



Generalized Mean as a Robust Estimator of Population Mean

Name : Ankita Pal

Course : MSC SEM IV

Roll No. : MSc(SEM-IV)-02

Reg. No. : VB-1212 of 2018-19

Supervisor : Prof. Soumalya Mukhopadhyay

Acknowledgements

This project has taken a considerable amount of time and resources and I would like to acknowledge the help of those who have made the project possible. In particular I would like to thank my supervisor Prof. Soumalya Mukhopadhyay for his time, patience and guidance, and also for allowing the idea to be pursued originally.

Further I would like to thank all of many thousands of people who have worked on the basics and whose papers or articles without whom the project would not have been possible. A list of papers and articles is given in [References](#).

Contents

Acknowledgements	1
Contents	2
1 Introduction	3
1.1 Aims & Objective of the Project	5
1.2 Overview of the Project Progress	5
1.3 Overview of this Report	6
2 Generalized Mean : A Potential Candidate	6
2.1 Asymptotic Bias & MSE of the Generalized Mean	7
2.2 Optimal α : The Tuning Parameter	8
2.2.1 Approach : Obtain Optimal α	9
2.3 Properties	11
2.3.1 Link of Generalized Mean with M-Estimation	11
2.3.2 Influence Function of Generalized Mean	11
3 Some Illustrative Data Examples	12
3.1 Simulation Data Analysis	12
3.2 Real Data Analysis	15
3.2.1 Example 1 : Snapper's Length Data	15
3.2.2 Example 2 : Newcomb Data	15
4 Conclusion	16
References	18

1 Introduction

This document is a report for the project “Robust Estimation for Population Mean”. The project is an attempt to provide a simple and stable estimate at center of the dominating population, eliminating the effect of outliers. Classical estimation methods rely heavily on assumptions which are often not met in practice. We know that when there are outliers in the data, classical estimators often have very poor performance with respect to their bias and MSE. Robust estimation techniques play a crucial role to provide improved estimators with respect to their bias and MSE. Robustness can be treated as one of the most desirable properties of an estimation procedure, especially of the population mean. The literature is abundant with study of robust statistical method for estimation of population mean. The problem of estimating the population mean from a sample containing a few ‘very large’ observations has been faced by most sampling practitioners. Trimmed and Winsorized means are among the most popular estimators of location parameter (Stigler, 1977). They overcome the extreme sensitivity of the mean as well as improve the efficiency of the median at light tailed distribution. The trimmed (and Winsorized) means can efficiently balance between robustness and efficiency, the two crucial requirements of any statistical procedure (Tukey (1962)). The trimmed mean has been an extremely popular estimator of location parameters over the years (Tukey and McLaughlin (1963), Bickel (1965), Huber (1972), Stigler (1973)). For an outlier prone symmetric distribution, the trimmed mean is a popular estimator for the location parameter (Caperaa and Rivest, 1995). On the other hand, when the problem is to estimate the mean of a skewed distribution, the Winsorized Mean is found to be very reasonable estimate by Rivest, 1994. These two robust averages are used by Jose and Winkler (2008) to determine the average forecast. Another important part of the development of modern robust statistics is M-estimation which can be regarded as a generalized of ML estimation. It was introduced by Huber (1964), who illustrated the asymptotic properties of M-estimators. Liang and Zeger (1986) helped to popularize M-estimators in the biostatistics literature under the name ‘generalized estimating equation’. Rousseeuw and Yohai (1984) proposed the S-estimation technique which is based on residual scale of M-estimation. Another variant is the MM-estimation procedure, in which the regression parameter is estimated using S-estimation which minimizes the scale of residual from M-estimation and then proceed with M-estimation. MM-estimation aims to obtain the estimates that have a high breakdown value and more efficiency. Basu et al. (1998) have developed a minimum divergence estimation method for robust parameter estimation. The proposed approach uses new density-based divergences which, unlike existing methods such as minimum Hellinger distance estimation, avoid the use of nonparametric density estimation and consequent complications such as bandwidth selection. The proposed class of ‘density power divergences’ is indexed by a single parameter α which controls the trade-off between robustness and efficiency. The robust estimator of the location parameter obtained by the method of Basu et. al. has similarity with the generalized mean, as in both the cases, the parameter α (although has a different interpretation in both the concepts), controls the trade-off between robustness and efficiency. Apart from these robust methods, there is another way to convert an unbiased raw estimate to an improved biased one, called the Shrinkage estimation. Unlike the robust estimation, a contaminated setup

is not necessary for the shrinkage estimation. Such a procedure for minimum variance unbiased linear estimator (MVULE) of the mean of a population was suggested by [Thompson \(1968\)](#). They employed the technique of shrinkage of the MVULE towards a natural origin in the parameter space by multiplying it with a shrinking factor and illustrated for the Normal and the Binomial population. Later, [Mehta and Srinivasan \(1971\)](#) studied the problem of shrinkage of the MLE of the population mean of various populations (Normal, Gamma, Poisson and Binomial) towards a natural origin, when some prior information regarding the location is available. In Bayesian framework, a useful shrinkage estimator for the mean of a multivariate normal population is Stein's estimator ([Stein, 1981](#)) which performs better than the usual MLE. It improves upon the total MLE risk, i.e., the sum of the expected errors of each component. However, any particular component would improve for some parameter values, and deteriorate for others. Stein's idea was later modified by James, Efron and Morris to provide the James-Stein estimator, which is a natural fit in estimation of multiple asset returns in portfolio ([Jorion, 1986](#)). A nice application of shrinkage estimation in linear models was suggested by [Tibshirani \(1996\)](#).

The present paper explores the robustness of the well known statistic generalized mean, which is defined as $M_\alpha = \left(\frac{1}{n} \sum_{i=1}^n x_i^\alpha\right)^{\frac{1}{\alpha}}$, for a set of given positive numbers x_1, x_2, \dots, x_n . It involves the three common means, i.e., the arithmetic mean (AM), the harmonic mean (HM) and the geometric mean (GM) as its particular cases, according as $\alpha = 1, \alpha = -1$ and $\alpha \rightarrow 0+$. Similar to the AM, GM and the HM, the trimmed and Winsorized means are also special cases of the generalized mean. For some suitably chosen tuning parameter α , M_α equals the trimmed mean of the Winsorized mean with a certain trimming proportion. However, that α may or may not be the same as for the α for which M_α is the closest to the mean of the true population. It also has some similarity with the robust estimator of the location parameter obtained by the method of [Basu et. al. \(1998\)](#) as in both the cases, the family of estimators are indexed by a parameter α (although has a different interpretation in both the concepts). The tuning parameter α controls the trade-off between robustness and efficiency. The generalized means are also regarded as particular cases of M-estimators. We found some applications of the generalized mean in machine learning ([Tripathi and Kalra, 2011](#); [Attia et. al., 2012](#)). However, the merit of generalized mean as an estimator was very inadequately explored till date. It was [Mukhopadhyay et. al. \(2019\)](#) who first explored the potential of generalized mean in estimating the mean of the dominating population more accurately eliminating the effect of the outliers. If the data are generated by a mixture population, involving a major (dominating) component and a minor (outlying) component and if our focus is on estimating the mean of the dominating component, then this simple but appealing tool serves as a reasonable estimate of out target parameter ([Mukhopadhyay et. al. \(2019\)](#)). It outperforms the existing sophisticated robust techniques very often and might be regarded as their potential competitor, which has the merits of simplicity and the scope of choosing the tuning parameter optimally.

1.1 Aims & Objective of the Project

Suppose the data are generated by a mixture population, involving a major (dominating) component and a minor (outlying) component. If we are interested in estimating the mean of the entire population, we can use the maximum likelihood estimator (MLE), which could, among other things, be the AM, GM or the HM, depending on the structure of the parent distribution. However, if our primary interest is in estimating the mean of the dominating component eliminating the effect of the outlying component, the usual MLE will normally perform poorly (unless, of course, the outlying component is entirely absent or we take the mixture structure into account when doing parametric modeling). So, the aim of the project is to obtain a suitable and stable estimator for estimating the population mean for a dominating population, eliminating the effect of outliers. Also on finding out the most suitable and stable estimate, we shall be comparing between Generalized Mean[2] and the method used in Basu et. al.(1998). Here we shall refer Basu, Harris & Hjort's method, i.e., the method proposed in Basu et. al. as the BHH method.

1.2 Overview of the Project Progress

Here we are interested in estimating the mean of the dominating population with a suitable π (proportion of dominating population). The sample observations may not come from a single population, they may come from a mixture where the majority of the observations belong to the dominating component and a few observations come from a smaller, outlying component. A formal view to the progress of the project was taken and resulted in a systematic development of the required components. In order to meet the stated aims, the following steps were undertaken:

The project was split into four separate parts, where the first part was designed independently, while the others depended on each other.

1. Simulation - Here k distinct samples each of size n is simulated from a Normal Mixture Distribution, where the dominating component follows $N(10, 1)$ and the outlying component follows $N(20, 1)$.
2. Evaluation - It consists of two parts : Calculations for estimating the population mean and, Calculating the MSE. Each of these parts are explained below:
 - (a) Estimation - A number of robust estimators were used to estimate the population mean of the dominant component, such as, Trimmed Mean, Winsorized Mean, Median, AM, Generalized Mean[2] and the method used in Basu et. al.(1998).
 - (b) Mean Square Error(MSE) - Since we would estimate the Population Mean using different estimators, so the most appropriate way to find the best estimate is by finding the MSE of each estimate.
3. Implementation - Implementation shall be done on a real-life data, to check whether the suitable estimate can be applied in real-life.

On completion of the project this report was produced to document the decisions made in the above steps.

1.3 Overview of this Report

This report fully describes the project undertaken. However to control the length of the report, the reader is on occasion referred to a bibliographical reference if particular details are required in an area.

The report is split into four main sections :

1. Introduction - This gives an Introduction to the project, its aims, an overview of the work undertaken in the project and an overview of this report.
2. Generalized Mean - In this section a detailed description of Generalized Mean, it's properties and asymptotic bias and MSE. It also contains the procedure of obtaining the optimal α .
3. Illustrative Data Examples - This section contains two types of examples on data analysis:(i) Simulation Data Analysis, and (ii) Real Data Analysis. In case of real data analysis, two real-life data has been used to perform the analysis.
4. Conclusion - Here mainly we conclude about generalized mean is a much easier tool than sophisticated robust analysis and a brief conclusion about the comparison between generalized mean and BHH.

2 Generalized Mean : A Potential Candidate

Generalized means or power means represent a family of functions for aggregating sets of positive real numbers that include the arithmetic, geometric and the harmonic means as particular cases. The generalized mean involves a power transformation of the original data. Given positive numbers x_1, x_2, \dots, x_n , the generalized mean with exponent α is defined as $M_\alpha = \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^{\frac{1}{\alpha}}$. The generalized means can be expressed in the general form $M_\alpha = h^{-1}(\bar{h}(\underline{x}))$, where \underline{x} is the vector of the realized observations (x_1, x_2, \dots, x_n) corresponding to the random vector (X_1, X_2, \dots, X_n) , and $h(x)$ is a continuous monotone function; in most applications it is either x^λ or $\ln x$. [Berger and Casella \(1992\)](#) have derived the most general expression of the one-parameter exponential family density for which $h^{-1}(\bar{h}(\underline{x}))$ turns out to be the maximum likelihood estimate of the model parameter θ . In many cases, the parent distribution may be unknown and the exact sampling distribution of $h^{-1}(\bar{h}(\underline{x}))$ may be hard to derive ([Berger and Casella \(1992\)](#)).

In this context, these authors assumed that the bias of $h(X_1)$ for estimating $h(\theta)$ is negligible. In the presence of outliers in the data, the amount of bias in estimating $h(\theta)$ by $h(X_1)$ may not be negligible. Accordingly, the usual MLE might have a substantial bias as an estimator of the target parameter under our parametric model. However, there exists an “optimal” estimator among the class of estimators M_α , having the smallest MSE for estimating the mean of the dominating population, substantially eliminating the effect of outliers. But when outliers are present in the data, the optimal tuning parameter α may not be close to either of the three common values 1, -1 or 0. In other words, the optimal estimator, although a member of the class of generalized means, can be quite far apart from any of the common means, such as, the AM, GM or HM.

2.1 Asymptotic Bias & MSE of the Generalized Mean

When the form of the true data generating distribution is unknown or parametric modelling is difficult, the MLE is generally not a relevant concept. Even when reasonably accurate modelling is possible, a small proportion of outliers can still seriously damage the analysis based on the MLE. Let us assume that the distributional structure of the data generating mechanism is unknown and there is at most one outlying component in the system.

We should expect that if M_α is a robust estimate in comparison with the common means, then there exists an α such that the M_α corresponding to that α has the minimum bias and MSE for estimating the mean of the dominating population. For this, we need to evaluate the expressions of the asymptotic bias and MSE of the generalized mean. This can be achieved from the asymptotic distribution of M_α . As the distributional structure of the sample (X_1, X_2, \dots, X_n) is unknown, the of $\frac{1}{n} \sum_{i=1}^n X_i^\alpha$ (say, T) is also unknown, and we cannot apply delta method directly on M_α . So initially we have to prove the asymptotic normality of T through CLT, and for that the expressions of the mean and the variance of T are required. We evaluate the approximate expressions of the mean and variance of T using the binomial expansion technique. A detailed review is given in [Chakraborty et. al.\(2014\)](#).

Lemma 1: *Let X_1, X_2, \dots, X_n be a random number of size n from a distribution F with mean μ and variance σ^2 respectively. Then the approximate expressions for the expectation and variance of $T = \frac{1}{n} \sum_{i=1}^n nX_i^\alpha$ are given by:*

$$E\left(\frac{1}{n} \sum_{i=1}^n nX_i^\alpha\right) \approx \mu^\alpha \left[1 + \frac{\alpha(\alpha-1)\sigma^2}{2\mu^2}\right]$$

$$Var\left(\frac{1}{n} \sum_{i=1}^n nX_i^\alpha\right) \approx \frac{\mu^{2\alpha}}{n} \left[\left(1 + \frac{\alpha(2\alpha-1)\sigma^2}{\mu^2}\right) - \left(1 + \frac{\alpha(\alpha-1)\sigma^2}{2\mu^2}\right)^2\right]$$

provided the coefficient of variation of the distribution is less than $\frac{1}{3}$.

Lemma 2: *Suppose X_1, X_2, \dots, X_n are random samples from a population which consists of two components. The dominating population has mean μ_1 and variance σ_1^2 , while μ_2 and σ_2^2 are the mean and variances of the outlying population. Let $\pi (> 1/2)$ be the mixing proportion and let c_1, c_2 be the coefficients of variation of the dominating and outlying populations with the assumption $c_1, c_2 < 1/3$. Then*

$$\frac{\sqrt{n}(M_\alpha^\alpha - \theta^*)}{\sqrt{V^*}} \rightarrow Z \sim N(0, 1); n \rightarrow \infty$$

where

$$\theta^* = \pi\mu_1^\alpha \left[1 + \frac{\alpha(\alpha-1)\sigma_1^2}{2\mu_1^2}\right] + (1-\pi)\mu_2^\alpha \left[1 + \frac{\alpha(\alpha-1)\sigma_2^2}{2\mu_2^2}\right]$$

$$V^* = \left[\pi \mu_1^{2\alpha} \left(1 + \frac{\alpha(2\alpha-1)\sigma_1^2}{\mu_1^2} \right) + (1-\pi) \mu_2^{2\alpha} \left(1 + \frac{\alpha(2\alpha-1)\sigma_2^2}{\mu_2^2} \right) \right] \\ - \left[\pi \mu_1^\alpha \left(1 + \frac{\alpha(\alpha-1)\sigma_1^2}{2\mu_1^2} \right) + (1-\pi) \mu_2^\alpha \left(1 + \frac{\alpha(\alpha-1)\sigma_2^2}{2\mu_2^2} \right) \right]^2$$

Lemma 3: Under the assumptions stated in Lemma 3, we have

$$\frac{M_\alpha - \theta^{**}}{\sqrt{V^{**}}} \rightarrow Z_1 \sim N(0, 1); n \rightarrow \infty$$

where

$$\theta^{**} = \theta^{*\frac{1}{\alpha}} = \left[\pi \mu_1^\alpha \left(1 + \frac{\alpha(\alpha-1)\sigma_1^2}{2\mu_1^2} \right) + (1-\pi) \mu_2^\alpha \left(1 + \frac{\alpha(\alpha-1)\sigma_2^2}{2\mu_2^2} \right) \right]^{\frac{1}{\alpha}} \\ V^{**} = \frac{1}{n\alpha^2} \theta^{*\frac{2}{\alpha}-2} V$$

Remark 1: From Lemma 4, the bias $B(M_\alpha)$ of M_α in estimating μ_1 is approximately,

$$E_F(M_\alpha) - \mu \approx \left[\pi \mu_1^\alpha \left(1 + \frac{\alpha(\alpha-1)\sigma_1^2}{2\mu_1^2} \right) + (1-\pi) \mu_2^\alpha \left(1 + \frac{\alpha(\alpha-1)\sigma_2^2}{2\mu_2^2} \right) \right]^{\frac{1}{\alpha}} - \mu_1 \quad (1)$$

and

$$MSE(M_\alpha) = V^{**} + [B(M_\alpha)]^2$$

.

2.2 Optimal α : The Tuning Parameter

Straightforward calculations based on Equation (1) show that $B(M_\alpha) \gtrless 0$ when

$$\frac{\alpha(\alpha-1)}{2} \gtrless \frac{(1-\pi)(\mu_1^\alpha - \mu_2^\alpha)}{\pi \mu_1^\alpha c_1^2 + (1-\pi) \mu_2^\alpha c_2^2} \quad (2)$$

Remark 2. When $\alpha = 1$, we have $Bias(\bar{X}) = (1-\pi)(\mu_1 - \mu_2)$. So, for estimating μ_1 , \bar{X} may have a substantial bias, depending on the values of π and μ_2 . However, there exists some α at which the bias is zero for estimating the mean of the dominating population.

On the other hand, the bias of \bar{X} is zero for estimating $\pi\mu_1 + (1-\pi)\mu_2$. In particular, when $\pi = 1.0$, i.e., there is no contamination in the data, the bias of M_α turns out to be zero at $\alpha = 1$.

Remark 3. The bias $B(M_\alpha)$ equals to zero when we have equality in Equation (2). If $\mu_2 < \mu_1$ and $\alpha > 0$, $B(M_\alpha)$ will be equal to zero for some $\alpha \in (0, 1)$, because only in this case the sign of the LHS and the RHS will match. If $\mu_2 < \mu_1$ and $\alpha < 0$, the RHS and LHS of Equation (2) will have positive signs. For $\mu_1 < \mu_2$, on the other hand, the signs of the LHS and RHS can match only for $\alpha > 1$.

Remark 4. Note that the variance of M_α is of order $\frac{1}{n}$. Hence, for increasing n , the variance term tends to zero and, accordingly, the value of α which minimizes $B(M_\alpha)$ also approximately minimizes $MSE(M_\alpha)$ for sufficiently large n .

2.2.1 Approach : Obtain Optimal α

In case of simulation studies, one can minimize the empirical mean square error over the grid of α values used. But in the absence of any knowledge about the parent population distribution, it is difficult to construct an empirical estimate of the MSE and there is no clear rule for choosing an optimal α . Here we provide some strategies for a reasonable selection of α in the absence of such distributional information. In the following, α^* will denote the optimal α satisfying $B(M_{\alpha^*}) = 0$, i.e., Bias of M_{α^*} is equal to zero.

Theorem 1: Let c_1 and c_2 denote, respectively, the coefficients of variation of the dominating and outlying populations, and let $c = \min(c_1, c_2)$. Similarly, μ_1 and μ_2 denote the means of the dominating and outlying components. Then we have the following results.

- (a) Let $\mu_1 > \mu_2$. Then α^* belongs to the range $\left(\frac{1 - \sqrt{1 + \frac{8}{c^2}}}{2}, 1\right)$. In particular, $B(M_{\alpha^*}) \leq B(\bar{X})$.
- (b) Let $\mu_1 < \mu_2$. Then α^* belongs to the range $\left(1, \frac{1 + \sqrt{1 + \frac{8}{c^2}}}{2}\right)$. In particular, $B(M_{\alpha^*}) \leq B(\bar{X})$.

Proof: We have already observed that $B(M_\alpha) = 0$ for some α satisfying

$$\frac{\alpha(\alpha - 1)}{2} = \frac{(1 - \pi)(\mu_1^\alpha - \mu_2^\alpha)}{\pi\mu_1^\alpha c_1^2 + (1 - \pi)\mu_2^\alpha c_2^2}$$

As π represents the weight of the dominating component, we have $\frac{\pi}{1 - \pi} \leq 1$. Hence,

$$\frac{\alpha(\alpha - 1)}{2} \leq \frac{\mu_1^\alpha - \mu_2^\alpha}{\mu_1^\alpha c_1^2 + \mu_2^\alpha c_2^2} \leq \frac{\mu_1^\alpha - \mu_2^\alpha}{c^2(\mu_1^\alpha + \mu_2^\alpha)} \quad (3)$$

Again, as $c_1, c_2 < 1/3, \mu_1, \mu_2$ and hence $\mu_1^\alpha, \mu_2^\alpha$ are necessarily positive. The equation (3) implies that $\frac{\alpha(\alpha - 1)}{2} \leq \frac{1}{c^2}$ (using $\frac{x - y}{x + y} \leq 1$, for $x, y > 0$) and hence $\alpha^2 - \alpha - \frac{2}{c^2} \leq 0$. The roots of the quadratic equation $\alpha^2 - \alpha - \frac{2}{c^2} = 0$ are $\frac{1 \pm \sqrt{1 + \frac{8}{c^2}}}{2}$. Thus $B(M_\alpha) = 0$ for some α satisfying $\frac{1 - \sqrt{1 + \frac{8}{c^2}}}{2} \leq \alpha \leq \frac{1 + \sqrt{1 + \frac{8}{c^2}}}{2}$. Also, if $\mu_1 < \mu_2$, then $B(M_\alpha) = 0$ for some $\alpha < 1$. Combining this with the present finding, we can conclude that when $\mu_1 < \mu_2$, the value α^* corresponding to zero bias must lie within $\frac{1 - \sqrt{1 + \frac{8}{c^2}}}{2}$ and 1. As α^* is optimal, we have $B(M_{\alpha^*}) \leq B(M_1) = B(\bar{X})$. This proves part (a). Part (b) can be proved similarly.

In applying the results of the previous theorem to real data, we face two major hurdles. The first hurdle is that as the values of μ_1 and μ_2 are unknown,

some guidance is necessary to determine whether to use the interval in Theorem 1, part (a) or part (b). In order to retain the simplicity of our procedure, we make this decision based on the histogram of the data. If the second highest peak of the histogram is on the right of the highest peak, then we use the interval in Theorem 1, part (a); if it is on the left of the highest peak then we use the interval in Theorem 1, part (b).

The second hurdle is that the quantities c_1, c_2 and c cannot be explicitly determined from the data. The coefficient of variation estimated from the full data will represent the pooled coefficient of variation (c_p) of the entire (mixture) population. In the following, we propose a revised bound for α^* in terms of the pooled CV c_p . We begin by establishing a relationship between c_p , and the CVs c_1 and c_2 (Lemma 4).

Lemma 4. *Suppose we have a mixture of two populations (without any reference to a dominating or an outlying population) with respective means μ_1 and μ_2 , standard deviations σ_1 and σ_2 and coefficients of variation c_1 and c_2 . Without loss of generality, let $c_1 < c_2$. If c_p denotes the pooled coefficient of variation of the combined set, then either $c_p^2 > \max(c_1, c_2)$ or $c_1^2 < c_p^2 < c_2^2$. This indicates that c_p^2 is higher than c^2 .*

In actual practice, a data based estimate $c_p^{(0)} = s/\bar{x}$ of c_p is used, where s and \bar{x} are sample standard deviation and sample mean of the data. An alternative approach for estimating the optimal α could be to make use of quartile based estimate of the coefficient of variation. As the rank based estimates of location and scale are more stable against outliers compared to the classical estimates. Accordingly we propose $c_p^{(1)} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$ as a rank based estimate of CV where Q_1 and Q_3 are the first and third quartiles of the data.

Suppose that $\mu_1 < \mu_2$, i.e., the mean of the outlying component is larger than that of the dominating component. Then the actual bound for α^* (using c) is of the form $(a, 1)$, where $a < 1$ and our bound (using c_p) is of the form $(b, 1)$, where $a < b < 1$. Here a is unknown and we take $(b + 1)/2$ as our estimate of α^* . On the other hand, if $\mu_1 > \mu_2$, the actual bound of α^* (using c) is of the form $(1, a_1)$ and our bound (using c_p) is of the form $(1, b_1)$, where $1 < b_1 < a_1$. Here a_1 is unknown and we take $(b_1 + 1)/2$ as our estimate of α^* .

A step by step guideline for the selection of the optimal α given the data x_1, x_2, \dots, x_n is as follows:

- Construct the histogram of the data
- Determine whether the second highest peak of the histogram is on the right or on the left of the highest peak. If it is on the right, then the relevant interval for α^* is provided by Theorem 1, part (a). In this case, the interval has an upper bound at 1.
- If the second highest peak is on the left of the highest peak, the relevant interval is the one given by Theorem 1, part (b). In this case, the interval has a lower bound at 1.
- Depending on the chosen bound, the unknown end of the bound is determined from the relevant formula in Theorem 1 by using an appropriate estimate of the coefficient of variation.

- We have suggested two possible choices of estimates of coefficient of variation as $c_p^{(0)}$ and $c_p^{(1)}$.
- Finally the estimate of the optimal α is taken to be the midpoint of the constructed interval.

2.3 Properties

2.3.1 Link of Generalized Mean with M-Estimation

Generalized means arise as the MLEs for the one-parameter exponential family

$$f(x|\theta) = w(x)\exp(\theta h(x) - H(\theta))$$

where $h(x) = x^\alpha$ and $h(x) = \frac{d}{dx}H(x)$. The ML equation here we solve is

$$\sum_{i=1}^n (x_i^\alpha - \theta^\alpha) = 0$$

Thus, it is a particular case of M-estimation with $\psi(x_i, \theta) = x_i^\alpha - \theta^\alpha$.

2.3.2 Influence Function of Generalized Mean

The generalized mean for a population with distribution function F is given by

$$T(F) = [E_F(X^\alpha)]^{\frac{1}{\alpha}}$$

Define

$$F_\epsilon(x) = (1 - \epsilon)F + \epsilon\delta_x$$

where δ_x denotes the distribution degenerate at x .

Therefore,

$$\begin{aligned} E_{F_\epsilon}(X^\alpha) &= (1 - \epsilon)E_F(X^\alpha) + \epsilon x^\alpha \\ &= (1 - \epsilon)(T(F))^\alpha + \epsilon x^\alpha \end{aligned}$$

and

$$\begin{aligned} T(F_\epsilon(x)) &= [E_{F_\epsilon}(X^\alpha)]^{\frac{1}{\alpha}} \\ &= [(1 - \epsilon)(T(F))^\alpha + \epsilon x^\alpha]^{\frac{1}{\alpha}} \end{aligned}$$

For simplicity, we denote $T(F_\epsilon(x))$ by $T(F_\epsilon)$ and $E_F(X)$ by μ .

The influence function is defined as

$$\begin{aligned} IF(x, T, F) &= \lim_{\epsilon \rightarrow 0} \frac{T(F_\epsilon) - T(F)}{\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{[(1 - \epsilon)(T(F))^\alpha + \epsilon x^\alpha]^{\frac{1}{\alpha}} - T(F)}{\epsilon} \end{aligned}$$

where either $x > \mu$ and $\alpha < 1$ or $x < \mu$ and $\alpha > 1$. Applying L' Hôpital's rule, we have

$$\begin{aligned} IF(x, T, F) &= \lim_{\epsilon \rightarrow 0} \frac{1}{\alpha} [(1 - \epsilon)(T(F))^\alpha + \epsilon x^\alpha]^{\frac{1}{\alpha} - 1} [x^\alpha - (T(F))^\alpha] \\ &= \frac{1}{\alpha} [T(F)]^{1-\alpha} [x^\alpha - (T(F))^\alpha] \\ &= \frac{1}{\alpha} \left[x^\alpha (E_F(X^\alpha))^{\frac{1-\alpha}{\alpha}} - (E_F(X^\alpha))^{\frac{1}{\alpha}} \right] \end{aligned}$$

We have considered the model $N(10,1)$. To draw the influence curve for the model, the range of x is taken as $(0,40)$. The figure of the influence function for different choices of α are as follows.

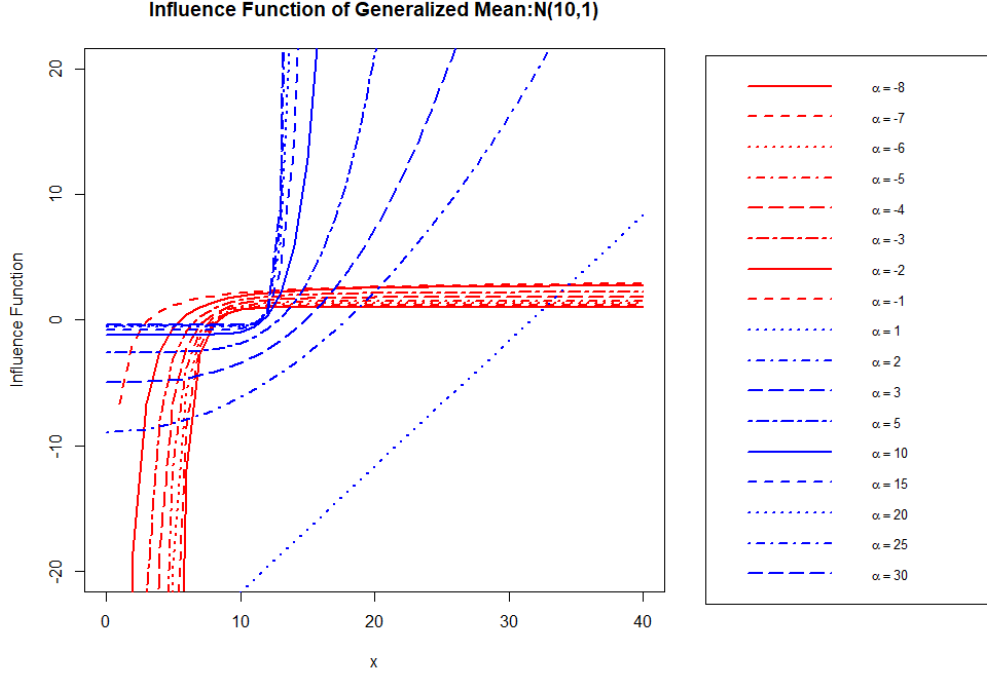


Figure 1: Influence Function of the Generalized Mean: $Normal(10,1)$ model

3 Some Illustrative Data Examples

3.1 Simulation Data Analysis

In order to illustrate the potential applications of the generalized mean, we perform a simulation study under the normal model. The normal model is in some sense representative of very common distribution scenario involving symmetric distribution.

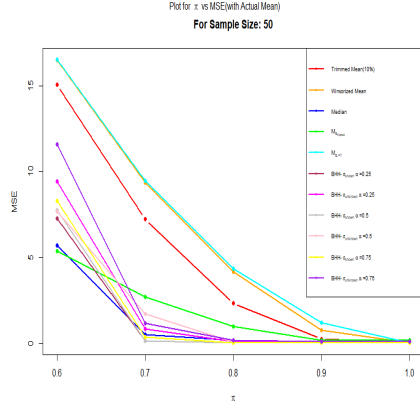
In our simulation example, the outlier population is chosen to belong to the same family as the dominating component, but the location parameters of the dominating and the outlying components are taken to be sufficiently apart. We consider two distinct normal populations having $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ distributions, with $\mu_1 \neq \mu_2$. In our notation, μ_1 and μ_2 will represent the mean of the dominating and outlying components, respectively. A consideration for choosing the means (μ_1 and μ_2) and the standard deviation σ is that the means should be positive and several standard deviations away from zero to essentially eliminate the possibility of having negative values in the sample. Accordingly, we chose the $N(10,1)$ and $N(20,1)$ distributions. The relative proportions of the $N(10,1)$ and $N(20,1)$ populations are taken to be 10:0, 9:1, 8:2, 7:3 and 6:4. In each of the cases, $k = 1000$ distinct samples of size $n = 50, 100$ and 500

are drawn. The parameter α varies over the intervals depending on the relative position of the outlying population with respect to the dominating population. For each of the k samples, the following estimators are evaluated, they are, Trimmed Mean(10%), Winsorized Mean, Median, M_α is evaluated only for the optimal α values and $\alpha = 1$ (i.e., AM) and BHH (known and unknown σ) for $\alpha = 0, 0.25, 0.5, 0.75$. The optimal α values are obtained by using Theorem 1. The mean square error for each of the above estimates is calculated, as the sum of the variance and the squared bias, both by using the actual mean as well as the pilot estimator $\hat{\mu}$ of the population mean. Here we use the sample median as the pilot estimator. In case of BHH (Basu et. al.'s) method for unknown standard deviation, σ we estimate it by $\hat{\sigma} = \frac{\text{median}(|X_i - \text{median}(\underline{x})|)}{0.6745}$, which is known robust estimation method for standard deviation. The following we will be showing the graphs of MSE vs π (proportion of dominating population) for each of the samples 50, 100 and 500.

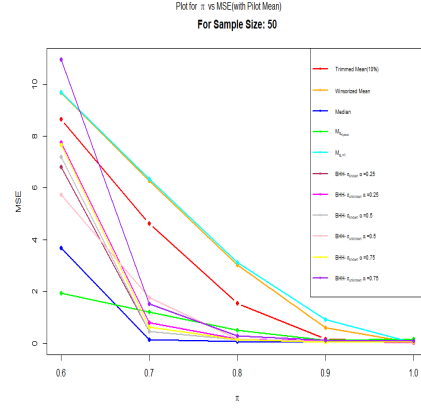
From Figure 2 the following points can be noted for comparing the proposed (Generalized Mean) method and BHH (Basu et. al.) method :-

- We can clearly see that as the π (proportion of dominating observations) increases the mean square error decreases and finally becomes zero when $\pi = 1.0$ in about all the sample sizes taken above.
- As the sample size increases from 50 to 500, the performance of the estimators becomes better.
- Now on specially **comparing M_α (for optimal α) and BHH**, we found the following results :-
 1. When $\pi = 0.6$, we find the MSE values of $M_{\alpha_{optimal}}$ is sometimes less or more than the MSE values of BHH (known and unknown σ) for different values of α .
 2. But as the value of π increases from 0.7, the MSE of both the estimators has a small difference between them.
 3. Finally when π ranges from 0.9 to 1.0, the MSE values for both these estimators become quite equal.

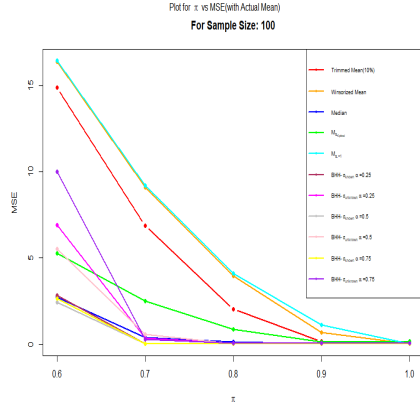
Finally we may say that **our proposed method (Generalized Mean) has a performance near to Basu et. al.'s method. Hence has the ability to estimate the location paramter of a contaminated population nearly as good as Basu et. al.'s method.**



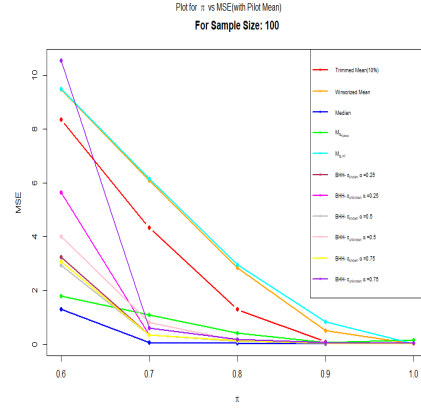
(a) MSE comparison with actual mean: Sample size 50



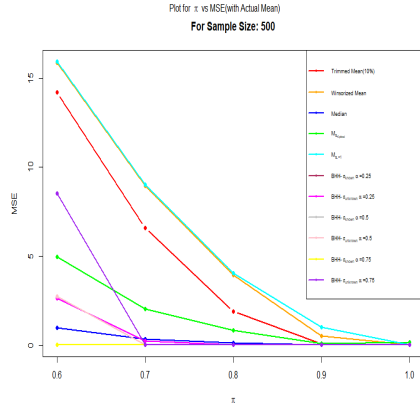
(b) MSE comparison with pilot mean: Sample size 50



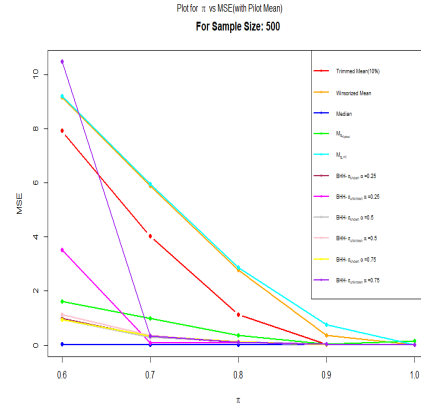
(c) MSE comparison with actual mean: Sample size 100



(d) MSE comparison with pilot mean: Sample size 100



(e) MSE comparison with actual mean: Sample size 500



(f) MSE comparison with pilot mean: Sample size 500

Figure 2: Variation of MSE of different estimators with varying proportion of dominating observations and sample size

3.2 Real Data Analysis

3.2.1 Example 1 : Snapper's Length Data

We have compared the estimates of the true population location obtained by the density power divergence method with the generalized mean for the suitably chosen α , based on the data on Snapper's length.

In the Figure 3a, we have shown the fitted probability curve, for various values of α by following the approach of Basu et. al.(1998), over the histogram of the data. The case of $\alpha = 0$ corresponds to MLE and the figure depicts the gradual shifting of probability curve through increase in the value of α till the value 1, the maximum value. In the Figure 3b, we have shown the best fitted probability curve for the best choice of α along with the probability curve corresponding to MLE over the sample histogram. Although α can vary within 0 to 1 by following Basu et. al.(1998), the best choice of α in the sense of best fitted probability curve of the data may not be possible. However, in the proposed method, the best choice of α may be suggested according to the thumb rule and the probability curve corresponding to that choice is found to capture the true data in an equally efficient way as any choice of Basu et. al.(1998).

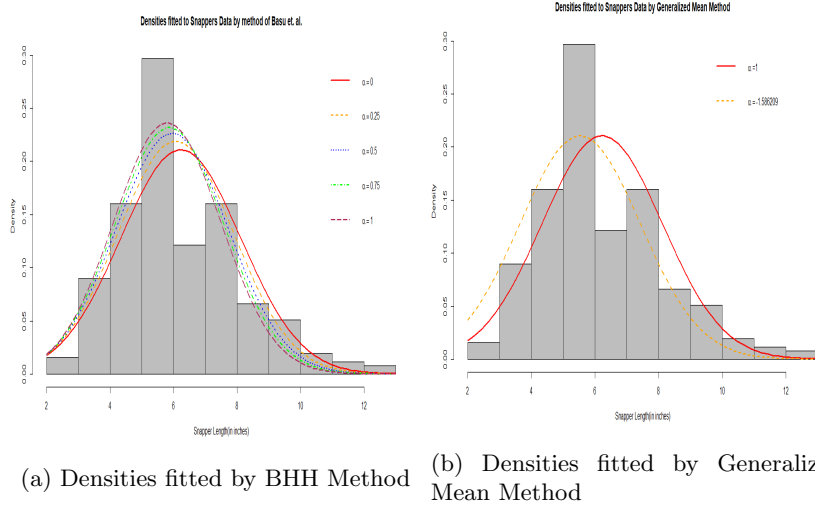


Figure 3: Probability Curves fitted to the Snapper's Data by Basu's Method and Generalized Mean Method

3.2.2 Example 2 : Newcomb Data

In this example again we will be comparing generalized mean with Basu et. al.'s method(BHH method) using another real-life data, i.e., Newcomb data. Since this data contains some negative values so we shall be shifting the origin and scale. In Figure 4a, we have shown the fitted probability curve, for various values of α by following the method of Basu et. al., superimposed on the histogram of the data. On the other hand, the case of $\alpha = 0$ we have considered MLE for estimating the mean of the data. So, except when the MLE is used, all the other probability curves fit the main body of the histogram quite well. In

Figure 4b, probability density curves have been superimposed on the histogram of the data following our proposed method. Similarly, as in Example 3.2.1, for Basu et. al.'s method we will not be able to obtain a probability curve for the best choice of α , whereas for our proposed method, the best choice of α can be preferred according to the thumb rule.

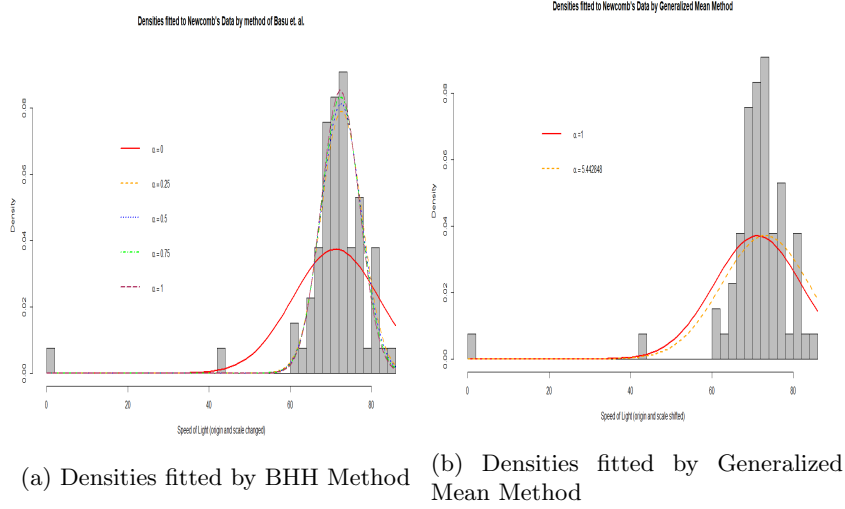


Figure 4: Probability Curves fitted to the Newcomb's Data by Basu's Method and Generalized Mean Method

4 Conclusion

The generalized mean is a simple and intuitive tool providing a natural summary for a set of positive numbers. However, the immense utility of the generalized mean in producing outlier stable estimates of location has not been exploited in the literature so far.

The influence of extreme observations have a diminishing impact on the generalized mean through suitable power transformations of the data, as the outlying observations in the original scale conform to the rest of the observations in the transformed scale. However, choosing the optimal α which provides the best conformity can be a challenging problem in case of real data. Some theoretical results have been provided in this connection (Theorem 1) and also some strategies have been provided for choosing the optimal α . Our method provides a simple, easy and appealing tool for the estimation of the mean of the dominating population under the presence of outliers without getting into sophisticated robust analysis.

For instance, if we consider the experimental biologists, they perform inferences regarding the growth of a species mostly on the basis of contaminated data. Their procedures are based on the common means and because of the high sensitivity of the common means to the outliers, the conclusions are prone to be erroneous. If inferences are performed using the generalized means, it is expected to lead them towards a more realistic scenario.

Our method may not be exactly accurate as the BHH method, but from this report we can at least conclude that the generalized mean performances very close to [Basu et. al.](#)'s method. Since our method is simple and easy it is very easy to understand and use. Here in this project we have considered only one outlying component, but we expect that it will be possible to use our method to estimate the central moments of the dominating population in case of mixtures. Therefore, considering more than one outlying components could be our future endeavor.

References

- (1) Attia, M.A., Sallam, E.A., Fahmy, M.M., 2012. A proposed generalized mean single multiplicative neuron model. *Computer Engineering and Systems*, 73-78. doi:[10.1109/ICCP.2012.6356163](https://doi.org/10.1109/ICCP.2012.6356163).
- (2) Basu, A., Harris, I., Hjort, L., Jones, M., 1998. Robust and efficient estimation by minimising a density power divergence. *Biometrika* 85(3), 549-559.
- (3) Berger, R.L., Casella, G., 1992. Deriving generalized means as least squares and maximum likelihood estimates. *The American Statisticians* 46(4), 279-282.
- (4) Bickel, P.J., 1965. On some robust estimates of location. *Annals of Mathematical Statistics* 36, 847-858.
- (5) Caperaa, P., Rivest, L.P., 1995. On the variance of the trimmed mean. *Statistics and Probability Letters* 22, 79-85.
- (6) Chakraborty, B., Bhattacharya, S., Basu, A., Bandhopadhyay, S., Bhattacharjee, A., 2014. Goodness-of-fit testing for the Gompertz growth curve model. *Metron* 72 (1):45-64.
- (7) Cressie, N. & Read, T. R. C., 1984. Multinomial goodness-of-fit tests. *J. R. Statist. Soc. B*(46), 440-464.
- (8) Huber, P., 1964. Robust estimation of a location parameter. *The Annals of Mathematical Statistics* 35(1), 73-101.
- (9) Huber, P., 1972. The 1972 wald lecture robust statistics: A review. *The Annals of Mathematical Statistics* 43, 1041-1067.
- (10) Jorion, P., 1986. Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* 21, 279-292.
- (11) Jose, V.R.R., Winkler, R.L., 2008. Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting* 24, 163-169.
- (12) Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13-22.
- (13) Lindsay, B. G., 1994. Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *Ann. Statist.* 22, 1081-1114.
- (14) Mehta, J.S., Srinivasan, R., 1971. Estimation of the mean by shrinkage to a point. *Journal of the American Statistical Association* 66, 86-90.
- (15) Mukhopadhyay, S., Das, A.J., Basu, A., Chatterjee, A., Bhattacharya, S., 2019. Does the generalized mean have the potential to control outliers?. *Communications in Statistics - Theory and Methods*.doi:[10.1080/03610926.2019.1652320](https://doi.org/10.1080/03610926.2019.1652320).
- (16) Neyman, J., 1949. Contribution to the theory of χ^2 tests. In *Proceedings of the First Berkley Symposium in Mathematics and Statistics*, 239-273. Berkeley, CA: University of California Press.

- (17) Rao, C. R., 1963. Criteria of estimation in large samples. *Sankhyā* A(25), 189-206.
- (18) Rivest, L.P., 1994. Statistical properties of winsorized means for skewed distributions. *Biometrika* 81(2), 373-383.
- (19) Rousseeuw, P., Yohai, V., 1984. Robust regression by means of s-estimators. *Robust and Non-Linear Time Series Analysis: Lecture*
- (20) Stein, C.M., 1981. Estimation of the mean of a multivariate normal population. *The Annals of Statistics* 9, 1135-1151.
- (21) Stigler, S.M., 1973. The asymptotic distribution of the trimmed mean. *The Annals of Statistics* 1, 472-477.
- (22) Stigler, S.M., 1977. Do robust estimators work with real data? *The Annals of Statistics* 5, 1055-1098.
- (23) Thompson, J.R., 1968. Some shrinkage techniques for estimating the mean. *Journal of the American Statistical Association* 63, 113-122.
- (24) Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 267-288.
- (25) Tripathi, B.K., Kalra, P.K., 2011. On efficient learning machine with root-power mean neuron in complex domain. *IEEE Transaction on Neural Networks* 22(5), 727-738.
- (26) Tukey, J.W., 1962. The future data analysis. *Annals of Mathematical Statistics* 33, 1-67.
- (27) Tukey, J.W., McLaughlin, D.H., 1963. Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization 1. *Sankhyā: The Indian Journal of Statistics, Series A(1961-2002)* 25, 331-352.
- (28) Thompson, J.R., 1968. Some shrinkage techniques for estimating the mean. *Journal of the American Statistical Association* 63, 113-122.
- (29) Victoria-Feser, M. P. & Ronchetti, E., 1997. Robust estimation for grouped data. *J. R. Statist. Soc. B*(57), 599-609.