International Journal of Contemporary Issues in Social Sciences
Volume 3, Issue 2, 2024 ISSN(P):2959-3808 | 2959-2461

PARAMETER ESTIMATION FOR GAMMA DISTRIBUTIONS BY MOMENT GENERATING FUNCTIONS IN THE PRESENCE OF OUTLIERS

Hafiz Bilal Ahmad[*1], Amir Shahzad[*2], Nazakat Ali[*3]

[*1,*2,*3]Master of Philosophy in Statistics, Department of Mathematics & Statistics, University of Agriculture, Faisalabad

[*1]bilal.uafstat@gmail.com; [*2]amir.statistics07@gmail.com; [*3]nazakatalibandesha@gmail.com

Corresponding Author: *

Received: 28 March, 2024 **Revised:** 28 April, 2024 **Accepted:** 20 May, 2024 **Published:** 28 May, 2024

ABSTRACT

The presence of outliers within datasets can significantly alter their behavior, leading to substantial errors in estimated results. Consequently, accurate parameter estimation necessitates the identification and mitigation of outlier effects. Various methodologies, including Method of Moments Estimation (MME) and Maximum Likelihood Estimation (MLE), are employed to estimate parameters in distributions affected by outliers. This research explores the dynamic behavior of parameters in the presence of outliers and devises strategies to mitigate their influence. Emphasis is placed on understanding the multifaceted impact of outliers on data sets, including their potential to distort formulas, mischaracterize parameters, and skew summary statistics. A simulated study is conducted to assess the robustness of test statistics in outlier detection. The Gamma distribution is examined, particularly in the context of size-biased and area-biased functions, with parameters calculated accordingly. Discordancy tests are employed to identify outliers, with graphical representations aiding in the delineation of scale and shape parameter behaviors. Furthermore, this research extends to the detection and analysis of both single and dual outliers, facilitating a comprehensive understanding of their effects and subsequent recalibration of results.

Keywords: Generalized Gamma Distribution, Probability Density Function, Detect Outliers, Size & Area Biased

INTRODUCTION

variability in the data, experimental error, or instrumental error. Whether to include or exclude outliers depends on the context of the analysis. A single outlier can have a substantial impact on statistical results, making the detection and management of outliers a crucial aspect of data analysis, particularly in finance research.

Detecting outliers is the first step in many data-mining processes. There are various methods for identifying and handling outliers, which can be categorized into univariate and multivariate techniques, as well as parametric and non-parametric methods. Large datasets often contain numerous events, some of which may be outliers. Effective data analysis begins with identifying these outliers. While outliers are sometimes viewed as noise or errors, they can also carry significant information.

Detected outliers are potential sources of anomalous data that can lead to model misspecification, biased parameter estimation, and incorrect results. Therefore, it is crucial to identify them before conducting modeling and analysis.

Currently, the sample mean and sample variance provide accurate estimates of the data's location and distribution, provided that the influence of outliers is minimized. An outlier indicates a deviation that appears to be inaccurate or inconsistent with the rest of the data points.

There are various definitions of outliers, one of which is:

"An observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data" (Barnett and Lewis, 1994).

Outliers are a prominent topic of discussion in the field of statistics. It is rare to encounter a dataset without at least one outlier. Often, outliers are seen as disturbances in the data. However, this research aims to demonstrate that outliers can hold valuable information once they have been properly identified. Outliers can cause deviations in parameters, impacting the accuracy of results. Therefore, it is crucial to detect and understand the effects of these observations.

Edgeworth (1887) discussed types of outliers and their real-life comparisons. Natural phenomena are coded into data from which conclusions are drawn. Experimental data, as noted by Anscombe (1960) and Beckman & Cook (1983), are subject to errors. Identifying these errors is crucial to avoid misleading results. Astronomers since the 18th century suggested removing extreme values to maintain accuracy.

Moment Distribution

In observational studies related to human life, wildlife, plants, insects, and forest populations, there is no well-defined sampling plan. Equal probability sampling is often impossible, leading to biased recorded observations. Ignoring this bias can result in incorrect conclusions. In natural and biological work, observations often fall into non-experimental, unsystematic, and non-random categories. Addressing model selection and data interpretation is crucial. Researchers have proposed ad hoc solutions, such as weighted distributions, to correct bias. Moment distributions are significant in probability and statistics when random sampling is infeasible, especially in environmental and biological studies, to ensure accurate data analysis (Kagan, 2002).

Size Biased Distribution

A one-sided distribution is a special case of weighted distribution, as noted by Fisher (1934), who developed such distributions to address bias in models. These biased distributions occur when observations from a random process do not have an equal probability of being recorded, but are instead recorded according to a weighting function. This concept is particularly relevant in fields like forestry, medical sciences, and psychology. When the probability of selecting an individual in a population is proportional to its magnitude, it is known as length-biased sampling. Consequently, observations selected based on their length result in a length-biased distribution.

An observer records nature's phenomena, noting deterministic patterns. Documented perceptions vary until properly categorized, with weighted and size-biased methods commonly used in research. Weighted distributions, constrained by unity, are prevalent in various models, including likelihood testing and visibility bias in data compilation. Notable works, such as those by Patil and Rao (1978), explore these distributions' implications, particularly in population data and wildlife management. Differences in weighted distributions' values are often subject to hypothesis testing, shedding light on diverse phenomena and informing research methodologies.

Area Biased Distribution

The classification of distributions includes length-biased studies, where sampling is proportional to length, causing bias. Cox (1962) introduced the length-biased distribution, applied notably in biomedical fields like predictive modeling. Works in weighted distributions, stemming from this concept, are termed area-biased distributions. Patill and Rao developed basic distributions and their weighted forms, including beta, gamma, and pareto distributions. Researchers have subsequently applied these weighted forms to lifetime distributions, such as length-biased weighted generalized Rayleigh, Bayes estimation of length-one-sided Weibull, and length-biased beta distributions.

How Size Biased Distribution to identify the outliers:

There are some methods or test statistic to identify the outliers. Some of these tests statistic are given below:

i. The Dixon type test
ii. Likelihood of maximum ratio (MLR) statistic test
iii. Zerbet as well as Nikulin (ZN) test statistic
iv. Shadrokh and Pazira test statistic
v. Lalitha and Kumar test statistic
vi. Gap test statistic
vii. Tietjen-Moore test statistic

Objectives:

- To estimate Gamma Distribution and the parameters when outliers in probability distribution.
- Evaluate the effect of outliers in Gamma Distribution moment distribution.

Literature Review

(Rohlf, 1975) suggest a new method for multivarate outlier dectection named generalized gap test. Observtion based on distribution of edges length of minimum spanning trees. It very diffcult to the observation presence sperated from the main cloud point of multiavte. The separated edge lenth fellow the gamma distribution closely if data in multvarate normal distributed.

(Kimber, 1979) perform three different test for single outlier which apply on gamma sampled data and parameters are unknown, first was principle maximum liklihhod based and the others was transformed based to normal approximation. The performance of the teste are studied when one outlier in the data set.

Kimber (1982) a procedure which sequential for the test of k upper outlier in exponential sample is propsed, K=2,3,4 is under considration and critaical values are tabulated. Existing test are also compared and flexibality is also discussed, low outliers are also considered in study.

Kimber (1983) introducted procedure for testing gamma sampled outliers in which shape paramter is unknown and also scale parameter is unknown. New Critical value table is not necessary. Tests properties of the upper outliers is investigated and low outlier is also study briefly.

(Akinsete et al., 2008) study about the different four parameters of the beta pareto distribution. The beta pareto distrubtion is found to be unimodel or decreasing hazard rate. The study also study about the distribution mean, mean deviation, variance, kurtosis, skewness and relation of these are also under considration. The technique of the maximum liklihood is also proposed and applied on two different data sets.

(Alexander et al., 2012) proposed Generalized Beta Generated(GBG) distribution, sub- model involve the all classical beta-generaters, exponential distrubution and kumaraswamy generated. Under three conditionds there are maximum entropy distribution, which are indecating the skewness of the generator of classical beta is used only to control tail-entropy and addional shape paramters is necessary for addition in the entropy center of the fountain distribution. Without differentiating tail weights this parameter control the skewness. Different derivations for moments, quantils and generating functions. GBG also has tractable properties. Maximum lillihood is used for the estimation of the parameters and new class is applied on the real life data.

(Alizadeh et al., 2013) discussed uniformly minimum variance unbiased(UMVU), percentile (PC), method of maximum liklihood, least square (LSE) also weighted least square (WLSE) probability distribution function estimations also cumulative density function are calculated for the rayleigh distrubution. Given model is useful and effective in model strength also in modelling general lifetime data. This study shows that (MLE) is good than the UMVUE, and UMVUE is good than the others.

(Hariharan et al., 2014) detection of the outlier is one of the important aspect of the data mining which find the observation that behaviour is different from the other observation.Detection of the outlier is also called outlier mining. In this study literature of the outlier and analysis of the outlier is given, different method for the outlier are comparied in this study.

(Tripathi et al., 2014) consider the issue of estimation of the pareto parameters under quadratic loss funtion (QLF) when parameter of the scale is limited. Zidek's (1974), the difference of risk integral expression, kubokawa approch is used for enhanced estimators compared to normal estimators, results are given.

(Gogoi and Das, 2015) this study is for the the comparsion of some statistics empirical power for detection of many outliers are in sample which is exponential distributed under alternative slippage method. Different degree of discordancy parameters are also investgated in this study. The results of the simulation suggest that maximum liklihood is good and rest of other statistics followed by dixon form test in the exponentail sample for the treatment of the upper outliers. Maximum liklihood is also better than the Lalitha and kumar (2012) test.

(Nasiri, 2016) lomax distribution in the occurrence of outliers addressed these problems in estimation of the parameter of $R = P(y < x)$ using help of two methods i.e. moments estimation and maximum likelihood method also find their joint

distribution and, in both cases, when λ is known and in case of unknown λ the MLE is used for function of R when K-Outliers are found.

(Jones et al., 2014) applies the generalized beta of the second kind (GB2) distribution to English hospital inpatient cost data. Using a quasi-experimental design, it compares the GB2 with nested and limiting cases, finding B2 and GG distributions perform better. The GB2 aids in selecting parametric distributions for healthcare costs.

A novel family of skewed distributions, termed modified beta distributions, is introduced. This new family of distributions extends the flexibility of the traditional beta distribution to better model asymmetric data. Key properties of the modified beta distributions are derived, offering insights into their behavior and potential applications (Nadarajah et al., 2014).

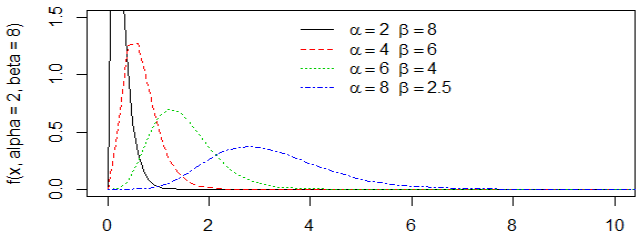
Results and Methodology

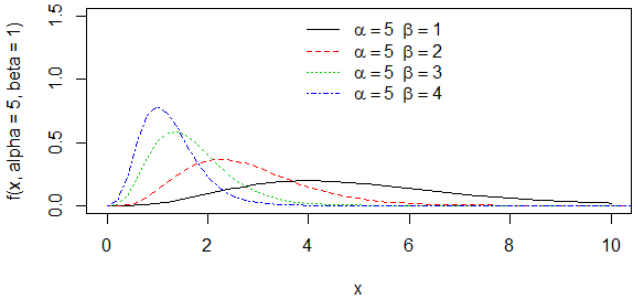
Gamma Distribution:

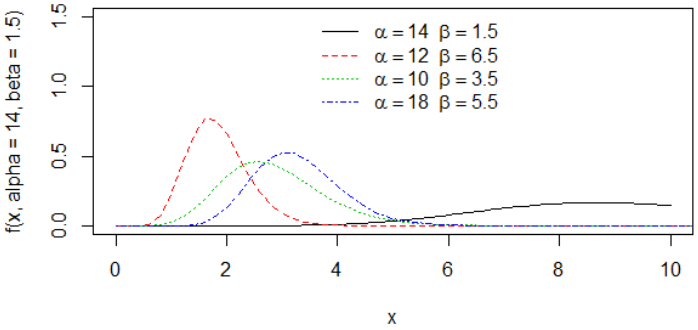
$$f(x;\alpha,\beta)=\frac{\beta^{\alpha}}{\Gamma\alpha}x^{\alpha-1}e^{-\beta x}$$

$$\beta_1=\frac{4}{\alpha} \qquad \gamma_1=\frac{2}{\sqrt{\alpha}}$$

$$\beta_2=\frac{2(\alpha+2)}{\alpha} \qquad \gamma_2=\frac{(4-\alpha)}{\alpha}$$







Moment Gamma distribution (MGD):

The pdf of two parameters MGD is defined as:

$$g(x;\alpha,\beta)=\frac{x^h f(x;\alpha,\beta)}{\mu'_h}\ ; \qquad x>0,\alpha>0,\beta>0,h=1,2,3,\ldots$$

Where $\mu'_h=E(x^h)=\int_{-\infty}^{\infty}x^h.f(x;\alpha,\beta)\,dx$

$$\mu'_h=\int_0^{\infty}x^h.\frac{\beta^{\alpha}}{\Gamma\alpha}x^{\alpha-1}e^{-\beta x}\,dx$$

$$\mu'_h=\frac{\beta^{\alpha}}{\Gamma\alpha}\int_0^{\infty}x^{h+\alpha-1}e^{-\beta x}\,dx$$

$$\mu'_h=\frac{\beta^{\alpha}}{\Gamma\alpha}\int_0^{\infty}(\frac{v}{\beta})^{h+\alpha-1}e^{-v}\frac{dv}{\beta}$$

$$\mu'_h=\frac{\beta^{\alpha-1}}{\Gamma\alpha}\int_0^{\infty}\frac{v^{h+\alpha-1}}{\beta^{h+\alpha-1}}e^{-v}\,dv$$

$$\mu'_h=\frac{\beta^{\alpha-1}}{\beta^{h+\alpha-1}\Gamma\alpha}\int_0^{\infty}v^{h+\alpha-1}e^{-v}\,dv$$

$$\mu'_h=\frac{1}{\beta^{h}\Gamma\alpha}\int_0^{\infty}v^{(h+\alpha)-1}e^{-v}\,dv$$

$$\mu'_h=\frac{\Gamma(\mathrm{h}+\alpha)}{\beta^{h}\Gamma\alpha}$$

As

$$g(x;\alpha,\beta)=\frac{x^h f(x;\alpha,\beta)}{\mu'_h}$$

So,

$$g(x;\alpha,\beta)=\frac{x^h\frac{\beta^{\alpha}}{\Gamma\alpha}x^{\alpha-1}e^{-\beta x}}{\frac{\Gamma(\mathrm{h}+\alpha)}{\beta^{h}\Gamma\alpha}}$$

$$g(x;\alpha,\beta)=\frac{x^h\beta^h\Gamma\alpha.\frac{\beta^{\alpha}}{\Gamma\alpha}x^{\alpha-1}e^{-\beta x}}{\Gamma(\mathrm{h}+\alpha)}$$

$$g(x;\alpha,\beta)=\frac{\beta^{s+\alpha}.x^{h+\alpha-1}e^{-\beta x}}{\Gamma(\mathrm{h}+\alpha)} \qquad \ldots\ldots (1)$$

$$\mu'_r=E(x^r)=\int_{-\infty}^{\infty}x^r.g(x)\,dx$$

$$\mu'_r = \int_0^\infty x^r . \frac{\beta^{h+\alpha} . x^{h+\alpha-1} e^{-\beta x}}{\Gamma(h+\alpha)} dx$$

$$= \frac{\beta^{h+\alpha-r-h-\alpha+1-1}}{\Gamma(h+\alpha)} \int_0^\infty (u)^{(r+h+\alpha)-1} e^{-u}\, du$$

$$= \frac{\beta^{-r}}{\Gamma(h+\alpha)} \Gamma(h+r+\alpha)$$

$$\mu'_r = \frac{\Gamma(h+r+\alpha)}{\beta^r \Gamma(h+\alpha)} \quad \ldots\ldots. (B)$$

By using $r = 1,2,3,4$ in equation (B) and obtained the four moments about origin are:

$$\mu'_1 = \frac{\Gamma(h+\alpha+1)}{\beta\Gamma(h+\alpha)}$$

$$\mu'_1 = \frac{(h+\alpha)}{\beta}$$

$$\mu'_2 = \frac{(h+\alpha)(h+\alpha+1)}{\beta^2}$$

$$\mu'_3 = \frac{(h+\alpha)(h+\alpha+1)(h+\alpha+2)}{\beta^3}$$

$$\mu'_4 = \frac{(h+\alpha)(h+\alpha+1)(h+\alpha+2)(h+\alpha+3)}{\beta^4}$$

$$\mu_1 = 0$$

$$\mu_2 = \frac{(h+\alpha)}{\beta^2}$$

$$\mu_3 = \frac{2(h+\alpha)}{\beta^3}$$

$$\mu_4 = \frac{3\{(\alpha+h)^2 + 2(h+\alpha)\}}{\beta^4}$$

$$\beta_1 = \frac{4}{h+\alpha} \qquad \gamma_1 = \frac{2}{\sqrt{h+\alpha}}$$

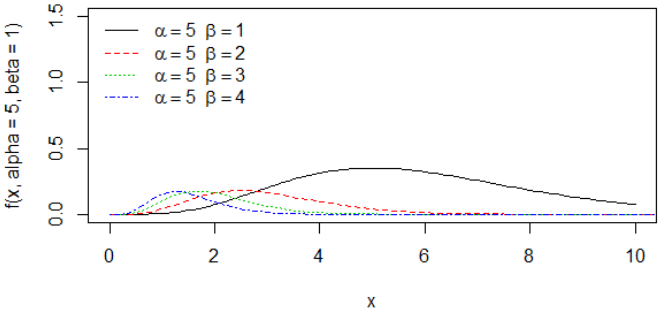
$$\beta_2 = \frac{3(h+\alpha+2)}{[h+\alpha]^3} \qquad \gamma_2 = \frac{3(h+\alpha+2)-3[h+\alpha]^3}{[h+\alpha]^3}$$

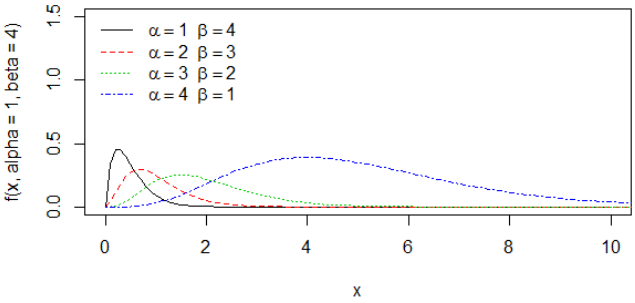
Size Biased Gamma Distribution:

The pdf of two parameters SBGD is estimated by taking $h = 1$ in eq. (1)

$$g(x;\alpha,\beta) = \frac{\beta^{\alpha+1} . x^{1+\alpha-1} e^{-\beta x}}{\Gamma(\alpha+1)}$$

$$g(x;\alpha,\beta) = \frac{\beta^{\alpha+1} . x^{\alpha} e^{-\beta x}}{\Gamma(\alpha+1)}$$





$$\mu'_r = E(x^r) = \int_{-\infty}^{\infty} x^r . g(x;\alpha,\beta)\, dx$$

$$\mu'_r = \int_0^\infty x^r . \frac{\beta^{\alpha+1} . x^{\alpha} e^{-\beta x}}{\Gamma(\alpha+1)}\, dx$$

$$\mu'_r = \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} \int_0^\infty x^{\alpha+r} e^{-\beta x}\, dx$$

$$\mu'_r = \frac{\Gamma(\alpha+r+1)}{\beta^r \Gamma(\alpha+1)} \qquad \ldots\ldots. (C)$$

By using $r = 1,2,3,4$ in equation (C) and get:

$$\mu'_1 = \frac{(\alpha+1)}{\beta} \qquad \mu'_2 = \frac{(\alpha+1)(\alpha+2)}{\beta^2}$$

$$\mu'_2 = \frac{(\alpha+1)(\alpha+2)(\alpha+3)}{\beta^3}$$

$$\mu'_4 = \frac{(\alpha+1)(\alpha+2)(\alpha+3)(\alpha+4)}{\beta^4}$$

First four moments about mean are:

$$\mu_1 = 0$$

$$\mu_2 = \frac{(\alpha+1)}{\beta^2} \qquad \mu_3 = \frac{2(\alpha+1)}{\beta^3}$$

$$\mu_4 = \frac{3\{(\alpha+1)^2 + 2(\alpha+1)\}}{\beta^4}$$

$$\beta_1 = \frac{4}{1+\alpha} \qquad \gamma_1 = \frac{2}{\sqrt{1+\alpha}}$$

$$\beta_2 = \frac{3(\alpha+3)}{[1+\alpha]^3} \qquad \gamma_2 = \frac{3(\alpha+3)-3[1+\alpha]^3}{[1+\alpha]^3}$$

Discordancy Test for Single Upper Outlier (k=1)

The pdf of SBGD is,

$$g(x;\alpha,\beta)=\frac{\beta^{\alpha+1}.x^{\alpha}e^{-\beta x}}{\Gamma(\alpha+1)}\ ; \qquad x>0,\alpha>0,\beta>0$$

To evaluate the test statistic here fix $\alpha=1$ in size biased Gamma distribution. To test of a single upper outlier which is called x_n in a SBG sample. The hypothesis is

$H_0: x_i \epsilon G \qquad i=1,2,3,\dots,n$

Expose that total observations belong to the given distribution *G* with the density

$$g(x;\beta)=\beta^2 x e^{-\beta x}\ ; \quad x>0,\beta>0$$

Consider have a slippage alternative hypothesis H_1 that $(n-1)$ of values are belonging to *G* and single observation called x_n to SBGD $G_1$with density

$$g(x;\beta)=\frac{\beta^2\theta^2 x e^{-\beta\theta x}}{\Gamma(2)};\quad x>0,\theta<1$$

Here θ is a slippage parameter may written as

$H_0: \theta=1$

$H_1: \theta<1$

The log likelihood function under H_0 is

$$\hat{L}_{H_0}(\beta)=2nln2-2nln\bar{x}+\sum ln\,x_i+2n$$

The log likelihood function under H_1 is

$$L_1(\beta,\theta)=\prod \beta^2\theta^2 x e^{-\beta\theta x}$$

$$L_{H_1}(\beta,\theta)=2nln\beta+2ln\theta+\sum ln\,x_i-\beta(n-1)\bar{x}'-\theta\beta x_n$$

Where $\bar{x}'$ is sample mean of (n-1) observations. $L_{H_1}(\beta)$ is maximized at $\hat{\beta}=\frac{2}{\bar{x}'}$ and $\hat{\theta}=\frac{x_n}{\bar{x}'}$ if $x_n>\bar{x}'$ then,

$$\hat{L}_{H_1}(\beta,\theta)=2nln2-2nln\bar{x}'+2lnx_n+\sum ln\,x_i-2ln\bar{x}'+2n$$

The log likelihood ratio test is $\hat{\Lambda}=(\hat{L}_{H_1}-\hat{L}_{H_0})$. Given ratio equal to zero, if $x_n<\bar{x}'$ while if $x_n>\bar{x}'$ then

$$\hat{\Lambda}=2nln\bar{x}-2(n+1)ln\bar{x}'+2lnx_n$$

Discordancy test for two upper outliers (k=2)

The pdf of SBGD is ,

$$g(x;\alpha,\beta)=\frac{\beta^{\alpha+1}.x^{\alpha}e^{-\beta x}}{\Gamma(\alpha+1)}\ ; \qquad x>0,\alpha>0,\beta>0$$

To develop the test statistic we fix $\alpha=1$ in SBGD. For testing of two upper outlier called x_n in a SBG sample. Our hypothesis is

$H_0: x_i \epsilon G \qquad i=1,2,3,\dots,n$

Declaring that all the observations belong to the distribution *G* with the density

$$g(x;\beta)=\beta^2 x e^{-\beta x}\ ; \quad x>0,\beta>0$$

Consider having a slippage alternative hypothesis H_1 that $(n-2)$ of obtained values are

	Significance level			Significance level	
n	5%	1%	n	5%	1%
5	2.4562	2.4495	18	6.995	7.8346
6	2.9836	2.9478	19	7.2471	8.2111
7	3.3227	3.4243	20	7.4992	8.5876
8	3.7562	3.9264	25	8.6745	9.7680
9	4.1603	4.3173	30	9.6342	10.9422
10	4.5404	4.7676	35	10.5939	12.0431
11	8.0633	9.1528	40	11.5536	13.1706
12	5.1698	5.5517	50	12.7330	15.0267
13	5.4589	5.8881	60	13.8447	16.7249
14	5.7276	6.3714	70	14.7999	17.4936
15	6.1182	6.7692	80	15.8601	18.2731
16	6.4908	7.0816	90	16.4811	20.6997
17	6.7429	7.4541	100	18.1383	21.2467

belonging to *G* and one observation say x_{n-1} to SBGD $G_1$with density

$$g(x;\beta)=\frac{\beta^2\theta^2 x e^{-\beta\theta x}}{\Gamma(2)};\quad x>0,\theta<1$$

Here θ is a slippage parameter may write as

$H_0: \theta=1$

$H_1: \theta<1$

The log likelihood function under H_0 is

$$L_{H_0}(\beta)=2nln\beta+\sum ln\,x_i-\beta n\bar{x}$$

Where $\bar{x}$ is sample mean of total values. $L_{H_0}(\beta)$ is maximized at $\beta=\frac{2}{\bar{x}}$

So,

$$\hat{L}_{H_0}(\beta)=2nln2-2nln\bar{x}+\sum ln\,x_i-2n$$

The log likelihood function under H_1 is

$$L_1(\beta)=\prod \beta^2\theta^2 x e^{-\beta\theta x}$$

$$L_{H_1}(\beta,\theta)=2nln\beta+4ln\theta+\sum ln\,x_i-\beta(n-2)\bar{x}'-2\theta\beta\bar{x}''$$

Where $\bar{x}'$ is sample mean of (n-2) values and, $\bar{x}''=\frac{x_n+x_{n-1}}{2}$. $L_{H_1}(\beta,\theta)$ Is maximized at $\hat{\beta}=\frac{2}{\bar{x}'}$ and $\hat{\theta}=\frac{\bar{x}''}{\bar{x}'}$ if $\bar{x}''>\bar{x}'$ then,

$$\hat{L}_{H_1}(\beta,\theta)=2nln2-2ln\bar{x}'+4ln\bar{x}''+\sum ln\,x_i-2ln\bar{x}'+2n$$

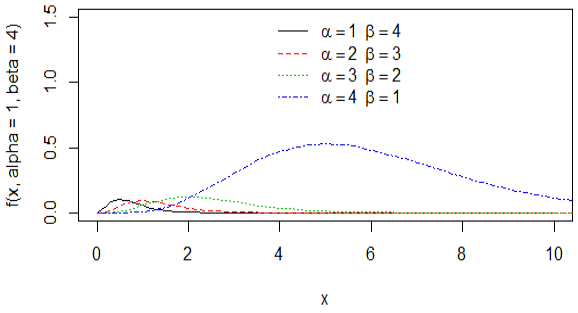
The LLRT is, $\hat{\Lambda} = (\hat{L}_{H_1} - \hat{L}_{H_0})$. Given ratio equal to zero if $\bar{x}'' < \bar{x}'$ while if $\bar{x}'' > \bar{x}'$ then

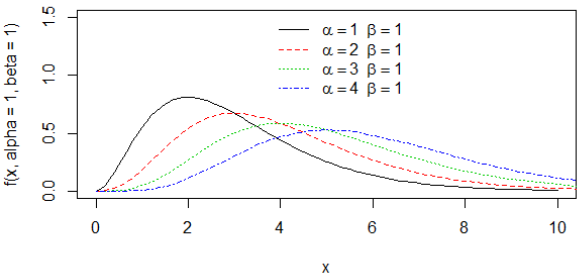
$$\hat{\Lambda} = 2nln\bar{x} - 4ln\bar{x}' + 4ln\bar{x}'' - 2ln\bar{x}'$$
$$\hat{\Lambda} = 2nln\bar{x} - 2(2n+1)ln\bar{x}' + 4ln\bar{x}''$$

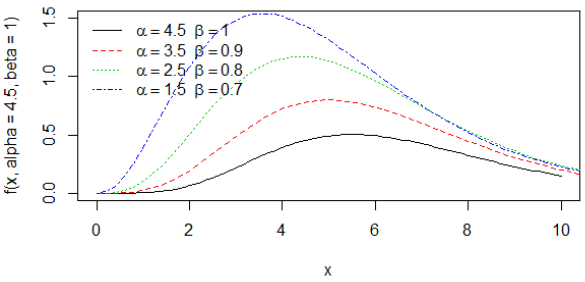
Area Biased Gamma Distribution:

The pdf of two parameters ABGD is obtained by taking $h = 2$ in eq. (1)

$$g(x;\alpha,\beta) = \frac{\beta^{\alpha+2}.x^{2+\alpha-1}e^{-\beta x}}{\Gamma(\alpha+2)}$$
$$g(x;\alpha,\beta) = \frac{\beta^{\alpha+2}.x^{\alpha+1}e^{-\beta x}}{\Gamma(\alpha+2)}$$







$$\mu_r' = E(x^r) = \int_{-\infty}^{\infty} x^r.g(x;\alpha,\beta)\,dx$$
$$\mu_r' = \int_0^{\infty} x^r.\frac{\beta^{\alpha+2}.x^{\alpha+1}e^{-\beta x}}{\Gamma(\alpha+2)}\,dx$$
$$\mu_r' = \frac{\beta^{\alpha+2}}{\Gamma(\alpha+2)}\int_0^{\infty} x^{\alpha+r+1}e^{-\beta x}\,dx$$
$$\mu_r' = \frac{\beta^{\alpha+2}}{\beta^{\alpha+r+2}\Gamma(\alpha+2)}\int_0^{\infty} w^{(\alpha+r+2)-1}e^{-w}\,dw$$
$$\mu_r' = \frac{\Gamma(\alpha+r+2)}{\beta^r\Gamma(\alpha+2)} \quad \ldots\ldots (D)$$

By using $r = 1,2$ in equation (D) and get:

$$\mu_1' = \frac{(\alpha+2)}{\beta}$$
$$\mu_2' = \frac{(\alpha+2)(\alpha+3)}{\beta^2}$$
$$\mu_3' = \frac{(\alpha+2)(\alpha+3)(\alpha+4)}{\beta^3}$$
$$\mu_4' = \frac{(\alpha+2)(\alpha+3)(\alpha+4)(\alpha+5)}{\beta^4}$$

First four moments about mean are:

$$\mu_1 = 0$$
$$\mu_2 = \frac{(\alpha+2)(\alpha+3)}{\beta^2} - (\frac{\alpha+2}{\beta})^2$$
$$\mu_2 = \frac{(\alpha+2)}{\beta^2}$$
$$\mu_3 = \frac{2(\alpha+2)}{\beta^3}$$
$$\mu_4 = \frac{3(\alpha^2+8\alpha+8)}{\beta^4}$$
$$\beta_1 = \frac{4}{\alpha+2} \qquad \gamma_1 = \frac{2}{\sqrt{\alpha+2}}$$
$$\beta_2 = \frac{3(\alpha+4)}{[\alpha+2]^3} \qquad \gamma_2 = \frac{3(\alpha+4)-3[\alpha+2]^3}{[\alpha+2]^3}$$

Discordancy Test for Single Upper Outlier (k=1)

To develop the test statistic we fix $\alpha = 1$ in ABGD. To testing of upper single outlier called x_n in ABG sample. Null hypothesis is

$$H_0: x_i \in G \qquad i = 1,2,3,\ldots,n$$

Claimed that total observations belonging to distribution G with density

$$g(x;\alpha,\beta) = \frac{\beta^3x^2e^{-\beta x}}{2};\ x > 0, \beta > 0$$

Consider having a slippage alternative hypothesis H_1 which $(n-1)$ of data are belonging to G and single value called x_n to ABGD $G_1$with density

$$g(x;\alpha,\beta) = \frac{\beta^3\theta^3x^2e^{-\beta\theta x}}{2}; \qquad x > 0, \theta < 1$$

Here θ is a slippage parameter may write as
$H_0: \theta = 1$
$H_1: \theta < 1$
The log likelihood function under H_0 is

$$L_{H_0}(\beta) = 3nln\beta + 2\sum ln\, x_i - \beta n\bar{x} - nln2$$

Where $\bar{x}$ is sample mean of total values. $L_{H_0}(\beta)$ Is maximized at $\beta = \frac{3}{\bar{x}}$
So,

$$\hat{L}_{H_0}(\beta) = 3nln3 - 3nln\bar{x} + 2\sum ln\, x_i - 3n - nln2$$

The log likelihood function under H_1 is

$$L_1(\beta) = \prod \frac{\beta^3\theta^3x^2e^{-\beta\theta x}}{2}$$

$$L_{H_1}(\beta,\theta) = 3nln\beta + 3ln\theta + 2\sum ln\, x_i - \beta(n-1)\bar{x}' - \theta\beta x_n$$

Where $\bar{x}'$ is sample mean of (n-1) values. $L_{H_1}(\beta,\theta)$ Is maximized at $\hat{\beta} = \frac{3}{\bar{x}'}$ and $\hat{\theta} = \frac{x_n}{\bar{x}'}$ while $x_n > \bar{x}'$ then,

$$\hat{L}_{H_1}(\beta,\theta) = 3nln3 - 3ln\bar{x}' + 3lnx_n + 2\sum ln\, x_i - 3nln\bar{x}' - nln2 + 2n$$

The LLRT is$\hat{\Lambda} = (\hat{L}_{H_1} - \hat{L}_{H_0})$. Given ratio equal to zero, when $x_n < \bar{x}'$ but if $x_n > \bar{x}'$ then

$$\hat{\Lambda} = 3nln\bar{x} - 3(n+1)ln\bar{x}' + 3lnx_n$$

Sample size	Significance level			Significance level	
n	5%	1%	n	5%	1%
5	4.9587	4.9743	18	15.2535	16.7497
6	5.8583	5.8801	19	15.3531	17.6555
7	6.7579	6.7859	20	16.3962	18.5613
8	7.6575	7.6917	25	19.1354	22.0986
9	8.5571	8.5975	30	21.8746	25.7284
10	9.4567	9.5033	35	24.6138	29.1742
11	10.3563	10.4091	40	27.353	31.9256
12	11.2559	11.3149	50	30.9522	37.5621
13	12.1555	12.2207	60	35.8477	43.1957
14	13.0551	13.1265	70	39.7432	48.1465
15	13.2547	14.0323	80	44.0587	58.7472
16	13.7543	14.9381	90	45.0142	59.8342
17	14.3539	15.8439	100	48.8697	63.1538

Discordancy Test for Single Upper Outlier (k=2)
To develop the test statistic we fix $\alpha = 1$ in Area Biased Gamma distribution. For testing of two upper outlier say x_n in Area Biased Gamma sample. Our hypothesis is

$$H_0: x_i \in G \qquad i = 1,2,3,\dots,n$$

Claimed that total observations belonging to the distribution G with density

$$g(x;\alpha,\beta) = \frac{\beta^3x^2e^{-\beta x}}{2};\ x > 0, \beta > 0$$

Consider having a slippage alternative hypothesis H_1 which is $(n-2)$ of given data belonging to G and one observation say x_{n-1} to ABGD $G_1$with density

$$g(x;\alpha,\beta) = \frac{\beta^3\theta^3x^2e^{-\beta\theta x}}{2}; \qquad x > 0, \theta < 1$$

Here θ is a slippage parameter may write as
$H_0: \theta = 1$
$H_1: \theta < 1$
The log likelihood function under H_0 is

$$L_{H_0}(\beta) = 3nln\beta + 2\sum ln\, x_i - \beta n\bar{x} - nln2$$

Where $\bar{x}$ is sample mean of all observations. $L_{H_0}(\beta)$ is maximized at $\beta = \frac{3}{\bar{x}}$
So,

$$\hat{L}_{H_0}(\beta) = 3nln3 - 3nln\bar{x} + 2\sum ln\, x_i - 3n - nln2$$

The log likelihood function under H_1 is

$$L_1(\beta) = \prod \frac{\beta^3\theta^3x^2e^{-\beta\theta x}}{2}$$

$$L_{H_1}(\beta,\theta) = 3nln\beta + 6ln\theta + 2\sum ln\, x_i - \beta(n-2)\bar{x}' - 2\theta\beta\bar{x}''$$

Where $\bar{x}'$ is sample mean of (n-2) observations and $\bar{x}'' = \frac{x_n + x_{n-1}}{2}$. $L_{H_1}(\beta,\theta)$ is maximized at $\hat{\beta} = \frac{3}{\bar{x}'}$ and $\hat{\theta} = \frac{\bar{x}''}{\bar{x}'}$ if $\bar{x}'' > \bar{x}'$ then,

$$\hat{L}_{H_1}(\beta,\theta) = 3nln3 - 6ln\bar{x}' + 4ln\bar{x}'' + 2\sum ln\, x_i - 3ln\bar{x}' + 3n$$

The LLRT is $\hat{\Lambda} = (\hat{L}_{H_1} - \hat{L}_{H_0})$. Given ratio equal to zero, if $\bar{x}'' < \bar{x}'$ while if $\bar{x}'' > \bar{x}'$ then

$$\hat{\Lambda} = 3nln\bar{x} - 3(2n+1)ln\bar{x}' + 6\ln(\frac{x_n + x_{n-1}}{2})$$

Sample size	Significance level		Sample size	Significance level	
n	5%	1%	n	5%	1%
5	2.4998	2.4898	18	8.7177	8.6984
6	2.9781	2.9869	19	9.196	9.1204
7	3.4564	3.484	20	9.6743	9.6424
8	3.9347	3.9811	25	10.9415	11.7644
9	4.413	4.4782	30	12.7587	13.936
10	4.8913	4.9753	35	14.5759	15.7775
11	5.3696	5.4724	40	16.2931	17.7365
12	5.8479	5.9695	50	19.1224	21.2529
13	6.3262	6.3666	60	21.7517	24.2729
14	6.8045	6.8637	70	24.1781	27.8965
15	7.2828	7.3608	80	26.9103	31.2145
16	7.7611	7.8579	90	28.5396	33.4378
17	8.2394	8.2455	100	30.3689	36.1837

Summary

This research underscores the importance of simulation and discordancy tests in assessing the impact of outliers. Moment distributions, particularly relevant for forestry and engineering, offer robust analytical tools. The Gamma distribution stands out in forestry research due to its wide usage. Instrumental errors can distort data, affecting its shape and scale, potentially leading to misinterpretation and flawed results if outliers are not properly addressed.

Recognizing outliers before proceeding with analysis and modeling is crucial. While much attention has been given to univariate probability distributions, this study focuses on identifying outliers in univariate moment distributions. Two discordancy tests are devised for outlier detection, supported by simulation studies for critical value determination.

Graphical representations illustrate how data behavior varies with parameter values. Moment distributions, often used as size-biased and aera-biased in research, are employed here, specifically the Gamma distribution, for outlier detection and parameter estimation. The study compares the outlier detection capabilities of four strategies through simulation, using tests of different sizes from various moment distributions. This comprehensive approach acknowledges that a single test may not fully capture strategy performance, necessitating a comparative analysis.

Ultimately, this research contributes to refining outlier detection methodologies in moment distributions, providing valuable insights for fields reliant on accurate data analysis. By leveraging simulation and rigorous testing, it enhances the reliability and robustness of analytical techniques, facilitating more accurate inferences and decision-making processes.

References

AKINSETE, A., FAMOYE, F. & LEE, C. 2008. The beta-Pareto distribution. *Statistics,* 42**,** 547-563.

ALEXANDER, C., CORDEIRO, G. M., ORTEGA, E. M. & SARABIA, J. M. 2012. Generalized beta-generated distributions. *Computational Statistics & Data Analysis,* 56**,** 1880-1897.

ALIZADEH, M., BAGHERI, F. & M KHALEGHY MOGHADDAM, M. 2013. Efficient estimation of the density and cumulative distribution function of the generalized Rayleigh distribution. *Journal of Statistical Research of Iran JSRI,* 10**,** 1-22.

GOGOI, B. & DAS, M. K. 2015. Detection of Multiple Upper Outliers in Exponential Sample under Slippage Alternative. *statistics*, 2.

HARIHARAN, B., ARBELÁEZ, P., GIRSHICK, R. & MALIK, J. Simultaneous detection and segmentation. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13, 2014. Springer, 297-312.

JONES, A. M., LOMAS, J. & RICE, N. 2014. Applying beta-type size distributions to healthcare cost regressions. *Journal of Applied Econometrics,* 29**,** 649-670.

KAGAN, Y. Y. 2002. Seismic moment distribution revisited: I. Statistical results. *Geophysical Journal International,* 148**,** 520-541.

KIMBER, A. 1979. Tests for a single outlier in a gamma sample with unknown shape and scale parameters. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 28**,** 243-250.

Kimber, A. C. (1982). Tests for Many Outliers in an Exponential Sample. *Journal of the Royal Statistical Society*, 263-271.

Kimber, A. C. (1983). Discordancy Testing in Gamma Samples with Both Parameters Unknown. *Journal of the Royal Statistical Society*, 304-310.

NADARAJAH, S., TEIMOURI, M. & SHIH, S. H. 2014. Modified beta distributions. *Sankhya B,* 76**,** 19-48.

NASIRI, P. Estimation parameter of R= P (Y< X) for Lomax distribution with presence of outliers. International Mathematical Forum, 2016. 239-248.

ROHLF, F. J. 1975. Generalization of the gap test for the detection of multivariate outliers. *Biometrics***,** 93-101.

TRIPATHI, Y. M., KUMAR, S. & PETROPOULOS, C. 2014. Improved estimators for parameters of a Pareto distribution with a restricted scale. *Statistical Methodology,* 18**,** 1-13.