**JIRSS (2007)**
**Vol. 6, No. 2, pp 173-177**

# A Sharp Inequality for Medians of L-Statistics in a Nonparametric Statistical Model

**Ryszard Zieliński**

Institute of Mathematics, Polish Academy of Science, Warszawa, Poland.
(R.Zielinski@impan.gov.pl)

**Abstract.** Sharp bounds for medians of L-statistics in the nonparametric statistical model with all continuous and strictly increasing distribution functions are given. As a corollary we conclude that L-statistics are very poor nonparametric quantile estimators.

## 1 Result

Let $X_1, \ldots, X_n$ be a sample from a distribution $F \in \mathcal{F}$, where $\mathcal{F}$ is the class of all continuous and strictly increasing distribution functions on their supports. Let $X_{1:n}, \ldots, X_{n:n}$ be the order statistics, let $T = \sum_{j=1}^{n} \lambda_j X_{j:n}$; $\lambda_j \geq 0$, $j = 1, 2, \ldots, n$; $\sum_{j=1}^{n} \lambda_j = 1$, be a nontrivial $L$-statistic (at least two $\lambda$'s are positive). Let $S = S(X_1, \ldots, X_n)$ be any function of observations $X_1, \ldots, X_n$ and let $Med(F, S)$ denote a median (of the distribution) of $S$ if the sample comes from the distribution $F$. Our primary interest are functions of the form $S(.) = F(T(.))$.

*Key words and phrases:* Harrell-Davis estimator, Kaigh-Cheng estimator, $L$-statistics, quantiles, quantile estimators.

**Theorem 1.1.** *If* $T = \sum_{j=k}^{m} \lambda_j X_{j:n}$ *is an L-statistic such that* $\lambda_k > 0$, $\lambda_m > 0$, $k < m$, *and* $\lambda_k + \lambda_{k+1} + \ldots + \lambda_m = 1$, *then*

$$(*) \qquad m(U_{k:n}) \le Med(F, F(T)) \le m(U_{m:n}),$$

*where* $m(U_{k:n})$ *and* $m(U_{m:n})$ *are medians of order statistics* $U_{k:n}$ *and* $U_{m:n}$ *from a sample of size n from the uniform* $U(0,1)$ *parent distribution. The bounds are sharp in the sense that for every* $\varepsilon > 0$ *there exists* $F \in \mathcal{F}$ *such that* $Med(F, F(T)) > m(U_{m:n}) - \varepsilon$ *and for every* $\eta > 0$ *there exists* $G \in \mathcal{F}$ *such that* $Med(G, G(T)) < m(U_{k:n}) + \eta$.

**Proof.** The first statement follows easily from the fact that $X_{k:n} < T < X_{m:n}$ and hence for every $F \in \mathcal{F}$ we have $U_{k:n} = F(X_{k:n}) < F(T) < F(X_{m:n}) = U_{m:n}$. To prove the second part of the theorem it is enough to construct families of distributions $F_\alpha, \alpha > 0$, and $G_\alpha, \alpha > 0$, such that $Med(F_\alpha, F_\alpha(T)) \to m(U_{m:n})$ and $Med(G_\alpha, G_\alpha(T)) \to m(U_{k:n})$, as $\alpha \to 0$.

Consider the family of power distributions $F_\alpha(x) = x^\alpha$, $0 < x < 1$, $\alpha > 0$. Then $X_{j:n} = F_\alpha^{-1}(U_{j:n}) = U_{j:n}^{1/\alpha}$ and

$$\begin{aligned} F_\alpha(T) &= \Big(\lambda_k U_{k:n}^{1/\alpha} + \lambda_{k+1} U_{k+1:n}^{1/\alpha} + \ldots + \lambda_{m-1} U_{m-1:n}^{1/\alpha} + \lambda_m U_{m:n}^{1/\alpha}\Big)^\alpha \\ &= U_{m:n}\Big[\lambda_k \Big(\frac{U_{k:n}}{U_{m:n}}\Big)^{1/\alpha} + \lambda_{k+1}\Big(\frac{U_{k+1:n}}{U_{m:n}}\Big)^{1/\alpha} + \ldots \\ &\quad + \lambda_{m-1}\Big(\frac{U_{m-1:n}}{U_{m:n}}\Big)^{1/\alpha} + \lambda_m\Big]^\alpha \end{aligned}$$

If $\alpha \to 0$ then $F_\alpha(T) \to U_{m:n}$ and $Med(F_\alpha, F_\alpha(T)) \to m(U_{m:n})$.

Now consider the family $G_\alpha$ with $G_\alpha(x) = 1 - (1-x)^\alpha$; in full analogy to the above we conclude that then $G_\alpha(T) \to U_{k:n}$ and $Med(G_\alpha, G_\alpha(T)) \to m(U_{k:n})$ as $\alpha \to 0$. □

**Corollary 1.1.** *If an L-statistic* $T = \sum_{j=k}^{m} \lambda_j X_{j:n}$, $\lambda_k > 0$, $\lambda_m > 0$, $\lambda_k + \lambda_{k+1} + \ldots + \lambda_m = 1$, $k < m$, *and* $\lambda_j = \lambda_j(q)$, $j = k, \ldots, m$, *is considered as a nonparametric estimator of the q-th quantile* $x_q(F) = F^{-1}(q)$ *of an unknown distribution* $F \in \mathcal{F}$, *then the error of estimation may be arbitrarily large in the sense that for every* $C > 0$ *there exists a distribution* $F \in \mathcal{F}$ *such that* $|Med(F, T) - x_q(F)| > C$.

**Proof.** Suppose that $q < m(U_{m:n})$. The case that $q > m(U_{k:n})$ can be considered in full analogy.

Choose $\varepsilon > 0$ such that $m(U_{m:n}) - \varepsilon > q$. By the Theorem there exists a distribution $F \in \mathcal{F}$ such that $Med(F, F(T)) > m(U_{m:n}) - \varepsilon > q$. By the obvious equality that states that $Med(F, F(T)) = F(Med(F, T))$ we obtain that $Med(F, T) - x_q(F) > 0$. For an $\sigma > 0$ consider the distribution $F_\sigma \in \mathcal{F}$ defined by the formula $F_\sigma(x) = F(x/\sigma)$. Then $x_q(F_\sigma) = \sigma \cdot x_q(F)$ and, due to the fact that $T$ is scale equivariant, $Med(F_\sigma, T) = \sigma \cdot Med(F, T)$. Hence $Med(F_\sigma, T) - x_q(F_\sigma) = \sigma \cdot (Med(F, T) - x_q(F))$ which by a suitable choice of $\sigma > 1$ may be arbitrarily large. □

## 2 Numerical illustrations (simulations)

To demonstrate that $L$-statistics may produce very large errors in estimating quantiles in the nonparametric model $\mathcal{F}$ with all continuous and strictly increasing distribution functions we decided to present the problem of estimating the median of an unknown $F \in \mathcal{F}$ with the following well known estimators:

*Davis and Steinberg (1986)*

$$X_{(n+1)/2:n}, \quad \text{if } n \text{ is odd}; \qquad \left(X_{n/2:n} + X_{n/2+1:n}\right)/2, \quad \text{if } n \text{ is even},$$

*Harrell and Davis (1982)*

$$HD = \frac{n!}{[(\frac{n-1}{2})!]^2} \sum_{j=1}^{n} \left[ \int_{(j-1)/n}^{j/n} [u(1-u)]^{(n-1)/2} du \right] X_{j:n},$$

*Kaigh and Cheng (1991) for n odd*

$$KC = \frac{1}{\binom{2n-1}{n}} \sum_{j=1}^{n} \binom{\frac{n-3}{2} + j}{\frac{n-1}{2}} \binom{\frac{3n-1}{2} - j}{\frac{n-1}{2}} X_{j:n}.$$

As the distributions for studying our problem we have chosen

*Pareto with cdf*

$$1 - \frac{1}{x^\alpha}, \quad x > 1, \quad \text{heavy tails, no moments of order } k \geq \alpha,$$

*Power (special case of Beta) with cdf*

$$x^\alpha, \quad x \in (0, 1), \quad \text{no tails, all moments },$$

*Exponential with cdf*

$$1 - exp\{-\alpha x\}, \quad x > 0, \quad \text{very regular} ,$$

all distributions for $\alpha = 1/2, 1/4$, and $1/8$.

Results of our numerical investigations for samples of size $n = 9$ (Harrell-Davis and Kaigh-Cheng) or for samples of size $n = 10$ (Davis-Steinberg statistic $(X_{5:10} + X_{6:10})/2$) are presented in the Table below. The number of simulated samples, and consequently the number of simulated values of the estimator under consideration, was $N = 9,999$, and the median from the sample of size $N = 9,999$ has been taken as an estimator of the median of the distribution of the estimator under consideration. Observe that $m(U_{n:n}) - m(U_{1:n})$ increases with $n$ so that errors of estimators with $k = 1$ and $m = n$ (e.g. HD and KC) increase with $n$.

Simulated medians of estimators

| Distribution | Median | HD | KC | $\frac{X_{5:10}+X_{6:10}}{2}$ |
|---|---|---|---|---|
| Pareto | | | | |
| $\alpha = 1/2$ | 4 | 7.72 | 13.71 | 4.13 |
| $\alpha = 1/4$ | 16 | 255 | 1107 | 18.45 |
| $\alpha = 1/8$ | 256 | $3.3 \times 10^6$ | $2.8 \times 10^7$ | 383 |
| Power | | | | |
| $\alpha = 1/2$ | 0.25 | 0.2780 | 0.2919 | 0.2535 |
| $\alpha = 1/4$ | 0.0625 | 0.1055 | 0.1286 | 0.0692 |
| $\alpha = 1/8$ | 0.0039 | 0.0241 | 0.0432 | 0.0053 |
| Exponential | | | | |
| $\alpha = 1/2$ | 1.3863 | 1.5138 | 1.6235 | 1.4079 |
| $\alpha = 1/4$ | 2.7726 | 3.0571 | 3.2731 | 2.8036 |
| $\alpha = 1/8$ | 5.5452 | 6.0595 | 6.4897 | 5.6143 |

## 3 A remark

A reason for the bad behavior of nontrivial $L$-statistics as quantile estimators is that they are not equivariant under monotonic transformation of data while the class $\mathcal{F}$ of all continuous and strictly increasing distribution functions allows such transformations. In some parametric families of distributions L-statistics may perform excellently. The problem is discussed thoroughly in a Technical Report (Zieliński 2005).

## Acknowledgment

The author wish to express his gratitude to referees for their helpful comments which improved the presentation of the paper, especially the proof of Corollary 1.1.

## References

Davis, C. E. and Steinberg, S. M. (1986), Quantile estimation. In Encyclopedia of Statistical Sciences., **7**, New York: Wiley.

Harrell, F. E. and Davis, C. E. (1982), A new distribution-free quantile estimator, Biometrika, **69**, 635-640.

Kaigh, W. D. and Cheng, C. (1991), Subsampling quantile estimators and uniformity criteria. Commun. Statist. Theor. Meth., **20**, 539-560.

Zieliński, R. (2005), L-statistics as nonparametric quantile estimators. IMPAN, Preprint 657, June 2005. Available at www.impan.gov.pl/˜rziel.