Semiparametric robust mean estimations based on the orderliness of quantile averages

Tuban Lee

This manuscript was compiled on June 11, 2023

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

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 $\forall 0 \leq \epsilon_1 \leq \epsilon_2 \leq \frac{1}{1+\gamma}, \mathrm{TM}_{\epsilon_1,\gamma} \geq \mathrm{TM}_{\epsilon_2,\gamma}$. 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 .1. *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, $0 \leq \epsilon \leq \frac{1}{1+\gamma}$.*

Proof. According to the definition of the γ-trimming inequality: $\forall 0 \leq \epsilon \leq \frac{1}{1+\gamma}$, $\frac{1}{1-\epsilon-\gamma\epsilon+2\delta}\int_{\gamma\epsilon-\delta}^{1-\epsilon+\delta} Q(u)\,du \geq \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$, where δ is an infinitesimal positive quantity. Subsequently, rewriting the inequality gives $\int_{\gamma\epsilon-\delta}^{1-\epsilon+\delta} Q(u)\,du - \frac{1-\epsilon-\gamma\epsilon+2\delta}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du \geq 0 \Leftrightarrow \int_{1-\epsilon}^{1-\epsilon+\delta} Q(u)\,du + \int_{\gamma\epsilon-\delta}^{\gamma\epsilon} Q(u)\,du - \frac{2\delta}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du \geq 0$. Since $\delta \to 0^+$, $\frac{1}{2\delta}\left(\int_{1-\epsilon}^{1-\epsilon+\delta} Q(u)\,du + \int_{\gamma\epsilon-\delta}^{\gamma\epsilon} Q(u)\,du\right) = \frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2} \geq \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$, 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 .2. *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, provided that $0 \leq \gamma \leq 1$. If assuming γ-orderliness, the inequality is valid for any non-negative γ.*

Proof. According to Theorem .1, $\frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2} \geq \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du \Leftrightarrow \gamma\epsilon\left(Q(\gamma\epsilon)+Q(1-\epsilon)\right) \geq \left(\frac{2\gamma\epsilon}{1-\epsilon-\gamma\epsilon}\right)\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$. 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 $0 \leq \gamma \leq 1$, 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, 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 .3. *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, 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

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

yields $\int_{\gamma\epsilon}^{2\gamma\epsilon}\left(\mathrm{BWM}l\left(x\right)+\mathrm{BWM}u\left(y\left(x\right)\right)\right)dx=\int_{\gamma\epsilon}^{2\gamma\epsilon}Q\left(u\right)du+\int_{\gamma\epsilon}^{2\gamma\epsilon}Q\left(1-\frac{x}{\gamma}\right)dx=\int_{\gamma\epsilon}^{2\gamma\epsilon}Q\left(u\right)du+\int_{1-2\epsilon}^{1-\epsilon}Q\left(u\right)du$, 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 $0\leq\gamma\leq1$ is necessary for the desired inequality to be valid.

$\square$

From the second γ-orderliness, $\frac{\partial^2\mathrm{QA}}{\partial^2\epsilon}\geq0\Rightarrow\forall0\leq a\leq2a\leq3a\leq\frac{1}{1+\gamma},\frac{1}{a}\left(\frac{(\mathrm{QA}(3a,\gamma)-\mathrm{QA}(2a,\gamma))}{a}-\frac{(\mathrm{QA}(2a,\gamma)-\mathrm{QA}(a,\gamma))}{a}\right)\geq0\Rightarrow$ if $0\leq\gamma\leq1$, $\mathcal{B}_i-2\mathcal{B}_{i+1}+\mathcal{B}_{i+2}\geq0$. 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, the recurrence relation of the derivatives naturally produces the alternating binomial coefficients,

$$(-1)^\nu\frac{\partial^\nu\mathrm{QA}}{\partial\epsilon^\nu}\geq0\Rightarrow\forall0\leq a\leq\ldots\leq(\nu+1)a\leq\frac{1}{1+\gamma},$$

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

$$\geq0\Leftrightarrow\frac{(-1)^\nu}{a^\nu}\left(\sum_{j=0}^{\nu}(-1)^j\binom{\nu}{j}\mathrm{QA}\left(\left(\nu-j+1\right)a,\gamma\right)\right)\geq0$$

$$\Rightarrow\text{if }0\leq\gamma\leq1,\sum_{j=0}^{\nu}(-1)^j\binom{\nu}{j}\mathcal{B}_{i+j}\geq0.$$

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

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

where $\mathfrak{B}_{i_j}=\sum_{l=n\gamma\epsilon(j+(i-1)(\nu+1))+1}^{n\epsilon(j+(i-1)(\nu+1)+1)}\left(X_l+X_{n-l+1}\right)$. 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 $\epsilon\to0$, $\mathrm{BM}\to\mu$. 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 Text). This indicates that their robustness to departures from the symmetry assumption is practically similar, despite being based on different orders of orderliness.

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.