

Discrimination Between the weibull and the Inverse Gaussian Distributions

Zakariya Yahya Algamal*, Haithem Taha Mohammad Ali

Department of Statistics and Informatics, University of Mosul, Mosul, Iraq

College of Computers and Information Technology, Nawroz University, Kurdistan region, Iraq

Abstract: One of the problems that appear in reliability and survival analysis is how we choose the best distribution that fitted the data. Sometimes we see that the handle data have two fitted distributions. Both inverse Gaussian and Weibull distributions have been used among many well-known failure time distributions with positively skewed data. The problem of selecting between them is considered. We used the logarithm of maximum likelihood ratio as a test for discriminating between these two distributions. The test has been carried out on some different data sets.

Keywords: Inverse Gaussian distribution, Weibull distribution, Ratio maximum likelihood, Discrimination.

I. Introduction

It is well known that the inverse Gaussian distribution (IG) and Weibull distribution (WEIBULL) are used to analyze asymmetric positively data. In reliability and survival analysis we need these distributions on modeling the failure time data. Sometimes we see that the both distributions fit our data. So, the question is: which one will be preferable than the other?. To answer to this question we use in this paper the likelihood ratio test to discriminating between the IG and WEIBULL distributions. Seven data sets have been taken to prove our test. Discriminating between any two general probability distribution function was studied by Atkinson (1969, 1970), Dumonceaux et al (1973), Dumonceaux and Antle (1973), and Kundu and Manglick (2004, 2005). This paper is organized as follows. Section 2 and section 3 show the properties of the IG and WEIBULL distributions, respectively. In section 4 the description of the likelihood ratio test is mentioned. Seven data sets are analyzed in section 5.

II. The Inverse Gaussian Distribution

The inverse Gaussian distribution is used to model nonnegative skewed data. This distribution referred to the theory of Brownian motion because the distribution of the first passage time of a Brownian motion belongs to the inverse Gaussian (Cklikara & Floks 1988).

Inverse Gaussian distribution has many applications and uses especially in reliability (survival analysis), and in the area on natural and social sciences. Since it is a positively skewed distribution, it has advantage over some other skewed distributions like lognormal, gamma, and weibull.

The p.d.f of an inverse Gaussian r.v X is

$$f(X; \mu, \lambda) = \left(\frac{\lambda}{2\pi}\right)^{1/2} x^{-3/2} \text{Exp}\left[-\frac{\lambda}{2\mu^2} \frac{(X - \mu)^2}{X}\right], \quad X > 0 \quad \dots (1)$$

Where $\mu > 0$ and $\lambda > 0$. The parameter μ represents the mean of the distribution and λ represents the scale parameter. There are three other forms of (2.1) (Tweedie 1957).

The likelihood function of (2.1) is

$$L(\mu, \lambda) = \left(\frac{\lambda}{2\pi}\right)^{n/2} \prod_{i=1}^n x_i^{-3/2} \text{Exp}\left[-\frac{\lambda}{2\mu^2} \sum_{i=1}^n \left\{\frac{(X_i - \mu)^2}{X_i}\right\}\right] \quad \dots (2)$$

And the natural logarithm of (2.2) is,

$$\ln L(\mu, \lambda) = \frac{n}{2} \ln \lambda - \frac{n}{2} \ln(2\pi) - \frac{3}{2} \ln\left(\prod_{i=1}^n x_i\right) - \frac{\lambda}{2\mu^2} \sum_{i=1}^n x_i + \frac{n\lambda}{\mu} - \frac{\lambda}{2} \sum_{i=1}^n \left(\frac{1}{x_i}\right) \quad \dots (3)$$

From (2.3) one can obtain the m.l.e for μ and λ (Tweedie 1956) as following:

$$\hat{\mu} = \bar{x} \quad \dots\dots\dots(4)$$

$$\hat{\lambda} = \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{x_i} - \frac{1}{\bar{x}} \right) \right]^{-1} \quad \dots\dots\dots(5)$$

III. The Weibull Distribution

Weibull distribution has been used in many different applications and for solving a variety of problems from many different disciplines and it is widely used to model continuous variables that are always positive and have skewed distributions. It is also a flexible life distribution model that may offer a good fit to some sets of failure data. The density function of the Weibull distribution with shape parameter α and the scale parameter β will be

$$f(X; \alpha, \beta) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right)^\alpha, \quad x > 0, \alpha > 0, \beta > 0 \quad \dots\dots(6)$$

The likelihood function of the weibull p.d.f is,

$$L(\alpha, \beta) = \alpha^n \beta^{-n\alpha} \prod_{i=1}^n x_i^{\alpha-1} \prod_{i=1}^n \exp\left(-\frac{x}{\beta}\right)^\alpha \quad \dots\dots(7)$$

The natural logarithm of (3.2) is,

$$\ln L(\alpha, \beta) = -n\alpha - n\alpha \ln \beta + (\alpha - 1) \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \left(\frac{x}{\beta}\right)^\alpha \quad \dots\dots(8)$$

By solving $\frac{\partial \ln(\alpha, \beta)}{\partial \beta} = 0$ for (3.3), we get

$$\hat{\beta} = \left(\frac{\sum_{i=1}^n x_i^{\hat{\alpha}}}{n} \right)^{1/\hat{\alpha}} \quad \dots\dots\dots(9)$$

And solving $\frac{\partial \ln(\alpha, \beta)}{\partial \alpha} = 0$ we get

$$\hat{\alpha} = \left(\left(\sum_{i=1}^n x_i^{\hat{\alpha}} \ln x_i \right) \left(\sum_{i=1}^n x_i^{\hat{\alpha}} \right)^{-1} - \frac{\sum_{i=1}^n \ln x_i}{n} \right)^{-1} \quad \dots\dots\dots(10)$$

Where (3.4) and (3.5) represent the m.l.e for α and β . (Johnson & Kotz, 1995, p.187)

IV. Likelihood Ratio Test

A likelihood ratio test (LRT) is a statistical test relying on a test statistics computed by taking the ratio of the maximum value of the likelihood function.

Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ are i.i.d random variables from a known distribution (with p.d.f). Recall that the likelihood function and its logarithm are given, then the LRT (let us denoted it here by L) is defined as:

$$L = \frac{L_1(\hat{\theta}_1, \hat{\theta}_2)}{L_2(\hat{\lambda}_1, \hat{\lambda}_2)} \dots\dots\dots (11)$$

Where $L_1(\hat{\theta}_1, \hat{\theta}_2)$ and $L_2(\hat{\lambda}_1, \hat{\lambda}_2)$ are the likelihood function of a known different p.d.f, and $\hat{\theta}_1, \hat{\theta}_2, \hat{\lambda}_1$ and $\hat{\lambda}_2$ are the m.l.e of $\theta_1, \theta_2, \lambda_1$ and λ_2 , respectively.

Now, from our problem, we rewrite (4.1) as:

$$L = \frac{L_{IG}(\hat{\mu}, \hat{\lambda})}{L_{WEIBULL}(\hat{\alpha}, \hat{\beta})} \dots\dots\dots (12)$$

By taking the natural logarithm of (4.2) and from (2.3), (2.4), (2.5), (3.3), (3.4), and (3.5), one can get

$$\ln L = \frac{n}{2} \left\{ \ln(\hat{\lambda}) - \ln(2\pi) - \bar{G} + \hat{\lambda} \left(\frac{1}{\bar{x}} - \frac{1}{\bar{H}} \right) + 2 \ln \hat{\alpha} + 2 \hat{\alpha} \ln \hat{\beta} - 2 \hat{\alpha} \bar{G} + \frac{2}{n} \sum_{i=1}^n \left(\frac{x}{\hat{\beta}} \right)^{\hat{\alpha}} \right\} \dots\dots\dots (13)$$

Where \bar{X} , \bar{G} and \bar{H} is the arithmetic, geometric, and harmonic mean, respectively.

The hypothesis test will be

\mathbf{H}_0 = The data belong to the IG distribution.

\mathbf{H}_1 = The data belong to the WEIBILL distribution.

Our decision to choose whether the data belong to the IG or to the WEIBULL distribution is based on the value of (4.3). If $\ln L > 0$ we choose the IG distribution as a fitted to the data, elsewhere ($\ln L < 0$) we prefer the WEIBULL distribution as a fitted to the data.

V. Analysis Of Data

In this section we have taken seven data sets in order to apply the formula (3.4) to discriminating between the two mentioned distributions.

5-1 Data Set (1)

Murthy and et al (2004) gave the following data which represent the time between failures for repairable item : [0.11,0.3,0.4,0.45,0.59,0.63,0.7,0.71,0.74,0.77,0.94,1.06,1.17,1.23,1.23,1.24,1.43,1.46,1.49,1.74,1.82,1.86,1.97, 2.23,2.37,2.46,2.63,3.46,4.36,4.73]

Table 1: The m.l.e for both distribution parameters and kolmogrove- Smirnov (K-S) statistic

WEIBULL	IG
$\hat{\alpha} = 1.4233$	$\hat{\mu} = 1.5427$
$\hat{\beta} = 1.6286$	$\hat{\lambda} = 1.6327$
K-S = 0.04756	K-S = 0.1112

Both K-S values are significant (i.e. the data belong to the both distributions). But the value of $\ln L$ is $-2.4212 < 0$, therefore the WEIBULL distribution is more suitable than IG distribution. Also, the K-S distance of WEIBULL is less than the K-S of IG.

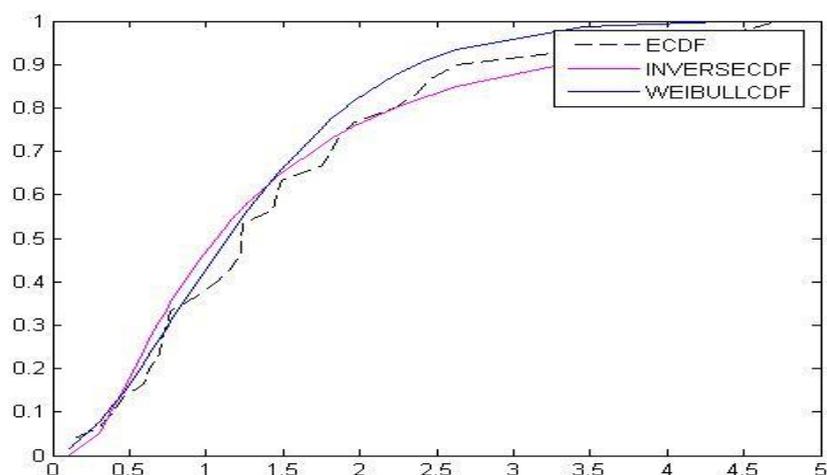


Figure 1: The CDF for both distributions and the ECDF (kolmogrove- Smirnov CDF) for data set (1)

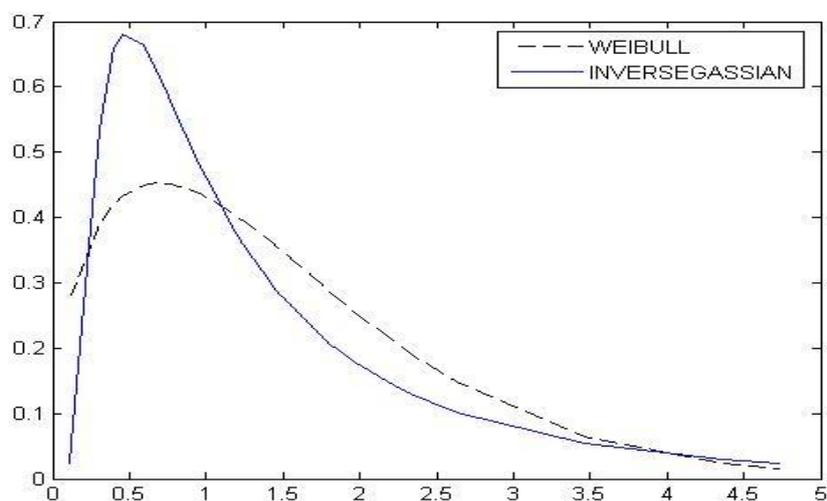


Figure 2: The p.d.f for both distributions for data set (1).

5-2 Data Set (2)

The second set gives data of the failure times of 24 mechanical components (Murthy 2004):
 [10.24,10.67,12.26,14.1,14.7,14.97,16.62,17.12,18.14,18.51,18.84,19.08,19.4,22.38,22.57,22.85,25.43,27.47,27.98,29.93,30.94,36.02,49.56,51.56]

Table 2: The m.l.e for both distribution parameters and kolmogrove- Smirnov (K-S) statistic

WEIBULL	IG
$\hat{\alpha} = 2.7566$	$\hat{\mu} = 22.9725$
$\hat{\beta} = 24.44$	$\hat{\lambda} = 122.4372$
K-S = 0.1309	K-S = 0.022

According to the values of K-S test of the two distributions, we conclude that the data are very well described by these two distributions. But $\ln L = 4.86 > 0$, we prefer that the IG distribution will be more reasonable.

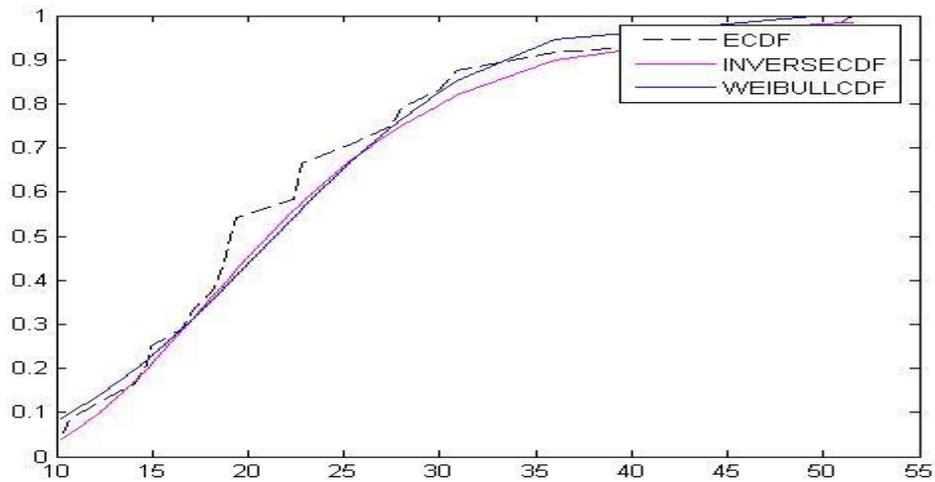


Figure 3: The CDF for both distributions and the ECDF (kolmogrove- Smirnov CDF) for data set (2)

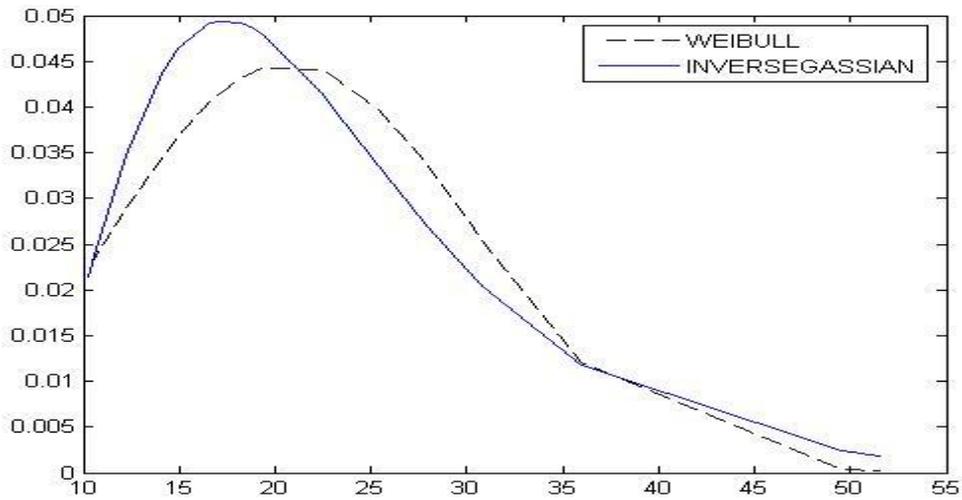


Figure 4: The p.d.f for both distributions for data set (2).

5-3 Data Set (3)

Murthy 2004, show this data:

1.92,2.99,5.52,5.73,5.85,7.75,8.86,9.17,9.37,9.6,10.2,11.24,11.5,12.74,13.36,16.58,18.92,22.48,30.42,38.14

Table 3: The m.l.e for both distribution parameters and kolmogrove- Smirnov (K-S) statistic

WEIBULL	IG
$\hat{\alpha}=1.6129$	$\hat{\mu}=12.617$
$\hat{\beta}=12.947$	$\hat{\lambda}=19.9475$
K-S = 0.1	K-S = 0.1032

Both K-S values are significant. But the value of $\ln \mathbf{L}$ is $0.566 > 0$, therefore the IG distribution is more suitable than WEIBULL distribution.

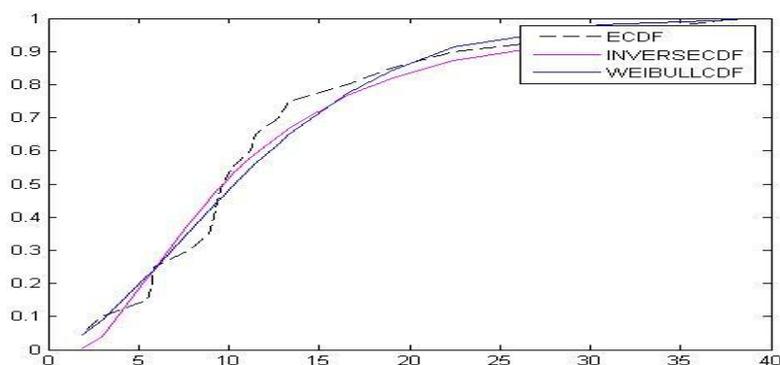


Figure 5: The CDF for both distributions and the ECDF (kolmogrove- Smirnov CDF) for data set (3)

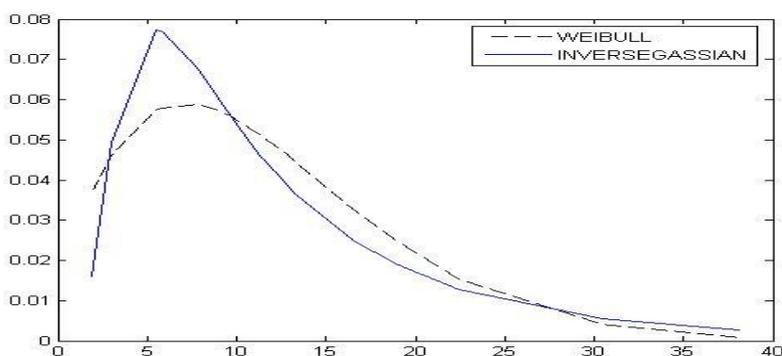


Figure 6: The p.d.f for both distributions for data set (3).

5-4 Data Set (4)

The data below are the voltage levels at which failures occurred (Lawless 2003), [32,35.4,36.2,39.8,41.2,43.3,45.5,46,46.2,46.4,46.5,46.8,47.3,47.3,47.6,49.2,50.4,50.9,52.4,56.3].

Table 4: The m.l.e for both distribution parameters and kolmogrove- Smirnov (K-S) statistic

WEIBULL	IG
$\hat{\alpha} = 7.97$	$\hat{\mu} = 45.335$
$\hat{\beta} = 47.41$	$\hat{\lambda} = 2441.37$
K-S = 0.1634	K-S = 0.1877

According to the values of K-S test of the two distributions, we conclude that the data are very well described by these two distributions. But $\ln L = -1.28 < 0$, we prefer that the WEIBULL distribution will be more reasonable.

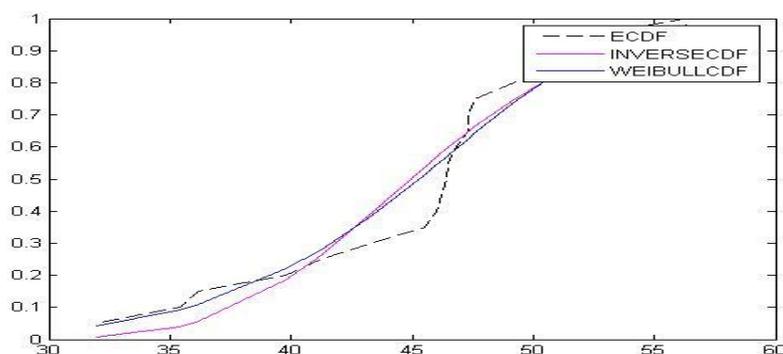


Figure 7: The CDF for both distributions and the ECDF (kolmogrove- Smirnov CDF) for data set (4)

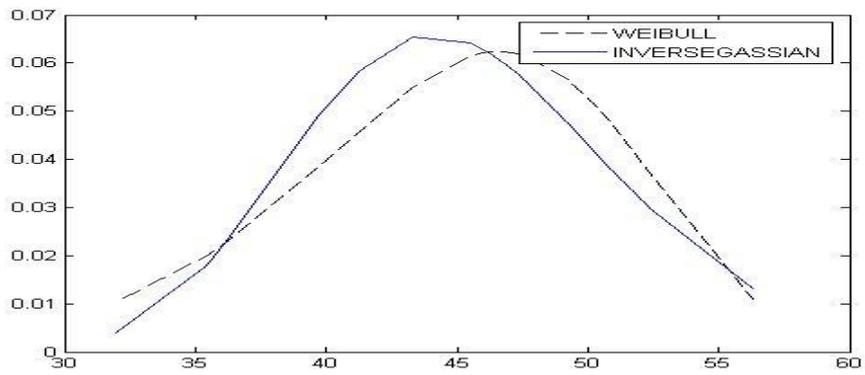


Figure 8: The p.d.f for both distributions for data set (4).

5-5 Data Set (5)

The data below are the voltage levels at which failures occurred (Lawless 2003), [39.4,45.3,49.2,49.4,51.3,52,53.2,53.2,54.9,55.5,57.1,57.2,57.5,59.2,61,62.4,63.8,64.3,67.3,67.7].

Table 5: The m.l.e for both distribution parameters and kolmogrove- Smirnov (K-S) statistic

WEIBULL	IG
$\hat{\alpha} = 8.225$	$\hat{\mu} = 56.045$
$\hat{\beta} = 58.6$	$\hat{\lambda} = 3188.48$
K-S = 0.075	K-S = 0.072

Both K-S values are significant. But the value of $\ln L$ is $-0627 < 0$, therefore the WEIBULL distribution is more suitable than IG distribution.

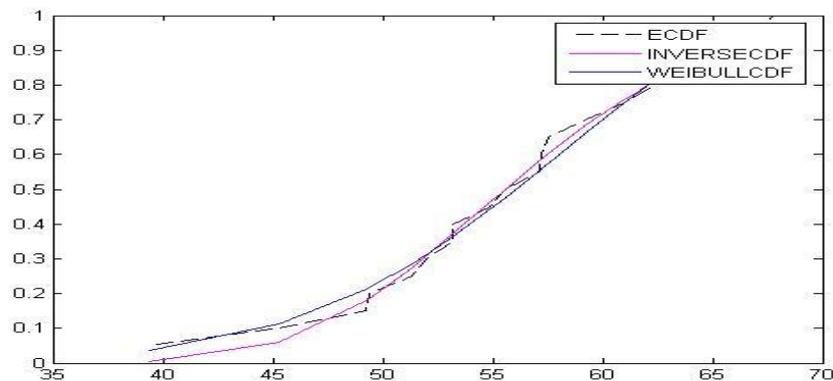


Figure 9: The CDF for both distributions and the ECDF (kolmogrove- Smirnov CDF) for data set (5).

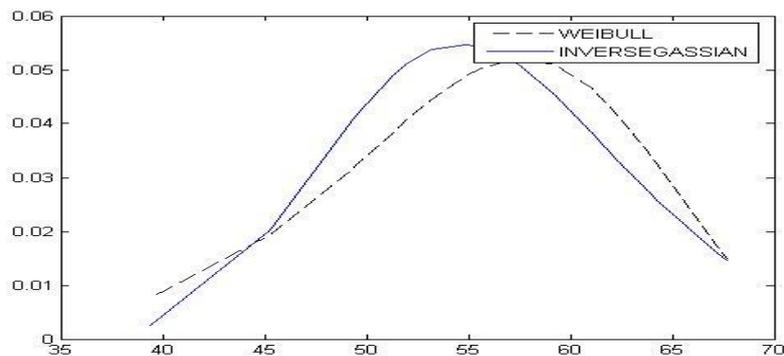


Figure 10: The p.d.f for both distributions for data set (5).

5-6 Data Set (6)

Kumagai et al (1989) presented the following time series data for toluene exposure concentrations (8 hr TWAs) for a worker doing stain removing.

0.9,1.1,1.9,2.1,2.6,2.9,3.1,3.2,4.9,4.9,5.2,5.8,6.2,6.9,7.8,8.3,8.7,10.5,11.1,13.6,16.6,17.4,20.4,21.9,22.4,50.9,57.4,58.3,58.6,66.9

Table 6: The m.l.e for both distribution parameters and kolmogrove- Smirnov (K-S) statistic

WEIBULL	IG
$\hat{\alpha} = 0.95$	$\hat{\mu} = 16.75$
$\hat{\beta} = 14.253$	$\hat{\lambda} = 6.464$
K-S = 0.1017	K-S = 0.072

Both K-S values are significant. But the value of $\ln L$ is 2.588 > 0, therefore the IG distribution is more suitable than WEIBULL distribution.

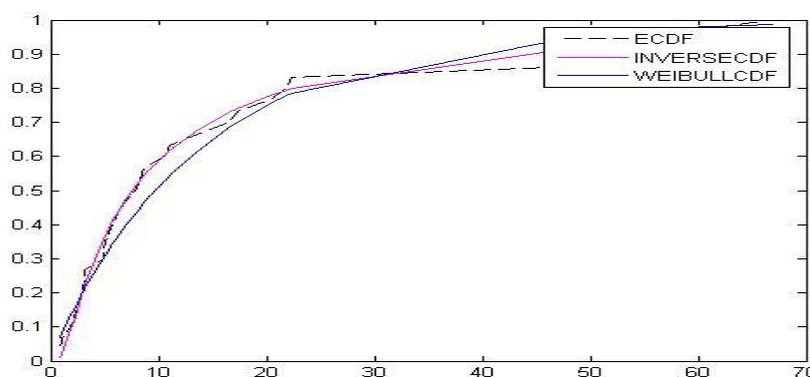


Figure 11: The CDF for both distributions and the ECDF (kolmogrove- Smirnov CDF) for data set (6).

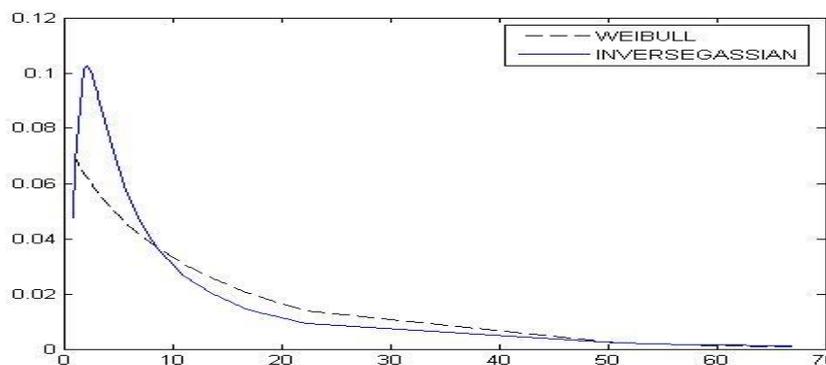


Figure 12: The p.d.f for both distributions for data set (6).

5-7 Data Set (7)

Kumagai and Matsunaga (1995) give these data

1.5,1.7,2.1,2.2,2.4,2.5,2.6,3.8,3.8,4.2,4.3,5.6,6,7,7.5,9.3,9.9,10.2,10.6,12.3,12.9,13.7,14.1,17.8,27.6,31,42,45.6,51.9,91.3,131.8

Table 7: The m.l.e for both distribution parameters and kolmogrove- Smirnov (K-S) statistic

WEIBULL	IG
$\hat{\alpha} = 0.974$	$\hat{\mu} = 19.006$
$\hat{\beta} = 14.369$	$\hat{\lambda} = 7.2326$
K-S = 0.116	K-S = 0.095

Both K-S values are significant. But the value of $\ln L$ is $7.246 > 0$, therefore the IG distribution is more suitable than WEIBULL distribution.

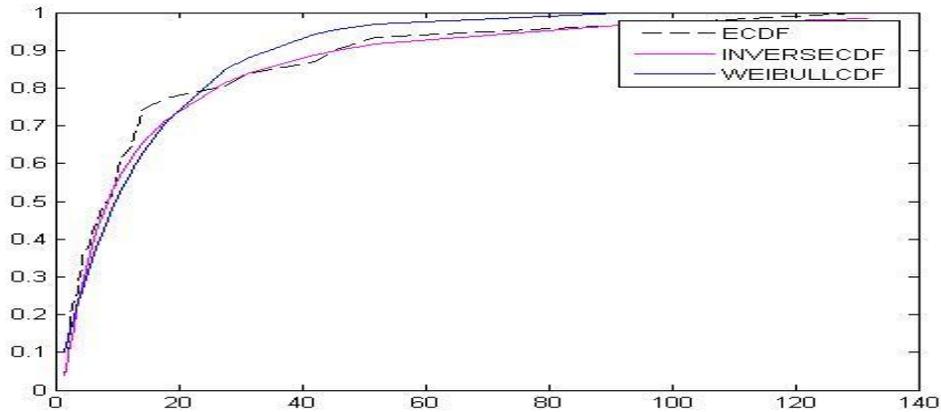


Figure 13: The CDF for both distributions and the ECDF (kolmogrove- Smirnov CDF) for data set (7).

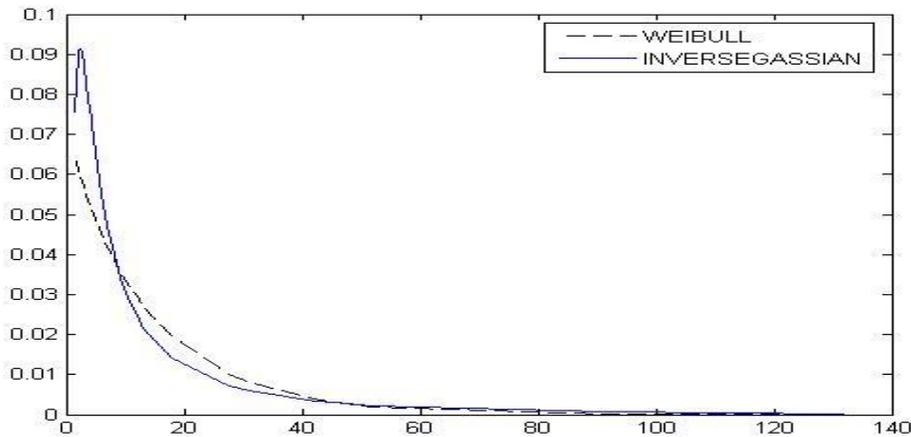


Figure 14: The p.d.f for both distributions for data set (7).

VI. Conclusions

1- Through Tables 1, 4, and 5 we see that these data have the same distributions according to the value of K-S test but the value of $\ln L$ suggests that these data to have the WEIBULL distribution rather than the IG distribution.

2- From Tables 2, 3, 6, and 7 the data have both distributions, but according to the value of $\ln L$ the IG distribution is more suitable than WEIBULL distribution.

References

- [1]. Atkinson, A. (1969), "A Test of Discriminating between Models," *Biometrika*, 56, 337-341.
- [2]. ——— (1970), "A Method for Discriminating between Models" (with Discussion), *Journal of Royal Statistical Society, Ser. B*, 32, 323-353.
- [3]. Chhikara, R. S., and Folks, J. L. (1977), "The Inverse Gaussian Distribution as a Lifetime Model," *Technometrics*, 19, 461-468.
- [4]. ——— (1988), "Inverse Gaussian Distribution: Theory, Methodology, and Applications" Marcel Dekker, Inc., New York.
- [5]. Dumonceaux, R., Antle, C. E., and Hass, G. (1973), "Likelihood Ration Test for Discriminating between Two Models with Unknown Location and Scale Parameters," *Technometrics*, 15, 19-31.
- [6]. Dumonceaux, R., Antle, C. E. (1973), "Discriminating between the Log-Normal and Weibull Distribution," *Technometrics*, 15, 923-926.
- [7]. Johnson, N. L., and Kotz, S. (1995), "Continuous Univariate Distributions-1," 2nd Ed., Wiley, New York.
- [8]. Kundua, D., and Manglick, A. (2004), "Discriminating between the Log-Normal and Gamma Distributions," *Noval Research Logistic*, 51, 893-905.
- [9]. ———(2005), "Discriminating between the Log-Normal and Gamma Distributions," *Journal of Applied Statistical Sciences*, 14, 175-187.

- [10]. Kumagai, S., Matsunaga, I., Sugimoto, K., Kusaka, Y., and Shirakawa, T. (1989), "Assessment of occupational Exposures to Industrial Hazardous Substances (III) on the Frequency Distribution of daily Exposure Averages (8 hr TWA)," *Japanese Journal of Industrial Health*, 31, 216-226.
- [11]. Kumagai, S., Matsunaga, I. (1995), "Changes in the Distribution of Short-Term Exposure Concentration with Different Averaging times," *American Industrial Hygiene Association Journal*, 54, 24-31.
- [12]. Lawless, J. F. (2003), "Statistical Models and Methods for Lifetime Data," 2nd Ed., Wiley, New Jersey.
- [13]. Murthy, D.N. and et al (2004), "Weibull Models", Wiley, New Jersey.
- [14]. Tweedie M. C. K. (1957a), "Statistical Properties of Inverse Gaussian Distribution. I," *Annals Mathematical Statistics*, 28, 362-377.
- [15]. ———(1957b), "Statistical Properties of Inverse Gaussian Distribution. II," *Annals Mathematical Statistics*, 28, 696-705.
- [16]. Algamal, Z. Y. (2015), "Using Maximum Likelihood Ratio Test to Discriminate Between the Inverse Gaussian and Gamma Distributions", *International Journal of Statistical Distributions* 1 (1), 27-32.
- [17]. Algamal, Z. Y., & Lee, M. H. (2015). Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. *Expert Systems with Applications*, 42(23), 9326-9332.
- [18]. Al-Fakih, A. M., Aziz, M., Abdallah, H. H., Algamal, Z. Y., Lee, M. H., & Maarof, H. (2015). High dimensional QSAR study of mild steel corrosion inhibition in acidic medium by furan derivatives. *Int. J. Electrochem. Sci*, 10, 3568-3583.
- [19]. Algamal, Z. Y., & Lee, M. H. (2015). Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification. *Computers in biology and medicine*, 67, 136-145.
- [20]. AL-Jammal, Z. Y. (2008). Exponentiated Exponential distribution as a failure time distribution. *IRAQI Journal of Statistical science*, 14, 63-75.