

ON ASYMMETRIC ROBUST LINEAR ESTIMATORS

Al-Kassab. M. , M & Alkhazaleh, A. M. H.

Faculty of Science
AL-Albyt University
Al-Mafraq Jordan

Abstract

A linear trimmed mean estimator for symmetric and asymmetric distributions has been developed. Statistics such as skewness and tail length, all describe the distribution characteristics. Usually the problem of the intervention of the mean estimation is to truncate the discarded extreme values. In this paper, we propose a new measure estimator by dividing the whole datasets into groups. The boundaries of the groups are derived using Neyman allocation in order to determine the proportions of trimming on both sides of the datasets. The proposed method examines via a simulation study of seven hinge adaptive estimators for seven distributions. The objective is concentrated on these linear asymmetric robust estimators to demonstrate the efficiency of these estimators. The results are tested over seven distributions: Uniform (0,1), Beta (1,6), Beta (6,1), Chi square (2), Chi square (4), exponential (6) and gamma (2,3). The main advantage of the proposed method depends on the statistical base. Whereas, the partition of other methods should be specified previously. It has been shown that the proposed estimator for the groups from three to six do as well as the seven hinge estimators with better efficiency.

Keywords: robust estimation, probability density function, hinge estimators, Neyman allocation

1 Introduction

The most common basic statistical procedure in experimental science is calculating the average. It is used for estimating the central tendency in the presence of random variations in the observations. Data variations can be the result of variations in the experience of interest or of some certain measuring errors. The problem of sensitivity of the data points that deviate from the pattern set by the majority of the data set (Hampel et al., 1986). They have lead to the development of robust location measures. Robustness of an estimator is measured by the breakdown value, which tells us how many data points need to be replaced by arbitrary values in order to make the estimator blow up or blow down. For example, the arithmetic mean has 0% breakdown whilst the median is very robust with a breakdown value of 50% (Hampel et al., 1986). Usually, the problem of the intervention the mean estimation is to truncate the discarded extreme values. The Princeton Robust Study (PRS) has come up with the best of the robust estimates of location (Andrews, 1972). They proposed α -trimmed mean to choose the value of trimming proportion to minimize the estimated variance, and those estimators have proved the suitability of efficiency for some chosen symmetric distributions. In addition, the PRS was an extensive study of the finite sample behavior of 68 estimators. They examined many classes of estimators via simulation procedures over a variety of symmetric and contaminated symmetric distribution. Nevertheless, in practical applications, there is no agreement that the observed sources are symmetric. Therefore, a proper asymmetric truncations need to be made (Lee, 2003). Hogg (1974) suggested the asymmetric trimmed

algorithms. These algorithms were tested regarding to the tail length and skewness at the first phase, and then a ratio relating to the left and right trimmed (Hogg, 1974). Subsequently, Reed and Stark (1996) adopted the same procedure at the first step by using the hinge type adaptive location estimators to choose the ratio of truncation. In this paper, new classes of estimators are proposed depending on the probability density function by computing the boundaries of the groups by minimizing the variance inside the groups. Thus, the whole datasets are divided into three to six groups, more than six groups approximately give similar results. Then, seven symmetric and asymmetric distributions are used to compare the proposed methods with the seven hinge estimators proposed by Reed and Stark (1996).

2 Asymmetric Linear Estimators

Distributional characteristics for data sets are largely ignored even though many statistics that describe these data are readily available. Statistics such as skewness and tail length, both describe the distribution characteristics. Linear estimators are defined as a linear combination of the sample order statistics with fixed coefficients (Jaekel,1971a). While, the estimator varies depending on the sample observations, the estimator is called an adaptive or an adaptive linear estimator (Reed and Stark, 1994). Adaptive estimators commonly adapt the value of the weights of the estimator, depending on one or more estimates of distributional criteria. This common class is further subdivided by difficulty. Simple adaptive estimators are selected from several linear estimators, depending on the value of some useful distributional criterion, such as tail length or skewness (Reed & Stark, 1994). Derived from the work of Hogg (1974, 1982), Reed & Stark (1996) defined seven adaptive location estimators based on measuring of tail length and skewness for a set of n observations. They adopted the notation of Hogg (1974, 1982) to define these measures. Based on the ordered values, they assumed $L(\alpha)$ be the mean of the smallest $[\alpha n]$ observations, $[\alpha n]$ means that αn is rounded down to the nearest integer where α is the proportion value, and $U(\alpha)$ be the mean of the largest $[\alpha n]$ observations. When $\alpha = 0.05$ and, therefore, $L(0.05)$ is the mean of the smallest $0.05n$ observations, B is the mean of the next largest $0.15n$ observations, C is the mean of the next largest $0.30n$ observations, D is the mean of the next largest $0.30n$ observations, E is the mean of the next largest $0.15n$ observations and $U(0.05)$ be the mean of the largest $0.05n$ observations. Hogg (1974) defined two measures of tail length, Q and Q_1 , where

$$Q = (U_{(0.05)} - L_{(0.05)}) / (U_{(0.50)} - L_{(0.50)})$$

$$Q_1 = (U_{(0.20)} - L_{(0.20)}) / (U_{(0.50)} - L_{(0.50)})$$

where U denoted the upper, and L denoted the lower; Later Hogg (1982) introduced another measure of tail length, namely:

$$H_3 = (U_{(0.20)} - L_{(0.20)}) / (E - B)$$

where, E and B are as defined above. With this measure, if the values of $H_3 < 1.26$ it suggests that the tails of the distribution are similar to a uniform distribution, while, values between 1.26 and 1.76 suggest a normal distribution and values greater than 1.76 suggest

the tails that are similar to those of a double exponential distribution (Keselman et al. 2008). Reed & Stark (1996) define four measures of skewness as:

$$\begin{aligned}
 Q_2 &= (U_{(0.05)} - T_{(0.25)}) / (T_{(0.25)} - L_{(0.05)}) \\
 H_1 &= (U_{(0.05)} - D) / (C - L_{(0.05)}) \\
 SK_2 &= (X_{(1)} - XMD) / (XMD - X_{(n)}) \\
 SK_5 &= (X_{(1)} - XM) / (XM - X_{(n)})
 \end{aligned}$$

where, D and C are as defined above and XMD is the median, XM is the arithmetic mean, $T_{(0.25)}$ is the 0.25-trimmed mean ($T_{(\alpha)}$) given below and $X(1)$ and $X(n)$ are, the first and last ordered observations respectively. Then, the α -trimmed mean ($T_{(\alpha)}$) is defined as

$$T_{\alpha} = \frac{1}{n(1-2\alpha)} \left[\sum_{i=k+1}^{n-k} X_i + (k - \alpha n)(X_k + X_{n-k+1}) \right]$$

where X_i is the i -th ordered observation and $k = [\alpha n] + 1$. Based on the former definition of tail length and skewness, Reed & Stark (1996) proposed a set of adaptive linear estimators ‘that have the facility of asymmetric trimming’. They defined a general proposal for their approach as follows:

- (1) They set the value for the total amount of trimming from the sample, which is the value of α .
- (2) They determined the proportion to be trimmed from the lower end of the sample (α_L) by the proportion $\alpha_L = \alpha[UW_X / (UW_X + LW_X)]$, where UW_X and LW_X are the numerator and denominator portions of the previously defined selector statistics.
- (3) They defined the upper trimming proportion by $\alpha_2 = \alpha - \alpha_1$. Based on this general system, Reed & Stark (1996) defined seven ‘hinge’ estimators, which are trimmed means, as the following:

$$\begin{aligned}
 HQ & \quad \alpha_1 = \alpha[UW_Q / (UW_Q + LW_Q)], \\
 HQ_1 & \quad \alpha_1 = \alpha[UW_{Q_1} / (UW_{Q_1} + LW_{Q_1})], \\
 HH_3 & \quad \alpha_1 = \alpha[UW_{H_3} / (UW_{H_3} + LW_{H_3})], \\
 HQ_2 & \quad \alpha_1 = \alpha[UW_{Q_2} / (UW_{Q_2} + LW_{Q_2})], \\
 HH_1 & \quad \alpha_1 = \alpha[UW_{H_1} / (UW_{H_1} + LW_{H_1})], \\
 HSK_2 & \quad \alpha_1 = \alpha[UW_{SK_2} / (UW_{SK_2} + LW_{SK_2})], \text{ and} \\
 HSK_5 & \quad \alpha_1 = \alpha[UW_{SK_5} / (UW_{SK_5} + LW_{SK_5})],
 \end{aligned}$$

They found from a simulation study, that $T_{0.10}$, $T_{0.15}$, HSK_2 and HSK_5 are the most efficient estimators when the distribution is symmetric. When the distribution is asymmetric, they found that ‘HQ, HQ₁, HQ₂, HH₁, HSK₂ and HSK₅ [were] always among the top four estimators, with HQ₁ and HQ₂ in the top three’ (Reed & Stark, 1996). The

asymmetric trimming mean is defined by assuming that $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$ characterize the ordered observations. Then, an α -trimmed mean is defined by:

$$T(\alpha_1, \alpha_2) = \frac{1}{h} \sum_{i=g_1+1}^{n-g_2} x_{(i)}$$

where α is usually selected so that $g_1 = [n\alpha_1]$ and $g_2 = [n\alpha_2]$, where $h = n - g_1 - g_2$ in the notation of Reed and Stark, α_1 and α_2 would be α_L and α_U respectively.

3 New linear estimators

Let $f(x)$ be a probability density (*pdf*). Let x_0 and x_L be the smallest and the largest values of x in the whole dataset. Then, in order to determine the boundaries y_1, y_2, \dots, y_{L-1} of the groups, we need to compute the relative frequency W_k , the population mean μ_k and the population variance S_k^2 . Then, the relative frequency can be computed as the follows:

$$W_k = \int_{y_{k-1}}^{y_k} f(x) dx$$

The mean μ_k is:

$$\mu_k = \frac{1}{W_k} \int_{y_{k-1}}^{y_k} x f(x) dx$$

And, the variance S_k^2 is:

$$S_k^2 = \frac{1}{W_k} \int_{y_{k-1}}^{y_k} x^2 f(x) dx - \mu_k^2$$

While, the variance depending on Neyman allocation is :

$$V_{Ney}(\bar{x}_{st}) = \frac{1}{n} \left(\sum_{k=1}^L W_k \sigma_k \right)^2$$

By assuming $\{y_k\}$ to be the stratum boundaries on the domain (a, b) and by minimizing the variance according to Neyman allocation and taking the derivative of the equation with respect to y_k , yield the equations for determining the boundaries as :

$$\frac{(y_k - \mu_k)^2 + S_k^2}{S_k} = \frac{(y_k - \mu_{k+1})^2 + S_{k+1}^2}{S_{k+1}}$$

where $k = 1, 2, \dots, L-1$. (Cochran, 1977).

These equations are difficult to solve, since μ_k and S_k^2 depend on y_k and y_k also depends on μ_k and S_k^2 . The iterative method is used to solve these equations. A computer program written in R language for three up to six groups is used to solve these equations. For example, in three groups, group one is (y_0, y_1) , (y_1, y_2) is group two and the last group is (y_2, y_3) , where $x_0 = y_0$ is the smallest value in the dataset, and the largest value in the dataset is $x_k = y_3$. We repeat the procedure for four, five and six groups. Now, we define the new estimator which is the asymmetric trimming mean as follows: assume that x_1, x_2, \dots, x_n are the ordered dataset and let y_1, y_2, \dots, y_{L-1} be the boundaries of the group computed according to Neyman allocation. Let $x_1 = y_0$ and $x_n = y_L$ be the smallest and largest values. We use the following notation:

LG the mean of the observations among the first group (y_0, y_1) .

MG the mean of the observations among the middle group[s].

UG the mean of the observations among the last group (y_{L-1}, y_L) . Then, the new linear measure estimator according to Neyman allocation is defined as:

$$HN = [UG - MG / MG - LG]$$

If the dataset is divided into three groups then the LG is the mean of the lowest group which is the first group, MG is the mean of the middle group which is second group, and UG is the mean of the last group.

In order to determine a proportion to be trimmed from the lower end of the sample (α_L), we use the following proportion

$$\alpha_L = \alpha [UHN / (UHN + LHN)]$$

where, UHN and LHN are the numerator and the denominator of the estimator HN respectively, α is the total trimming proportion to be trimmed from the sample and the upper trimming proportion can be defined by $\alpha_U = \alpha - \alpha_L$. Let $\alpha_L = r$ and $\alpha_U = s$, then, the asymmetric trimming mean is calculated as follows

$$T_{(r,s)} = \frac{1}{n - r - s} \left[\sum_{i=r+1}^{n-s} X_i \right]$$

Tail length, skewness and new estimators derived from Neyman allocations are statistics for the symmetric and asymmetric distributions are given in Table 1. For the symmetric distribution all of these skewness measures are basically equal to 1 (range 1.01-1.02), while the new estimators also have equal value. A quick examination of Table 1 shows that 3G, 4G, 5G and 6G are increasing as the number of groups increase except at Beta (6,1) since it is left skewed.

Table 1 mean selector statistics

Distribution	New estimators				Tail length			Skewness			
	3G	4G	5G	6G	Q	Q1	H3	Q2	H1	SK2	SK5

Uniform(0,1)	1.01	1.01	1.01	1.01	1.91	1.60	1.01	1.02	1.01	1.02	1.01
Gamma(2,3)	1.63	1.88	2.07	2.21	2.65	1.77	2.27	5.30	2.27	0.29	0.38
Beta(1,6)	1.65	1.90	2.08	2.22	2.51	1.74	2.53	5.55	2.53	0.24	0.34
Beta(6.1)	1.27	0.94	0.79	0.70	2.51	1.74	0.40	0.92	0.40	4.34	3.03
Chi sq. (2)	1.94	2.34	2.67	2.93	2.81	1.80	3.29	7.41	3.29	0.16	0.26
Chi sq. (4)	1.64	1.90	2.09	2.24	2.67	1.77	2.27	5.33	2.28	0.29	0.37
Exp(6)	1.27	1.76	2.14	2.41	2.83	1.80	3.35	7.55	3.35	0.16	0.25

4 Simulation Methods

Reed and Stark (1996) introduced the natural logarithm of the relative efficiency of the estimator. The natural logarithm has a 0 reference point and is symmetrical in that $\ln 2 = -\ln(\frac{1}{2})$. Also, in the Princeton Robust Study, efficiency was introduced to provide a basis for comparing two estimators. The relative efficiency (RE) of an estimator relative to a standard estimator is defined as the ratio of the variance of the standard estimator to the variance of the estimator in question (Reed and Stark, 1996). They also computed the natural logarithm of the relative efficiency. While, in this paper we computed the relative efficiency (RE) for the proposed asymmetric estimators and the asymmetric hinge estimators as suggested by Reed and Stark relative to the minimum estimator of the proposed method, this is defined as the ratio of the variance of all hinge estimators to the variance of the minimum estimator proposed by our method (from three to six groups with proportions 0.05, 0.10, 0.15, and 0.20), the relative efficiency can be defined as :

$$RE = \frac{\text{var}(all\ methods)}{\text{var}(Minimum\ proposed\ Estimator)}$$

Random observations were generated from each of the seven distributions using R program. The distributions are the Uniform (0,1), Beta (1,6), Beta (6,1), Chi square (2), Chi square (4), exponential (6) and gamma (2,3). Sample sizes (n) of 50, 100, and 150 were generated for each iteration (N) of 1000, 2500, and 5000. Since similar results were found for all iteration values, we tabulate the results for the case $N = 2500$ iterations. The estimators were ordered within each table based on the type of the distribution.

Table 2 Linear estimators for Uniform (0,1) distribution (2500) iterations

Est.	Prop.	Value of Est.	variance	RE	Bias
------	-------	---------------	----------	----	------

<i>3G</i>	0.05	0.49867	0.08237	1	-0.0242
	0.10	0.49862	0.08240	1.00033	-0.0242
	0.15	0.49857	0.08242	1.00044	-0.0241
	0.20	0.49852	0.08243	1.00070	-0.0240
<i>4G</i>	0.05	0.49867	0.08237	1	-0.0242
	0.10	0.49868	0.08240	1.00028	-0.0242
	0.15	0.49857	0.08242	1.00033	-0.0241
	0.20	0.49857	0.08242	1.00052	-0.0241
<i>5G</i>	0.05	0.49867	0.08237	1	-0.0242
	0.10	0.49865	0.08239	1.00024	-0.0242
	0.15	0.49859	0.08240	1.00027	-0.0241
	0.20	0.49854	0.08241	1.00039	-0.0241
<i>6G</i>	0.05	0.49867	0.08237	1	-0.0242
	0.10	0.49864	0.08239	1.00021	-0.0242
	0.15	0.49849	0.08239	1.00023	-0.0240
	0.20	0.49849	0.08240	1.00030	-0.0240
<i>HQ</i>	0.05	0.50857	0.08339	1.01239	-0.0341
	0.10	0.51346	0.0826	1.00270	-0.0390
	0.15	0.51987	0.08285	1.00574	-0.0454
	0.20	0.52998	0.08339	1.01239	-0.0555
<i>HQ1</i>	0.05	0.50837	0.08247	1.00115	-0.0339
	0.10	0.51326	0.08259	1.00265	-0.0388
	0.15	0.51815	0.08277	1.00478	-0.0437
	0.20	0.52307	0.08300	1.00755	-0.0486
<i>HH1</i>	0.05	0.50813	0.08247	1.00110	-0.0337
	0.10	0.51379	0.08261	1.00284	-0.0393
	0.15	0.52163	0.08291	1.00651	-0.0472
	0.20	0.52925	0.08332	1.01147	-0.0548
<i>HQ2</i>	0.05	0.49868	0.08237	1	-0.0242
	0.10	0.50357	0.08240	1.00035	-0.0291
	0.15	0.50845	0.08248	1.00132	-0.0340
	0.20	0.51138	0.08258	1.00244	-0.0369
<i>HH3</i>	0.05	0.49867	0.08237	1	-0.0242
	0.10	0.49866	0.08238	1.00012	-0.0242
	0.15	0.4988	0.08238	1.00009	-0.0243
	0.20	0.49861	0.08238	1.00009	-0.0241
<i>HSK2</i>	0.05	0.49868	0.08238	1.00007	-0.0242
	0.10	0.49851	0.08243	1.00072	-0.0240
	0.15	0.49831	0.08250	1.00151	-0.0238
	0.20	0.49813	0.08259	1.00266	-0.0237
<i>HSK5</i>	0.05	0.49868	0.08237	1	-0.0242
	0.10	0.49844	0.08242	1.00060	-0.0240
	0.15	0.49834	0.08244	1.00082	-0.0239
	0.20	0.49836	0.08250	1.00148	-0.0239

Table 3 Linear estimators for Beta (1,6) distribution (2500) iterations

Est.	Prop.	Value of Est.	variance	RE	Bias
<i>3G</i>	0.05	0.14148	0.01515	1.00040	0.01757
	0.10	0.13876	0.01516	1.00140	0.02029
	0.15	0.13721	0.01518	1.00251	0.02184
	0.20	0.13625	0.01519	1.00339	0.02280
<i>4G</i>	0.05	0.14251	0.01514	1.00006	0.01654
	0.10	0.13999	0.01515	1.00072	0.01906
	0.15	0.13918	0.01516	1.00117	0.01987
	0.20	0.13881	0.01516	1.00141	0.02024
<i>5G</i>	0.05	0.14273	0.01514	1.00001	0.01632
	0.10	0.14082	0.01515	1.00044	0.01823
	0.15	0.14053	0.01515	1.00057	0.01852
	0.20	0.14051	0.01515	1.00059	0.01854
<i>6G</i>	0.05	0.14279	0.01514	1	0.01626
	0.10	0.14156	0.01515	1.00029	0.01749
	0.15	0.14147	0.01515	1.00029	0.01758
	0.20	0.14147	0.01515	1.00030	0.01735
<i>HQ</i>	0.05	0.14281	0.01515	1.00037	0.01623
	0.10	0.14385	0.01515	1.00018	0.01520
	0.15	0.14305	0.01515	1.00017	0.01600
	0.20	0.14435	0.01515	1.00037	0.01470
<i>HQ1</i>	0.05	0.14281	0.01514	1.00060	0.01623
	0.10	0.13991	0.01515	1.00073	0.01914
	0.15	0.13794	0.01517	1.00194	0.02111
	0.20	0.13665	0.01519	1.00305	0.02240
<i>HH1</i>	0.05	0.14845	0.01517	1.00206	0.01060
	0.10	0.14991	0.01519	1.00333	0.00914
	0.15	0.15294	0.01525	1.00690	0.00611
	0.20	0.15678	0.01534	1.01305	0.00227
<i>HQ2</i>	0.05	0.14224	0.01514	1.00010	0.01681
	0.10	0.13937	0.01516	1.00094	0.01967
	0.15	0.13762	0.01517	1.00204	0.02143
	0.20	0.13664	0.01519	1.00290	0.02241
<i>HH3</i>	0.05	0.14282	0.01514	1	0.01623
	0.10	0.14329	0.01514	1.00013	0.01576
	0.15	0.14338	0.01514	1.00012	0.01567
	0.20	0.14422	0.01515	1.00020	0.01483
<i>HSK2</i>	0.05	0.13034	0.01531	1.01118	0.02871
	0.10	0.12073	0.01565	1.03357	0.03832
	0.15	0.11265	0.01608	1.06210	0.04640
	0.20	0.1054	0.01659	1.09526	0.05365
<i>HSK5</i>	0.05	0.13267	0.01525	1.00721	0.02638
	0.10	0.1231	0.01555	1.02675	0.03595
	0.15	0.11581	0.01590	1.04997	0.04324
	0.20	0.10934	0.01630	1.07643	0.04971

Table 4 Linear estimators for Beta (6,1) distribution

Est.	Prop.	Value of Est.	variance	RE	Bias
<i>3G</i>	0.05	0.86326	0.01513	1.00225	-0.0294
	0.10	0.86925	0.01525	1.00973	-0.0354
	0.15	0.87386	0.01539	1.01905	-0.0400
	0.20	0.87770	0.01554	1.02925	-0.0438
<i>4G</i>	0.05	0.86219	0.01512	1.00136	-0.0283
	0.10	0.86620	0.01518	1.00538	-0.0323
	0.15	0.86948	0.01526	1.01048	-0.0356
	0.20	0.87226	0.01534	1.01613	-0.0384
<i>5G</i>	0.05	0.86146	0.01511	1.00101	-0.0276
	0.10	0.86428	0.01515	1.00326	-0.0304
	0.15	0.86684	0.01520	1.00646	-0.0330
	0.20	0.86885	0.01525	1.00984	-0.0350
<i>6G</i>	0.05	0.86032	0.01511	1	-0.0265
	0.10	0.86302	0.01513	1.00217	-0.0292
	0.15	0.86503	0.01516	1.00429	-0.0312
	0.20	0.8666	0.0152	1.00655	-0.0327
<i>HQ</i>	0.05	0.86715	0.01611	1.06696	-0.0333
	0.10	0.87638	0.01547	1.02471	-0.0425
	0.15	0.8823	0.01573	1.04214	-0.0484
	0.20	0.88879	0.01611	1.06696	-0.0549
<i>HQ1</i>	0.05	0.86715	0.01519	1.00624	-0.0333
	0.10	0.87321	0.01535	1.01686	-0.0393
	0.15	0.87826	0.01554	1.02949	-0.0444
	0.20	0.88272	0.01576	1.04359	-0.0489
<i>HH1</i>	0.05	0.86236	0.01512	1.00144	-0.0285
	0.10	0.86575	0.01517	1.00464	-0.0319
	0.15	0.86904	0.01524	1.00910	-0.0352
	0.20	0.87160	0.01530	1.01358	-0.0377
<i>HQ2</i>	0.05	0.86663	0.01519	1.00584	-0.0328
	0.10	0.87273	0.01534	1.01613	-0.0389
	0.15	0.87794	0.01554	1.02897	-0.0441
	0.20	0.88262	0.01576	1.04364	-0.0488
<i>HH3</i>	0.05	0.85732	0.01509	0.99996	-0.0235
	0.10	0.85691	0.01509	0.99990	-0.0230
	0.15	0.85675	0.01509	0.99999	-0.0229
	0.20	0.85596	0.01510	0.99997	-0.0221
<i>HSK2</i>	0.05	0.86972	0.01526	1.01071	-0.0359
	0.10	0.87934	0.01560	1.03310	-0.0455
	0.15	0.88746	0.01603	1.06183	-0.0536
	0.20	0.89455	0.01652	1.09435	-0.0607
<i>HSK5</i>	0.05	0.86740	0.01520	1.00673	-0.0335
	0.10	0.87694	0.01549	1.02616	-0.0431
	0.15	0.88437	0.01585	1.04987	-0.0505
	0.20	0.89070	0.01624	1.07590	-0.0568

Table 5 Linear estimators for Chi sq(2) distribution (2500) iterations

Est.	Prop.	Value of Est.	variance	RE	Bias
<i>3G</i>	0.05	1.92224	4.03218	1.00174	0.12212
	0.10	1.85645	4.04774	1.00561	0.18791
	0.15	1.80679	4.06612	1.01017	0.23756
	0.20	1.76837	4.08460	1.01477	0.27599
<i>4G</i>	0.05	1.96111	4.02691	1.00043	0.08325
	0.10	1.90439	4.03557	1.00259	0.13996
	0.15	1.87232	4.04418	1.00472	0.17203
	0.20	1.84931	4.05269	1.00684	0.19505
<i>5G</i>	0.05	1.97655	4.02553	1.00009	0.06781
	0.10	1.93406	4.03024	1.00126	0.11030
	0.15	1.9105	4.03535	1.00253	0.13386
	0.20	1.89974	4.03913	1.00347	0.14462
<i>6G</i>	0.05	1.98256	4.02523	1	0.0618
	0.10	1.95201	4.02794	1.00069	0.09235
	0.15	1.93854	4.03038	1.00130	0.10582
	0.20	1.93077	4.03323	1.00200	0.11359
<i>HQ</i>	0.05	1.97863	4.02584	1.00017	0.06573
	0.10	1.98692	4.02535	1.00005	0.05743
	0.15	1.98005	4.02562	1.00011	0.06431
	0.20	1.9808	4.02584	1.00017	0.06356
<i>HQ1</i>	0.05	1.97832	4.02519	1.00001	0.06604
	0.10	1.9085	4.03525	1.00250	0.13586
	0.15	1.86247	4.04769	1.00559	0.18189
	0.20	1.84253	4.05395	1.00715	0.20183
<i>HH1</i>	0.05	2.0874	4.03051	1.00133	-0.04300
	0.10	2.10544	4.03589	1.00266	-0.06110
	0.15	2.16432	4.05039	1.00627	-0.12000
	0.20	2.21486	4.06940	1.01099	-0.17050
<i>HQ2</i>	0.05	1.97709	4.02526	1.00002	0.06727
	0.10	1.91598	4.03305	1.00196	0.12838
	0.15	1.89050	4.03845	1.00330	0.15386
	0.20	1.87327	4.04322	1.00448	0.17108
<i>HH3</i>	0.05	1.99954	4.02558	1.00010	0.04481
	0.10	2.00426	4.02545	1.00008	0.04010
	0.15	2.01264	4.02536	1.00006	0.03172
	0.20	2.02607	4.02522	1	0.01829
<i>HSK2</i>	0.05	1.74612	4.09474	1.01728	0.29824
	0.10	1.58351	4.20987	1.04589	0.46085
	0.15	1.44542	4.34897	1.08044	0.59894
	0.20	1.32796	4.49714	1.11726	0.71639
<i>HSK5</i>	0.05	1.77468	4.08304	1.01438	0.26968
	0.10	1.61529	4.18429	1.03953	0.42907
	0.15	1.48892	4.30180	1.06873	0.55544
	0.20	1.38209	4.42629	1.09965	0.66226

Table 6 Linear estimators for Chi sq(4) distribution (2500) iterations

Est.	Prop.	Value of Est.	variance	RE	Bias
3G	0.05	3.9751	7.93431	1.00038	-0.2204
	0.10	3.92132	7.94066	1.00118	-0.1666
	0.15	3.89554	7.9463	1.00189	-0.1408
	0.20	3.87999	7.95055	1.00243	-0.1253
4G	0.05	4.00045	7.93175	1.00006	-0.2458
	0.10	3.9526	7.93536	1.00051	-0.1979
	0.15	3.94081	7.93713	1.00074	-0.1861
	0.20	3.93712	7.93772	1.00081	-0.1824
5G	0.05	4.00734	7.93133	1.00001	-0.2526
	0.10	3.97262	7.93338	1.00026	-0.2179
	0.15	3.97215	7.93351	1.00028	-0.2175
	0.20	3.97881	7.93309	1.00023	-0.2241
6G	0.05	4.00944	7.93128	1	-0.2547
	0.10	3.98917	7.93242	1.00014	-0.2345
	0.15	3.99586	7.93224	1.00012	-0.2412
	0.20	4.00538	7.93226	1.00012	-0.2507
HQ	0.05	4.0099	7.93952	1.00104	-0.2552
	0.10	4.05775	7.93463	1.00042	-0.3031
	0.15	4.05154	7.93450	1.00041	-0.2968
	0.20	4.08608	7.93952	1.00104	-0.3314
HQ1	0.05	4.00944	7.93128	1	-0.2552
	0.10	3.95045	7.93615	1.00061	-0.1957
	0.15	3.91442	7.94327	1.00151	-0.1597
	0.20	3.89911	7.94734	1.00202	-0.1444
HH1	0.05	4.14636	7.95102	1.00249	-0.3917
	0.10	4.18095	7.96205	1.00388	-0.4263
	0.15	4.24948	7.99158	1.00760	-0.4948
	0.20	4.33148	8.03856	1.01353	-0.5768
HQ2	0.05	4.00729	7.93129	1	-0.2526
	0.10	3.95005	7.93575	1.00056	-0.1954
	0.15	3.92901	7.93897	1.00097	-0.1743
	0.20	3.92167	7.94117	1.00125	-0.1670
HH3	0.05	4.00944	7.93128	1	-0.2549
	0.10	3.99802	7.93219	1.00011	-0.2433
	0.15	4.00666	7.93168	1.00005	-0.2520
	0.20	4.02055	7.93175	1.00006	-0.2659
HSK2	0.05	3.74521	8.00740	1.00960	0.00948
	0.10	3.53796	8.16363	1.02930	0.21673
	0.15	3.37511	8.35012	1.05281	0.37959
	0.20	3.23343	8.55640	1.07882	0.52126
HSK5	0.05	3.76895	7.99182	1.00763	-0.01430
	0.10	3.57335	8.13077	1.02515	0.18135
	0.15	3.42922	8.28340	1.04440	0.32548
	0.20	3.29900	8.45716	1.06630	0.45570

Table 7 Linear estimators for Exp(6) distribution (2500) iterations

Est.	Prop.	Value of Est.	variance	RE	Bias
<i>3G</i>	0.05	0.15866	0.02768	1.00235	0.01111
	0.10	0.15291	0.02781	1.00724	0.01686
	0.15	0.14822	0.02798	1.01332	0.02155
	0.20	0.14465	0.02814	1.01921	0.02512
<i>4G</i>	0.05	0.16272	0.02763	1.00054	0.00705
	0.10	0.15767	0.02770	1.00308	0.01210
	0.15	0.15486	0.02777	1.00556	0.01491
	0.20	0.15275	0.02784	1.00809	0.01703
<i>5G</i>	0.05	0.16419	0.02762	1.00008	0.00559
	0.10	0.16041	0.02765	1.00148	0.00936
	0.15	0.15853	0.02769	1.00270	0.01125
	0.20	0.15744	0.02772	1.00379	0.01233
<i>6G</i>	0.05	0.16458	0.02761	1	0.00519
	0.10	0.16228	0.02763	1.00069	0.00749
	0.15	0.16104	0.02765	1.00135	0.00873
	0.20	0.16051	0.02767	1.00190	0.00926
<i>HQ</i>	0.05	0.16446	0.02762	1.00014	0.00531
	0.10	0.16525	0.02761	1.00005	0.00452
	0.15	0.16444	0.02762	1.00013	0.00533
	0.20	0.16482	0.02762	1.00014	0.00495
<i>HQ1</i>	0.05	0.16444	0.02761	1.00002	0.00533
	0.10	0.15877	0.02768	1.00239	0.01100
	0.15	0.15497	0.02776	1.00536	0.01480
	0.20	0.15341	0.02780	1.00679	0.01637
<i>HH1</i>	0.05	0.17336	0.02765	1.00137	-0.00360
	0.10	0.17475	0.02769	1.00262	-0.00500
	0.15	0.17990	0.02779	1.00644	-0.01010
	0.20	0.18393	0.02792	1.01108	-0.01420
<i>HQ2</i>	0.05	0.16432	0.02761	1.00004	0.00545
	0.10	0.15935	0.02767	1.00190	0.01042
	0.15	0.15706	0.02771	1.00334	0.01271
	0.20	0.15554	0.02774	1.00456	0.01423
<i>HH3</i>	0.05	0.16593	0.02762	1.00011	0.00384
	0.10	0.16645	0.02761	1.00010	0.00332
	0.15	0.16715	0.02761	1.00010	0.00263
	0.20	0.16837	0.02761	1.00010	0.00140
<i>HSK2</i>	0.05	0.14528	0.02808	1.01698	0.02449
	0.10	0.13175	0.02887	1.04551	0.03802
	0.15	0.12003	0.02985	1.08098	0.04974
	0.20	0.11027	0.03088	1.11816	0.05950
<i>HSK5</i>	0.05	0.14778	0.02800	1.01394	0.02199
	0.10	0.13443	0.02869	1.03911	0.03535
	0.15	0.12383	0.02951	1.06860	0.04594
	0.20	0.11488	0.03037	1.09997	0.05489

Table 8 Linear estimators for Gamma (2,3) distribution (2500) iterations

Est.	Prop.	Value of Est.	variance	RE	Bias
<i>3G</i>	0.05	5.94215	17.7747	1.00042	0.37751
	0.10	5.86170	17.7898	1.00127	0.45796
	0.15	5.82338	17.8027	1.00200	0.49628
	0.20	5.80360	17.8109	1.00246	0.51606
<i>4G</i>	0.05	5.98313	17.7684	1.00006	0.33653
	0.10	5.91281	17.7768	1.00054	0.40685
	0.15	5.89681	17.7801	1.00072	0.42285
	0.20	5.89336	17.7811	1.00078	0.42630
<i>5G</i>	0.05	5.99417	17.7673	1.00001	0.32549
	0.10	5.94571	17.7717	1.00025	0.37395
	0.15	5.94285	17.7720	1.00027	0.37681
	0.20	5.95190	17.7716	1.00024	0.36776
<i>6G</i>	0.05	5.99795	17.7672	1	0.32171
	0.10	5.97087	17.7695	1.00013	0.34879
	0.15	5.97523	17.7694	1.00012	0.34443
	0.20	5.99407	17.7693	1.00012	0.32559
<i>HQ</i>	0.05	5.99854	17.7856	1.00104	0.32112
	0.10	6.06802	17.7746	1.00041	0.25164
	0.15	6.06382	17.7744	1.00040	0.25584
	0.20	6.11229	17.7856	1.00104	0.20737
<i>HQ1</i>	0.05	5.99795	17.7672	1	0.32112
	0.10	5.90893	17.7787	1.00064	0.41073
	0.15	5.85571	17.7946	1.00154	0.46395
	0.20	5.83150	17.8045	1.00210	0.48816
<i>HH1</i>	0.05	6.20374	17.8112	1.00247	0.11592
	0.10	6.25357	17.8351	1.00382	0.06609
	0.15	6.35826	17.9026	1.00762	-0.03860
	0.20	6.48173	18.0082	1.01356	-0.16210
<i>HQ2</i>	0.05	5.99338	17.7674	1.00001	0.32628
	0.10	5.90829	17.7776	1.00058	0.41137
	0.15	5.87748	17.7848	1.00099	0.44218
	0.20	5.86926	17.7882	1.00118	0.45040
<i>HH3</i>	0.05	5.99762	17.7672	1	0.32204
	0.10	5.98146	17.7690	1.00010	0.33820
	0.15	5.99276	17.7681	1.00005	0.32690
	0.20	6.01381	17.7680	1.00004	0.30584
<i>HSK2</i>	0.05	5.5966	17.9442	1.00996	0.72306
	0.10	5.28888	18.2947	1.02969	1.03078
	0.15	5.04708	18.7099	1.05305	1.27258
	0.20	4.83695	19.1677	1.07882	1.48271
<i>HSK5</i>	0.05	5.63704	17.9043	1.00772	0.68262
	0.10	5.35052	18.2093	1.02488	0.96914
	0.15	5.12946	18.5574	1.04447	1.19020
	0.20	4.93947	18.9363	1.06580	1.38019

5 Results

For these classes of estimators, the obvious interest is the behavior of our estimator that gives the minimum variance. Within the symmetric distribution which is the Uniform distribution (0,1) (Table 2), the values of the estimators are closed, so that the values of the variances of the estimators and relative efficiency are approximately equal for the different groups. Also the values of the estimators (*HH3*, *HSK2*, and *HSK5*) are closed to the value of the new estimator. While, for the estimators (*HQ*, *HQ1*, *HH1*, and *HQ2*), the new estimator have minimum variance compared with the variances of these estimators. Within the asymmetric distributions Gamma (2,3), Chi square (2), Chi square (4) , Beta (1,6), and Exponential (6), the values of the estimators increased slowly as the number of groups increased, and decreased as the values of proportions increased, the behavior of the estimator for Beta (6,1) is counterproductive since it is negative skewed. The optimal number of groups that gives minimum variance when $\alpha = 0.05, 0.10, 0.15,$ and 0.20 are when the data divided into 6 groups.

6. Conclusion

Most of the previous studies on the problem of the intervention of the mean estimation data sets are largely ignored the type of the distribution of the datasets. But statistics such as skewness and tail length, both describe the distribution characteristics. In this paper, we proposed a new measure estimator by dividing the whole datasets into groups with respect to the probability density function. The boundaries of the groups derived using Neyman allocation in order to determine the proportions of trimming on both sides of the datasets. These boundaries are divided regarding the type of the probability density function. The proposed method was tested via a simulation study over the seven hinge adaptive estimators for seven distributions. The relative efficiency (RE) is defined as the ratio of the variance of the minimum estimator of the proposed method (from three to six groups with proportions 0.05, 0.10, 0.15, and 0.20) to the variance that derived from other hinge estimators with the same values of proportions. The variances of the proposed method are approximately lower than the variances derived from the hinge estimators (*HQ*, *HQ1*, *HH1*, *HQ2*, *HH3*, *HSK2*, *HSK5*) or equal. This means that the relative efficiency for these estimators are greater than or equal to one.

References:

- Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., & Tukey, J. W., (1972). Robust estimates of location. Princeton, NJ: Princeton University Press.
- Cochran, WG., (1977). Sampling techniques, 3rd edition. New York: John Wiley, sons.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A., (1986). Robust statistics. John Wiley & Sons, New York.
- Hogg R. V., (1982), On adaptive statistical inferences. Comm. Stat.-Theory Method 11, 2531-254.
- Hogg, R.V. (1974) Adaptive robust procedures: a partial review and some suggestions for future application and theory, J. Am. Stat. Assoc. 69, 909-923.
- Jaekel, L. A., (1971a) robust estimates of location: symmetry and asymmetric contamination, Annals of mathematical statistics, 42, 1020-1034.

Keselman H. J. , Wilcox R. R., Algina J., Othman A. R. & Fradette K. (2008). A comparative study of robust tests for spread: Asymmetric trimming strategies British Journal of Mathematical and Statistical Psychology (2008), 61, 235–253.

Lee J.,Y., (2003) Low bias mean estimator for symmetric and asymmetric unimodal distributions Computer Methods and Programs in Biomedicine 72, 99-107.

Reed J.F. & Stark D.R. (1994) Robustness estimators of location: a classification of linear and adaptive estimators. Journal of applied statistics, 21, No 3, 95-124.

Reed J.F. & Stark D.R. (1996) Hinge estimators of location: Robust to asymmetry Computer Methods and Programs in Biomedicine 49, 11-17.