Semiparametric robust mean estimations based on the orderliness of quantile averages

Tuban Lee

This manuscript was compiled on June 13, 2023

semiparametric | mean-median-mode inequality | asymptotic | unimodal | Hodges–Lehmann estimator

Remarkably, in 2018, Li, Shao, Wang, Yang (1) proved the bias bound of any quantile for arbitrary continuous distributions with finite second moments. Here, let $\mathcal{P}_{\mu,\sigma}$ denotes the set of continuous distributions whose mean is μ and standard deviation is σ. The bias upper bound of the quantile average for $P \in \mathcal{P}_{\mu=0,\sigma=1}$ is given in the following theorem.

Theorem .1. *The bias upper bound of the quantile average for any continuous distribution whose mean is zero and standard deviation is one is*

$$\sup_{P\in\mathcal{P}_{\mu=0,\sigma=1}} QA(\epsilon,\gamma) = \frac{1}{2}\left(\sqrt{\frac{\gamma\epsilon}{1-\gamma\epsilon}} + \sqrt{\frac{1-\epsilon}{\epsilon}}\right), \qquad [1]$$

where $0 \le \epsilon \le \frac{1}{1+\gamma}$.

Proof. Since $\sup_{P\in\mathcal{P}_{\mu=0,\sigma=1}} \frac{1}{2}(Q(\gamma\epsilon) + Q(1-\epsilon)) \le \frac{1}{2}(\sup_{P\in\mathcal{P}_{\mu=0,\sigma=1}} Q(\gamma\epsilon) + \sup_{P\in\mathcal{P}_{\mu=0,\sigma=1}} Q(1-\epsilon))$, the assertion follows directly from the Lemma 2.6 in (1). □

In 2020, Bernard et al. (2) further refined these bounds for unimodal distributions and derived the bias bound of the symmetric quantile average. Here, the bias upper bound of the quantile average, $0 \le \gamma < 5$, for $P \in \mathcal{P}_U \cap \mathcal{P}_{\mu=0,\sigma=1}$ is given as

$$\sup_{P\in\mathcal{P}_U\cap\mathcal{P}_{\mu=0,\sigma=1}} \mathrm{QA}(\epsilon,\gamma) = \begin{cases} \frac{1}{2}\left(\sqrt{\frac{4}{9\epsilon}-1} + \sqrt{\frac{3\gamma\epsilon}{4-3\gamma\epsilon}}\right) & 0 \le \epsilon \le \frac{1}{6} \\ \frac{1}{2}\left(\sqrt{\frac{3(1-\epsilon)}{4-3(1-\epsilon)}} + \sqrt{\frac{3\gamma\epsilon}{4-3\gamma\epsilon}}\right) & \frac{1}{6} < \epsilon \le \frac{1}{1+\gamma} \end{cases}.$$

The proof based on the bias bounds of any quantile (2) and the $\gamma \ge 5$ case are given in the SI Text. Subsequent theorems reveal the safeguarding role these bounds play in defining estimators based on νth γ-orderliness. The proof of Theorem .2 is provided in the SI Text.

Theorem .2. $\sup_{P\in\mathcal{P}_{\mu=0,\sigma=1}} QA(\epsilon,\gamma)$ *is monotonic decreasing with respect to* ϵ *over the interval* $[0, \frac{1}{1+\gamma}]$, *when* $0 \le \gamma \le 1$.

Theorem .3. $\sup_{P\in\mathcal{P}_U\cap\mathcal{P}_{\mu=0,\sigma=1}} QA(\epsilon,\gamma)$ *is monotonic decreasing with respect to* ϵ *over the interval* $[0, \frac{1}{1+\gamma}]$, *when* $0 \le \gamma \le 1$.

Proof. When $0 \le \epsilon \le \frac{1}{6}$, $\frac{\partial \sup \mathrm{QA}}{\partial \epsilon} = \frac{\gamma}{\sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}}(4-3\epsilon\gamma)^2} - \frac{1}{3\sqrt{\frac{4}{\epsilon}-9\epsilon^2}}$. When $\gamma = 0$, $\frac{\gamma}{\sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}}(4-3\epsilon\gamma)^2} = \frac{\sqrt{\gamma}}{\sqrt{\frac{\epsilon}{12-9\epsilon\gamma}}(4-3\epsilon\gamma)^2} = 0$, $\frac{\partial \sup \mathrm{QA}}{\partial \epsilon} = -\frac{1}{3\sqrt{\frac{4}{\epsilon}-9\epsilon^2}} < 0$. When $\epsilon \to 0^+$, $\lim_{\epsilon\to 0^+}\left(\frac{\gamma}{(4-3\gamma\epsilon)^2\sqrt{\frac{\epsilon\gamma}{12-9\gamma\epsilon}}} - \frac{1}{3\sqrt{\frac{4}{\epsilon}-9\epsilon^2}}\right) = \lim_{\epsilon\to 0^+}\left(\frac{\gamma\sqrt{3}}{\sqrt{4^3\epsilon\gamma}} - \frac{1}{6\sqrt{\epsilon^3}}\right) \to -\infty$, so, $\frac{\partial \sup \mathrm{QA}}{\partial \epsilon} < 0$.

□

Inequalities related to weighted averages

So far, it is quite natural to hypothesize that the value of ϵ,γ-trimmed mean should be monotonically related to the breakdown point in a semiparametric distribution, since it is a linear combination of quantile averages as shown in Section **??**. Analogous to the γ-orderliness, the γ-trimming inequality for a right-skewed distribution is defined as $\forall 0 \le \epsilon_1 \le \epsilon_2 \le \frac{1}{1+\gamma}$, $\mathrm{TM}_{\epsilon_1,\gamma} \ge \mathrm{TM}_{\epsilon_2,\gamma}$. γ-orderliness is a sufficient condition for the γ-trimming inequality, as proven in the SI Text. The next theorem shows a relation between the ϵ,γ-quantile average and the ϵ,γ-trimmed mean under the γ-trimming inequality, suggesting the γ-orderliness is not a necessary condition for the γ-trimming inequality.

Theorem .4. *For a distribution that is right-skewed and follows the* γ*-trimming inequality, it is asymptotically true that the quantile average is always greater or equal to the corresponding trimmed mean with the same* ϵ *and* γ, *for all* $0 \le \epsilon \le \frac{1}{1+\gamma}$.

Proof. According to the definition of the γ-trimming inequality: $\forall 0 \le \epsilon \le \frac{1}{1+\gamma}$, $\frac{1}{1-\epsilon-\gamma\epsilon+2\delta}\int_{\gamma\epsilon-\delta}^{1-\epsilon+\delta} Q(u)\,du \ge \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$, where δ is an infinitesimal positive quantity. Subsequently, rewriting the inequality gives $\int_{\gamma\epsilon-\delta}^{1-\epsilon+\delta} Q(u)\,du - \frac{1-\epsilon-\gamma\epsilon+2\delta}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du \ge 0 \Leftrightarrow \int_{1-\epsilon}^{1-\epsilon+\delta} Q(u)\,du + \int_{\gamma\epsilon-\delta}^{\gamma\epsilon} Q(u)\,du - \frac{2\delta}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du \ge 0$. Since $\delta \to 0^+$, $\frac{1}{2\delta}\left(\int_{1-\epsilon}^{1-\epsilon+\delta} Q(u)\,du + \int_{\gamma\epsilon-\delta}^{\gamma\epsilon} Q(u)\,du\right) = \frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2} \ge \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$, the proof is complete. □

An analogous result about the relation between the ϵ,γ-trimmed mean and the ϵ,γ-Winsorized mean can be obtained in the following theorem.

Theorem .5. *For a right-skewed distribution following the* γ*-trimming inequality, asymptotically, the Winsorized mean is always greater or equal to the corresponding trimmed mean with the same* ϵ *and* γ, *for all* $0 \le \epsilon \le \frac{1}{1+\gamma}$, *provided that* $0 \le \gamma \le 1$. *If assuming* γ*-orderliness, the inequality is valid for any non-negative* γ.

T.L. designed research, performed research, analyzed data, and wrote the paper.

The author declares no competing interest.

[1] To whom correspondence should be addressed. E-mail: tl@biomathematics.org

Proof. According to Theorem .4, $\frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2} \geq \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du \Leftrightarrow \gamma\epsilon\left(Q(\gamma\epsilon)+Q(1-\epsilon)\right) \geq \left(\frac{2\gamma\epsilon}{1-\epsilon-\gamma\epsilon}\right)\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$. Then, if $0 \leq \gamma \leq 1$, $\left(1-\frac{1}{1-\epsilon-\gamma\epsilon}\right)\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du + \gamma\epsilon\left(Q(\gamma\epsilon)+Q(1-\epsilon)\right) \geq 0 \Rightarrow \int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du + \gamma\epsilon Q(\gamma\epsilon) + \epsilon Q(1-\epsilon) \geq \int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du + \gamma\epsilon\left(Q(\gamma\epsilon)+Q(1-\epsilon)\right) \geq \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$, the proof of the first assertion is complete. The second assertion is established in Theorem 0.3. in the SI Text. □

Replacing the TM in the γ-trimming inequality with WA forms the definition of the γ-weighted inequality. The γ-orderliness also implies the γ-Winsorization inequality when $0 \leq \gamma \leq 1$ for a right-skewed distribution, as proven in the SI Text. To construct weighted averages based on the νth γ-orderliness and satisfying the corresponding weighted inequality, when $0 \leq \gamma \leq 1$, let $\mathcal{B}_i = \int_{i\epsilon}^{(i+1)\epsilon} \mathrm{QA}(u,\gamma)\,du$, $ka = k\epsilon + c$. From the γ-orderliness for a right-skewed distribution, it follows that, $-\frac{\partial \mathrm{QA}}{\partial \epsilon} \geq 0 \Leftrightarrow \forall 0 \leq a \leq 2a \leq \frac{1}{1+\gamma}, -\frac{(\mathrm{QA}(2a,\gamma)-\mathrm{QA}(a,\gamma))}{a} \geq 0 \Rightarrow \mathcal{B}_i - \mathcal{B}_{i+1} \geq 0$, if $0 \leq \gamma \leq 1$. Suppose that $\mathcal{B}_i = \mathcal{B}_0$. Then, the ϵ,γ-block Winsorized mean, is defined as

$$\mathrm{BWM}_{\epsilon,\gamma,n} := \frac{1}{n}\left(\sum_{i=n\gamma\epsilon+1}^{(1-\epsilon)n} X_i + \sum_{i=n\gamma\epsilon+1}^{2n\gamma\epsilon+1} X_i + \sum_{i=(1-2\epsilon)n}^{(1-\epsilon)n} X_i\right),$$

which is double weighting the leftest and rightest blocks having sizes of $\gamma\epsilon n$ and ϵn, respectively. As a consequence of $\mathcal{B}_i - \mathcal{B}_{i+1} \geq 0$, the γ-block Winsorization inequality is valid, provided that $0 \leq \gamma \leq 1$. The block Winsorized mean uses two blocks to replace the trimmed parts, not two single quantiles. The subsequent theorem provides an explanation for this difference.

Theorem .6. *Asymptotically, for a right-skewed γ-ordered distribution, the Winsorized mean is always greater than or equal to the corresponding block Winsorized mean with the same ϵ and γ, for all $0 \leq \epsilon \leq \frac{1}{1+\gamma}$, if $0 \leq \gamma \leq 1$.*

Proof. From the definitions of BWM and WM, the statement necessitates $\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du + \gamma\epsilon Q(\gamma\epsilon) + \epsilon Q(1-\epsilon) \geq \int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du + \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(u)\,du + \int_{1-2\epsilon}^{1-\epsilon} Q(u)\,du \Leftrightarrow \gamma\epsilon Q(\gamma\epsilon) + \epsilon Q(1-\epsilon) \geq \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(u)\,du + \int_{1-2\epsilon}^{1-\epsilon} Q(u)\,du$. Define $\mathrm{WM}l(x) = Q(\gamma\epsilon)$ and $\mathrm{BWM}l(x) = Q(x)$. In both functions, the interval for x is specified as $[\gamma\epsilon, 2\gamma\epsilon]$. Then, define $\mathrm{WM}u(y) = Q(1-\epsilon)$ and $\mathrm{BWM}u(y) = Q(y)$. In both functions, the interval for y is specified as $[1-2\epsilon, 1-\epsilon]$. The function $y : [\gamma\epsilon, 2\gamma\epsilon] \to [1-2\epsilon, 1-\epsilon]$ defined by $y(x) = 1 - \frac{x}{\gamma}$ is a bijection. $\mathrm{WM}l(x) + \mathrm{WM}u(y(x)) = Q(\gamma\epsilon) + Q(1-\epsilon) \geq \mathrm{BWM}l(x) + \mathrm{BWM}u(y(x)) = Q(x) + Q\left(1-\frac{x}{\gamma}\right)$ is valid for all $x \in [\gamma\epsilon, 2\gamma\epsilon]$, according to the definition of γ-orderliness. Integration of the left side yields, $\int_{\gamma\epsilon}^{2\gamma\epsilon} \left(\mathrm{WM}l(u) + \mathrm{WM}u(y(u))\right) du = \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(\gamma\epsilon)\,du + \int_{y(\gamma\epsilon)}^{y(2\gamma\epsilon)} Q(1-\epsilon)\,du = \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(\gamma\epsilon)\,du + \int_{1-2\epsilon}^{1-\epsilon} Q(1-\epsilon)\,du = \gamma\epsilon Q(\gamma\epsilon) + \epsilon Q(1-\epsilon)$, while integration of the right side yields $\int_{\gamma\epsilon}^{2\gamma\epsilon} \left(\mathrm{BWM}l(x) + \mathrm{BWM}u(y(x))\right) dx = \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(u)\,du + \int_{\gamma\epsilon}^{2\gamma\epsilon} Q\left(1-\frac{x}{\gamma}\right) dx = \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(u)\,du + \int_{1-2\epsilon}^{1-\epsilon} Q(u)\,du$, which are the left and right sides of the desired inequality. Given that the upper limits and lower limits of the integrations are different for each term, the condition $0 \leq \gamma \leq 1$ is necessary for the desired inequality to be valid. □

From the second γ-orderliness for a right-skewed distribution, $\frac{\partial^2 \mathrm{QA}}{\partial^2 \epsilon} \geq 0 \Rightarrow \forall 0 \leq a \leq 2a \leq 3a \leq \frac{1}{1+\gamma}, \frac{1}{a}\left(\frac{(\mathrm{QA}(3a,\gamma)-\mathrm{QA}(2a,\gamma))}{a} - \frac{(\mathrm{QA}(2a,\gamma)-\mathrm{QA}(a,\gamma))}{a}\right) \geq 0 \Rightarrow$ if $0 \leq \gamma \leq 1$, $\mathcal{B}_i - 2\mathcal{B}_{i+1} + \mathcal{B}_{i+2} \geq 0$. SM_ϵ can thus be interpreted as assuming $\gamma = 1$ and replacing the two blocks, $\mathcal{B}_i + \mathcal{B}_{i+2}$ with one block $2\mathcal{B}_{i+1}$. From the νth γ-orderliness for a right-skewed distribution, the recurrence relation of the derivatives naturally produces the alternating binomial coefficients,

$$(-1)^\nu \frac{\partial^\nu \mathrm{QA}}{\partial \epsilon^\nu} \geq 0 \Rightarrow \forall 0 \leq a \leq \ldots \leq (\nu+1)a \leq \frac{1}{1+\gamma},$$

$$\frac{(-1)^\nu}{a}\left(\frac{\frac{\mathrm{QA}(\nu a+a,\gamma)^{\cdot^{\cdot^{\cdot}}}}{a} - \frac{^{\cdot^{\cdot^{\cdot}}}\mathrm{QA}(2a,\gamma)}{a}}{a} - \frac{\frac{\mathrm{QA}(\nu a,\gamma)^{\cdot^{\cdot^{\cdot}}}}{a} - \frac{^{\cdot^{\cdot^{\cdot}}}\mathrm{QA}(a,\gamma)}{a}}{a}\right)$$

$$\geq 0 \Leftrightarrow \frac{(-1)^\nu}{a^\nu}\left(\sum_{j=0}^{\nu} (-1)^j \binom{\nu}{j} \mathrm{QA}\left((\nu-j+1)a,\gamma\right)\right) \geq 0$$

$$\Rightarrow \text{if } 0 \leq \gamma \leq 1, \sum_{j=0}^{\nu} (-1)^j \binom{\nu}{j} \mathcal{B}_{i+j} \geq 0.$$

Based on the νth orderliness, the ϵ,γ-binomial mean is introduced as

$$\mathrm{BM}_{\nu,\epsilon,\gamma,n} := \frac{1}{n}\left(\sum_{i=1}^{\frac{1}{2}\epsilon^{-1}(\nu+1)^{-1}} \sum_{j=0}^{\nu} \left(1-(-1)^j \binom{\nu}{j}\right) \mathfrak{B}_{i_j}\right),$$

where $\mathfrak{B}_{i_j} = \sum_{l=n\gamma\epsilon(j+(i-1)(\nu+1))+1}^{n\epsilon(j+(i-1)(\nu+1)+1)} (X_l + X_{n-l+1})$. If ν is not indicated, it defaults to $\nu = 3$. Since the alternating sum of binomial coefficients equals zero, when $\nu \ll \epsilon^{-1}$ and $\epsilon \to 0$, $\mathrm{BM} \to \mu$. The solutions for the continuity of the breakdown point is the same as that in SM and not repeated here. The equalities $\mathrm{BM}_{\nu=1,\epsilon} = \mathrm{BWM}_\epsilon$ and $\mathrm{BM}_{\nu=2,\epsilon} = \mathrm{SM}_{\epsilon,b=3}$ hold, when $\gamma = 1$ and their respective ϵs are identical. Interestingly, the biases of the $\mathrm{SM}_{\epsilon=\frac{1}{9},b=3}$ and the $\mathrm{WM}_{\epsilon=\frac{1}{9}}$ are nearly indistinguishable in common asymmetric unimodal distributions such as Weibull, gamma, lognormal, and Pareto (SI Dataset S1). This indicates that their robustness to departures from the symmetry assumption is practically similar under unimodality, even though they are based on different orders of orderliness. If single quantiles are used, based on the second γ-orderliness, the stratified quantile mean can be defined as

$$\mathrm{SQM}_{\epsilon,\gamma,n} := 4\epsilon \sum_{i=1}^{\frac{1}{4\epsilon}} \frac{1}{2}\left(\hat{Q}_n\left((2i-1)\gamma\epsilon\right) + \hat{Q}_n\left(1-(2i-1)\epsilon\right)\right),$$

$\mathrm{SQM}_{\epsilon=\frac{1}{4}}$ is the Tukey's midhinge (3). In fact, SQM is a subcase of SM when $\gamma = 1$ and $b \to \infty$, so the solution for the continuity of the breakdown point, $\frac{1}{\epsilon} \bmod 4 \neq 0$, is identical. However, since the definition is based on the empirical quantile function, no decimal issues related to order statistics will arise. The next theorem explains another advantage.

Theorem .7. *For a right-skewed second γ-ordered distribution, asymptotically, $SQM_{\epsilon,\gamma}$ is always greater or equal to the corresponding $BM_{\nu=2,\epsilon,\gamma}$ with the same ϵ and γ, for all $0 \leq \epsilon \leq \frac{1}{1+\gamma}$, if $0 \leq \gamma \leq 1$.*

Proof. For simplicity, suppose the order statistics of the sample are distributed into $\epsilon^{-1} \in \mathbb{N}$ blocks in the computation of both $\mathrm{SQM}_{\epsilon,\gamma}$ and $\mathrm{BM}_{\nu=2,\epsilon,\gamma}$. The computation of $\mathrm{BM}_{\nu=2,\epsilon,\gamma}$ alternates between weighting and non-weighting, let '0' denote the block assigned with a weight of zero and '1' denote the block assigned with a weighted of one, the sequence indicating the weighted or non-weighted status of each block is: $0, 1, 0, 0, 1, 0, \ldots$. Let this sequence be denoted by $a_{\mathrm{BM}_{\nu=2,\epsilon,\gamma}}(j)$, its formula is $a_{\mathrm{BM}_{\nu=2,\epsilon,\gamma}}(j) = \left\lfloor \frac{j \bmod 3}{2} \right\rfloor$. Similarly, the computation of $\mathrm{SQM}_{\epsilon,\gamma}$ can be seen as positioning quantiles (p) at the beginning of the blocks if $0 < p < \frac{1}{1+\gamma}$, and at the end of the blocks if $p > \frac{1}{1+\gamma}$. The sequence of denoting whether each block's quantile is weighted or not weighted is: $0, 1, 0, 1, 0, 1, \ldots$. Let the sequence be denoted by $a_{\mathrm{SQM}_{\epsilon,\gamma}}(j)$, the formula of the sequence is $a_{\mathrm{SQM}_{\epsilon,\gamma}}(j) = j \bmod 2$. If pairing all blocks in $\mathrm{BM}_{\nu=2,\epsilon,\gamma}$ and all quantiles in $\mathrm{SQM}_{\epsilon,\gamma}$, there are two possible pairings of $a_{\mathrm{BM}_{\nu=2}}(j)$ and $a_{\mathrm{SQM}_{\epsilon,\gamma}}(j)$. One pairing occurs when $a_{\mathrm{BM}_{\nu=2,\epsilon,\gamma}}(j) = a_{\mathrm{SQM}_{\epsilon,\gamma}}(j) = 1$, while the other involves the sequence $0, 1, 0$ from $a_{\mathrm{BM}_{\nu=2,\epsilon,\gamma}}(j)$ paired with $1, 0, 1$ from $a_{\mathrm{SQM}_{\epsilon,\gamma}}(j)$. By leveraging the same principle as Theorem .6 and the second γ-orderliness (replacing the two quantile averages with one quantile average between them), the desired result follows. $\square$

The biases of $\mathrm{SQM}_{\epsilon=\frac{1}{8}}$, which is based on the second orderliness with a quantile approach, are notably similar to those of $\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{8}}$, which is based on the third orderliness with a block approach, in common asymmetric unimodal distributions (Figure **??**).

Data Availability. Data for Figure **??** are given in SI Dataset S1. All codes have been deposited in GitHub.

ACKNOWLEDGMENTS. I sincerely acknowledge the insightful comments from the editor which considerably elevated the lucidity and merit of this paper.

1. L Li, H Shao, R Wang, J Yang, Worst-case range value-at-risk with partial information. *SIAM J. on Financial Math.* **9**, 190–218 (2018).
2. C Bernard, R Kazzi, S Vanduffel, Range value-at-risk bounds for unimodal distributions under partial information. *Insur. Math. Econ.* **94**, 9–24 (2020).
3. JW Tukey, *Exploratory data analysis.* (Reading, MA) Vol. 2, (1977).