How Normal is Normal: How Symmetric is Symmetric: How Local is the Location for a Symmetric Distribution

Silvia Irin Sharna, Mian Arif Shams Adnan, Rahmatullah Imon
Department of Mathematical Sciences, Ball State University, Muncie, IN 47304

Abstract
The typical terms "Symmetric as well as Normal" in Statistics require some classification, at least, including arithmetically symmetric and/or normal, geometrically symmetric and/or normal, or harmonically symmetric and/or normal. Tests for classifying the pattern and symmetricity and/or normality of the data; finding proper location and other characteristics, Box-plots, cut off points; outlier(s) detection criteria, criticism of the calculations of the location of the existing probability distributions, etc have been developed.

Key Words: Arithmetically Normal, Arithmetically Symmetric, Geometrically Asymmetric, Geometrically Abnormal, Harmonically Normal, Harmonically Symmetric.

1. Introduction

Statistics is a science of going from particular to general. Since, population along with characteristics might not be known to a researcher or an analyst before carrying out the prophesy, the analyst tries to figure out the characteristics or the shapes or tail behavior or the pattern of the distribution using various Meta analyses based sample estimates from sample(s). Unfortunately, some normality assumptions with bigger sample sizes ensemble some basic questions which are very invisibly and pragmatically pertinent as (i) How big sample size is big? (ii) What is optimum number and optimum length of the class interval of the histogram? (iii) What is the type of the sequence of the data? and so on. Data can be of arithmetic or geometric or harmonic pattern with their respective arithmetic or geometric or harmonic moments. It is better to discuss first about the pattern of the sequence of the data for the symmetric data rather than the asymmetric ones. At first it is needed to represent the data set with a location as for example one dimensional location or mean of the data since the successive higher dimensional locations are some functions of the very basic first location. So, it is important to have a biopsy of how local or representative is the first order moment to entire data of the sample.

Although we are not considering order statistics, but still we represent all data of one variable on a real line which is some sort of representation from lower to upper values. That is, in traditional statistics, instead of assuming the dependent ordered observations we are assuming that all the observations are independent among themselves whereas we are presenting all the observations (along with their corresponding frequencies) on an ordered real line. So, we should not overlook the approach of ordering the data in (order free) traditional statistics. This is also evident in checking the normality for a data set using p-p plot.

After ordering all the data, if we treat it as a sequence of several ordered observations, we will observe that the sequence follows either arithmetic or geometric or harmonic progression. After checking goodness of fit test between observed quantiles and expected progression's quantiles, we can guess the type of sequence of the data as either arithmetically or geometrically or harmonically Symmetric (or Asymmetric).

Since we are entitled only for the symmetric distribution, we should only discuss here on the new way of classifying symmetric distributions is three types including (i) Arithmetically Symmetric, (ii) Geometrically Symmetric and (iii) Harmonically Symmetric.

The aim of this paper is to develop the methods of (i) figuring out that the data comes from which type of patterned symmetric distributions: arithmetically symmetric or geometrically symmetric or harmonically symmetric? (ii) What would be the form of expected quantiles and moments of each of the typically symmetric data? (iii) What would be the graphical approaches of checking the Symmetricity of the type of the symmetric data using Box Plots, Histograms and P-P plots. (iv) What are the graphical and theoretical methods of Outlier detection using proper cutoff points for each of the symmetric patterned data? (v) What are the existing examples of the various types of traditional distributions along with the less local location?

2. Statistical Methods and Methodologies for detecting the type of Normality

Let $x_{1,}x_2$ **,…,** x_n be a random sample of size n from a population. We want to develop the goodness of fit criteria for deciding the data come from which distribution based on quantiles. Several series of new tests can be brought to light for Quartiles, Octiles, Deciles, Percentiles or any quantiles with chi-square test statistic with one or more than one degree of freedom. Ones we can figure out the progression of the ordered data, we can say data follows a specific type of symmetric distribution.

2.1 Quartiles

The concerned hypotheses are

$$H_0\!: The\ data\ follows\ Arithmetically\ symmetric\ distributions.$$
$$H_1\!: The\ data\ follows\ Geometrically\ symmetric\ distributions.$$
$$H_2\!: The\ data\ follows\ Harmonically\ symmetric\ distributions.$$

After ordering the observations we may find the observed quartiles say Q_1, Q_2, Q_3 respectively. Now if the data comes from arithmetically or geometrically or harmonically symmetric distributions the expected 2nd quartile will be $\frac{Q_1+Q_3}{2}$, $\sqrt{Q_1Q_3}$ or $\frac{2}{\frac{1}{Q_1}+\frac{1}{Q_3}}$, respectively i.e., $O_2 = \frac{O_1+O_3}{2}$, $O_2 = \sqrt{O_1O_3}$, $Q_2 = \frac{2}{\frac{1}{Q_1}+\frac{1}{Q_3}}$.

2.1.1 Goodness of Fit Test for Quartiles

Now for the concerned sequences or progression-ed typed data the following statistics are true:

Arithmetic Fit (1): $$\frac{\left(Q_2-\frac{Q_1+Q_3}{2}\right)^2}{\frac{Q_1+Q_3}{2}} \sim \chi_1^2,$$

Geometric Fit (2): $$\frac{\left(Q_2-\sqrt{Q_1Q_3}\right)^2}{\sqrt{Q_1Q_3}} \sim \chi_1^2,$$

Harmonic Fit (3): $$\frac{\left(Q_2-\frac{2}{\frac{1}{Q_1}+\frac{1}{Q_3}}\right)^2}{\frac{2}{\frac{1}{Q_1}+\frac{1}{Q_3}}} \sim \chi^2_1.$$

(Qi) If $\chi^2 < \chi^2_1$, we will accept H. Otherwise we will rejects H.

(Qii) If any two of the aforesaid hypotheses are accepted, then

$$\frac{\chi^2_1}{\chi^2_1} \sim F_{1,1}$$

If $F < F_{1,1}$, we will accept the hypothesis concerned to the denominator of the F statistic; otherwise accept that of the numerator of the F statistic.

2.2 Octiles

After ordering the observations we may find the observed octiles say $O_1, O_2, O_3, O_4, O_5, O_6, O_7$ respectively.

O_1 O_2 O_3 O_4 O_5 O_6 O_7

Now if the data comes from arithmetically symmetric distributions the expected octiles will be respectively as follows.

$$O_2 = \frac{O_1+O_3}{2},$$

$$O_3 = \frac{O_2+O_4}{2} = \frac{O_1+O_5}{2} = \frac{O_1+O_2+O_4+O_5}{4},$$

$$O_4 = \frac{O_1+O_7}{2} = \frac{O_2+O_6}{2} = \frac{O_3+O_5}{2} = \frac{O_2+O_3+O_5+O_6}{4} = \frac{O_1+O_2+O_6+O_7}{4} = \frac{O_1+O_3+O_5+O_7}{4} = \frac{O_1+O_2+O_3+O_5+O_6+O_7}{6},$$

$$O_5 = \frac{O_4+O_6}{2} = \frac{O_3+O_7}{2} = \frac{O_3+O_4+O_6+O_7}{4},$$

$$O_6 = \frac{O_5+O_7}{2}$$

For geometrically symmetric distribution, the expected octiles will be as below.

$$O_2 = \sqrt{O_1 O_3},$$

$$O_3 = \sqrt{O_2 O_4} = \sqrt{O_1 O_5} = \sqrt[4]{O_1 O_2 O_4 O_5},$$

$$O_4 = \sqrt{O_3 O_5} = \sqrt{O_2 O_6} = \sqrt{O_1 O_7} = \sqrt[4]{O_2 O_3 O_5 O_6} = \sqrt[4]{O_1 O_2 O_6 O_7} = \sqrt[4]{O_1 O_3 O_5 O_7} = \sqrt[6]{O_1 O_2 O_3 O_5 O_6 O_7},$$

$$O_5 = \sqrt{O_4 O_6} = \sqrt{O_3 O_7} = \sqrt[4]{O_3 O_4 O_6 O_7},$$

$$O_6 = \sqrt{O_5 O_7}.$$

Harmonically symmetric distribution will leave the expected octiles as below:

$$O_2 = \frac{2}{\frac{1}{Q_1}+\frac{1}{Q_3}},$$

$$O_3 = \frac{2}{{O_2}^{-1}+{O_4}^{-1}} = \frac{2}{{O_1}^{-1}+{O_5}^{-1}} = \frac{4}{{O_1}^{-1}+{O_2}^{-1}+{O_4}^{-1}+{O_5}^{-1}},$$

$$O_4 = \frac{2}{{O_1}^{-1}+{O_7}^{-1}} = \frac{2}{{O_2}^{-1}+{O_6}^{-1}} = \frac{2}{{O_3}^{-1}+{O_5}^{-1}} = \frac{4}{{O_2}^{-1}+{O_3}^{-1}+{O_5}^{-1}+{O_6}^{-1}} = \frac{4}{{O_1}^{-1}+{O_2}^{-1}+{O_6}^{-1}+{O_7}^{-1}} = \frac{4}{{O_1}^{-1}+{O_3}^{-1}+{O_5}^{-1}+{O_7}^{-1}} = \frac{6}{{O_1}^{-1}+{O_2}^{-1}+{O_3}^{-1}+{O_5}^{-1}+{O_6}^{-1}+{O_7}^{-1}},$$

$$O_5 = \frac{2}{{O_4}^{-1}+{O_6}^{-1}} = \frac{2}{{O_3}^{-1}+{O_7}^{-1}} = \frac{4}{{O_3}^{-1}+{O_4}^{-1}+{O_6}^{-1}+{O_7}^{-1}},$$

$$O_6 = \frac{2}{\frac{1}{Q_5}+\frac{1}{Q_7}}.$$

2.2.1 Goodness of Fit Test for Octiles

For testing whether the distribution is arithmetically symmetric or not, we can have various types of hypotheses checked due to the properties of the octiles distance squares.

2.2.1.1a Test of Arithmetic Symmetricity about the origin O_4

A bunch of hypotheses have been disclosed immediately to check the level of arithmetic symmetricity.

H_{A1}: *The two sided middle* 25 % *data follows arithmetically symmetric distributions.*
H_{A2}: *The two sided extreme* 25 % *data follows arithmetically symmetric distribution.*
H_{A3}: *The* 25 % *two sided data inbetween middle and extreme follows arithmetically symmetric distribution.*
H_{A4}: *The two sided middle* 50 % *data follows averagely arithmetically symmetric distribution.*
H_{A5}: *The two sided extreme* 50 % *data follows averagely arithmetically symmetric distribution.*
H_{A6}: *The* 50 % *two sided data inbetween middle and extreme follows averagely arithmetically symmetric distribution.*
H_{A7}: *The two sided middle* 75 % *data follows averagely arithmetically symmetric distribution.*
H_{A8}: *Two side any middle* 25 % *data follows arithmetically symmetric distributions.*
H_{A9}: *Two sided any middle* 50 % *data follows arithmetically symmetric distributions.*
H_{A10}: *The two sided data follows arithmetically symmetric distribution.*

All squared distanced octiles follow chi-square statistic with 1 degree of freedom. As such the explicit forms of all of the squared distanced octiles are presented below.

$$\frac{\left(O_4-\frac{O_1+O_7}{2}\right)^2}{\frac{O_1+O_7}{2}}, \frac{\left(O_4-\frac{O_2+O_6}{2}\right)^2}{\frac{O_2+O_6}{2}}, \frac{\left(O_4-\frac{O_3+O_5}{2}\right)^2}{\frac{O_3+O_5}{2}}, \frac{\left(O_4-\frac{O_2+O_3+O_5+O_6}{4}\right)^2}{\frac{O_2+O_3+O_5+O_6}{4}},$$

$$\frac{\left(O_4-\frac{O_1+O_2+O_6+O_7}{4}\right)^2}{\frac{O_1+O_2+O_6+O_7}{4}}, \frac{\left(O_4-\frac{O_1+O_3+O_5+O_7}{4}\right)^2}{\frac{O_1+O_3+O_5+O_7}{4}}, \frac{\left(O_4-\frac{O_1+O_2+O_3+O_5+O_6+O_7}{6}\right)^2}{\frac{O_1+O_2+O_3+O_5+O_6+O_7}{6}} \sim \chi^2_1,$$

(O Ai) For the following inequality, the concerned hypothesis will be accepted

$$\frac{\left(O_4-\frac{O_3+O_5}{2}\right)^2}{\frac{O_3+O_5}{2}} < \chi^2_1,$$

H_{A1}: *The two sided middle* 25 % *data follows arithmetically symmetric distributions.*

(O Aii) For the next inequality, the post stated hypothesis will be accepted

$$\frac{\left(O_4-\frac{O_1+O_7}{2}\right)^2}{\frac{O_1+O_7}{2}} < \chi^2_1,$$

H_{A2}: *The two sided extreme* 25 % *data follows arithmetically symmetric distribution.*

(O Aiii) For the following inequality, the next hypothesis will be accepted

$$\frac{\left(O_4-\frac{O_2+O_6}{2}\right)^2}{\frac{O_2+O_6}{2}} < \chi^2_1,$$

H_{A3}: *The* 25 % *two sided data inbetween middle and extreme follows arithmetically symmetric distribution.*

(O Aiv) For the next inequality, the following hypothesis will be accepted

$$\frac{\left(O_4-\frac{O_2+O_3+O_5+O_6}{4}\right)^2}{\frac{O_2+O_3+O_5+O_6}{4}}<\chi^2_1,$$

H_{A4}: *The tw sided middle* 50 % *data follows averagely arithmetically symmetric distribution.*

(O Av) For the next stated inequality, the concerned hypothesis will be accepted

$$\frac{\left(O_4-\frac{O_1+O_2+O_6+O_7}{4}\right)^2}{\frac{O_1+O_2+O_6+O_7}{4}}<\chi^2_1,$$

H_{A5}: *The two sided extreme* 50 % *data follows averagely arithmetically symmetric distribution.*

(O Avi) For the following inequality, the posterior hypothesis will be accepted

$$\frac{\left(O_4-\frac{O_1+O_3+O_5+O_7}{4}\right)^2}{\frac{O_1+O_3+O_5+O_7}{4}}<\chi^2_1,$$

H_{A6}: *The* 50 % *two sided data inbetween middle and extreme follows*

averagely arithmetically symmetric distribution.

(O Avii) For the inequality, the hypothesis will be accepted

$$\frac{\left(O_4-\frac{O_1+O_2+O_3+O_5+O_6+O_7}{6}\right)^2}{\frac{O_1+O_2+O_3+O_5+O_6+O_7}{6}}<\chi^2_1,$$

H_{A7}: *The two sided middle* 75 % *data follows averagely arithmetically symmetric distribution.*

(O Aviii) For the following inequality, the concerned hypothesis will be accepted

$$\frac{\left(O_4-\frac{O_1+O_7}{2}\right)^2}{\frac{O_1+O_7}{2}}+\frac{\left(O_4-\frac{O_2+O_6}{2}\right)^2}{\frac{O_2+O_6}{2}}+\frac{\left(O_4-\frac{O_3+O_5}{2}\right)^2}{\frac{O_3+O_5}{2}}<\chi^2_3,$$

H_{A8}: *Two sided any middle* 25 % *data follows arithmetically symmetric distributions.*

(O Aix) For the next inequality, the following hypothesis will be accepted

$$\frac{\left(O_4-\frac{O_2+O_3+O_5+O_6}{4}\right)^2}{\frac{O_2+O_3+O_5+O_6}{4}}+\frac{\left(O_4-\frac{O_1+O_2+O_6+O_7}{4}\right)^2}{\frac{O_1+O_2+O_6+O_7}{4}}+\frac{\left(O_4-\frac{O_1+O_3+O_5+O_7}{4}\right)^2}{\frac{O_1+O_3+O_5+O_7}{4}}<\chi^2_3,$$

H_{A9}: *Two sided any middle* 50 % *data follows arithmetically symmetric distributions.*

(O Ax) If

$$\left[\frac{\left(O_4-\frac{O_1+O_7}{2}\right)^2}{\frac{O_1+O_7}{2}}+\frac{\left(O_4-\frac{O_2+O_6}{2}\right)^2}{\frac{O_2+O_6}{2}}+\frac{\left(O_4-\frac{O_3+O_5}{2}\right)^2}{\frac{O_3+O_5}{2}}+\frac{\left(O_4-\frac{O_2+O_3+O_5+O_6}{4}\right)^2}{\frac{O_2+O_3+O_5+O_6}{4}}+\frac{\left(O_4-\frac{O_1+O_2+O_6+O_7}{4}\right)^2}{\frac{O_1+O_2+O_6+O_7}{4}}+\frac{\left(O_4-\frac{O_1+O_3+O_5+O_7}{4}\right)^2}{\frac{O_1+O_3+O_5+O_7}{4}}+\frac{\left(O_4-\frac{O_1+O_2+O_3+O_5+O_6+O_7}{6}\right)^2}{\frac{O_1+O_2+O_3+O_5+O_6+O_7}{6}}\right]<\chi^2_7,$$

we will accept the hypothesis

H_{A10}: *The two sided data follows arithmetically symmetric distribution.*

(O Ai) For the following inequality, the concerned hypothesis will be accepted

$$\frac{\left(O_4-\frac{O_3+O_5}{2}\right)^2}{\frac{O_3+O_5}{2}}<\chi^2_1,$$

H_{A1}: *The two sided middle* 25 % *data follows arithmetically symmetric distributions.*

(O Aii) For the next inequality, the post stated hypothesis will be accepted

$$\frac{\left(O_4-\frac{O_1+O_7}{2}\right)^2}{\frac{O_1+O_7}{2}}<\chi^2_1,$$

H_{A2}: *The two sided extreme* 25 % *data follows arithmetically symmetric distribution.*

(O Aiii) For the following inequality, the next hypothesis will be accepted

$$\frac{\left(O_4-\frac{O_2+O_6}{2}\right)^2}{\frac{O_2+O_6}{2}}<\chi^2_1,$$

H_{A3}: *The* 25 % *two sided data inbetween middle and extreme follows*

$$arithmetically\ symmetric\ distribution.$$

(O Aiv) For the next inequality, the following hypothesis will be accepted

$$\frac{\left(O_4 - \frac{O_2+O_3+O_5+O_6}{4}\right)^2}{\frac{O_2+O_3+O_5+O_6}{4}} < \chi^2_1,$$

H_{A4}: $The\ tw\ sided\ middle$ 50 % $data\ follows\ averagely\ arithmetically\ symmetric\ distribution.$

(O Av) For the next stated inequality, the concerned hypothesis will be accepted

$$\frac{\left(O_4 - \frac{O_1+O_2+O_6+O_7}{4}\right)^2}{\frac{O_1+O_2+O_6+O_7}{4}} < \chi^2_1,$$

H_{A5}: $The\ two\ sided\ extreme$ 50 % $data\ follows\ averagely\ arithmetically\ symmetric\ distribution.$

(O Avi) For the following inequality, the posterior hypothesis will be accepted

$$\frac{\left(O_4 - \frac{O_1+O_3+O_5+O_7}{4}\right)^2}{\frac{O_1+O_3+O_5+O_7}{4}} < \chi^2_1,$$

H_{A6}: The 50 % $two\ sided\ data\ inbetween\ middle\ and\ extreme\ follows$
$averagely\ arithmetically\ symmetric\ distribution.$

(O Avii) For the inequality, the hypothesis will be accepted

$$\frac{\left(O_4 - \frac{O_1+O_2+O_3+O_5+O_6+O_7}{6}\right)^2}{\frac{O_1+O_2+O_3+O_5+O_6+O_7}{6}} < \chi^2_1,$$

H_{A7}: $The\ two\ sided\ middle$ 75 % $data\ follows\ averagely\ arithmetically\ symmetric\ distribution.$

(O Aviii) For the following inequality, the concerned hypothesis will be accepted

$$\frac{\left(O_4 - \frac{O_1+O_7}{2}\right)^2}{\frac{O_1+O_7}{2}} + \frac{\left(O_4 - \frac{O_2+O_6}{2}\right)^2}{\frac{O_2+O_6}{2}} + \frac{\left(O_4 - \frac{O_3+O_5}{2}\right)^2}{\frac{O_3+O_5}{2}} < \chi^2_3,$$

H_{A8}: $Two\ sided\ any\ middle$ 25 % $data\ follows\ arithmetically\ symmetric\ distributions.$

(O Aix) For the next inequality, the following hypothesis will be accepted

$$\frac{\left(O_4 - \frac{O_2+O_3+O_5+O_6}{4}\right)^2}{\frac{O_2+O_3+O_5+O_6}{4}} + \frac{\left(O_4 - \frac{O_1+O_2+O_6+O_7}{4}\right)^2}{\frac{O_1+O_2+O_6+O_7}{4}} + \frac{\left(O_4 - \frac{O_1+O_3+O_5+O_7}{4}\right)^2}{\frac{O_1+O_3+O_5+O_7}{4}} < \chi^2_3,$$

H_{A9}: $Two\ sided\ any\ middle$ 50 % $data\ follows\ arithmetically\ symmetric\ distributions.$

(O Ax) If

$$\left[\frac{\left(O_4 - \frac{O_1+O_7}{2}\right)^2}{\frac{O_1+O_7}{2}} + \frac{\left(O_4 - \frac{O_2+O_6}{2}\right)^2}{\frac{O_2+O_6}{2}} + \frac{\left(O_4 - \frac{O_3+O_5}{2}\right)^2}{\frac{O_3+O_5}{2}} + \frac{\left(O_4 - \frac{O_2+O_3+O_5+O_6}{4}\right)^2}{\frac{O_2+O_3+O_5+O_6}{4}} + \frac{\left(O_4 - \frac{O_1+O_2+O_6+O_7}{4}\right)^2}{\frac{O_1+O_2+O_6+O_7}{4}} + \frac{\left(O_4 - \frac{O_1+O_3+O_5+O_7}{4}\right)^2}{\frac{O_1+O_3+O_5+O_7}{4}} + \frac{\left(O_4 - \frac{O_1+O_2+O_3+O_5+O_6+O_7}{6}\right)^2}{\frac{O_1+O_2+O_3+O_5+O_6+O_7}{6}}\right] < \chi^2_7,$$

we will accept the hypothesis

H_{A10}: $The\ two\ sided\ data\ follows\ arithmetically\ symmetric\ distribution.$

Similar types of hypothesis can be checked for geometrically and harmonically symmetric distributions. For any twin of the triplet of the aforesaid sixteen triplet hypotheses

$$\frac{\chi^2_n}{\chi^2_n} \sim F_{n,n}$$

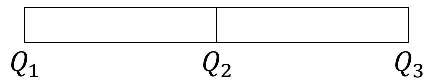
If $F < F_{n,n}$, we will accept the hypothesis concerned to the denominator of the F statistic; otherwise accept that of the numerator of the F statistic. The aforesaid tests can be generalized for Deciles, Percentiles or any quantiles with chi-square test statistic with one or more than one degree of freedom.

3 Graphical Representations for Assessing the type of Symmetricity

We may use three types of Box Plots named Arithmetic Box Plot, Geometric Box Plot and Harmonic Box Plot.

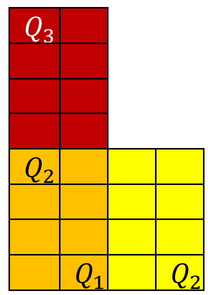
3.1 Box Plot for Quartiles

Now for Arithmetic Symmetric Distribution the Arithmetic Box Plot will be of the following type



to check $Q_3 - Q_2 = Q_2 - Q_1$.

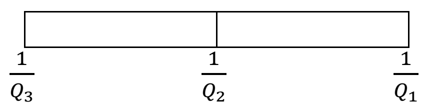
In case of, Geometrically Symmetric Distribution the Geometric Box Plot will be of the following type



to verify

$$\frac{Q_3}{Q_2} = \frac{Q_2}{Q_1}.$$

For Harmonically Symmetric Distribution the Harmonic Box Plot will be of the following type



iff

$$\frac{1}{Q_2} - \frac{1}{Q_3} = \frac{1}{Q_1} - \frac{1}{Q_2}.$$

4. Outlier Detection

4.1 Outlier Detection for Quartiles

In case of arithmetically symmetric or normal distribution we term one observation will be an outlier if it falls out of the following fences:

$$\text{Lower Fence} = Q_1 - 1.5(Q_3 - Q_1), \text{ Upper Fence} = Q_3 + 1.5(Q_3 - Q_1).$$

For geometrically symmetric or normal distribution we term one observation will be treated as outlier if it falls out of the following fences:

$$\text{Lower Fence} = 2\left(\frac{Q_3}{Q_1}\right), \text{ Upper Fence} = \frac{1}{2}\left(\frac{Q_3}{Q_1}\right)$$

Due to harmonically symmetric or normal distribution one observation will be an outlier if it cannot maintain the following limits:

$$\text{Lower Fence} = \frac{1}{Q_3} - 1.5\left(\frac{1}{Q_1} - \frac{1}{Q_3}\right), \text{ Upper Fence} = \frac{1}{Q_1} + 1.5\left(\frac{1}{Q_1} - \frac{1}{Q_3}\right).$$

4.2 Outlier Detection for Octiles

In case of arithmetically symmetric or normal distribution we term one observation will be an outlier if it falls out of the following fences:

$$\text{Lower Fence} = O_1 - \frac{5}{6}(O_7 - O_1), \text{ Upper Fence} = O_7 + \frac{5}{6}(O_7 - O_1).$$

For geometrically symmetric or normal distribution we term one observation will be treated as outlier if it falls out of the following fences:

$$\text{Lower Fence} = 2\left(\frac{O_3}{O_1}\right), \text{ Upper Fence} = \frac{1}{2}\left(\frac{O_3}{O_1}\right)$$

Due to harmonically symmetric or normal distribution one observation will be an outlier if it cannot maintain the following limits:

$$\text{Lower Fence} = \frac{1}{O_7} - \frac{8}{6}\left(\frac{1}{O_1} - \frac{1}{O_7}\right), \text{ Upper Fence} = \frac{1}{O_1} + \frac{8}{6}\left(\frac{1}{O_1} - \frac{1}{O_7}\right).$$

5. Statistical Methods and Methodologies for detecting the type of Normality

5.1 Quantiles based assessment

For a set of data $x_{1,}x_2$ **,…,** x_n as a random sample of size n from a population, we order the data $x_{(1),}x_{(2)}$ **,…,** $x_{(n)}$ and find the quantiles $q_{(1)}, q_{(2)}, \dots, q_{(k).}$ If the successive differences $q_{(2)-}\ q_{(1)}$, $q_{(3)-}\ q_{(2),}$**,…,** $q_{(k)-}\ q_{(k-1)}$ are constant, then the data are said to be arithmetically symmetric. As for example, for octiles we have the following arithmetic relations

$O_2 = \frac{O_1+O_3}{2} \Rightarrow O_2 - O_1 = O_3 - O_2 \dots (i)$, $O_3 = \frac{O_2+O_4}{2} \Rightarrow O_3 - O_2 = O_4 - O_3 \dots (ii)$, $O_4 = \frac{O_3+O_5}{2} \Rightarrow O_4 - O_3 = O_5 - O_4 \dots (iii)$, $O_5 = \frac{O_4+O_6}{2} \Rightarrow O_5 - O_4 = O_6 - O_5 \dots (iv)$, $O_6 = \frac{O_5+O_7}{2} \Rightarrow O_6 - O_5 = O_7 - O_6 \dots (v)$

From $(i) - (v)$ we get the constant arithmetic successive octiles difference relation for arithmetic octiles as

$$O_2 - O_1 = O_3 - O_2 = O_4 - O_3 = O_5 - O_4 = O_6 - O_5 = O_7 - O_6.$$

The data are said to be geometrically symmetric if the successive ratios $\frac{q_{(2)}}{q_{(1)}}$**,** $\frac{q_{(3)}}{q_{(2)}}$**,…,** $\frac{q_{(n)}}{q_{(n-1)}}$ are constant since the following geometric relations are carried by the octiles.

$O_2 = \sqrt{O_1 O_3} \Rightarrow \frac{O_2}{O_1} = \frac{O_3}{O_2} \dots (vi)$, $O_3 = \sqrt{O_2 O_4} \Rightarrow \frac{O_3}{O_2} = \frac{O_4}{O_3} \dots (vii)$, $O_4 = \sqrt{O_3 O_5} \Rightarrow \frac{O_4}{O_3} = \frac{O_5}{O_4} \dots (viii)$, $O_5 = \sqrt{O_4 O_6} \Rightarrow \frac{O_5}{O_4} = \frac{O_6}{O_5} \dots (ix)$, $O_6 = \sqrt{O_5 O_7} \Rightarrow \frac{O_6}{O_5} = \frac{O_7}{O_6} \dots (x)$

From $(vi) - (ix)$ we get the constant geometric successive octiles ratio relation for geometric octiles as

$$\frac{O_2}{O_1} = \frac{O_3}{O_2} = \frac{O_4}{O_3} = \frac{O_5}{O_4} = \frac{O_6}{O_5} = \frac{O_7}{O_6}.$$

The successive differences of the ratios $\frac{1}{q_{(1)}} - \frac{1}{q_{(2)}}, \frac{1}{q_{(2)}} - \frac{1}{q_{(3)}}, \dots, \frac{1}{q_{(n-1)}} - \frac{1}{q_{(n)}}$ are constant. For octiles , the following harmonic relations are evident

$O_2 = \frac{2}{\frac{1}{Q_1}+\frac{1}{Q_3}} \Rightarrow \frac{1}{Q_1} - \frac{1}{Q_2} = \frac{1}{Q_2} - \frac{1}{Q_3} \ldots (xi),\ O_3 = \frac{2}{O_2^{-1}+O_4^{-1}} \Rightarrow \frac{1}{Q_2} - \frac{1}{Q_3} = \frac{1}{Q_3} - \frac{1}{Q_4} \ldots (xii),$
$O_4 = \frac{2}{O_3^{-1}+O_5^{-1}} \Rightarrow \frac{1}{Q_3} - \frac{1}{Q_4} = \frac{1}{Q_4} - \frac{1}{Q_5} \ldots (xiii), O_5 = \frac{2}{O_4^{-1}+O_6^{-1}} \Rightarrow \frac{1}{Q_4} - \frac{1}{Q_5} = \frac{1}{Q_5} - \frac{1}{Q_6} \ldots (xiv), O_6 = \frac{2}{\frac{1}{Q_5}+\frac{1}{Q_7}} \Rightarrow \frac{1}{Q_5} - \frac{1}{Q_6} = \frac{1}{Q_6} - \frac{1}{Q_7} \ldots (xv)$

From $(xi) - (xv)$ we get the constant harmonic successive inverse octiles difference relation for harmonic octiles as

$$\frac{1}{Q_1} - \frac{1}{Q_2} = \frac{1}{Q_2} - \frac{1}{Q_3} = \frac{1}{Q_3} - \frac{1}{Q_4} = \frac{1}{Q_4} - \frac{1}{Q_5} = \frac{1}{Q_5} - \frac{1}{Q_6} = \frac{1}{Q_6} - \frac{1}{Q_7}.$$

6. Comparison among the arithmetically, geometrically and harmonically symmetric or normal distributions

For arithmetically symmetric or normal distribution, the first four moments are
$\mu = E(x) = \sum xp(x), \sigma^2 = \sum (x-\mu)^2 p(x)$

For geometrically symmetric or normal distribution, the first two moments are

$$\hat{\mu} = G(x) = (x_1 x_2 \ldots x_n)^{\frac{1}{n}}, \sigma = \left(\frac{x_2}{x_1}\frac{x_3}{x_2}\ldots\frac{x_n}{x_{n-1}}\right)^{\frac{1}{n-1}} = \left(\frac{x_n}{x_1}\right)^{\frac{1}{n-1}}$$

For the grouped data,

$$\hat{\mu} = G(x) = x_1^{P(x_1)} x_2^{P(x_2)} \ldots x_n^{P(x_n)},$$

$$\hat{\sigma} = \left(\frac{x_2^{P(x_2)}}{x_1^{P(x_1)}}\frac{x_3^{P(x_3)}}{x_2^{P(x_2)}}\ldots\frac{x_n^{P(x_n)}}{x_{n-1}^{P(x_{n-1})}}\right)^{\frac{1}{n-1}} = \left(\frac{x_n^{P(x_n)}}{x_1^{P(x_1)}}\right)^{\frac{1}{n-1}}$$

For Harmonically symmetric or normal distribution, the first two moments are

$$\hat{\mu} = H(x) = \frac{n}{x_1^{-1} + x_2^{-1} + \cdots + x_n^{-1}}$$

$$\widehat{\sigma^2} = \frac{n}{\left(x_1^{-1} - \frac{n}{x_1^{-1}+x_2^{-1}+\cdots+x_n^{-1}}\right)^2 + \left(x_2^{-1} - \frac{n}{x_1^{-1}+x_2^{-1}+\cdots+x_n^{-1}}\right)^2 + \cdots + \left(x_n^{-1} - \frac{n}{x_1^{-1}+x_2^{-1}+\cdots+x_n^{-1}}\right)^2}$$

7. Existing examples of some traditional distributions along with the less local locations

7.1 Geometric Distribution

Geometric distributions is a discrete distribution that present the discrete waiting time preceding 1st success. The traditional expected waiting time is proportional to the probability of success which means that if the probability of success is greater the traditional mean waiting time is greater and vice versa. This is not logically correct since if probability of success is greater, naturally the waiting time to get 1st success will be relative less than the waiting time preceding 1st success for the probability of success with less value. The probability mass function of the Geometric distribution is as below:

$$P(x) = (1-p)^{x-1}p, \qquad x = 1, 2, \ldots$$

$$E(x) = \frac{p}{1-p}.$$

If we consider the probability mass function as the general term of the several frequency terms of a Geometric sequence of the following form:

$$P(x_1) = p, P(x_2) = (1-p)p, P(x_3) = (1-p)^2p, P(x_4) = (1-p)^3p, \dots$$
$$\therefore \widehat{GM(x)} = {x_1}^{P(x_1)}{x_2}^{P(x_2)} \dots . {x_n}^{P(x_n)} = 1^p \times 2^{(1-p)p} \times 3^{(1-p)^2p} \times \dots . \times n^{(1-p)^{(n-1)}p}$$
$$= (1.2 \dots . n)^p \times (2.3 \dots n)^{(1-p)} \times (3.4 \dots n)^{(1-p)} \times \dots \times n^{(1-p)}$$
$$\therefore \widehat{GM(x)} = 2^p(2.3 \dots n)^{(1-p)} \times 3^p(3.4 \dots n)^{(1-p)} \times \dots \times n^p n^{(1-p)}$$

Since, each of the n-terms of the right hand side of the aforesaid equation decreases as p increases. Therefore, if the probability of successes, p, increases in the Geometric distribution, the Geometric average decreases which is logical since if probability of successes increases, naturally the average waiting time will decrease. If n increases, Geometric distribution tend to symmetric one, so let us assume value of n with a small one.

If $p = 0.2$ and n = 3, then the sample Geometric Mean waiting time,
$\widehat{GM(x)} = 1^p \times 2^{p(1-p)} \times 3^{p(1-p)^2} = 1^{0.2} \times 2^{0.2(1-0.2)} \times 3^{0.2(1-0.2)^2} = 1.29$,
but the sample Arithmetic Mean waiting time, $E(x) = \frac{p}{1-p} = \frac{0.2}{1-0.2} = 0.25$.

If $p = 0.8$ and n= 3, then the sample Geometric Mean waiting time
$\widehat{GM(x)} = 1^p \times 2^{p(1-p)} \times 3^{p(1-p)^2} = 1^{0.8} \times 2^{0.8(1-0.8)} \times 3^{0.8(1-0.8)^2} = 1.16$
and the population Arithmetic Mean waiting time = E(x) = $\frac{p}{1-p} = \frac{0.8}{1-0.8} = 4$.

So, $\frac{\widehat{G(x;p=0.2)}}{\widehat{G(x;p=0.8)}} = \frac{1.29}{1.16} = 1.11$ whereas $\frac{E(x;p=0.2)}{E(x;p=0.8)} = \frac{0.25}{2} = 0.0625$ which means that in Geometric Distribution, considering Geometric Mean if probability of success increases, waiting time decreases whereas for Arithmetic Mean if probability of success increases, waiting time increases. Therefore, in Geometric distribution, Geometric Mean gives logically correct answer than Arithmetic Distribution since for higher value of probability of successes, the waiting time be relatively lower. Moreover, it is seen here that for p = 0.2, the Arithmetic Mean is 0.25, where are average waiting time for getting a success can never be less than 1 since least number of trial or success is 1.

7.2 Negative Binomial Distribution

Negative Binomial distributions is a discrete distribution that present the discrete waiting time preceding *r*th success. The traditional expected waiting time is proportional to the probability of success which means that if the probability of success is greater the traditional mean waiting time is greater and vice versa. This is not logically correct since if probability of success is greater, naturally the waiting time to get *r*th success will be relative less than the waiting time preceding *r*th success for the probability of success with less value. The probability mass function of the Negative Binomial distribution is as

$$P(x) = \binom{x-1+r-1}{r-1}(1-p)^{x-1}p^r, \qquad x = 1, 2, \dots$$
$$E(x) = \frac{rp}{1-p}.$$

If we consider the probability mass function as the general term of the several frequency terms of a Geometric sequence of the following form:

$$P(x_1) = p^r, P(x_2) = r(1-p)p^r, P(x_3) = \frac{r(r+1)}{2}(1-p)^2p^r, P(x_4) = \frac{r(r+1)(r+2)}{6}(1-p)^3p^r, \dots$$
$$\therefore \widehat{GM(x)} = {x_1}^{P(x_1)}{x_2}^{P(x_2)} \dots . {x_n}^{P(x_n)} = 1^{p^r} \times 2^{r(1-p)p^r} \times 3^{\frac{r(r+1)}{2}(1-p)^2p^r} \times \dots . \times n^{\frac{r(r+1)(r+2)\dots(r+n-2)}{n-1}(1-p)^{n-1}p^r}$$

$$= (1.2 \ldots . n)^{p^r} \times (2.3 \ldots n)^{r(1-p)} \times (3.4 \ldots n)^{(r+1)(1-p)} \times \ldots \times n^{(r+n-2)(1-p)}$$
$$\therefore \widehat{GM(x)} = 2^{p^r}(2.3 \ldots n)^{r(1-p)} \times 3^{p^r}(3.4 \ldots n)^{(r+1)(1-p)} \times \ldots \times n^{p^r} n^{(r+n-2)(1-p)}$$

Since, each of the n-terms of the right hand side of the aforesaid equation decreases as p increases. Therefore, if the probability of successes, p, increases in the Negative Binomial distribution, the Geometric average decreases which is logical since if probability of successes increases, naturally the average waiting time will decrease. If n or r increases, Negative Binomial distribution tend to symmetric one, so let us assume value of both n and with small ones. If $p = 0.2$ and n = 5, r = 2 then the sample Negative Binomial Mean waiting time,

$\widehat{GM(x)}$= $1^{p^r} \times 2^{r(1-p)p^r} \times 3^{\frac{r(r+1)}{2}(1-p)^2 p^r} \times 4^{\frac{r(r+1)(r+2)}{6}(1-p)^3 p^r} \times 5^{\frac{r(r+1)(r+2)(r+3)}{24}(1-p)^4 p^r}$

= $1^{(0.2)^2} \times 2^{2(1-0.2)0.2^2} \times 3^{\frac{2(2+1)}{2}(1-0.2)^2 0.2^2} \times 4^{\frac{2(2+1)(2+2)}{6}(1-0.2)^3 0.2^2} \times 5^{\frac{2(2+1)(2+2)(2+3)}{24}(1-0.2)^4 0.2^2}$=1.45,

but the sample Arithmetic Mean waiting time, E(x)= $\frac{rp}{1-p}$= $\frac{2*0.2}{1-0.2}$ = 0.50.

If $p = 0.8$ and n= 5, r = 2 then the sample Negative Binomial Mean waiting time

$\widehat{GM(x)}$ = $1^{p^r} \times 2^{r(1-p)p^r} \times 3^{\frac{r(r+1)}{2}(1-p)^2 p^r} \times 4^{\frac{r(r+1)(r+2)}{6}(1-p)^3 p^r} \times 5^{\frac{r(r+1)(r+2)(r+3)}{24}(1-p)^4 p^r}$

= $1^{(0.8)^2} \times 2^{2(1-0.8)0.8^2} \times 3^{\frac{2(2+1)}{2}(1-0.8)^2 0.8^2} \times 4^{\frac{2(2+1)(2+2)}{6}(1-0.8)^3 0.8^2} \times 5^{\frac{2(2+1)(2+2)(2+3)}{24}(1-0.8)^4 0.8^2}$

=1.35,

and the population Arithmetic Mean waiting time = E(x) = $\frac{rp}{1-p} = \frac{2*0.8}{1-0.8} = 8$. So, $\frac{\widehat{G(x;p=0.2)}}{\widehat{G(x;p=0.8)}} = \frac{1.45}{1.35} = 1.07$ whereas $\frac{E(x;p=0.2)}{E(x;p=0.8)} = \frac{0.50}{8}$ = 0.0625 which means that in Negative Binomial Distribution, considering Geometric Mean if probability of success increases, waiting time decreases whereas for Arithmetic Mean if probability of success increases, waiting time increases. Therefore, in Negative Binomial distribution, Geometric Mean gives logically correct answer than Arithmetic Distribution since for higher value of probability of successes, the waiting time be relatively lower. Moreover, it is seen here that for r = 2 and p = 0.2, the Arithmetic Mean is 0.50, where are average waiting time for getting a success can never be less than 1 since least number of trial or success is 1. If r accelerated increases compared to n, the effect of probability of success drastically decreases and the ratio of two geometric means behave opposite behavior just like that of two arithmetic means. This is due to the reason is that, if r increases , the geometric distribution tend to normal distribution that require arithmetic mean to be more appropriate for each of the two means.

7.3 Exponential Distribution

For the simplest form of the exponential distribution with the probability density function as $f(x) = e^{-x}, x > 0$, $E(x) = \int_0^\infty xe^{-x}dx = 1$ which means that the arithmetic mean or average waiting time is 1 time unit whereas $\int_0^1 e^{-x}dx = \frac{e-1}{e} = 0.632$. So, for the exponential distribution with the aforesaid pdf, we are assuming that $E(x) = 1$ as the center under which there are 63 % observations. The distribution's Geometric mean is

$G(x) = e^{\int_0^\infty e^{-x}\ln(x)dx}$ =0.56 which means that the geometric mean or average waiting time is 0.56 unit time whereas $\int_0^{0.56} e^{-x}dx = 0.43$. So, geometric mean $G(x)$ =0.56 as the center leaves 43 % observations to its left side. Since, $\int_0^{0.695} e^{-x}dx = 0.50$ so the median of the referred exponential distribution is 0.695. The additive or arithmetic or horizontal or linear drift from median toward the arithmetic mean $E(x)$ is 1.0 - 0.695 =0.305 and that to the geometric mean $G(x)$ is 0.695 – 0.56 = 0.135. The multiplicative or geometric or vertical or scale drift from median to the arithmetic mean point is $\frac{1}{0.695} = 1.44$ whereas that of geometric mean to the median point is $\frac{0.695}{0.56} = 1.24$. So, for the aforesaid distribution with the following inequality $GM = 0.56 < Median = 0.695 < AM = 1.0$, we come to know that median is 1.44 times of geometric mean whereas arithmetic means is 1.24 times of median. There are several reasons to warrant geometric mean = 0.56 to be more appropriate as the center of the exponential distribution compared to the arithmetic mean = 1.0. At first, a better center should be median or be as close as to the median of a distribution especially for an asymmetric one. Here the exponential distribution has asymmetric form and exponential or geometric pattern. So, geometric mean is more appropriate to present the center of the distribution. Secondly, geometric mean has 43 % observation to the left hand side of the distribution whereas arithmetic mean has 37 % observation to the right hand side of the distribution. So, geometric mean is more close (with distance = 50% - 43% = 7%) to the 50th percentile of the distribution compared to the arithmetic mean (with distance 63% - 50% = 50% - 37% = 13%). Thirdly, within first (0.695-0 =) tiny interval of (0, 0.695) of length 0.695 of the horizontal axis of the distribution, 50% observations lie whereas within the rest of the interval [0.695, ∞) of infinite length of the real line axis of the same, 50% observations belong. Since the first interval not only includes the modal value (mode is close to 0) but also the 50 % of the data within very short interval, the center of the distribution should belong to this first interval rather than the second interval. Geometric mean = 0.56 belong to the first interval but arithmetic mean = 1 does not belong to the first interval. So, it is geometric mean, not arithmetic mean, which is more credit worthy to present the locality of the exponential distribution. Fourthly, for the first interval, the height if density is the highest compared to that for the second interval with relatively consistent flat tail. A center of a distribution belong to that interval of the x axis for which the distribution has higher height of the density. For the various specifications of the parameter of the exponential distributions the arithmetic mean waiting times are $E(x) = \int_0^\infty xe^{-x}dx = 1$, $E(x) = \int_0^\infty x2e^{-2x}dx = \frac{1}{2} = 0.50$ and $E(x) = \int_0^\infty x4e^{-4x}dx = \frac{1}{4} = 0.25$ and the geometric mean waiting times are $G(x) = e^{\int_0^\infty e^{-x}\ln(x)dx} =$ 0.56, $G(x) = e^{\int_0^\infty 2e^{-2x}\ln(x)dx} = 0.28$ and $G(x) = e^{\int_0^\infty 4e^{-4x}\ln(x)dx} = 0.14$ along with the probability density functions $f(x) = e^{-x}, x > 0$, $f(x) = 2e^{-2x}, x > 0$ and $f(x) = 4e^{-4x}, x > 0$ respectively. So, $E(x, \lambda = 1) = 2E(x, \lambda = 2) = 4E(x, \lambda = 4)$ and $G(x, \lambda = 1) = 2G(x, \lambda = 2) = 4G(x, \lambda = 4)$.

7.4 Gamma Distribution

For the simplest form of the gamma distribution (the form other than exponential distribution) with the probability density function as $f(x) = xe^{-x}, x > 0$, $E(x) = \int_0^\infty x^2e^{-x}dx = 2$ which means that the arithmetic mean or average waiting time is 2 time unit whereas $\int_0^2 xe^{-x}dx = 0.59$. So, for the exponential distribution with the aforesaid pdf, we are assuming that $E(x) = 2$ as the center under which there are 59 %

observations. The distribution's Geometric mean is $G(x) = e^{\int_0^{\infty} xe^{-x}\ln(x)dx} = 1.53$ which means that the geometric mean or average waiting time is 1.53 unit time whereas $\int_0^{1.53} xe^{-x}dx = 0.45$. Therefore, geometric mean $G(x)$ =1.53 as the center leaves 45 % observations to its left side. Since, $\int_0^{0.695} e^{-x}dx = 0.50$ so the median of the referred exponential distribution is 1.68. The additive or arithmetic or horizontal or linear drift from median toward the arithmetic mean $E(x)$ is 2.0 - 1.68 =0.32 and that to the geometric mean $G(x)$ is 1.68 – 1.53 = 0.15. The multiplicative or geometric or vertical or scale drift from median to the arithmetic mean point is $\frac{2}{1.68} = 1.2$ whereas that of geometric mean to the median point is $\frac{1.68}{1.53} = 1.1$. So, for the aforesaid distribution with the following inequality $GM = 1.53 < Median = 1.68 < AM = 2.00$, we come to know that median is 1.10 times of geometric mean whereas arithmetic means is 1.20 times of median. There are several reasons to warrant geometric mean = 1.53 to be more appropriate as the center of the exponential distribution compared to the arithmetic mean = 2.0. At first, the gamma distribution has asymmetric form and exponential or geometric pattern. So, geometric mean is more appropriate to present the center of the distribution. Secondly, geometric mean has 45 % observation to the left hand side of the distribution whereas arithmetic mean has 41 % observation to the right hand side of the distribution. So, geometric mean is more close (with distance = 50% - 45% = 5%) to the 50th percentile of the distribution compared to the arithmetic mean (with distance 59% - 50% = 50% - 41% = 9%). Thirdly, within first tiny interval of (0, 1.68) of length (1.68 – 0=) 1.68 of the horizontal axis of the distribution, 50% observations lie whereas within the rest of the interval $[1.68, \infty)$ of infinite length of the real line axis of the same, 50% observations belong. Since the first interval not only includes the modal value (mode is close to 0 and 1) but also the 50 % of the data within very short interval, the center of the distribution should belong to this first interval rather than the second interval. Geometric mean = 1.53 belong to the first interval but arithmetic mean = 2 does not belong to the first interval. So, it is geometric mean, not arithmetic mean, which is more credit worthy to present the locality of the gamma distribution. Fourthly, for the first interval, the height if density is the highest compared to that for the second interval with relatively consistent flat tail. A center of a distribution belong to that interval of the x axis for which the distribution has higher height of the density. For the various specifications of the parameter of the gamma distributions the arithmetic mean waiting times are $E(x) = \int_0^{\infty} x^2e^{-x}dx = 2$, $E(x) = \int_0^{\infty} x\frac{xe^{-2x}2^2}{1!}dx = 1$ and $E(x) = \int_0^{\infty} x\frac{xe^{-4x}4^2}{1!}dx = 0.50$ and the geometric mean waiting times are $G(x) = e^{\int_0^{\infty} xe^{-x}\ln(x)dx} = 1\ .53$, $G(x) = e^{\int_0^{\infty}\frac{xe^{-2x}2^2}{1!}\ln(x)dx} = 0.76$ and $G(x) = e^{\int_0^{\infty}\frac{xe^{-4x}4^2}{1!}\ln(x)dx} = 0.38$ along with the probability density functions $f(x) = xe^{-x}, x > 0$, $f(x) = \frac{xe^{-2x}2^2}{1!}, x > 0$ and $f(x) = \frac{xe^{-2x}2^2}{1!}, x > 0$ respectively. So, $E(x, n = 2, \lambda = 1) = 2E(x, n = 2, \lambda = 2) = 4E(x, n = 2, \lambda = 4) = 2$ and $G(x, n = 2, \lambda = 1) = 2G(x, n = 2, \lambda = 2) = 4G(x, n = 2, \lambda = 4) = 1.53$. So, $\frac{E(x,n=2,\lambda=1)}{G(x,n=2,\lambda=1)} = \frac{2}{1.53} \approx \frac{4}{3}$. The arithmetic mean waiting times are $E(x) = \int_0^{\infty} x\frac{x^2e^{-x}}{2!}dx = 3$, $E(x) = \int_0^{\infty} x\frac{x^2e^{-2x}2^3}{2!}dx = 1.5$ and $E(x) = \int_0^{\infty} x\frac{x^2e^{-4x}4^3}{2!}dx = 0.75$ and the geometric mean waiting times are $G(x) = e^{\int_0^{\infty}\frac{x^2e^{-x}}{2!}\ln(x)dx} = 2\ \ .52$, $G(x) = e^{\int_0^{\infty}\frac{x^2e^{-2x}2^3}{2!}\ln(x)dx} = 1.23$ and $G(x) = e^{\int_0^{\infty}\frac{x^2e^{-4x}4^3}{2!}\ln(x)dx} = 0.63$ along with the probability density functions $f(x) =$

$\frac{x^2 e^{-x}}{2!}, x > 0$, $f(x) = \frac{x^2 e^{-2x} 2^3}{2!}, x > 0$ and $f(x) = \frac{x^2 e^{-4x} 4^3}{2!}, x > 0$ respectively. So, $E(x, n = 3, \lambda = 1) = 2E(x, n = 3, \lambda = 2) = 4E(x, n = 3, \lambda = 4) = 3$ and $G(x, n = 3, \lambda = 1) = 2G(x, n = 3, \lambda = 2) = 4G(x, n = 3, \lambda = 4)$=2.52. So, $\frac{E(x,n=3,\lambda=1)}{G(x,n=3,\lambda=1)} = \frac{3}{2.52} \approx \frac{6}{5}$. The arithmetic mean waiting times are $E(x) = \int_0^\infty x \frac{x^{19} e^{-x}}{19!} dx = 20$, $E(x) = \int_0^\infty x \frac{x^{19} e^{-2x} 2^{20}}{19!} dx = 10$ and $E(x) = \int_0^\infty x \frac{x^{19} e^{-4x} 4^{20}}{19!} dx = 5$ and the geometric mean waiting times are $G(x) = e^{\int_0^\infty \frac{x^9 e^{-x}}{9!} \ln(x) dx} = 9.50$, $G(x) = e^{\int_0^\infty \frac{x^9 e^{-2x} 2^{10}}{9!} \ln(x) dx} = 4.75$ and $G(x) = e^{\int_0^\infty \frac{x^9 e^{-4x} 4^{10}}{9!} \ln(x) dx} = 2.38$ along with the probability density functions $f(x) = \frac{x^2 e^{-x}}{2!}, x > 0$, $f(x) = \frac{x^2 e^{-2x} 2^3}{2!}, x > 0$ and $f(x) = \frac{x^2 e^{-4x} 4^3}{2!}, x > 0$ respectively. So, $E(x, n = 10, \lambda = 1) = 2E(x, n = 10, \lambda = 2) = 4E(x, n = 10, \lambda = 4) = 10$ and $G(x, n = 10, \lambda = 1) = 2G(x, n = 10, \lambda = 2) = 4G(x, n = 10, \lambda = 4)$ = 9.50. So, $\frac{E(x,n=10,\ \lambda=1)}{G(x,n=10,\ \lambda=1)} = \frac{10}{9.50} \approx \frac{20}{19}$. Therefore, as shape parameter (n) increases the ratio of the arithmetic mean and the geometric mean tends to 1. This is true since if the shape parameter increases, the Gamma distribution tends to normal distribution. Moreover, $E(x) = \int_0^\infty x \frac{x^9 e^{-25x} 25^{10}}{9!} dx = 0.40$, $G(x) = e^{\int_0^\infty \frac{x^9 e^{-30x} 30^{10}}{9!} \ln(x) dx} = 0.40$. So, $E(x, n = 10, \lambda = 25) = G(x, n = 10, \lambda = 25) = 0.40$. If the value of scale parameter (λ) increases, the exponential behavior of Gamma distribution reduces or tends to arithmetic behavior. This property is consistent to the property of the Gamma distribution since Gamma distribution approaches to Normal distribution as scale parameter is sufficiently large.

7.5 Chi-square Distribution

All the new properties developed in this paper for the Gamma distribution are also true for the Chi-square distribution since chi-square is a special type of Gamma distribution.

7.6 Weibull distribution

For the form of the exponential distribution with the probability density function as $f(x) = \frac{a}{b}\left(\frac{x}{b}\right)^{a-1} e^{-\left(\frac{x}{b}\right)^a}, x \geq 0$, $E(x,\ \mathrm{a} = \mathrm{b} = 2) = \int_0^\infty x \frac{2}{2}\left(\frac{x}{2}\right)^{2-1} e^{-\left(\frac{x}{2}\right)^2} dx = 1$, $E(x, \mathrm{a} = \mathrm{b} = 10) = 9.51$, $E(x, \mathrm{a} = \mathrm{b} = 100) = 62.42$, $E(x, \mathrm{a} = 2, \mathrm{b} = 10) = 39.89$, $E(x, \mathrm{a} = 2, \mathrm{b} = 100) = 32.12$, $E(x, \mathrm{a} = 10, \mathrm{b} = 2) = 1.98$, $E(x, \mathrm{a} = 10, \mathrm{b} = 100) = 56.05$, $E(x, \mathrm{a} = 100, \mathrm{b} = 2) = 1.99$, $E(x, \mathrm{a} = 100, \mathrm{b} = 10) = 6.24$. If *a* and *b* contemporarily increases to same extent, E(x) increases. But if a<b, E(x) increases and the rate of increase of E(x) decreases as *b* increases. But for a>b, E(x) decreases, whereas the rate of decrease of E(x) remain same as *a* increases for smaller *b* but E(x) increases when *a* increases with greater *b*. $G(x,\ \mathrm{a} = \mathrm{b} = 2) = \int_0^\infty (lnx) \frac{2}{2}\left(\frac{x}{2}\right)^{2-1} e^{-\left(\frac{x}{2}\right)^2} dx = 0.40$, $G(x, \mathrm{a} = \mathrm{b} = 10) = 2.24$, $G(x, \mathrm{a} = \mathrm{b} = 100) = 4.60$, $G(x, \mathrm{a} = 2, \mathrm{b} = 10) = 2.01$, $G(x, \mathrm{a} = 2, \mathrm{b} = 100) = 4.32$, $G(x, \mathrm{a} = 10, \mathrm{b} = 2) = 0.64$, $G(x, \mathrm{a} = 10, \mathrm{b} = 100) = 4.55$, $G(x, \mathrm{a} = 100, \mathrm{b} = 2) = 0.69$, $G(x, \mathrm{a} = 100, \mathrm{b} = 10) = 2.30$. If *a* and *b* contemporarily increases to same extent, G(x) increases. But if a<b, E(x) increases and the rate of increase of E(x) increases as *b* increases. But for a>b, E(x) decreases, whereas the rate of decrease of E(x) remain same as *a* increases for smaller *b* but E(x) increases when *a* increases with greater *b*.

$\frac{E(x,a=b=2)}{G(x,a=b=2)} = \frac{1}{0.40} = 2.50 = \frac{100}{40}, \frac{E(x,a=b=10)}{G(x,a=b=10)} = \frac{9.51}{2.24} = 4.25 = \frac{100}{24}, \frac{E(x,a=b=100)}{G(x,a=b=100)} = \frac{62.42}{4.60} = 13.57 = \frac{100}{7}, \frac{E(x,a=2,b=10)}{G(x,a=2,b=10)} = \frac{39.89}{2.01} = 19.85 = \frac{100}{5}, \frac{E(x,a=2,b=100)}{G(x,a=2,b=100)} = \frac{32.12}{4.32} = 7.44 = \frac{100}{13}, \frac{E(x,a=10,b=2)}{G(x,a=10,b=2)} = \frac{1.98}{0.64} = 3.09 = \frac{100}{32}, \frac{E(x,a=10,b=100)}{G(x,a=10,b=100)} = \frac{56.05}{4.55} = 12.32 = \frac{100}{8}, \frac{E(x,a=100,b=2)}{G(x,a=100,b=2)} = \frac{1.99}{0.69} = 2.88 = \frac{100}{35}, \frac{E(x,a=100,b=10)}{G(x,a=100,b=10)} = \frac{6.24}{2.30} = 2.71 = \frac{100}{37}$. When a and b both increases equally, the ratio of arithmetic mean and geometric mean acceleratory increases more and more. If $a < b$, then for smaller value of a, the ratio decreases if b increases. If $a > b$, then for bigger value of a, the ratio remains almost same if b increases. If a is in between smaller and bigger value, then for medium value of a, the ratio increases if b increases. For smaller a, the ratio is bigger. If a increases small value to larger ones, the ratio become dropping down since for larger a the exponential behavior changes to normal one.

7.7 Beta distribution

For the form of the exponential distribution with the probability density function as

$$f(x) = \frac{x^{m-1}(1-x)^{n-1}}{\beta(m,n)}, 0 \le x \le 1,$$

$E(x, m = n) = \frac{m}{m+n} = 0.50$, $E(x, m = 1, n = 2) = \frac{1}{1+2} = 0.33$, $E(x, m = 2, n = 3) = \frac{2}{2+3} = 0.40$, $E(x, m = 2, n = 10) = \frac{2}{2+10} = 0.17$, $E(x, m = 10, n = 2) = \frac{10}{10+2} = 0.83$, $E(x, m = 20, n = 2) = \frac{20}{20+2} = 0.91$, $E(x, m = 20, n = 5) = \frac{20}{20+5} = 0.80$, $E(x, m = 2, n = 20) = \frac{2}{2+20} = 0.09$, $E(x, m = 5, n = 20) = \frac{5}{5+20} = 0.20$, $G(x, m = n = 2) = \int_0^{\infty} (lnx) \frac{x^{2-1}(1-x)^{2-1}}{\beta(2,2)} dx = 0.43$, $G(x, m = n = 10) = 0.49$, $G(x, m = n = 13) = 0.35$, $G(x, m = 1, n = 2) = 0.22$, $G(x, m = 2, n = 3) = 0.34$, $G(x, m = 2, n = 10) = 0.13$, $G(x, m = 10, n = 2) = 0.83$, $G(x, m = 20, n = 2) = 0.91$, $G(x, m = 20, n = 5) = 0.77$, $G(x, m = 2, n = 20) = 0.07$, $G(x, m = 5, n = 20) = 0.18$. If m and n contemporarily increases to same extent, E(x) remains same ie, E(x) is invariant to the same extent of both parameter. But a center should be sensitive to the change of various realizations of its parameters. But G(x) is sensitive to the change of both parameters to same extent. However, if m and m both remain small, the ratio of arithmetic mean and geometric mean remain greater. But this ratio tend to 1 when any of the shape parameters tend to large. It happens since for any of the large parameter Beta distribution tends to Normal and Beta distribution is asymmetric or exponential when the parameters are small.

Conclusion

Since a statistical data analyst does not know the population before carrying out a study, assuming arithmetic normality assumptions is frequently misleading the statistician to obtain the true results. Statisticians should not only base on the three types of means for three types of progressions, but also should know what is core increment and exponent of sequence of the entire data along with its general term which is the driving force of deciding various features of the said distribution.

References

Hogg, Mckean and Craig. Introduction to Mathematical Statistics. 2013. Seventh Edition. Pearson publisher.