A Note on Empirical Sample Distribution of Journal Impact Factors in Major Discipline Groups

SK Mishra
Dept. of Economics
North-Eastern Hill University
Shillong, Meghalaya (India)
Contact: mishrasknehu@yahoo.com

1. Introduction: Mansilla et al. (2007) observed that journal impact factors (JIF), irrespective of the discipline, exhibit their adherence to a specified rank-size rule. Egghe (2009) made an attempt to give a theoretical explanation for the JIF rank-order distributions observed by Mansilla et al. (2007). Waltman and Eck (2009), while concluding that Egghe's analysis relies on the unrealistic assumption that the articles published in a journal can be regarded as a random sample from the population of all articles published in a field (and since Waltman-Eck's observations deny the agreement of Egghe's analysis with empirical data and hence Egghe could not give a satisfactory explanation for JIF rank-order distributions), observed:

> "*Egghe interprets the [J]IF of a journal as the average of a number of independent and identically distributed random variables. Each random variable represents the number of citations of one of the articles published in the journal. Using the central limit theorem, Egghe's interpretation implies that the [J]IF of a journal is a random variable that is (approximately) normally distributed. Egghe also makes the assumption that for a given scientific field each journal in this field can be considered as a random sample in the total population of all articles in the field. This assumption has the implication that the [J]IFs of all journals in a field follow the same normal distribution*."

Mishra (2009) found that the empirical $\log_{10}$(JIF) distributions in different major discipline groups (such as biology, chemistry, engineering and physics for the year 2006 and psychology and social sciences for the year 2002) are Pearsons's type-IV. Thus, the empirical evidences support the criticism of Egghe's arguments made by Waltman and Eck (2009) and as a consequence one cannot assert that the distributions of JIFs across the discipline groups could be more or less identical or normal. As a matter of fact, the empirical distributions of $\log_{10}$(JIF) are asymmetric, non-mesokurtic and often with too thick or too short tails.

The objectives of this paper are to search for the best fit statistical distributions to the JIF data for various major discipline groups such as biology, chemistry, engineering and physics for the year 2006 (source: http://www.icast.org.in/Impact/subject2006.html), psychology and social sciences for the year 2002 (source: http://www.staff.city.ac.uk/~sj361/here_you_can_see_an_excel_spread.htm) and economics (for 2009; source: http://ideas.repec.org/top/top.journals.simple.html), and point out whether these empirical distributions have some notable general features. To each discipline group data (in its natural form as well as common logarithmic transformation) a number of statistical distributions have been fitted and their fitness is judged on three test statistics. The best-fit distributions have been reported in case of each discipline group.

Suggested Citation: Mishra, S. K. (2010) "A Note on Empirical Sample Distribution of Journal Impact Factors in Major Discipline Groups", (February 14, 2010). Available at SSRN: http://ssrn.com/abstract=1552723

2. Description of Statistical Distributions found to fit Best to the JIF/Log(JIF) Data: We fitted numerous distributions to the JIF data (and their $\log_{10}$ transforms) to find the three best-fits in each case. Overall, some of the theoretical distributions were found most suitable among them.

2.1. The List of Theoretical Distributions that Best fit the Data: We describe now the theoretical distributions that fitted the JIF and Log_{10}(JIF) data most. It may be noted that many other theoretical distributions were fitted to the data, but they were not fitting as best as the ones described below.

i. Beta Distribution: With the support random variable $x: a \le x \le b$ where $(a < b)$, the probability density function (pdf) of Beta distribution is given as:

$$f(x) = \frac{1}{B(\alpha_1, \alpha_2)} \frac{(x-a)^{\alpha_1 - 1}(b-x)^{\alpha_2 - 1}}{(b-a)^{\alpha_1 + \alpha_2 - 1}}$$

α_1 - continuous shape parameter ($\alpha_1 > 0$)
α_2 - continuous shape parameter ($\alpha_2 > 0$)
a, b - continuous boundary parameters
B is the Beta Function

ii. Burr-XII Distribution: It is also known as 4-parameter generalized Beta-II distribution with unit shape parameter, Singh-Maddala distribution (Singh and Maddala, 1976) as well as the Pareto-IV distribution (Kleiber and Kotz, 2003). With the support random variable $x: \gamma \le x < +\infty$, the probability density function (pdf) of Burr 4-parameters (4p) distribution is given as:

$$f(x) = \frac{\alpha k \left(\frac{x-\gamma}{\beta}\right)^{\alpha - 1}}{\beta \left(1 + \left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right)^{k+1}}, \text{ where}$$

$k, \alpha > 0$ are the two shape parameters
$\beta > 0$ is the scale parmeter
γ is the location parameter
If γ=0, then the distribution is 3p

iii. Dagum (Inverse Burr-III) Distribution: With the support random variable $x: \gamma \le x < +\infty$, the probability density function (pdf) of Dagum 4-parameters (4p) distribution is given as:

$$f(x) = \frac{\alpha k \left(\frac{x-\gamma}{\beta}\right)^{\alpha k - 1}}{\beta \left(1 + \left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right)^{k+1}}, \text{ where}$$

$k, \alpha > 0$ are the two shape parameters
$\beta > 0$ is the scale parmeter
γ is the location parameter
If γ=0, then the distribution is 3p

iv. Generalized Extreme Value (GEV) Distribution: With the support random variable $x: -\infty < x < +\infty$ for $k = 0$ and $1 + k(x-\mu)/\sigma > 0$ otherwise, the probability density function (pdf) of GEV distribution is given as:

$$f(x) = \begin{cases} \frac{1}{\sigma} \exp\left(-(1+kz)^{-1/k}\right)(1+kz)^{-1-1/k} & k \ne 0 \\ \frac{1}{\sigma} \exp(-z - \exp(-z)) & k = 0 \end{cases}, \text{ where}$$

k is the shape parameter
σ>0 is the scale parameter
μ is the location parameter
$z = (x - \mu)/\sigma$

v. Generalized Gamma Distribution: With the support random variable $x: \gamma \le x < +\infty$, the probability density function (pdf) of Generalized Gamma 4-parmeters (4p) distribution is given as:

$$f(x) = \frac{k(x-\gamma)^{k\alpha-1}}{\beta^{k\alpha}\Gamma(\alpha)} \exp(-((x-\gamma)/\beta)^k)$$

If γ=0, the distribution is 3p

, where

$\alpha, k > 0$ are the two shape parameters

β>0 is the scale parameter

γ is the location parameter

vi. Generalized Normal (or Error) Distribution: With the support random variable $x: -\infty < x < +\infty$, the probability density function (pdf) of Generalized Normal distribution (also called the error distribution) is given as:

$$f(x) = c_1\sigma^{-1}\exp(-|c_0 z|^k)$$

$$z = \frac{x-\mu}{\sigma}$$

k - continuous shape parameter

σ - continuous scale parameter ($\sigma > 0$)

μ - continuous location parameter

vii. Hyperbolic Secant Distribution: With the support random variable $x: -\infty < x < +\infty$, the probability density function (pdf) of Hyper-Secant distribution is given as:

$$f(x) = \frac{\operatorname{sech}\left(\frac{\pi(x-\mu)}{2\sigma}\right)}{2\sigma}$$

σ - continuous scale parameter ($\sigma > 0$)

μ - continuous location parameter

$\pi = 4\arctan(1)$

viii. Inverse Gaussian Distribution: With the support random variable $x: 0 < x < +\infty$, the probability density function (pdf) of Inverse Gaussian 3-parmeters (3p) distribution is given as:

$$f(x) = \sqrt{\frac{\lambda}{2\pi(x-\gamma)^3}} \exp\left(-\frac{\lambda(x-\gamma-\mu)^2}{2\mu^2(x-\gamma)}\right), \text{ where}$$

If γ=0, it gives 2p Inverse Gaussian Distribution

$\lambda > 0$ is the continuous parameter

μ>0 is the continuous parameter

γ is the continuous location parameter

ix. Johnson SB Distribution: With the support random variable $x: \zeta \le x \le \zeta + \lambda$, the probability density function (pdf) of Johnson SB distribution is given as:

$$f(x) = \frac{\delta}{\lambda\sqrt{2\pi}\, z(1-z)} \exp\left(-\frac{1}{2}\left(\gamma + \delta \ln\left(\frac{z}{1-z}\right)\right)\right), \text{ where}$$

λ is the location parameter and $z = (x - \zeta)/\lambda$

γ is the shape parameter

δ>0 is another shape parameter

$\lambda > 0$ is the scale parmeter

x. Johnson SU Distribution: With the support random variable $x: -\infty < x < +\infty$, the probability density function (pdf) of Johnson SU distribution is given as:

$$f(x) = \frac{\delta}{\lambda\sqrt{2\pi}\sqrt{z^2+1}} \exp\left(-\frac{1}{2}\left(\gamma + \delta \ln\left(z + \sqrt{z^2+1}\right)\right)^2\right), \text{where}$$

ζ is the location parameter

$z = (x - \zeta)/\lambda$

γ is the shape parameter

δ>0 is another shape parameter

$\lambda > 0$ is the scale parmeter

xi. Kumaraswamy Distribution: With the support random variable $x: a \le x \le b$, the probability density function (pdf) of Kumaraswamy distribution is given as:

$$f(x) = \frac{\alpha_1 \alpha_2 z^{\alpha_1 - 1}(1 - z^{\alpha_1})^{\alpha_2 - 1}}{(b - a)}, \text{where}$$

$z = (x - a)/(b - a)$

$\alpha_1, \alpha_2 > 0$ are the two shape parameters

a,b: a<b are the boundary parameters

xii. Log-Logistic Distribution: With the support random variable $x: \gamma \le x < +\infty$, the probability density function (pdf) of log-logistic distribution is given as:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1}\left(1 + \left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right)^{-2}, \text{where}$$

$\gamma = 0$ gives 2p log-logistic distribution.

$\alpha > 0$ is the shape parameter

β>0 is the scale parameter

γ is the location parameter

xiii. Log-Normal Distribution: With the support random variable $x: \gamma < x < +\infty$, the probability density function (pdf) of log-normal distribution is given as:

$$f(x) = \frac{\exp\left(-\frac{1}{2}\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)^2\right)}{(x-\gamma)\,\sigma\sqrt{2\pi}}, \text{where}$$

$\gamma = 0$ yields the 2p Lognormal distribution

σ - continuous parameter ($\sigma > 0$)

μ - continuous parameter

γ - continuous location parameter

xiv. Log-Pearson III (LP3) Distribution: With the support random variable $x: 0 < x \le e^{\gamma}$ for $\beta < 0$ and $e^{\gamma} \le x < +\infty$ for $\beta > 0$, the probability density function (pdf) of LP3 distribution is given as:

$$f(x) = \frac{1}{x\,|\beta|\,\Gamma(\alpha)}\left(\frac{\ln(x)-\gamma}{\beta}\right)^{\alpha-1} \exp\left(-\frac{\ln(x)-\gamma}{\beta}\right), \text{where}$$

$\alpha > 0$, $\beta \ne 0$ and γ are the parameters

xv. Weibull Distribution: With the support random variable $x: \gamma \le x < +\infty$, the probability density function (pdf) of Weibull 3-parmeters (3p) distribution is given as:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right), \text{where}$$

If γ=0, the distribution is 2p

$\alpha > 0$ is the shape parameter

$\beta > 0$ is the scale parmeter

γ is the location parameter

2.2. Goodness of Fit Tests: The goodness of fit tests measure the compatibility of a random sample with a theoretical probability distribution function, showing how well the chosen distribution fits the data being analyzed. The general procedure consists of defining a test statistic which is some function of the data measuring the distance between the hypothetical and empirical values, and then calculating the probability of obtaining data which have a still larger value of this test statistic than the value observed, assuming the hypothesis is true. This probability is called the confidence level. Small probabilities indicate a poor fit while higher probabilities indicate a better fit. In this study we have applied three tests. In these tests, the null hypothesis is that the data follow the specific distribution. The alternative hypothesis is that the data do not follow the hypothesized distribution. The null hypothesis is rejected at the chosen significance level (α) if the test statistic, D, is greater than the critical value obtained from the table compiled for a particular test obtainable from published sources (Chakravarti et al, 1967; Stephens, 1974, 1976, 1977-a, 1977-b, 1979). Chi Squared tables are available in almost all statistics books that deal with the testing of statistical hypothesis.

(i) The Kolmogorov-Smirnov Test: This test is based on the largest vertical difference between $F(x)$, the theoretical distribution function, and $F_n(x) = (\text{no.of observations} \le x)/n$. The KS statistic is defined as: $D_n = KS = \sup_x |(F_n(x) - F(x)|$. The K-S test is distribution free (in the sense that the critical values do not depend on the specific distribution being tested).

(ii) **Anderson-Darling Test**: This test gives more weight to the tails of the distribution than the Kolmogorov-Smirnov test does. It has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated for each distribution. It is based on the comparison between the observed cumulative distribution function (cdf) and the expected cdf. In this test, the statistic, D, is defined as:

$$D = AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\log_e(F(X_i)) + \log_e(1 - F(X_{n-i+1}))]$$

(iii) **Chi-Squared Test**: Let O_i be the observed frequency and E_i be the expected frequency in a class i in the limits x_{i1} and x_{i2}, such that $E_i = F(x_{i2}) - F(x_{i1})$, then the chi-squared statistic, D, is defined as:

$$D = \chi^2 = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i}$$

The null hypothesis regarding the distributional form is rejected at the chosen significance level (α) if the test statistic is greater than the critical value, defined as $\chi^2_{(1-\alpha,\ k-1)}$.

It may be noted that since the three goodness-of-fit tests are defined differently, it is not necessary that the null hypothesis rejected (accepted) by any one test is also rejected (accepted) by the other test or tests.

3. The Biology Group of Disciplines: In this group of disciplines we have JIF values for 1043 journals (year 2006). The frequency distribution of JIF as well as $\log_{10}$(JIF) is skewed and with positive excess kurtosis (leptokurtic), indicating sharper peak and longer, fatter tails. Descriptive statistics for JIF(biology) and $\log_{10}$(JIF(Biology)) are presented in Table-1.1

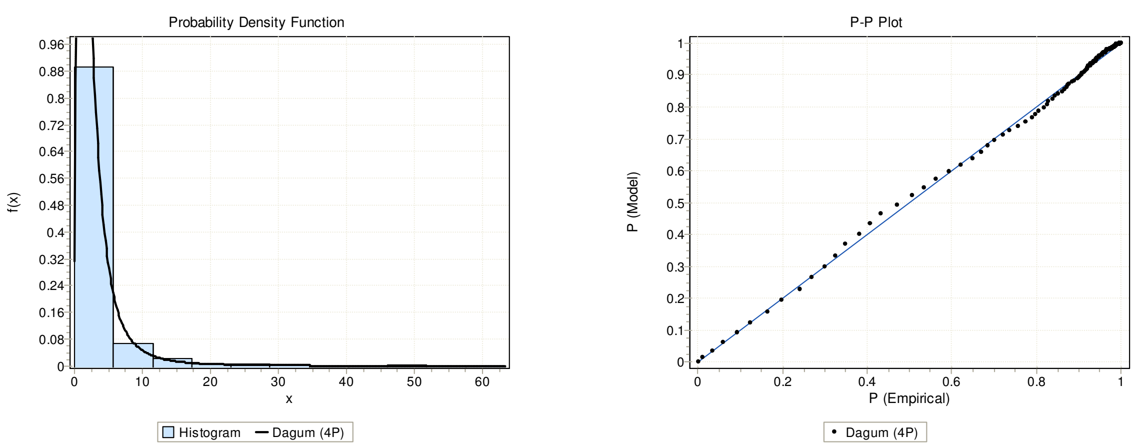
Table-1.1: Descriptive Statistics regarding the Journal Impact Factor for the Biology Group (year 2006)							
For the Natural Value of Journal Impact Factor				**For the Common Log Value of Journal Impact Factor**			
Statistic	**Value**	**Percentile**	**Value**	**Statistic**	**Value**	**Percentile**	**Value**
Sample Size	1043	Min	0.036	Sample Size	1043	Min	-1.4437
Range	63.306	5%	0.3702	Range	3.2454	5%	-0.43159
Mean	3.2541	10%	0.5626	Mean	0.29848	10%	-0.2498
Variance	21.914	25% (Q1)	1.094	Variance	0.18811	25% (Q1)	0.03902
Std. Deviation	4.6812	50% (Median)	2.161	Std. Deviation	0.43372	50% (Median)	0.33465
Coef. of Variation	1.4386	75% (Q3)	3.541	Coef. of Variation	1.4531	75% (Q3)	0.54913
Std. Error	0.14495	90%	5.9788	Std. Error	0.01343	90%	0.77661
Skewness	5.9781	95%	9.9736	Skewness	-0.312	95%	0.99885
Excess Kurtosis	52.114	Max	63.342	Excess Kurtosis	1.3483	Max	1.8017

The distributions best fitted to the JIF(Biology)/$\log_{10}$(JIF(Biology)) data are as follows.

i. **The natural Scale JIF Data (Biology) Group**: Three best fit distributions to the natural scale JIF data (for the year 2006) are: (a) Dagum 4p, (b) Dagum 3p, and (c) Burr 4p/Burr 3p. The details are given in Table 1.2.

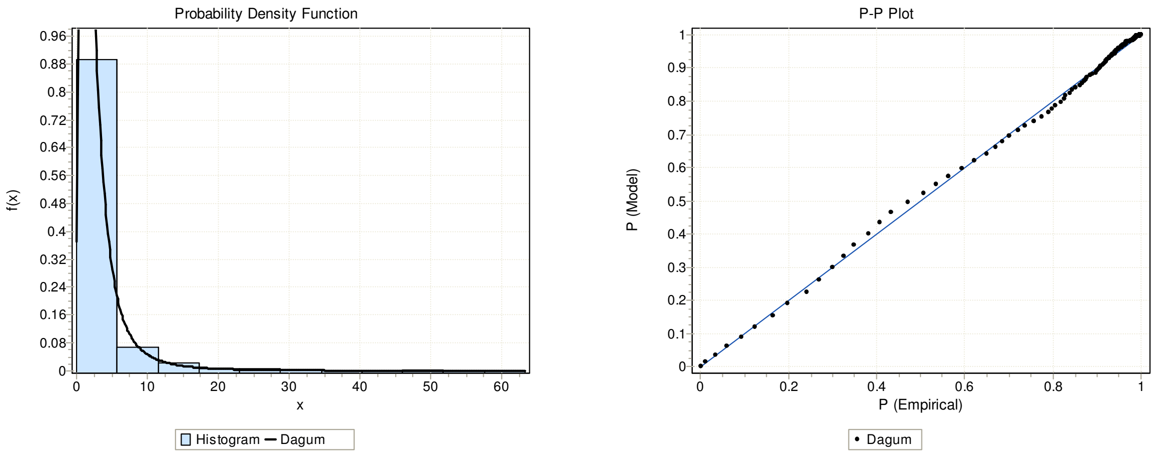
Table-1.2: Estimated Parameters and Goodness of Fit Statistics for Natural Scale JIF Data (Biology Group)				
Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Dagum 4p	k=0.65768; α=2.1501; β=2.7667; γ=0.02365	0.03302 (1) [0.20099]	1.0223 (2)	20.413 (1) [0.02558]
Dagum 3p	k=0.71122; α=2.1203; β=2.6457	0.03328 (2) [0.19403]	1.0187 (1)	20.711 (2) [0.0232]
Burr 4p	k=1.2824; α=1.7199; β=2.5214; γ=-0.00863	0.03854 (3) [0.08797]	1.3924 (3)	24.619 (4) [0.00612]
Burr 3p	k=1.3114; α=1.6966; β=2.5603	0.0388 (4) [0.08432]	1.424 (4)	24.39 (3) [0.00663]

Fig.1.1: Histogram, pdf and P-P plot of Dagum 4p Distribution fitted to Natural Scale JIF Data (Biology Group)



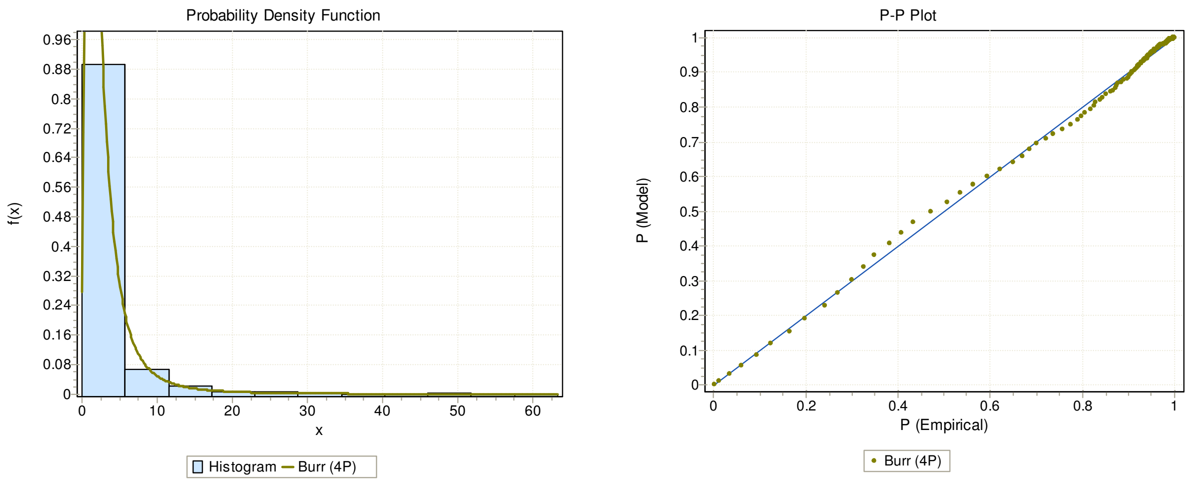
...

Fig.1.2: Histogram, pdf and P-P plot of Dagum 3p Distribution fitted to Natural Scale JIF Data (Biology Group)



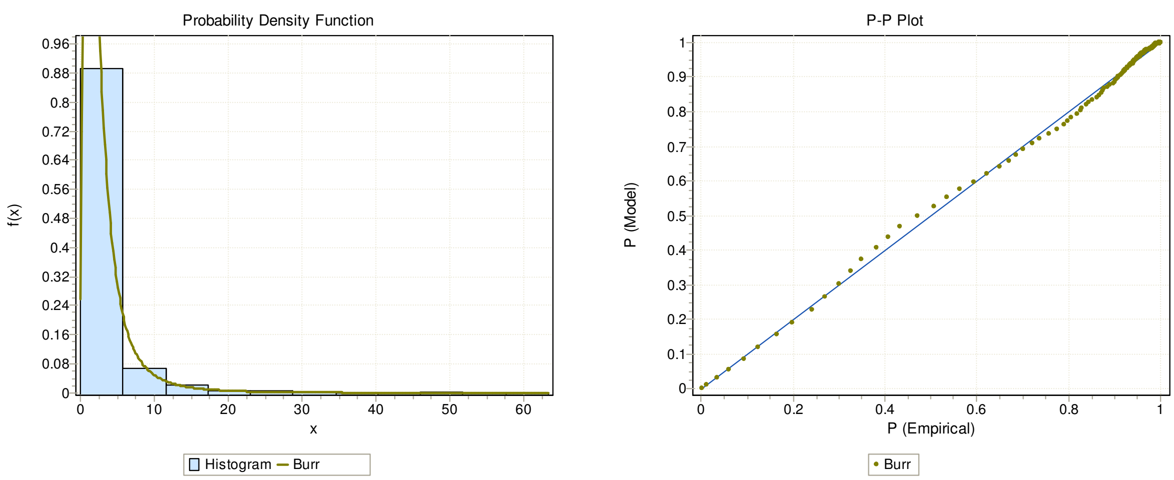
...

Fig.1.3: Histogram, pdf and P-P plot of Burr 4p Distribution fitted to Natural Scale JIF Data (Biology Group)



...

Fig.1.4: Histogram, pdf and P-P plot of Burr 3p Distribution fitted to Natural Scale JIF Data (Biology Group)

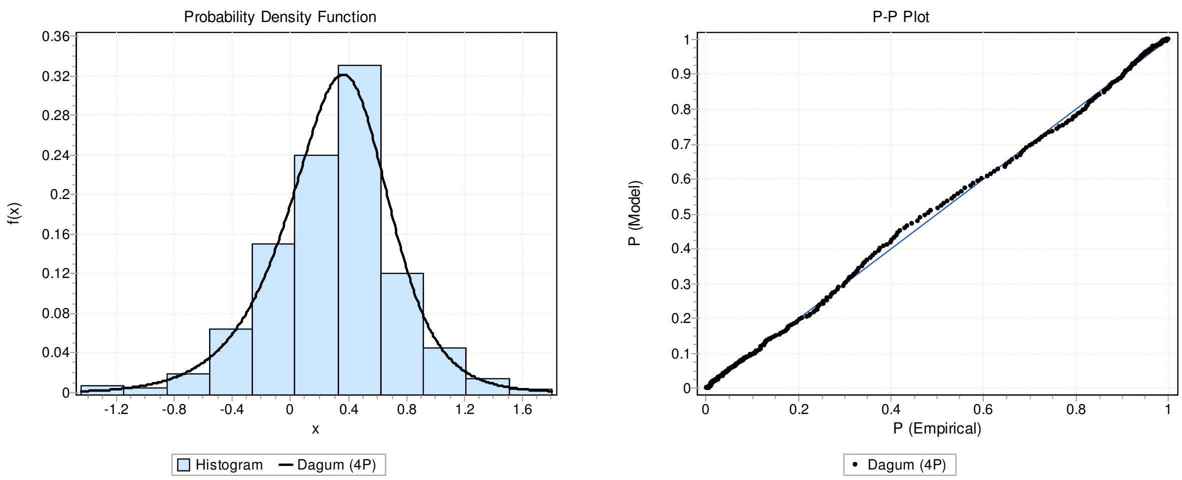


ii. **The Logarithmic JIF Data**: Three best fit distributions to the log(JIF) data (for the year 2006) are: (a) Dagum 4p, (b) Burr 4p, and (c) Johnson SU. The details are given in Table 1.3.

Table-1.3: Estimated Parameters and Goodness of Fit Statistics for Log_{10}(JIF) Data (Biology Group)

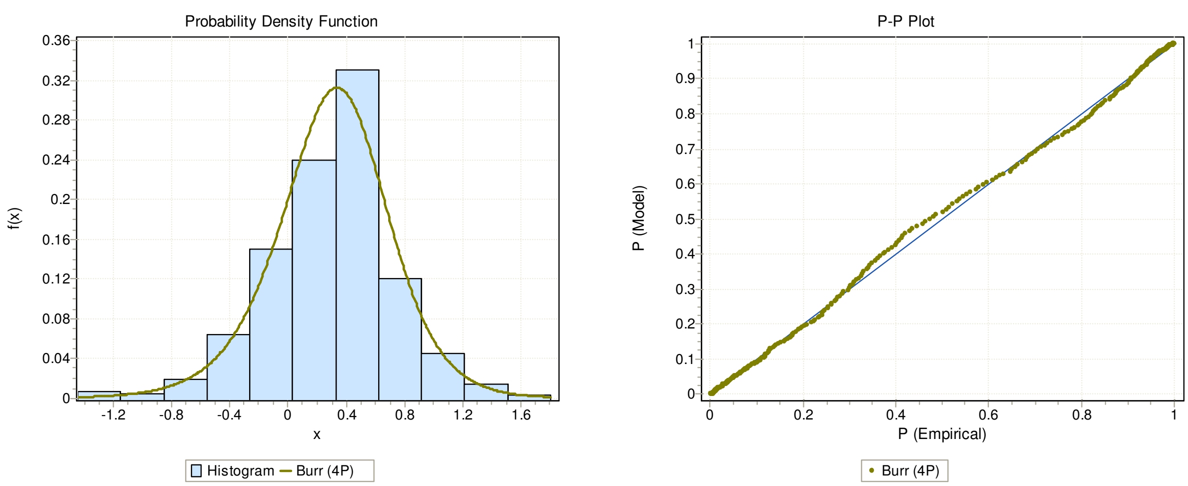
Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Dagum 4p	k=0.55244; α=29.113; β=5.4948; γ=-5.0032	0.03249 (1) [0.21621]	0.91381 (1)	19.369 (1) [0.03582]
Burr 4p	k=1.279; α=1.1329E+8 β=2.8662E+7 ; γ=-2.8662E+7	0.03883 (2) [0.08384]	1.4286 (2)	25.907 (3) [0.00387]
Johnson SU	γ=0.39595; δ=2.1606 λ=0.82337; ξ=0.46737	0.03972 (3) [0.07247]	1.5906 (3)	25.625 (2) [0.00428]

Fig.1.5: Histogram, pdf and P-P plot of Dagum 4p Distribution fitted to Log_{10}(JIF) Data (Biology Group)



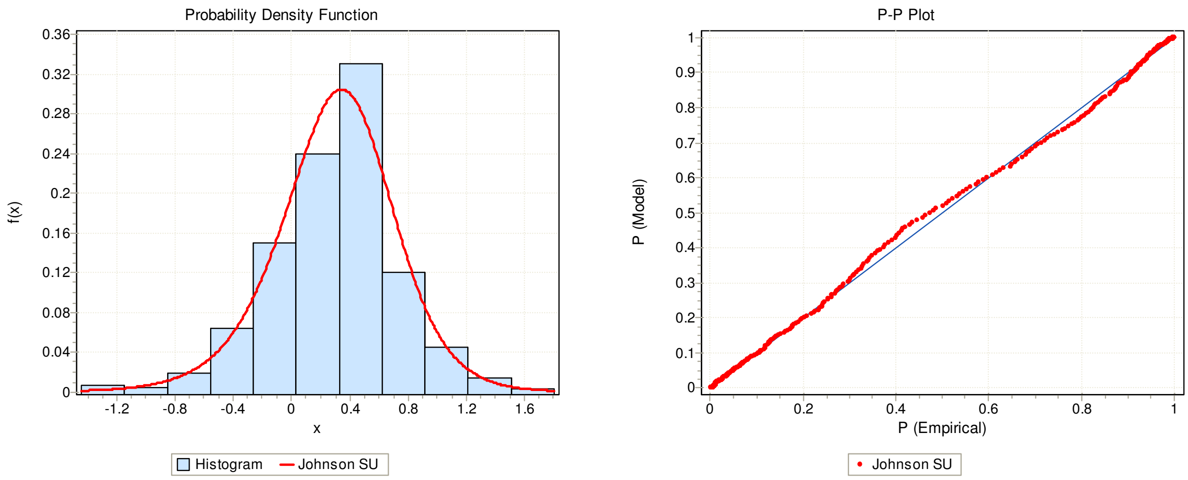
..

Fig.1.6: Histogram, pdf and P-P plot of Burr 4p Distribution fitted to Log_{10}(JIF) Data (Biology Group)



...

Fig.1.7: Histogram, pdf and P-P plot of Johnson SU Distribution fitted to Log_{10}(JIF) Data (Biology Group)



4. Economics and Statistics Group: In this group of disciplines we have JIF values for 1043 journals (year 2009, Source RePEc, IDEAS). It is pertinent to mention here that *Thomson Scientific* does not include many journals in its documentation and therefore many journals in economics do not have the journal impact factor published by *Thomson Scientific*. The Research Papers in Economics Project (RePEc) and Internet Documents in Economics Access Service (IDEAS) fill in this gap and provide the JIF data for the journals included in the project. It includes many statistics journals also. It is systematically and regularly updated. In this study we have used the journal Impact Factor data from this source. We find that for Ecostat group *t*he frequency distribution of JIF is skewed and with positive excess kurtosis (leptokurtic), indicating sharper peak and longer, fatter tails, while $\log_{10}$(JIF) frequency distribution is skewed and with negative excess (but only meager) kurtosis (platykurtic), indicating slightly flatter peak and shorter, thinner tails. Descriptive statistics for JIF(Ecostat) and $\log_{10}$(JIF(Ecostat)) are presented in Table-2.1

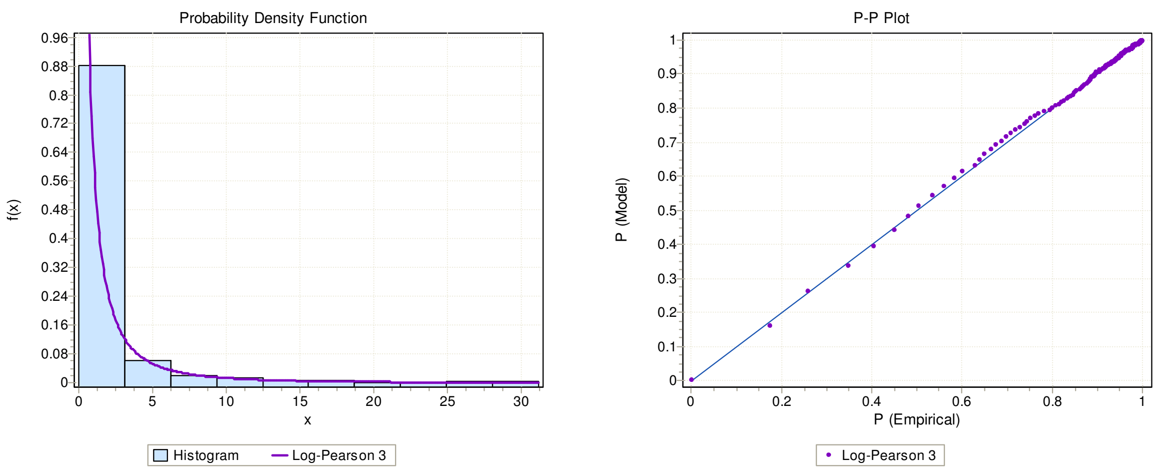
Table-2.1: Descriptive Statistics regarding the Journal Impact Factor for the Ecostat Group (year 2009)							
For the Natural Value of Journal Impact Factor				**For the Common Log Value of Journal Impact Factor**			
Statistic	**Value**	**Percentile**	**Value**	**Statistic**	**Value**	**Percentile**	**Value**
Sample Size	796	Min	0.001	Sample Size	796	Min	-3
Range	31.168	5%	0.017	Range	4.4937	5%	-1.7696
Mean	1.5157	10%	0.0307	Mean	-0.41921	10%	-1.5129
Variance	10.409	25% (Q1)	0.12025	Variance	0.62836	25% (Q1)	-0.91992
Std. Deviation	3.2264	50% (Median)	0.4175	Std. Deviation	0.79269	50% (Median)	-0.37934
Coef. of Variation	2.1287	75% (Q3)	1.4733	Coef. of Variation	-1.8909	75% (Q3)	0.16828
Std. Error	0.11436	90%	3.8969	Std. Error	0.0281	90%	0.59072
Skewness	4.7276	95%	6.8605	Skewness	-0.24343	95%	0.83636
Excess Kurtosis	28.722	Max	31.169	Excess Kurtosis	-0.18429	Max	1.4937

The distributions best fitted to the JIF(Ecostat)/$\log_{10}$(JIF(Ecostat)) data are as follows.

i. **The natural Scale JIF Data (Ecostat Group)**: Three best fit distributions to the natural scale JIF data (for the year 2009) are: (a) Log Pearson-III, (b) Log Normal 3p, and (c) Burr 4p/Burr 3p or Generalized Gamma 4p. The details are given in Table 2.2.

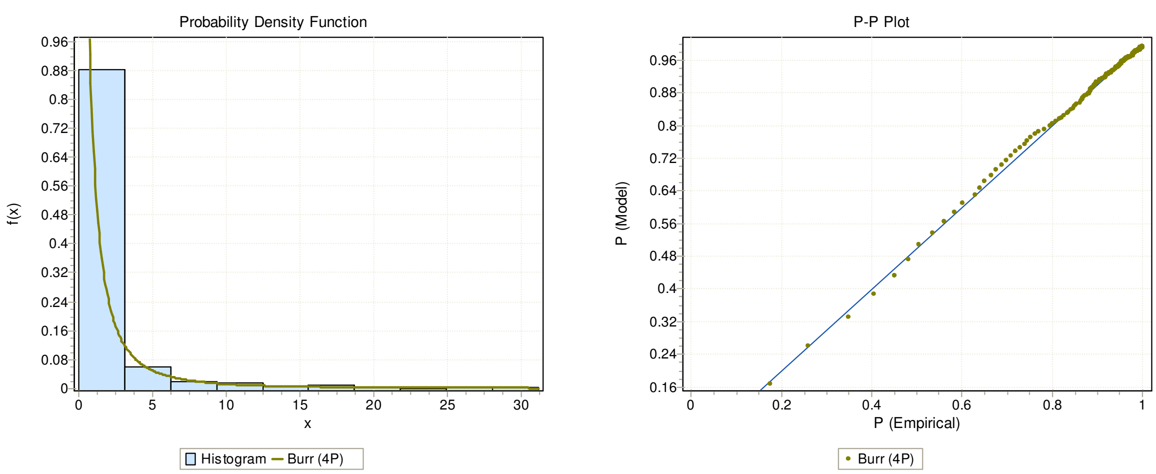
Table-2.2: Estimated Parameters and Goodness of Fit Statistics for Natural Scale JIF Data (Ecostat Group)				
Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Log Pearson III	α=67.499; β=-0.22216; γ=14.03	0.01799 (1) [0.95485]	0.35564 (1)	8.7523 (1) [0.46044]
Burr 4p	k=1.969; α=0.76809; β=1.2971; γ=0.001	0.02472 (2) [0.7057]	4.6434 (14)	NA
Log Normal 3p	σ=1.8182; μ=-0.96207; γ=-2.1565E-4	0.02773 (3) [0.56349]	1.0066 (4)	10.474 (2) [0.31348]
Gen. Gamma 4p	k=0.29597; α=3.7396; β=0.00709; γ=-7.1996E-4	0.03257 (5) [0.35962]	0.86905 (2)	13.902 (5) [0.12584]
Burr 3p	k=1.967; α=0.79588; β=1.2807	0.03433 (9) [0.29825]	0.99414 (3)	10.793 (3) [0.29014]

Fig.2.1: Histogram, pdf and P-P plot of Log Pearson-III Distribution fitted to Natural Scale JIF Data (Ecostat Group)



...

Fig.2.2: Histogram, pdf and P-P plot of Burr 4p Distribution fitted to Natural Scale JIF Data (Ecostat Group)



...

Fig.2.3: Histogram, pdf and P-P plot of Log-Normal 3p Distribution fitted to Natural Scale JIF Data (Ecostat Group)

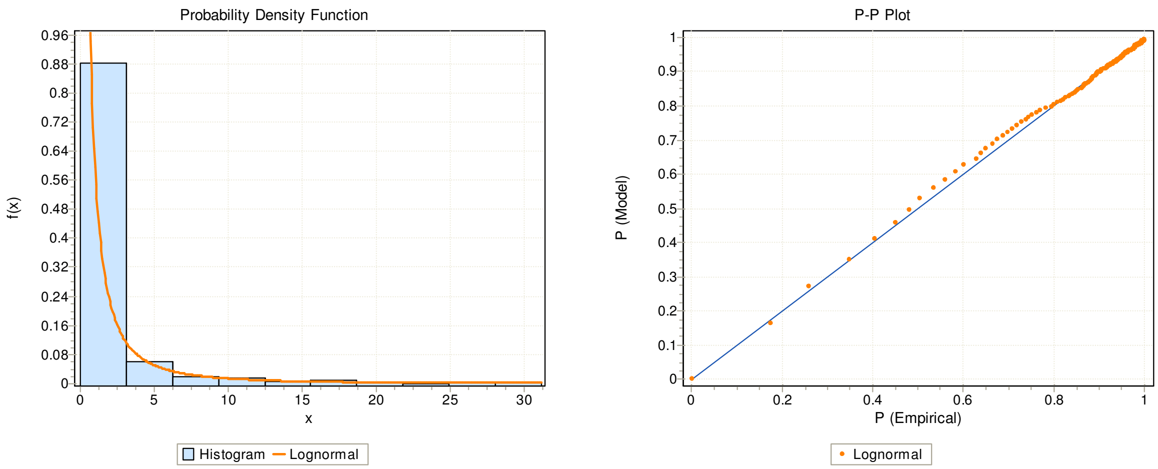
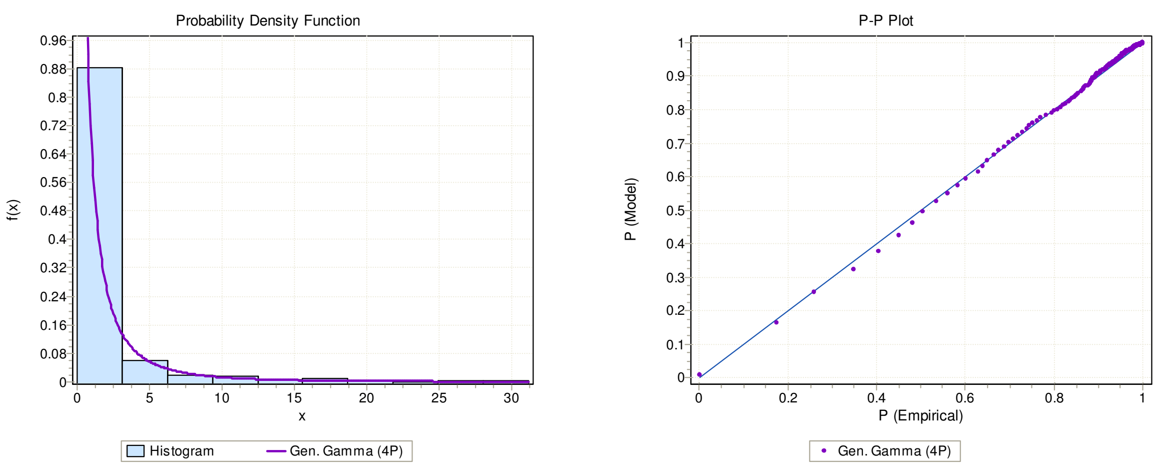
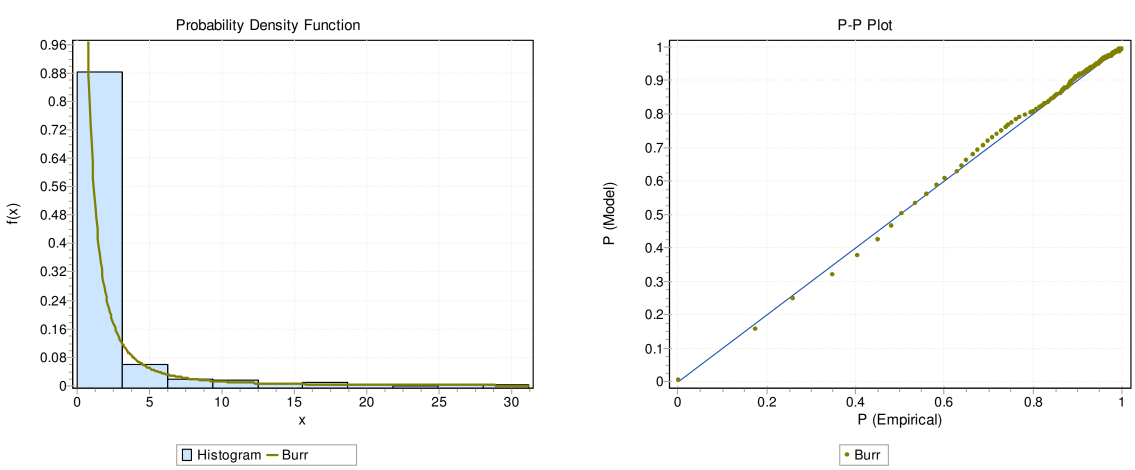


Fig.2.4: Histogram, pdf and P-P plot of Generalize Gamma 4p Distribution fitted to Natural Scale JIF Data (Ecostat Group)



..

Fig.2.5: Histogram, pdf and P-P plot Burr-3p Distribution fitted to Natural Scale JIF Data (Ecostat Group)

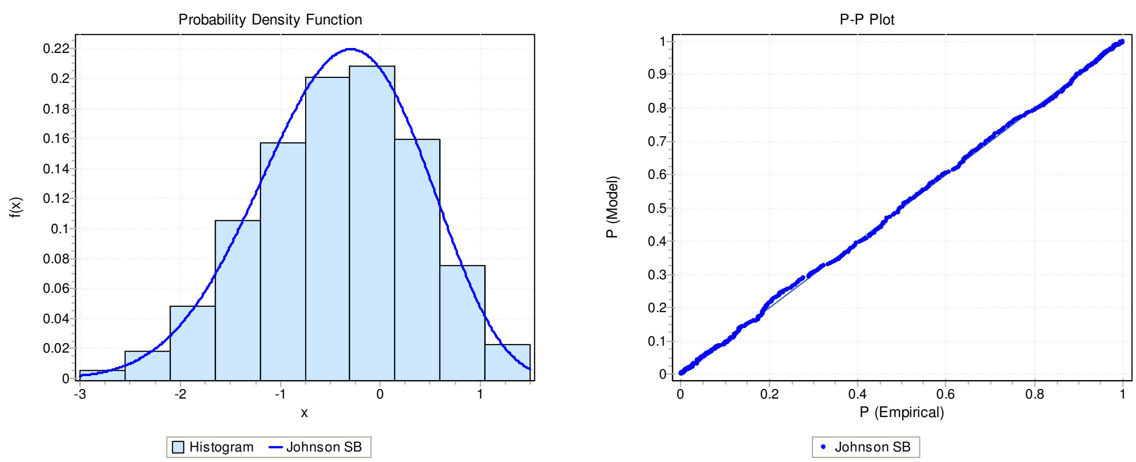


ii. **The Logarithmic JIF Data (Ecostat Group)**: Three best fit distributions to the $\log_{10}$(JIF) data (for the year 2009) are: (a) Johnson SB, (b) Burr 4p, and (c) Weibull 3p. The details are given in Table 2.3. Although the Kumaraswamy distribution fits well to the data, but we have no reason to assume the fixed lower and upper limits (a and b) on $\log_{10}$(JIF(Ecostat)) data. Hence, we reject it.

Table-2.3: Estimated Parameters and Goodness of Fit Statistics for Log_{10}(JIF) Data (Ecostat Group)				
Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Johnson SB	γ=-1.0887; δ=2.3375; λ=8.0953; ξ=-5.3557	0.01769 (1) [0.96074]	0.26769 (1)	8.4083 (3) [0.49358]
Burr 4p	k=141.05; α=4.9526; β=10.1; γ=-3.8341	0.01803 (2) [0.95401]	0.26807 (2)	8.2837 (1) [0.50583]
Weibull 3p	α=4.9119; β=3.7024; γ=-3.8144	0.01841 (4) [0.94544]	0.26925 (3)	8.2917 (2) [0.50504]
Kumaraswamy	α_1=4.352; α_2=13.638; a=-3.5658; b=2.7954	0.01826 (3) [0.94895]	0.28028 (4)	8.8839 (4) [0.44806]

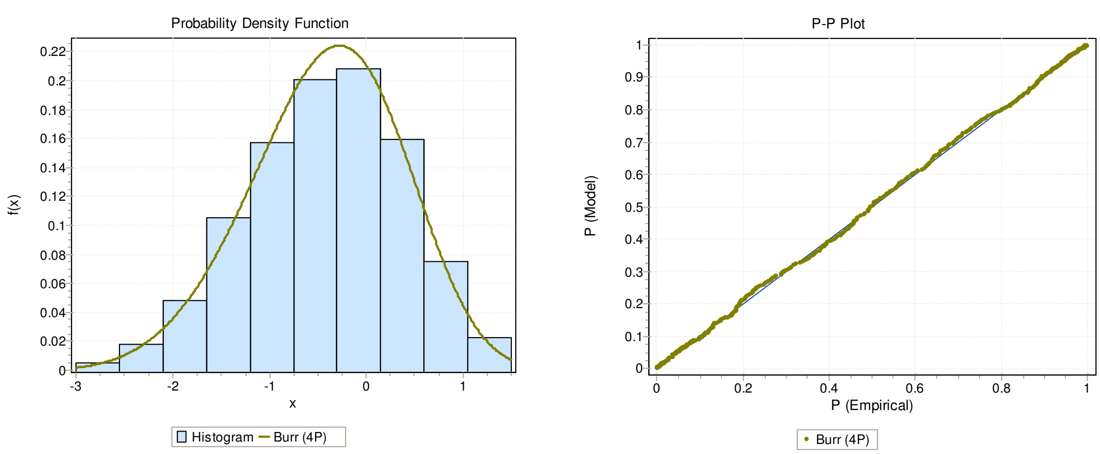
...

Fig.2.6: Histogram, pdf and P-P plot Johnson SB Distribution fitted to Log_{10}(JIF) Data (Ecostat Group)



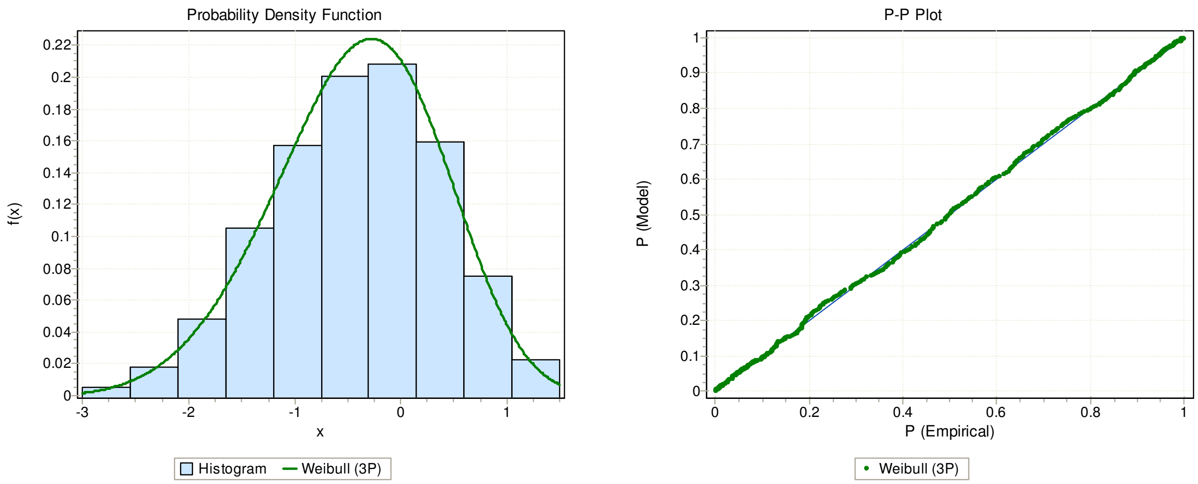
.

Fig.2.7: Histogram, pdf and P-P plot Burr 4p Distribution fitted to Log_{10}(JIF) Data (Ecostat Group)



..

Fig.2.8: Histogram, pdf and P-P plot Weibull 3p Distribution fitted to Log_{10}(JIF) Data (Ecostat Group)



5. Chemistry Group: In this group of disciplines we have JIF values for 433 journals (year 2006). The frequency distribution of JIF as well as log_{10}(JIF) is skewed and with positive excess kurtosis (leptokurtic), indicating sharper peak and longer, fatter tails. Descriptive statistics for JIF(Chemistry) and log_{10}(JIF(Chemistry)) are presented in Table-3.1

Table-3.1: Descriptive Statistics regarding the Journal Impact Factor for the Chemistry Group (year 2006)							
For the Natural Value of Journal Impact Factor				**For the Common Log Value of Journal Impact Factor**			
Statistic	**Value**	**Percentile**	**Value**	**Statistic**	**Value**	**Percentile**	**Value**
Sample Size	433	Min	0.051	Sample Size	433	Min	-1.2924
Range	26.003	5%	0.2429	Range	2.7083	5%	-0.6148
Mean	2.0454	10%	0.3512	Mean	0.10382	10%	-0.45445
Variance	6.3888	25% (Q1)	0.6315	Variance	0.18198	25% (Q1)	-0.19963
Std. Deviation	2.5276	50% (Median)	1.256	Std. Deviation	0.4266	50% (Median)	0.09899
Coef. of Variation	1.2357	75% (Q3)	2.544	Coef. of Variation	4.109	75% (Q3)	0.40552
Std. Error	0.12147	90%	4.153	Std. Error	0.0205	90%	0.61836
Skewness	4.2725	95%	6.0311	Skewness	-0.03347	95%	0.7804
Excess Kurtosis	27.769	Max	26.054	Excess Kurtosis	0.01634	Max	1.4159

The distributions best fitted to the JIF(Chemistry)/log_{10}(JIF(Chemistry)) data are as follows.

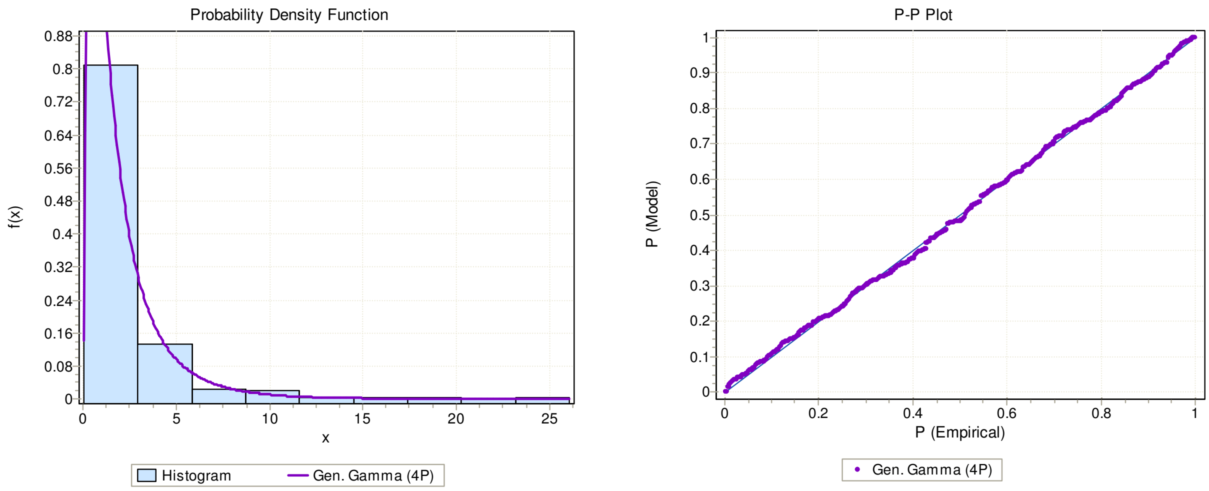
i. **The natural Scale JIF Data (Chemistry Group)**: Three best fit distributions to the natural scale JIF data (for the year 2006) are: (a) Gen Gamma 4p, (b) Inv. Gaussian 3p/Log-Pearson-III, and (c) Lognormal 3p or 2p. The details are given in Table 3.2.

Table-3.2: Estimated Parameters and Goodness of Fit Statistics for Natural Scale JIF Data (Chemistry Group)				
Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Gen. Gamma 4p	k=0.37857; α=6.754;	0.02363 (1) [0.96433]	0.33907 (6)	2.8368 (1) [0.94418]

	β=0.00938; γ=0.03765			
Inv. Gaussian 3p	λ=1.7818; μ=2.1346; γ=-0.08917	0.0242 (2) [0.9563]	0.2885 (4)	4.1448 (6) [0.84383]
Log-Pearson-III	α=3571.2; β=-0.01644; γ=58.939	0.0248 (3) [0.94667]	0.22902 (1)	3.4221 (2) [0.90515]
Lognormal 3p	σ=0.97341; μ=0.24688; γ=-0.00618	0.02604 (4) [0.92326]	0.24365 (2)	3.6386 (3) [0.88817]
Lognormal 2p	σ=0.98114; μ=0.23905	0.02618 (5) [0.92023]	0.24797 (3)	3.7747 (5) [0.87686]

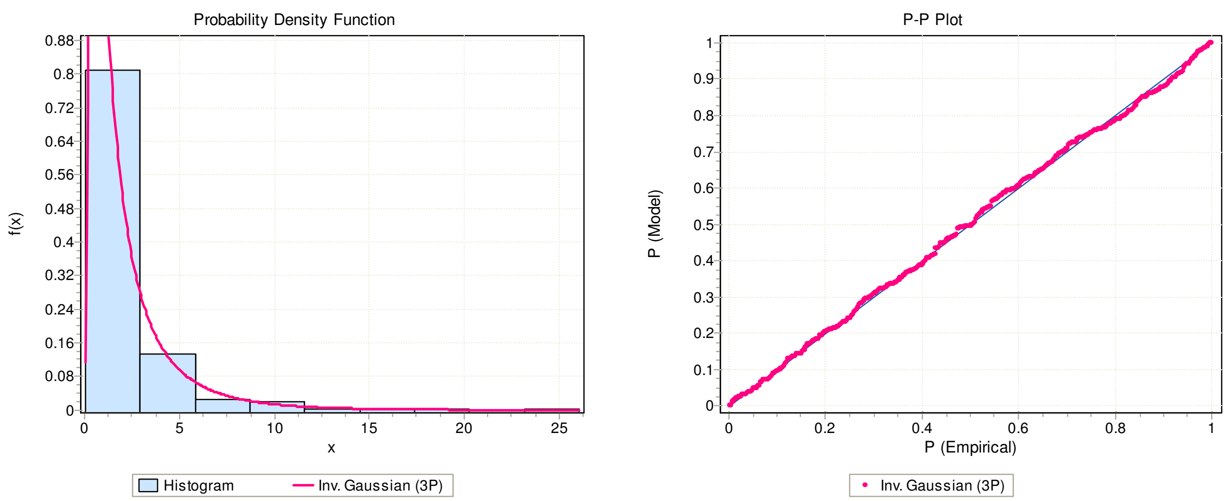
..

Fig.3.1: Histogram, pdf and P-P plot of Gen. Gamma 4p Distribution fitted to Natural Scale JIF Data (Chemistry Group)



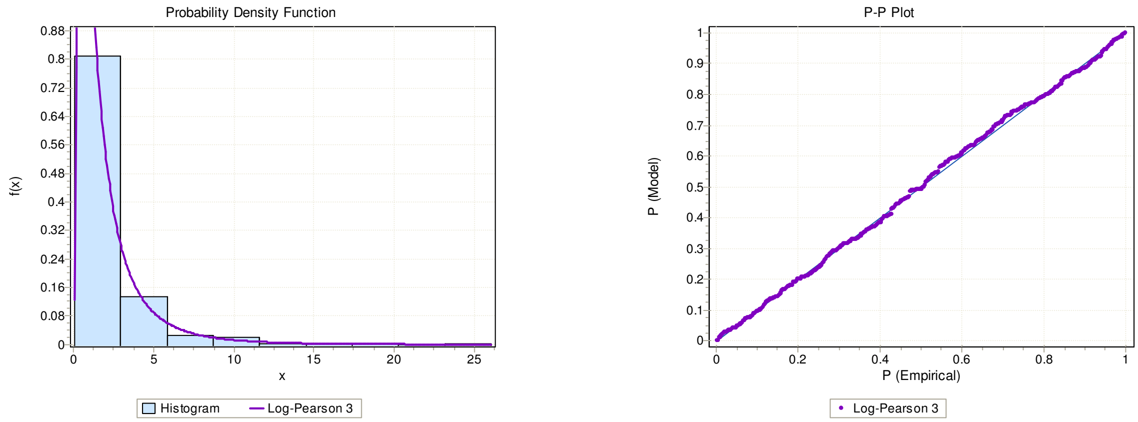
...

Fig.3.2: Histogram, pdf and P-P plot of Inverse Gaussian 3p Distribution fitted to Natural Scale JIF Data (Chemistry Group)



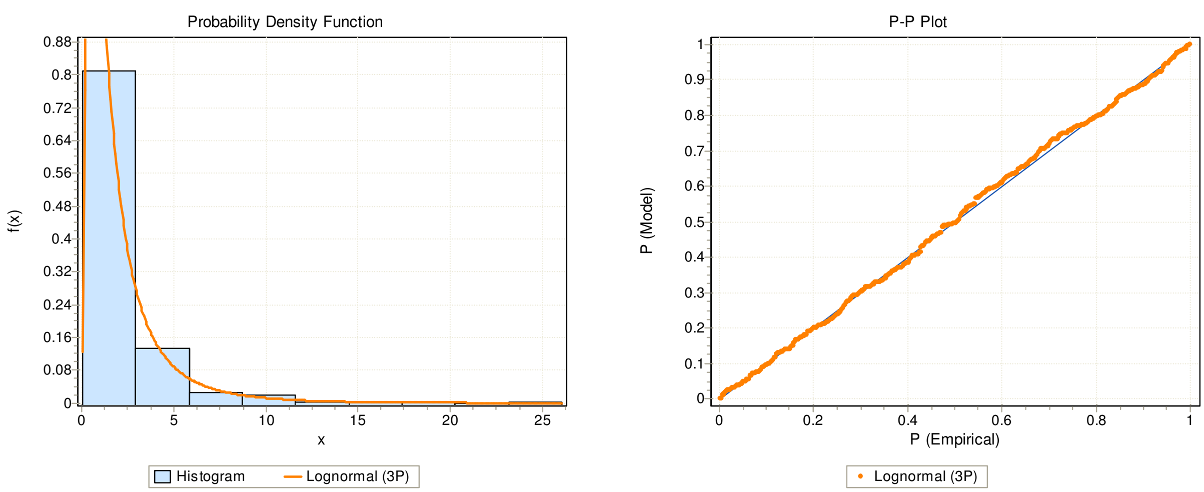
..

Fig.3.3: Histogram, pdf and P-P plot of Log-Pearson-III Distribution fitted to Natural Scale JIF Data (Chemistry Group)



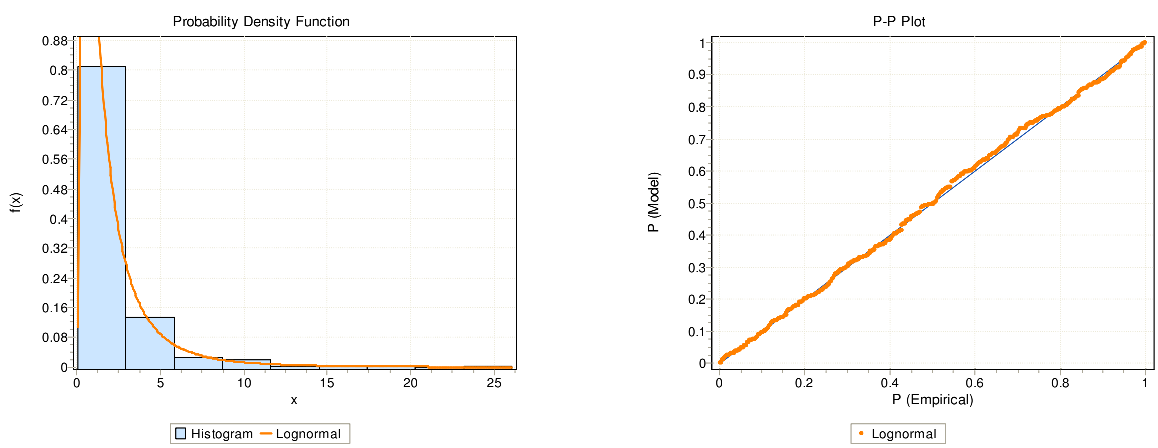
..

Fig.3.4: Histogram, pdf and P-P plot of Log-Normal 3p Distribution fitted to Natural Scale JIF Data (Chemistry Group)



..

Fig.3.5: Histogram, pdf and P-P plot of Log-Normal 2p Distribution fitted to Natural Scale JIF Data (Chemistry Group)



..

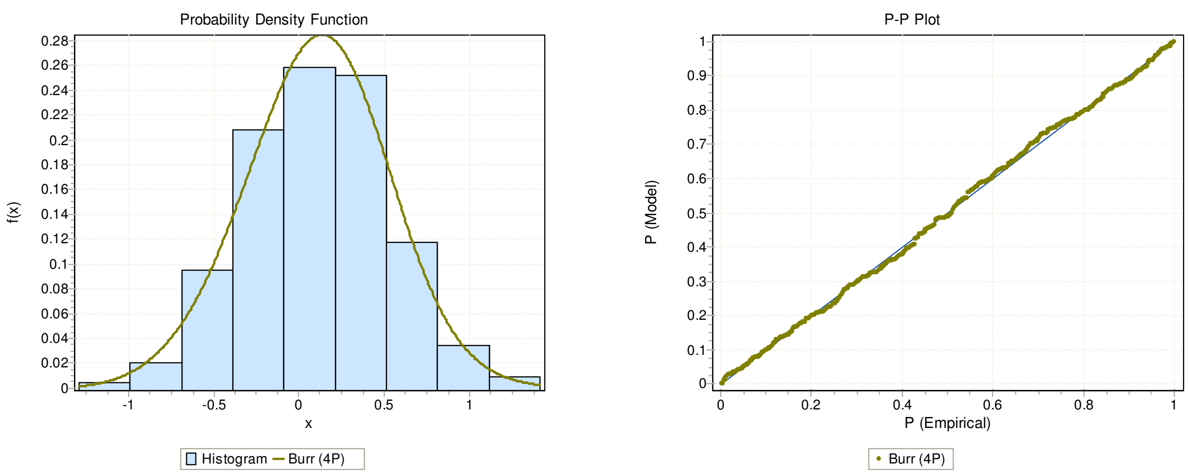
ii. **The Logarithmic JIF Data (Chemistry Group)**: Three best fit distributions to the $\log_{10}$(JIF) data (for the year 2006) are: (a) Burr 4p, (b) Johnson SU and (c) Weibull 3p. The details are given in Table 3.3. Although the Kumaraswamy and General Extreme Value distributions fit well to the data (rank 2 and 3 respectively according to KS criterion), but we reject them on other goodness of fit criteria.

Table-3.3: Estimated Parameters and Goodness of Fit Statistics for Log_{10}(JIF) Data (Chemistry Group)

Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Burr 4p	k=5.4133; α=5.4602; β=2.7074; γ=-1.7688	0.02488 (4) [0.94537]	0.22912 (1)	2.9592 (1) [0.93689]
Johnson SU	γ=3.0468; δ=16.463; λ=6.8914; ξ=1.3889	0.02513 (5) [0.94091]	0.23211 (2)	3.4258 (2) [0.90487]
Weibull 3p	α=4.2074; β=1.7776 γ=-1.5139	0.02102 (1) [0.98889]	0.31535 (9)	3.9722 (9) [0.85962]

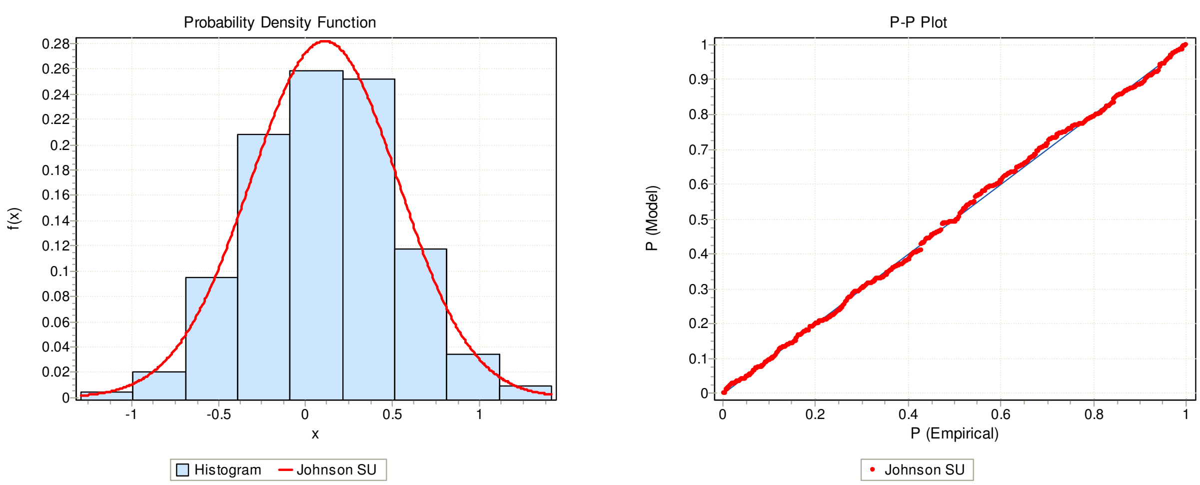
.

Fig.3.6: Histogram, pdf and P-P plot Burr 4p Distribution fitted to Log_{10}(JIF) Data (Chemistry Group)



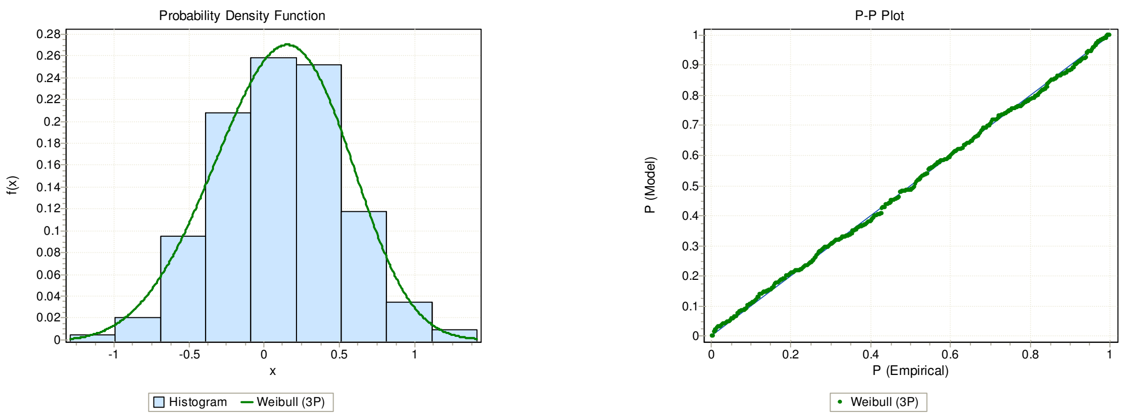
.

Fig.3.7: Histogram, pdf and P-P plot Johnson SU Distribution fitted to Log_{10}(JIF) Data (Chemistry Group)



..

Fig.3.8: Histogram, pdf and P-P plot Weibull 3p Distribution fitted to Log_{10}(JIF) Data (Chemistry Group)



6. Engineering Group: In this group of disciplines we have JIF values for 706 journals (year 2006). The frequency distribution of JIF as well as $\log_{10}$(JIF) is skewed and with positive excess kurtosis (leptokurtic), indicating sharper peak and longer, fatter tails. Descriptive statistics for JIF(Engineering) and $\log_{10}$(JIF(Engineering)) are presented in Table-4.1

Table-4.1: Descriptive Statistics regarding the Journal Impact Factor for the Engineering Group (year 2006)							
For the Natural Value of Journal Impact Factor				**For the Common Log Value of Journal Impact Factor**			
Statistic	**Value**	**Percentile**	**Value**	**Statistic**	**Value**	**Percentile**	**Value**
Sample Size	706	Min	0.001	Sample Size	706	Min	-3
Range	10.532	5%	0.0797	Range	4.0226	5%	-1.0986
Mean	0.87214	10%	0.1485	Mean	-0.24762	10%	-0.82833
Variance	0.75291	25% (Q1)	0.333	Variance	0.20848	25% (Q1)	-0.47756
Std. Deviation	0.8677	50% (Median)	0.645	Std. Deviation	0.4566	50% (Median)	-0.19044
Coef. of Variation	0.99491	75% (Q3)	1.1098	Coef. of Variation	-1.8439	75% (Q3)	0.04522
Std. Error	0.03266	90%	1.8322	Std. Error	0.01718	90%	0.26297
Skewness	3.5129	95%	2.5093	Skewness	-1.0626	95%	0.39956
Excess Kurtosis	24.995	Max	10.533	Excess Kurtosis	2.8965	Max	1.0226

The distributions best fitted to the JIF(engineering)/$\log_{10}$(JIF(Engineering)) data are as follows.

i. **The natural Scale JIF Data (Engineering Group)**: Three best fit distributions to the natural scale JIF data (2006) are: (a) Dagum 3p, (b) Burr 4p, and (c) Gen. Extreme Value. The details are given in Table 4.2.

Table-4.2: Estimated Parameters and Goodness of Fit Statistics for Natural Scale JIF Data (Engineering Group)				
Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Dagum 3p	k=0.45393; α=2.6009; β=1.0606	0.01841 (1) [0.96685]	0.21175 (1)	4.6778 (2) [0.86144]
Dagum 4p	k=0.45041; α=2.6067; β=1.065; γ=4.1339E-4	0.0186 (2) [0.96383]	0.21661 (2)	4.8777 (3) [0.84484]
Burr 4p	k=3.1813; α=1.4014; β=1.7533; γ=-0.00195	0.02032 (3) [0.92685]	0.39581 (3)	6.9301 (5) [0.6444]
Gen. Extreme Value	k=0.26357; σ=0.42276; μ=0.48092	0.0217 (4) [0.88634]	0.52096 (5)	2.5618 (1) [0.97917]

Fig.4.1: Histogram, pdf and P-P plot of Dagum 3p Distribution fitted to Natural Scale JIF Data (Engineering Group)

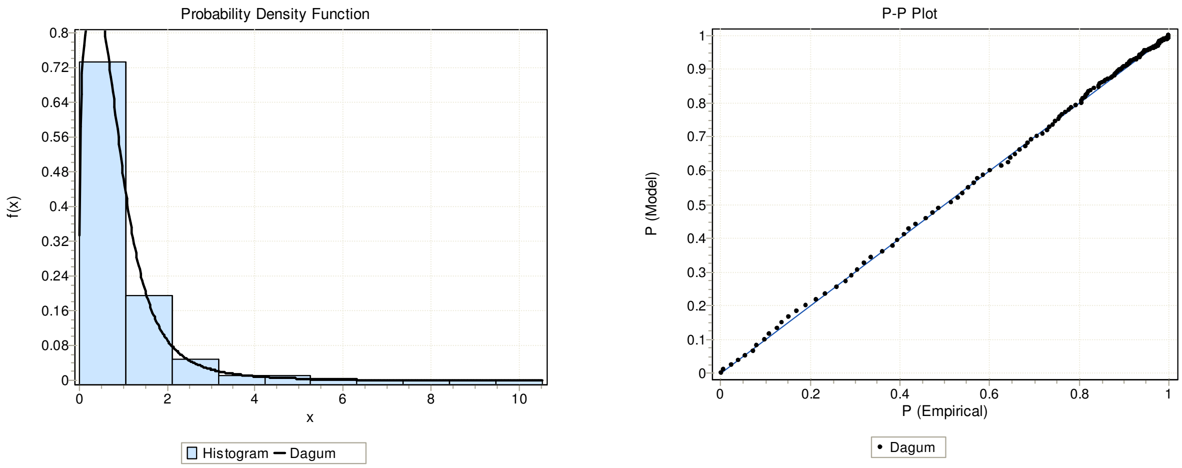


Fig.4.2: Histogram, pdf and P-P plot of Burr 4p Distribution fitted to Natural Scale JIF Data (Engineering Group)

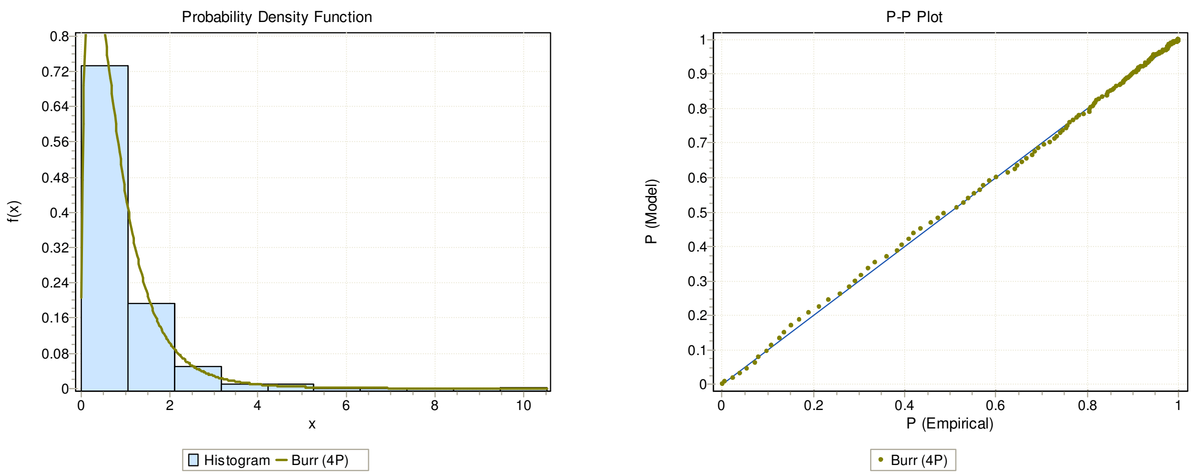
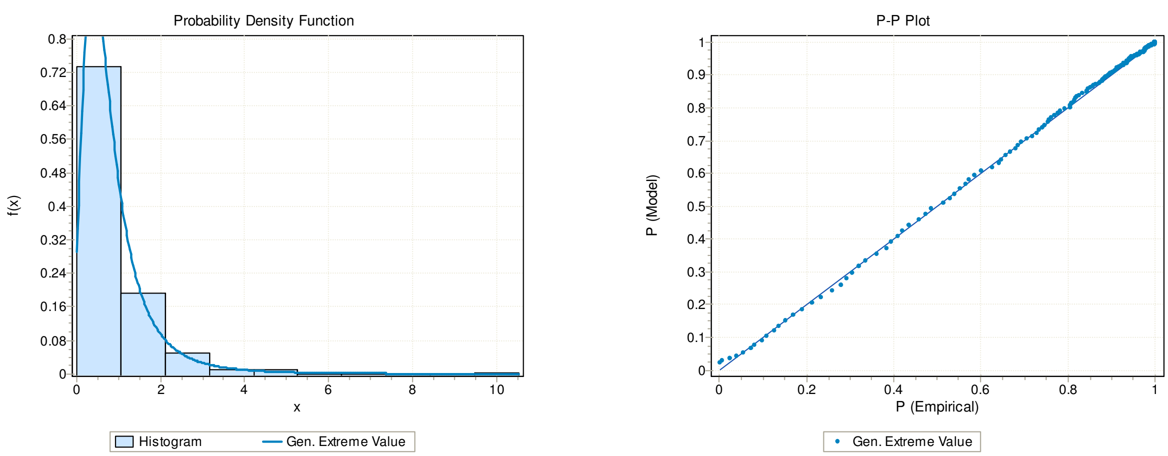


Fig.4.3: Histogram, pdf and P-P plot of Gen. Extreme Value Distribution fitted to Natural Scale JIF Data (Engineering Group)



ii. **The Logarithmic JIF Data (Engineering Group)**: Three best fit distributions to the $\log_{10}$(JIF) data (for the year 2006) are: (a) Dagum 4p, (b) Johnson SU and (c) Log Logistic 3p. The details are given in Table 4.3.

Table-4.3: Estimated Parameters and Goodness of Fit Statistics for Log_{10}(JIF) Data (Engineering Group)

Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Dagum 4p	k=0.44956; α=3.0816E+7; β=5.1118E+6; γ=-5.1118E+6	0.01879 (1) [0.96045]	0.21938 (1)	4.6815 (1) [0.86114]
Johnson SU	γ=1.5001; δ=2.1023; λ=0.66062; ξ=0.32618	0.02786 (2) [0.63339]	0.44514 (2)	7.6222 (2) [0.57262]
Log Logistic 3p	α=5.1604E+8; β=1.2559E+8 ; γ=-1.2559E+8	0.03748 (3) [0.26767]	2.7469 (8)	13.412 (4) [0.14484]
Burr 4p	k=5.7185; α=9.5049E+7; β=3.3308E+7; γ=-3.3308E+7	0.04024 (5) [0.19763]	1.3982 (3)	12.844 (3) [0.16979]

..

Fig.4.4: Histogram, pdf and P-P Dagum 4p Distribution fitted to Log_{10}(JIF) Data (Engineering Group)

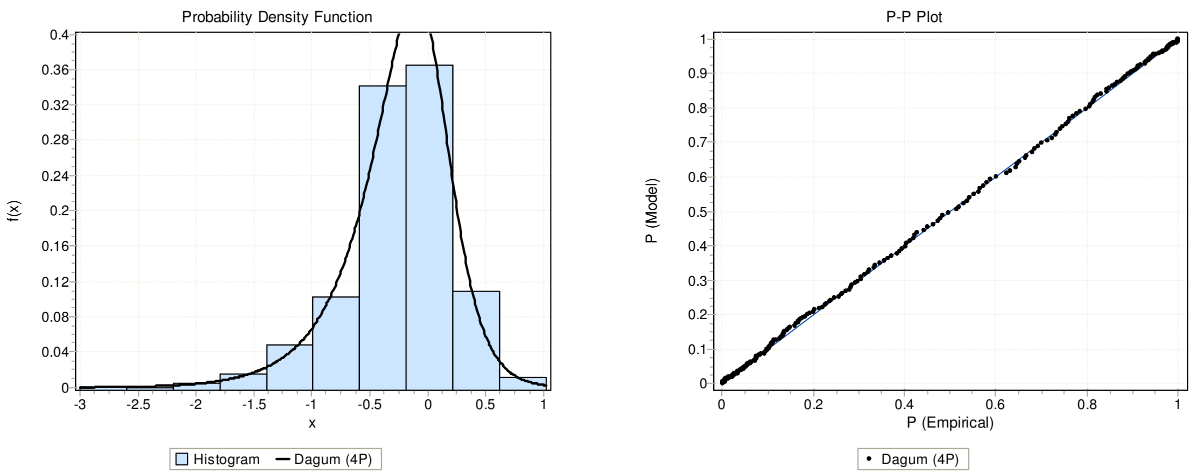
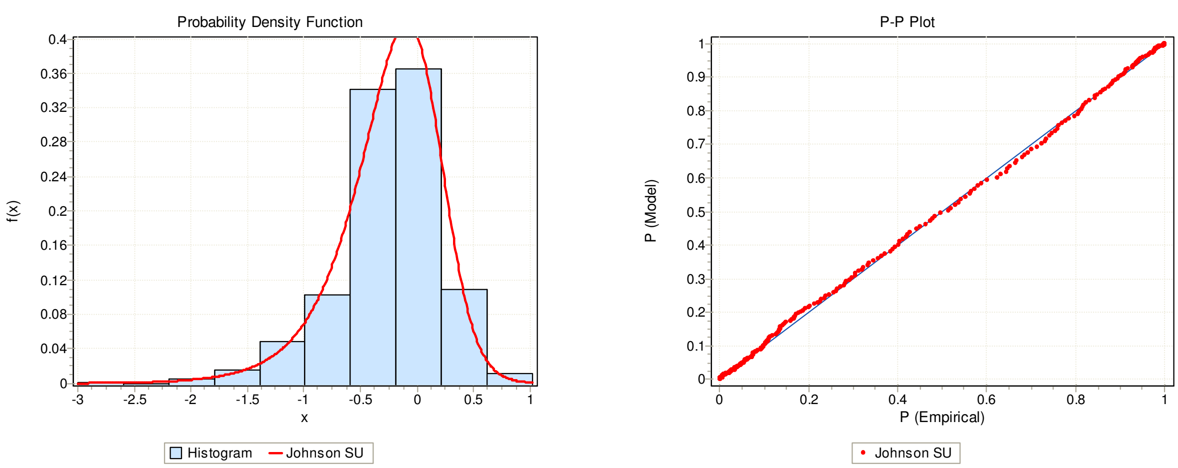
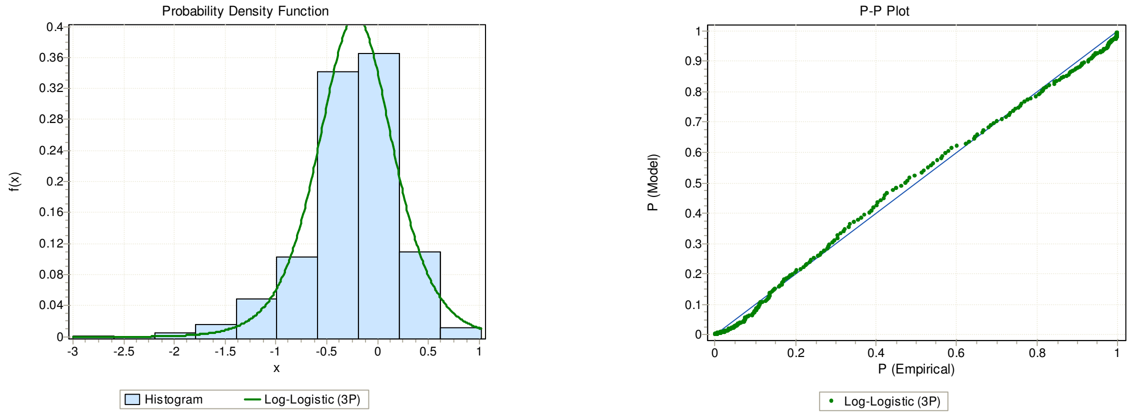


Fig.4.5: Histogram, pdf and P-P Johnson SU Distribution fitted to Log_{10}(JIF) Data (Engineering Group)



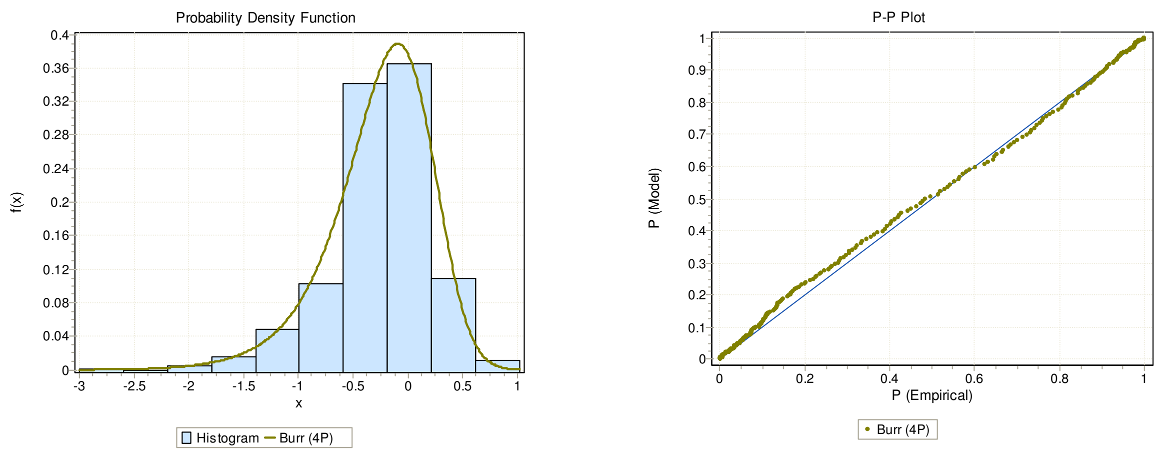
..

Fig.4.6: Histogram, pdf and P-P Log Logistic 3p Distribution fitted to Log_{10}(JIF) Data (Engineering Group)



..

Fig.4.7: Histogram, pdf and P-P Burr 4p Distribution fitted to Log_{10}(JIF) Data (Engineering Group)



7. Physics Group: In this group of disciplines we have JIF values for 294 journals (year 2006). The frequency distribution of JIF as well as $\log_{10}$(JIF) is skewed and with positive excess kurtosis (leptokurtic), indicating sharper peak and longer, fatter tails. Descriptive statistics for JIF(Physics) and $\log_{10}$(JIF(Physics)) are presented in Table-5.1

Table-5.1: Descriptive Statistics regarding the Journal Impact Factor for the Physics Group (year 2006)							
For the Natural Value of Journal Impact Factor				**For the Common Log Value of Journal Impact Factor**			
Statistic	**Value**	**Percentile**	**Value**	**Statistic**	**Value**	**Percentile**	**Value**
Sample Size	294	Min	0.044	Sample Size	294	Min	-1.3565
Range	33.464	5%	0.2915	Range	2.8817	5%	-0.53538
Mean	1.9872	10%	0.4015	Mean	0.09424	10%	-0.39641
Variance	8.7989	25% (Q1)	0.71075	Variance	0.15731	25% (Q1)	-0.14833
Std. Deviation	2.9663	50% (Median)	1.224	Std. Deviation	0.39662	50% (Median)	0.08778
Coef. of Variation	1.4927	75% (Q3)	2.058	Coef. of Variation	4.2088	75% (Q3)	0.31344
Std. Error	0.173	90%	3.861	Std. Error	0.02313	90%	0.5867
Skewness	5.9652	95%	6.1847	Skewness	0.23731	95%	0.79132
Excess Kurtosis	50.074	Max	33.508	Excess Kurtosis	1.0215	Max	1.5251

The distributions best fitted to the JIF(Physics)/$\log_{10}$(JIF(Physics)) data are as follows.

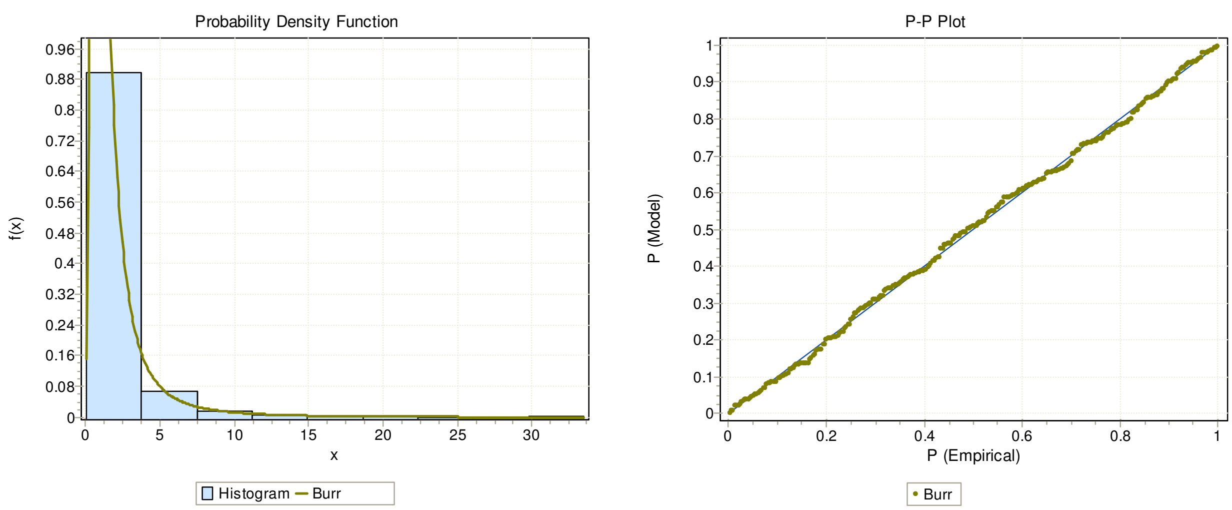
i. **The natural Scale JIF Data (Physics Group)**: Three best fit distributions to the natural scale JIF data (for the year 2006) are: (a) Burr 3p/4p, (b) Log Logistic 3p/2p, and (c) Dagum 3p/Gen. Extreme Value. Overall, the fitness of Dagum 4p may not be considered better than that of Dagum 3p. The details are given in Table 5.2.

Table-5.2: Estimated Parameters and Goodness of Fit Statistics for Natural Scale JIF Data (Physics Group)

Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Burr 3p	k=0.79698; α=2.1745; β=1.0359	0.02633 (1) [0.98376]	0.18774 (2)	4.5969 (4) [0.79966]
Burr 4p	k=0.8159; α=2.1356; β=1.0441; γ=0.00746	0.02674 (2) [0.98091]	0.18691 (1)	4.6066 (5) [0.79868]
Log Logistic 3p	α=1.9425; β=1.1938; γ=0.02213	0.02721 (3) [0.97735]	0.21815 (5)	4.0085 (2) [0.85636]
Log Logistic 2p	α=1.9692; β=1.2284	0.03051 (7) [0.93935]	0.26653 (7)	3.8246 (1) [0.87259]
Dagum 3p	k=1.2813; α=1.8427; β=1.0061	0.02895 (5) [0.9602]	0.20395 (3)	4.9664 (7) [0.76117]
Gen. Extreme Value	k=0.49476; σ=0.70743; μ=0.90853	0.02951 (6) [0.95327]	0.26351 (6)	4.3594 (3) [0.82333]

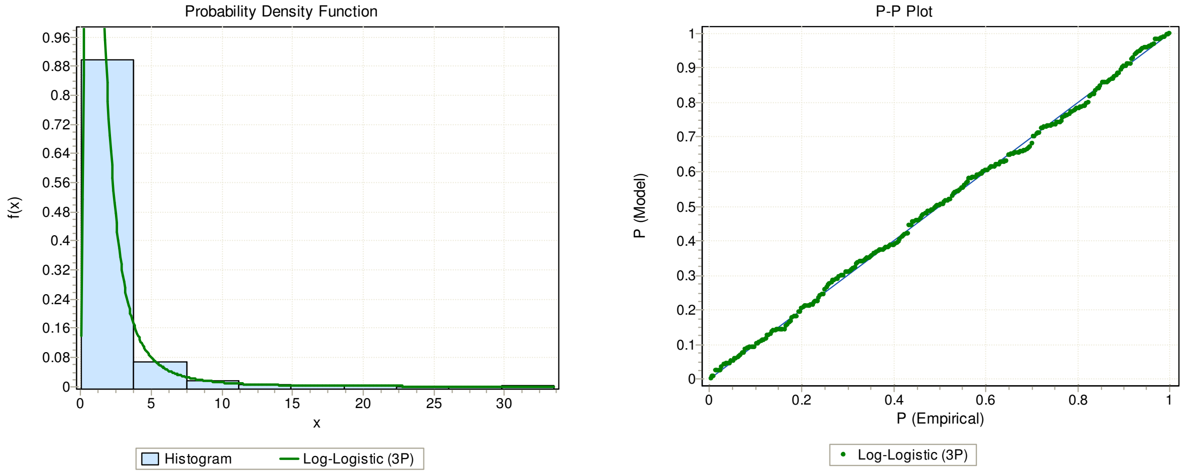
..

Fig.5.1: Histogram, pdf and P-P plot of Burr 3p Distribution fitted to Natural Scale JIF Data (Physics Group)



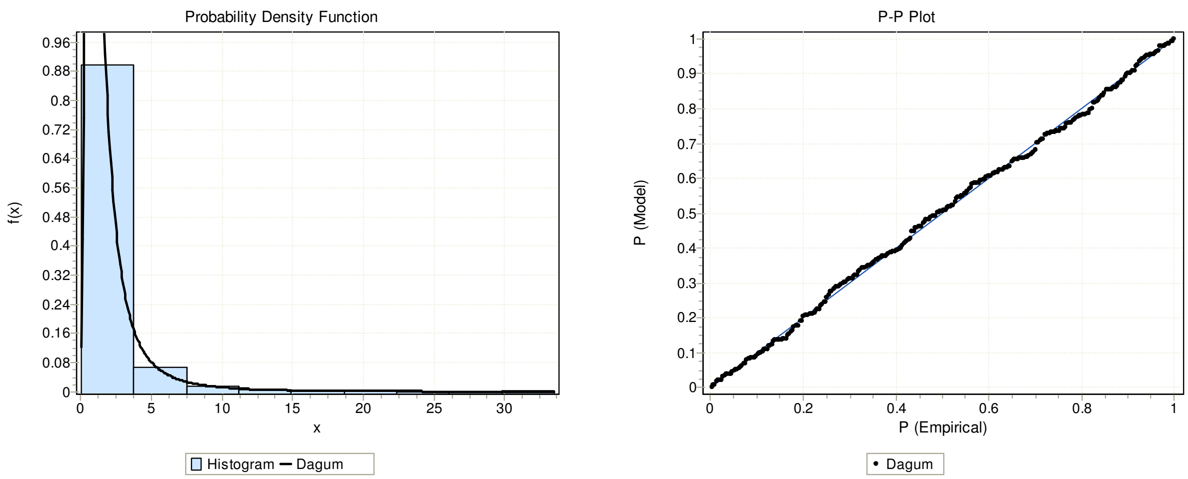
..

Fig.5.2: Histogram, pdf and P-P plot of Log Logistic 3p Distribution fitted to Natural Scale JIF Data (Physics Group)



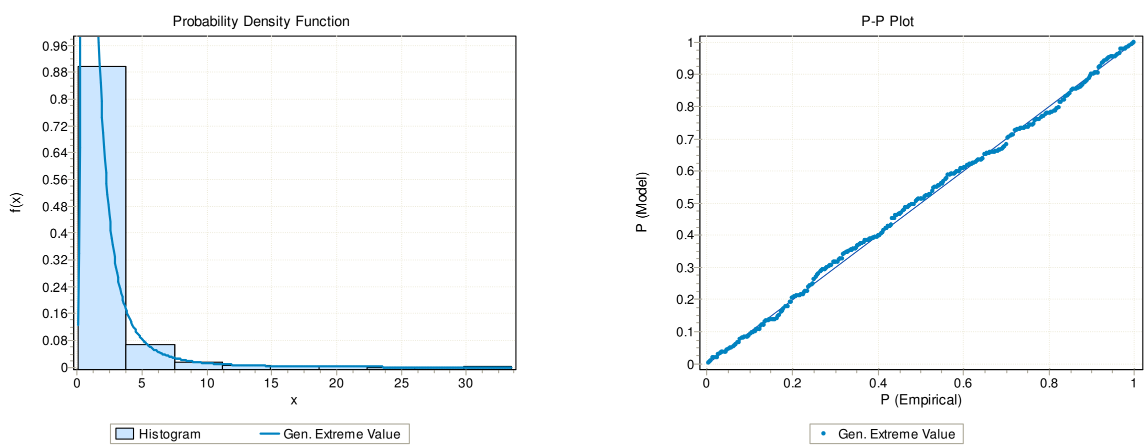
..

Fig.5.3: Histogram, pdf and P-P plot of Dagum 3p Distribution fitted to Natural Scale JIF Data (Physics Group)



..

Fig.5.4: Histogram, pdf and P-P plot of Gen. Extreme Value Distribution fitted to Natural Scale JIF Data (Physics Group)



..

ii. **The Logarithmic JIF Data (Physics Group)**: Three best fit distributions to the $\log_{10}$(JIF) data (for the year 2006) are: (a) Burr 4p, (b) Log Logistic, and (c) Dagum 4p/Generalized Normal (Error)/Hypersecant. The details are given in Table 5.3.

Table-5.3: Estimated Parameters and Goodness of Fit Statistics for Log_{10}(JIF) Data (Physics Group)

Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Burr 4p	k=0.8013; α=1116.0; β=223.34; γ=-223.32	0.02636 (1) [0.98351]	0.18746 (2)	4.5977 (7) [0.79958]
Log Logistic 3p	α=29.614; β=6.4432; γ=-6.3615	0.02665 (2) [0.98157]	0.1819 (1)	3.8772 (3) [0.86803]
Dagum 4p	k=1.1242; α=53.367; β=12.038; γ=-11.993	0.02769 (3) [0.97327]	0.19146 (3)	4.2999 (4) [0.8291]
Error (Gen. Normal)	k=1.3987; σ=0.39662; μ=0.09424	0.03319 (4) [0.89153]	0.29158 (5)	3.0799 (1) [0.92925]
Hypersecant	σ=0.39662; μ=0.09424	0.03632 (7) [0.81899]	0.38035 (7)	3.1581 (2) [0.92405]

..

Fig.5.5: Histogram, pdf and P-P Burr 4p Distribution fitted to Log_{10}(JIF) Data (Physics Group)

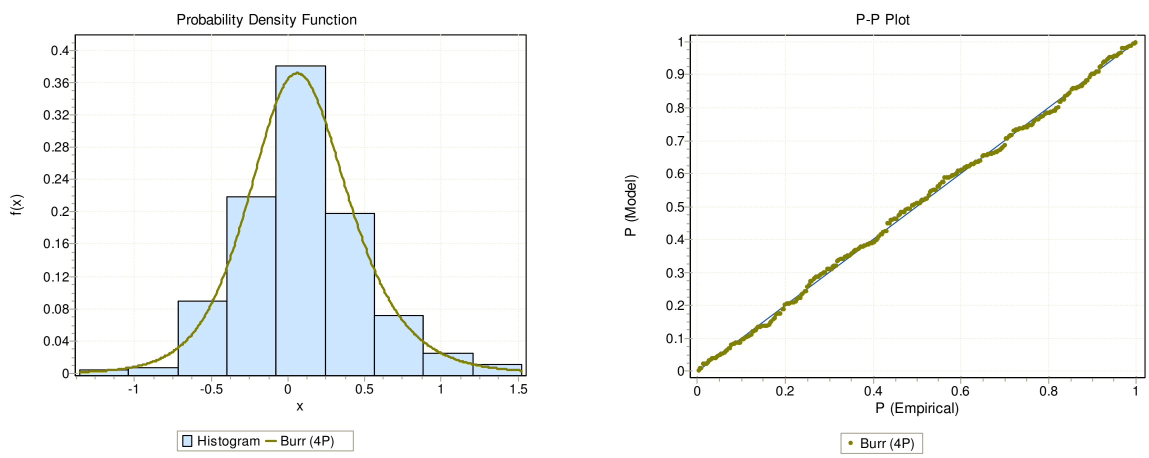
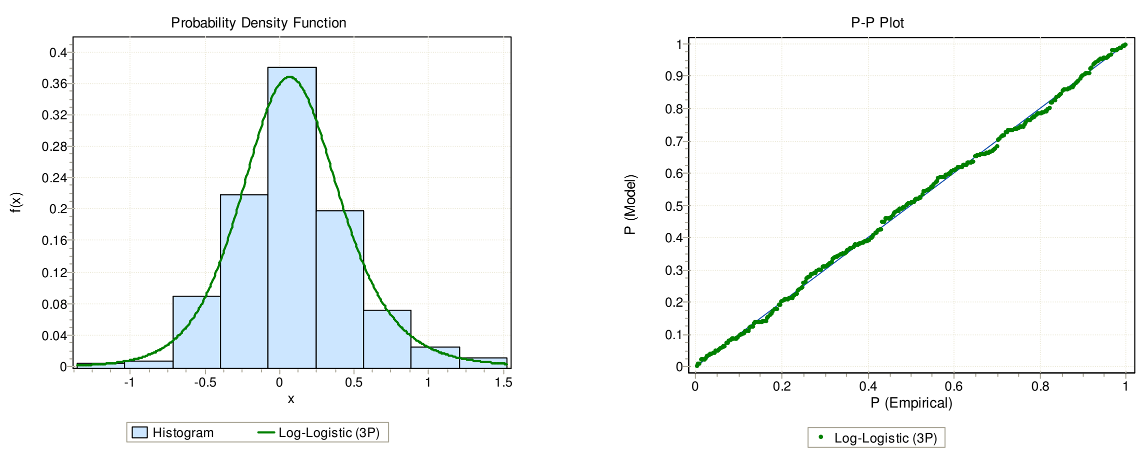
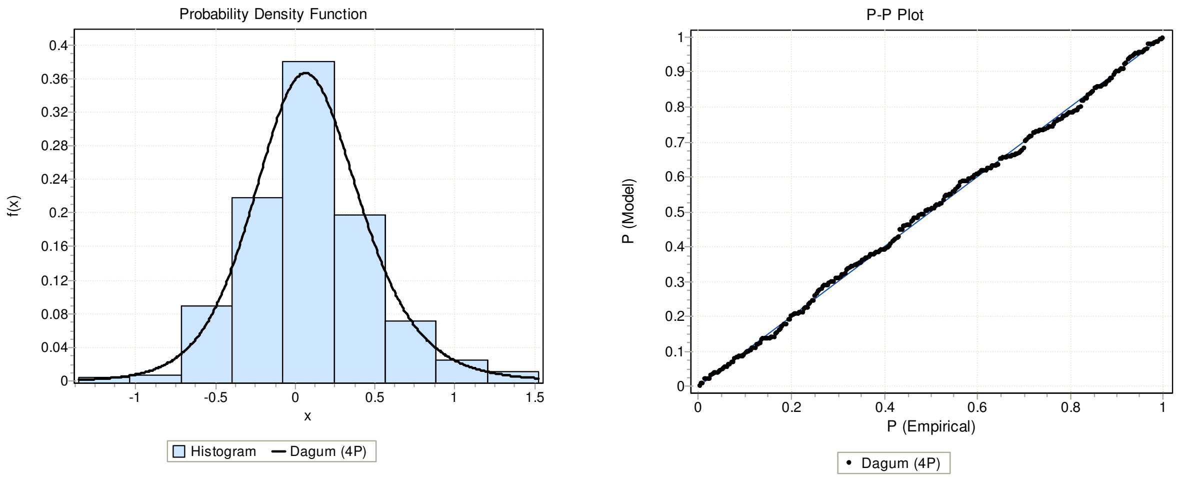


Fig.5.6: Histogram, pdf and P-P Log Logistic 3p Distribution fitted to Log_{10}(JIF) Data (Physics Group)



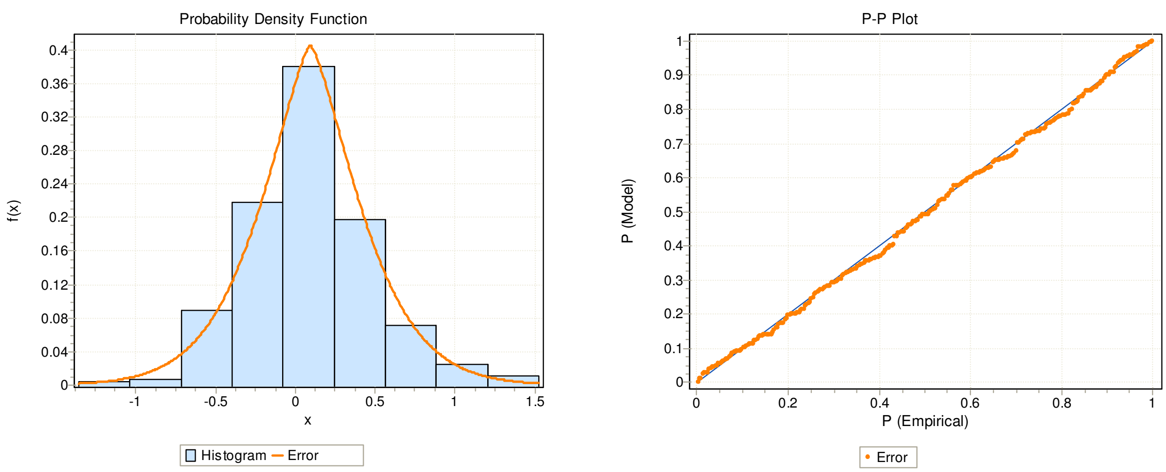
..

Fig.5.7: Histogram, pdf and P-P Dagum 4p Distribution fitted to Log_{10}(JIF) Data (Physics Group)



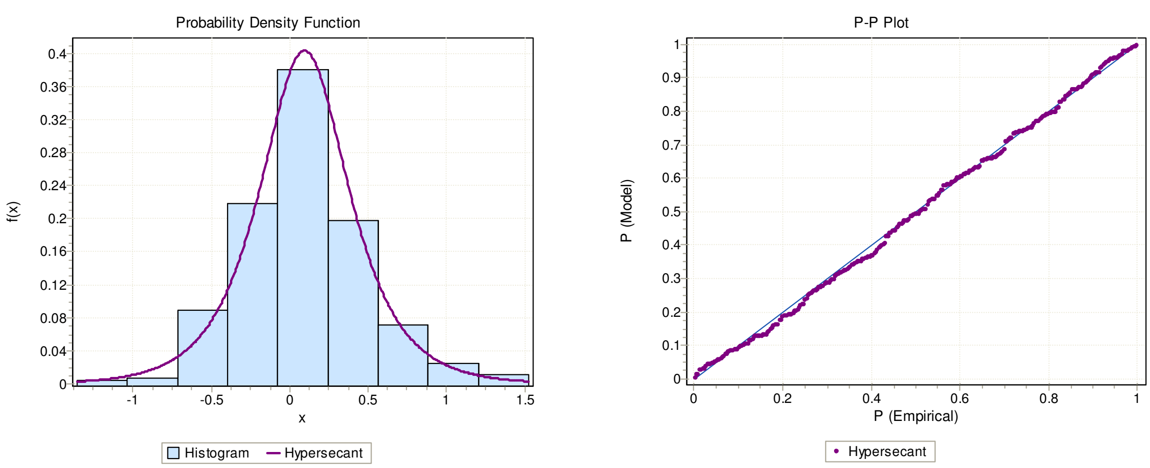
..

Fig.5.8: Histogram, pdf and P-P Gen. Normal (Error) Distribution fitted to Log_{10}(JIF) Data (Physics Group)



..

Fig.5.9: Histogram, pdf and P-P Hypersecant Distribution fitted to Log_{10}(JIF) Data (Physics Group)



..

8. Psychology Group: In this group of disciplines we have JIF values for 421 journals (year 2002). The frequency distribution of JIF as well as $\log_{10}$(JIF) is skewed and with positive excess kurtosis (leptokurtic), indicating sharper peak and longer, fatter tails. Descriptive statistics for JIF(Psychology) and $\log_{10}$(JIF(Psychology)) are presented in Table-6.1

Table-6.1: Descriptive Statistics regarding the Journal Impact Factor for the Psychology Group (year 2002)							
For the Natural Value of Journal Impact Factor				**For the Common Log Value of Journal Impact Factor**			
Statistic	**Value**	**Percentile**	**Value**	**Statistic**	**Value**	**Percentile**	**Value**
Sample Size	421	Min	0.031	Sample Size	421	Min	-1.5086
Range	8.699	5%	0.2042	Range	2.4497	5%	-0.68995
Mean	1.1966	10%	0.2582	Mean	-0.08117	10%	-0.58804
Variance	1.4284	25% (Q1)	0.494	Variance	0.14698	25% (Q1)	-0.30631
Std. Deviation	1.1952	50% (Median)	0.86	Std. Deviation	0.38338	50% (Median)	-0.0655
Coef. of Variation	0.99876	75% (Q3)	1.5445	Coef. of Variation	-4.7229	75% (Q3)	0.18879
Std. Error	0.05825	90%	2.3686	Std. Error	0.01868	90%	0.37448
Skewness	3.0556	95%	3.2216	Skewness	-0.31512	95%	0.50807
Excess Kurtosis	12.528	Max	8.73	Excess Kurtosis	0.69005	Max	0.94101

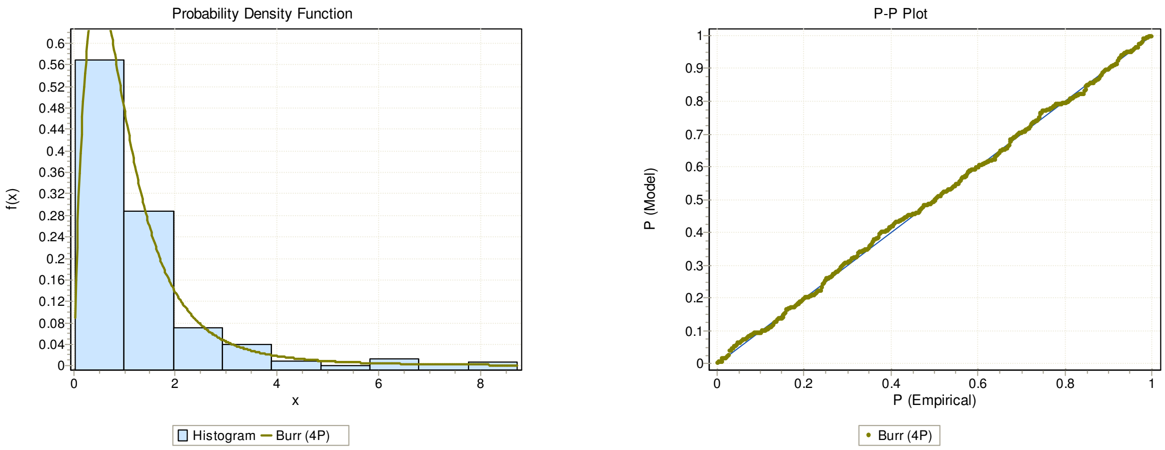
The distributions best fitted to the JIF(Psychology)/$\log_{10}$(JIF(Psychology)) data are as follows.

i. **The natural Scale JIF Data (Psychology Group)**: Three best fit distributions to the natural scale JIF data (for the year 2002) are: (a) Burr 3p/4p, (b) Dagum 3p/4p, and (c) Gen. Extreme Value/Gen. Gamma 4p/Log Pearson-III. Overall, the degrees of goodness of fit on KS/AD and Chi-Square criteria run opposite to each other leading to difficulties in judgment. The results are presented in Table 6.2.

Table-6.2: Estimated Parameters and Goodness of Fit Statistics for Natural Scale JIF Data (Psychology Group)				
Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Burr 4p	k=1.6637; α=1.7059; β=1.2491; γ=0.01753	0.02214 (1) [0.98319]	0.2325 (3)	6.2055 (8) [0.62422]
Burr 3p	k=1.4808; α=1.8131; β=1.151	0.02256 (2) [0.97966]	0.23109 (1)	5.8232 (7) [0.66703]
Dagum 4p	k=0.62425; α=2.3656; β=1.1435; γ=0.02459	0.02479 (3) [0.95235]	0.232 (2)	7.3641 (11) [0.4979]
Dagum 3p	k=0.73631; α=2.3; β=1.053	0.02703 (4) [0.90968]	0.26516 (4)	9.1872 (13) [0.32675]
Gen. Extreme Value	k=0.31309; σ=0.52419; μ=0.66211	0.02798 (6) [0.88732]	0.34545 (7)	3.2426 (1) [0.91822]
Gen. Gamma 4p	k=0.35379; α=10.941; β=0.0011; γ=-0.00151	0.03556 (11) [0.64814]	0.51441 (11)	3.506 (2) [0.89873]
Log Pearson III	α=40.281; β=-0.13909; γ=5.4158	0.0381 (12) [0.56099]	0.61135 (12)	4.6246 (3) [0.79684]

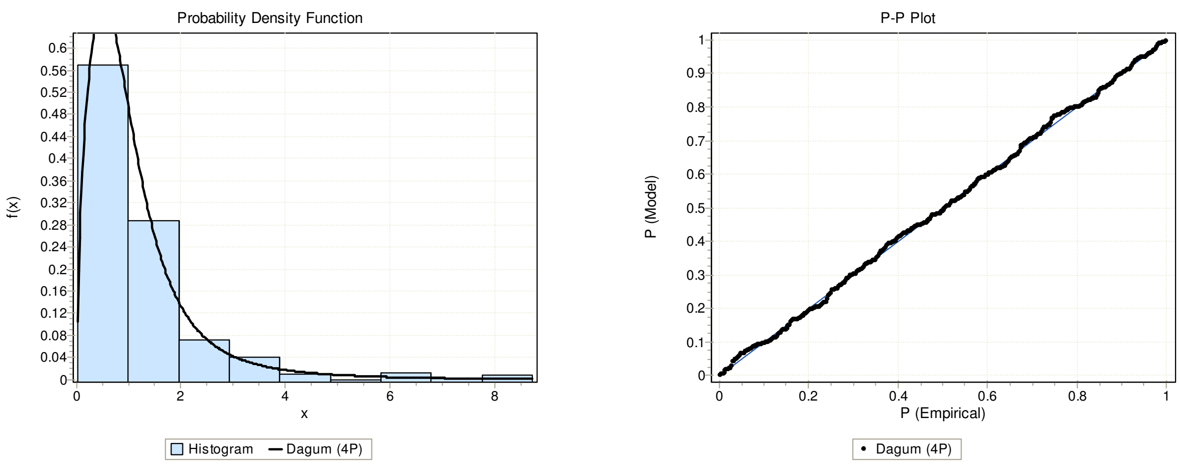
..

Fig.6.1: Histogram, pdf and P-P plot of Burr 4p Distribution fitted to Natural Scale JIF Data (Psychology Group)



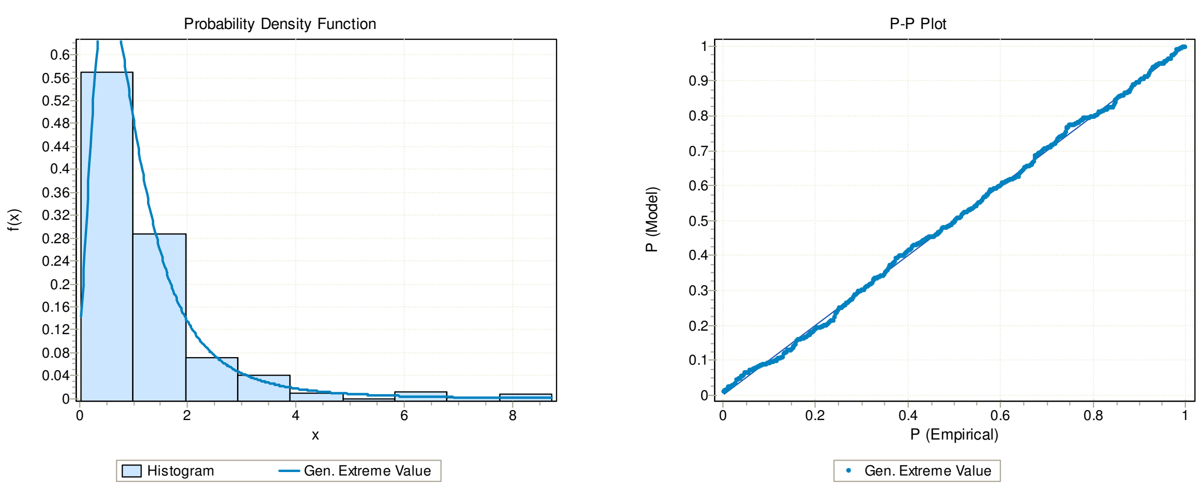
..

Fig.6.2: Histogram, pdf and P-P plot of Burr 4p Distribution fitted to Natural Scale JIF Data (Psychology Group)



..

Fig.6.3: Histogram, pdf and P-P plot of Gen. Extr. Value Distribution fitted to Natural Scale JIF Data (Psychology Group)



...

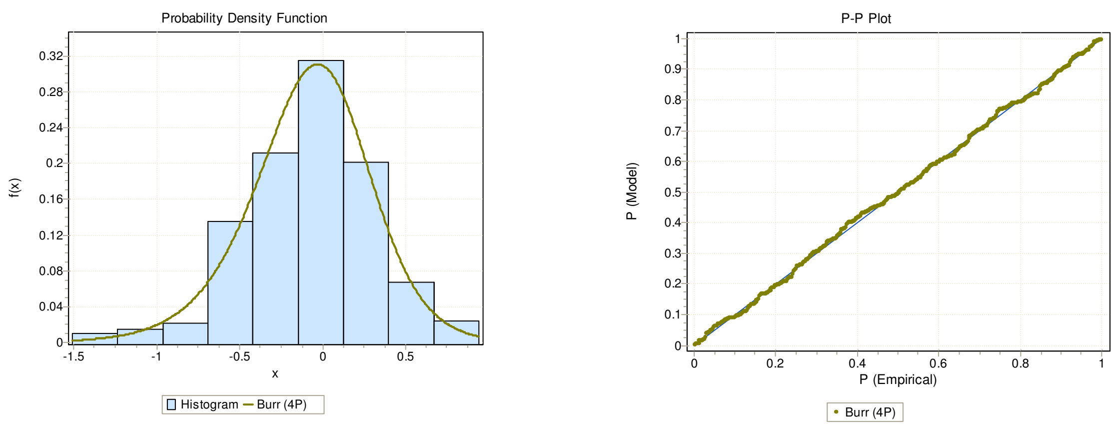
ii. **The Logarithmic JIF Data (Psychology Group)**: The best fit distributions to the $\log_{10}$(JIF) data (2002) are: (a) Burr 4p/Johnson SU, and (b) Dagum 4p. The Beta and the Kumaraswamy distributions are rejected as the lower /upper limits of $\log_{10}$(JIF) are not fixed. The details are given in Table 6.3.

Table-6.3: Estimated Parameters and Goodness of Fit Statistics for Log_{10}(JIF) Data (Psychology Group)

Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Burr 4p	k=1.9136; α=27.104; β=7.1369; γ=-6.9849	0.02186 (1) [0.9853]	0.23045 (1)	6.3578 (4) [0.60722]
Dagum 4p	k=0.46399; α=18.366 β=3.0058; γ=-2.8674	0.02567 (2) [0.93752]	0.23166 (2)	8.9895 (6) [0.34318]
Johnson SU	γ=0.83227; δ=2.9398; λ=1.0191; ξ=0.22862	0.02708 (3) [0.90857]	0.28202 (3)	4.4264 (1) [0.81675]
Beta	α_1=6086.9; α_2=74.665; a=-270.85; b=3.2408	0.03706 (6) [0.59652]	0.58791 (5)	5.9633 (2) [0.65134]
Kumaraswamy	α_1=5.409; α_2=343.8; a=-1.9247; b=3.9452	0.04277 (8) [0.41305]	1.0162 (12)	6.1567 (3) [0.62968]

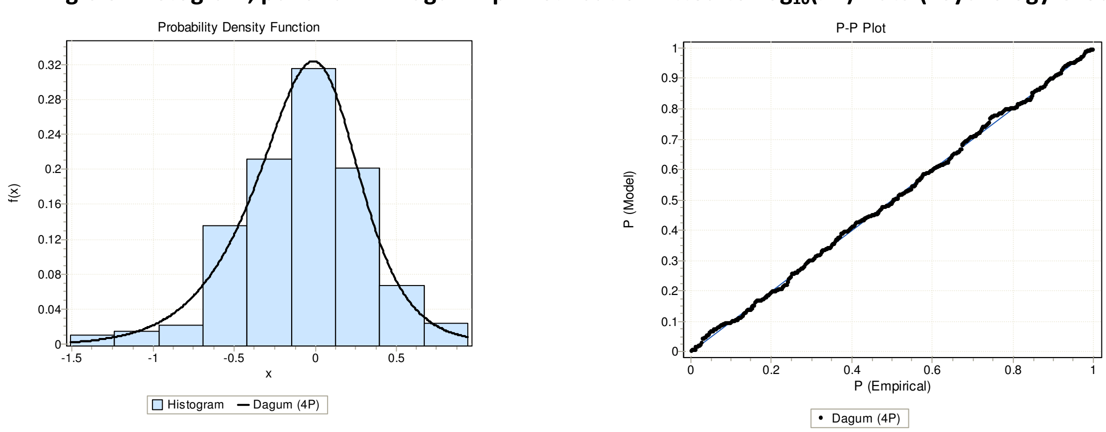
..

Fig.6.4: Histogram, pdf and P-P Burr 4p Distribution fitted to Log_{10}(JIF) Data (Psychology Group)



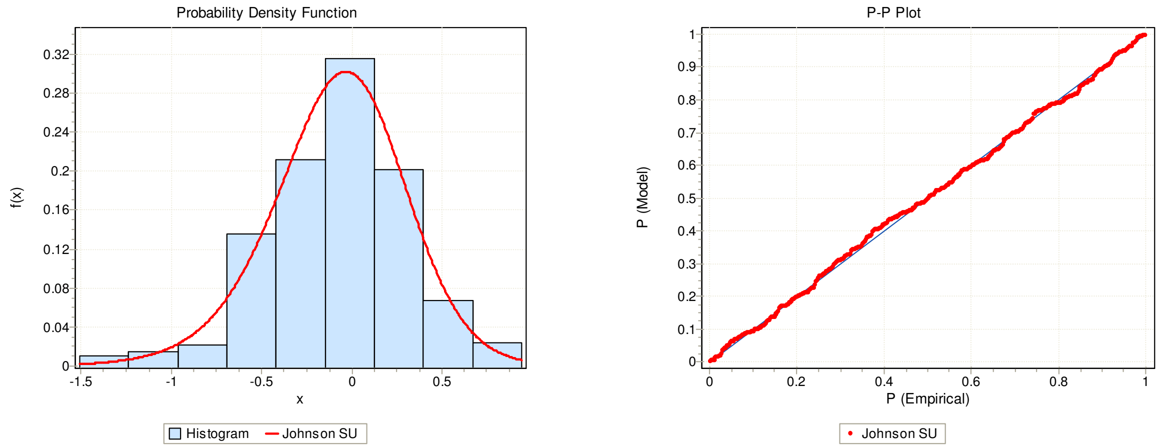
..

Fig.6.5: Histogram, pdf and P-P Dagum 4p Distribution fitted to Log_{10}(JIF) Data (Psychology Group)



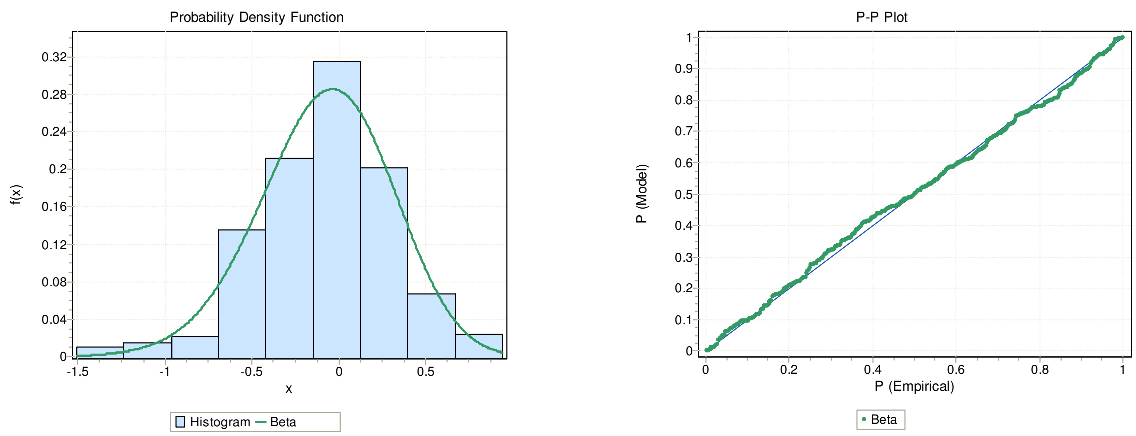
..

Fig.6.6: Histogram, pdf and P-P Johnson SU Distribution fitted to Log_{10}(JIF) Data (Psychology Group)



..

Fig.6.7: Histogram, pdf and P-P Beta Distribution fitted to Log_{10}(JIF) Data (Psychology Group)



..

9. Social Sciences Group: In this group of disciplines we have JIF values for 1301 journals (year 2002). The frequency distribution of JIF as well as $\log_{10}$(JIF) is skewed and with positive excess kurtosis (leptokurtic), indicating sharper peak and longer, fatter tails. Descriptive statistics for JIF(Social Sc) and $\log_{10}$(JIF(Social Sc)) are presented in Table-7.1

Table-7.1: Descriptive Statistics regarding the Journal Impact Factor for the Social Sc. Group (year 2002)							
For the Natural Value of Journal Impact Factor				**For the Common Log Value of Journal Impact Factor**			
Statistic	**Value**	**Percentile**	**Value**	**Statistic**	**Value**	**Percentile**	**Value**
Sample Size	1301	Min	0.011	Sample Size	1301	Min	-1.9586
Range	11.611	5%	0.1034	Range	3.0239	5%	-0.98551
Mean	0.88949	10%	0.1812	Mean	-0.23111	10%	-0.74184
Variance	0.92394	25% (Q1)	0.349	Variance	0.17336	25% (Q1)	-0.45717
Std. Deviation	0.96122	50% (Median)	0.621	Std. Deviation	0.41636	50% (Median)	-0.20691
Coef. of Variation	1.0806	75% (Q3)	1.0835	Coef. of Variation	-1.8016	75% (Q3)	0.03483
Std. Error	0.02665	90%	1.808	Std. Error	0.01154	90%	0.2572
Skewness	3.9466	95%	2.5423	Skewness	-0.49189	95%	0.40523
Excess Kurtosis	25.748	Max	11.622	Excess Kurtosis	1.0088	Max	1.0653

The distributions best fitted to the JIF(Social Sc)/$\log_{10}$(JIF(Social Sc)) data are as follows.

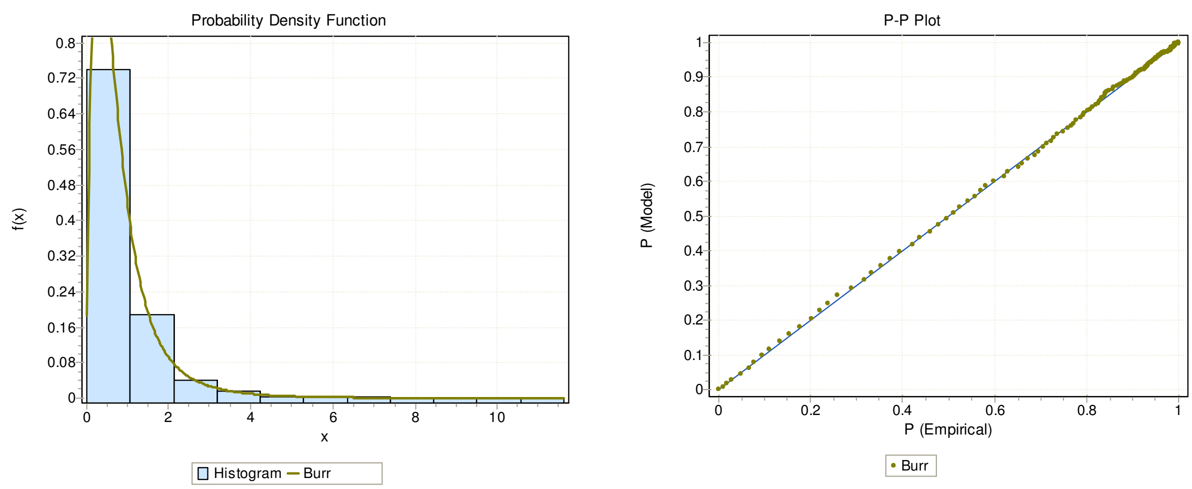
i. **The natural Scale JIF Data (Social Sc Group)**: Three best fit distributions to the natural scale JIF data (for the year 2002) are: (a) Burr 3p/4p, (b) Dagum 3p/4p, and (c) Gen. Extreme Value/Gen. Gamma 4p/Log Pearson-III. The details are presented in Table 7.2. Overall, the degrees of goodness of fit on KS/AD and Chi-Square criteria run opposite to each other leading to difficulties in judgment.

Table-7.2: Estimated Parameters and Goodness of Fit Statistics for Natural Scale JIF Data (Social Sc Group)

Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Burr 3p	k=1.7319; α=1.6253; β=0.96771	0.01503 (1) [0.92602]	0.2542 (2)	5.1155 (6) [0.88333]
Burr 4p	k=1.8533; α=1.5759; β=1.024; γ=0.0055	0.01709 (2) [0.83535]	0.323 (4)	5.7574 (7) [0.83522]
Gen. Extreme Value	k=0.32704; σ=0.39771; μ=0.47236	0.01789 (3) [0.79206]	0.53709 (5)	3.9649 (2) 0.94892
Dagum 3p	k=0.65549; α=2.2568; β=0.82956	0.01844 (6) [0.76089]	0.24649 (1)	3.4381 (1) [0.96916]
Dagum 4p	k=0.5919; α=2.3023; β=0.87711; γ=0.00944	0.01844 (5) [0.76095]	0.29016 (3)	5.0364 (5) [0.88873]
Log Logistic 3p	α=2.0082; β=0.63673; γ=-0.02315	0.01835 (4) [0.76611]	0.71091 (10)	4.5113 (3) [0.92135]

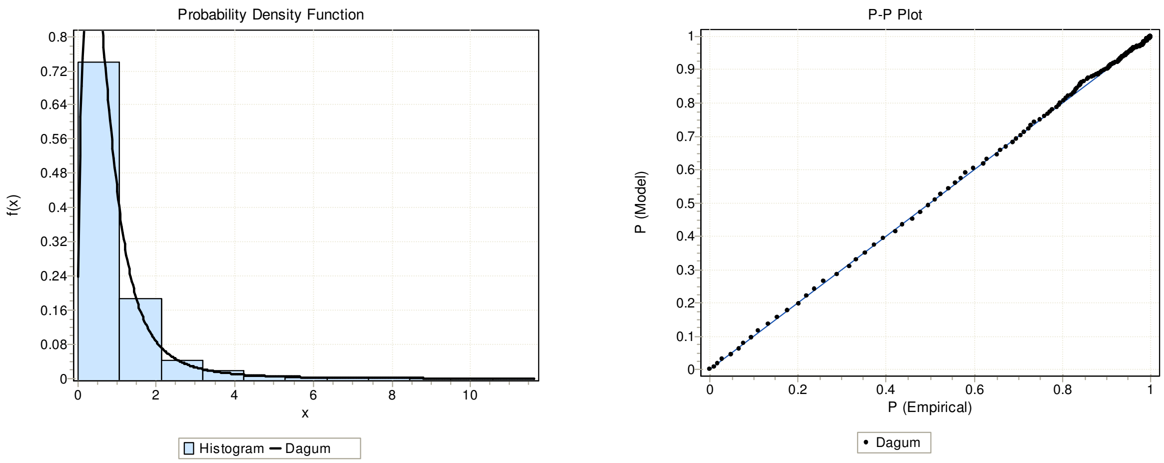
..

Fig.7.1: Histogram, pdf and P-P plot of Burr 3p Distribution fitted to Natural Scale JIF Data (Social Sc Group)



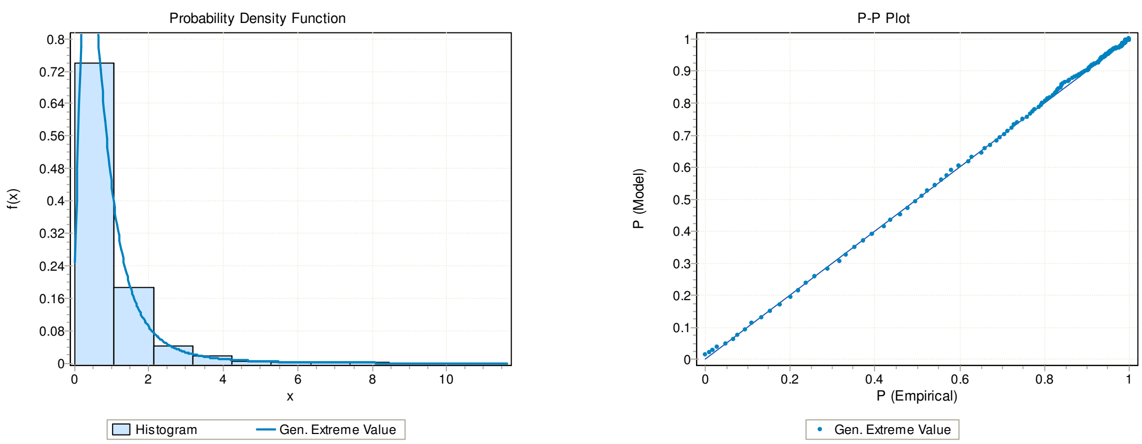
..

Fig.7.2: Histogram, pdf and P-P plot of Dagum 3p Distribution fitted to Natural Scale JIF Data (Social Sc Group)



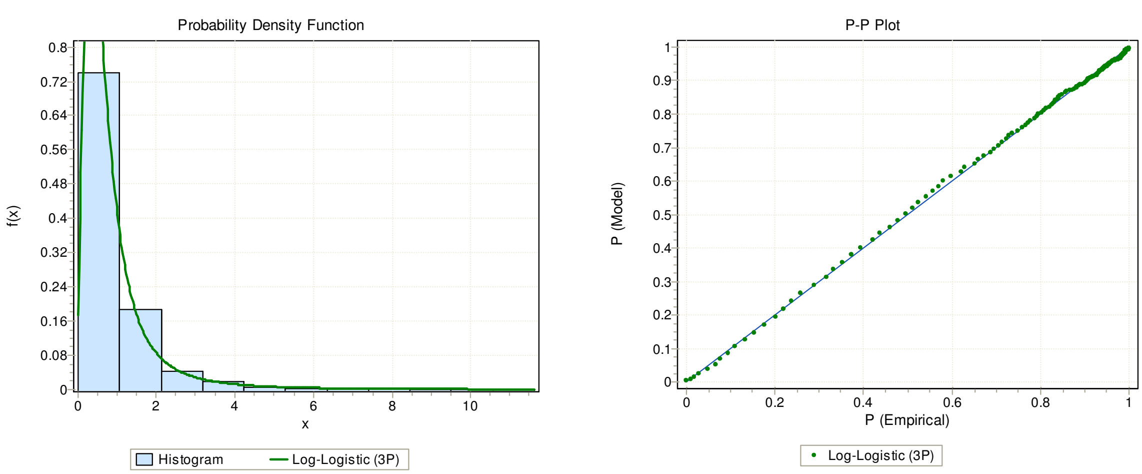
..

Fig.7.3: Histogram, pdf and P-P plot of Gen Extr. Value Distribution fitted to Natural Scale JIF Data (Social Sc Group)



...

Fig.7.4: Histogram, pdf and P-P plot of Log Logistic 3p Distribution fitted to Natural Scale JIF Data (Social Sc Group)



..

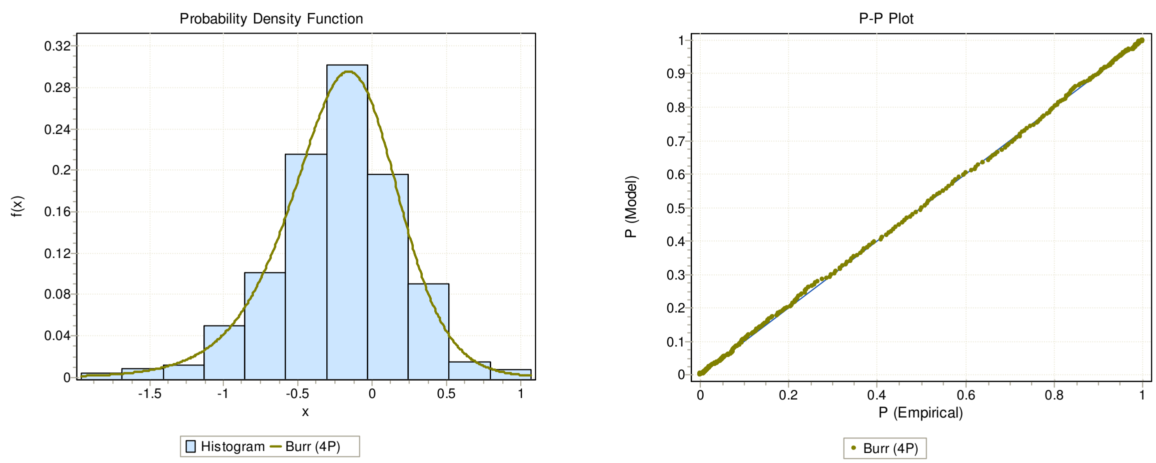
ii. **The Logarithmic JIF Data (Social Sc Group)**: The best fit distributions to the $\log_{10}$(JIF) data (for the year 2002) are: (a) Burr 4p/Johnson SU, and (b) Dagum 4p/Beta. The Kumaraswamy distribution fits well to the data but it may not be acceptable on theoretical grounds. The details are given in Table 7.3.

Table-7.3: Estimated Parameters and Goodness of Fit Statistics for Log_{10}(JIF) Data (Social Sc Group)

Best Fit Distribution	Estimated Parameters	Goodness of Fit Statistic for the Distribution		
		KS (rank) [prob]	AD(rank)[prob]	χ^2 (rank)[prob]
Burr 4p	k=1.7536; α=83117.0; β=22312.0; γ=-22312.0	0.01554 (1) [0.90685]	0.26447 (2)	5.8631 (2) [0.82663]
Dagum 4p	k=0.64935; α=8.1316E+5; β=1.5561E+5; γ=-1.5561E+5	0.01895 (2) [0.73121]	0.25938 (1)	3.5748 (1) [0.9645]
Johnson SU	γ=1.1144; δ=2.7065; λ=0.96272; ξ=0.20539	0.02051 (3) [0.63673]	0.52945 (3)	7.5805 (3) [0.66974]

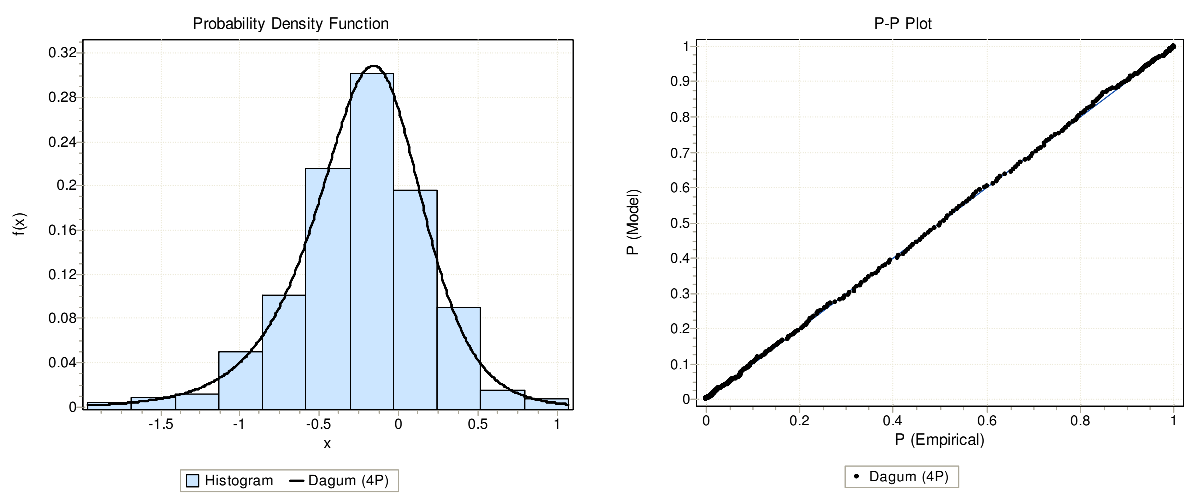
..

Fig.7.5: Histogram, pdf and P-P Burr 4p Distribution fitted to Log_{10}(JIF) Data (Social Sc Group)



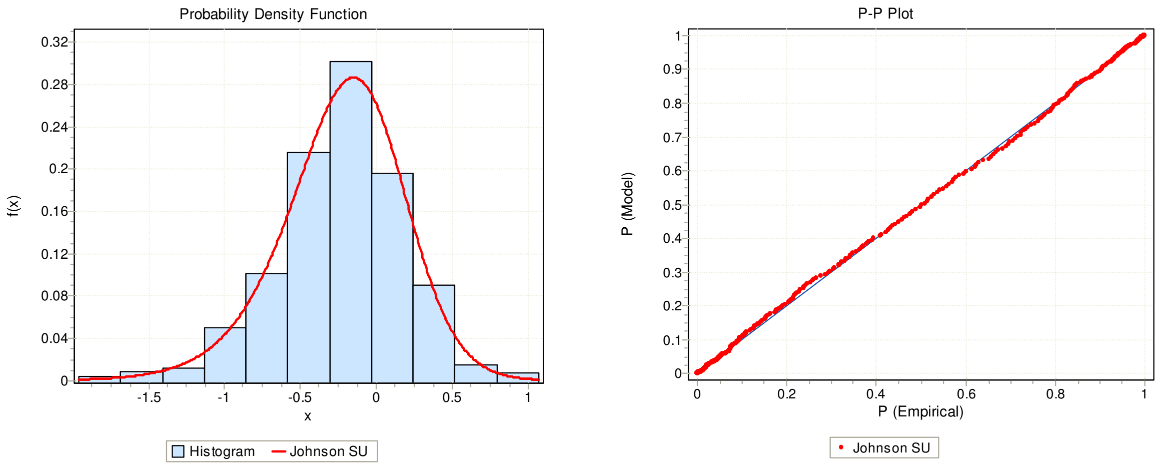
..

Fig.7.6: Histogram, pdf and P-P Dagum 4p Distribution fitted to Log_{10}(JIF) Data (Social Sc Group)



..

Fig.7.7: Histogram, pdf and P-P Johnson SU Distribution fitted to Log_{10}(JIF) Data (Social Sc Group)



10. Some Observations: In the past, researchers have hypothesized various types of statistical distributions underlying the generation mechanism of journal impact factors. These are: negative exponential (Brookes, 1970), combination of exponentials (Avramescu, 1979), Poisson (Brown, 1980), generalized inverse Gaussian-Poisson (Sichel, 1985; Burrell and Fenton, 1993), lognormal (Matricciani, 1991; Egghe and Rao, 1992), Weibull (Hurt and Budd, 1992; Rousseau and West-Vlaanderen, 1993), gamma (Sahoo and Rao, 2006), negative binomial (Bensman, 2008), approximately normal (Stringer et al., 2008), normal (Egghe, 2009), generalized Waring (Glänzel, 2009; see Panaretos and Xekalaki, 1986; Irwin, 1975), Pearson's type IV (Mishra, 2009), etc. However, in the present study, we have frequently encountered Burr-XII, inverse Burr-III (Dagum), Johnson SU, and a few other distributions closely related to Burr distribution to best fit the JIF data.

Tadikamalla (1980) gives a comprehensive idea about the Burr (types XII, III and II) and the related distributions such as Lomax, exponential-gamma (Dubey, 1966, 1970), compound Weibull, Weibull, logistic, log-logistic, and 2p-kappa family of distributions and concludes that the Burr type III and type XII distributions can be used to fit almost any unimodal data and are comparable to the Pearson and the Johnson systems of distributions. Moreover, they have the added advantage in having their inverse distribution function in simple closed forms. It is pertinent to note that the major characteristics of JIF data lay in the asymmetry and non-mesokurticity. Burr distributions take care of them very well. ***However, it may be noted that no theoretical distribution fits so well to the JIF data in the biology group of disciplines. It may be conjectured that either this group has a mixed distribution or some sort of theoretical distribution that was not included in list of 'the likely distributions' considered by us.***

All said and done, a search for a probability distribution underlying the mechanism of generation of JIF data is based on the presumption that only the search-quality-cite factors determine the JIF data pattern. On the other hand, in view of the findings of Rossner et al. (2007), a search for the generation mechanism and the underlying probability distribution of JIF data published or provided by Thomson Scientific would be of no avail. To quote Rossner et al. (2007),

> *"It became clear that Thomson Scientific could not or (for some as yet unexplained reason) would not sell us the data used to calculate their published impact factor. If an author is unable*

to produce original data to verify a figure in one of our papers, we revoke the acceptance of the paper. We hope this account will convince some scientists and funding organizations to revoke their acceptance of impact factors as an accurate representation of the quality - or impact - of a paper published in a given journal. Just as scientists would not accept the findings in a scientific paper without seeing the primary data, so should they not rely on Thomson Scientific's impact factor, which is based on hidden data. As more publication and citation data become available to the public through services like PubMed, PubMed Central, and Google Scholar®, we hope that people will begin to develop their own metrics for assessing scientific quality rather than rely on an ill-defined and manifestly unscientific number. "

A persistent lack of transparency may easily rouse a question as to integrity. This is loudly resonant in Rossner et al. (2007). If that is the fact too, and if the JIF data is generated or partially influenced by any mechanism other than 'search-quality-cite mechanism', this is certainly not a healthy state of affairs. Once the journal impact factor is biased or mutilated, it gathers forces to further accentuation it due to the Mathew effect (Larivière and Gingras, 2009). Such biases can make an average journal (or author publishing therein) a super-average journal (or author) and vice versa. It is to be noted that the journal impact factor has a strong influence on the scientific community, affecting decisions on where to publish, whom to promote or hire, how much to pay and how much to finance the research projects proposed by the scientists.

References

1. Avramescu, A. (1979) "Actuality and Obsolescence of Scientific Literature", *Journal of the American Society for Information Science*, 30(5): 296-303.
2. Bensman, S. J. (2008) "Distributional Differences of the Impact Factor in the Sciences Versus the Social Sciences: An Analysis of the Probabilistic Structure of the 2005 Journal Citation Reports", Journal of the American Society for Information Science and Technology, 59(9): 1366–1382.
3. Brookes, B. C. (1970) "The Growth, Utility, and Obsolescence of Scientific Periodical Literature", *Journal of Documentation*, 26(4): 283-294.
4. Brown, P. (1980) "The Half-life of Chemical Literature", *Journal of the American Society for Information Science*, 31(1): 61-63.
5. Burrell, Q. and Fenton, M. R. (1993) "Yes, the GIGP Really Does Work – and is Workable", *Journal of the American Society for Information Science*, 44(2): 61-69.
6. Chakravarti, I. M., Laha, R. G. and Roy, J. (1967) *Handbook of Methods of Applied Statistics, Volume I*, John Wiley and Sons, NY.
7. Dubey, S. D. (1966) "Characterization Theorems for Several Distributions and their Applications", *Journal of Industrial Mathematical Society*, 16(1): 1-12.
8. Dubey, S. D. (1970) "Compound Gamma, Beta and F Distributions", *Metrica*, 16(1): 27-31.
9. Egghe, L. (2009) "Mathematical Derivation of the Impact Factor Distribution", *Journal of Informetrics*, 3(4): 290-295.
10. Egghe, L. and Rao, I. K. (1992) "Citation Age of Data and the Obsolescence Function: Fits and Explanations", *Information Processing and Management*, 28(2): 201-217.
11. Glänzel, W. (2009) "The Multi-dimensionality of Journal Impact", *Scientometrics*, 78(2): 355-374.
12. Hurt, C. D. and Budd, J. M. (1992) "Modeling the Literature of Superstring Theory: A Case of Fast Literature", *Scientometrics*, 24(3): 471-480.

13. Irwin, J. O. (1975) The Generalized Waring Distribution. Part I, *Journal of the Royal Statistical Society. Series A (General)*, 138: 18–31.
14. Kleiber, C. and Kotz, S. (2003) *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley, Hoboken, NJ.
15. Larivière, V. and Gingras, Y. (2009) "The Impact factor's Matthew Effect: A Natural Experiment in Bibliometrics", http://www.ost.uqam.ca/Portals/0/docs/articles/2009/MathhewsEffect.pdf
16. Mansilla, R., Köppen, E., Cocho, G., & Miramontes, P. (2007) "On the Behavior of Journal Impact Factor Rank-order Distribution", *Journal of Informetrics*, *1*(2): 155–160.
17. Matricciani, E. (1991) "The Probability Distribution of the Age of References in Engineering Papers", *IEEE Transactions on Professional Communication*, 34(1): 7-12.
18. Mishra, S. K. (2009) "Does the Journal Impact Factor Help Make a Good Indicator of Academic Performance?", Available at *SSRN*: http://ssrn.com/abstract=1485868
19. Panaretos, J. and Xekalaki, E. (1986) "The Stuttering Generalized Waring Distribution", *Statistics & Probability Letters*, 4(1986) 313-318.
20. Rossner, M., Van Epps, H. and Hill, E. (2007) "Show Me the Data", *The Journal of Cell Biology*, 179(6): 1091-1092
21. Rousseau, R. and West-Vlaanderen, K. I. H. (1993) "A Note on Maximum Impact Factors", Available at http://www.cais-acsi.ca/proceedings/1993/Rousseau_1993.pdf
22. Sichel, H.S. (1985) "A Bibliometric Distribution which Really Works", *Journal of the American Society for Information Science*, 36: 314-321.
23. Sahoo, B. B. and Rao, I. K. R. (2006) "A Distribution of Impact Factors of Journals in the Area of Software: An Empirical Study", Information Processing & Management, 42(6): 1465-1470.
24. Singh, S. K., and Maddala, G. S. (1976) "A Function for Size Distribution of Incomes", *Econometrica*, 44(5): 963-970.
25. Stephens, M. A. (1974) "EDF Statistics for Goodness of Fit and Some Comparisons", *Journal of the American Statistical Association*, 69 (347): 730-737.
26. Stephens, M. A. (1976) "Asymptotic Results for Goodness-of-Fit Statistics with Unknown Parameters", *Annals of Statistics*, 4(2): 357-369.
27. Stephens, M. A. (1977-a) "Goodness of Fit for the Extreme Value Distribution", *Biometrika*, 64(3): 583-588.
28. Stephens, M. A. (1977-b) "Goodness of Fit with Special Reference to Tests for Exponentiality", *Technical Report No. 262*, Department of Statistics, Stanford University, Stanford, CA.
29. Stephens, M. A. (1979) "Tests of Fit for the Logistic Distribution Based on the Empirical Distribution Function", *Biometrika*, 66(3): 591-595.
30. Stringer, M. J., Sales-Pardo, M. and Amaral, L. A. N. (2008) Effectiveness of Journal Ranking Schemes as a Tool for Locating Information. *PLoS ONE* 3(2): e1683. doi:10.1371/ journal.pone.0001683: 1-8.
31. Tadikamalla, P. R. (1980) "A Look at the Burr and Related Distributions", *International Statistical Review*, 48(3): 337-344.