ON THE ROBUST ESTIMATOR OF THE POPULATION MOMENTS

Marlon S. Frias
https://orcid.org/0000-0001-8891-3805
Bukidnon State University
Malaybalay City, Philippines

ABSTRACT

This paper proposed the Huber's M-estimator for estimating the population moments m_k. Eleven features of the population moments from twelve probability distributions, mean squared error, and bias were computed to compare the results of Huber's and the classical estimator. For the Huber's M-estimator, we used the trimming proportions p = 0.05, 0.10, 0.15, 0.20 and 0.25 and compute the estimates for n=20, 30, and 100. The result confirms that the appropriate estimator of the population moments of the symmetric distributions is classical while the estimator with trimming proportion p = 0.25 is appropriate for the asymmetric distributions. Furthermore, the estimators with higher trimming proportions give smaller variances as compared to the estimators that have lower trimming proportions.

Keywords: *Huber m-estimation, population moments, robust estimator, trimmed mean*

1.0 Introduction

The moments describe the nature of the distribution. Some features such as the mean, the variance, the skewness, etc. can characterize any distribution. The *k*th population moment, or *k*th moment of the distribution *f(x)*, is $E(x^k)$. Since it is often difficult or impossible to obtain the population observations, one should only get a sample from a population of interest and infer the characteristics of the population based on the sample data gathered. Because of this, the statistic m_k is introduced to estimate the population moments. The basic idea of the m_k is averaging the *k*th power of the observation and is represented by the formula: $m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k,\ k = 1, 2, \ldots$

The notion of a moment is fundamental for describing the features of a population. For example, the population mean (population average), usually denoted μ, measures the central tendency and the population variance which is usually denoted σ^2 or Var(y), is defined as the second moment of *y* centered about its mean: $Var(y) = E[(y-\mu)^2]$. The variance is widely used as a measure of spread in a distribution. Another feature of a population moment is the skewness that is referred to as the third population moment, which is commonly denoted as γ_1 and is mathematically defined as $\gamma_1 = \frac{\sum_{i=1}^{N} z_i^3}{N}$ where we take the z-score for each, cube it, sum across all *N* individuals, and then divide by the number of individuals *N*.

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. Further, the next feature of the population moments is the kurtosis that is referred to as the fourth population moment. Kurtosis is conceptually defined as the "peakedness" of a distribution and is denoted by γ_2. Some distributions are rather flat, and others have a rather sharp peak. A very peaked probability distribution is known as leptokurtic while a relatively flat distribution is platykurtic. However, a distribution that is somewhere in between is known as mesokurtic. One can determine the normality of the data by just looking at its histogram or by getting the value of its skewness and kurtosis. If, however, the value of the skewness and kurtosis of certain data is like the normal distribution, that is $\gamma_1 = 0$ and $\gamma_2 = 0$, then the data is considered as normally distributed. This is important because many statistical tests require normality assumption.

It is clear that m_k estimates μ_k for all k and by the law of large numbers, that is $m_k \to \mu_k$ for all $n \to \infty$. Moreover, by Central Limit Theorem (CLT), $\sqrt{n}(m_k - \mu_k) \to N(0, v^2)$ as $n \to \infty$, v^2 is the asymptotic variance. However, according to Slutsky's Theorem, the variance of the kth moment is $v^2 = k^2\mu_k^{2(k-1)}\sigma^2$ which is quite large for the higher order moments and makes the density estimate undesirable. It should also be noted that the estimates of μ_k using the m_k is affected by the extreme values or by the presence of the outliers, especially in the higher order moments. These are the reasons why we proposed an alternative estimator of the population moments that will obtain a robust estimate and smaller variance. The concept of the proposed estimator is to trim both ends of the tail of the distribution to obtain robust estimates that would yield a smaller variance. This estimator is derived by using the Huber's m-estimation. In this study, five trimming proportions (p = 0.05, 0.10, 0.15, 0.20, and 0.25) are considered and then compared to the classical estimator. An estimator that obtains the smallest mean-squared error (MSE) is chosen to estimate the population moments and will be considered as the appropriate estimator.

2.0 Basic Concepts

Huber (1964) proposed a method of estimating a location parameter: the generalization of the Maximum Likelihood Estimation (MLE). This estimation method will obtain estimates that are robust, that is insensitive to departures from underlying assumptions caused by, for example, *outliers. Robust estimators should have good performance under the underlying assumptions, and the performance deteriorates gracefully as the situation departs from the assumptions.* The basic idea of this paper is to make use of the Huber's M-estimation in estimating the population moments of the symmetric probability distributions. The essential form of the M-estimation problem is the following:

Given a set of n independent and identically distributed random samples $x_1, x_2, \dots, x_n$ with some probability density function f parameterized by θ, the problem is to estimate the location parameter θ. First, consider a distance function $\rho(x_i, \theta)$ that is given by

$$\rho(x_i, \theta) = (x_i - \theta)^2 \qquad (1)$$

Since we have *n* random samples, then our function becomes

$$\sum_{i=1}^{n} \rho(x_i, \theta) = \sum_{i=1}^{n} (x_i - \theta)^2 \qquad (2)$$

and our goal in here is to get an estimate of the parameter θ that minimizes this function. Hence, we have to differentiate the equation (2) with respect to θ (denote this as $\rho(x_i, \theta)$) and solve for the root of θ. Thus, we have

$$\sum_{i=1}^{n} \varphi(x_i, \theta) = -2\sum_{i=1}^{n} (x_i - \theta) \qquad (3)$$

Equate the result to zero and solve for θ to get the estimate. Thus we have

$$-2\sum_{i=1}^{n} (x_i - \theta) = 0 \qquad \Rightarrow \qquad \hat{\theta} = \bar{x} \qquad (4)$$

The obtained estimate of θ is the sample mean that is affected by the outlier. Hence, consider another distance function:

$$\rho(x_i, \theta) = |x_i - \theta| \qquad (5)$$

Similar to equation (2), our goal is to minimize the equation:

$$\sum_{i=1}^{n} \rho(x_i, \theta) = \sum_{i=1}^{n} |x_i - \theta| \qquad (6)$$

Solving for the derivative with respect to θ, thus, we have

$$\sum_{i=1}^{n} \varphi(x_i, \theta) = -2\sum_{i=1}^{n} \mathrm{sgn}(x_i - \theta) \qquad (7)$$

Equate the result to zero and solve for θ to get the estimate. Thus, we have

$$-2\sum_{i=1}^{n} \mathrm{sgn}(x_i - \theta) = 0 \qquad \Leftrightarrow \qquad \hat{\theta} = \text{median } \{ x_i \} \qquad (8)$$

which is highly robust, but unfortunately not very efficient.

Again, consider another ρ function

$$\rho(x_i) = \begin{cases} 0, & x_i \le -c \\ x_i^2, & -c \le x_i < c \\ 0, & x_i \ge c \end{cases} \quad (9)$$

Hence,

$$\sum_{i=1}^{n} \rho(x_i, \theta) = \begin{cases} -\theta, & x_i \le -c \\ (x_i - \theta)^2, & -c \le x_i < c \\ -\theta, & x_i \ge c \end{cases} \quad (10)$$

So, we have

$$\sum_{i=1}^{n} \varphi(x_i, \theta) = \begin{cases} 0, & x_i \le -c \\ -2(x_i - \theta), & -c \le x_i < c \\ 0, & x_i \ge c \end{cases} \quad (11)$$

Hence, the obtained estimator for θ is

$$\hat{\theta} = \frac{1}{[n-2r]} \sum_{i=[r+1]}^{[n-r]} x_i^k$$ where r is the number of trimmed samples (12)

3.0 Methods

This paper assessed the performance of the proposed alternative robust estimator. Simulation of observations in 12 distributions for sample sizes n = 20, 30 and 100 was done to determine the estimates of the population moments. The 12 probability distributions were chosen according to (1) type of distribution; (2) symmetry; and (3) kurtosis. First, the original data were standardized and computed the first up to the 11^{th} sample moments of every probability distribution using the classical and the trimmed means with trimming proportions p=0.05, 0.10, 0.15, 0.20 and 0.25. The process was repeated 100 times. The mean squared errors and bias of every *kth* moment (1^{st} up to 11^{th}) were derived. The variance of the classical estimator and the five (5) alternative estimators were also obtained. Finally, the averages of the MSE's, bias, and variances of the estimates were computed. The estimator that yields the smallest MSE is regarded as the appropriate estimator, and the estimator that has the smallest variance is the most efficient estimator of a certain population moment.

4. 1 Parameter Estimation

Definition 1.

Let x_1, x_2, x_3 ..., be a random sample from a *pmf* or *pdf f(x)*. For k = 1, 2, 3,… the kth population moment, or *k*th moment of the distribution *f(x)*, is $E(x^k)$. The kth sample moment is

$$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k, \qquad k = 1,2,... \tag{13}$$

It is clear that m_k estimates μ_k for all k and by the law of large numbers:

$$m_k \to \mu_k \quad \text{for all } n \to \infty \tag{14}$$

Moreover, by the Central Limit Theorem:

$$\sqrt{n}(m_k - \mu_k) \to N(0, v^2) \quad \text{as } n \to \infty \tag{15}$$

where v^2 is the asymptotic variance.

Further, the Slutsky's Theorem states that if a sequence of numbers a_n tends to infinity as n increases such that:

$$a_n(Y_n - \mu) \to Y \qquad \text{in distribution} \tag{16}$$

and if $g(\cdot)$ is a continuous and continuously differentiable function, then:

$$a_n(g(Y_n) - g(\mu_n)) \to g'(\mu)Y \quad \text{in distribution.} \tag{17}$$

Hence, the large sample variance of $g(Y_n)$ is approximately equal to :

$$Var[g(Y_n)] = [g'(\mu)]^2 Var(Y) = [g'(\mu)]^2\sigma^2 \tag{18}$$

If we take $g(x) = x^k$, then $\mu = E(x^k) = \mu_k$ and $g(\mu) = \mu_k^k$, and hence, $g'(\mu) = k\mu_k^{k-1}$. The large sample variance of the kth moment is therefore approximately:

$$v^2 = k^2\mu_k^{2(k-1)}\sigma^2 \tag{19}$$

which will be large for higher order moments.

The expected value of the kth sample moment is given by

$$E(m_k) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i^k\right) = \frac{1}{n}\sum_{i=1}^{n} \mu_k$$

$$= \frac{1}{n}\cdot n\mu_k$$

$$= \mu_k \qquad (20)$$

Hence, the estimator is unbiased for the kth population moment. However, by using the Slutsky's theorem, the variance of the estimates m_k of μ_k is approximately equal to:

$$V(m_k) = \frac{1}{n^2}\sum_{i=1}^{n} V\left(X_i^k\right)$$

$$= \frac{1}{n^2}\cdot nV\left(X_i^k\right)$$

$$= \frac{1}{n}\cdot k^2 \mu_k^{2(k-1)} \sigma^2 \qquad (21)$$

which is very large, thereby making the density estimate undesirable. It is because of this that we propose to make use of the Huber's m-estimation to obtain a more robust estimate of the μ_k and to minimize its variance. Hence, from equation (12), the proposed robust estimator is given by

$$\hat{m}_k = \frac{1}{[n-2r]}\sum_{i=[r+1]}^{[n-r]} x_i^k \ , \qquad k = 1,2,... \qquad (22)$$

4.2 Numerical Simulation

Table 4.1 Estimators of the Population Moments: Sample size n = 20

PROBABILITY DISTRIBUTIONS	Estimator	MSE	BIAS
Normal	Classical	3416115.61	-69.95
Student's t	Classical	3416115.61	-69.95
Uniform (Continuous)	Classical	836.07	1.53
Weibull	Trimmed (p = 0.25)	0.56	0.98
Exponential	Trimmed (p = 0.25)	0.65	1.04
Chi-Square	Trimmed (p = 0.25)	177792.50	167.95
Binomial	Classical	508256.96	-89.94
Hypergeometric	Classical	253845.13	-117.20
Uniform (Discrete)	Classical	2140.19	4.52
Negative Binomial	Trimmed (p = 0.25)	6309657.74	920.18
Geometric	Trimmed (p = 0.25)	0.11	0.53
Bernoulli	Trimmed (p = 0.25)	2.05	0.75

Table 4.2 Estimators of the Population Moments: Sample size n = 30

PROBABILITY DISTRIBUTIONS	Estimator	MSE	BIAS
Normal	Classical	5804254.20	-10.99
Student's t	Classical	5804254.20	-10.99
Uniform (Continuous)	Classical	971.51	1.46
Weibull	Trimmed (p = 0.25)	0.45	0.95
Exponential	Trimmed (p = 0.25)	0.47	0.99
Chi-Square	Trimmed (p = 0.25)	156098.75	157.94
Binomial	Classical	174598.73	-104.57
Hypergeometric	Classical	180866.44	-44.60
Uniform (Discrete)	Classical	2029.25	-1.01
Negative Binomial	Trimmed (p = 0.25)	4805302.32	808.39
Geometric	Trimmed (p = 0.25)	0.09	0.58
Bernoulli	Trimmed (p = 0.25)	0.01	0.58

Table 4.3 Estimators of the Population Moments: Sample size n = 100

PROBABILITY DISTRIBUTIONS	Estimator	MSE	BIAS
Normal	Classical	1663280.12	-2.05
Student's t	Classical	1663280.12	-2.05
Uniform (Continuous)	Classical	60.33	0.59
Weibull	Trimmed (p = 0.25)	0.42	1.21
Exponential	Trimmed (p = 0.25)	0.50	1.20
Chi-Square	Trimmed (p = 0.25)	40379.00	83.75
Binomial	Classical	1837835.54	32.25
Hypergeometric	Classical	3741886.37	-78.04
Uniform (Discrete)	Classical	49.47	0.00
Negative Binomial	Trimmed (p = 0.25)	1157071.65	407.67
Geometric	Trimmed (p = 0.25)	0.05	0.82
Bernoulli	Trimmed (p = 0.25)	0.75	0.59

It is important to note that the best estimator of the symmetric distributions (continuous or discrete) is the classical for it gives consistently smallest MSE's in all sample sizes. On the other hand, the estimator with trimming proportion p = 0.25 is appropriate in estimating the population moments of all skewed distributions that are considered in this study. These probability distributions are: Weibull, Exponential, Chi-Square, Negative Binomial, Geometric, and Bernoulli. The finding simply implies that the proposed robust estimator performs well as compared to the classical when the distribution of the data is asymmetric.

This result coincides with the findings of Tukey that if the distribution of the data deviates (skewed) from the normal distribution caused by the outliers, the trimmed and Winsorized means will provide robust estimates. This is important because in reality the distribution of the data may not be perfectly symmetric and some are contaminated by the outliers which cause the skewness.

The proposed estimator which is derived using Huber's estimation is robust in estimating the features of the population moments. Through this estimator, the estimates are not affected by the extreme values since the data were trimmed at both the lower and upper tails of its distribution hence obtain smaller variance compared to the classical estimator. The performances of the estimators were determined through the MSE criterion derived from simulation experiments by Monte Carlo method.

5. Conclusions

Based on the results, the appropriate estimator of the symmetric distributions is the classical estimator since it gives more stable estimates as manifested by its MSE compared to the proposed estimator. However, the estimator with trimming proportion p = 0.25 is more precise in estimating the population moments of the skewed distributions.

As to the efficiency of the estimators, the estimator with trimming proportion p = 0.25 has the smallest variance for asymmetric distributions. Thus, the proposed estimator is the most efficient estimator under this condition. However, the most efficient estimator for Weibull and exponential distributions is the robust estimator at p = 0.15.

References

Boyer, J.E., (1981). *Variances for Adaptive Trimmed Means.*

Brischetto, A. and Richards, A., (2007). *The Performance of Trimmed Mean Measures of Underlying Inflation.*

Collins, J.R. (1986). *Maximum Asymptotic Variances of Trimmed Means Under Asymmetric Contamination.*

Evans, M. (2000). *Statistical Distributions Third Edition.*

Hall, P. (1981). *Large Sample Properties of Jaeckel's Adaptive Trimmed Mean.*

Huber, P.J. (1963). *Robust Estimation of a Location Parameter.*

Joreskog, K.g. (1999). *Formulas for Skewness and Kurtosis.*

Kay, S.M. (1997). *Fundamentals of Statistical Signal Processing: Estimation Theory.*

Kim, S.J., (1992). *The Metrically Trimmed Mean as a Robust Estimator of Location.*

Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions with Applications.*

Lee, J.Y. (2004). *Adaptive Choice of Trimming Proportions for Location Estimation of the Mean.*

Lomax, R.G., (2007). *An Introduction to Statistical Concepts.*

Magallanes, A.B. (2007). *Generalized Method of Moments Estimation on a Linear Panel Data Model of a Clinical Trial.*

Robbins, H. et al., (1983). *Adaptive choice of mean or median in estimating the center of a symmetric distribution.*

Saucier, R. (2000). *Computer Generation of the Statistical Distributions.*

Shao, J. (2003). *Mathematical Statistics (2nd Edition).*

Stefanski, L.A. et. Al., (2001). *The Calculus of M-Estimation.*

Stigler, S.M. (1972). *The Asymptotic Distribution of the Trimmed Mean.*

Teruel, J.E., (2010). *Identification of Probability Distribution: Fingerprinting Method.*

Websites:

http://math.uprm.edu/wrolke/esma6665/robust.htm

http://math.uprm.edu/wrolke/esma6665/robust.htm

http://www.itl.nist.gov/div898/handbook/eda/section3/eda356.htm

http://faculty.smu.edu/millimet/classes/eco7377/papers/wooldridge.pdf

http://www.experiencefestival.com/a/Kurtosis_-Terminology_and_examples

http://www.mathstatica.com/examples/MOM/

http://www.itl.nist.gov/div898/handbook/eda/section3/eda366.htm

http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm