More Efficient L-Estimates

Joshua D. Naranjo
Western Michigan University

Abstract

We propose more efficient L-estimates by using pairwise averages of the observations instead of the observations themselves. For instance, we show that minimum variance quantile estimation of the mean parameter in the exponential distribution improves from 65% to 88%. Simulations show similar improvements in frequently used scale and location estimators like the interquartile range, MAD, and trimmed mean.

Key words: L-estimates; Walsh Averages; Quantiles; Efficiency.

1 Introduction

L-estimators are statistics based on the sample quantiles, or order statistics. The appeal of L-estimators in applications lies in their relative simplicity and robustness (e.g. the median and interquartile range). However, their low efficiency often discourages potential users.

In this paper we propose more efficient L-estimates by calculating *the same estimates* on the ordered pairwise averages (Walsh averages) instead of the regular ordered observations. The estimates are then restandardized for consistency. The procedure is straightforward, the gains in efficiency often dramatic, and minimal user simplicity is sacrificed. An illustration of the procedure follows.

Let $\underline{X} = (x_1, \cdots, x_n)$ be a random sample from the cdf F. Imagine that the set of $\binom{n}{2}$ Walsh averages $W = \{(x_i + x_j)/2,\ i < j\}$ is a random sample from the cdf F_2 of $(X_1 + X_2)/2$. Note that if F is symmetric with mean μ and variance σ^2 then F_2 is symmetric with mean μ and variance $\sigma^2/2$. Now if $L(\underline{X})$ is an L-estimator for μ, we propose using $L(\underline{W})$ instead. If $S(\underline{X})$ is an L-estimator for σ, we propose using $\sqrt{2}S(\underline{W})$ instead (although in Section 3, we propose a correction factor for nonindependence of the Walsh averages in constructing scale estimates based on W).

More generally, consider the empirical cdf $F_n(x)$, the step function with jumps of size $1/n$ at the observations $x_1, \cdots, x_n$. For $0 < p < 1$ define the pth sample quantile as $F_n^{-1}(p) = \inf\{x : F_n(x) \geq p\}$. The sample quantile $F_n^{-1}(p)$ is consistent for the distribution quantile $F^{-1}(p)$, and is asymptotically normal with variance $p(1-p)/[F'(F^{-1}(p))]^2 n$, assuming the derivative exists. Now consider the empirical cdf $F_{2,n}(x)$ with jumps of size $\binom{n}{2}^{-1}$ at the Walsh averages. The empirical cdf $F_{2,n}(x)$ is consistent for $F_2(x)$, the cdf of the average $(X_1 + X_2)/2$, and $F_{2,n}^{-1}$ is similarly consistent for F_2^{-1} (under weak conditions). If both F and F_2 are indexed by a parameter θ then estimation of θ may be based on either F_n or $F_{2,n}$. For instance, the center of symmetry may be estimated by either median $F_n^{-1}(1/2)$ or $F_{2,n}^{-1}(1/2)$. At the normal distribution, the asymptotic relative efficiencies are 64% and 95%, respectively, a considerable difference (see e.g. Hettmansperger, 1984, Sec. 2.6).

In Example 1 of Section 3, we show an equally dramatic improvement in minimum variance quantile estimation of the exponential parameter θ. A simulation study in Example 2 suggests similar improvements for the interquartile range and median absolute deviation in the estimation of the standard deviation σ. In addition to efficiency, bias is decreased when the underlying distribution has thicker tails than the assumed normal distribution.

The outline of the article is as follows. In Section 2 we derive the asymptotic variance and prove asymptotic normality of the quantiles of $F_{2,n}$. Section 3 presents applications to L-estimates.

2 Asymptotic Normality of Quantiles

The results of this section rely heavily on the fact that the empirical cdf $F_{2,n}(t) = \binom{n}{2}^{-1} \sum\sum_{i<j} u((X_i + X_j)/2 \leq t)$ is a U-statistic with kernel $h(x_1, x_2) = u\left(\frac{1}{2}(x_1 + x_2) \leq t\right)$, where u is the indicator function.

The following two lemmas state consistency and asymptotic normality of $F_{2,n}(x)$. Let $\xrightarrow{wp1}$ and $\xrightarrow{d}$ denote convergence with probability 1 and convergence in distribution, respectively.

Lemma 2.1 $F_{2,n}(x) \xrightarrow{wp1} F_2(x)$.

Lemma 2.2 $\sqrt{n}\{F_{2,n}(t) - F_2(t)\} \xrightarrow{d} N(0, 4VarF(2t - X))$.

The next lemma puts a bound on the rate of convergence to normality of the distribution of $F_{2,n}(t)$ (a Berry-Esseen theorem).

Lemma 2.3 *For a fixed t,*

$$\sup_{-\infty<s<\infty} \left| P\left(\frac{\sqrt{n}(F_{2,n}(t) - F_2(t))}{2\sqrt{VarF(2t - X)}} \leq s\right) - \Phi(s) \right| \leq \frac{C\ [VarF(2t - X)]^{-3/2}}{8\sqrt{n}}$$

where Φ is the standard normal cdf and C is an absolute constant.

Lemmas 2.1 - 2.3 follow directly from standard asymptotic results on U-statistics. See for example Serfling (1980, Sec. 5.5).

The following two theorems establish consistency and asymptotic normality of the quantiles. We use the notation $q_p = F_2^{-1}(p)$ and $\hat{q}_{pn} = F_{2,n}^{-1}(p)$.

Theorem 2.1 *If q_p is the unique solution to $F_2(x-) \leq p \leq F_2(x)$, then $\hat{q}_{pn} \xrightarrow{a.s.} q_p$.*

PROOF: By definition of q_p and the uniqueness condition, for $\epsilon > 0$, we have $F_2(q_p - \epsilon) < p < F_2(q_p + \epsilon)$. By Lemma 2.1,

$$P[F_{2,j}(q_p - \epsilon) < p < F_{2,j}(q_p + \epsilon),\ all\ \ j \geq n] \to 1, \quad n \to \infty.$$

Since $F_{2,j}(q_p - \epsilon) < p$ implies $q_p - \epsilon < \hat{q}_{pj}$, and $F_{2,j}(q_p + \epsilon) > p$ implies $q_p + \epsilon \geq \hat{q}_{pj}$, we have $P[q_p - \epsilon < \hat{q}_{pj} < q_p + \epsilon,\ all\ \ j \geq n] \to 1, \quad n \to \infty$. This proves Theorem 2.1.

Theorem 2.2 *Suppose that F_2 is continuous and differentiable (with derivative f_2) at q_p. Then*

$$\sqrt{n}\{\hat{q}_{pn} - q_p\} \xrightarrow{d} N\left(0, \frac{4\gamma(p)}{[f_2(q_p)]^2}\right) \tag{2.1}$$

where $\gamma(p) = \ VarF(2q_p - X)$.

PROOF: Let $H(t) = P\left(\frac{\sqrt{n}f_2(q_p)}{2\sqrt{\gamma(p)}}(\hat{q}_{pn} - q_p) \leq t\right) = P\left(p \leq F_{2,n}(q_p + \frac{2\sqrt{\gamma(p)}}{\sqrt{n}f_2(q_p)}t)\right)$. We want to show $|H(t) - \Phi(t)| \to 0$, where Φ is the standard normal cdf. Now, let $a_n = \frac{2\sqrt{\gamma(p)}}{\sqrt{n}f_2(q_p)}$, $b_{n,t} = q_p + a_n t$, and $c_{n,t} = \frac{\sqrt{n}[p - F_2(b_{n,t})]}{2\sqrt{VarF(2b_{n,t} - X)}}$. Then

$$\begin{aligned}\Phi(t) - H(t) &= P(F_{2,n}(q_p + a_n t) < p) - 1 + \Phi(t) \\ &= P\left(\frac{\sqrt{n}[F_{2,n}(b_{n,t}) - F_2(b_{n,t})]}{2\sqrt{VarF(2(b_{n,t}) - X)}} \leq c_{n,t}\right) - 1 + \Phi(t) \\ &= P(Z_n(b_{n,t}) < c_{n,t}) - \Phi(c_{n,t}) - \Phi(-c_{n,t}) + \Phi(t)\end{aligned}$$

where $Z_n(y) = \{\sqrt{n}[F_{2,n}(y) - F_2(y)]\}/\{2\sqrt{VarF(2(y) - X)}\}$. By Lemma 2.3, $|P(Z_n(b_{n,t}) < c_{n,t}) - \Phi(c_{n,t})| \leq \frac{CVarF(2q_p - X)}{8\sqrt{n}} \to 0$. It remains to show $|\Phi(-c_{n,t}) - \Phi(t)| \to 0$, or equivalently $-c_{n,t} \to t$, as $n \to \infty$. Now,

$$-c_{n,t} = t\frac{2\sqrt{\gamma(p)}}{\sqrt{n}f_2(q_p)} \frac{\sqrt{n}}{2\sqrt{VarF(2(b_{n,t}) - X)}} \frac{F_2(q_p + a_n t) - p}{a_n t} \to t$$

since $a_n t \to 0$ and $b_{n,t} \to q_p$ as $n \to \infty$. This proves Theorem 2.2.

The variance of the pth sample quantile $F_{2,n}^{-1}(p)$ depends on the underlying distribution $F(x)$ through $\gamma(p) = Var\ F(2F_2^{-1}(p) - X)$ and through $f_2(F_2^{-1}(p))$. The asymptotic variance in (2.1) reduces to $1/12(\int f^2)^2$ for p=1/2 and f symmetric. The term $\int f^2$ may also be written as $E_f(f(X)$ or the expected height of the density, and is an inverse measure of spread. In Example 1 of the next section we illustrate the usefulness of (2.1) in constructing minimum variance parameter estimates based on ordered Walsh averages.

3 More Efficient L-estimates

3.1 Example 1. Exponential(θ)

Consider estimating the mean θ of the exponential distribution. Given a random sample $X_1, \cdots, X_n$, let $X_{(k)}$ be the kth order statistic and let δ_k be such that $\delta_k X_{(k)}$ is unbiased for θ. Siddiqui (1963) has shown that the minimum asymptotic variance among quantile estimators $\delta_k X_{(k)}$ is attained at $k \approx .8n$. For large n, the minimum variance is $1.544\,\theta^2/n$ so that the asymptotic efficiency of this uniformly minimum variance unbiased estimator (UMVU) relative to the MLE $\overline{X}$ is 64.8%.

Now let $W_{(i)},\ i = 1, \cdots, K$ denote the ordered Walsh averages, where $K = n(n-1)/2$. We want the UMVU estimator among the estimators $\tau_k W_{(k)}$, where τ_k is a constant. Let $[y]$ denote the the smallest integer larger than or equal to y. The pth sample quantile $W_{([pK])}$ is consistent for $q_p = F_2^{-1}(p)$. Using the fact that $(X_1 + X_2)/2$ has a gamma distribution with parameters $\alpha = 2$ and $\beta = \theta/2$, we have $p = F_2(q_p) = 1 - (1 + \frac{2q_p}{\theta})e^{-2q_p/\theta}$ which has solution

$$q_p = h_p \theta \tag{3.1}$$

where h_p satisfies

$$1 - p = (1 + 2h_p)e^{-2h_p}. \tag{3.2}$$

Thus quantile estimators of θ have the form

$$T_n(p) = \frac{1}{h_p} W_{(pK)}$$

where h_p satisfies (3.2). From (2.1), $T_n(p)$ has asymptotic variance $\frac{4\gamma(p)}{nh_p^2[f_2(q_p)]^2}$

Table 1: Comparison of Estimators of Exponential Mean Based on a Simulation of 1000 Sets of n=10 Samples.

	Mean	SD	Median	Min	Max
$.6998X_{(8)}$	10.0	3.822	9.4	2.6	29.0
$.7435W_{(34)}$	10.0	3.336	9.7	2.9	25.7
$\overline{X}$	10.0	3.143	9.6	2.7	23.6

where

$$\gamma(p) = VarF(2q_p - X) = Var[(1 - e^{-(2q_p - X)/\theta})u(0 \leq 2q_p - X < \infty)].$$

Direct integration yields

$$\gamma(p) = 2e^{-2q_p/\theta}\left(1 - e^{-2q_p/\theta} - \frac{2q_p}{\theta}e^{-2q_p/\theta} - \frac{2q_p^2}{\theta^2}e^{-2q_p/\theta}\right). \qquad (3.3)$$

Noting that f_2 is the pdf of Gamma($\alpha = 2,\ \beta = \theta/2$), $T_n(p)$ has asymptotic variance $[8e^{-2h_p}(1 - e^{-2h_p} - 2h_p e^{-2h_p} - 2h_p^2 e^{-2h_p})]/[nh_p^2\,[4h_p e^{-2h_p}/\theta]^2]$. This is minimized at $h_p \approx 1.345$, or equivalently $p \approx .750$. The minimum asymptotic variance is $1.134\theta^2/n$ so that our estimator has asymptotic efficiency 88.2%.

A simulation was conducted to confirm the results. 1000 samples of size $n = 10$ were generated from an exponential distribution with mean θ=10. For $n = 10$, the UMVU among regular quantile estimators is $.6998X_{(8)}$ (Harter, 1961). The UMVU among ordered Walsh averages is $.7435W_{(34)}$. For each sample, $X_{(8)}$, $W_{(34)}$, and the MLE $\overline{X}$ were calculated. Table 1 shows results of the simulation.

The ratios of sample variances in Table 1 closely match the asymptotic efficiencies, i.e. $(3.143)^2/(3.822)^2 = .676$ for $X_{(8)}$ and $(3.143)^2/(3.336)^2 = .888$ for $W_{(34)}$. The true asymptotic efficiencies are 64.8% and 88.2%, respectively. Despite the small sample size, the asymptotic scaling factor .7435 for W works quite well, as evidenced by a sample median (over the 1000 samples) closer to the true value $\theta = 10$ than the MLE $\overline{X}$. W also has median, minimum and maximum values closer to $\theta = 10$ than $X_{(8)}$.

3.2 Example 2. A Simulation Study

Under varying tail thickness of the underlying distribution, we conducted simulations to assess the performance of the median and trimmed mean of Walsh averages for estimating the center of symmetry μ, and the interquartile range (IQR) and median absolute deviation (MAD) for estimating the standard deviation σ. Consider the normal, logistic, and double exponential distributions standardized to have mean μ=0 and standard deviation σ=1. The three distributions are ordered according to increasing tail thickness (Hettmansperger, 1984, p.113). We use the logistic and double exponential to study behavior of the estimates at alternatives to the assumed normal distribution. From each distribution, 1000 samples of size 10 were generated. For each sample, the eight statistics in Table 2 were computed. Med(W) is the median of the $\binom{10}{2} = 45$ Walsh averages. Med(X) is the regular sample median. The estimators $\overline{W}_{50} = \sum_{i=12}^{34} W_{(i)}$ and $\overline{X}_{50} = \sum_{i=3}^{8}$ are the 50% trimmed means of the Walsh averages and regular observations, respectively. The estimator $IQR(W) = \frac{(W_{(35)} - W_{(11)})}{1.4476}\sqrt{2}\sqrt{9/8}$ is the interquartile range of the Walsh averages standardized to be unbiased for σ at the normal distribution. Note that if we treat the Walsh averages as a random sample of size 45 from $N(0,\ \sigma^2/2)$, then $(W_{(35)} - W_{(11)})/1.4476$ is unbiased for $\sigma/\sqrt{2}$. The term $\sqrt{9/8}$ is an adjustment factor for nonindependence of the Walsh averages motivated as follows. Let $S^2(F)$ be the variance functional so $S^2(F_n) = \sum(x_i - \overline{x})^2/n$. It can be shown that $S^2(F_{2,n}) = \sum(w_k - \overline{w})^2/\binom{n}{2} = \frac{\sum(x_i-\overline{x})^2}{2n}\frac{n-2}{n-1}$. Thus $S(F_n) = \sqrt{(n-1)/(n-2)}\sqrt{2}S(F_{2,n})$ so that $\sqrt{(n-1)/(n-2)}$ is a correction factor for nonindependence of the Walsh averages in estimation of the scale parameter σ. We apply this adjustment factor analogously to IQR(W) and MAD(W). The statistic IQR(X) in Table 2 is $IQR(X) = (X_{(8)} - X_{(3)})/1.3122$, again standardized to be unbiased for σ at the normal distribution. The estimator $MAD(X) = med\{|X_i - med\{X_j\}|\}1.483$ is

Table 2: Comparison of Variances Between W- and X-Based Estimators from a Simulation of 1000 Sets of n=10 Samples

	Normal	Logistic	Dbl. Exp.
Var[Med(W)]/Var[Med(X)]	.75	.85	1.14
Var[$\overline{W}_{50}$]/Var[$\overline{X}_{50}$]	.90	.96	1.11
Var[IQR(W)]/Var[IQR(X)]	.52	.69	.89
Var[MAD(W)]/Var[MAD(X)]	.63	.74	.97

the median absolute deviation estimator of scale, where *med* is the median function. Finally, $MAD(W) = med\{|W_k - med\{W_j\}|\}1.483\sqrt{2}\sqrt{9/8}$ is the median absolute deviation estimator based on the Walsh averages. The constant 1.483 is the recommended standardization for MAD in the literature (see e.g. Staudte and Sheather, 1990, p.132), and achieves consistency at the normal distribution.

For each L-estimate, Table 2 shows the simulated relative efficiency between the X- and the W-based estimators. The L-estimates based on Walsh averages have better efficiencies at the normal distribution. The reduction in variance is slight for the trimmed mean, moderate for the median, and quite large for the scale estimators IQR and MAD. Similar improvements hold at a lesser degree for sampling from the logistic distribution. It should be noted that Med(W) is efficient at the logistic distribution (Hettmansperger, 1984, Sec. 2.9). At the more extreme double exponential distribution, Med(X) and $\overline{X}_{50}$ are slightly better than their W counterparts. This is expected since Med(X) is the maximum likelihood estimator for the double exponential, and hence, efficient.

It is common practice to standardize scale estimators according to the normal distribution, as we have done here. How reliable are the estimates when the underlying distribution has thicker tails? Table 3 shows that the scale estimators based on Walsh averages are more robust. At the logistic distribution, the average of IQR(W) over 1000 samples is .95 compared to .91 for the regular IQR (both are estimating $\sigma = 1.0$). Similarly, MAD(W) exhibits less bias than

Table 3: Average of Scale Estimators over 1000 Sets of n=10 Samples When the True Standard Deviation is σ=1.0. (S.E. = .01 in all cases.)

	Normal	Logistic	Dbl. Exp.
IQR(W)	1.00	.95	.89
IQR(X)	1.03	.91	.78
MAD(W)	.99	.93	.86
MAD(X)	.93	.84	.71

MAD(X). At the even thicker tailed double exponential distribution, IQR(W) had an average value of .89 compared to .78 for IQR(X), while MAD(W) had an average value of .86 compared to .71 for MAD(X). All standard errors in the above comparisons are approximately .01. Table 3 also shows that the commonly used asymptotic standardization factor of 1.483 for MAD(X) is quite deficient for moderate sample sizes. At $n = 10$, MAD(X) underestimates σ considerably, even at the correct normal distribution. The bias is smaller at larger sample sizes (see Table 5 for $n = 20$). On the other hand, the asymptotic standardization for MAD(W) works well even at $n = 10$.

A second set of simulations were run using samples of size $n = 20$. The results are given in Tables 4 and 5 which are analogues of Tables 2 and 3, respectively. Table 4 shows that the W-based estimators do even better efficiency-wise against the X-based estimators at the larger sample size. Bias of the scale estimators remain about the same, except for an improvement in MAD(X) at the normal distribution.

4 Discussion

There are many estimates in the literature based on order statistics, or linear combinations of order statistics. Some of these methods (like the median and MAD) achieve less sensitivity to gross errors at the expense of efficiency. Un-

Table 4: Comparison of Variances Between W- and X-Based Estimators from a Simulation of 1000 Sets of n=20 Samples

	Normal	Logistic	Dbl. Exp.
Var[Med(W)]/Var[Med(X)]	.71	.77	1.07
Var[$\overline{W}_{50}$]/Var[$\overline{X}_{50}$]	.88	.93	1.15
Var[IQR(W)]/Var[IQR(X)]	.40	.55	.77
Var[MAD(W)]/Var[MAD(X)]	.49	.63	.83

Table 5: Average of Scale Estimators over 1000 Sets of n=20 Samples When the True Standard Deviation is σ=1.0. (S.E. = .01 in all cases.)

	Normal	Logistic	Dbl. Exp.
IQR(W)	1.00	.93	.86
IQR(X)	1.01	.90	.73
MAD(W)	1.00	.92	.85
MAD(X)	.97	.86	.72

fortunately too much efficiency is often lost, reducing the estimator's role in applications. We have shown that substantial efficiency may be regained by constructing estimates from ordered pairwise averages instead of the regular order statistics. The cost in complexity is minimal, since the procedure basically entails regular L-estimation of the parameters of F_2 instead of F. And since the same L-estimator is computed, the appeal to users of the original X-based estimator may not be lost.

References

Harter,H.L. (1961). "Estimating the Parameters of Negative Exponential Population from One or Two Order Statistics. *Ann. Math. Stat.*, 32, 1078-1090.

Hettmansperger,T.P. (1984) *Statistical Inference Based on Ranks.* John Wiley & Sons, New York.

Serfling,R.J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

Siddiqui,M.M. (1963). "Optimum Estimators of the Parameters of Negative Exponential Distributions from One or Two Order Statistics," *Ann. Math. Stat.*, 34, 117-121.

Staudte,R.G. and Sheather,S.J. (1990). *Robust Estimation and Testing.* John Wiley & Sons, New York.