Semiparametric robust mean estimations based on the orderliness of quantile averages

Tuban Lee

12

20

21

22

23

27

31

This manuscript was compiled on May 31, 2023

As one of the most fundamental problems in statistics, the robust location estimation has many prominent solutions, such as the symmetric trimmed mean, symmetric Winsorized mean, Hodges–Lehmann estimator, Huber M-estimator, and median of means. Recent studies suggest that their maximum biases concerning the mean can be quite different in asymmetric distributions, but the underlying mechanisms and average performance remain largely unclear. This study establishes several forms of orderliness among quantile averages, similar to the mean-median-mode inequality, within a wide range of semi-parametric distributions, particularly highlighting the unique role of γ -symmetric distributions. From this, a sequence of advanced robust mean estimators emerges, which also explains why the Winsorized mean and median of means typically have smaller biases compared to the trimmed mean. Building on the U-orderliness, the superiority of the median Hodges–Lehmann mean is discussed.

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

n 1823, Gauss (1) proved that for any unimodal distribution, $|m-\mu| \leq \sqrt{\frac{3}{4}}\omega$ and $\sigma \leq \omega \leq 2\sigma$, where μ is the population mean, m is the population median, ω is the root mean square deviation from the mode, and σ is the population standard deviation. This pioneering work revealed that despite potential bias in robust mean estimates, the deviation remains bounded in units of a scale parameter under certain assumptions. Bernard, Kazzi, and Vanduffel (2020) (2) further derived asymptotic bias bounds of any quantile for unimodal distributions with finite second moments, by reducing this optimization problem to a parametric one, which can be solved analytically. They showed that m has the smallest maximum distance to μ among all symmetric quantile averages (SQA_c). Daniell, in 1920, (3) analyzed a class of estimators, linear combinations of order statistics, and identified that the ϵ -symmetric trimmed mean (STM_{ϵ}) belongs to this class. Another popular choice, the ϵ -symmetric Winsorized mean (SWM $_{\epsilon}$), named after Winsor and introduced by Tukey (4) and Dixon (5) in 1960, is also an L-estimator. Bieniek (2016) derived exact bias upper bounds of the Winsorized mean based on Danielak and Rychlik's work (2003) on the trimmed mean for any distribution with a finite second moment and confirmed that the former is smaller than the latter (6, 7). In 1963, Hodges and Lehmann (8) proposed a class of nonparametric location estimators based on rank tests and, from the Wilcoxon signedrank statistic (9), deduced the median of pairwise means as a robust location estimator for a symmetric population. Both L-statistics and R-statistics achieve robustness essentially by removing a certain proportion of extreme values. In 1964, Huber (10) generalized maximum likelihood estimation to the minimization of the sum of a specific loss function, which measures the residuals between the data points and the model's parameters. Some L-estimators are also M-estimators, e.g., the sample mean is an M-estimator with a squared error loss

function, the sample median is an M-estimator with an absolute error loss function (10). The Huber M-estimator is obtained by applying the Huber loss function that combines elements of both squared error and absolute error to achieve robustness against gross errors and high efficiency for contaminated Gaussian distributions (10). Sun, Zhou, and Fan (2020) examined the concentration bounds of the Huber M-estimator (11). Mathieu (2022) (12) further derived the concentration bounds of M-estimators and demonstrated that, by selecting the tuning parameter which depends on the variance, the Huber M-estimator can also be a sub-Gaussian estimator. The concept of the median of means $(MoM_{k,b=\frac{n}{k},n})$ was first introduced by Nemirovsky and Yudin (1983) in their work on stochastic optimization (13). Given its good performance even for distributions with infinite second moments, the MoM has received increasing attention over the past decade (14-17). Devroye, Lerasle, Lugosi, and Oliveira (2016) showed that $MoM_{k,b=\frac{n}{h},n}$ nears the optimum of sub-Gaussian mean estimation with regards to concentration bounds when the distribution has a heavy tail (15). Laforgue, Clemencon, and Bertail (2019) proposed the median of randomized means $(MoRM_{k,b,n})$ (16), wherein, rather than partitioning, an arbitrary number, b, of blocks are built independently from the sample, and showed that $MoRM_{k,b,n}$ has a better nonasymptotic sub-Gaussian property compared to $MoM_{k,b=\frac{n}{r},n}$. In fact, asymptotically, the Hodges-Lehmann (H-L) estimator is equivalent to $MoM_{k=2,b=\frac{n}{h}}$ and $MoRM_{k=2,b}$, and they can be seen as the pairwise mean distribution is approximated by the sampling without replacement and bootstrap, respectively. When $k \ll n$, the difference between sampling with replacement and without replacement is negligible. For the asymptotic validity, readers are referred to the foundational works of Efron (1979) (18), Bickel and Freedman (1981, 1984) (19, 20), and Helmers, Janssen, and Veraverbeke (1990) (21).

35

36

37

41

42

43

44

45

46

47

50

51

52

53

57

59

60

61

62

64

65

66

67

Significance Statement

In 1964, van Zwet introduced the convex transformation order for comparing the skewness of two distributions. This paradigm shift played a fundamental role in defining robust measures of distributions, from spread to kurtosis. Here, instead of examining the stochastic ordering between two distributions, the orderliness of quantile averages within a distribution is investigated. By classifying distributions through the signs of derivatives, a series of sophisticated robust mean estimators is deduced. Nearly all common nonparametric robust location estimators are found to be special cases thereof.

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

Here, the ϵ,b -stratified mean is defined as

69

70

71

72

73

74

75

76

77

78

79

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103 104

105

106

107

108

109

110

111

112

$$\mathrm{SM}_{\epsilon,b,n} \coloneqq \frac{b}{n} \left(\sum_{j=1}^{\frac{b-1}{2b\epsilon}} \sum_{i_j = \frac{(2bj-b+1)n\epsilon}{b-1}}^{\frac{(2bj-b+1)n\epsilon}{b-1}} X_{i_j} \right),$$

where $X_1 \leq \ldots \leq X_n$ denote the order statistics of a sample of n independent and identically distributed random variables X_1, \ldots, X_n . $b \in \mathbb{N}, b \geq 3$. The definition was further refined to guarantee the continuity of the breakdown point by incorporating an additional block in the center when $\lfloor \frac{b-1}{2b\epsilon} \rfloor \mod 2 = 0$, or by adjusting the central block when $\lfloor \frac{b-1}{2b\epsilon} \rfloor \mod 2 = 1$ (SI Text). If the subscript n is omitted, only the asymptotic behavior is considered. If b is omitted, b = 3 is assumed. $\mathrm{SM}_{\epsilon,b=3}$ is equivalent to STM_{ϵ} , when $\epsilon > \frac{1}{6}$. When $\frac{b-1}{2\epsilon} \in \mathbb{N}$ and $b \mod 2 = 1$, the basic idea of the stratified mean is to distribute the data into $\frac{b-1}{2\epsilon}$ equal-sized non-overlapping blocks according to their order. Then, further sequentially group these blocks into b equal-sized strata and compute the mean of the middle stratum, which is the median of means of each stratum. In situations where $i \mod 1 \neq 0$, a potential solution is to generate multiple smaller samples that satisfy the equality by sampling without replacement, and subsequently calculate the mean of all estimations. The details of determining the smaller sample size and the number of sampling times are provided in the SI Text. Although the principle resembles that of the median of means, $SM_{\epsilon,b,n}$ is different from $MoM_{k=\frac{n}{L},b,n}$ as it does not include the random shift. Additionally, the stratified mean differs from the mean of the sample obtained through stratified sampling methods, introduced by Nevman (1934) (22) or ranked set sampling (23), introduced by McIntyre in 1952, as these sampling methods aim to obtain more representative samples or improve the efficiency of sample estimates, but the sample means based on them are not robust. When $b \mod 2 = 1$, the stratified mean can be regarded as replacing the other equal-sized strata with the middle stratum, which, in principle, is analogous to the Winsorized mean that replaces extreme values with less extreme percentiles. Furthermore, while the bounds confirm that the Winsorized mean and median of means outperform the trimmed mean (6, 7, 15) in worst-case performance, the complexity of bound analysis makes it difficult to achieve a complete and intuitive understanding of these results. Also, a clear explanation for the average performance of them remains elusive. The aim of this paper is to define a series of semiparametric models using the signs of derivatives, reveal their elegant interrelations and connections to parametric models, and show that by exploiting these models, a set of sophisticated mean estimators can be deduced, which exhibit strong robustness to departures from assumptions.

Quantile Average and Weighted Average

The symmetric trimmed mean, symmetric Winsorized mean, and stratified mean are all L-estimators. More specifically, they are symmetric weighted averages, which are defined as

$$\mathrm{SWA}_{\epsilon,n} \coloneqq \frac{\sum_{i=1}^{\lceil \frac{n}{2} \rceil} \frac{X_i + X_{n-i+1}}{2} w_i}{\sum_{i=1}^{\lceil \frac{n}{2} \rceil} w_i},$$

where w_i s are the weights applied to the symmetric quantile averages according to the definition of the corresponding L-estimators. For example, for the ϵ -symmetric trimmed mean,

 $w_i = \begin{cases} 0, & i < n\epsilon \\ 1, & i \ge n\epsilon \end{cases}$, provided that $n\epsilon \in \mathbb{N}$. The mean and median are indeed two special cases of the symmetric trimmed mean.

To extend the symmetric quantile average to the asymmetric case, two definitions for the ϵ, γ -quantile average (QA_{ϵ, γ, n}) are proposed. The first definition is:

$$\frac{1}{2}(\hat{Q}_n(\gamma\epsilon) + \hat{Q}_n(1-\epsilon)), \qquad [1]$$

118

121

122

123

124

128

129

130

131

132

133

135

136

138

139

141

142

143

144

145

Lee

and the second definition is:

$$\frac{1}{2}(\hat{Q}_n(\epsilon) + \hat{Q}_n(1 - \gamma \epsilon)), \qquad [2] \quad {}_{125}$$

where $\hat{Q}_n(p)$ is the empirical quantile function; γ is used to adjust the degree of asymmetry, $\gamma \geq 0$; and $0 \leq \epsilon \leq \frac{1}{1+\gamma}$. For trimming from both sides, [1] and [2] are essentially equivalent. The first definition along with $\gamma \geq 0$ and $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ are assumed in the rest of this article unless otherwise specified, since many common asymmetric distributions are right-skewed, and [1] allows trimming only from the right side by setting $\gamma = 0$.

Analogously, the weighted average can be defined as

$$WA_{\epsilon,\gamma,n} := \frac{\int_{\epsilon_0=0}^{\frac{1}{1+\gamma}} QA(\epsilon_0,\gamma,n) w_{\epsilon_0}}{\int_{\epsilon_0=0}^{\frac{1}{1+\gamma}} w_{\epsilon_0}}.$$

For any weighted average, if γ is omitted, it is assumed to be 1. The ϵ, γ -trimmed mean $(\mathrm{TM}_{\epsilon, \gamma, n})$ is a weighted average with a left trim size of $\gamma \epsilon n$ and a right trim size of ϵn , where $w_{\epsilon_0} = \begin{cases} 0, & \epsilon_0 < \epsilon \\ 1, & \epsilon_0 \geq \epsilon \end{cases}.$ Using this definition, the TM computation remains the same, regardless of whether $\gamma \epsilon n \notin \mathbb{N}$ or $\epsilon n \notin \mathbb{N}$, since this definition is based on the empirical quantile function. However, in this article, considering the computational cost in practice, non-asymptotic definitions of various types of weighted averages are primarily based on order statistics. Unless stated otherwise, the solution to their decimal issue is the same as that in SM.

Classifying Distributions by the Signs of Derivatives

Let \mathcal{P}_{Υ} denote the set of all continuous distribution over \mathbb{R} . Without loss of generality, all the classes discussed below are confined to the intersection with this nonparametric class of distributions. Besides fully and smoothly parameterizing them by a Euclidean parameter or merely assuming regularity conditions, there are many ways to classify distributions. In 1956, Stein initiated the problem of estimating parameters in the presence of an infinite dimensional nuisance shape parameter (24). A notable example discussed in his groundbreaking work was the estimation of the center of symmetry for an unknown symmetric distribution. In 1993, Bickel, Klaassen, Ritov, and Wellner published an influential semiparametrics textbook (25), which systematically categorized many common models into three classes: parametric, nonparametric, and semiparametric. Yet, there is another old and commonly encountered class of distributions that receives little attention in semiparametric literature: the unimodal distribution. It is a very unique semiparametric model because its definition is based on the signs of derivatives, i.e., $(f'(x) > 0 \text{ for } x \leq M) \land$

2 |

 $(f'(x) < 0 \text{ for } x \ge M)$, where f(x) is the probability density function (pdf) of a random variable X, M is the mode. Let \mathcal{P}_U denote the set of all unimodal distributions. There was a widespread misbelief that the median of an arbitrary unimodal distribution always lies between its mean and mode until Runnenburg (1978) and van Zwet (1979) (26, 27) endeavored to determine sufficient conditions for the inequality to hold, thereby implying the possibility of its violation. The class of distributions that satisfy the mean-median-mode inequality constitutes a subclass of \mathcal{P}_U . By analogy, a right-skewed distribution is called γ -ordered, if and only if

$$\forall 0 \le \epsilon_1 \le \epsilon_2 \le \frac{1}{1+\gamma}, \mathrm{QA}(\epsilon_1, \gamma) \ge \mathrm{QA}(\epsilon_2, \gamma).$$

The necessary and sufficient condition below hints at the relation between the mean-median-mode inequality and the γ -orderliness.

Theorem .1. A distribution is γ -ordered if and only if its pdf satisfies the inequality $f(Q(\gamma \epsilon)) \geq f(Q(1-\epsilon))$ for all $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ or $f(Q(\gamma \epsilon)) \leq f(Q(1-\epsilon))$ for all $0 \leq \epsilon \leq \frac{1}{1+\gamma}$.

Proof. Without loss of generality, consider the case of right-skewed distribution. From the definition of γ -orderliness, it is deduced that $\frac{Q(\gamma\epsilon-\delta)+Q(1-\epsilon+\delta)}{2} \geq \frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2} \Leftrightarrow Q(\gamma\epsilon-\delta)-Q(\gamma\epsilon) \geq Q(1-\epsilon)-Q(1-\epsilon+\delta) \Leftrightarrow Q'(1-\epsilon) \geq Q'(\gamma\epsilon)$, where δ is an infinitesimal positive quantity. Observing that the quantile function is the inverse function of the cumulative distribution function $(\operatorname{cdf}), Q'(1-\epsilon) \geq Q'(\gamma\epsilon) \Leftrightarrow F'(Q(\gamma\epsilon)) \geq F'(Q(1-\epsilon))$, thereby completing the proof, given that the derivative of cdf is pdf.

According to Theorem .1, if a probability distribution is right-skewed and monotonic, it will always be γ -ordered. For a right-skewed unimodal distribution, if $Q(\gamma \epsilon) > M$, then the inequality $f(Q(\gamma \epsilon)) \geq f(Q(1 - \epsilon))$ holds. The principle is extendable to unimodal-like distributions. Suppose there is a right-skewed multimodal distribution following the mean- γ median-first mode inequality with several smaller modes on the right side, with the first mode, M_1 , having the greatest probability density, and the γ -median, $Q(\frac{\gamma}{1+\gamma})$, falling within the first dominant mode (i.e., if $x > Q(\frac{\gamma}{1+\gamma}), f(Q(\frac{\gamma}{1+\gamma})) \ge$ f(x)), then if $Q(\gamma \epsilon) > M_1$, the inequality $f(Q(\gamma \epsilon)) \geq f(Q(1 - \epsilon))$ ϵ)) also holds. In other words, even though a distribution following the mean- γ -median-mode inequality may not be strictly γ -ordered, the inequality defining γ -orderliness remains valid for most quantile averages. The mean- γ -median-mode inequality can also indicate possible bounds for γ in practice, e.g., for any distributions, when $\gamma \to \infty$, the γ -median will be greater than the mean and the mode, when $\gamma \to 0$, the γ -median will be smaller than the mean and the mode.

Consider the sign of the derivative of the quantile average with respect to the breakdown point; the above definition of γ -orderliness can also be expressed as

$$\forall 0 \le \epsilon \le \frac{1}{1+\gamma}, \frac{\partial QA}{\partial \epsilon} \le 0.$$

The left-skewed case can be obtained by reversing the inequality $\frac{\partial QA}{\partial \epsilon} \leq 0$ to $\frac{\partial QA}{\partial \epsilon} \geq 0$ and employing the second definition of QA, as given in [2]. For simplicity, it will be omitted in the following discussion. If $\gamma=1$, the γ -ordered distribution is referred to as ordered.

Furthermore, many common right-skewed distributions are partial bounded, indicating a convex behavior of the QA function when $\epsilon \to 0$. By further assuming convexity, the second γ -orderliness can be defined for a right-skewed distribution as follows.

$$\forall 0 \le \epsilon \le \frac{1}{1+\gamma}, \frac{\partial^2 QA}{\partial \epsilon^2} \ge 0 \land \frac{\partial QA}{\partial \epsilon} \le 0.$$

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

208

209

210

212

213

214

215

217

218

219

220

221

222

223

224

225

226 227

228

230

Analogously, the ν th γ -orderliness of a right-skewed distribution can be defined as $(-1)^{\nu} \frac{\partial^{\nu} QA}{\partial \epsilon^{\nu}} \geq 0 \wedge \ldots \wedge - \frac{\partial QA}{\partial \epsilon} \geq 0$. If $\gamma = 1$, the ν th γ -orderliness is referred as to ν th orderliness. Let \mathcal{P}_O denote the set of all distributions that are ordered and $\mathcal{P}_{O_{\nu}}$ and $\mathcal{P}_{\gamma O_{\nu}}$ represent the sets of all distributions that are ν th ordered and ν th γ -ordered, respectively. When the shape parameter of the Weibull distribution, α , is smaller than 3.258, it can be shown that the Weibull distribution belongs to $\mathcal{P}_U \cap \mathcal{P}_O \cap \mathcal{P}_{O_2} \cap \mathcal{P}_{O_3}$ (SI Text). At $\alpha \approx 3.602$, the Weibull distribution is symmetric, and as $\alpha \to \infty$, the skewness of the Weibull distribution approaches 1. Therefore, the parameters that prevent it from being included in the set correspond to cases when it is near-symmetric, as shown in the SI Text. Nevertheless, computing the derivatives of the QA function is often intricate and, at times, challenging. The following theorems establish the relationship between \mathcal{P}_O , $\mathcal{P}_{O_{\nu}}$, and $\mathcal{P}_{\gamma O_{\nu}}$, and a wide range of other semi-parametric distributions. They can be used to quickly identify some parametric distributions in \mathcal{P}_O , $\mathcal{P}_{O_{\nu}}$, and $\mathcal{P}_{\gamma O_{\nu}}$.

Theorem .2. For any random variable X whose probability distribution function belongs to a location-scale family, the distribution is ν th γ -ordered if and only if the family of probability distributions is ν th γ -ordered.

Proof. Let Q_0 denote the quantile function of the standard distribution without any shifts or scaling. After a location-scale transformation, the quantile function becomes $Q(p) = \lambda Q_0(p) + \mu$, where λ is the scale parameter and μ is the location parameter. According to the definition of the ν th γ -orderliness, the signs of derivatives of the QA function are invariant after this transformation. As the location-scale transformation is reversible, the proof is complete.

Theorem .2 demonstrates that in the analytical proof of the ν th γ -orderliness of a parametric distribution, both the location and scale parameters can be regarded as constants. It is also instrumental in proving other theorems.

Theorem .3. Define a γ -symmetric distribution as one for which the quantile function satisfies $Q(\gamma \epsilon) = 2Q(\frac{\gamma}{1+\gamma}) - Q(1-\epsilon)$ for all $0 \le \epsilon \le \frac{1}{1+\gamma}$. Any γ -symmetric distribution is ν th γ -ordered.

Proof. The equality implies that $\frac{\partial Q(\gamma\epsilon)}{\partial \epsilon} = \gamma Q'(\gamma\epsilon) = \frac{\partial (-Q(1-\epsilon))}{\partial \epsilon} = Q'(1-\epsilon)$. From the definition of QA, the QA function of the γ -symmetric distribution is a horizontal line, since $\frac{\partial QA}{\partial \epsilon} = \gamma Q'(\gamma\epsilon) - Q'(1-\epsilon) = 0$. So, the ν th order derivative of QA is always zero.

Theorem .4. A symmetric distribution is a special case of a γ -symmetric distribution when $\gamma = 1$.

147

148

149

150

151

152

153

155

156

157

158

159

161

162

163

164

165

166

167

168

169

171

172

173

174

175

Proof. Without loss of generality, assuming continuity. A symmetric distribution is a probability distribution such that for all x, f(x) = f(2m-x). Its cdf satisfies F(x) = 1 - F(2m-x). Let x = Q(p), then, F(Q(p)) = p = 1 - F(2m - Q(p)) and $F(Q(1-p)) = 1 - p \Leftrightarrow p = 1 - F(Q(1-p))$. Therefore, F(2m - Q(p)) = F(Q(1-p)). Since the cdf is monotonic, 2m - Q(p) = Q(1-p).

As a consequence of Theorem .3, and because the generalized Gaussian distribution is symmetric around the median, it is found to be ν th ordered.

238

240

242

249

250

251

252

253

255

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

279

280

281

Theorem .5. Any right-skewed distribution whose quantile function Q satisfies $Q^{(\nu)}(p) \geq 0 \wedge \dots Q^{(i)}(p) \geq 0 \dots \wedge Q^{(2)}(p) \geq 0$, $i \mod 2 = 0$, is ν th γ -ordered, provided that $0 \leq \gamma \leq 1$.

 $\begin{array}{ll} \text{245} & \textit{Proof.} \; \operatorname{Since} \; (-1)^i \frac{\partial^i \operatorname{QA}}{\partial \epsilon^i} = \frac{1}{2} ((-\gamma)^i Q^i (\gamma \epsilon) + Q^i (1 - \epsilon)) \; \text{and} \; 1 \leq \\ \text{246} & i \leq \nu, \; \text{when} \; i \; \text{mod} \; 2 = 0, \; (-1)^i \frac{\partial^i \operatorname{QA}}{\partial \epsilon^i} \geq 0 \; \text{for all} \; \gamma \geq 0. \; \; \text{When} \\ \text{247} & i \; \text{mod} \; 2 = 1, \; \text{if further assuming} \; 0 \leq \gamma \leq 1, \; (-1)^i \frac{\partial^i \operatorname{QA}}{\partial \epsilon^i} \geq 0, \\ \text{248} & \; \text{since} \; Q^{(i+1)} \; (p) \geq 0. \end{array}$

This result makes it straightforward to show that the Pareto distribution follows the ν th γ -orderliness, provided that $0 \le \gamma \le 1$, since the quantile function of the Pareto distribution is $Q(p) = x_m(1-p)^{-\frac{1}{\alpha}}$, where $x_m > 0$, $\alpha > 0$, and so $Q^{(\nu)}(p) \ge 0$ for all $\nu \in \mathbb{N}$ according to the chain rule.

Theorem .6. A right-skewed distribution with a monotonic decreasing pdf is second γ -ordered.

Proof. Given that a monotonic decreasing pdf implies $f'(x) = F^{(2)}(x) \leq 0$, let x = Q(F(x)), then by differentiating both sides of the equation twice, one can obtain $0 = Q^{(2)}(F(x))(F'(x))^2 + Q'(F(x))F^{(2)}(x) \Leftrightarrow Q^{(2)}(F(x)) = -\frac{Q'(F(x))F^{(2)}(x)}{(F'(x))^2} \geq 0$, since $Q'(p) \geq 0$. The desired result is derived from Theorem .1 and .5.

Theorem .6 provides valuable insights into the relation between modality and orderliness. The conventional definition states that a distribution with a monotonic pdf is still considered unimodal. However, within its supported interval, the mode number is zero. The number of modes and their magnitudes within a distribution are closely related to the possibility of orderliness being valid, although counterexamples can always be constructed for non-monotonic distributions. It can be easily established that the gamma distribution is second γ -ordered when $\alpha \leq 1$, as the pdf of the gamma distribution is $f(x) = \frac{\lambda^{-\alpha} x^{\alpha-1} e^{-\frac{x}{\lambda}}}{\Gamma(\alpha)}$, where $x \ge 0$, $\lambda > 0$, $\alpha > 0$, and Γ represents the gamma function. This pdf is a product of two monotonic decreasing functions under constraints. For $\alpha > 1$, analytical analysis becomes challenging. Numerical results show that orderliness is valid until $\alpha > 140$, the second orderliness is valid until $\alpha > 78$, and the third orderliness is valid until $\alpha > 55$ (SI Text). It is instructive to consider that when $\alpha \to \infty$, the gamma distribution converges to a Gaussian distribution with mean $\mu = \alpha \lambda$ and variance $\sigma = \alpha \lambda^2$. The skewness of the gamma distribution, $\frac{\alpha+2}{\sqrt{\alpha(\alpha+1)}}$, is monotonic with respect to α , since $\frac{\partial \tilde{\mu}_3(\alpha)}{\partial \alpha} = \frac{-3\alpha-2}{2(\alpha(\alpha+1))^{3/2}} < 0$. When $\alpha = 55$, $\tilde{\mu}_3(\alpha) = 1.027$. Theorefore, similar to the Weibull

distribution, the parameters which make these distributions

fail to be included in $\mathcal{P}_U \cap \mathcal{P}_O \cap \mathcal{P}_{O_2} \cap \mathcal{P}_{O_3}$ also correspond to cases when it is near-symmetric.

286

287

288

289

290

294

295

297

304

305

306

307

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

334

335

336

337

338

339

Theorem .7. Consider a γ -symmetric random variable X. Let it be transformed using a function $\phi(x)$ such that $\phi^{(2)}(x) \geq 0$ over the interval supported, the resulting convex transformed distribution is γ -ordered. Moreover, if the quantile function of X satisfies $Q^{(2)}(p) \leq 0$, the convex transformed distribution is second γ -ordered.

 $\begin{array}{ll} Proof. \ \ \mathrm{Let} \ \phi \mathrm{QA}(\epsilon,\gamma) = \frac{1}{2} (\phi(Q(\gamma\epsilon)) + \phi(Q(1-\epsilon))). \ \ \mathrm{Then}, \\ \frac{\partial \phi \mathrm{QA}}{\partial \epsilon} = \frac{1}{2} \left(\gamma \phi' \left(Q\left(\gamma \epsilon \right) \right) Q' \left(\gamma \epsilon \right) - \phi' \left(Q\left(1-\epsilon \right) \right) Q' \left(1-\epsilon \right) \right) = \\ \frac{1}{2} \gamma Q' \left(\gamma \epsilon \right) \left(\phi' \left(Q\left(\gamma \epsilon \right) \right) - \phi' \left(Q\left(1-\epsilon \right) \right) \right) \leq 0, \ \ \mathrm{since} \ \ \mathrm{for} \ \ \mathrm{a} \ \ \gamma - \mathrm{symmetric} \ \mathrm{distribution}, \ Q(\frac{1}{1+\gamma}) - Q\left(\gamma \epsilon \right) = Q\left(1-\epsilon \right) - Q(\frac{1}{1+\gamma}), \\ \mathrm{differentiating} \ \ \mathrm{both} \ \ \mathrm{sides}, \ -\gamma Q' \left(\gamma \epsilon \right) = -Q'(1-\epsilon), \ \ \mathrm{where} \\ Q' \left(p \right) \geq 0, \phi^{(2)} \left(x \right) \geq 0. \ \ \ \mathrm{If} \ \ \mathrm{further} \ \ \mathrm{differentiating} \ \ \mathrm{the} \\ \mathrm{equality}, \ \ \gamma^2 Q^{(2)} \left(\gamma \epsilon \right) = -Q^{(2)}(1-\epsilon). \ \ \ \mathrm{Since} \ \ \frac{\partial^{(2)} \phi \mathrm{QA}}{\partial \epsilon^{(2)}} = \\ \frac{1}{2} \left(\gamma^2 \phi^2 \left(Q\left(\gamma \epsilon \right) \right) \left(Q' \left(\gamma \epsilon \right) \right)^2 + \phi^2 \left(Q\left(1-\epsilon \right) \right) \left(Q' \left(1-\epsilon \right) \right)^2 \right) \ \ + \\ \frac{1}{2} \left(\gamma^2 \phi' \left(Q\left(\gamma \epsilon \right) \right) \left(Q^2 \left(\gamma \epsilon \right) \right) + \phi' \left(Q\left(1-\epsilon \right) \right) \left(Q^2 \left(1-\epsilon \right) \right) \right) \\ = \\ \frac{1}{2} \left(\left(\phi^{(2)} \left(Q\left(\gamma \epsilon \right) \right) + \phi^{(2)} \left(Q\left(1-\epsilon \right) \right) \right) \left(\gamma^2 Q' \left(\gamma \epsilon \right) \right)^2 \right) \ \ + \\ \frac{1}{2} \left(\left(\phi' \left(Q\left(\gamma \epsilon \right) \right) - \phi' \left(Q\left(1-\epsilon \right) \right) \right) \gamma^2 Q^{(2)} \left(\gamma \epsilon \right) \right). \ \ \mathrm{If} \ \ Q^{(2)} \left(p \right) \leq 0, \\ \frac{\partial^{(2)} \phi \mathrm{QA}}{\partial \epsilon^{(2)}} \geq 0. \ \ \Box$

The mean-median-mode inequality for distributions of the powers and roots of the variates of a given distribution was investigated by Henry Rietz in 1927 (28), but the most straightforward solution is the exponential transformation, since the derivatives are invariably positive. An application of Theorem .7 is that the lognormal distribution is ordered as it is exponentially transformed from the Gaussian distribution. The quantile function of the Gaussian distribution meets the condition $Q^{(2)}(p) = -2\sqrt{2}\pi\sigma e^{2\mathrm{erfc}^{-1}(2p)^2}\mathrm{erfc}^{-1}(2p) \leq 0$, where σ is the standard deviation of the Gaussian distribution and erfc denotes the complementary error function. Thus, the lognormal distribution is second ordered. Numerical results suggest that it is also third ordered, although analytically proving this result is challenging.

Theorem .7 also reveals a relation between convex transformation and orderliness, since ϕ is the non-decreasing convex function in van Zwet's trailblazing work Convex transformations of random variables (29) if adding an additional constraint that $\phi'(x) \geq 0$. Consider a near-symmetric distribution S, such that SQA_{ϵ} as a function of ϵ fluctuates from 0 to $\frac{1}{2}$, with $\mu = m$. By definition, S is not ordered. Let s be the pdf of S. Applying the transformation $\phi(x)$ to S decreases $s(Q_S(\epsilon))$, and the decrease rate, due to the order, is much smaller for $s(Q_S(1-\epsilon))$. As a consequence, as the second derivative of $\phi(x)$ increases, eventually, after a point, $s(Q_S(\epsilon))$ becomes greater than $s(Q_S(1-\epsilon))$ even if it was not previously. Thus, the SQA function becomes monotonically decreasing, and S becomes ordered. Accordingly, in a family of distributions that differ by a skewness-increasing transformation in van Zwet's sense, violations of orderliness typically occur only when the distribution is near-symmetric.

Pearson proposed using the mean-median difference $\mu-m$ as a measure of skewness after standardization in 1895 (30). Bowley (1926) proposed a measure of skewness based on the SQA-median difference SQA_{ϵ} – m (31). Groeneveld and Meeden (1984) (32) generalized these measures of skewness based on van Zwet's convex transformation (29) while exploring their

4 | Lee

properties. A distribution is called monotonically right-skewed if and only if $\forall 0 \le \epsilon_1 \le \epsilon_2 \le \frac{1}{2}$, $SQA_{\epsilon_1} - m \ge SQA_{\epsilon_2} - m$. Since m is a constant, the monotonic skewness is equivalent to the orderliness. For a nonordered distribution, the signs of $SQA_{\epsilon} - m$ with different breakdown points might be different, implying that some skewness measures indicate left-skewed distribution, while others suggest right-skewed distribution. Although it seems reasonable that such a distribution is likely be generally near-symmetric, counterexamples can be constructed. For example, consider the Weibull distribution, when $\alpha > \frac{1}{1-\ln(2)}$, it is near-symmetric and nonordered, the non-monotonicity of the SQA function arises when ϵ is close to $\frac{1}{2}$. Replacing the third quartile with one from a right-skewed heavy-tailed distribution leads to a right-skewed, heavy-tailed, and nonordered distribution. Therefore, the validity of robust measures of skewness based on the SQA-median difference is closely related to the orderliness of the distribution.

Remarkably, in 2018, Li, Shao, Wang, Yang (33) 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 .8. The bias upper bound of the quantile average for any continuous distribution whose mean is zero and standard deviation is one is

$$\sup_{P \in \mathcal{P}_{\mu=0,\sigma=1}} QA(\epsilon, \gamma) = \frac{1}{2} \left(\sqrt{\frac{\gamma \epsilon}{1 - \gamma \epsilon}} + \sqrt{\frac{1 - \epsilon}{\epsilon}} \right), \quad [3]$$

where $0 \le \epsilon \le \frac{1}{1+\alpha}$. 369

342

343

344

345

348

349

350

351

352

355

356

357

358

359

360

362

363

364

365

370

373

374

375

376

377

Proof. Since
$$\sup_{P \in \mathcal{P}_{\mu=0,\sigma=1}} \frac{1}{2}(Q(\gamma\epsilon) + Q(1-\epsilon)) \leq \frac{1}{2}(\sup_{P \in \mathcal{P}_{\mu=0,\sigma=1}} Q(\gamma\epsilon) + \sup_{P \in \mathcal{P}_{\mu=0,\sigma=1}} Q(1-\epsilon)),$$
 the assertion follows directly from the Lemma 2.6 in (33).

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

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

reveal the safeguarding role these bounds play in defining estimators based on ν th γ -orderliness. The proof of Theorem .9 is provided in the SI Text.

Theorem .9. $\sup_{P \in \mathcal{P}_{\mu=0,\sigma=1}} QA(\epsilon, \gamma)$ is monotonic decreasing with respect to ϵ over the interval $[0, \frac{1}{1+\gamma}]$, when $0 \le \gamma \le 1$. 378 379

380 382

Proof. When
$$0 \le \epsilon \le \frac{1}{6}$$
, $\frac{\partial \sup QA}{\partial \epsilon} = \frac{\gamma}{\sqrt{\frac{\epsilon \gamma}{12 - 9\epsilon \gamma}}(4 - 3\epsilon \gamma)^2} - \frac{1}{3\sqrt{\frac{4}{3} - 9\epsilon^2}}$. When $\gamma = 0$, $\frac{\partial \sup QA}{\partial \epsilon} = -\frac{1}{3\sqrt{\frac{4}{3} - 9\epsilon^2}} \le 0$

When $\epsilon \to 0^+$, $\lim_{\epsilon \to 0^+} \left(\frac{\gamma}{(4-3\gamma\epsilon)^2 \sqrt{\frac{\epsilon\gamma}{12-9\gamma\epsilon}}} - \frac{1}{3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2} \right)$ $\lim_{\epsilon \to 0^+} \left(\frac{\gamma \sqrt{3}}{\sqrt{4^3 \epsilon \gamma}} - \frac{1}{6\sqrt{\epsilon^3}} \right) \to -\infty.$ Assuming $\epsilon > 0$, when $0 < \gamma \le 1$, to prove $\frac{\partial \sup QA}{\partial \epsilon} \le 0$, it is equivalent to showing $\frac{\sqrt{\frac{\epsilon \gamma}{12 - 9\epsilon \gamma}}(4 - 3\epsilon \gamma)^2}{\gamma} \ge 3\sqrt{\frac{\epsilon}{\epsilon} - 9}\epsilon^2$. Define $L(\epsilon, \gamma) = \frac{\sqrt{\frac{\epsilon \gamma}{12 - 9\epsilon \gamma}}(4 - 3\epsilon \gamma)^2}{\gamma}$, $R(\epsilon, \gamma) = 3\sqrt{\frac{\epsilon}{\epsilon} - 9}\epsilon^2$. $\frac{L(\epsilon, \gamma)}{\epsilon^2} = \frac{\sqrt{\frac{\epsilon \gamma}{12 - 9\epsilon \gamma}}(4 - 3\epsilon \gamma)^2}{\gamma \epsilon^2} = \frac{1}{\gamma} \left(\frac{\epsilon}{\epsilon} - 3\gamma\right)^2 \sqrt{\frac{1}{\frac{12}{\epsilon \gamma} - 9}}$, $\frac{R(\epsilon, \gamma)}{\epsilon^2} = 3\sqrt{\frac{\epsilon}{\epsilon} - 9}$. Then, $\frac{L(\epsilon,\gamma)}{\epsilon^2} \ge \frac{R(\epsilon,\gamma)}{\epsilon^2} \Leftrightarrow \frac{1}{\gamma} \sqrt{\frac{1}{\frac{1}{2\gamma}-9}} \left(\frac{4}{\epsilon} - 3\gamma\right)^2 \ge 3\sqrt{\frac{4}{\epsilon}-9} \Leftrightarrow$ 391 $\frac{1}{\gamma} \left(\frac{4}{\epsilon} - 3\gamma \right)^2 \geq 3\sqrt{\frac{12}{\epsilon\gamma} - 9}\sqrt{\frac{4}{\epsilon} - 9}.$ Let $LmR\left(\frac{1}{\epsilon} \right) =$ $\begin{array}{l} \frac{1}{\gamma^2} \left(\frac{4}{\epsilon} - 3\gamma\right)^4 - 9 \left(\frac{12}{\epsilon\gamma} - 9\right) \left(\frac{4}{\epsilon} - 9\right). \ \frac{\partial LmR(1/\epsilon)}{\partial (1/\epsilon)} = \frac{16 \left(\frac{4}{\epsilon} - 3\gamma\right)^3}{\gamma^2} - \\ 36 \left(\frac{12}{\epsilon\gamma} - 9\right) - \frac{108 \left(4\frac{4}{\epsilon} - 9\right)}{\gamma} = \frac{4 \left(4 \left(\frac{4}{\epsilon} - 3\gamma\right)^3 - 27\gamma \left(\frac{4}{\epsilon} - 3\gamma\right) + 27(9 - \frac{4}{\epsilon})\gamma\right)}{\gamma^2} = \\ \frac{4 \left(256 \frac{1}{\epsilon}^3 - 576 \frac{1}{\epsilon}^2 \gamma + 432 \frac{1}{\epsilon} \gamma^2 - 216 \frac{1}{\epsilon} \gamma - 108\gamma^3 + 81\gamma^2 + 243\gamma\right)}{\gamma^2}. \end{array}$ Since 395 $\begin{array}{c} \gamma^2 \\ 256\frac{1}{\epsilon}^3 - 576\frac{1}{\epsilon}^2\gamma + 432\frac{1}{\epsilon}\gamma^2 - 216\frac{1}{\epsilon}\gamma - 108\gamma^3 + 81\gamma^2 + 243\gamma \geq \\ 1536\frac{1}{\epsilon}^2 - 576\frac{1}{\epsilon}^2 + 432\frac{1}{\epsilon}\gamma^2 - 216\frac{1}{\epsilon}\gamma - 108\gamma^3 + 81\gamma^2 + 243\gamma \geq \\ 924\frac{1}{\epsilon}^2 + 36\frac{1}{\epsilon}^2 - 216\frac{1}{\epsilon} + 432\frac{1}{\epsilon}\gamma^2 - 108\gamma^3 + 81\gamma^2 + 243\gamma \geq \\ 924\frac{1}{\epsilon}^2 + 36\frac{1}{\epsilon}^2 - 216\frac{1}{\epsilon} + 513\gamma^2 - 108\gamma^3 + 243\gamma > 0, \end{array}$ 397 398 399 $\frac{\partial LmR(1/\epsilon)}{\partial (1/\epsilon)} > 0. \text{ Also, } LmR(6) = \frac{81(\gamma - 8)\left((\gamma - 8)^3 + 15\gamma\right)}{\gamma^2} > 0 \iff \gamma^4 - 32\gamma^3 + 399\gamma^2 - 2168\gamma + 4096 > 0. \text{ Since } \gamma^4 > 0,$ 400 401 if $0 < \gamma \le 1$, then $32\gamma^3 < 256$, it suffices to prove that 402 $399\gamma^2 - 2168\gamma + 4096 > 256$. Applying the quadratic 403 formula demonstrates the validity of this inequality. Hence, 404 $LmR\left(\frac{1}{\epsilon}\right) \geq 0$ for $\epsilon \in (0,\frac{1}{\epsilon}]$, provided that $0 < \gamma \leq 1$. The 405 first part is finished. 406

 $\sqrt{3} \left(\frac{\gamma}{\sqrt{\gamma \epsilon (4 - 3\gamma \epsilon)^{\frac{3}{2}}}} - \frac{1}{\sqrt{1 - \epsilon (3\epsilon + 1)^{\frac{3}{2}}}} \right). \quad \text{When} \quad \gamma = \frac{\gamma}{\sqrt{\gamma \epsilon (4 - 3\gamma \epsilon)^{\frac{3}{2}}}} = \frac{\sqrt{\gamma}}{\sqrt{\epsilon (4 - 3\gamma \epsilon)^{\frac{3}{2}}}} = 0, \quad \frac{\partial \sup QA}{\partial \epsilon} < 0.$ other cases, to determine whether $\frac{\partial \sup QA}{\partial \epsilon} \leq 0$, since $\sqrt{1-\epsilon} (3\epsilon+1)^{\frac{3}{2}} > 0$ and $\sqrt{\gamma\epsilon} (4-3\gamma\epsilon)^{\frac{3}{2}} > 0$, show- $\operatorname{ing} \frac{\sqrt{\gamma \epsilon (4-3\gamma \epsilon)^{\frac{3}{2}}}}{2} \geq \sqrt{1-\epsilon} (3\epsilon+1)^{\frac{3}{2}} \Leftrightarrow \frac{\gamma \epsilon (4-3\gamma \epsilon)^{3}}{2} \geq (1-\epsilon)^{\frac{3}{2}}$ $\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 \leq \epsilon \leq \frac{1}{6} & \epsilon \right) \left(3\epsilon+1\right)^3 \Leftrightarrow -27\gamma^2\epsilon^4 + 108\gamma\epsilon^3 + \frac{64\epsilon}{\gamma} + 21\epsilon^2 - 102\epsilon & -6\epsilon - 1 \leq \epsilon \\ \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 \leq \frac{1}{1+\gamma} \\ \frac{1}{1+\gamma} & \gamma \leq \frac{1}{\epsilon} - 1. \end{cases} \text{ Also, as } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq \frac{1}{1+\gamma} < 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq \frac{1}{1+\gamma} < 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0 < \gamma \leq 1, \text{ the inequality can be further } 0$ $\frac{\partial h(\gamma)}{\partial \gamma} = 108\epsilon^3 - \frac{64\epsilon}{\gamma^2}$. When $\gamma \leq \sqrt{\frac{64\epsilon}{18\epsilon^3}}$, $\frac{\partial h(\gamma)}{\partial \gamma} \geq 0$, when $\gamma \geq \sqrt{\frac{64\epsilon}{18\epsilon^3}}$, $\frac{\partial h(\gamma)}{\partial \gamma} \leq 0$, therefore, the minimum of $h(\gamma)$ must be when $\dot{\gamma}$ is equal to the boundary point of the domain. When $\frac{1}{6} < \epsilon \le \frac{1}{2}$, $0 < \gamma \le 1$, since $h(0) \to \infty$, $h(1) = 108\epsilon^3 + 64\epsilon$, the minimum occurs at the boundary point $\gamma = 1, 108\gamma\epsilon^3 + \frac{64\epsilon}{2} - 162\epsilon^2 - 8\epsilon - 1 > 108\epsilon^3 + 56\epsilon - 162\epsilon^2 - 1.$ Theorem .10. $\sup_{P \in \mathcal{P}_U \cap \mathcal{P}_{\mu=0,\sigma=1}} QA(\epsilon, \gamma)$ is monotonic decreasing with respect to ϵ over the interval $[0, \frac{1}{1+\gamma}]$, when $0 \le \gamma \le 1$.

Let $g(\epsilon) = 108\epsilon^3 + 56\epsilon - 162\epsilon^2 - 1$. $g'(\epsilon) = 324\epsilon^2 - 324\epsilon + 56$, when $\epsilon \le \frac{2}{9}$, $g'(\epsilon) \ge 0$, when $\epsilon \le \frac{2}{9}$, $g'(\epsilon) \ge 0$, when $\epsilon \le \frac{2}{9}$, $g'(\epsilon) \ge 0$, the simplified inequality is satisfied. When $\epsilon \le \frac{2}{9}$, $\epsilon \le 0$, the satisfied inequality is satisfied. When $\epsilon \le 0$ is satisfied. When $\epsilon \le 0$ is satisfied inequality $\epsilon \le 0$. The proof is $\epsilon \le 0$, the samplified inequality $\epsilon \le 0$. The proof is satisfied inequality $\epsilon \le 0$. The proof is $\epsilon \le 0$, the samplified inequality $\epsilon \le 0$, the proof is satisfied. When $\epsilon \le 0$ is satisfied inequality $\epsilon \le 0$, the proof is $\epsilon \le 0$, the proof is satisfied. When $\epsilon \le 0$ is satisfied inequality $\epsilon \le 0$, the proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$. The proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$. The proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$. The proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$. The proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$. The proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$. The proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$. The proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$. The proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$. The proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$. The proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$. The proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$. The proof is $\epsilon \le 0$, the proof is $\epsilon \le 0$. The proof i

407

409

410

411

412

418

419

420

421

422

423

424

425 426

427

Let $nu(\epsilon) = -108\epsilon^4 + 54\epsilon^3 - 18\epsilon^2 + 7\epsilon + 1$, then $nu'(\epsilon) = -432\epsilon^3 + 162\epsilon^2 - 36\epsilon + 7$, $nu''(\epsilon) = -1296\epsilon^2 + 324\epsilon - 36 < 0$.

Since $nu'(\epsilon = \frac{1}{2}) = -\frac{49}{2} < 0$, $nu'(\epsilon) < 0$. Also, $nu(\epsilon = \frac{1}{2}) = 0$,

so $nu(\epsilon) \le 0$, the simplified inequality is also satisfied. As a result, the simplified inequality is also valid within the range of $\frac{1}{6} < \epsilon \le \frac{1}{1+\gamma}$, provided that $0 < \gamma \le 1$. Then, it validates $\frac{\partial \sup QA}{\partial \epsilon} \le 0$ for the same range of ϵ and ϵ .

The first and second formulae, when $\epsilon = \frac{1}{6}$, are all equal

to
$$\frac{1}{2} \left(\frac{\sqrt{\frac{\gamma}{4 - \frac{\gamma}{2}}}}{\sqrt{2}} + \sqrt{\frac{5}{3}} \right)$$
. It follows that $\sup \mathrm{QA}(\epsilon, \gamma)$ is continuous every $[0, \frac{1}{3}]$. Hence, $\frac{\partial \sup \mathrm{QA}}{\partial s} < 0$ holds for the entire

uous over $[0, \frac{1}{1+\gamma}]$. Hence, $\frac{\partial \sup QA}{\partial \epsilon} \leq 0$ holds for the entire range $0 \leq \epsilon \leq \frac{1}{1+\gamma}$, when $0 \leq \gamma \leq 1$, which leads to the assertion of this theorem.

Let \mathcal{P}_{Υ}^k denote the set of all continuous distributions whose moments, from the first to the kth, are all finite. For a right-skewed distribution, it suffices to consider the upper bound. The monotonicity of $\sup_{P \in \mathcal{P}_{\Upsilon}^2} \mathrm{QA}$ with respect to ϵ implies that the extent of any violations of the γ -orderliness, if $0 \leq \gamma \leq 1$, is bounded for any distribution with a finite second moment, e.g., for a right-skewed distribution in \mathcal{P}_{Υ}^2 , if $\exists 0 \leq \epsilon_1 \leq \epsilon_2 \leq \epsilon_3 \leq \frac{1}{1+\gamma}$, then $\mathrm{QA}_{\epsilon_2,\gamma} \geq \mathrm{QA}_{\epsilon_3,\gamma} \geq \mathrm{QA}_{\epsilon_1,\gamma}$, $\mathrm{QA}_{\epsilon_2,\gamma}$ will not be too far away from $\mathrm{QA}_{\epsilon_1,\gamma}$, since $\sup_{P \in \mathcal{P}_{\Upsilon}^2} \mathrm{QA}_{\epsilon_1,\gamma} > \sup_{P \in \mathcal{P}_{\Upsilon}^2} \mathrm{QA}_{\epsilon_2,\gamma} > \sup_{P \in \mathcal{P}_{\Upsilon}^2} \mathrm{QA}_{\epsilon_3,\gamma}$. Moreover, a stricter bound can be established for unimodal distributions. The violation of ν th γ -orderliness, when $\nu \geq 2$, is also bounded as it corresponds to the higher-order derivatives of the QA function with respect to ϵ .

Inequalities related to weighted averages

The bias bound of the ϵ -symmetric trimmed mean also exhibits monotonicity for $\mathcal{P}_U \cap \mathcal{P}_{\Upsilon}^2$, as proven in the SI Text by applying the formulae provided in Bernard et al.'s paper (2). So far, it appears clear that the bias of a reasonable estimator is closely related to its degree of robustness. In a right-skewed unimodal distribution, it is often observed that $\mu \geq \text{STM}_{\epsilon} \geq m$. Analogous to the γ -orderliness, the γ -trimming inequality for a right-skewed distribution is defined as $\forall 0 \leq \epsilon_1 \leq \epsilon_2 \leq \frac{1}{1+\gamma}, \text{TM}_{\epsilon_1,\gamma} \geq \text{TM}_{\epsilon_2,\gamma}$. While γ -orderliness is a sufficient condition for the γ -trimming inequality, as proven in the SI Text, it is not necessary.

Theorem .11. 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 .

473 Proof. According to the definition of the γ -trimming in474 equality: $\frac{1}{1-\epsilon-\gamma\epsilon+2\delta}\int_{\gamma\epsilon-\delta}^{1-\epsilon+\delta}Q(u)\,du\geq\frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon}Q(u)\,du$,
475 where δ is an infinitesimal positive quantity. Subse476 quently, rewriting the inequality gives $\int_{\gamma\epsilon-\delta}^{1-\epsilon+\delta}Q(u)\,du$ 477 $\frac{1-\epsilon-\gamma\epsilon+2\delta}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon}Q(u)\,du\geq0$ \Leftrightarrow $\int_{1-\epsilon}^{1-\epsilon+\delta}Q(u)\,du$ +
478 $\int_{\gamma\epsilon-\delta}^{\gamma\epsilon}Q(u)\,du-\frac{2\delta}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon}Q(u)\,du\geq0$. Since δ \to 479 $0^+, \frac{1}{2\delta}\left(\int_{1-\epsilon}^{1-\epsilon+\delta}Q(u)\,du+\int_{\gamma\epsilon-\delta}^{\gamma\epsilon}Q(u)\,du\right)=\frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2}\geq$ 480 $\frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon}Q(u)\,du$, the proof is complete.

An analogous result can be obtained in the following theorem.

Theorem .12. 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 $0 \le \gamma \le 1$. If assuming γ -orderliness, the inequality is valid for any non-negative γ .

Proof. According to Theorem .11, $\frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2} \geq \frac{1}{1-\epsilon-\gamma\epsilon} \int_{\gamma\epsilon}^{1-\epsilon} Q\left(u\right) du \quad \Leftrightarrow \quad \gamma\epsilon\left(Q\left(\gamma\epsilon\right)+Q\left(1-\epsilon\right)\right) \quad \geq \\ \left(\frac{2\gamma\epsilon}{1-\epsilon-\gamma\epsilon}\right) \int_{\gamma\epsilon}^{1-\epsilon} Q\left(u\right) du. \qquad \text{Then, if } 0 \leq \gamma \leq \\ 1, \left(1-\frac{1}{1-\epsilon-\gamma\epsilon}\right) \int_{\gamma\epsilon}^{1-\epsilon} Q\left(u\right) du + \gamma\epsilon\left(Q\left(\gamma\epsilon\right)+Q\left(1-\epsilon\right)\right) \quad \geq \\ 0 \Rightarrow \int_{\gamma\epsilon}^{1-\epsilon} Q\left(u\right) du + \gamma\epsilon Q\left(\gamma\epsilon\right) + \epsilon Q\left(1-\epsilon\right) \geq \int_{\gamma\epsilon}^{1-\epsilon} Q\left(u\right) du + \gamma\epsilon\left(Q\left(\gamma\epsilon\right)+Q\left(1-\epsilon\right)\right) \geq \\ \int_{\gamma\epsilon}^{1-\epsilon} Q\left(u\right) du + \gamma\epsilon Q\left(\gamma\epsilon\right) + \epsilon Q\left(1-\epsilon\right) \geq \int_{\gamma\epsilon}^{1-\epsilon} Q\left(u\right) du + \gamma\epsilon Q\left(\gamma\epsilon\right) + Q\left(1-\epsilon\right) \leq C \left(1-\epsilon\right) \leq C \left(1-\epsilon\right) + C \left(1-\epsilon\right) \leq C \left(1-\epsilon\right) + C \left(1-\epsilon\right) + C \left(1-\epsilon\right) \leq C \left(1-\epsilon\right) + C \left(1-\epsilon\right) + C \left(1-\epsilon\right) \leq C \left(1-\epsilon\right) + C \left(1-\epsilon\right$

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

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

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 .13. 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 , given that $0 \le \gamma \le 1$.

Proof. From the definitions of BWM and WM, the statement necessitates $\int_{\gamma_{\epsilon}}^{1-\epsilon}Q\left(u\right)du+\gamma\epsilon Q\left(\gamma\epsilon\right)+\epsilon Q\left(1-\epsilon\right)\geq \int_{\gamma_{\epsilon}}^{1-\epsilon}Q\left(u\right)du+\int_{\gamma_{\epsilon}}^{2\gamma\epsilon}Q\left(u\right)du+\int_{1-2\epsilon}^{1-\epsilon}Q\left(u\right)du\Leftrightarrow \gamma\epsilon Q\left(\gamma\epsilon\right)+\epsilon Q\left(1-\epsilon\right)\geq \int_{\gamma_{\epsilon}}^{2\gamma\epsilon}Q\left(u\right)du+\int_{1-2\epsilon}^{1-\epsilon}Q\left(u\right)du.$ Define WMl(x) = Q(\gamma\epsilon) and BWMl(x) = Q(x). In both functions, the interval for x is specified as $[\gamma\epsilon,2\gamma\epsilon]$. Then, define WMu(y) = Q(1-\epsilon) and BWMu(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. WMl(x) + WMu(y(x)) = Q(\gamma\epsilon)+Q(1-\epsilon)\) bigodifferes BWMl(x) + BWMu(y(x)) = Q(x)+Q(1-\frac{x}{\gamma}) is valid for all $x\in[\gamma\epsilon,2\gamma\epsilon]$, according to the definition of γ -orderliness. Integration of the left side yields, $\int_{\gamma\epsilon}^{2\gamma\epsilon}\left(WMl(u)+WMu(y(u))\right)du=\int_{\gamma\epsilon}^{2\gamma\epsilon}Q\left(\gamma\epsilon\right)du+\int_{y(\gamma\epsilon)}^{y(2\gamma\epsilon)}Q\left(1-\epsilon\right)du=\int_{\gamma\epsilon}^{2\gamma\epsilon}Q\left(\gamma\epsilon\right)du+\int_{1-2\epsilon}^{1-2\epsilon}Q\left(1-\epsilon\right)du=\gamma\epsilon Q\left(\gamma\epsilon\right)+\epsilon Q\left(1-\epsilon\right),$ while integration of the right side

yields $\int_{\gamma_{\epsilon}}^{2\gamma_{\epsilon}} (BWMl(x) + BWMu(y(x))) dx = \int_{\gamma_{\epsilon}}^{2\gamma_{\epsilon}} Q(u) du +$ $\int_{\gamma\epsilon}^{2\gamma\epsilon}Q\left(1-\frac{x}{\gamma}\right)dx=\int_{\gamma\epsilon}^{2\gamma\epsilon}Q\left(u\right)du+\int_{1-2\epsilon}^{1-\epsilon}Q\left(u\right)du,\text{ which are the left and right sides of the desired inequality. Given that the }$ upper limits and lower limits of the integrations are different for each term, the condition $0 \le \gamma \le 1$ is necessary for the desired inequality to be valid.

524

528

From the second γ -orderliness, $\frac{\partial^2 QA}{\partial^2 \epsilon} \geq 0 \Rightarrow \forall 0 \leq a \leq 2a \leq 3a \leq \frac{1}{1+\gamma}, \frac{1}{a} \left(\frac{(QA(3a,\gamma)-QA(2a,\gamma))}{a} - \frac{(QA(2a,\gamma)-QA(a,\gamma))}{a} \right) \geq 0 \Rightarrow \text{if } 0 \leq \gamma \leq 1, \ \mathcal{B}_i - 2\mathcal{B}_{i+1} + \mathcal{B}_{i+2} \geq 0. \ \text{SM}_{\epsilon} \ \text{can thus}$ 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 ν th γ -orderliness, the recurrence relation of the derivatives naturally produces the alternating binomial coefficients,

$$(-1)^{\nu} \frac{\partial^{\nu} QA}{\partial \epsilon^{\nu}} \geq 0 \Rightarrow \forall 0 \leq a \leq \ldots \leq (\nu+1)a \leq \frac{1}{1+\gamma} \text{ at the end of the blocks if } p > \frac{1}{1+\gamma}. \text{ The sequence of denoting whether each block's quantile is weighted or not weighted is:} \\ \frac{(-1)^{\nu}}{a} \begin{pmatrix} \frac{QA(\nu a + a, \gamma) \cdot \cdot \cdot}{a} - \frac{\cdot \cdot \cdot QA(2a, \gamma)}{a} - \frac{QA(\nu a, \gamma) \cdot \cdot \cdot}{a} - \frac{\cdot \cdot \cdot \cdot QA(a, \gamma)}{a} \end{pmatrix} \xrightarrow{a} \text{ the end of the blocks if } p > \frac{1}{1+\gamma}. \text{ The sequence of denoting whether each block's quantile is weighted or not weighted is:} \\ 0, 1, 0, 1, 0, 1, \dots. \text{ Let the sequence be denoted by } a_{SQM_{\epsilon, \gamma}}(j), \\ 0, 1, 0, 1, 0, 1, \dots. \text{ Let the sequence of denoting whether each block's quantile is weighted or not weighted is:} \\ 0, 1, 0, 1, 0, 1, \dots. \text{ Let the sequence be denoted by } a_{SQM_{\epsilon, \gamma}}(j), \\ 0, 1, 0, 1, 0, 1, \dots. \text{ Let the sequence of denoting whether each block's quantile is weighted or not weighted is:} \\ 0, 1, 0, 1, 0, 1, \dots. \text{ Let the sequence of denoting whether each block's quantile is weighted or not weighted is:} \\ 0, 1, 0, 1, 0, 1, \dots. \text{ Let the sequence of denoting whether each block's quantile is weighted or not weighted is:} \\ 0, 1, 0, 1, 0, 1, \dots. \text{ Let the sequence of denoting whether each block's quantile is weighted or not weighted or not weighted is:} \\ 0, 1, 0, 1, 0, 1, \dots. \text{ Let the sequence is } a_{SQM_{\epsilon,\gamma}}(j) = j \text{ mod 2. If pairing all blocks in } BM_{\nu=2,\epsilon,\gamma}$$
 and all quantiles in $SQM_{\epsilon,\gamma}$ there are two possible pairings of $a_{BM_{\nu=2},\epsilon,\gamma}(j) = a_{SQM_{\epsilon,\gamma}}(j)$. One pairing occurs when $a_{BM_{\nu=2,\epsilon,\gamma}}(j) = a_{SQM_{\epsilon,\gamma}}(j)$ and $a_{SQM_{\epsilon,\gamma}}(j) = a_{SQM_{\epsilon,\gamma}}(j)$ are two possible pairings of $a_{BM_{\nu=2,\epsilon,\gamma}}(j) = a_{SQM_{\epsilon,\gamma}}(j)$ and $a_{SQM_{\epsilon,\gamma}}(j) = a_{SQM_{\epsilon,\gamma}}(j)$ are two possible pairings of $a_{BM_{\nu=2,\epsilon,\gamma}}(j) = a_{SQM_{\epsilon,\gamma}}(j)$ and $a_{SQM_{\epsilon,\gamma}}(j) = a_{SQM_{\epsilon,\gamma}}(j)$ are two possible pairings of $a_{BM_{\nu=2,\epsilon,\gamma}}(j) = a_{SQM_{\epsilon,\gamma}}(j)$ and $a_{SQM_{\epsilon,\gamma}}(j) = a_{SQM_{\epsilon,\gamma}}(j)$ and $a_{SQM_{\epsilon,\gamma}}(j) = a_{SQM_{\epsilon,\gamma}}(j)$ and $a_{SQM_{\epsilon,\gamma}}(j) = a_{SQM_{\epsilon,\gamma}}(j)$ and $a_{SQM_{\epsilon,\gamma}}(j) = a_{SQM_{\epsilon,\gamma}}(j)$ and $a_{SQM_$

Based on the ν th orderliness, the ϵ, γ -binomial mean is intro-

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

where $\mathfrak{B}_{i_j} = \sum_{l=n\gamma \epsilon(j+(i-1)(\nu+1)+1)}^{n\epsilon(j+(i-1)(\nu+1)+1)} (X_l + X_{n-l+1})$. If ν is not indicated, it defaults to $\nu=3$. As the alternating sum of binomial coefficients equals zero, when $\nu \ll \epsilon^{-1}$, $\epsilon \to 0$, BM $\to \mu$. The solutions for the continuity of the breakdown point is the same as that in SM and not repeated here. The equality $BM_{\nu=1,\epsilon} = BWM_{\epsilon}$ holds. Similarly, $BM_{\nu=2,\epsilon} = SM_{\epsilon,b=3}$, when $\gamma = 1$ and their respective ϵ s are identical. Interestingly, the biases of the $\mathrm{SM}_{\epsilon=\frac{1}{q},b=3}$ and the $\mathrm{WM}_{\epsilon=\frac{1}{q}}$ 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. If single quantiles are used, based on the second γ -orderliness, the stratified quantile mean can be defined as

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

 $SQM_{\epsilon=\frac{1}{2}}$ is the Tukey's midhinge (34). 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, $\frac{1}{6}$ mod $4 \neq 0$, 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 .14. 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 , provided that

537

538

540

543

544

545

546

547

548

549

550

551

552

553

554

557

558

559

560

561

562

563

565

566

567

568

569

570

571

572

573

574

575

576

577

579

580

581

582

583

Proof. For simplicity, suppose the order statistics of the sample are distributed into $\epsilon^{-1} \in \mathbb{N}$ blocks in the computation of both $SQM_{\epsilon,\gamma}$ and $BM_{\nu=2,\epsilon,\gamma}$. 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 $a_{\mathrm{BM}_{\nu=2,\epsilon,\gamma}}(j) = \left| \frac{j \bmod 3}{2} \right|$. Similarly, the computation of $\mathrm{SQM}_{\epsilon,\gamma}$ can be seen as positioning quantiles (p) at the beginning of the blocks if 0 , andat 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 $a_{\text{SQM}_{\epsilon,\gamma}}(j)$, the formula of the sequence is $a_{\mathrm{SQM}_{\epsilon,\gamma}}(j) = j \mod 2$. If pairare two possible pairings of $a_{\text{BM}_{\nu=2}}(j)$ and $a_{\text{SQM}_{\epsilon,\gamma}}(j)$. One pairing occurs when $a_{\mathrm{BM}_{\nu=2,\epsilon,\gamma}}(j) = a_{\mathrm{SQM}_{\epsilon,\gamma}}(j) = 1$, while the $\geq 0 \Leftrightarrow \frac{(-1)^{\nu}}{a^{\nu}} \left(\sum_{j=0}^{\nu} (-1)^{j} {\nu \choose j} \operatorname{QA} \left((\nu - j + 1) a, \gamma \right) \right) \geq 0 \text{ ther involves the sequence } 0, 1, 0 \text{ from } a_{\operatorname{BM}_{\nu=2,\epsilon,\gamma}}(j) = 1, \text{ while the with } 1, 0, 1 \text{ from } a_{\operatorname{SQM}_{\epsilon,\gamma}}(j) = 1, \text{ while the least sequence } 1, 1, 0, 1 \text{ from } a_{\operatorname{BM}_{\nu=2,\epsilon,\gamma}}(j) = 1, \text{ while the least sequence } 1, 1, 0, 1 \text{ from } a_{\operatorname{SQM}_{\epsilon,\gamma}}(j) = 1, \text{ while the least sequence } 1, 1, 0, 1,$ with 1,0,1 from $a_{\text{SQM}_{\epsilon,\gamma}}(j)$. By leveraging the same principle as Theorem .13 and the second γ -orderliness (replacing the two \Rightarrow if $0 \le \gamma \le 1$, $\sum_{i=0}^{\nu} (-1)^j \binom{\nu}{j} \mathcal{B}_{i+j} \ge 0$ quantile averages with one quantile average between them), the desired result follows. the desired result follows.

> The biases of $SQM_{\epsilon=\frac{1}{2}}$, which is based on the second orderliness with a quantile approach, are notably similar to those of $BM_{\nu=3,\epsilon=\frac{1}{9}}$, which is based on the third orderliness with a block approach, in common asymmetric unimodal distributions (Figure ??).

Hodges-Lehmann inequality and U-orderliness

The Hodges-Lehmann estimator stands out as a very unique robust location estimator due to its definition being substantially dissimilar from conventional symmetric weighted averages. In their landmark paper, Estimates of location based on rank tests, Hodges and Lehmann (8) proposed two methods to compute the H-L estimator: the Wilcoxon score R-estimator and the median of pairwise means, with time complexities of O(nlog(n)) and $O(n^2)$, respectively. The Wilcoxon score R-estimator is an estimator based on signed-rank test, or R-estimator, (8) and was later independently discovered by Sen (35, 36). However, the median of pairwise means is a generalized L-statistic and a trimmed U-statistic, as classified by Serfling in his novel conceptualized study in 1984 (37). Serfling further advanced the understanding by generalizing the H-L kernel as $hl_k = \frac{1}{k} \sum_{i=1}^k x_i$, where $k \in \mathbb{N}$ (37). Here,

the H-L kernel as
$$hl_k = \frac{1}{k} \sum_{i=1}^k x_i$$
, where $k \in \mathbb{N}$ (37). He the weighted H-L kernel is defined as $whl_k = \frac{\sum_{i=1}^k x_i \mathbf{w}_i}{\sum_{i=1}^k \mathbf{w}_i}$.

By using the whl_k kernel and the L-estimator, it is now clear that the Hodges-Lehmann estimator is an LL-statistic, the definition of which is provided as follows:

$$LL_{k,\epsilon,\gamma,n} := L_{\epsilon_0,\gamma,n} \left(\left(whl_k \left(X_{N_1}, \cdots, X_{N_k} \right) \right)_{N=1}^{\binom{n}{k}} \right),$$

where $L_{\epsilon_0,\gamma,n}(Y)$ denotes the L-estimator with the sequence $(whl_k(X_{N_1},\cdots,X_{N_k}))_{N=1}^{\binom{n}{k}}$ as an input, the upper breakdown

531

532

533

point of the L-estimator is ϵ_0 , the lower breakdown point is 587 $\gamma \epsilon_0$. The upper asymptotic breakdown point of $LL_{k,\epsilon,\gamma}$ is 588 $\epsilon = 1 - (1 - \epsilon_0)^{\frac{1}{k}}$, as proven in another relevant paper. There 589 are two ways to adjust the breakdown point: either by setting 590 k as a constant and adjusting ϵ_0 , or by setting ϵ_0 as a constant 591 and adjusting k. The bootstrap method can be applied to 592 ensure the continuity of k, also making the breakdown point 593 continuous. Specifically, let the bootstrap size be denoted 594 by b, then first sampling the original sample (1 - k + |k|)b595 times with each sample size of |k|, and then subsequently 596 sampling $(1 - \lceil k \rceil + k)b$ times with each sample size of $\lceil k \rceil$. 597 The corresponding kernels are computed separately, and the 598 pooled sequence is used as an input for the L-estimator. Indeed, 599 for any finite sample, when k = n, the whl_k kernel distribution becomes a single point, $\frac{\sum_{i=1}^{n} X_{i} \mathbf{w}_{i}}{\sum_{i=1}^{n} \mathbf{w}_{i}}$, since $\binom{n}{n} = 1$, when k = n, there is only one possible combination. When $\mathbf{w}_{i} = 1$, the minimum of the hl_{k} kernel distribution is $\frac{1}{k} \sum_{i=1}^{k} X_{i}$, due 602 603 to the property of order statistics. The maximum of it is 604 $\frac{1}{k}\sum_{i=1}^{k}X_{n-i+1}$. The monotonicity of the order statistics 605 implies the monotonicity of the extrema with respect to k, i.e., 606 the support of the hl_k kernel distribution shrinks monotonically. 607 For unequal \mathbf{w}_i s, the shrinkage of the support of the whl_k 608 kernel distribution might not be strictly monotonic, but the general trend is the same if $\mathbf{w}_i \geq 0$. That means all LL-610 statistics converges to the same point as $k \to n$. 611

Data Availability. Data for Figure ?? are given in SI Dataset 612 S1. All codes have been deposited in GitHub. 613

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

614

615

616 617

618

619 620

621

622

623

624 625

626

627 628

629

630

631

632

633

636 637

638

639

640 641

642

643

645

653

654

655

- 1. CF Gauss, Theoria combinationis observationum erroribus minimis obnoxiae. (Henricus Dieterich), (1823),
- C Bernard, R Kazzi, S Vanduffel, Range value-at-risk bounds for unimodal distributions under partial information. Insur. Math. Econ. 94, 9-24 (2020).
- P Daniell, Observations weighted according to order. Am. J. Math. 42, 222-236 (1920).
- JW Tukey. A survey of sampling from contaminated distributions in Contributions to probability and statistics. (Stanford University Press), pp. 448-485 (1960).
- WJ Dixon, Simplified Estimation from Censored Normal Samples, The Annals Math. Stat. 31, 385 - 391 (1960).
- K Danielak, T Rychlik, Theory & methods: Exact bounds for the bias of trimmed means. Aust. & New Zealand J. Stat. 45, 83-96 (2003).
- M Bieniek, Comparison of the bias of trimmed and winsorized means. Commun. Stat. Methods 45 6641-6650 (2016)
- J Hodges Jr, E Lehmann, Estimates of location based on rank tests. The Annals Math. Stat 34 598-611 (1963)
- F Wilcoxon, Individual comparisons by ranking methods. Biom. Bull. 1, 80-83 (1945).
- PJ Huber, Robust estimation of a location parameter. Ann. Math. Stat. 35, 73-101 (1964).
- Q Sun, WX Zhou, J Fan, Adaptive huber regression. J. Am. Stat. Assoc. 115, 254-265 (2020). 634 635
 - T Mathieu, Concentration study of m-estimators using the influence function. Electron. J. Stat 16. 3695-3750 (2022)
 - AS Nemirovskij, DB Yudin, Problem complexity and method efficiency in optimization. (Wiley-Interscience), (1983).
 - D Hsu, S Sabato, Heavy-tailed regression with a generalized median-of-means in International Conference on Machine Learning. (PMLR), pp. 37-45 (2014).
 - L Devroye, M Lerasle, G Lugosi, RI Oliveira, Sub-gaussian mean estimators. The Annals Stat. 44, 2695-2725 (2016).
 - P Laforgue, S Clémençon, P Bertail, On medians of (randomized) pairwise means in Interna tional Conference on Machine Learning. (PMLR), pp. 1272-1281 (2019).
 - G LECUÉ, M LERASLE, Robust machine learning by median-of-means: Theory and practice The Annals Stat. 48, 906-931 (2020).
- 647 B Efron, Bootstrap methods: Another look at the jackknife. The Annals Stat. 7, 1-26 (1979)
 - PJ Bickel, DA Freedman, Some asymptotic theory for the bootstrap. The annals statistics 9, 1196-1217 (1981).
- PJ Bickel, DA Freedman, Asymptotic normality and the bootstrap in stratified sampling. The 651 annals statistics 12, 470-482 (1984).
- R Helmers, P Janssen, N Veraverbeke, Bootstrapping U-quantiles. (CWI. Department of 652 Operations Research, Statistics, and System Theory [BS]), (1990).
 - J Neyman, On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. J. Royal Stat. Soc. 97, 558-606 (1934).

23. G McIntyre, A method for unbiased selective sampling, using ranked sets. Aust. journal agricultural research 3, 385-390 (1952).

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

- CM Stein, Efficient nonparametric testing and estimation in Proceedings of the third Berkeley symposium on mathematical statistics and probability, Vol. 1, pp. 187-195 (1956).
- 25. P Bickel, CA Klaassen, Y Ritov, JA Wellner, Efficient and adaptive estimation for semiparametric models. (Springer) Vol. 4, (1993).
- JT Runnenburg, Mean, median, mode. Stat. Neerlandica 32, 73-79 (1978). 26
- 27. Wy Zwet, Mean, median, mode ii. Stat. Neerlandica 33, 1-5 (1979).
- H Rietz, On certain properties of frequency distributions of the powers and roots of the variates of a given distribution. Proc. Natl. Acad. Sci. 13, 817-820 (1927).
- WR van Zwet, Convex Transformations of Random Variables: Nebst Stellingen. (1964)
- K Pearson, X. contributions to the mathematical theory of evolution.—ii. skew variation in homogeneous material. Philos. Transactions Royal Soc. London.(A.) 186, 343-414 (1895).
- AL Bowley, Elements of statistics. (King) No. 8, (1926).
- RA Groeneveld, G Meeden, Measuring skewness and kurtosis. J. Royal Stat. Soc. Ser. D (The Stat. 33, 391-399 (1984).
- 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)
- JW Tukey, Exploratory data analysis. (Reading, MA) Vol. 2, (1977).
- PK Sen, On the estimation of relative potency in dilution (-direct) assays by distribution-free methods. Biometrics pp. 532-552 (1963).
- M Ghosh, MJ Schell, PK Sen, A conversation with pranab kumar sen. Stat. Sci. pp. 548-564 (2008).
- 37. RJ Serfling, Generalized I-, m-, and r-statistics. The Annals Stat. 12, 76-86 (1984).

8 | Lee