On the Derivation of the SIA-log-Student-t Distribution for Achieving Better Fit

Ayesha Butt
Student of MPhil Statistics
Kinnaird College for Women
Lahore, Pakistan
e-mail: ayeshabashr@gmail.com

Saleha Naghmi Habibullah
Professor of Statistics
Kinnaird College for Women
Lahore, Pakistan
e-mail: salehahabibullah@gmail.com

***Abstract*— In order to accomplish appropriate decision-making and policy-formulation, data-sets are required on qualitative and quantitative variables. Sets of quantitative data pertaining to business, health, social disciplines and other fields are often modeled by distributions of non-negative continuous random variables such as the lognormal, Weibull and gamma distributions. The log-student-t distribution is also defined on the positive half-line and this distribution possesses the property of being invariant under the reciprocal transformation or Self-Inverse at Unity (SIU). In order to achieve a *generalization* of this distribution, in this paper, we develop the SIA-version of this distribution. We derive the basic properties of the SIA-log-Student-t distribution including its mid-quartile range, inter-quartile range, Bowley's coefficient of skewness and percentile coefficient of kurtosis for various values of the parameter k, one of the two parameters of this distribution. As well, we present the expressions of the reliability function and hazard function of this distribution for various values of k. SIA-distributions possess a property by which it is possible to develop estimators of distribution parameters that are *more efficient* than their non-SIA counterparts which carries implications for achieving a better-fitting model for data on non-negative random variables. As such, we believe that the SIA-log-Student-t distribution developed in this paper (along with its fundamental properties) will form the basis for the development of an SIA-estimator of k that will lead to *more accurate modeling* of data-sets distributed according to the SIA-log-Student-t distribution.**

Keywords: Log-t-distribution, Quantile function, Self-inverse at A, Bowley's coefficient of kurtosis, Percentile coefficient, Mid-Quartile Range, Inter-Quartile range, Cumulative distribution Function.

I. INTRODUCTION

The Student t distribution is a member of a family of continuous probability distributions which was developed by William Sealy Gosset under the pseudonym Student. It arises when the sample size is small and the population standard deviation is unknown and the mean of a normally distributed population is being estimated. It plays a role in a number of widely used statistical analyses such as the construction of confidence intervals for the difference between two population means and the t-test for evaluating the statistical significance of the difference between two sample means. The Student's t-distribution is a special case of the generalized hyperbolic distribution.

The student t-distribution has the following pdf:

$$h(z)=\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}}\left(1+\frac{z^2}{k}\right)^{-\frac{k+1}{2}}, \quad -\infty<z<\infty \qquad (1)$$

II. SIU AND SIA-DISTRIBUTIONS

The SIU log-Student-t distribution has been verified by some of the authors See [1], [2] and [3]. This SIU log-Student-t distribution is described as if we put the transformation 1/x which is the reciprocal of x then the transformation does not effect this distribution and the SIU log-Student-t distribution remain unchanged. After applying the transformation we got the SIU version of log-Student-t-distribution which is:

$$g(y)=\frac{\Gamma\left(\frac{k+1}{2}\right)\left(1+\frac{(\ln y)^2}{k}\right)^{\frac{-(k+1)}{2}}}{y\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}}, 0<y<\infty \qquad (2)$$

It is worthwhile to note that the log-Student-t distribution given in (2) is an SIU distribution. Further it was assessed that the distribution in SIA version which is an enhancement towards the better fitting model See [4]. The assessment of the log-Student-t distribution has motivated us to derive the SIA version of log-Student-t distribution, which makes this more appropriate one.

III. Derivation of the SIA log-Student-t Distribution

In this section, we derive the SIA version of the log-Student-t distribution.

Applying the transformation X=AY, ($A>0$) on the pdf of the log-Student-t distribution given in (1.1), we have:

$$Y=\frac{X}{A}\Rightarrow\frac{dy}{dx}=\frac{1}{A}\Rightarrow\left|\frac{dy}{dx}\right|=\frac{1}{A}\quad ,Since\, A>0$$

So:

$$f(x)=\frac{\Gamma\left(\frac{k+1}{2}\right)\left(1+\frac{\left(ln\frac{x}{A}\right)^2}{k}\right)^{\frac{-(k+1)}{2}}}{\frac{x}{A}\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}}\frac{1}{A}$$

Hence the SIA-log-Student-t distribution is given by

$$f(x)=\frac{\Gamma\left(\frac{k+1}{2}\right)}{x\Gamma\left(\frac{k}{2}\right)\sqrt{(k\pi)}}\left[1+\frac{1}{k}\left\{ln\left(\frac{x}{A}\right)\right\}^2\right]^{\frac{-(k+1)}{2}}$$

where,

$$0<x<\infty \qquad (3)$$

A. *Different PDF's for the different values of A and k=1and 2*

A\k	1	2
0.2	$\frac{\left[1+\left\{\ln\left(\frac{x}{0.2}\right)\right\}^2\right]^{-1}}{x\pi}$	$\frac{\left[1+\frac{1}{2}\left\{\ln\left(\frac{x}{0.2}\right)\right\}^2\right]^{-\frac{3}{2}}}{2x\sqrt{2}}$
0.5	$\frac{\left[1+\left\{\ln\left(\frac{x}{0.5}\right)\right\}^2\right]^{-1}}{x\pi}$	$\frac{\left[1+\frac{1}{2}\left\{\ln\left(\frac{x}{0.5}\right)\right\}^2\right]^{-\frac{3}{2}}}{2x\sqrt{2}}$
1	$\frac{\left[1+\left\{\ln\left(x\right)\right\}^2\right]^{-1}}{x\pi}$	$\frac{\left[1+\frac{1}{2}\left\{\ln\left(x\right)\right\}^2\right]^{-\frac{3}{2}}}{2x\sqrt{2}}$

B. *Different PDF's for the different values of A and k=1and 2*

A\k	1	2
2	$\frac{\left[1+\left\{\ln\left(\frac{x}{2}\right)\right\}^2\right]^{-1}}{x\pi}$	$\frac{\left[1+\frac{1}{2}\left\{\ln\left(\frac{x}{2}\right)\right\}^2\right]^{-\frac{3}{2}}}{2x\sqrt{2}}$
5	$\frac{\left[1+\left\{\ln\left(\frac{x}{5}\right)\right\}^2\right]^{-1}}{x\pi}$	$\frac{\left[1+\frac{1}{2}\left\{\ln\left(\frac{x}{5}\right)\right\}^2\right]^{-\frac{3}{2}}}{2x\sqrt{2}}$

C. *Different PDF's for the different values of A and k=5 and 10*

A\k	5	10
0.2	$\frac{8\left[1+\frac{1}{5}\left\{\ln\left(\frac{x}{0.2}\right)\right\}^2\right]^{-3}}{3x\pi\sqrt{5}}$	$\frac{945\left[1+\frac{1}{10}\left\{\ln\left(\frac{x}{0.2}\right)\right\}^2\right]^{-\frac{11}{2}}}{768x\sqrt{10}}$
0.5	$\frac{8\left[1+\frac{1}{5}\left\{\ln\left(\frac{x}{0.5}\right)\right\}^2\right]^{-3}}{3x\pi\sqrt{5}}$	$\frac{945\left[1+\frac{1}{10}\left\{\ln\left(\frac{x}{0.5}\right)\right\}^2\right]^{-\frac{11}{2}}}{768x\sqrt{10}}$
1	$\frac{8\left[1+\frac{1}{5}\left\{\ln\left(x\right)\right\}^2\right]^{-3}}{3x\pi\sqrt{5}}$	$\frac{945\left[1+\frac{1}{10}\left\{\ln\left(x\right)\right\}^2\right]^{-\frac{11}{2}}}{768x\sqrt{10}}$
2	$\frac{8\left[1+\frac{1}{5}\left\{\ln\left(\frac{x}{2}\right)\right\}^2\right]^{-3}}{3x\pi\sqrt{5}}$	$\frac{945\left[1+\frac{1}{10}\left\{\ln\left(\frac{x}{2}\right)\right\}^2\right]^{-\frac{11}{2}}}{768x\sqrt{10}}$
5	$\frac{8\left[1+\frac{1}{5}\left\{\ln\left(\frac{x}{5}\right)\right\}^2\right]^{-3}}{3x\pi\sqrt{5}}$	$\frac{945\left[1+\frac{1}{10}\left\{\ln\left(\frac{x}{5}\right)\right\}^2\right]^{-\frac{11}{2}}}{768x\sqrt{10}}$

$$\frac{\Gamma\left(\frac{k+1}{2}\right)\log\left(\frac{x}{A}\right)\,{}_2F_1\left(\frac{1}{2},\frac{k+1}{2};\frac{3}{2};\frac{-\log^2\left(\frac{x}{A}\right)}{k}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}} \qquad (4)$$

Where ${}_2F_1$ is a hyper geometric function.

D. *Graphs*

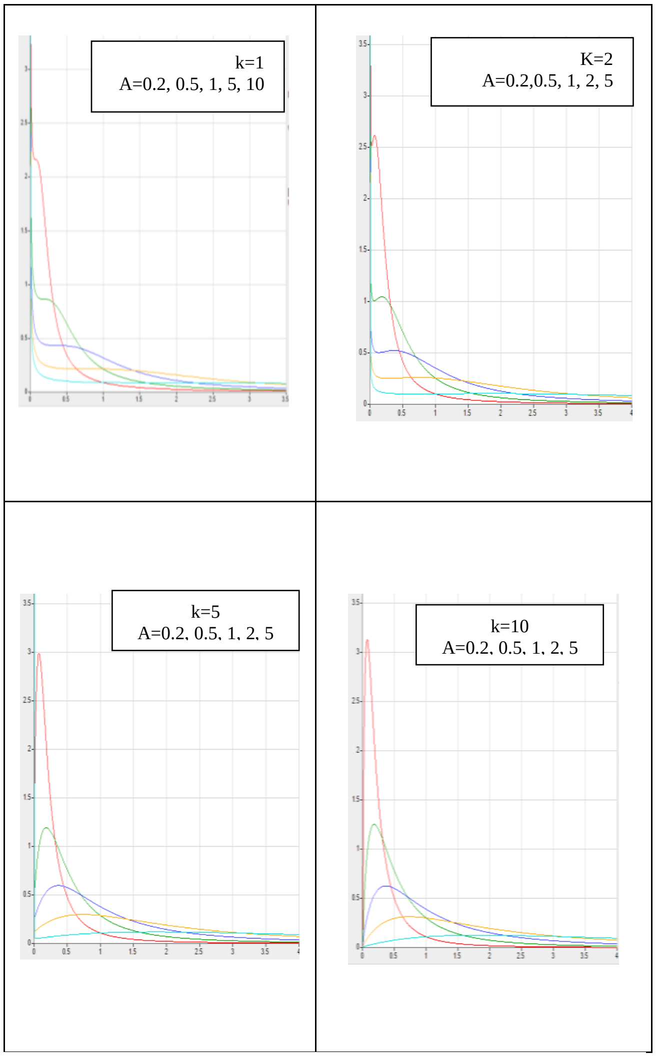


Figure1: The above graphs shows the different PDFs of SIA log-Student-t- distribution for the values A=0.2,0.5,1,2,5 and k=1,2,5,10.

IV. CDF of the SIA log-t distribution

The Cumulative distribution function of the SIA-log-Student-t distribution is given by

A. *The different CDFs for different PDFs of SIA Log-Student-t-Distribution*

At k=1,2

A	F(x)	
	k=1	k=2
0.2	$\frac{\tan^{-1}\left[\ln\left\{\frac{x}{0.2}\right\}\right]}{\pi}$	$\frac{\ln\left\{\frac{x}{0.2}\right\}+\sqrt{\left\{\ln\left(\frac{x}{0.2}\right)\right\}^2+2}}{2\sqrt{\left\{\ln\left(\frac{x}{0.2}\right)\right\}^2+2}}$
0.5	$\frac{\tan^{-1}\left[\ln\left\{\frac{x}{0.5}\right\}\right]}{\pi}$	$\frac{\ln\left\{\frac{x}{0.5}\right\}+\sqrt{\left\{\ln\left(\frac{x}{0.5}\right)\right\}^2+2}}{2\sqrt{\left\{\ln\left(\frac{x}{0.5}\right)\right\}^2+2}}$
1	$\frac{\tan^{-1}\left[\ln\left\{x\right\}\right]}{\pi}$	$\frac{\ln\left\{x\right\}+\sqrt{\left\{\ln\left(x\right)\right\}^2+2}}{2\sqrt{\left\{\ln\left(x\right)\right\}^2+2}}$
2	$\frac{\tan^{-1}\left[\ln\left\{\frac{x}{2}\right\}\right]}{\pi}$	$\frac{\ln\left\{\frac{x}{2}\right\}+\sqrt{\left\{\ln\left(\frac{x}{2}\right)\right\}^2+2}}{2\sqrt{\left\{\ln\left(\frac{x}{2}\right)\right\}^2+2}}$
5	$\frac{\tan^{-1}\left[\ln\left\{\frac{x}{5}\right\}\right]}{\pi}$	$\frac{\ln\left\{\frac{x}{5}\right\}+\sqrt{\left\{\ln\left(\frac{x}{5}\right)\right\}^2+2}}{2\sqrt{\left\{\ln\left(\frac{x}{5}\right)\right\}^2+2}}$

V. Reliability Function for the SIA Log-Student-t-distribution

The reliability function for the SIA log-Student-t-distribution is given as:

$$R(x)=\frac{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}-\Gamma\left(\frac{k+1}{2}\right)\log\left(\frac{x}{A}\right)\,{}_2F_1\left(\frac{1}{2},\frac{k+1}{2};\frac{3}{2};\frac{-\log^2\left(\frac{x}{A}\right)}{k}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}} \qquad (5)$$

The reliability function for the different values of A and k are calculated as:

k=1

A	R(x)
0.2	$\frac{\pi-\tan^{-1}\left[\ln\left\{\frac{x}{0.2}\right\}\right]}{\pi}$
0.5	$\frac{\pi-\tan^{-1}\left[\ln\left\{\frac{x}{0.5}\right\}\right]}{\pi}$
1	$\frac{\pi-\tan^{-1}\left[\ln\left\{x\right\}\right]}{\pi}$
2	$\frac{\pi-\tan^{-1}\left[\ln\left\{\frac{x}{2}\right\}\right]}{\pi}$
5	$\frac{\pi-\tan^{-1}\left[\ln\left\{\frac{x}{5}\right\}\right]}{\pi}$

k=2

A	R(x)
0.2	$\frac{2\sqrt{\left\{\ln\left(\frac{x}{0.2}\right)\right\}^2+2}-\left(\ln\left\{\frac{x}{0.2}\right\}+\sqrt{\left\{\ln\left(\frac{x}{0.2}\right)\right\}^2+2}\right)}{2\sqrt{\left\{\ln\left(\frac{x}{0.2}\right)\right\}^2+2}}$
0.5	$\frac{2\sqrt{\left\{\ln\left(\frac{x}{0.5}\right)\right\}^2+2}-\left(\ln\left\{\frac{x}{0.5}\right\}+\sqrt{\left\{\ln\left(\frac{x}{0.5}\right)\right\}^2+2}\right)}{2\sqrt{\left\{\ln\left(\frac{x}{0.5}\right)\right\}^2+2}}$
1	$\frac{2\sqrt{\left\{\ln\left(x\right)\right\}^2+2}-\left(\ln\left\{x\right\}+\sqrt{\left\{\ln\left(x\right)\right\}^2+2}\right)}{2\sqrt{\left\{\ln\left(x\right)\right\}^2+2}}$
2	$\frac{2\sqrt{\left\{\ln\left(\frac{x}{2}\right)\right\}^2+2}-\left(\ln\left\{\frac{x}{2}\right\}+\sqrt{\left\{\ln\left(\frac{x}{2}\right)\right\}^2+2}\right)}{2\sqrt{\left\{\ln\left(\frac{x}{2}\right)\right\}^2+2}}$

VI. Hazard function:

The hazard function for the SIA log-Student-t distribution can be given as:

$$h(x)=\frac{\Gamma\left(\frac{k+1}{2}\right)\left[1+\frac{1}{k}\left\{\ln\left(\frac{x}{A}\right)\right\}^2\right]^{\frac{-(k+1)}{2}}}{x\left[\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}-\Gamma\left(\frac{k+1}{2}\right)\log\left(\frac{x}{A}\right)\,{}_2F_1\left(\frac{1}{2},\frac{k+1}{2};\frac{3}{2};\frac{-\log^2\left(\frac{x}{A}\right)}{k}\right)\right]} \qquad (6)$$

The hazard function for the different values of A and k can be given as:

k=1

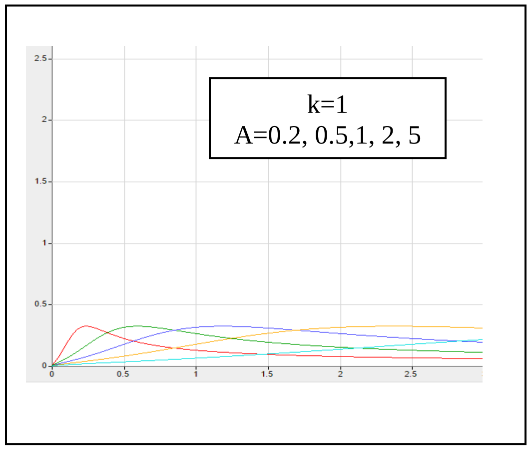
A	h(x)
0.2	$\frac{\left[1+\left\{\ln\left(\frac{x}{0.2}\right)\right\}^2\right]^{-1}}{\pi-\tan^{-1}\left[\ln\left\{\frac{x}{0.2}\right\}\right]}$

0.5	$\dfrac{\left[1+\left\{\ln\left(\frac{x}{0.5}\right)\right\}^{2}\right]^{-1}}{\pi-\tan^{-1}\left[\ln\left\{\frac{x}{0.5}\right\}\right]}$
1	$\dfrac{\left[1+\left\{\ln\left(x\right)\right\}^{2}\right]^{-1}}{\pi-\tan^{-1}\left[\ln\left\{x\right\}\right]}$
2	$\dfrac{\left[1+\left\{\ln\left(\frac{x}{2}\right)\right\}^{2}\right]^{-1}}{\pi-\tan^{-1}\left[\ln\left\{\frac{x}{2}\right\}\right]}$
5	$\dfrac{\left[1+\left\{\ln\left(\frac{x}{5}\right)\right\}^{2}\right]^{-1}}{\pi-\tan^{-1}\left[\ln\left\{\frac{x}{5}\right\}\right]}$

k=2

A	**h(x)**
0.2	$\dfrac{2\left[1+\frac{1}{2}\left\{\ln\left(\frac{x}{0.2}\right)\right\}^{2}\right]^{\frac{3}{2}}\sqrt{\left\{\ln\left(\frac{x}{0.2}\right)\right\}^{2}+2}}{2\sqrt{\left\{\ln\left(\frac{x}{0.2}\right)\right\}^{2}+2}-\left(\ln\left\{\frac{x}{0.2}\right\}+\sqrt{\left\{\ln\left(\frac{x}{0.2}\right)\right\}^{2}+2}\right)}$
0.5	$\dfrac{2\left[1+\frac{1}{2}\left\{\ln\left(\frac{x}{0.5}\right)\right\}^{2}\right]^{\frac{3}{2}}\sqrt{\left\{\ln\left(\frac{x}{0.5}\right)\right\}^{2}+2}}{2\sqrt{\left\{\ln\left(\frac{x}{0.5}\right)\right\}^{2}+2}-\left(\ln\left\{\frac{x}{0.5}\right\}+\sqrt{\left\{\ln\left(\frac{x}{0.5}\right)\right\}^{2}+2}\right)}$
1	$\dfrac{2\left[1+\frac{1}{2}\left\{\ln\left(x\right)\right\}^{2}\right]^{\frac{3}{2}}\sqrt{\left\{\ln\left(x\right)\right\}^{2}+2}}{2\sqrt{\left\{\ln\left(x\right)\right\}^{2}+2}-\left(\ln\left(x\right)+\sqrt{\left\{\ln\left(x\right)\right\}^{2}+2}\right)}$
2	$\dfrac{2\left[1+\frac{1}{2}\left\{\ln\left(\frac{x}{2}\right)\right\}^{2}\right]^{\frac{3}{2}}\sqrt{\left\{\ln\left(\frac{x}{2}\right)\right\}^{2}+2}}{2\sqrt{\left\{\ln\left(\frac{x}{2}\right)\right\}^{2}+2}-\left(\ln\left\{\frac{x}{2}\right\}+\sqrt{\left\{\ln\left(\frac{x}{2}\right)\right\}^{2}+2}\right)}$
5	$\dfrac{2\left[1+\frac{1}{2}\left\{\ln\left(\frac{x}{5}\right)\right\}^{2}\right]^{\frac{3}{2}}\sqrt{\left\{\ln\left(\frac{x}{5}\right)\right\}^{2}+2}}{2\sqrt{\left\{\ln\left(\frac{x}{5}\right)\right\}^{2}+2}-\left(\ln\left\{\frac{x}{5}\right\}+\sqrt{\left\{\ln\left(\frac{x}{5}\right)\right\}^{2}+2}\right)}$

A. Graph



VII. Properties of the SIA log-t distribution:

The mean, variance and higher moments of the SIS log t distribution do not exist. In this section, we present the different properties of SIA log-t-distributions:

A. Quantile Functions

The quantile functions for the different values of A and k is given as follows:

A\k → ↓	**1**	**2**
0.2	$0.2e^{\tan\left[\pi\left\{p-\frac{1}{2}\right\}\right]}$	$0.2e^{(2p-1)\sqrt{\frac{1}{2p(1-p)}}}$
0.5	$0.5e^{\tan\left[\pi\left\{p-\frac{1}{2}\right\}\right]}$	$0.5e^{(2p-1)\sqrt{\frac{1}{2p(1-p)}}}$
1	$e^{\tan\left[\pi\left\{p-\frac{1}{2}\right\}\right]}$	$e^{(2p-1)\sqrt{\frac{1}{2p(1-p)}}}$

2	$\frac{\tan\left[\pi\left\{p-\frac{1}{2}\right\}\right]}{2e}$	$\frac{(2p-1)}{2e}\sqrt{\frac{1}{2p(1-p)}}$
5	$\frac{\tan\left[\pi\left\{p-\frac{1}{2}\right\}\right]}{5e}$	$\frac{(2p-1)}{5e}\sqrt{\frac{1}{2p(1-p)}}$

B. 1st Quartiles

A\k → ↓	**1**	**2**
0.2	0.07357	0.08839
0.5	0.18393	0.22098
1	0.36787	0.44197
2	0.73575	0.88395
5	1.83939	2.20988

C. 3rd Quartile

A	**k**	
	1	**2**
0.2	0.54365	0.45251
0.5	1.35914	1.13127
1	2.71829	2.26255
2	5.43658	4.52511
5	13.59145	11.31279

D. 2nd Quartiles

	k	
0.2	0.2	0.2
0.5	0.5	0.5
1	1	1
2	2	2
5	5	5

E. Mid Quartile Range

A\k → ↓	**1**	**2**
0.2	0.30861	0.27045
0.5	0.77154	0.67613
1	1.54308	1.35226
2	3.08616	2.70453
5	7.71542	6.76134

F. Inter-Quartile Range

A\k → ↓	**1**	**2**
0.2	0.47008	0.36411
0.5	1.17520	0.91029
1	2.35041	1.82058
2	4.70082	3.64116
5	11.75206	9.10291

G. Bowley's co-efficient of Kurtosis

A\k → ↓	**1**	**2**
0.2	0.46211	0.38698
0.5	0.46211	0.38698
1	0.46211	0.38698
2	0.46211	0.38698
5	0.46211	0.38698

H. 1st percentile

A\k → ↓	**1**	**2**
0.2	0.00921	0.03034
0.5	0.02303	0.07586
1	0.04606	0.15173
2	0.09212	0.31470
5	0.23032	0.78676

I. 9th percentile

A\k → ↓	**1**	**2**
0.2	4.34161	1.31808
0.5	10.85402	3.29521
1	21.70805	6.59042
2	43.41611	13.18085
5	108.54028	32.95213

J. Percentile Co-Efficient

A\k → ↓	**1**	**2**
0.2	0.07123	0.21002
0.5	0.07123	0.21002
1	0.07123	0.21002
2	0.07123	0.21020
5	0.07123	0.21020

VI. Concluding Remarks

In order to be able to take appropriate decisions, and to formulate the policy, data-sets are required on qualitative and quantitative variables. Sets of quantitative data pertaining to business, health, social disciplines and other fields are often modeled by distributions of non-negative continuous random variables such as the lognormal, Weibull and gamma distributions. In order to achieve the *generalization* of this distribution, we have developed the SIA-version of this distribution, which is an extension of the log-Student-t-distribution under the transformation X=AY. By taking different values of 'k' and 'A'. We have derived the basic properties of the SIA-log-Student-t distribution including its mid-quartile range, inter-quartile range, Bowley's coefficient of skewness and percentile coefficient of kurtosis for various values of the parameter k, one of the two parameters of this distribution. Reliability function and hazard function of this distribution for various values of k have been presented as well.

SIA-distributions possess a property where estimators of distribution parameters that are *more efficient* than their non-SIA counterparts and we believe that the SIA-log-Student-t distribution will form the basis for the development of an SIA-estimator of k that will lead to *more accurate modeling* of data-sets distributed according to the SIA-log-Student-t distribution.

VII. References

[1] Seshadri, V. (1965). On Random Variables which have the Same Distributions as their Reciprocals. *Can. Math. Bull*., 8(6), 819-824.

[2] Saunders, S.C. (1974). A Family of Random Variables Closed Under Reciprocation . *J. Amer. Statist Assoc*., 69(346),533-539.

[3] Habibullah, S.N., Memon, A.Z. and Ahmad, M. (2010).*On a Class of Distributions Closed Under Inversion,* Lambert Academic Publishing (LAP), ISBN 978-3-8383-4868-1

[4] Habibullah, S.N. and Fatima, S.S. (2015), "On a Newly Developed Estimator for More Accurate Modeling with an Application to Civil Engineering", Proceedings of the 12th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP12) organized by CERRA (Vancouver, BC, Canada, July 12-15, 2015). *Sponsoring Agency: Higher Education Commission, Pakistan.*