PNAS
www.pnas.org

Supporting Information for

Robust estimations based on distribution structures: Moments

Tuobang Li.
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

Theorem B.3. $\psi_{\mathbf{k}}\left(x_1=\lambda x_1+\mu, \cdots, x_{\mathbf{k}}=\lambda x_{\mathbf{k}}+\mu\right)=\lambda^{\mathbf{k}} \psi_{\mathbf{k}}\left(x_1, \cdots, x_{\mathbf{k}}\right)$.

Proof. $\psi_{\mathbf{k}}$ can be divided into $\mathbf{k}$ groups. From 1st to $\mathbf{k}-1$th group, the gth group has $\binom{\mathbf{k}}{g}\binom{g}{1}$ terms having the form $(-1)^{g+1} \frac{1}{\mathbf{k}-g+1} x_{i_1}^{\mathbf{k}-g+1} x_{i_2} \ldots x_{i_g}$. The final $\mathbf{k}$th group is the term $(-1)^{\mathbf{k}-1}(\mathbf{k}-1) x_1 \cdots x_{\mathbf{k}}$.

The first choice is letting $x_{i_1}=x_1$, $\mathbf{k} \neq g$, the gth group of $\psi_{\mathbf{k}}$ has $\binom{\mathbf{k}-l}{g-l}$ terms having the form $(-1)^{g+1} \frac{1}{\mathbf{k}-g+1} x_1^{\mathbf{k}-g+1} x_2 \cdots x_l x_{i_1} \cdots x_{i_{g-l}}$, where $x_1,\ x_2, \cdots, x_l$ are fixed, $x_{i_1}, \cdots, x_{i_{g-l}}$ are selected such that $i_1, \cdots,\ i_{g-l} \neq 1,2, \cdots, l$ and $i_1 \neq \ldots \neq i_{g-l}$. Define another function $\Psi_{\mathbf{k}}\left(x_1, x_2, \cdots, x_l, x_{i_1}, \cdots, x_{i_{g-l}}\right)=\left(\lambda x_1+\mu\right)^{\mathbf{k}-g+1}\left(\lambda x_2+\mu\right) \cdots\left(\lambda x_l+\mu\right)\left(\lambda x_{i_1}+\mu\right) \cdots\left(\lambda x_{i_{g-l}}+\mu\right)$, the first group of $\Psi_{\mathbf{k}}$ is $\lambda^{\mathbf{k}} x_1 \cdots x_l x_{i_1} \cdots x_{i_{g-l}}$, the hth group of $\Psi_{\mathbf{k}}$, $h>1$, has $\binom{\mathbf{k}-g+1}{\mathbf{k}-h-l+2}$ terms having the form $\lambda^{\mathbf{k}-h+1} \mu^{h-1} x_1^{\mathbf{k}-h-l+2} x_2 \cdots x_l$. Transforming $\psi_{\mathbf{k}}$ by $\Psi_{\mathbf{k}}$, then combing all terms with $\lambda^{\mathbf{k}-h+1} \mu^{h-1} x_1^{\mathbf{k}-h-l+2} x_2 \cdots x_l$, $\mathbf{k}-h-l+2>1$, the summed coefficient is $S1_l=\sum_{g=l}^{h+l-1}(-1)^{g+1} \frac{1}{\mathbf{k}-g+1}\binom{\mathbf{k}-g+1}{\mathbf{k}-h-l+2}\binom{\mathbf{k}-l}{g-l}=\sum_{g=l}^{h+l-1}(-1)^{g+1} \frac{(\mathbf{k}-l)!}{(h+l-g-1)!(\mathbf{k}-h-l+2)!(g-l)!}=0$, since the summation is starting from l, ending at $h+l-1$, the first term includes the factor $g-l=0$, the final term includes the factor $h+l-g-1=0$, the terms in the middle are also zero due to the factorial property.

Another possible choice is letting one of $x_{i_2} \ldots x_{i_g}$ equal to x_1, the gth group of $\psi_{\mathbf{k}}$ has $(\mathbf{k}-h)\binom{h-1}{g-\mathbf{k}+h-1}$ terms having the form $(-1)^{g+1} \frac{1}{\mathbf{k}-g+1} x_1 x_2 \ldots x_j^{\mathbf{k}-g+1} \ldots x_{\mathbf{k}-h+1} x_{i_1} \ldots x_{i_{g-\mathbf{k}+h-1}}$, provided that $\mathbf{k} \neq g$, $2 \leq j \leq \mathbf{k}-h+1$, where $x_1, \ldots, x_{\mathbf{k}-h+1}$ are fixed, $x_j^{\mathbf{k}-g+1}$ and $x_{i_1}, \cdots, x_{i_{g-\mathbf{k}+h-1}}$ are selected such that $i_1, \cdots,\ i_{g-\mathbf{k}+h-1} \neq 1,2, \cdots, \mathbf{k}-h+1$ and $i_1 \neq \ldots \neq i_{g-\mathbf{k}+h-1}$. Transforming these terms by $\Psi_{\mathbf{k}}\left(x_1, x_2, \ldots, x_j, \ldots, x_{\mathbf{k}-h+1}, x_{i_1}, \ldots, x_{i_{g-\mathbf{k}+h-1}}\right)=$ $\left(\lambda x_1+\mu\right)\left(\lambda x_2+\mu\right) \cdots\left(\lambda x_j+\mu\right)^{\mathbf{k}-g+1} \cdots\left(\lambda x_{\mathbf{k}-h+1}+\mu\right)\left(\lambda x_{i_1}+\mu\right) \cdots\left(\lambda x_{i_{g-\mathbf{k}+h-1}}+\mu\right)$, then there are $\mathbf{k}-g+1$ terms having the form $\lambda^{\mathbf{k}-h+1} \mu^{h-1} x_1 x_2 \ldots x_{\mathbf{k}-h+1}$. Transforming the final $\mathbf{k}$th group of $\psi_{\mathbf{k}}$ by $\Psi_{\mathbf{k}}\left(x_1, \ldots, x_{\mathbf{k}}\right)=\left(\lambda x_1+\mu\right) \cdots\left(\lambda x_{\mathbf{k}}+\mu\right)$, then, there is one term having the form $(-1)^{\mathbf{k}-1}(\mathbf{k}-1) \lambda^{\mathbf{k}-h+1} \mu^{h-1} x_1 x_2 \ldots x_{\mathbf{k}-h+1}$. Another possible combination is that the gth group of $\psi_{\mathbf{k}}$ contains $(g-\mathbf{k}+h-1)\binom{h-1}{g-\mathbf{k}+h-1}$ terms having the form $(-1)^{g+1} \frac{1}{\mathbf{k}-g+1} x_1 x_2 \ldots x_{\mathbf{k}-h+1} x_{i_1} \ldots x_{i_j}^{\mathbf{k}-g+1} \ldots x_{i_{g-\mathbf{k}+h-1}}$. Transforming these terms by $\Psi_{\mathbf{k}}\left(x_1, x_2, \ldots, x_{\mathbf{k}-h+1}, x_{i_1}, \ldots, x_{i_j}, \ldots, x_{i_{g-\mathbf{k}+h-1}}\right)=$ $\left(\lambda x_1+\mu\right)\left(\lambda x_2+\mu\right) \cdots\left(\lambda x_{\mathbf{k}-h+1}+\mu\right)\left(\lambda x_{i_1}+\mu\right) \cdots\left(\lambda x_{i_j}+\mu\right)^{\mathbf{k}-g+1} \cdots\left(\lambda x_{i_{g-\mathbf{k}+h-1}}+\mu\right)$, then there is only one term having the form $\lambda^{\mathbf{k}-h+1} \mu^{h-1} x_1 x_2 \ldots x_{\mathbf{k}-h+1}$. The above summation $S1_l$ should also be included, i.e., $x_1^{\mathbf{k}-h-l+2}=x_1$, $\mathbf{k}=h+l-1$. So, combing all terms with $\lambda^{\mathbf{k}-h+1} \mu^{h-1} x_1 x_2 \ldots x_{\mathbf{k}-h+1}$, according to the binomial theorem, the summed coefficient is $S2_l=\sum_{g=\mathbf{k}-h+1}^{\mathbf{k}-1}(-1)^{g+1}\binom{h-1}{g-\mathbf{k}+h-1}\left(\mathbf{k}-h+1+\frac{g-\mathbf{k}+h-1}{\mathbf{k}-g+1}\right)+(-1)^{\mathbf{k}-1}(\mathbf{k}-1)=(\mathbf{k}-h+1) \sum_{g=\mathbf{k}-h+1}^{\mathbf{k}-1}(-1)^{g+1}\binom{h-1}{g-\mathbf{k}+h-1}+$ $\sum_{g=\mathbf{k}-h+1}^{\mathbf{k}-1}(-1)^{g+1}\binom{h-1}{g-\mathbf{k}+h-1}\left(\frac{g-\mathbf{k}+h-1}{\mathbf{k}-g+1}\right)+(-1)^{\mathbf{k}-1}(\mathbf{k}-1)=(-1)^{\mathbf{k}}(\mathbf{k}-h+1)+(h-2)(-1)^{\mathbf{k}}+(-1)^{\mathbf{k}-1}(\mathbf{k}-1)=0$. The summation identities required are $\sum_{g=\mathbf{k}-h+1}^{\mathbf{k}-1}(-1)^{g+1}\binom{h-1}{g-\mathbf{k}+h-1}=(-1)^{\mathbf{k}}$ and $\sum_{g=\mathbf{k}-h+1}^{\mathbf{k}-1}(-1)^{g+1}\binom{h-1}{g-\mathbf{k}+h-1}\left(\frac{g-\mathbf{k}+h-1}{\mathbf{k}-g+1}\right)=$ $(h-2)(-1)^{\mathbf{k}}$. These two summation identities are proven in Lemma B.4 and B.5. The result is the same if replacing x_1 with x_i, where i is from 2 to $\mathbf{k}$, and replacing x_l with other x_i. Thus, all terms including μ can be canceled out. The proof is complete by noticing that the remaining part is $\lambda^{\mathbf{k}} \psi_{\mathbf{k}}\left(x_1, \cdots, x_{\mathbf{k}}\right)$.

□

Lemma B.4. $\sum_{g=\mathbf{k}-h+1}^{\mathbf{k}-1}(-1)^{g+1}\binom{h-1}{g-\mathbf{k}+h-1}=(-1)^{\mathbf{k}}$.

Proof. Let $u=\mathbf{k}-h+1$, then the above identity becomes $\sum_{g=u}^{\mathbf{k}-1}(-1)^{g+1}\binom{\mathbf{k}-u}{g-u}=(-1)^{\mathbf{k}}$. Then, by deducing,

$$
\begin{aligned}
\sum_{g=u}^{\mathbf{k}-1}(-1)^{g+1}\binom{\mathbf{k}-u}{g-u} &= \sum_{i=0}^{\mathbf{k}-u-1}(-1)^{i+u+1}\binom{\mathbf{k}-u}{i} && \text{(Substitute } i=g-u\text{)} \\
&= (-1)^{\mathbf{k}+2}+\sum_{i=0}^{\mathbf{k}-u}(-1)^{i+u+1}\binom{\mathbf{k}-u}{i} \\
&= (-1)^{\mathbf{k}+2}+(-1)^{u+1} \sum_{i=0}^{\mathbf{k}-u}(-1)^{i}\binom{\mathbf{k}-u}{i} \\
&= (-1)^{\mathbf{k}} && \text{(Apply the alternating sum identity),}
\end{aligned}
$$

the proof is complete.

□

Lemma B.5. $\sum_{g=\mathbf{k}-h+1}^{\mathbf{k}-1}(-1)^{g+1}\binom{h-1}{g-\mathbf{k}+h-1}\left(\frac{g-\mathbf{k}+h-1}{\mathbf{k}-g+1}\right)=(h-2)(-1)^{\mathbf{k}}$.

Proof. Let $u = \mathbf{k} - h + 1$, then the above identity becomes $\sum_{g=u}^{\mathbf{k}-1} (-1)^{g+1} \binom{\mathbf{k}-u}{g-u-1} = (-1)^{\mathbf{k}} (\mathbf{k} - u - 1)$. Then by deducing,

$$\begin{aligned}
\sum_{g=u}^{\mathbf{k}-1}(-1)^{g+1}\binom{\mathbf{k}-u}{g-u-1} &= \sum_{i=-1}^{\mathbf{k}-u-2}(-1)^{u+i+2}\binom{\mathbf{k}-u}{i} && \text{(Substitute } i = g-u-1\text{)}\\
&= \sum_{i=0}^{\mathbf{k}-u}(-1)^{u+i+2}\binom{\mathbf{k}-u}{i} - \sum_{i=\mathbf{k}-u-1}^{\mathbf{k}-u}(-1)^{u+i+2}\binom{\mathbf{k}-u}{i} && \text{(Apply the alternating sum identity)}\\
&= -\sum_{i=\mathbf{k}-u-1}^{\mathbf{k}-u}(-1)^{u+i+2}\binom{\mathbf{k}-u}{i}\\
&= (-1)^{\mathbf{k}+2}\binom{\mathbf{k}-u}{\mathbf{k}-u-1} + (-1)^{\mathbf{k}+3}\binom{\mathbf{k}-u}{\mathbf{k}-u}\\
&= (-1)^{\mathbf{k}+2}(\mathbf{k}-u) + (-1)^{\mathbf{k}+3}\\
&= (-1)^{\mathbf{k}+2}(\mathbf{k}-u-1),
\end{aligned}$$

the proof is complete. □

Theorem F.1. *Given a U-statistic associated with a symmetric kernel of degree* $\mathbf{k}$. *Then, assuming that as* $n \to \infty$, $\mathbf{k}$ *is a constant, the upper breakdown point of the LU-statistic is* $1 - (1 - \epsilon_0)^{\frac{1}{\mathbf{k}}}$, *where* ϵ_0 *is the upper breakdown point of the corresponding LL-statistic.*

Proof. Suppose m arbitrary large contaminants are added to the sample. The fraction of bad values in the sample can be represented as $\epsilon_{U_{\mathbf{k}}} = \frac{m}{n+m}$, where n denotes the original number of data points that remain unaffected. In the kernel distribution, $\binom{n}{\mathbf{k}}$ out of a total of $\binom{n+m}{\mathbf{k}}$ points are not corrupted. Then, the breakdown can be avoided if the following inequality holds

$$\binom{n}{\mathbf{k}} > \left(\frac{1}{\epsilon_0} - 1\right) \times \left(\binom{n+m}{\mathbf{k}} - \binom{n}{\mathbf{k}}\right).$$

Since ϵ_0 is the upper breakdown point of the corresponding LL-statistic, $0 \leq \epsilon_0 \leq \frac{1}{1+\gamma}$,

$$\frac{1}{1-\epsilon_0} > \frac{\binom{n+m}{\mathbf{k}}}{\binom{n}{\mathbf{k}}} = \frac{(n+m)(n+m-1)\dots(n+m-\mathbf{k}+1)}{n(n-1)\dots(n-\mathbf{k}+1)}.$$

Assuming $n \to \infty$, $\mathbf{k}$ is a constant, $\lim_{n\to\infty}\left(\frac{n+m-\mathbf{k}+1}{n-\mathbf{k}+1}\right) = \frac{n+m}{n}$, then the above inequality does not hold when $\frac{n+m}{n} \geq \left(\frac{1}{1-\epsilon_0}\right)^{\frac{1}{\mathbf{k}}}$. So, the upper asymptotic breakdown point of the LU-statistic is $\epsilon_{U_{\mathbf{k}}} = \frac{m}{n+m} = 1 - \frac{n}{n+m} = 1 - (1-\epsilon_0)^{\frac{1}{\mathbf{k}}}$. □

$\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}}$ for the exponential distribution

For a continuous distribution, $\mathrm{TM}_{y,z}=\frac{\int_{F^{-1}(y)}^{F^{-1}(z)}xf(x)dx}{\int_{F^{-1}(y)}^{F^{-1}(z)}f(x)dx}$. For the exponential distribution, it is $\frac{\lambda(-y+(y-1)\ln(1-y)+z-(z-1)\ln(1-z))}{z-y}$. Then,

$$\begin{aligned}
\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}} &= \frac{1}{24}\Big(4\mathrm{TM}_{\frac{1}{24},\frac{2}{24}}-2\mathrm{TM}_{\frac{2}{24},\frac{3}{24}}+2\mathrm{TM}_{\frac{3}{24},\frac{4}{24}}+0\mathrm{TM}_{\frac{4}{24},\frac{5}{24}}+4\mathrm{TM}_{\frac{5}{24},\frac{6}{24}}-2\mathrm{TM}_{\frac{6}{24},\frac{7}{24}}+2\mathrm{TM}_{\frac{7}{24},\frac{8}{24}}\\
&\quad+0\mathrm{TM}_{\frac{8}{24},\frac{9}{24}}+4\mathrm{TM}_{\frac{9}{24},\frac{10}{24}}-2\mathrm{TM}_{\frac{10}{24},\frac{11}{24}}+2\mathrm{TM}_{\frac{11}{24},\frac{12}{24}}+2\mathrm{TM}_{\frac{12}{24},\frac{13}{24}}-2\mathrm{TM}_{\frac{13}{24},\frac{14}{24}}+4\mathrm{TM}_{\frac{14}{24},\frac{15}{24}}\\
&\quad+0\mathrm{TM}_{\frac{15}{24},\frac{16}{24}}+2\mathrm{TM}_{\frac{16}{24},\frac{17}{24}}-2\mathrm{TM}_{\frac{17}{24},\frac{18}{24}}+4\mathrm{TM}_{\frac{18}{24},\frac{19}{24}}+0\mathrm{TM}_{\frac{19}{24},\frac{20}{24}}+2\mathrm{TM}_{\frac{20}{24},\frac{21}{24}}-2\mathrm{TM}_{\frac{21}{24},\frac{22}{24}}+4\mathrm{TM}_{\frac{22}{24},\frac{23}{24}}\Big)\\
&= \frac{1}{24}\Big(4\lambda\left(1-22\ln\left(\frac{12}{11}\right)+23\ln\left(\frac{24}{23}\right)\right)-2\lambda\left(1-21\ln\left(\frac{8}{7}\right)+22\ln\left(\frac{12}{11}\right)\right)+2\lambda\left(1-20\ln\left(\frac{6}{5}\right)+21\ln\left(\frac{8}{7}\right)\right)\\
&\quad+4\lambda\left(1-18\ln\left(\frac{4}{3}\right)+19\ln\left(\frac{24}{19}\right)\right)-2\lambda\left(1+18\ln\left(\frac{4}{3}\right)-17\ln\left(\frac{24}{17}\right)\right)+2\lambda\left(1-16\ln\left(\frac{3}{2}\right)+17\ln\left(\frac{24}{17}\right)\right)\\
&\quad+4\lambda\left(1+15\ln\left(\frac{8}{5}\right)-14\ln\left(\frac{12}{7}\right)\right)-2\lambda\left(1+14\ln\left(\frac{12}{7}\right)-13\ln\left(\frac{24}{13}\right)\right)+2\lambda\left(1-12\ln(2)+13\ln\left(\frac{24}{13}\right)\right)\\
&\quad+2\lambda\left(1+\ln(4096)-11\ln\left(\frac{24}{11}\right)\right)-2\lambda\left(1-10\ln\left(\frac{12}{5}\right)+11\ln\left(\frac{24}{11}\right)\right)+4\lambda\left(1-9\ln\left(\frac{8}{3}\right)+10\ln\left(\frac{12}{5}\right)\right)\\
&\quad+2\lambda\left(1+8\ln(3)-7\ln\left(\frac{24}{7}\right)\right)-2\lambda\left(1-6\ln(4)+7\ln\left(\frac{24}{7}\right)\right)+4\lambda\left(1+\ln\left(\frac{3125}{1944}\right)\right)\\
&\quad+2\lambda\left(1+\ln\left(\frac{81}{32}\right)\right)-2\lambda\left(1+\ln\left(\frac{32}{9}\right)\right)+4\lambda(1+\ln(6))\Big)\\
&= \frac{2\lambda}{24}\Big(12-12\ln(2)+8\ln(3)+6\ln(4)+\ln(36)+\ln(4096)-16\ln\left(\frac{3}{2}\right)-54\ln\left(\frac{4}{3}\right)-18\ln\left(\frac{8}{3}\right)-20\ln\left(\frac{6}{5}\right)\\
&\quad+30\ln\left(\frac{8}{5}\right)+30\ln\left(\frac{12}{5}\right)+42\ln\left(\frac{8}{7}\right)-42\ln\left(\frac{12}{7}\right)-14\ln\left(\frac{24}{7}\right)-\ln\left(\frac{32}{9}\right)-66\ln\left(\frac{12}{11}\right)-22\ln\left(\frac{24}{11}\right)\\
&\quad+26\ln\left(\frac{24}{13}\right)+34\ln\left(\frac{24}{17}\right)+38\ln\left(\frac{24}{19}\right)+46\ln\left(\frac{24}{23}\right)+\ln\left(\frac{81}{32}\right)+2\ln\left(\frac{3125}{1944}\right)\Big)\\
&= \lambda\left(1+\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right).
\end{aligned}$$

Methods

A. d Value Calibration. Asymptotic d values for the invariant moments for the exponential distribution ($\lambda=1$) were approximated by a quasi-Monte Carlo study (1, 2). The study was conducted using the R programming language (version 4.3.1) with the following libraries: randtoolbox (3), Rcpp (4), Rfast (5), matrixStats (6), foreach (7), and doParallel (8). A large quasi-random sample was generated, with a sample size of approximately 1.8 million, from the exponential distribution. This sample was then quasi-subsampled about 1.8$\mathbf{k}$ million times to approximate the kernel distributions. Consequently, computations were made for the $\mathbf{k}$th moment ($\mathbf{k}m$), the symmetric weighted Hodges-Lehmann $\mathbf{k}$th moment (SWHL$\mathbf{k}m$), the median $\mathbf{k}$th moment ($m\mathbf{k}m$), and the corresponding quantiles. The d values of recombined/quantile moments were obtained by the formulae $d_{r\mathbf{k}m}=\frac{\mathbf{k}m-\mathrm{SWHL}\mathbf{k}m}{\mathrm{SWHL}\mathbf{k}m-m\mathbf{k}m}$ and $d_{q\mathbf{k}m}=\frac{\hat{F}_{n,\psi_{\mathbf{k}}}(\mathbf{k}m)-\hat{F}_{n,\psi_{\mathbf{k}}}(\mathrm{SWHL}\mathbf{k}m)}{\hat{F}_{n,\psi_{\mathbf{k}}}(\mathrm{SWHL}\mathbf{k}m)-\frac{1}{2}}$. The accuracy of the estimates was verified by comparing the quasi-bootstrap central moments to their asymptotic values, yielding errors of ≈ 0.0003, ≈ 0.001, and ≈ 0.03 for the second, third, and fourth central moments, respectively. The standard deviations of these central moments kernel distributions were 2.234, 9.627, and 60.064, respectively, resulting in standardized errors for the values that were all smaller than 0.001, thus ensuring the accuracy implied in the number of significant digits of the values in Table 1 in the Main Text.

For finite sample, the d values were estimated using 1000 pseudorandom samples with sample size $n=4096$ with a quasi-bootstrap size of 18000. To estimate the errors of d value estimations of recombined mean in this way, first consider the first order Taylor approximation of the d value function, $d=\frac{x_1-x_2}{x_2-x_3}\approx d^0+\frac{\partial d}{\partial x_1}x_1+\frac{\partial d}{\partial x_2}x_2+\frac{\partial d}{\partial x_3}x_3$. Then, by applying Bienaymé's identity, the variance of d can be approximated by $\sigma_d^2\approx\left|\frac{\partial d}{\partial x_1}\right|^2\sigma_{x_1}^2+\left|\frac{\partial d}{\partial x_2}\right|^2\sigma_{x_2}^2+\left|\frac{\partial d}{\partial x_3}\right|^2\sigma_{x_3}^2+2\left|\frac{\partial d}{\partial x_1}\right|\left|\frac{\partial d}{\partial x_2}\right|Cov\left(X_1,X_2\right)+2\left|\frac{\partial d}{\partial x_1}\right|\left|\frac{\partial d}{\partial x_3}\right|Cov\left(X_1,X_3\right)+2\left|\frac{\partial d}{\partial x_2}\right|\left|\frac{\partial d}{\partial x_3}\right|Cov\left(X_2,X_3\right)=\left|\frac{1}{x_2-x_3}\right|^2\sigma_{x_1}^2+\left|-\frac{x_1-x_2}{(x_2-x_3)^2}-\frac{1}{x_2-x_3}\right|^2\sigma_{x_2}^2+\left|\frac{x_1-x_2}{(x_2-x_3)^2}\right|^2\sigma_{x_3}^2+2\left|\frac{1}{x_2-x_3}\right|\left|-\frac{x_1-x_2}{(x_2-x_3)^2}-\frac{1}{x_2-x_3}\right|Cov\left(X_1,X_2\right)+2\left|\frac{1}{x_2-x_3}\right|\left|\frac{x_1-x_2}{(x_2-x_3)^2}\right|Cov\left(X_1,X_3\right)+2\left|-\frac{x_1-x_2}{(x_2-x_3)^2}-\frac{1}{x_2-x_3}\right|\left|\frac{x_1-x_2}{(x_2-x_3)^2}\right|Cov\left(X_2,X_3\right)$. Since for the recombined mean, $\sigma_{x_1}^2=0$, so, $\sigma_{d_{rm}}^2\approx\left|-\frac{x_1-x_2}{(x_2-x_3)^2}-\frac{1}{x_2-x_3}\right|^2\sigma_{x_2}^2+\left|\frac{x_1-x_2}{(x_2-x_3)^2}\right|^2\sigma_{x_3}^2+$

$2\left(-\frac{x_1-x_2}{(x_2-x_3)^2}-\frac{1}{x_2-x_3}\right)\left(\frac{x_1-x_2}{(x_2-x_3)^2}\right)Cov\left(X_2,X_3\right)$, where x_1 is the expected value, x_2 is the weighted L-statistic used, x_3 is the median. For quantile mean, since $\sigma_{x_3}^2=0$, $\sigma_{d_{qm}}^2 \approx \left|\frac{1}{x_2-x_3}\right|^2\sigma_{x_1}^2+\left|-\frac{x_1-x_2}{(x_2-x_3)^2}-\frac{1}{x_2-x_3}\right|^2\sigma_{x_2}^2+2\left(\frac{1}{x_2-x_3}\right)\left(-\frac{x_1-x_2}{(x_2-x_3)^2}-\frac{1}{x_2-x_3}\right)Cov\left(X_1,X_2\right)$, where x_1 is the percentile of the expected value, x_2 is the percentile of the weighted L-statistic used, x_3 is the percentile of median, $\frac{1}{2}$. Finally, the errors were estimated by the corresponding sample statistics. The results of error estimation were included in the SI Dataset S1.

B. ASAB, ASB, and SSE. The computations of ASABs for invariant central moments were described in the Main Text. ASBs are the same, besides under finite sample scenarios. The SSE was computed by approximating the sampling distribution with 1000 pseudorandom samples for $n=4096$ and 30 pseudorandom samples for $n=1.8\times10^6$. Common random numbers were used for better comparison. Analogous to the asymptotic bias, the scaled standard error can be standardized, averaged, and weighted. It should be noted that, in Table 1, for symmetric distributions, the generalized Gaussian, the standard errors were used for location and asymmetry estimators, since when the mean value is close to zero, the scaled standard error will approach infinity and therefore be too sensitive to small changes. The errors of ASB and SSE were estimated by $se\left(\bar{x}\right)=\frac{\sigma}{\sqrt{n}}\approx\frac{usb}{\sqrt{n}}$, $se\left(sd\right)\approx\frac{1}{2\sigma}se\left(var\right)=\sqrt{\frac{\mu_4}{4n\sigma^2}-\frac{n-3}{4n(n-1)}\sigma^2}\approx\sqrt{\frac{fm}{4nvar}-\frac{n-3}{4n(n-1)}var}$, where usb is unbiased standard deviation of the sampling distribution with normality assumption (9). The computational methods used for two-parameter distributions were identical. The computations of invariant moments were described in the Main Text. The results of error estimation were included in the SI Dataset S1.

C. The Impact of Bootstrap Size on Variance. The study of the impact of the bootstrap size, from $n=1.8\times10^2$ to $n=1.8\times10^4$, on the variance for the exponential distribution was done the same as above.

D. Comparisons to Unbiased Central Moments, M-Estimators, and Marks Percentile Estimator. Within the same kurtosis range and five two-parametric distributions as the above, algorithms for unbiased central moment estimation proposed by Gerlovina and Hubbard (10) were used for estimating unbiased central moments. Then, within the same kurtosis range and four two-parametric distributions (except the generalized Gaussian distribution, since the logarithmic function does not produce results for negative values), the percentile estimators were computed using the method proposed by Marks (2005) (11) (consistent for the Weibull distribution) and the parameter setting proposed by Boudt, Caliskan, and Croux (2011) (12). The robust M-estimators were also computed in the same way using the methods proposed by Huber (13) (consistent for the Gaussian distribution) and He and Fung (1999) (14) (consistent for the Weibull distribution). Bisection is used to find the solution of the key equation in (14), while the results from the percentile estimator were used as initial values (-0.3 and +0.3). The results of He and Fung M-Estimator and Marks Percentile Estimator were then transformed to the first four moments to compute ASABs, ASBs, and SSEs. The ASABs, ASBs, and SSEs of unbiased central moments and Huber M-estimator were processed similarly.

E. Maximum Asymptotic Biases. For simplicity, a brute force approach was used to estimate the maximum biases of SWHLMs and SWHL**k**ms for five unimodal distributions. From the minimum kurtosis, a wide range was set to roughly estimate the parameter ranges in which the maximum bias might occur (the corresponding maximum kurtoses are all larger than 500). Then, the parameter range was broken to 100 parts, the maximum among all estimates was determined to be very close to the true maximum. Pseudo-maximum bias was described in the Main Text.

The brute force approach is generally valid, i.e., the maximum is the global maximum, not local maximum, even when the the corresponding maximum kurtosis is finite. Because all five distributions here have the property that, as the kurtosis of the distribution increases to infinity, the standardized biases of SWHLMs approach zero.

For example, for the Perato distribution,

$$B_{\mathrm{Q}}(\epsilon,\alpha)=\frac{x_m\left(1-\epsilon\right)^{-\frac{1}{\alpha}}-\frac{\alpha x_m}{\alpha-1}}{\sqrt{\frac{\alpha {x_m}^2}{(1-\alpha)^2(\alpha-2)}}}.$$

$\lim_{\alpha\to2}B_{\mathrm{Q}}(\epsilon,\alpha)=\lim_{\alpha\to2}\frac{x_m(1-\epsilon)^{-\frac{1}{\alpha}}-\frac{\alpha x_m}{\alpha-1}}{\sqrt{\frac{\alpha {x_m}^2}{(1-\alpha)^2(\alpha-2)}}}=\lim_{\alpha\to2}\frac{(1-\epsilon)^{-\frac{1}{2}}-2}{\sqrt{\frac{2}{(-1)^2(2-2)}}}=0$. Since SWHLMs are quantile combinations, their standardized biases all approach zero.

In SMRM I, it is shown that 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. That means for the SWAs based on the orderliness, the distribution will follow the mean-SWA-median inequality as the skewness approaches infinity, and therefore as the kurtosis approaches infinity since they are correlated. Thus, proving the limits of the ratios between μ and σ, as well as m and σ is enough.

For example, for the Weibull distribution, the ratio of μ and σ is $\lim_{\alpha\to0}\frac{\Gamma\left(1+\frac{1}{\alpha}\right)}{\sqrt{\Gamma\left(\frac{\alpha+2}{\alpha}\right)}}=\lim_{\alpha\to0}\frac{\left(1+\frac{1}{\alpha}-1\right)!}{\sqrt{\left(\frac{\alpha+2}{\alpha}-1\right)!}}=\lim_{\alpha\to0}\frac{\left(\frac{1}{\alpha}\right)!}{\sqrt{\left(2\times\frac{1}{\alpha}\right)!}}=0$, the ratio of m and σ is $\lim_{\alpha\to0^+}\frac{\sqrt[\alpha]{\ln(2)}}{\sqrt{\Gamma\left(\frac{\alpha+2}{\alpha}\right)}}=\lim_{\alpha\to0^+}\frac{\sqrt[\alpha]{\ln(2)}}{\sqrt{\left(\frac{\alpha+2}{\alpha}-1\right)!}}=\lim_{x\to\infty}\frac{e^{x\ln(\ln(2))}}{\sqrt{(2x)!}}$, where $x=\frac{1}{\alpha}$. Applying Stirling's approximation for the factorial gives:

$$\lim_{x\to\infty}\frac{e^{x\ln(\ln(2))}}{\sqrt{(2x)!}}=\lim_{x\to\infty}\frac{e^{x\ln(\ln(2))}}{\sqrt{\left(\frac{2x}{e}\right)^{(2x)}\sqrt{2\pi(2x)}}}=\lim_{x\to\infty}\frac{e^{x\ln(\ln(2))}}{\sqrt{2}\sqrt[4]{\pi}\sqrt{2^{2x}e^{-2x}}\sqrt{x^{2x+\frac{1}{2}}}}=\lim_{x\to\infty}\frac{e^{x\ln(\ln(2))}}{\sqrt{2}\sqrt[4]{\pi}2^{x}e^{-x}\sqrt{x^{2x+\frac{1}{2}}}}=\lim_{x\to\infty}\frac{e^{x(\ln(\ln(2))-1)}}{2^{x}\sqrt{2}\sqrt[4]{\pi}x^{x}\sqrt{x}}.$$

Since $(\ln(\ln(2))-1)\approx -1.367$, the numerator goes to zero as $x\to\infty$. Obviously, the denominator is monotonic increasing and goes to infinity as $x\to\infty$, therefore, $\lim_{\alpha\to 0^+}\frac{\sqrt[\alpha]{\ln(2)}}{\sqrt{\Gamma\left(\frac{\alpha+2}{\alpha}\right)}}=0$.

Similarly, for the gamma distribution, the ratio of μ and σ is $\lim_{\alpha\to 0}\frac{\alpha}{\sqrt{\alpha}}=\lim_{\alpha\to 0}\frac{1}{\sqrt{\alpha}}=0$, the ratio of m and σ is $\lim_{\alpha\to 0}\frac{P^{-1}\left(\alpha,\frac{1}{2}\right)}{\sqrt{\alpha}}=0$ (15).

The lognormal distribution is the same, the ratio of μ and σ is $\lim_{\sigma\to\infty}\frac{e^{\mu+\frac{\sigma^2}{2}}}{\sqrt{\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}}}=\lim_{\sigma\to\infty}\frac{e^{\mu+\frac{\sigma^2}{2}}}{\sqrt{e^{2\mu+2\sigma^2}}}=\lim_{\sigma\to\infty}\frac{e^{\frac{\sigma^2}{2}}}{e^{\sigma^2}}=0$, the ratio of m and σ is $\lim_{\sigma\to\infty}\frac{e^{\mu}}{\sqrt{\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}}}=0$.

As demonstrated, the growth rate of the standard deviation greatly exceeds that of the mean and that of the median. This phenomenon is closely tied to the Taylor's law and is more widespread than these examples suggest.

F. Language Refinement and Mathematical Expressions. ChatGPT, an AI language model developed by OpenAI, was used to improve the grammatical accuracy of the manuscript. To deduce and verify complex mathematical expressions, both Wolfram Alpha and ChatGPT were utilized.

SI Dataset S1 (dataset_one.xlsx)

Raw data of Table 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. MC Dutang, Package 'randtoolbox' (2015).
4. D Eddelbuettel, R François, Rcpp: Seamless r and c++ integration. *J. statistical software* **40**, 1–18 (2011).
5. M Papadakis, et al., Package 'rfast' (2023).
6. H Bengtsson, et al., Package 'matrixstats' (2023).
7. S Weston, , et al., foreach: provides foreach looping construct (2019).
8. S Weston, R Calaway, Getting started with doparallel and foreach (2015).
9. CR Rao, *Linear statistical inference and its applications.* (John Wiley & Sons Inc), (1965).
10. I Gerlovina, AE Hubbard, Computer algebra and algorithms for unbiased moment estimation of arbitrary order. *Cogent mathematics & statistics* **6**, 1701917 (2019).
11. NB Marks, Estimation of weibull parameters from common percentiles. *J. applied Stat.* **32**, 17–24 (2005).
12. K Boudt, D Caliskan, C Croux, Robust explicit estimators of weibull parameters. *Metrika* **73**, 187–209 (2011).
13. PJ Huber, Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964).
14. X He, WK Fung, Method of medians for lifetime data with weibull models. *Stat. medicine* **18**, 1993–2009 (1999).
15. MA Chaudhry, SM Zubair, *On a class of incomplete gamma functions with applications.* (Chapman and Hall/CRC), (2001).