teacher and dignified supervisor **Dr. Ishfaq Ahmad**, for his keen interest, benevolent attention, vital and intellectual suggestions throughout the research work. I am highly indebted to him for his graciously offered help to complete this tedious task. May Allah bless him with health, happiness and long life. I also want to express my unbounded thanks to my respected teacher **Dr. Zahid Iqbal** Assistant Professor, Department of Mathematic and Statistic for encouraging me. My deepest gratitude and regards to my family for their affection, prayers, financial and moral support throughout our educational career, whose emotional, intellectual and fiscal abet and numerous sacrifices enabled me to stand where I do today. I am unable to pay thanks for the kindness and loving attitude of my elder brother **Muhammad Umar** who devoted his days and nights for my success. The completion of this research work is not possible without his prayers, financial support and encouragement.

Furthermore, I cannot ignore the cooperation, emotional support, entertainment, love and care of my friends **Syed Aftab Mehmood** and **Inam Ullah** throughout my research.

**Muhammad Alam**

# DECLARATION

I hereby declare that this thesis, neither as a whole nor a part of it, has been copied out from any source. It is further declared that I have prepared this dissertation entirely on the basis of my personal efforts made under the supervision of my supervisor **Dr. Ishfaq Ahmad**. No portion of the work, presented in this dissertation, has been submitted in the support of any application for any degree or qualification of this or any other learning institute.

**Signature:____________________**

**Muhammad Alam**

**MS (Statistics)**

**Reg. No 01-FBAS/MSST/F12**

**Department of Mathematics and Statistics,**

**Faculty of Basic and Applied Sciences,**

**International Islamic University Islamabad, Pakistan.**

**Table of Contents**

## List of Figures

## List of tables

# Abstract

In this dissertation, we estimate the three parametric lognormal distribution parameters for income data in Pakistan using different methods of estimation. The data are obtained from Pakistan Social and Living Standard Measurement Survey (PSLSMS) of 2010-2011, Pakistan Bureau of Statistics (PBS) Islamabad. We considered three distributions, generalized gamma, generalized pareto and three parametric lognormal distribution to find the best fit distribution by different goodness of fit tests and L-moments ratio diagram. The best fit distribution is three parametric lognormal distribution for income data in Pakistan. We estimate the parameters of lognormal distribution by the method of maximum likelihood, method of moments and method of L-moments for the income of male group, female group and four provinces. For assessment of different estimation methods, we compare the results with the help of Absolute Bias and Root Mean Square Error (RMSE). On the basis of Absolute Bias and RMSE we found that the L-moments method is the best for small sample size and maximum likelihood method is best for large samples.

# CHAPTER 1

## 1.1 Introduction

The importance of income throughout the world is undeniable. It is a backbone for a country especially, country like Pakistan. It is as important as breathing air. Without air there is no hope of life and without having a reasonable source of earning there cannot be even a single ray of hope for survival. The distribution of income centers around an important matter in a country. The level to which government should rearrange income from the peoples with more income to those with less.

The calculation of public income is essential information for a state. By the national income we can find that the financial system is increasing or decreasing. The increasing of the economy brings prosperity in the country. The income of a country provides the estimation of the performance of a state in economic years. We can find the individual income and know about the financial development of a country. Income reveals the living standard of the peoples and compares the life style with other countries of the world.

Without income growth, a country cannot achieve its goals and cannot improve the living standard of citizens. The improvement in education is the first source for the development of income of a national income. The industrial sector development as the second source for increasing the income of a country. It provides large number of jobs to the peoples. The third source for the betterment of income of a country is agriculture. It provides an opportunity of employment, reducing poverty ratio, supply foods, provide raw material such as sugarcane, cotton, tobacco, oil seed, rice and milk. Due to good income, we can promote education, health, monitoring, transport, agriculture, defense technologies, industrial development etc. It measures the strength of the economy.

The ratio of crimes is inversely proportional to income. The citizens of Pakistan are not equal economically. Some groups of peoples are very rich and some are very poor. There is an inequality in the income. And due to this inequality, unethical activities such as corruption, killing, kidnapping, robbery, burglary, bribery, agitation etc. takes birth. Income inequality also leads to increase poverty and dissimilarity. There are some reasons due to which income dissimilarity arises, that are the education pattern, not only the years of education, but also the excellence of education play an important position in income variation. Another reason is the gender discrimination that has female labor earned less than male labor. It also arises due to the regions such as rural and urban areas people's income. Every country has the goal to reduce poverty, which is possible by the increase in income. This income brings prosperity in a country.

Wages and income of the peoples are of great interest for economists. In developed countries around the globe, there is a continuous struggle for objective solutions to problems pertaining to the standard of living and their general wellbeing. Income models can be utilized separately from feature of studying life, or for reaching at an inter-regional, international comparison and dimensions of standards of life.

The "income distribution" terminology is a statistical idea. Not a single individual is distributing income. Rather, people's decisions about work, saving and asset give birth to their income distributions. They relate through markets and influenced by the tax system. In early 1990s and early 2000s emerged a growing body of work. Significantly and minutely precise, establishing how the income distribution has altered.

It is a prescriptive question that government should redistribute income, and each individual answer depends on his/her values. But most of the people require the knowledge that how the facts and figures regarding the existing income distribution for answering the narrow questions. To view in the long run, the roadmap of income inequality over the last (twentieth) century is marked by two main events. A steep decline in inequality around the outbreak of World War II. And a skyrocketing and the astronomical rise in inequality, which started in the mid-1970s and further gained momentum in the 1980s. In the current scenario, income inequality is as higher as it was in the 1920s.In other words, a wholesome knowledge and strategy of the overall income distribution over a particular period for a country. And also income distributions among particular subgroups of people are essential to overcome the crisis.

These distributions, on the other hand, could be utilized in analytical terms of income discriminations or income inequalities in the public. The wage models can be used to evaluate interregional comparison. For accurate estimate of the wage related parts of the standard of living as well as for exact decisions regarding interventions in this sphere. We have to go through the pros and cons of the whole wage distribution in a particular period, i.e. every wage group. Entrepreneurs should pay special attention to the wage distributions when analyzing the market probabilities. These distributions can be used for tax load estimation etc. It has to do concern with the process improvements and other statistical variables correlated with income. Day to day instances of household expenses and facilities, time consumption, purchasing purposes as well as various positions within the circle of sociological research are immensely helpful. In this study, we discuss about the income in Pakistan gender wise and province wise. An appropriate probability distribution will be taken and its unknown parameters have to be estimated, when the analyzed income distribution is described by a probability model. Incomes within countries generally adopt a skewed distribution with a long heavy tail. Lognormal distribution has been found best fit on the average for income distributions. This can also be proved through the goodness of fit test such as Chi square test, Anderson Darling test.

## 1.2Generalized Gamma Distribution

The generalized gamma is a continuous probability distribution. The probability density function of the generalized gamma distribution is given as

$$f(x;\alpha,d,p) = \frac{(\frac{p}{\alpha^{d}})x^{d-1}e^{-(\frac{x}{\alpha})^{p}}}{\Gamma(\frac{d}{p})} \qquad (1.2.1)$$

Where $0< x < \infty,\ \ \alpha > 0, d > 0,\ p > 0$

The generalized gamma distribution is used in several fields of life for estimation of parameters, e.g. they are used, for the analysis of skewed data, to find the amount of rainfall, aggregate insurance claims and flow of items in manufacturing etc.

## 1.3 Generalized Pareto Distribution

The generalized pareto is a continuous probability distribution. The probability density function of the generalized pareto distribution is given as

$$f(x;\varepsilon,\sigma,k) = \alpha^{-1}e^{-(1-k)y} \qquad (1.3.1)$$

Where $y = -k^{-1}\frac{\log\{1-k(x-\varepsilon)\}}{\alpha}$

Where $-\infty < \varepsilon < \infty,\ \ -\infty < k < \infty\ ,\ \alpha > 0.$

With parameters $\varepsilon$ = location parameter, $\alpha$ = scale parameter, k = shape parameter. The generalized pareto distribution is used in several fields of life for estimation of parameters, e.g. they are used to describe the allocation of wealth among individuals,to find the amount of rainfall etc.

## 1.4 Lognormal Distribution

The lognormal distribution is a continuous probability distribution. It is also called Galton distribution. The probability density function of the three parametric lognormal distribution is given as $f(x;\ \mu,\sigma,\gamma) = 1/((x-\gamma)\sigma\sqrt{2\pi})e^{\{\frac{-[\ln(x-\gamma)-\mu]^2}{2\sigma^2}\}}$ (1.4.1)

Where $0 < x < \infty, 0 \leq \gamma < x, -\infty < \mu < \infty, \sigma > 0$. With parameters $\mu$ = scale parameter, $\sigma$ = shape parameter, $\gamma$ = location parameter or threshold parameter. The lognormal distribution is used in several fields of life for estimation of parameters, e.g. they are used in Statistics, Geology, Medical Science, Environmental Science, Technology, Ecology, Social science and income. The Lognormal distribution is used to claim that the distribution is positively skewed or not, The distribution of income is also positively skewed. So we use the lognormal distribution for income data. The two and four parametric lognormal distribution is also used as an income distribution in practice. But the three parametric lognormal distribution is used as an income distribution most frequently.

The skewness and kurtosis of the lognormal distribution increases if the value of shape parameter $\sigma$ increases. The skewness will increase if the value of $\sigma$ increases for a given value of $\mu$. But when the value of scale parameter $\mu$ increases, then the dispersion of the distribution will be increased. The threshold parameter $\gamma$ shifts the probability density function of the distribution. If we put $\gamma = 0$ then the three parametric lognormal distribution become two parametric lognormal distribution. The very important property of the lognormal distribution is that as the value of shape parameter $\sigma$ decreases, then the shape of the distribution will be close to Normal distribution. We use the lognormal distribution for parameter estimation using L-moment method. The parameters obtained as a result of L-moments method are often even more precise than other methods of parameters estimation. The accuracy of L-moment method can be compared with other methods of estimation such as moments method and maximum likelihood method.

## 1.5 L-moments

L-moments are the outlook of certain linear combinations of order statistics. And they can be defined for every random variable whose mean exists. The L-moments can be used for summarization and explanation of theoretical probability distributions. Also used for summarization and description of observed data samples, estimation of parameters, quantiles of probability distributions and hypothesis testing. L-moments suffer less from the effect of sampling variability. L-moments are more powerful than conventional moments when there are outliers in the data. The L-moments are used for extreme events such as rainfall, flood, earth quick, droughts, heat waves, snow fall, famine and income distribution. For income modeling the L-moments are used because L-moments also contain some extreme values.The income model is used for assessment of living standard. The industrialist is using the income distribution for the market potential and for tax load estimation. The L-moments technique gives correct result in the estimation of parameters as compare to the maximum likelihood method and moments method.

## 1.6 Objective of the Study

- To find the best distribution for income data through the goodness of fit test.
- To estimate the parameters of the best fit distribution by different methods of estimation, such as maximum likelihood method, moments method and L-moments method.
- To compare the results of different methods of estimation through Absolute Bias and RMSE for different sample sizes.
- To emphasize the variation in income inequality between different groups.

# CHAPTER 2

## Literature Review

A brief review of the income distribution, using L-moments method.

A. R. Kemal (2006) used the data of Household Income and Expenditure Survey (HIES) and Pakistan Integrated Household Survey (PIHS) from 1963-2002 of Pakistan, rural and urban. And decided that the main goal of economic growth is to improve the living standard of ordinary men. But only economic growth is not sufficient for the standard of living for men. The improvement in the income distribution is also play an important role in the betterment of the human life. But unluckily the ratio of poverty becomes increase in 1990s. The economists focused to find the ratio of poor population instead of improving measures of income distribution. The income inequalities can also be estimated by Pakistan Integrated Household Survey (PIHS) and Household Income and Expenditure Survey (HIES).

Bilkova Diana (2008) used the wage data of Czech Republic from 2004-2005. They applied different methods of parameter estimation such as quantile method, moments method and maximum likelihood method. Also apply lognormal distribution and decided that the results of lognormal distribution is best.

Hafeez Ur Rehman et al (2008) analyzed the distribution of income, growth and its development in Pakistan. And decided that Income inequality is a main financial problem to all the word. In the early periods several attempts have been made to establish a relation between income inequality and economic growth. But there are no such conclusions found between income inequality and economic growth. So they concluded that there is a reverse correlation between income and economic growth.

Bilkova Diana (2011) used the data of household income per capita in Czech Republic from 1992-2008. They decided that L-moments method for estimation of parameters give us the correct consequence as compare to other methods, such maximum likelihood method and moments method in case of ungrouped data. But for grouped data these methods give similar results.

Bilkova Diana (2011) used the income data from Statistical surveys "SILC" and Micro census of the Czech statistical office. And wage data taken from the Czech Statistical office website. They use the three parametric lognormal distribution and decide that L-moments technique gives the correct result in the estimation of parameters as compared to maximum likelihood method, moments method and quantile method for individual data. But in case of grouped data all the four methods give similar result.

Bilkova Diana (2011) used the income data of Czech Republic. And suggested that L-moments method for estimation of parameters provides precise consequence as compare to the other methods of estimation such as moments method, maximum likelihood method in case of ungrouped data. But for grouped data these methods gives same results.

Bilkova Diana (2011) used the income data of household per capita in the Czech Republic, and proposed that for income and wage distribution, L-moments is the best method for estimation of parameters as compared with moments method, quantile method and maximum likelihood method.

Diana Bilkova and Ivana Mala (2011) used the data of Czech Republic household income from 1992-2008. Applied lognormal distribution and four methods of parameters estimation that are L-moments method, quantile method, maximum likelihood method and moments method. They decided that L-moments method gives more accurate results than other methods.

Diana Bilkova (2011) used two types of data. The first one is the Czech Republic per capita household income in years 1992,1996 and 2002. And the second is the statistical survey micro census of years 2005,2006,2007 and 2008. They used the three parameters lognormal distribution. And conclude that the accuracy of L-moments method is better than other methods of estimation that are maximum likelihood method, quantile method and moments method for individual data. But for grouping data these four methods give similar consequences.

Diana Bilkova (2011) used the data of income and wages of Czech Republic from 2002-2010, apply lognormal distribution and for estimation of parameters different methods are used such as quantile method, moments method and L-moments method. They decided that L-moments method gives accurate results than other methods of estimation for small as well as for large sample size.

Ivana Mala (2011) used the household income data of Czech Republic in 2008. The lognormal distribution is applied. And then find the parameters by maximum likelihood method, and concluded that the distribution of income and wage is positively skewed.

Jakub Nedved and Ivana Mala (2011) used the income data of Czech Republic from 2000-2010. The model used three parameters lognormal distribution. Different methods of parameters estimation are used such as quantile method, maximum likelihood method and moments method. And decided that the lognormal distribution is best model for income data.

Jana Langhamrova and Diana Bilkova (2011) used the income data of Czech Republic from 2002-2009. They decided that lognormal distribution is most commonly used for income distribution modeling. It provides more information related to the population and its result is precise.

Jakub Nedved (2011) used the income data of a business of "TREXIMA" company from 2000-2010. Apply the three parameters lognormal distribution as an income distribution model for non-business and business regions in the Czech Republic. The distribution of lognormal model worth depends on the different method of parameters estimation. The maximum likelihood method is best suitable for the non-business region. And the quantile method is best for the business region.

Mala Ivana (2011) used the household per capita income of Czech Republic from 2005-2008. Apply the lognormal distribution and also the Dagum distribution. Find the parameters by the maximum likelihood method. And decided that lognormal and Dagum distribution is suitable for the modeling of income distribution.

Diana Bilkova and Ivana Mala (2012) used three methods of parameters estimation that are L-moments method, quantile method and maximum likelihood method for the data of net annual income in Czech Republic from 1992-2007, and proposed the estimation results of L-moments method are very reasonable and reliable as compare to the other's methods.

Diana Bilkova (2012) used the individual data from the net family income per capita in Czech Republic. And the group data were taken from the gross monthly wage in Czech. Applied the three parameters lognormal distribution. And decided that L-moments method of parameters estimation gives us correct result than other methods such as maximum likelihood method, quantile method and moments method for individual data. But for the group data all these four methods give similar result.

Diana Bilkova (2012) used the household income data of Czech from 1992-2008. They used different methods of estimation of parameters such as quantile method, maximum likelihood method and moments method. And decided that for both small and large sample the L-moments method results are correct as compared to other methods of estimation.

Diana Bilkova and Ivana Mala (2012) used the income data of Czech Republic from 1992-2007. They applied the three parameters lognormal distribution and two methods for parameters estimation that are L-moments method and maximum likelihood method. They decided that both method of estimation give accurate result, but slight different.

Diana Bilkova (2012) used the data of income of Czech Republic from 1992-2008. Used lognormal distribution and applied the methods of parameters estimation that are maximum likelihood method, moments method and L-moments method. They decided that the L-moments method gives best results than the other methods of estimation for individual data. But for grouping data these methods give similar results.

Diana Bilkova (2013) used the data of male and female wage of Czech Republic from 2003-2011. They compared the wages of male and female, decided that the average wage of men is greater than the average wage of female.

Mirza Naveed Shahzadand and Zahid Asghar (2013) used the data of income, wage and expenditure. They proposed that L-moments method gives strong results of estimation when there are outliers in the data. And the results obtained from the L-moments method are more capable than conventional moments.

# CHAPTER 3

## Material and Methods

### 3.1 Study Area and Data

The Income data are taken from Pakistan Social and Living Standard Measurement survey (PSLSMS) of 2010-2011 of the Pakistan Bureau of Statistics (PBS) Islamabad. The calculation of national income is important. From this calculation we can find the performance of a country during a year. From the national income we can also find weather the financial system is increasing or decreasing. The national income depends upon the wages, profits, rents, interest and business. Due to these factors we can improve our national income while from it we can improve the per capita income and also the standard of living.The income distribution also used to find the proportion of low, middle and high income workers in a country.

The main goal of economic growth is to improve the living standard of ordinary peoples. The economic growth is not sufficient for the standard of living. Improvement in the income distribution can also play an important role in the betterment of human life. But unluckily the ratio of poverty become increase in 1990s and the economist focuses to find the ratio of poor population instead of improving the measure of income distribution. Income inequality is a main financial problem to the whole word.

Income tax is a due that the employees, salaried groups of peoples, private's business men and firm owners paid to the government. So the income tax is an important source through which a government can improve its economy. The total tax revenues of Pakistan GDP are 11% and the

1.1% tax revenues are from the personal income. But in Pakistan only 2% peoples are registered that paying the tax.

Pakistan is a developing country. It contains different groups of peoples divided on the basis of income, i.e. rich, poor and below the line of poverty. These are because of the income inequality. Due to the income inequalities unethical activities arise in the society that are agitation, bribery, burglary, robbery, killing and kidnapping. The variation in income is also arises due to the pattern and quality of education. The other reason for income inequalities are the discrimination of gender in employment. That female earns less than male and the rural and urban regions are also responsible for income inequalities. The main responsibility of government is that to make policies for the income distribution. They should provide better opportunities for employment to the citizens.

## 3.2 Reasons of Income Inequality in Pakistan

i. Many researchers have studied different sets of data. Some studied the Household Income and Expenditure Surveys (HIES). While other used the income tax data.

ii. Some of them have studied the inequality in consumption, expenditure and income.

iii. Many of them have studied the inequality in urban and rural. While others in the whole Pakistan.

iv. Some studied the income inequalities in the population, while others in the household.

The statistical measures that are used to determine the income inequalities are Gini coefficient, Lorenz curve, coefficient of variation, lognormal distribution, Interquartile range and ratios. The income inequalities can also be estimated by Pakistan Integrated Household Survey (PIHS), Household Income and Expenditure Survey (HIES).

We cannot achieve the goal of improving without the income growth. For the development of income of a country the first source is to improve the education. The second source is industrial development. Industrial development provides large number of job opportunities. And the third source is agriculture. In early fifty years agriculture was not considered as profitable for financial progress in many developed and developing countries of the world. But in the last fifty years, agriculture is considered as main source for the development of economic growth. But unfortunately our agriculture is back from the rest of the world. The government has to improve agriculture of the country. Due to the agricultural improvement, we can improve per capita income, employment source, poverty reduction, food supply, national income source, raw material supply, industrial sector development, and improvement of the standard of living.

### 3.3 Methods of Estimation

Some of the parameters estimation methods are used that are maximum likelihood method, moments method and L-moments method. These methods are discussed below.

#### 3.3.1 Maximum Likelihood Method

Let suppose $X$ is a vector of observations in a sample which can be described by the model $f(x,\theta)$, where $\theta$ is the parameter value. The given model $f(x,\theta)$, will give us the probabilities of the various values of $X$. When we define the maximum likelihood then we have to define the likelihood function first. The likelihood function for a given value of $X$ is $L(\theta)$ as a function of $\theta$. The MLE is a reasonable method of estimation because it locates that value of $\theta$ for which the observed data are most probable. The MLE enjoys some good properties in large sample size such as normality, efficiency, consistency and unbiasness. Suppose a random sample

$X = (x_1, x_2, \ldots\ldots x_n)$ is drawn from a distribution with probability density function $f(x,\theta), \theta\epsilon\Theta$ where $\theta$ is parameter space. The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(xi, \theta) \qquad (3.3.1.1)$$

We define maximum likelihood estimator $\theta^{\wedge}$ of $\theta$ as the value of $\theta$ such that $L(\hat{\theta} \geq L(\theta), \forall\theta\epsilon\Theta$, where is $\theta$ the parameter space. Such a value may be obtained by solving $\frac{\partial l}{\partial \theta} = 0$ or $\frac{\partial \log l}{\partial \theta} = 0$ for $\theta$ and $\frac{\partial^2 l}{\partial \theta^2} < 0$

The maximum likelihood method of parameters estimation is widely used and simple to apply. When the maximum likelihood estimator is derived then we can find its standard error very easily. The main advantage of this method is i.e. it is often used for estimation of parameters. It is to be noted that MLE is used for the estimation of parameters in the case of large sample size. When the sample size increased it gives unbiased minimum variance estimators. The MLE can also be used for making of confidence limits, and for testing of hypothesis.

The most common advantages of the maximum likelihood method are that Maximum likelihood estimators are consistent, asymptotically normally distributed. These are also efficient estimators mean that its variance is smaller than other unbiased estimators in the case of large sample size. When we have a statistic and it is sufficient for parameter. Then the Maximum likelihood estimator of the parameter is a function of sufficient statistic. The main drawback of Maximum likelihood method is that for small sample size it gives biased estimators.

### 3.3.2 Method of Moments

Let suppose we have random samples $x_1, x_2, \dots. x_n$ from a distribution with probability density function $f(x, \underline{\theta})$, and $\underline{\theta} = (\theta 1, \dots, \theta n)$, $\underline{\theta} \epsilon \Theta$ is unknown vector parameter. We define the $r_{th}$ population moment about the origin '0' as, for continuous distribution

$$\acute{\mu} r = E(x^r) = \int_{-\infty}^{\infty} x^r f(x, \underline{\theta}) dx \qquad (3.3.2.1)$$

And for discrete distribution

$$\acute{\mu} r = E(x^r) = \sum_{r=1}^{n} f(x, \underline{\theta}), \quad r = 1,2 \dots . \qquad (3.3.2.2)$$

So $\acute{\mu} r$ in general is a function of θ. We may also define the $r_{th}$ sample moment as

$$\acute{\mu} r = \frac{\sum_{i=1}^{n} xi^r}{n}, \quad i = 1,2, \dots$$

So $\acute{m} r$ is an unbiased estimator of $\acute{\mu} r$

$$E(\acute{m} r) = \frac{\sum_{i=1}^{n} E(x^r)}{n} = \frac{n\acute{\mu} r}{n} = \acute{\mu} r \qquad (3.3.2.3)$$

Eq. (3.3.2.3) is called the moments equation, when we solve the equation for $\theta's$ it will give us estimators called Moments Equation Estimators (MEE). We denote the moment estimator by $\tilde{\theta}$. The advantages of moments method are that it is an old and simple technique used for estimation of parameters and very simple method for finding the estimators of parameters. Its estimators are consistent and best method in case of point estimation parameters. It is an easy method to find estimator and also in case when the other methods fail to find. It is used in case of large sample size. Its first moment equal to mean, second to variance, third to skewness and forth one is equal to kurtosis.

The drawbacks of the moments method are that this method is not available for every distribution. Moments method estimators are not necessary to be sufficient. Its estimators are stronger but less efficient. The moments method estimators are asymptotically unbiased, but although not the best estimator. It often gives us biased estimators.

### 3.3.3 Method of L-moments

L-moments are the outlook of certain linear combinations of order statistics. And they can be defined for every random variable whose mean exists. The L-moments can be used for summarization and explanation of theoretical probability distributions. Summarization and description of observed data samples, estimation of parameters and quantiles of probability distributions and hypothesis testing. The major advantages of L-moments over conventional moments, undergo less from the effect of sampling changeability. L-moments are stouter than conventional moments when there are outliers in the data. The L-moments are used for extreme events such as rainfall, flood, earth quick, droughts, heat waves, snow fall, famine and income. For income modeling the L-moments are used because L-moments also contain some extreme values.The income model is used for assessment of Living standard. The industrialists are using the income distribution for the market potential and for tax load estimation. It gives more accurate results than other methods of estimations of parameters such as moment method and maximum likelihood method especially for small sample size.

For description of the Statistical measures we use the usual moments, another technique which is based on the use of the other characteristics of moments called L-moments. The L-moments are similar to the usual moments but the difference is that they are linear combination of Order statistics; the L-moments are suitable for both hypothetical and realistic results. For small sample size of estimation the amount of bias is less than the conventional moments. The first four sample L-moments represents mean, variance, skewness and kurtosis of a distribution respectively.

The conventional moments are used to estimate the parameters of a distribution. But the conventional moments for small sample size are not always convenient. The sample conventional moments are used to summarize the mean, variance, skewness and kurtosis. Similarly the sample L-moments are used to summarize the location, scale, skewness, and kurtosis for a specified distribution.

Let $Y$ be a random variable with cumulative distribution function $F(Y)$ and quantile function $Y(F)$ and let $Y_{1:n} \leq Y_{2:n} \leq \cdots \leq Y_{n:n}$ be the order statistics of a random sample of size '$n$' taken from the distribution of $Y$. The L-moments of $Y$ can be defined

$$\lambda_1 = E(Y_{1:1})$$

$$\lambda_2 = \frac{1}{2}E(Y_{2:2} - Y_{1:2})$$

$$\lambda_3 = \frac{1}{3}E(Y_{3:3} - 2Y_{2:3} + Y_{1:3})$$

$$\lambda_4 = \frac{1}{4}E(Y_{4:4} - 3Y_{3:4} + 3Y_{2:4} - Y_{1:4})$$

The general term can be written as

$$\lambda_r = r^{-1}\sum_{i=0}^{r-1}(-1)^i \binom{r-1}{i} E(Y_{r-i:r}), \qquad r = 1,2,\ldots \qquad (3.3.3.1)$$

Now the expectation of an order statistic may be defined as

$$E(Y_{r:n}) = \frac{n!}{(r-1)!(n-r)!}\int_0^1 y(v)v^{r-1}(1-v)^{n-r}dv \qquad (3.3.3.2)$$

The L-moments ratios can be defined as:

$$\tau_r = \frac{\lambda_r}{\lambda_2}, \quad r = 3,4 \ldots\ldots$$

The L-CV is $\tau = \frac{\lambda_2}{\lambda_1}, \quad 0 < \tau < 1$

L-CV is similar to the coefficient of variation.

The probability weighted moments is defined as

$$M_{p,r,s} = E[Y^p\{F(Y)\}^r\{1 - F(Y)\}^s] \tag{3.3.3.3}$$

The Probability weighted moments can also be written as linear combinations of L-moments. So the methods based on L-moments and Probability weighted moments are the same. But the method of L-moments are more suitable for interpretations of measures the shape and scale of probability distributions. $\lambda_1 \ldots\ldots \lambda_r$ And $\tau_3 \ldots\ldots \tau_r$ are the valuable quantities for the probability distributions summarization. $\lambda_1$ is used for location parameter, and $\lambda_2$ for scale parameter, $\tau_3$ for measure of skewness and $\tau_4$ for kurtosis. The probability distribution of L-moments method can be discuss as, suppose that $Y$ be a random variable which can take the values real numbers. The occurrence frequency of these real value numbers is called probability distribution or frequency distribution. And its cumulative distribution function can be defined as $F(y) = \Pr[Y \le y]$, where the $Pr$ is called the probability and the increasing function of $y$ is $F(y)$ such that $0 \le F(y) \le 1 \; \forall y$. The quantile function of $Y$ is given as $F\big(y(v)\big) = v$ where $0 < v < 1$.

If we differentiate $F(y)$ then $\frac{d}{dy}F(y) = f(y)$

And the expected value of random variable $Y$ is expressed as

$$E(Y) = \int_{-\infty}^{\infty} y dF(y) = \int_{-\infty}^{\infty} y f(y) dy \tag{3.3.3.4}$$

If we transform $v = F(y)$ then

$$E(Y) = \int_0^1 y(v) dv \tag{3.3.3.5}$$

The variance of random variable $Y$ is given by

$var(Y) = E[\{Y - E(Y)\}^2]$

The covariance of two random variables $X$ and $Y$ can be expressed as

$cov(X, Y) = E[\{X - E(X)\}\{Y - E(Y)\}]$

And the correlation of $X$ and $Y$ is

$corr(X, Y) = cov(X, Y)/\sqrt{\{var(X)var(Y)}$

### 3.3.3.1 Population Conventional Moments

The shape of the probability distribution is defined by the moments, moments are given below

$\mu = E(y)$ is called mean and the higher moments are

$\mu_r = E(y - \mu)^r$ Where$r = 2,3, \ldots..$

The standard deviation is expressed as

$\sqrt{\mu_2} = \sqrt{\{E(y - \mu)^2\}} = \sigma^2$, And $var(y) = \sigma^2$

Coefficient of variation can be given as

$CV = \sigma/\mu$

And Skewness is $\gamma = {\mu^2}_3/{\mu^3}_2$

Kurtosis $\kappa = \mu_4/{\mu^2}_2$

### 3.3.3.2 L-moments of Probability Distribution

L-moments are another choice for expressing the probability distribution shape. The random variable $Y$ and its Probability weighted moments with cumulative distribution function $F(.)$ to be quantities.

$$M_{p,r,s} = E[Y^p\{F(Y)\}^r\{1 - F(Y)\}^s] \tag{3.3.3.2.1}$$

The Probability weighted moments special case are

$\alpha_r = M_{1,0,r}$ and $\beta_r = M_{1,r,0}$ for a distribution having the quantile function $y(v)$ are given as

$$\alpha_r = \int_0^1 y(v)(1 - v)^r dv \tag{3.3.3.2.2}$$

$$\beta_r = \int_0^1 y(v)v^r dv \tag{3.3.3.2.3}$$

The ordinary moments can be written as:

$$E(Y^r) = \int_0^1 \{y(v)\}^r dv \tag{3.3.3.2.4}$$

The $\alpha_r$ and $\beta_r$ which are the probability weighted moments can be used as a basis for parameters estimation of probability distribution. Certain linear combination of the probability weighted moments can be carried for estimation of scale parameter and skewness.

$\alpha_0 - 2\alpha_1$ or $2\beta_1 - \beta_0$ are the scale parameter and the skewness of a distribution can be expressed as $6\beta_2 - 6\beta_1 + \beta_0$

Let we have a polynomial $P^*{}_r(v),\ r = 0,1,2, \ldots\ldots$

i. $P^*{}_r(v)$ having degree $r$ in $v$

ii. $P^*_r(1) = 1$

iii. $\int_0^1 P^*_r(v)P^*_s(v)dv = 0 \;\; if\; r \neq s$

By Legendre polynomials

$$P^*_r(v) = \sum_{k=0}^{r} P^*_{r,k} v^k \tag{3.3.3.2.5}$$

$$P^*_{r,k} = (-1)^{r-k}\binom{r}{k}\binom{r+k}{k} = \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!} \tag{3.3.3.2.6}$$

Now if $Y$ is a random variable having quantile function $y(v)$, then the L-moments of $Y$ is defined as

$$\lambda_r = \int_0^1 y(v)P^*_{r-1}(v)dv \tag{3.3.3.2.7}$$

### 3.3.3.3 The L-moments in term of Probability Weighted Moments

$\lambda_1 = \alpha_0 = \beta_0$

$\lambda_2 = \alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0$

$\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2 = 6\beta_2 - 6\beta_1 + \beta_0$

$\lambda_4 = \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$

And the general form

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^{r} P^*_{r,k}\,\alpha_k = \sum_{k=0}^{r} P^*_{r,k}\,\beta_k \tag{3.3.3.3.1}$$

### 3.3.3.4 L-moments and Order Statistics

L-moments also find by taking into consideration the linear combinations of observations in a sample which is selected from the data and arrange in increasing order. Let $Y_{k:n}$ be a sample of size $n$ selected from the data. And then the samples in order are $Y_{1:n} \leq Y_{2:n} \leq \cdots \leq Y_{n:n}$.s

Where $Y_{1:1}$ having the information about the location of a distribution. It is a sample of size one which have single observation and $Y_{1:2}$ and $Y_{2:2}$ is the sample of size two having two observations. It has the information about the dispersion or scale of a distribution. For the skewness and kurtosis the sample will be of size three and four which have three and four observations respectively, that are $Y_{1:3} \leq Y_{2:3} \leq Y_{3:3}$ and $Y_{1:4} \leq Y_{2:4} \leq Y_{3:4} \leq Y_{4:4}$. So the L-moments are the expected values of linear combinations. The L-moments of a probability distribution are given as

$$\lambda_1 = E(Y_{1:1})$$

$$\lambda_2 = \frac{1}{2}E(Y_{2:2} - Y_{1:2})$$

$$\lambda_3 = \frac{1}{3}E(Y_{3:3} - 2Y_{2:3} + Y_{1:3})$$

$$\lambda_4 = \frac{1}{4}E(Y_{4:4} - 3Y_{3:4} + 3Y_{2:4} - Y_{1:4})$$

And for general

$$\lambda_r = r^{-1}\sum_{i=o}^{r-1}(-1)^i \binom{r-1}{i} E(Y_{r-i:r}) \tag{3.3.3.4.1}$$

### 3.3.3.5 Properties of L-moments

The $\lambda_1, \lambda_2$ which are L-moments and $\tau$ is L-CV and $\tau_3$ and $\tau_4$ are the L-moments ratios are the practical values for expressing the probability distributions. They have the following properties.

i. Existence: The L-moments for a distribution will be existing when mean of that distribution exists.

ii. Uniqueness: The L-moments exclusively describe distribution if its mean exists. And the different distributions have different L-moments.

iii. Terminology: For the mean of a distribution is $\lambda_1$,for the variance of a distribution is $\lambda_2$, L.CV is $\tau$, for the L-skewness $\tau_3$, for L-kurtosis $\tau_4$.

iv. Numerical value: The mean $\lambda_1$ can take any value. $\lambda_2$ which is the variance having no negative value that is $\lambda_2 \geq 0$ and $0 \leq \tau < 1$ for a distribution only takes positive values. And $|\tau_r| < 1$ for all $r \geq 3$ that is the L-moments ratio. And the bound for $\tau_3$ and $\tau_4$ given as $\frac{1}{4}(5\tau_3{}^2 - 1) \leq \tau_4 < 1$.

v. Linear transformation: Suppose that $X$ and $Y$ are the two random variables having L-moments $\lambda_r$ and $\lambda_{r^*}$, $Y = aX + b$, then $\lambda_{1^*} = a\lambda_1 + b$, $\lambda_{2^*} = |a|\lambda_2$ .

vi. Symmetry: Suppose $Y$ be a random variable which is symmetrical having mean 'μ' and $\Pr[Y \geq \mu + y] = \Pr[Y \leq \mu - y]$ $\forall\, y$, then all odd order L-moments ratios of $Y$ will be zero, that are $\tau_r = 0, r = 3,5 \ldots..$

### 3.3.3.6 Sample L-moments

Sample Probability weighted moment $b_r$ is an unbiased estimator of $\beta_r$. $\beta_r$ is called population probability weighted moment.

$b_r = n^{-1}\binom{n-1}{r}^{-1}\sum_{i=r+1}^{n}\binom{i-1}{r} y_{i:n}$ (3.3.3.6.1)

By (Landwehr et al.1979a) the sample probability weighted moments may be written as

$b_0 = n^{-1}\sum_{i=1}^{n} y_{i:n}$

$b_1 = n^{-1}\sum_{i=2}^{n}\frac{(i-1)}{(n-1)} y_{i:n}$

$b_2 = n^{-1}\sum_{i=3}^{n}\frac{(i-1)(i-2)}{(n-1)(n-2)} y_{j:n}$

And for the general term we may be written as

$b_r = n^{-1}\sum_{i=r+1}^{n}\frac{(i-1)(i-2)\ldots\ldots(i-r)}{(n-1)(n-2)\ldots\ldots(n-r)} y_{i:n}$ (3.3.3.6.2)

The sample L-moments may also be defined as

$l_1 = b_0$

$l_2 = 2b_1 - b_0$

$l_3 = 6b_2 - 6b_1 + b_0$

$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0$

And for the general term

$l_{r+1} = \sum_{k=0}^{r} P^{*}{}_{r,k}\, b_k$, where $r = 0,1,\ldots\ldots,n-1$ (3.3.3.6.3)

And the $l_r$ is the sample L-moment is the unbiased estimator of $\lambda_r$ .The sample L-moments

ratios are $t_r = \frac{l_r}{l_2}$

And the L-CV $t = \frac{l_2}{l_1}$

These are the estimators of $T_r$ and T. These are not the unbiased estimator, but have very small bias.

# Chapter 4

## Results and Discussion

### 4.1 Run test for Independence

The Run test is used to test the independence of data. It is used to decide that the data set from random process or not. The test statistics for run test is given as:

$$Z = \frac{R-\overline{R}}{s_R} \tag{4.1.1}$$

Where $R$ for observed runs and $\overline{R}$ for expected runs, $S_R$ is the standard deviation of runs. Where $\overline{R} = \frac{2n1n2}{n1+n2} + 1$ and $s^2{}_R = \frac{2n1n2(2n1n2-n1-n2)}{(n1+n2)^2(n1+n2-1)}$

Where $n1$ and $n2$ are the numbers of positive and negative runs respectively. We reject the null hypothesis if $|Z| > Z_{1-\alpha/2}$

$H0$: Data are randomly distributed, $Ha$: Data are not randomly distributed As the computed P-value is greater than the significance level $\alpha = 0.05$, so we cannot reject the null hypothesis $H0$.

### 4.2 Tests for Homogeneity

### 4.2.1 Mann- Whitney Test

The Mann-Whitney is a non-parametric test. It used to check the homogeneity of the two populations with respect to location parameter. It is used to check that the two populations have the same mean or not. So in our analysis the income data comes from the same distribution. The Mann- Whitney test statistic as

$$U = n_1 n_2 + \left[\frac{n_1(n_1+1)}{2}\right] - T \qquad (4.2.1.1)$$

Where $n_1$ *and* $n_2$ the first and second samples respectively. T is denotes the sum of ranks for the first sample.

### 4.2.2 Ansari Bradley Test

Ansari Bradley test is used to test the homogeneity of two populations with respect to scale parameter. It is used to check that the two populations come from the same distribution with identical variance. In our analysis the Pakistani income data come from the same distribution. The Ansari Bradley test statistic is given as

$$A_N = \sum_{l}^{N} |(l - \frac{N+1}{2})| Z_l \qquad (4.2.2.1)$$

## 4.3 Methods for Fitting Distributions by Goodness of Fit Tests

Different methods are used to test the goodness of fit of distributions. These are Chi Square test, Kolmogorov Smirnov test, Anderson Darling test and L-moment ratio diagram. We check the three distributions that are generalized gamma, generalized pareto and lognormal3 distribution by goodness of fit tests to find the best fit distribution for the income data in Pakistan. The goodness of fit tests are used to test whether the sample data come from the specified distribution or not.

Anderson Darling test statistic is $A^2 = -\mathrm{N} - \mathrm{S}$ (4.3.1)

Where $S = \sum_{i=1}^{N} \frac{(2i-1)}{N} [lnF(Y_i) + \ln(1 - F(Y_{N+1-i}))]$

Chi- Square test statistic is $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$ (4.3.2)

Where $O_i$ = Observed data, $E_i$= Expected data and $\sigma^2$= variance of the observation.

Kolmogorove-Smirnov test statistic is $D_n = \max|S_n(x) - F(x)|$ (4.3.3)

**Table 4.1** Fitting of Distribution by Goodness of fit tests.

| Distributions | Kolmogorov Smirnov | Anderson Darling | Chi- Square |
|---|---|---|---|
| Generalized Gamma | 0.14608 | 152.7 | 1029.40 |
| Generalized Pareto | 0.09251 | 1194.5 | NA |
| Lognormal3 | 0.06257 | 17.362 | 190.06 |

According to the criteria available in Easy fit software for selection of best fit distribution. we will finalized that distribution which shows minimum value of the test statistic, as shown in above table, using a particular methods of estimation. So the best fit distribution is three parameters lognormal distribution, because its values for all test statistics is less than remaining two other distributions.

## 4.4 L-moments Ratio Diagram

After the selection of best fit distribution by the goodness of fit test. Then we find the best fit distribution by the method of L-moments ratio diagram. L- moment ratio diagram is a graphical method used to compare L-skewness and L-kurtosis of different distributions. It gives us visual indication about the best fit distribution. It gives us the information that which distribution is best fit for the data. We can say that L-moments ratio diagram is used as a goodness of fit measure for a distribution. The L-moment ratio can be obtained by higher order L-moment divided by measure of dispersion that as $t_r = \frac{l_r}{l_2}$. The $t_3$ used for skewness and $t_4$ used for kurtosis. So in this case the best fit distribution is three parameters lognormal distribution that is shown in the diagram. After the selection of best fit distribution, we estimate the three parameters of lognormal distribution by the L-moments method, moments method and maximum likelihood that are shown below. The positive sign shows that the data follow three parametric lognormal distribution curve.

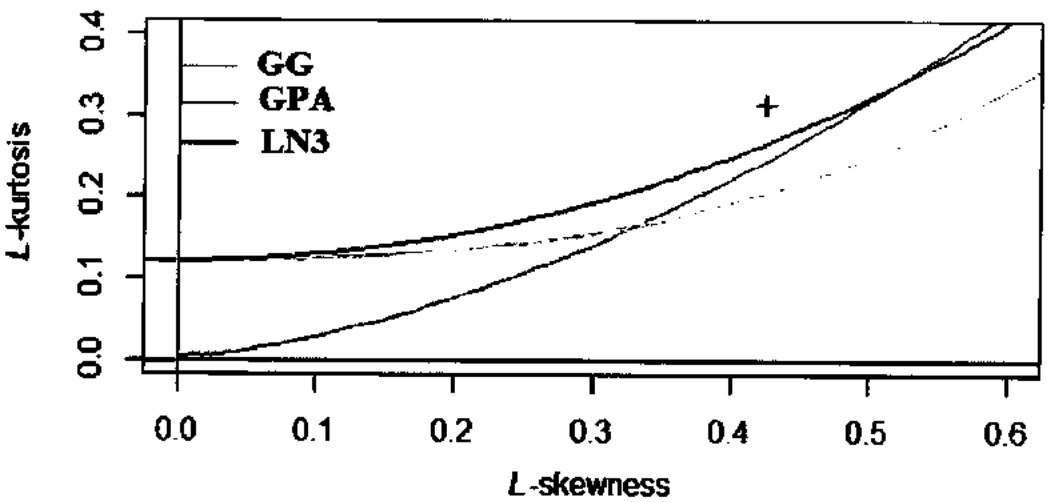



**Fig. 4.1** L-moments ratio diagram for selection of best fit distribution

## 4.5 Sample L-moment, L-moment Ratios and Quartiles of the Annual Income of Combined Samples (Male and Female) in Pakistan

**Table 4.2** L-moments ratios and quartiles for combined (male and female) income

| $l_1$ | $t$ | $t_3$ | $t_4$ | $Q_1$ | $Q_2$ | $Q_3$ |
|---|---|---|---|---|---|---|
| 127200 | 0.4223073 | 0.4244511 | 0.3135346 | 60000 | 96000 | 144000 |

In table 4.2, $l_1$ represent the mean of the data. $t$ is L-CV which is similar to the coefficient of variation. And the value of $t$ will be $0 \leq t < 1$. As shown in the above table the value of L-CV is 0.4223 which indicates that there is inequality in the income data. The skewness and kurtosis are used to describe the shape of a distribution. When the kurtosis is positive, then the distribution will be peaked. And when the kurtosis is negative the distribution will be flat. The $t_4$ is used for L-kurtosis, $t_3$ is the L-skewness and its value will be $-1 < t_3 < 1$. For symmetric distribution the value of $t_3 = 0$. And for negatively skewed distribution its value will be negative. For positively skewed distribution the value of $t_3$ will be positive. The distribution of income is a positively skewed so the value of $t_3$ =0.4245. Its mean that most of the peoples income will be on the left side of the mean. The extreme income means the high income will be on the right side of the mean. $Q_1$, $Q_2$, $Q_3$ are the three quartiles of the combined sample (male and female) income data as shown in the above table.

### 4.5.1 Parameters Estimation of the three Parameters Lognormal Distribution for Combined Samples (Male and Female) of Income Data

**Table 4.3** Parameters estimation by different methods of three parameters lognormal distribution for combined (male and female) income

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ |
|---|---|---|---|
| L-moments | 11.21 | 0.91 | 15216 |
| Moments | 11 | 1 | 26390.2 |
| Maximum likelihood | 11.5 | 0.7 | -5854 |

In table 4.3, $\hat{\mu}$ = estimate of scale parameter, $\hat{\sigma}$ = estimate of shape parameter, $\hat{\gamma}$ = estimate of location parameter. These are the estimates of three parameters of lognormal distribution for combined (male and female) income data in Pakistan. The skewness and kurtosis of the lognormal distribution increases if the value of shape parameter $\sigma$ increases. The skewness will increase if the value of $\sigma$ increases for a given value of scale parameter $\mu$. But when the value of scale parameter $\mu$ increases, then the dispersion of the distribution will be increased. The location parameter $\gamma$ shifts the probability density function of the distribution. If we put $\gamma = 0$ then the three parametric lognormal distribution become two parametric lognormal distribution. The very important property of the lognormal distribution is that as the value of shape parameter $\sigma$ decreases, then the shape of the distribution will be close to Normal distribution.

### 4.5.2 Probability Distribution Plot for Combined Samples (Male and Female) of Income Data

The Probability distribution plot is used for the interpretation of data. It gives us the information that the distribution is skewed or not. For the income data that are shown in the distribution plot, the horizontal axis represents the income of different groups of peoples and the vertical axis represents the probabilities. From distribution plot we also illustrate the dispersion in the data. In the following diagram some groups of peoples have more income and some are less income.

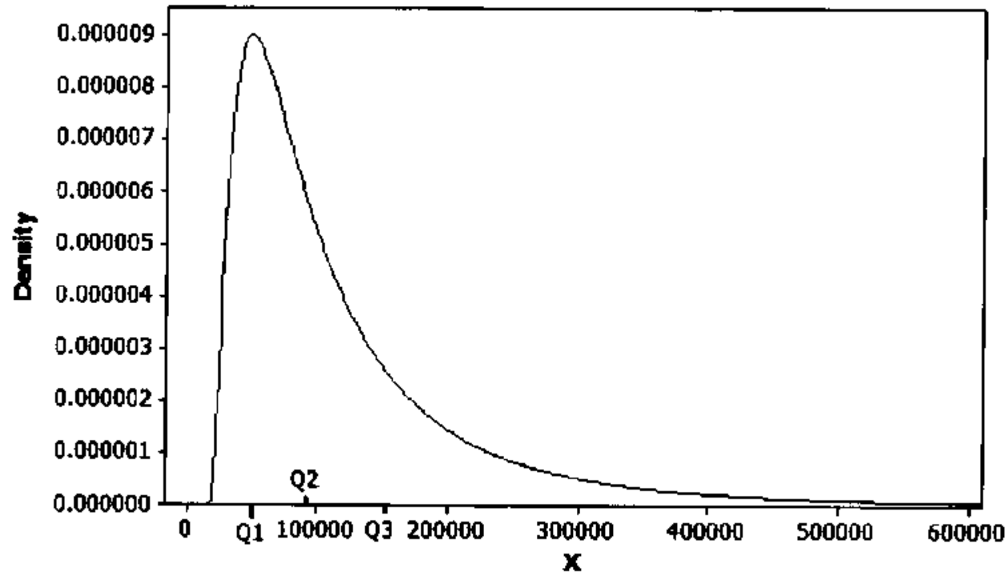



**Fig. 4.2** Probability distribution plot of combined samples (male and female) income

From the above plot we conclude that the distribution of the combined sample (male and female) income follows the positively skewed distribution. The 25% of the total population has income below 60000. And 50% of the total population has income less than100000. 75% of the total population income is below than 144000. And the last 25% of the total population that have income above than 145000. Very small number of people they have income above 400000.

## 4.6 Sample L-moment, L-moment Ratios and Quartiles of Male Income in Pakistan

**Table 4.4** L-moments ratios and quartiles for male income

| $l_1$ | $t$ | $t_3$ | $t_4$ | $Q_1$ | $Q_2$ | $Q_3$ |
|---|---|---|---|---|---|---|
| 131900 | 0.4047008 | 0.4399626 | 0.3190783 | 62400 | 96000 | 151200 |

In table 4.4, $l_1$ is the mean of the data. The value of $t$ is 0.4047 that indicates that there is inequality in the male income data. $t_3$ is the L-skewness and its value is 0.4399 which is positive.

So the male income data distribution will be positively skewed. It indicates that most of the peoples income will be on the left side of the mean. The extreme income means the high income will be on the right side of the mean, $Q_1$,$Q_2$,$Q_3$ are the three quartiles of the male income data as shown in the above table.

### 4.6.1 Parameters Estimation of the three Parameters Lognormal Distribution for Male Income Data

**Table 4.5** Parameters estimation by different methods of three parameters lognormal distribution for male income

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ |
|---|---|---|---|
| L-moments | 11.14 | 0.94 | 24366.17 |
| Moments | 11 | 1 | 31449.8 |
| Maximum likelihood | 11.5 | 0.7 | -2001 |

In table 4.5, $\hat{\mu}$ = estimate of scale parameter, $\hat{\sigma}$ = estimate of shape parameter, $\hat{\gamma}$ = estimate of location parameter. These are the estimates of three parameters lognormal distribution for male income data in Pakistan. The skewness and kurtosis of the lognormal distribution increases if the value of shape parameter $\sigma$ increases. when the value of scale parameter $\mu$ increases, then the dispersion of the distribution will be increased.

### 4.6.2 Probability Distribution Plot for Male Income Data

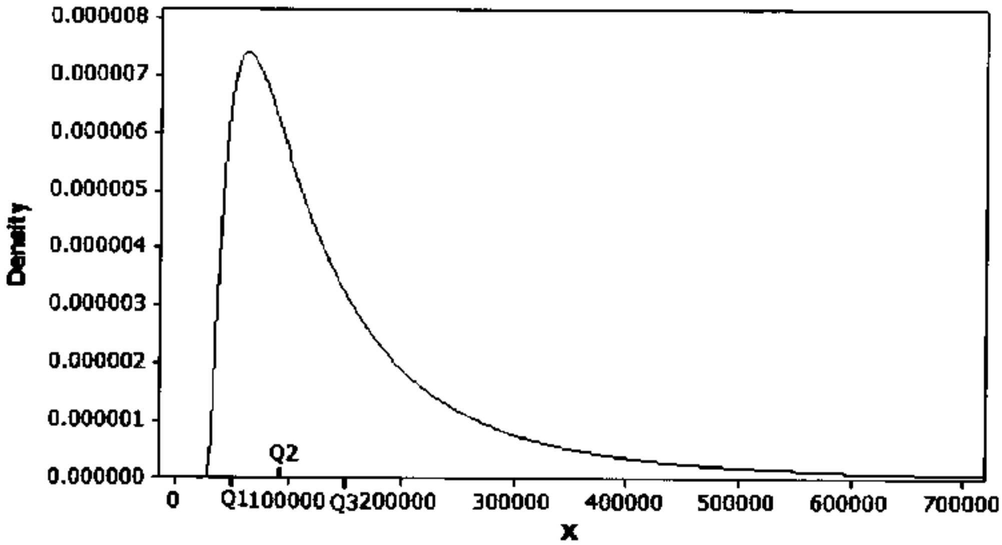



**Fig. 4.3** Probability distribution plot of male income

From the above figure we conclude that 25% 0f male income is less than 62000. 50% of people have income below than 96000. And 75% of the total population income is less 150000. The last 25% of the population that have income above than 150000.

## 4.7 Sample L-moment, L-moment Ratios and Quartiles of the Annual Income of Female in Pakistan

**Table 4.6** L-moments ratios and quartiles for female income

| $l_1$ | $t$ | $t_3$ | $t_4$ | $Q_1$ | $Q_2$ | $Q_3$ |
|---|---|---|---|---|---|---|
| 79352.53 | 0.5651828 | 0.5235219 | 0.3122305 | 24000 | 36000 | 96000 |

In table 4.6, the average income of female is less than the male income. The value of $t$ is 0.5652 that indicates that there is inequality in the female income data. As we compare the variation of male income with female income, then there is large variation in female income as compared with male income. The L-CV of male income is less than the female. The skewness of female income is greater than male income. $t_3$ is the L-skewness and its value is 0.5235 which is positive. So the female income data distribution will be positively skewed. It indicates that most of the female income will be on the left side of the mean. And the extreme income means the high income will be on the right side of the mean, $Q_1$,$Q_2$, $Q_3$ are the three quartiles of the female income data.

### 4.7.1 Parameters Estimation of three Parameters Lognormal Distribution for Female Income Data

**Table 4.7** Parameters estimation by different methods of three parameters lognormal distribution for female income

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ |
|---|---|---|---|
| L-moments | 10.58 | 1.15 | 2663.74 |
| Moments | 11.4 | 0.8 | -38872.5 |
| Maximum likelihood | 10.7 | 1.1 | 605 |

In table 4.7, $\hat{\mu}$ = estimate of scale parameter, $\hat{\sigma}$ = estimate of shape parameter, $\hat{\gamma}$ = estimate of location parameter. These are the estimates of three parameters lognormal distribution for female income data in Pakistan. The skewness and kurtosis of the lognormal distribution increases if the value of shape parameter $\sigma$ increases. But when the value of scale parameter $\mu$ increases, then the dispersion of the distribution will be increased.

### 4.7.2 Probability Distribution Plot for Female Income Data

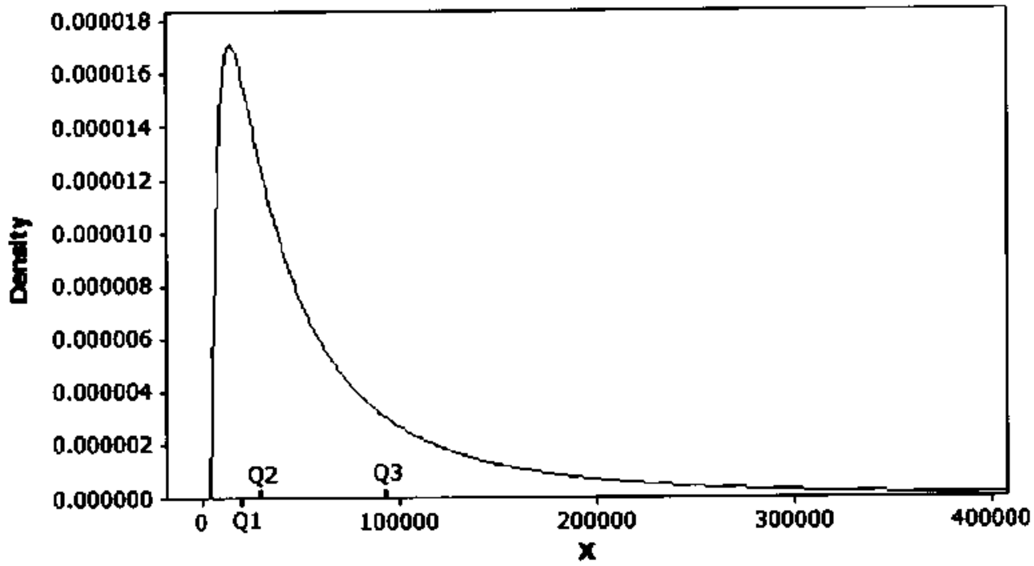



**Fig. 4.4** Probability distribution plot of female income

From the above female income plot we see that first 25% of the population has income below than 24000. And 50% of the population income less than 36000.The 75% of the total female population income less than 96000. And last 25% of the female population that have income greater than 96000. As we see that the male income is greater than female income in Pakistan.

## 4.8 Province Wise Income of Combined Samples (Male and Female)

### 4.8.1 Sample L-moment, L-moment Ratios and Quartiles of the Annual Income of Punjab Province

**Table 4.8** L-moments ratios and quartiles for Punjab province income

| $l_1$ | $t$ | $t_3$ | $t_4$ | $Q_1$ | $Q_2$ | $Q_3$ |
|---|---|---|---|---|---|---|
| 136516 | 0.4550447 | 0.4440456 | 0.3255626 | 60000 | 96000 | 156000 |

$l_1$ represents the mean of the data. $t$ is L-CV, $t_4$ is used for L-kurtosis and $t_3$ is the L-skewness. $Q_1$, $Q_2$, $Q_3$ represents the three quartiles.

### 4.8.1.1 Parameters Estimation of the three Parameters Lognormal Distribution for Income Data in Punjab Province

**Table 4.9** Parameters estimation by different methods of three parameters lognormal distribution for Punjab province income

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ |
|---|---|---|---|
| L-moments | 11.27 | 0.95 | 12472.48 |
| Moments | 11.2 | 1 | 15653.7 |
| Maximum likelihood | 11.5 | 0.8 | -6343 |

In table 4.9, $\hat{\mu}$ = estimate of scale parameter, $\hat{\sigma}$ = estimate of shape parameter, $\hat{\gamma}$ = estimate of location parameter. These are the estimates of three parameters lognormal distribution for Punjab province income data.

### 4.8.1.2 Probability Distribution Plot for Punjab Province Income Data

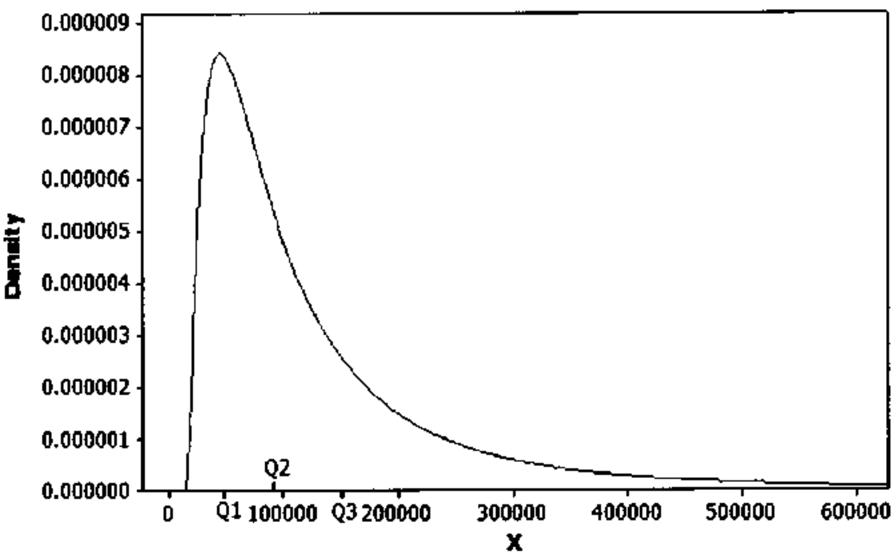



**Fig. 4.5** Probability distribution plot of Punjab province income

From the Punjab province income plot we see that first 25% of the population has income below than 60000. And 50% of the population income less than 96000.The 75% of the population income less than 156000. Last 25% of the Punjab province population have income greater than 156000. In Punjab province less than 5% of the total population that have income greater than 500000. So large numbers of peoples have minimum income.

### 4.8.2 Sample L-moment, L-moment Ratios and Quartiles of the Annual Income of Sindh Province

**Table 4.10** L-moments ratios and quartiles for Sindh province income

| $l_1$ | $t$ | $t_3$ | $t_4$ | $Q_1$ | $Q_2$ | $Q_3$ |
|---|---|---|---|---|---|---|
| 114050 | 0.4324113 | 0.4718103 | 0.3461704 | 54000 | 75000 | 124800 |

Where $l_1$ represents the mean of the data. $t$ is L-CV, the $t_4$ is used for L-kurtosis, $t_3$ is the L-skewness. $Q_1$, $Q_2$, $Q_3$ represents the three quartiles.

#### 4.8.2.1 Parameters Estimation of the three Parameters Lognormal Distribution for Sindh Province

**Table 4.11** Parameters estimation by different methods of three parameters lognormal distribution for Sindh province income

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ |
|---|---|---|---|
| L-moments | 10.92 | 1.02 | 21078.27 |
| Moments | 10.9 | 1.1 | 20241.3 |
| Maximum likelihood | 11.3 | 0.8 | 226.3 |

In table 4.11, $\hat{\mu}$ = estimate of scale parameter, $\hat{\sigma}$ = estimate of shape parameter, $\hat{\gamma}$ = estimate of location parameter. These are the estimates of three parameters lognormal distribution for Sindh province income data.

### 4.8.2.2 Probability Distribution Plot for Sindh Province Income Data

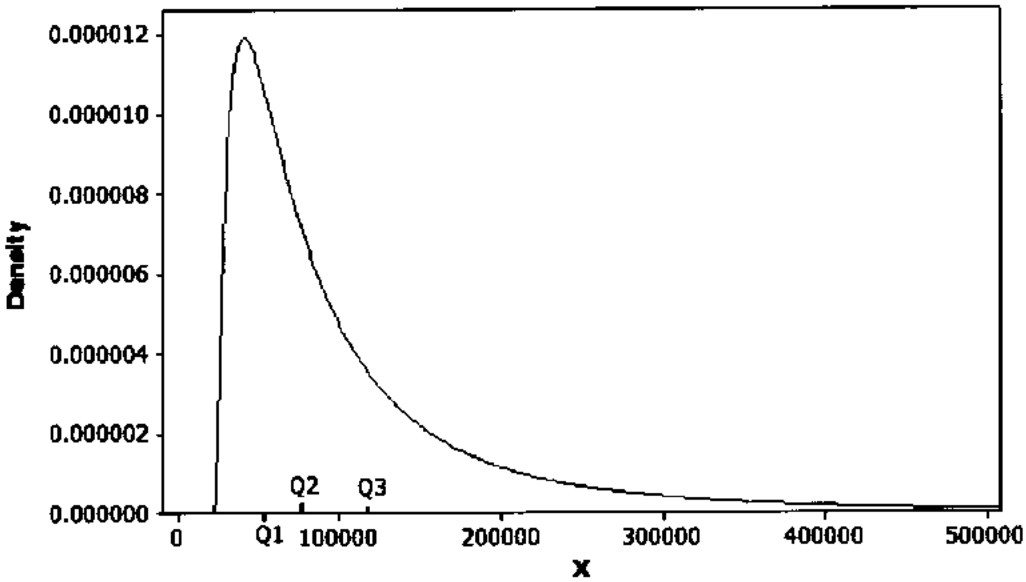



**Fig. 4.6** Probability distribution plot of Sindh province income

From Sindh province income plot we see that first 25% of the population has income less than 54000. And 50% of the population income below than 75000. The 75% of the population income less than 124000. The last 25% of the Sindh province population has income greater than 124000.

### 4.8.3 Sample L-moment, L-moment ratios and Quartiles of the annual income of KPK province

**Table 4.12** L-moments ratios and quartiles for KPK province income

| $l_1$ | $t$ | $t_3$ | $t_4$ | $Q_1$ | $Q_2$ | $Q_3$ |
|---|---|---|---|---|---|---|
| 128561 | 0.4097654 | 0.3784994 | 0.271213 | 60000 | 96000 | 156000 |

Where $l_1$ represents the mean of the data. $t$ is L-CV, the $t_4$ is used for L-kurtosis, $t_3$ is the L-skewness. $Q_1$, $Q_2$, $Q_3$ represents the three quartiles.

#### 4.8.3.1 Parameters Estimation of the three Parameters Lognormal Distribution for KPK Province

**Table 4.13** Parameters estimation by different methods of three parameters lognormal distribution for KPK province income

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ |
|---|---|---|---|
| L-moments | 11.39 | 0.80 | 5961.28 |
| Moments | 11.1 | 1 | 24876.7 |
| Maximum likelihood | 11.6 | 0.7 | -9481.5 |

In table 4.13, $\hat{\mu}$ = estimate of scale parameter, $\hat{\sigma}$ = estimate of shape parameter, $\hat{\gamma}$ = estimate of location parameter. These are the estimates of three parameters lognormal distribution for KPK province income data.

### 4.8.3.2 Probability Distribution Plot for KPK Province Income Data

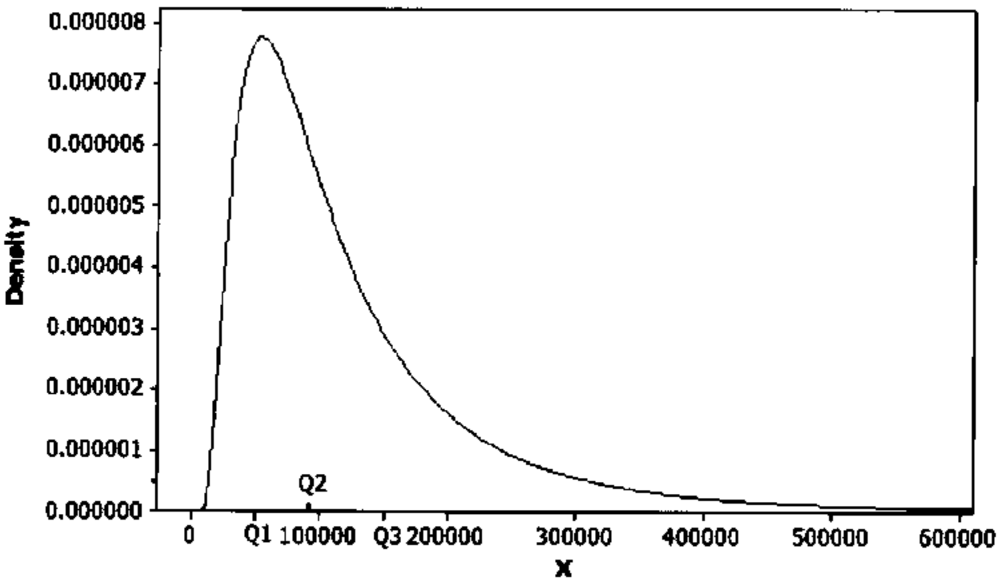



**Fig. 4.7** Probability distribution plot of KPK province income

From KPK province income plot as we see that first 25% of the population has income less than 60000. And 50% of the population income below than 96000. The 75% of the population income less than 155000. Last 25% of the KPK province population that have income greater than 155000.

### 4.8.4 Sample L-moment, L-moment Ratios and Quartiles of the Annual Income of Baluchistan Province

**Table 4.14** L-moments ratios and quartiles for Baluchistan province income

| $l_1$ | $t$ | $t_3$ | $t_4$ | $Q_1$ | $Q_2$ | $Q_3$ |
|---|---|---|---|---|---|---|
| 127360 | 0.3015520 | 0.3179764 | 0.2387240 | 78000 | 108000 | 154800 |

Where $l_1$ represents the mean of the data. $t$ is L-CV, the $t_4$ is used for L-kurtosis, $t_3$ is the L-skewness, $Q_1$, $Q_2$, $Q_3$ represents the three quartiles.

#### 4.8.4.1 Parameters Estimation of the three Parameters Lognormal Distribution for Baluchistan Province

**Table 4.15** Parameters estimation by different methods of three parameters lognormal distribution for Baluchistan province income

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ |
|---|---|---|---|
| L-moments | 11.35 | 0.67 | 21505.47 |
| Moments | 11.3 | 0.7 | 23618.6 |
| Maximum likelihood | 11.6 | 0.5 | -3981.4 |

In table 4.15, $\hat{\mu}$ = estimate of scale parameter, $\hat{\sigma}$ = estimate of shape parameter, $\hat{\gamma}$ = estimate of location parameter. These are the estimates of three parameters lognormal distribution for Baluchistan province income data.

#### 4.8.4.2 Probability Distribution Plot for Baluchistan Province Income Data

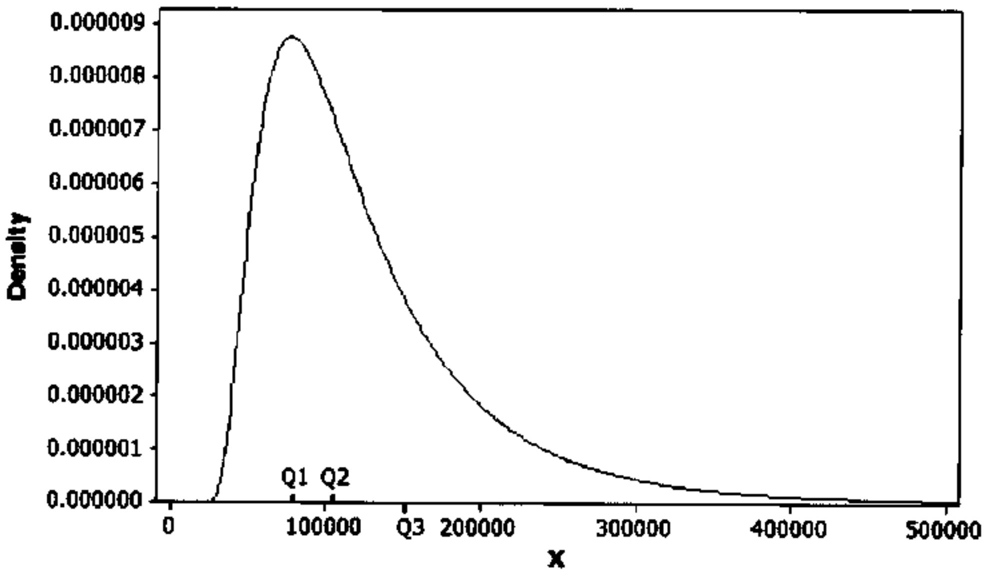



**Fig. 4.8** Probability distribution plot of Baluchistan province income

From the Baluchistan province income plot as we see that first 25% of the population has income below than 78000. 50% of the population income below than 108000. And the 75% of the population income less than 154000. Last 25% of the Baluchistan province population that have income greater than 154000.

## 4.9 Comparison of Estimation Methods

In last we compare the estimated parameters of lognormal distribution obtained by L-moments method, moments method and maximum likelihood method. We compare the estimates of parameters with respect to Absolute Bias and RMSE. The Bias of an estimator $\hat{\theta}$ is the difference between the expected value of an estimator and the parameter value. The Bias can be calculated as, $bias(\hat{\theta}\,) = E(\hat{\theta}) - \theta$. If an estimator has zero Bias its means it is an unbiased estimator. The RMSE is used to represent the difference between sample and population values. The RMSE is a measure of accuracy. The RMSE of an estimator $\hat{\theta}$ can be expressed as RMSE($\hat{\theta}$) = $\sqrt{MSE(\hat{\theta})} = \sqrt{E((\hat{\theta}-\theta)^2)}$.

We take different sample sizes from the data. Also take the random numbers of the same size that we selected from the real income data. Then we find the absolute Bias and RMSE. So the best method will be that whose Absolute Bias and RMSE are small. So for small sample size the best method is L-moments. And for large sample size the best method is the maximum likelihood.

**Table 4.16** Absolute bias and RMSE for different methods of estimation of parameters for sample size $n = 20$

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ | $\lvert Bias \rvert$ | RMSE |
|---|---|---|---|---|---|
| L-moments | 11.59 | 0.76 | -8390 | 35892 | 162128 |
| Moments | 12.40 | 0.40 | -127937 | 151951 | 231421 |
| Maximum likelihood | 11.64 | 0.72 | -11172 | 69622 | 211036 |

When n=20, small sample size the best method of estimation is L-moments. Because its absolute bias and RMSE are less than the moments method and maximum likelihood method.

**Table 4.17** Absolute bias and RMSE for different methods of estimation of parameters for sample size $n = 30$

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ | $\lvert Bias \rvert$ | RMSE |
|---|---|---|---|---|---|
| L-moments | 11.25 | 1.10 | 15673 | 39066 | 233841 |
| Moments | 12.06 | 0.73 | -68074 | 74522 | 293810 |
| Maximum likelihood | 11.22 | 1.18 | 17360 | 83800 | 479023 |

When n=30 then in this case the L-moments method for estimation of parameters is best. Its absolute bias and RMSE are less than the other methods.

**Table 4.18** Absolute bias and RMSE for different methods of estimation of parameters for sample size $n = 50$

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ | $\lvert Bias \rvert$ | RMSE |
|---|---|---|---|---|---|
| L-moments | 11.59 | 0.71 | -8310 | 44740 | 182374 |
| Moments | 12.34 | 0.39 | -116996 | 129186 | 192291 |
| Maximum likelihood | 11.48 | 0.78 | 980 | 12203 | 152556 |

For n=50 the best method of parameters estimation is the maximum likelihood. Its absolute bias and RMSE are less than the L-moments method and moments method as shown in the above table.

**Table 4.19** Absolute bias and RMSE for different methods of estimation of parameters for sample size $n = 100$

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ | $\lvert Bias \rvert$ | RMSE |
|---|---|---|---|---|---|
| L-moments | 11.51 | 0.70 | 380 | 13826 | 145428 |
| Moments | 11.18 | 0.88 | 23264 | 27413 | 157953 |
| Maximum likelihood | 11.62 | 0.63 | -8860 | 3176 | 119216 |

But when n=100 means when sample size increases the best method of estimation of parameters is maximum likelihood method.

**Table 4.20** Absolute bias and RMSE for different methods of estimation of parameters for sample size $n = 250$

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ | $\lvert Bias \rvert$ | RMSE |
|---|---|---|---|---|---|
| L-moments | 11.11 | 0.91 | 20654 | 17299 | 233256 |
| Moments | 11.10 | 0.98 | 14758 | 3962 | 199615 |
| Maximum likelihood | 11.52 | 0.65 | -6640 | 2095 | 158376 |

And when n=250 or above than 250 means by increases sample size the best method of estimation of parameters is maximum likelihood, than L-moments method and moments method. Because its absolute bias and RMSE is less.

**Table 4.21** Absolute bias and RMSE for different methods of estimation of parameters for sample size $n = 500$

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ | $\lvert Bias \rvert$ | RMSE |
|---|---|---|---|---|---|
| L-moments | 11.07 | 1.06 | 22652 | 12180 | 282549 |
| Moments | 11.29 | 1 | 3702 | 10710 | 242800 |
| Maximum likelihood | 11.45 | 0.79 | 333 | 5235 | 227193 |

**Table 4.22** Absolute bias and RMSE for different methods of estimation of parameters for sample size $n = 1000$

| Methods of estimation | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ | $\lvert Bias \rvert$ | RMSE |
|---|---|---|---|---|---|
| L-moments | 11.20 | 0.90 | 15784 | 15859 | 181929 |
| Moments | 11.30 | 0.90 | 10520 | 12502 | 173879 |
| Maximum likelihood | 11.50 | 0.80 | -2612 | 2717 | 171754 |

# CHAPTER 5

## Summary and Conclusions

Income reveals the living standard of the peoples of a country. We are then able to compare their lifestyle with the rest of the world. It gives the estimation about the performance of a state in economic years. Without income growth a country cannot achieve its goals to improve the living standard of citizens. The development of national income depends upon the improvement in education, industrial sector development and agricultural growth.

For current study, the Income data are taken from Pakistan Social and Living Standard Measurement Survey (PSLSMS) of 2010-2011 of the Pakistan Bureau of Statistics (PBS) Islamabad. First of all we apply the run test on income data for randomness. So there is randomness in the income data of Pakistan. We also apply the Mann-Whitney test for checking the homogeneity of the two populations with respect to location parameter. Also apply Ansari Bradley test for the homogeneity of the two populations with respect to scale parameter. After the randomness, we fit three distributions such as generalized gamma, generalized pareto and three parameters lognormal distributions. We check the best fit distribution by goodness of fit test that are Kolmogorov Smirnow, Anderson Darling and Chi square. The best fit distribution is the three parameter lognormal. Incomes within countries generally adopt a skewed distribution with a long heavy tail. Lognormal distribution has been found best fit on the average for income distribution in Pakistan.

After the selection of best fit distribution by the goodness of fit test. We also find the best fit distribution by the L-moments ratio diagram. So in this case the best fit distribution is lognormal3 distribution. We estimate the parameters of lognormal distribution by different methods of estimation. These are the method of L-moments, method of moments and method of maximum likelihood. We estiamte the parameters by these three methods for all male and female in the whole Pakistan. For male and female in four provinces separately. Also find the probability distribution plot of the income data. The citizens of Pakistan are not equal in income. A few groups of peoples are very rich and some are very poor. There is inequality in income. From the probability plot we conclude that the income of male is greater than female in Pakistan. Large number of people whose income is minimum. And few numbers of peoples whose income is maximum. The income of Punjab province is greater than all provinces preceding by Baluchistan, KPK and Sindh respectively.

Finally we compare the estimated parameters of lognormal distribution obtained by method of L-moments, method of moments and method of maximum likelihood, with respect to Absolute Bias and RMSE for different sample size. So the best method of parameters estimation will be that whose Absolute Bias and RMSE are less. So in case of small sample size, the best method of parameters estimation is the L-moments. And for large sample size the best method is maximum likelihood. For large sample size the method of maximum likelihood gives consistent, unbiased and efficient estimators. The method of moments is also used in case of large sample size, but its estimators are not efficient and not necessary to be sufficient.

# References


Aitchison, J., & Brown, J. A. (1957). "The Lognormal distribution with special reference to its uses in economics".

Basak, P., Basak, I., & Balakrishnan, N. (2009). "Estimation for the three-parameter Lognormal distribution based on progressively censored data". Computational Statistics & Data Analysis, 53(10), 3580-3592.

Bilkova, D. (2011). "L-moments and their use in modeling the distribution of income and wage", ISI held on, 21-26.

Bilkova, D. (2008). "Application of Lognormal curves in modeling of wage distribution", journal of Applied Mathematics, 1(2), 341-352.

Bilkova, D. (2011). "Analysis of the distribution of income in recent years in the Czech Republic by region", International days of Statistics and Economics, 286-297.

Bilkova, D. (2011). "Estimating parameters of Lognormal distribution use the method of L-moments", 04-09.

Bilkova, D. (2011). "Modelling of income and wage distribution using the method of L-moments of parameter estimation", International days of Statistics and Economics, 40-50.

Bilkova, D. (2011). "Three parametric Lognormal distribution and estimating its parameters using the methods of L-moments", 05-06.

Bilkova, D. (2011). "Uses of the L-moment method in modeling the wage distribution", journal of Applied Mathematics", Aplimat held on 01-04.

## References


Bilkova, D., and Mala, I. (2012). "Application of the Method of L-moments when modeling the income distribution in the Czech Republic", Austrian journal of Statistics, 125-132.

Bilkova, D. (2012). "Lognormal distribution and using L-moment method for estimation its parameters", International journal of Mathematical models and methods in Applied Sciences, 6(1), 30-44.

Bilkova, D. (2012). "Modeling of wage and income distribution using the method of L-moments", Journal of Mathematics and System Science, 2(1), 13-19.

Bilkova, D., and Mala, I. (2012). "Modeling the income distributions in the Czech Republic since 1992", Austrian journal of Statistics, 133-152.

Bilkova, D. (2013). "Impact of Global Economical Crisis on the development of wages and salaries of men and women in the Czech Republic, comparison of the development of these distributions in the period before the crisis and the time of crisis",.

Gutiérrez, R., Gutiérrez-Sánchez, R., Nafidi, A., & Ramos, E. (2009). "Three-parameter stochastic Lognormal distribution model: statistical computation and simulating annealing–application to real case". Journal of Statistical Computation and Simulation, 79(1), 25-38.

Hosking, J. R. M., & Wallis, J. R. (2005). "Regional frequency analysis: an approach based on L-moments". Cambridge University Press.

Kemal, A. R. (2006)."Income inequalities in Pakistan and a strategy to reduce income inequalities", Background paper for PRSP-II, PRSP secretariat, Government of Pakistan.

Mala, I. (2011). "Distribution of incomes per capita of the Czech households from 2005 to 2008", Journal of Applied Mathematics, 305-310.

## References


Mala, D. B. I. "Modelling income distribution in the Czech Republic with a special focus on income distribution of professionals in informatics", International days of Statistics and Economics, 51-60.

Myung, J. (2002). "Tutorial on Maximum likelihood estimation", journal of Mathematical Psychology 47(2003), 90-100.

Mala, I. (2011). "Conditional distributions of incomes and their characteristics", ISI, 1-6.

Mirza, N. S., and Zahid. A. (2013). "Comparing TL-moments and Conventional moments of Dagum distribution by simulated data", Revista Colombian De Estadistica Junio, 79-93.

Nedved, J., and Mala, I. (2012). "Analysis of incomes in the Czech Republic using the Lognormal distribution", Research Journal of Economics, Business and ICT, 4, 58-63.

Nedved, J. (2011). "The use of the Lognormal distribution in analyzing incomes", international days of Statistics and Economics, 451-463.

Peel, M. C., Wang, Q. J., Vogel, R. M., & McMAHON, T. A. (2001). "The utility of L-moment ratio diagrams for selecting a regional probability distribution," Hydrological Sciences Journal, 46(1), 147-155.

Rehman, H. U., Khan, S., and Ahmed, I. (2008). "Income distribution, growth and financial development across countries analysis," Pakistan Economic and Social Review, 116.

Singh, A. K., Singh, A., & Engelhardt, M. (1997). "The Lognormal distribution in Environmental Applications," Washington, DC: US Environmental Protection Agency, Office of Solid Waste and Emergency Response.

# References

Vogel, R. M., & Fennessey, N. M. (1993). "L-moments diagrams should replace product Moments diagrams," Water Resources Research, 29(6), 1745-1752.

# Appendix

## (1) Generalized Gamma Distribution

The probability density function $f(x;\mu,\sigma,\gamma) = \frac{(x-\varepsilon)^{\alpha-1} e^{\frac{-(x-\varepsilon)}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}$

Where $\alpha = \frac{4}{\gamma^2}$, $\beta = \frac{1}{2}\sigma|\gamma|$ , $\varepsilon = \frac{\mu-2\sigma}{\gamma}$

Where $\mu > 0,\ \gamma >,\ \sigma > 0$

With parameters $\mu$ = location parameter, $\sigma$ = scale parameter, $\gamma$ = shape parameter

The cumulative distribution function $F(x) = 1 - G\left(\alpha, \frac{\varepsilon - x}{\beta}\right)/\Gamma(\alpha)$

### L-moments

Mean $\lambda_1 = \varepsilon + \alpha\beta$

Variance $\lambda_2 = \frac{\beta\Gamma(\alpha+\frac{1}{2})}{\sqrt{\pi}\Gamma(\alpha)}$

Skewness $\tau_3 \approx \frac{1+E_1\alpha+E_2\alpha^2+E_3\alpha^3}{1+F_1\alpha+F_2\alpha^2+F_3\alpha^3}$

Where $E_1 = 2.3807576$, $E_2 = 1.5931792$, $E_3 = 1.1618371\times 10^{-1}$

$F_1 = 5.1533299$, $F_2 = 7.1425260$, $F_3 = 1.9745056$

Kurtosis $\tau_4 \approx \frac{1+G_1\alpha+G_2\alpha^2+G_3\alpha^3}{1+H_1\alpha+H_2\alpha^2+H_3\alpha^3}$

Where $G_1 = 2.1235833$, $G_2 = 4.1670213$, $G_3 = 3.1925299$

## Appendix

$H_1 = 9.0551443,\ H_2 = 2.6649995\times 10^1,\ H_3 = 2.6193668\times 10^1$

### Parameters

Location parameter $\mu = \lambda_1$

Scale parameter $\sigma = \lambda_2 \pi^{\frac{1}{2}} \alpha^{\frac{1}{2}} \Gamma\alpha / \Gamma(\alpha + \frac{1}{2})$

Shape parameter $\gamma = 2\alpha^{\frac{-1}{2}} \tau_3$

## (2) Generalized Pareto Distribution

The probability density function $f(x; \varepsilon, \sigma, k) = \alpha^{-1} e^{-(1-k)y}$

Where $y = -k^{-1} \frac{\log\{1-k(x-\varepsilon)\}}{\alpha}$, $-\infty < \varepsilon < \infty$, $-\infty < k < \infty$, $\alpha > 0$

With parameters $\varepsilon$ = location parameter, $\alpha$ = scale parameter, k = shape parameter

The cumulative distribution function $F(x) = 1 - e^{-y}$

The quantile function $x(F) = \varepsilon + \frac{\alpha\{1-(1-F)^k\}}{k}$

### L-moments

Mean $\lambda_1 = \frac{\varepsilon+\alpha}{1+k}$

Variance $\lambda_2 = \frac{\alpha}{\{(1+k)(2+k)\}}$

Skewness $\tau_3 = \frac{(1-k)}{(3+k)}$

Kurtosis $\tau_4 = \frac{\tau_3(1+5\tau_3)}{5+\tau_3}$

**Parameters**

Location parameter $\varepsilon = \lambda_1 - (2+k)\lambda_2$

Scale parameter $\alpha = (1+k)(2+k)\lambda_2$

Shape parameter $k = \frac{(1-3\tau_3)}{(1+\tau_3)}$

# (3) Lognormal Distribution

The probability density function $f(x;\, \alpha, k, \varepsilon) = \frac{e^{ky-\frac{y^2}{2}}}{\alpha\sqrt{2\pi}}$, where $y = -k^{-1}\log\{1 - \frac{k(x-\varepsilon)}{\alpha}\}$

Where $-\infty < \alpha < \infty$, $0 \le \varepsilon < x$, $k > 0$. With parameters $\alpha$ = scale parameter, $k$ = shape parameter and $\varepsilon$ = location parameter.

**L-moments**

Mean $\lambda_1 = \varepsilon + \alpha(1 - e^{\frac{k^2}{2}})/k$

Variance $\lambda_2 = \frac{\alpha}{k} e^{\frac{k^2}{2}} \left\{1 - 2\varphi\left(-\frac{k}{\sqrt{2}}\right)\right\}$

Skewness $\tau_3 \approx -k \frac{A_0 + A_1k^2 + A_2k^4 + A_3k^6}{1 + B_1k^2 + B_2k^4 + B_3k^6}$

Kurtosis $\tau_4 \approx {\tau_4}^0 + k^2 \frac{C_0 + C_1k^2 + C_2k^4 + C_3k^6}{1 + D_1k^2 + D_2k^4 + D_3k^6}$

Where ${\tau_4}^0 = 1.2260172 \times 10^{-1}$, $A_0 = 4.8860251 \times 10^{-1}$, $A_1 = 4.4493076 \times 10^{-3}$

$A_2 = 8.8027039\times 10^{-4}$, $A_3 = 1.1507084\times 10^{-6}$, $B_1 = 6.4662924\times 10^{-2}$

$B_2 = 3.3090406\times 10^{-3}$, $B_3 = 7.4290680\times 10^{-5}$, $C_0 = 1.8756590\times 10^{-1}$

$C_1 = -2.5352147\times 10^{-3}$, $C_2 = 2.6995102\times 10^{-4}$, $C_3 = -1.8446680\times 10^{-6}$

$D_1 = 8.2325617\times 10^{-2}$, $D_2 = 4.2681448\times 10^{-3}$, $D_3 = 1.1653690\times 10^{-4}$

**Parameters**

Scale parameter $\alpha = \frac{\lambda_2 k e^{\frac{-k^2}{2}}}{1-2\varphi(-\frac{k}{\sqrt{2}})}$

Shape parameter $k \approx -\tau_3 \frac{E_0+E_1\tau_3^2+E_2\tau_3^4+E_3\tau_3^6}{1+F_1\tau_3^2+F_2\tau_3^4+F_3\tau_3^6}$

Where $E_0 = 2.0466534$, $E_1 = -3.6544371$, $E_2 = 1.8396733$, $E_3 = -0.20360244$

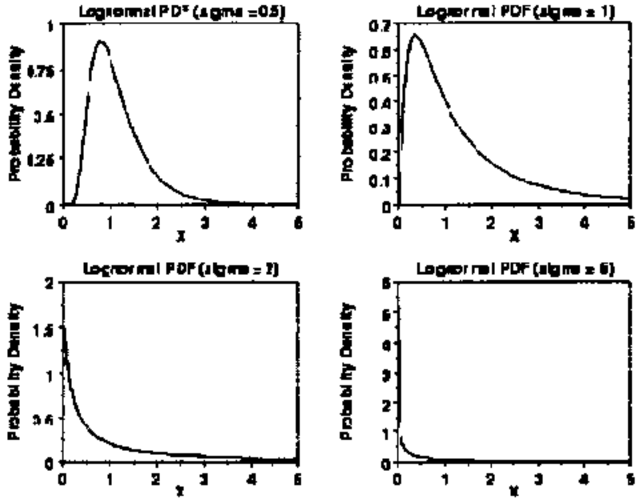
$F_1 = -2.0182173$, $F_2 = 1.2420401$, $F_3 = -0.21741801$

Location parameter $\varepsilon = \lambda_1 - \frac{\alpha}{k}(1 - e^{\frac{k^2}{2}})$

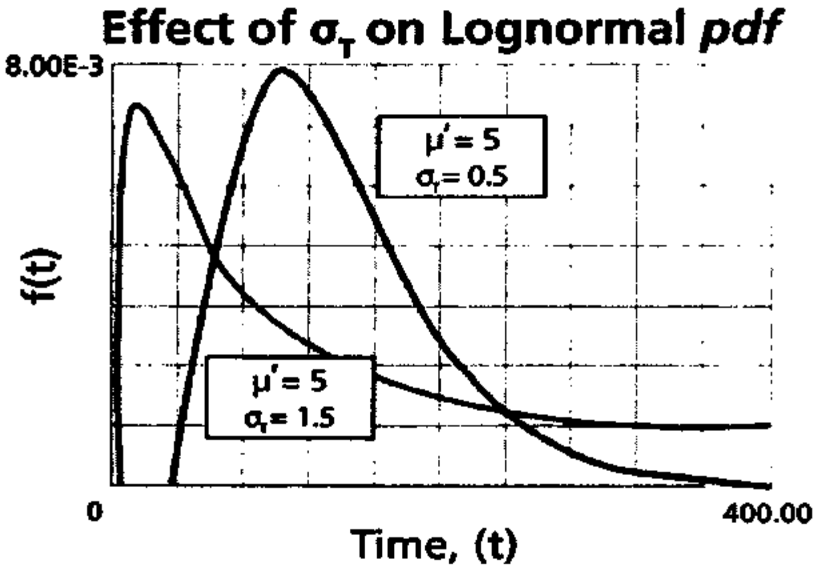
## Characteristics of the Lognormal Distribution

The following is the plot of lognormal probability density function for different values of shape parameter $\sigma$. (From, Engineering Statistics handbook). As the value of shape parameter $\sigma$ increases, the skewnss and the kurtosis of the distribution increases, that is given in figures (A).

(A)



The skewness and kurtosis of the lognormal distribution increases if the value of shape parameter $\sigma$ increases. The skewness will increase if the value of $\sigma$ increases for a given value of $\mu$. As shown in the figures (B). (B)



But when the value of scale parameter $\mu$ increases, then the dispersion of the distribution will be increases, as shown in figure (C).

(C)