Supporting Information for

Semiparametric robust mean estimation based on the orderliness of quantile averages

Tuban Lee.
E-mail: tl@biomathematics.org

This PDF file includes:

Supporting text
Legend for Dataset S1
SI References

Other supporting materials for this manuscript include the following:

Dataset S1

Supporting Information Text

Methods

The asymptotic biases of all robust location estimators proposed in this article for the Weibull, gamma, Pareto, and lognormal distributions were computed within the specified range of kurtosis and standardized by the standard deviations of the respective distributions. For the Weibull, gamma, and lognormal distributions, the kurtosis range is from 3 to 15 (there are two shape parameter solutions for the Weibull distribution, the lower one is used here). For the Pareto distribution, the range is from 9 to 21.

The asymptotic results were approximated by a quasi-Monte Carlo study (1, 2) based on generating a large quasi-random sample with sample size 1.8 million for corresponding distributions and quasi-subsampling the sample $1.8k$ million times (3) to approximate the distributions of the kernels of the weighted Hodges-Lehmann mean. The accuracies were checked by comparing the results to the asymptotic values for the exponential distribution (errors), showing that the errors were all smaller than 0.001σ.

Orderliness and weighted average inequality

Unlike the mean-median-mode inequality, whose necessary and sufficient condition is often hard to compute. The following result highlights another advantage of trimming inequality.

Theorem 0.1. *A necessary and sufficient condition of the γ-trimming inequality is the monotonic behavior of the bias of trimmed mean as a function of the breakdown point ϵ.*

Proof. Just considering the right-skewed distributions without loss of generality, from the definition of γ-trimming inequality, since $\forall \epsilon_1 \leq \epsilon_2 \leq \frac{1}{2}, \mathrm{TM}_{\epsilon_1,\gamma} \geq \mathrm{TM}_{\epsilon_2,\gamma}$, therefore

$$\frac{\mathrm{TM}_{\epsilon_1,\gamma}-\mu}{\sigma} \geq \frac{\mathrm{TM}_{\epsilon_2,\gamma}-\mu}{\sigma} \iff B_{\mathrm{TM}_{\epsilon_1,\gamma}} \geq B_{\mathrm{TM}_{\epsilon_2,\gamma}} \iff B_{\mathrm{TM}}\left(\epsilon_1,\gamma\right) \geq B_{\mathrm{TM}}\left(\epsilon_2,\gamma\right),$$

which implies the monotonicity of $B_{\mathrm{TM}}(\epsilon,\gamma)$ with respect to ϵ. $\square$

The bias function is free of scale parameter, so the derivatives are much easier to compute. A useful sufficient condition is the monotonicity of symmetric quantile average with respect to ϵ.

Theorem 0.2. *A sufficient condition of the trimming inequality for a right-skewed distribution is the monotonicity of the symmetric quantile average with respect to the breakdown point ϵ.*

Proof. The trimming inequality is equivalent to, $\forall 0<\epsilon<\frac{1}{2}$, $\frac{1}{1-2\epsilon+2\delta}\int_{\epsilon-\delta}^{1-\epsilon+\delta} Q\left(u\right)du \geq \frac{1}{1-2\epsilon}\int_{\epsilon}^{1-\epsilon} Q\left(u\right)du$, where δ is an infinitesimal quantity.

Then, by deducing $\frac{1}{1-2\epsilon+2\delta}\int_{\epsilon-\delta}^{1-\epsilon+\delta} Q\left(u\right)du = \frac{1}{\frac{1}{2}-\epsilon+\delta}\int_{\epsilon-\delta}^{\frac{1}{2}} \mathrm{SQA}\left(u\right)du = \lim_{n\to\infty}\left(\frac{1}{\left(\frac{1}{2}n-n\epsilon+1\right)}\sum_{i=n\epsilon-1}^{\frac{1}{2}n}\frac{X_i+X_{n-i}}{2}\right)$
$> \lim_{n\to\infty}\left(\frac{1}{\left(\frac{1}{2}n-n\epsilon+1\right)}\left(\left(\sum_{i=n\epsilon}^{\frac{1}{2}n}\frac{X_i+X_{n-i}}{2}\right)+\frac{X_{n\epsilon}+X_{n-n\epsilon}}{2}\right)\right)$
$= \lim_{n\to\infty}\left(\frac{1}{\left(\frac{1}{2}n-n\epsilon\right)}\left(\frac{\left(\frac{1}{2}n-n\epsilon\right)}{\left(\frac{1}{2}n-n\epsilon+1\right)}\left(\sum_{i=n\epsilon+1}^{\frac{1}{2}n}\frac{X_i+X_{n-i}}{2}\right)+\frac{2\left(\frac{1}{2}n-n\epsilon\right)}{\left(\frac{1}{2}n-n\epsilon+1\right)}\frac{X_{n\epsilon}+X_{n-n\epsilon}}{2}\right)\right)$
$= \lim_{n\to\infty}\left(\frac{1}{\left(\frac{1}{2}n-n\epsilon\right)}\left(\frac{\left(\frac{1}{2}n-n\epsilon\right)}{\left(\frac{1}{2}n-n\epsilon+1\right)}\left(\sum_{i=n\epsilon+1}^{\frac{1}{2}n}\frac{X_i+X_{n-i}}{2}\right)+\frac{X_{n\epsilon}+X_{n-n\epsilon}}{2}+\frac{\left(\frac{1}{2}n-n\epsilon-1\right)}{\left(\frac{1}{2}n-n\epsilon+1\right)}\frac{X_{n\epsilon}+X_{n-n\epsilon}}{2}\right)\right)$
$> \lim_{n\to\infty}\left(\frac{1}{\frac{1}{2}n-n\epsilon}\left(\frac{\frac{1}{2}n-n\epsilon}{\frac{1}{2}n-n\epsilon+1}\sum_{i=n\epsilon+1}^{\frac{1}{2}n}\frac{X_i+X_{n-i}}{2}+\frac{X_{n\epsilon}+X_{n-n\epsilon}}{2}+\frac{1}{\frac{1}{2}n-n\epsilon+1}\left(\frac{X_{n\epsilon+1}+X_{n-n\epsilon-1}}{2}+\cdots+\frac{X_{\frac{1}{2}n}+X_{\frac{1}{2}n}}{2}\right)\right)\right)$
$= \lim_{n\to\infty}\left(\frac{1}{\left(\frac{1}{2}n-n\epsilon\right)}\left(\left(\sum_{i=n\epsilon}^{\frac{1}{2}n}\frac{X_i+X_{n-i}}{2}\right)\right)\right) = \frac{1}{1-2\epsilon}\int_{\epsilon}^{1-\epsilon} Q\left(u\right)du$, the proof is complete. $\square$

Theorem 0.3. *A necessary and sufficient condition of the γ-orderliness is the monotonic behavior of the bias of the quantile average as a function of the breakdown point ϵ.*

Proof. The proof is analogous to Theorem 0.1, just replacing the TM as QA. $\square$

Orderliness is more fundamental than trimming inequality; for example, orderliness also implies Winsorization inequality.

Definition 0.1 (Winsorization inequality). A distribution follows the Winsorization inequality if and only if $\forall \epsilon_1 \leq \epsilon_2 \leq \frac{1}{2}, \mathrm{WM}_{\epsilon_1} \geq \mathrm{WM}_{\epsilon_2}$, or $\forall \epsilon_1 \leq \epsilon_2 \leq \frac{1}{2}, \mathrm{WM}_{\epsilon_1} \leq \mathrm{WM}_{\epsilon_2}$.

Theorem 0.4. *A sufficient condition of the Winsorization inequality for a right-skewed distribution is the monotonicity of the symmetric quantile average function with respect to the breakdown point ϵ.*

Proof. The Winsorization inequality in Definition 0.1 is equivalent to,

$\forall 0<\epsilon<\frac{1}{2}$, $\int_{\epsilon-\delta}^{1-\epsilon+\delta} Q(u)\,du+(\epsilon-\delta)(Q(\epsilon-\delta)+Q(1-\epsilon+\delta)) \geq \int_{\epsilon}^{1-\epsilon} Q(u)\,du+\epsilon(Q(\epsilon)+Q(1-\epsilon))$, where δ is an infinitesimal quantity.

Then, by deducing $\int_{\epsilon-\delta}^{1-\epsilon+\delta} Q(u)\,du+(\epsilon-\delta)(Q(\epsilon-\delta)+Q(1-\epsilon+\delta))=\int_{\epsilon-\delta}^{\frac{1}{2}} SQA(u)\,du+2(\epsilon-\delta)\,SQA(\epsilon-\delta)$

$=\lim_{n\to\infty}\left(\frac{2}{n}\left(\sum_{i=n\epsilon-1}^{\frac{1}{2}n}\frac{X_i+X_{n-i}}{2}+(n\epsilon-1)(\frac{X_{n\epsilon-1}+X_{n-n\epsilon+1}}{2})\right)\right)$

$>\lim_{n\to\infty}\left(\frac{2}{n}\left(\sum_{i=n\epsilon}^{\frac{1}{2}n}\frac{X_i+X_{n-i}}{2}+\frac{X_{n\epsilon}+X_{n-n\epsilon}}{2}+(n\epsilon-1)\left(\frac{X_{n\epsilon-1}+X_{n-(n\epsilon-1)}}{2}\right)\right)\right)$

$=\lim_{n\to\infty}\left(\frac{2}{n}\left(\sum_{i=n\epsilon}^{\frac{1}{2}n}\frac{X_i+X_{n-i}}{2}+(n\epsilon)\left(\frac{X_{n\epsilon}+X_{n-n\epsilon}}{2}\right)\right)\right)$

$=\int_{\epsilon}^{1-\epsilon} Q(u)\,du+(\epsilon)(Q(\epsilon)+Q(1-\epsilon))$, the proof is complete. □

Theorem 0.5. *A necessary and sufficient condition of the second orderliness is the monotonicity and convexity of the bias function of the symmetric quantile average with respect to the breakdown point ϵ.*

Proof. The proof is analogous to Theorem 0.1. □

Then, the orderliness for parametric distributions will be discussed. For simplicity, $0<\epsilon<\frac{1}{2}$ is assumed in the following proofs unless otherwise specified.

Theorem 0.6. *The Weibull distribution is ordered if the shape parameter $\alpha \leqslant \frac{1}{1-\ln(2)} \approx 3.259$.*

Proof. The pdf of the Weibull distribution is $f(x)=\frac{\alpha e^{-\left(\frac{x}{\lambda}\right)^{\alpha}}\left(\frac{x}{\lambda}\right)^{\alpha-1}}{\lambda}$, $x\geq 0$, the quantile function is $F^{-1}(p)=\lambda(-\ln(1-p))^{1/\alpha}$, $1\geq p\geq 0$, $\alpha>0,\lambda>0$. Then, the standardized bias of a symmetric quantile average with a breakdown point ϵ, is

$$B_{\mathrm{SQA}}(\epsilon,\alpha)=\frac{\mathrm{SQA}_{\epsilon}-\mu}{\sigma}=\frac{\frac{1}{2}\left(\lambda(-\ln(1-\epsilon))^{1/\alpha}+\lambda(-\ln(\epsilon))^{1/\alpha}\right)-\lambda\Gamma\left(1+\frac{1}{\alpha}\right)}{\sqrt{\lambda^2\left(\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma\left(1+\frac{1}{\alpha}\right)^2\right)}}.$$

$\frac{\partial B_{\mathrm{SQA}}}{\partial\epsilon}=\frac{\frac{(-\ln(1-\epsilon))^{\frac{1}{\alpha}-1}}{\alpha-\alpha\epsilon}+\frac{(-\ln(\epsilon))^{1/\alpha}}{\alpha\epsilon\ln(\epsilon)}}{2\sqrt{\Gamma\left(\frac{\alpha+2}{\alpha}\right)-\Gamma\left(1+\frac{1}{\alpha}\right)^2}}$. Let $g(\epsilon,\alpha)=\frac{(-\ln(1-\epsilon))^{\frac{1}{\alpha}-1}}{\alpha-\alpha\epsilon}+\frac{(-\ln(\epsilon))^{1/\alpha}}{\alpha\epsilon\ln(\epsilon)}=-(-\ln(1-\epsilon))^{\frac{1}{\alpha}}((1-\epsilon)(\ln(1-\epsilon)))^{-1}+(-\ln(\epsilon))^{\frac{1}{\alpha}}(\epsilon\ln(\epsilon))^{-1}$. Arranging the equation $g(\epsilon,\alpha)=0$, it can be shown that $\frac{\epsilon}{(1-\epsilon)}=\left(\frac{\ln(\epsilon)}{\ln(1-\epsilon)}\right)^{\frac{1}{\alpha}-1}$. Let $L(\epsilon)=\frac{\epsilon}{(1-\epsilon)}$, $R(\epsilon,\alpha)=\left(\frac{\ln(\epsilon)}{\ln(1-\epsilon)}\right)^{\frac{1}{\alpha}-1}$, $LmR(\epsilon,\alpha)=L(\epsilon)-R(\epsilon,\alpha)$, then $\frac{\partial LmR}{\partial\alpha}=\frac{\ln\left(\frac{\ln(\epsilon)}{\ln(1-\epsilon)}\right)\left(\frac{\ln(\epsilon)}{\ln(1-\epsilon)}\right)^{\frac{1}{\alpha}-1}}{\alpha^2}$. For $0<\epsilon<\frac{1}{2}$, $\frac{\partial LmR}{\partial\alpha}>0$, so $LmR(\epsilon,\alpha)$ is monotonic with respect to α. When $\alpha=\frac{1}{1-\ln(2)}$, $g(\epsilon)=-\frac{1}{\epsilon(-\ln(\epsilon))^{\ln(2)}}+\frac{1}{(1-\epsilon)(-\ln(1-\epsilon))^{\ln(2)}}$ Let $h(\epsilon)=\epsilon(-\ln(\epsilon))^{\ln(2)})$, $h'(\epsilon)=\frac{(-\ln(\epsilon))^{\ln(2)}\ln(2\epsilon)}{\ln(\epsilon)}$, for $0<\epsilon<e^{-ln(2)}=\frac{1}{2}$, $h'(\epsilon)>0$. As a result, $h(\epsilon)$ is monotonic increasing, $-h(1-\epsilon)$ is monotonic increasing, $h(\epsilon)-h(1-\epsilon)$ is also monotonic increasing. So, if $0<\epsilon<\frac{1}{2}$, $h(\epsilon)-h(1-\epsilon)<h\left(\frac{1}{2}\right)-h\left(1-\frac{1}{2}\right)=0$, $g(\epsilon,\alpha)<0$. So, $\frac{\partial B_{\mathrm{SQA}}}{\partial\epsilon}<0$, $B_{\mathrm{SQA}}(\epsilon,\alpha)$ is monotonic decreasing in ϵ when $\alpha\leqslant\frac{1}{1-\ln(2)}$. The assertion follows from Theorem 0.3. □

Remark. The Weibull distribution can be symmetric. Its skewness is $\tilde{\mu}_3=\frac{2\Gamma\left(1+\frac{1}{\alpha}\right)^3-3\Gamma\left(1+\frac{2}{\alpha}\right)\Gamma\left(1+\frac{1}{\alpha}\right)+\Gamma\left(1+\frac{3}{\alpha}\right)}{\left(\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma\left(1+\frac{1}{\alpha}\right)^2\right)^{3/2}}$. Denote the solution of $\tilde{\mu}_3=0$ as $\alpha_0\approx 3.602$. The above proof implies that when α is close to α_0, the bias function of SQA is no longer monotonic.

Then, the bias function of the trimmed mean for the Weibull distribution is

$$B_{\mathrm{TM}}(\epsilon,\alpha)=\frac{\frac{\Gamma\left(1+\frac{1}{\alpha},-\ln(1-\epsilon)\right)-\Gamma\left(1+\frac{1}{\alpha},-\ln(\epsilon)\right)}{1-2\epsilon}-\Gamma\left(1+\frac{1}{\alpha}\right)}{\sqrt{\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma\left(1+\frac{1}{\alpha}\right)^2}},$$

Numerical solutions and the plot in Figure **??** indicate that, when $\alpha=3.259,\epsilon=0.1$, $\frac{\partial B_{\mathrm{TM}}}{\partial\epsilon}\approx-0.164$, when $\epsilon=0.4$, $\frac{\partial B_{\mathrm{TM}}}{\partial\epsilon}\approx-0.001$, when $\epsilon=0.499$, $\frac{\partial B_{\mathrm{TM}}}{\partial\epsilon}\approx 7.66\times10^{-8}$. So, the bias function of the trimmed mean can also be non-monotonic when $\alpha\geq\frac{1}{1-\ln(2)}$.

SI Dataset S1 (dataset_one.xlsx)

Raw data of asymptotic bias of all estimators shown in Figure 1 in the main text.

References

1. RD Richtmyer, A non-random sampling method, based on congruences, for" monte carlo" problems, (New York Univ., New York. Atomic Energy Commission Computing and Applied . . .), Technical report (1958).
2. IM Sobol', On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki* **7**, 784–802 (1967).
3. KA Do, P Hall, Quasi-random resampling for the bootstrap. *Stat. Comput.* **1**, 13–22 (1991).