The Maximum Likelihood Estimation of a Longitudinal Data of Household Income in the Presence of Outlier Densities

3rd ICAPTA Conference (UNILAG 2016)

A.O. Bello[1], O.M. Adetutu[2], F.A. Oguntolu[3]
Department of Statistics[1,2],
Department of Mathematics[3]
Federal University of Technology,Minna

April 13, 2016

Abstract

This work reports on the use of maximum likelihood function and the probability graphical method to estimate the location parameter for a mini-metropolis household income(X) with heterogeneous social-economic composition.

The Easyfit software was used to fit the household income data to suggest the possible probability distribution(s) for the data. Some of the suggested distributions were taken as the functional form of the income's(X as a r.v) probability distribution and they were empirically solved using the maximum likelihood method of estimation(MLE) in comparison to traditional matching moment estimation(MME).

The estimate that is most consistent with the sample data were solved analytically based on the distribution function(s) suggested by easyfit software. We also compared the maximum likelihood estimates of obtained from each distribution functions graphically using R-programming language codes

Introduction

To accurately fit the probability distribution for a household income is vital to measuring location parameter for household income potential of a given geographical region.

This is important for the purpose of economic or developmental planning. It is reasonable that different distributions will be found for different household income in different geographical area.

Household income is often used as an economic indicator and the accurate fit of distribution will ensure accurate estimation of parameters which will aid correct economic decision making on such geographical region of which per capital income is of interest.

Household income refers not only to the salaries and benefits received but also to receipts from any personal business, investments, dividends and other income(Investopedia, 2015).

Fitting A Continuous Probability Distribution for Data

Vito Ricci (2005) and Zaven A. Karian and Dudewicz,(2011) unanimously suggests that in the search for a method of fitting a continuous probability distribution to data there is need to;

- first decide on what type (or family) of distributions to consider, also
- decide on what fitting method to use either method of moments, maximum likelihood, least squares etc. and conclusively
- develop a specific computational schemes to be invoked to estimate the parameters associated with the fit.

1

[1](See R, Zaven A. Karian Edward J. Dudewicz(2011) and Vito Ricci 2005, for details)

The Log-normal Distribution

Log-normal with parameters μ and δ both non-negative $0 < x < \infty$ which makes it more preferable to $-\infty < x < \infty$ of the popular normal distribution.

$$f(x) = \begin{cases} \frac{1}{x\delta\sqrt{2\Pi}} \exp\left[-\frac{(\ln(x)-\mu)^2}{2\delta^2}\right] & \text{if x > 0} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

The moments of the log-normal distribution can be obtained using analogic, moments or the moment matching estimation (MME) and maximum likelihood estimation (MLE).

$$\left.\begin{array}{l} \alpha_1 = e^{\mu+\delta^2/2} \\ \alpha_2 = (e^{\delta^2} - 1)e^{2\mu+\delta^2} \\ \alpha_3 = \sqrt{e^{4\delta^2} - 1}(e^{\delta^2} + 2) \end{array}\right\} \quad (2)$$

supporting $x \in (0, +\infty)$ with entropy $\log(\delta e^{\mu+\frac{1}{2}}\sqrt{2\Pi})$ and fisher information $\begin{pmatrix} 1/\delta^2 & 0 \\ 0 & 2/\delta^2 \end{pmatrix}$.

Log-normal Properties and Cumulative Distribution

The mode being a point of global maximum of a density function is obtainable by$mode[X] = e^{\mu - \delta^2}$; and median; $med[X] = e^{\mu}$. The log-normal parameters can be obtained if arithmetic mean and the arithmetic variance are known and simple to estimate by estimator;

$$\mu = \ln(E|X|) - \frac{1}{2}\ln(1 + \frac{Var[X]}{(E[X])^2}) = \ln(E[X]) - \frac{1}{2}\delta^2 \tag{3}$$

and

$$\delta^2 = \ln(1 + \frac{Var[X]}{(E[X])^2}) \tag{4}$$

The location and scale parameters of a log-normal distribution are better approach using geometric mean and geometric standard deviation respectively than the arithmetic approach.

$$F(x) = \frac{1}{\delta\sqrt{2\prod}}\int_0^x \frac{1}{t} e^{-\frac{1}{2}(\frac{\ln t - \mu}{\delta})^2 dt} = \frac{1}{\delta\sqrt{2\prod}}\int_0^{\ln x} e^{-\frac{1}{2}(\frac{x-\mu}{\delta})^2 dx} \tag{5}$$

Literature Suggestions of Suitable Distribution Function

Knowing that normal distribution may not be suitable for variable that are inherently non-negative such as we have in household income which is likely to be strongly skewed due to possible presence of outliers;

Log-normal is vital in description of natural phenomena. According to Ritzema et al (1994) log-normal distribution was employed in hydrology research to analyse extreme values of variables such as monthly, annual maximum value of daily rainfall and river discharge.

In social-economic research, there were conclusions made about over 97% of the populations income considered to follow a log-normal distribution (see Clement et al, 2005).

The distribution of higher-income individual follows a Pareto distribution (Wataru, 2002). This is similar to the presence of some extreme income in our case of consideration.

Pareto Distribution

The power law distribution is also named *Pareto distribution* after the Italian economist Vilfredo Pareto (1848–1923) who originally observed it studying the allocation of wealth among individuals: a larger share of wealth of any society (approximately 80%) is owned by a smaller fraction (about 20%) of the people in the society.

$$f(x) = \begin{cases} \frac{\beta\lambda^{\beta}}{x^{\beta+1}}, & \text{if x} > \lambda \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The Pareto distribution supports$x \in (\lambda + \infty)$; $\lambda>0$ and $\beta>0$. The Pareto distribution are characterized by cumulative density function; $1 - (\frac{\lambda_m}{x})^{\alpha}$ $\text{for x} \geq \lambda_{\text{m}}$. The praetor root of equation is unstable which allows the shape and scale parameters to be of two roots.

$$\mu = \begin{cases} \infty & \text{for } \beta \leq 1 \\ \frac{\beta\lambda_m}{\beta-1} & \text{for } \beta > 1 \end{cases} \tag{7}$$

and

$$\delta^2 = \begin{cases} \infty & \text{for } \beta \in \\ \frac{\lambda_m^2\beta}{(\beta-1)(\beta-2)} & \text{for } \beta > 2 \end{cases} \tag{8}$$

The mode λ_m, median and fisher information ; $\begin{pmatrix} \frac{\alpha}{x_m^2} & -\frac{1}{x_m} \\ \frac{1}{-x_m} & \frac{1}{\delta^2} \end{pmatrix}$

Extreme Value Distribution

This distribution is also called the smallest extreme value distribution with example of application of Weibull(1951) reported on the strength of a certain material that follow an extreme value distribution (See Christian Walck(2007), p 41).

The extreme value distribution is also used in life and failure data analysis, "weakest link" situations,temperature minima, rainfall in droughts, human mortality of the aged.
The Largest Extreme Value Distribution will be more appropriate with its parameters μ and $\delta > 0$ having the pdf;

$$f(x) = \frac{1}{\delta} \exp \pm^{[x-\mu]}/_{\delta} \;\; \exp \pm[-e\frac{-(x-\mu)}{\delta}] \tag{9}$$

The cumulative distribution of extreme value distribution

$F(x) = \int_{-\infty}^{x} f(u)du = \int_{-\infty}^{\pm\frac{x-\mu}{\delta}} \quad g(z)dz = G(\pm\frac{x-\mu}{\delta})$ where G(z) is the cumulative function of g(z) which is given by $G(z) = \int_{-\infty}^{z} e^{-\mu - e^{-\mu}} du = \int_{e^{-z}}^{\infty} e^{-y} dy = e^{-e^{-z}}$ where we have made the substitution y = e-u in simplifying the integral

Method of Estimation

Parameters of distribution function can be obtained using analogic moments or the moment matching estimation (MME),quantile matching estimation (QME), maximum goodness-of-fit estimation (MGE)and Maximum likelihood estimation(MLE)but MLE and moment matching estimations are certainly the most commonly used method for fitting distributions (Cullen and Frey, 1999).

- The Moment Method

Algebraic moments of order r are defined as the expectation value

$$\alpha_r^{'} = E(x^r) = \int_{-\infty}^{+\infty} x^r f(x)dx \tag{10}$$

For $\alpha_0^{'}$ =1 and $\alpha_1^{'}$ is the mean value of the distribution, $\alpha_2$as the variance of the model with α_3 and α_4 for the kurtosis and skewness respectively. Central moments of order r are defined as$\alpha_r = E((K - E(k))^r)$. The commonly used is μr is regarded as the unknown population mean.

- Error of Moment Method

Consider a sample with n observations $x_1, x_2, ..., x_n$ we define the moment-statistics for the algebraic and central moments $\alpha_r^{'}$ and α_r as $\alpha_r^{'} = 1/n \sum_{r=0}^{n} x^r$ and $\alpha_r^{'} = 1/n \sum_{r=0}^{n} (x^r - m_1^r)^r$. The covariance between an algebraic and a central moment is

$$Cov(\alpha_q, \alpha_r) = 1/n(\mu_{q+r} - \mu_q\mu_r + rq\mu_2\mu_{r-1}\mu_{q-1} - r\mu_{r-1}\mu_{q+1} - q\mu_{r+1}\mu_{q-1}) \tag{11}$$

Method of Maximum Likelihood

Having notable extreme values in any real life data can be worrisome, outliers raised the possible question of sudden system error (non random error), or observational error.

A case where it is not any of the mentioned, it becomes a problem to present a reliable estimate(s) for further estimation in alternative processes and presenting point estimates for predictions, budgeting and planning purposes.
This estimation obtained the value of location paremeter that maximize $log_e L(\theta)$, that is, the value of $\hat{\theta}$ that assign the highest possible probability to the data that was obtained.

- **Log-normal Distribution**

$$f(x) = \frac{1}{\sqrt{2\prod}\delta x} \exp - \frac{[\log(x) - \mu)^2]}{2\delta^2} \tag{12}$$

The likelihood function for n-observation household income$x_1, x_2, ..., x_n$, with parameters μ and δ^2

$$L(\mu, \delta^2) = \frac{1}{\sqrt{2\prod}\delta x_1} \exp - \frac{[\log(x_1) - \mu)^2]}{2\delta^2} + \frac{1}{\sqrt{2\prod}\delta x_2} \exp - \frac{[\log(x_2) - \mu)^2]}{2\delta^2} + \ldots + \frac{1}{\sqrt{2\prod}\delta x_n} \exp - \frac{[\log(x_n) - \mu)^2]}{2\delta^2}$$

$$L(\mu, \delta^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\prod}\delta x_i} \exp - \frac{[\log(x_i) - \mu)^2]}{2\delta^2} \tag{13}$$

$$L(\mu, \delta^2) = \frac{1}{(\sqrt{2\prod}\delta^2 x_i)^{\frac{n}{2}}} \exp - \frac{1}{2\delta^2} \sum_{i=1}^{n} [\log(x_i) - \mu)^2] \tag{14}$$

Maximum Likelihood Estimation

The value of μ and δ that maximize the likelihood function are maximum likelihood estimator (MLE) denoted by$\hat{\mu}$ and $\hat{\delta}$. The parameters were obtained by partial derivation of the log-likelihood function wrt toμ and δ

$$\log L(\mu,\ \delta) = \frac{-n}{2}\log_e 2\prod - \frac{n}{2}\log_e \delta^2 - \frac{1}{2\delta^2}\sum_{i=1}^{n}[\ln(x) - \mu] \tag{15}$$

$$\partial \log L(\mu,\ \delta)/_{\partial\mu} = \frac{1}{2\delta^2}\sum_{i=1}^{n}[\ln(x) - \mu]$$

$$\partial \log L(\mu,\ \delta)/_{\partial\mu} = 0$$

$$\tfrac{1}{2\delta^2}\textstyle\sum_{i=1}^{n}[\ln(x) - \mu] = 0$$
$$\textstyle\sum_{i=1}^{n}(\log(x) - \mu) = 0$$
$$\log\sum x = n\mu$$

$$\hat{\mu} = {}^{\log\sum x}\Big/_{n} \tag{16}$$

Partial differentiation of equation (15) wrt to δ^2yielded $\delta^2 = \sum_{i=1}^{n}(\log(x_i) - \mu)$

Maximum Likelihood Estimation

- **Pareto Distribution**

$$\mathrm{f(x)} = \frac{\beta\lambda_{\mathrm{m}}^{\beta}}{\mathrm{x_i^{\beta+1}}} \tag{17}$$

The likelihood functions for n-observation household income$x_1, x_2, ..., x_n$, considering Pareto pdf;

$$\mathrm{L}(\beta,\ \lambda) = \frac{\beta\lambda_{\mathrm{m}_1}^{\beta}}{\mathrm{x}_1^{\beta+1}} + \frac{\beta\lambda_{\mathrm{m}_2}^{\beta}}{\mathrm{x}_2^{\beta+1}} + \ldots + \frac{\beta\lambda_{\mathrm{m}_\mathrm{n}}^{\beta}}{\mathrm{x}_\mathrm{n}^{\beta+1}} \tag{18}$$

$$\mathrm{L}(\beta,\ \lambda) = \beta^{\mathrm{n}}\lambda_{\mathrm{m}}^{\mathrm{n}\beta}\prod_{\mathrm{i=1}}^{\mathrm{n}}\frac{1}{\mathrm{x_i^{\beta+1}}}$$

$$\log_e L(\beta,\lambda) = n\ln\beta + n\beta\lambda_m - (\beta+1)\sum_{i=1}^{n}\ln x_i \tag{19}$$

The partial derivation the Pareto logarithm likelihood function wrt to each parameters yielded

$\hat{\beta} = \frac{n}{\sum_i(\ln x_i - \ln\lambda_m)}$ with $\mathrm{s.e}(\hat{\beta}) = \hat{\beta}/\mathrm{n}$

Maximum Likelihood Estimation

- Extreme Value Distribution

$$L(\mu,\delta) = \frac{1}{\delta} e^{(x_1-\mu)\delta} \exp[-e^{(x_1-\mu)\delta}] + \frac{1}{\delta} e^{(x_2-\mu)\delta} \exp[-e^{(x_2-\mu)\delta}] + \ldots + \frac{1}{\delta} e^{(x_n-\mu)\delta} \exp[-e^{(x_n-\mu)\delta}] \tag{20}$$

$$L(\mu,\delta) = \prod_{i=1}^{n} \frac{1}{\delta} e^{(x_i-\mu)} \quad \exp\left[-e^{(x_i-\mu)\delta}\right]$$

$$L(\mu,\delta) = \tfrac{n}{\delta} e^{\sum(x-\mu)\delta} \exp[-e^{\sum(x-\mu)\delta}]$$

$$\log_e L(\mu,\delta) = n \log\delta + \sum(x-\mu)\delta - e^{\sum(x-\mu)\delta} \tag{21}$$

The parameters of the extreme distribution of any Types can be obtained by partial differential of $\log_e L(\mu,\delta)$ wrt to the parameter of interest.

The root of equation of this distribution is of multiple origin.

Graphical Evaluation

- **Theoretical Distribution Function vs. Empirical Distribution Model.**

The standard theoretical distribution function (t.d.f) shapes will be compared with empirical distribution function (e.d.f) fitted. It will be ideal to check how well the main body and tails of the e.d.f visually matched the t.d.f.? The histogram, density plots and Quantile-Quantile (Q-Q) plot shall be used for this comparisons.

The graphical technique is for determining if a data set comes from a known population distributions.

- **Goodness of Fit(GOF)**

Knowing how good our fit of the pdf to the data required a pragmatic approach. If only descriptive statistical information is available, then a simple visual comparison will be sufficient if the data is free of many extreme values. A further check after the parameter estimation process will be more reasonable.

The goodness of fit tests relevant to determining the most suitable statistical distribution for the data are Kolmogorov-Smirnov (KS), Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). The model with the minimum AIC value is selected as the best fit model. The AIC usually be used.

Computational Procedure

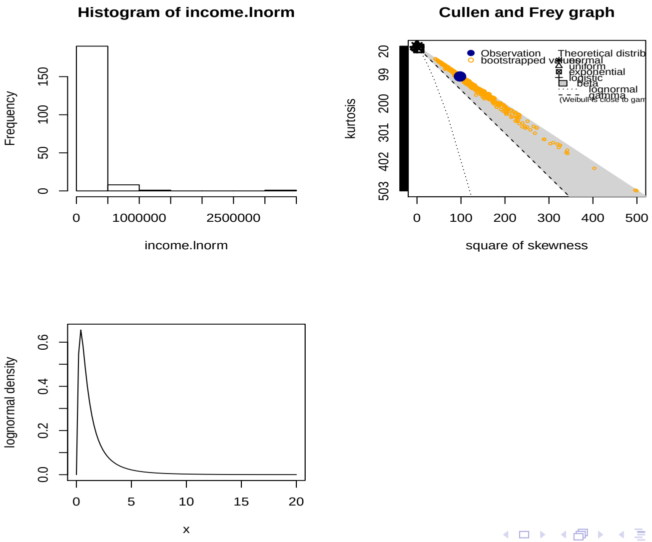
Fitting the most adequate distribution for the household income data begun with the exploratory data analysis (EDA)using histogram and skewness-kurtosis plot provided by the descdist function in R package which suggested the kind of pdf that suits the data, this was also compared with easyfit software suggested pdfs. The plots are to show presence or absent of skewness that is the behaviour in the tails and outliers in a data and skewness.

The probability curve on histograms can be graphically compared to the standard shapes associated to the suspecting probability distributions. A non-zero skewness reveals a lack of symmetry of the empirical distribution, while the kurtosis value quantities the weight of tails in comparison to the normal distribution for which the kurtosis equals 3.

Computational schemes to estimates parameters and graphically evaluate the three major distributions suspected was formulated in R-packages: library("stats4""fitdistrplus","descdist","MASS","fBasics")

Graphical Distribution Suggestions

Fig 1



Results and Discussion

Location Parameter Estimates ' lnorm '	matching moments Method	maximum likelihood Method
Mean	11.328105	11.44235452
Std Deviation	1.996814	0.9834032
Kolmogorov-Smirnov statistic	0.3300693	0.2327564
Cramer-von Mises statistic	15.6761084	4.5059558
Anderson-Darling statistic	81.9090422	29.5680999
Aikake's Information Criterion	13179.75	12848.56
Bayesian Information Criterion	13188.18	12856.99

Table 1

100 Samples			200 Samples		500 samples	
Distribution	Log-normal	SnPareto	Log-normal	SnPareto	Log-normal	SnPareto
$L(\theta)$	0.0004186823	5.9932e-05	0.0008037468	0.0001100977	0.0008037468	0.000260585
AIC	12178.70	16171.75	13179.75	17179.75	14179.75	23179.00

Table 2

Location Parameter	logmean	Logmode	logmedian	TrimmedMean
$f(x_1)$	2.375297e-06	6.667624e-09	2.441384e-06	2.442240e-06
$f(x_2)$	6.775384e-06	6.523536e-07	6.488569e-06	6.484578e-06
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
$f(x_n)$	6.509270e-06	4.435462e-07	6.276967e-06	6.273698e-06
Total($L(\theta)$)	**0.0007769066**	**0.0003445284**	**0.0007566447**	**0.0007563607**

Table 3

Graphical Evaluation of Pareto Distribution

* Plot of varying λ-the Shape parameter

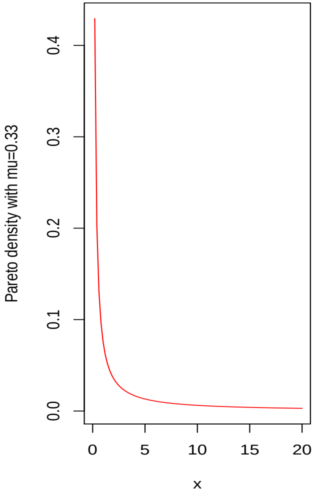


Fig 2

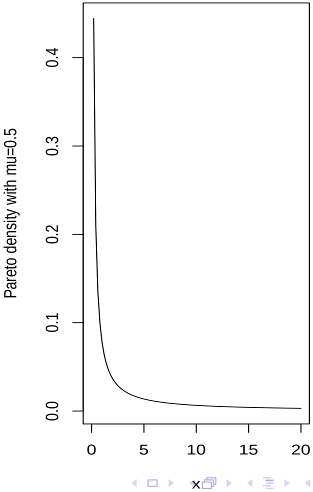


Fig 3

Plot of varying β-Parameter of Pareto Distribution)

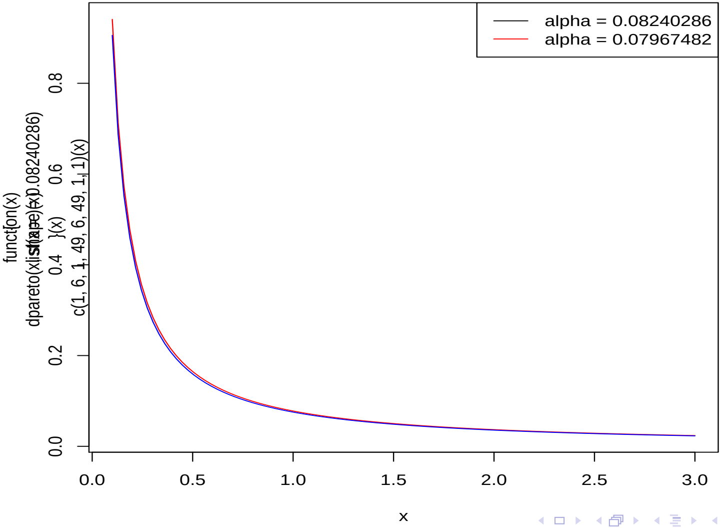


Fig 4

Graphical Evaluation

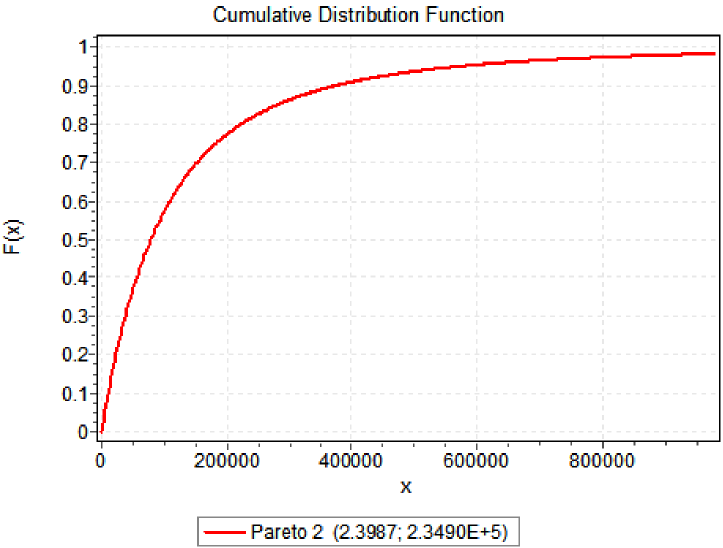
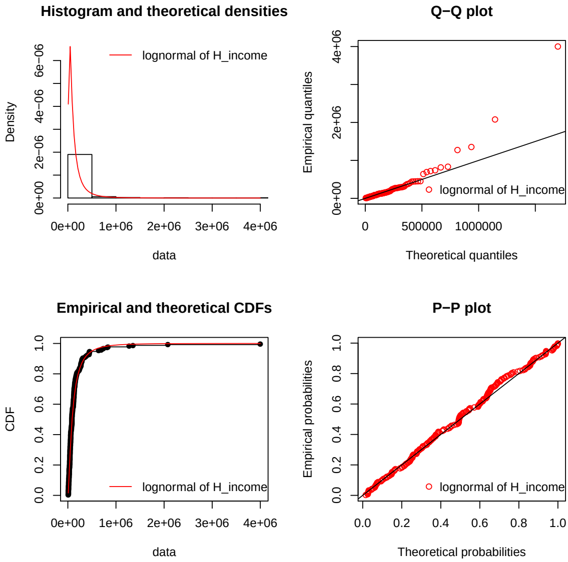


Figure: Fig 5:Pareto Fitted Parameter's CDF Plot

Graphical Evaluation of Log-normal Function Fit



[2]

[2] Fig 6

Graphical Evaluation of Matching Moment & Maximum Likelihood Estimations

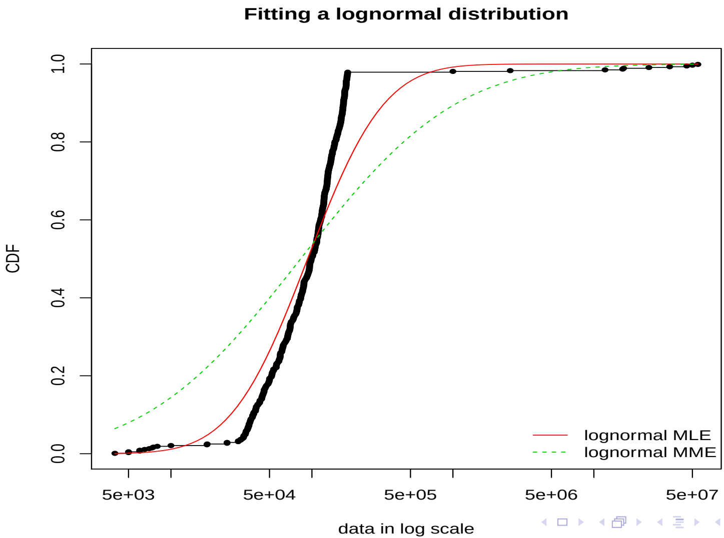


Fig 7

Graphical Evaluation Of Distribution Fits on Data

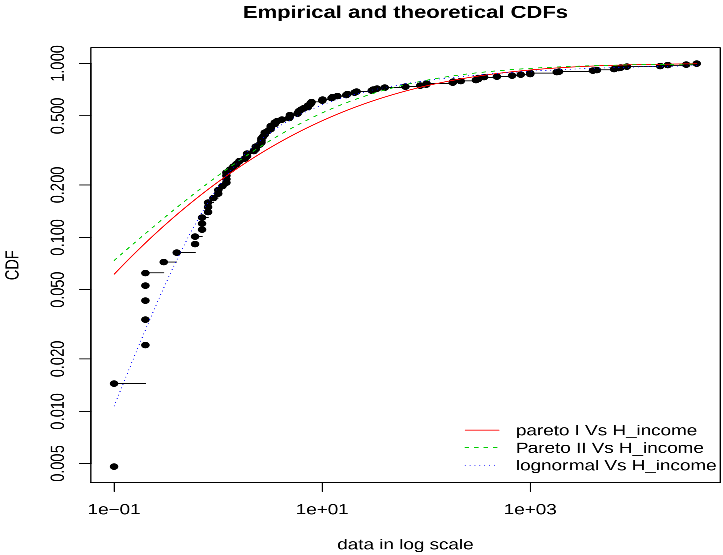


Fig 8

Conclusions

Cullen d Frey graph provided by the descdist function in R and Easyfit software both agree on significant fit of normal, log-normal, weibull, Pareto, gamma and beta distributions.
However, the underlying theoretical properties of lognormal, Pareto and Extreme value distribution relating to the household income data at hand guided our choice of their use.

From the Table 1; the matching moment method (MME) and maximum likelihood method (MLE) produced 11.328105 and 11.44235 logarithm arithmetic mean with standard deviation of 1.996814 and 0.9834032 respectively. Aikakes information criterion values put the MLE at 12848.56 less than 13188.18 for the MME.

The Method of maximum likelihood was used to estimate the parameters for the three distribution under consideration and the parameters for each model were used to randomly generated 100 samples, 200 samples and 500 samples each for the respective distributions. The likelihood value $L(\theta)$ of the parameters for each distribution was estimated and presented in table 2 above,showing that $l(\theta)$ value of log-normal distribution is larger at 100 samples, 200 samples and 500 samples implying that the log-normal function is much consistent with the sample data.

The root of partially differentiated log-likelihood function of Pareto distribution with respect to the shape and location parameters yielded (0.33,0.5) and (0.082402,0.07967)respectively. The plots of the fitted distribution with the two different roots show little deviation(see fig 2,3 &4).

The extreme value distribution breaks down at the verge of estimation.
Fig 6 shows the points of the P-p plot falling approximately along the reference line,suggestion compatibility of the log-normal distribution to the empirical income data. The Q-Q plot does not show any greater departure from the reference line(no lack-of-fit), only for some pockets of extreme data point that could not be well capture by the distribution function.

The empirical and theoretical cumulative density function show the adequacy of the log-normal distribution to the data fit.Also confirming log-normal to be more suitable in comparison to the Pareto is graphical evaluation of distribution fit to data, the dotted blue line captured the movement of the data black-pointed well compared to the red and dotted green line representing pareto of type I and II.

References

[1] Christian Walck(2007).*Hand-book on Statistical Distributions for experimentalists*.Particle Physics Group Fysikum, University of Stockholm; (e-mail: *walck@physto.se*)last modification 10 September 2007

[2] Clement, Fabio, Gallegati, Mauro(2005). *Pareto's Law of Income Distribution for Germans, the UK and US*. EconWPA

[3] Cullen, A. and Frey, H. (1999). *Probabilistic Techniques in Exposure Assessment*. Plenum Publishing Co., 1st edition.

[4] Investopedia(2015). Household Income Definition.$http://www.investopedia.com/terms/h/household_income.asp\#ixzz3z6m4srLw$

[5] Marie Laure Delignette-Muller and Christophe Dutang (2014); *fitdistrplus: An R Package for Fitting Distributions.* Universite de Lyon and Universite de Strasbourg ,October 2014

[6] Ritzema (E.D),H.P(1994). Frequency and Regression Analysis Publication 16. International Institute for land Reclamation and improvement(ILRI), the Netherlands pp 175-224 ISBN 90-70754-33-9

[7] Vito Ricci (2005).*Fitting Distribution With R*; $vito_ricci@yahoo.com$

[8] Wataru Souma(2002).*Physics of Personal income*. airxiv.ond.mat/0202388

[9] Zaven A. Karian ,Edward J. Dudewicz(2011). *Handbook of Fitting Statistical Distributions* ISBN 978-1-58488-711-9 Chapman & Hall/CRC Taylor & Francis Group 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742

THANK YOU ALL.

Thank you.........