PNAS
www.pnas.org

Supporting Information for

Near-consistent robust estimations of moments for unimodal distributions

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

This PDF file includes:

Supporting text
Legend for Dataset S1
SI References

Other supporting materials for this manuscript include the following:

Dataset S1

Supporting Information Text

Theorem. $\psi_k\left(x_1=\lambda x_1+\mu, \cdots, x_k=\lambda x_k+\mu\right)=\lambda^k \psi_k\left(x_1, \cdots, x_k\right)$.

Proof. ψ_k can be divided into k groups. From 1st to $k-1$th group, the gth group has $\binom{k}{g}\binom{g}{1}$ terms having the form $(-1)^{g+1} \frac{1}{k-g+1} x_{i_1}^{k-g+1} \ldots x_{i_g}$. The final kth group is the term $(-1)^{k-1}(k-1) x_1 \cdots x_k$. Let $x_{i_1}=x_1$, $k \neq g$, the gth group of ψ_k has $\binom{k-l}{g-l}$ terms having the form $(-1)^{g+1} \frac{1}{k-g+1} x_1^{k-g+1}\ x_2 \cdots x_l x_{i_1} \cdots x_{i_{g-l}}$, where $x_1,\ x_2, \cdots, x_l$ are fixed, $x_{i_1}, \cdots, x_{i_{g-l}}$ are selected such that $i_1, \cdots,\ i_{g-l} \neq 1,2, \cdots, l$. Let $\Psi_k\left(x_1, x_2, \cdots, x_l, x_{i_1}, \cdots, x_{i_{g-l}}\right)=\left(\lambda x_1+\mu\right)^{k-g+1}\left(\lambda x_2+\mu\right) \cdots\left(\lambda x_l+\mu\right)\left(\lambda x_{i_1}+\mu\right) \cdots \left(\lambda x_{i_{g-l}}+\mu\right)$, the first group of Ψ_k is $\lambda^k x_1 \cdots x_l x_{i_1} \cdots x_{i_{g-l}}$, the hth group of Ψ_k, $h>1$, has $\binom{k-g+1}{k-h-l+2}$ terms having the form $\lambda^{k-h+1} \mu^{h-1} x_1^{k-h-l+2} x_2 \cdots x_l$. Transforming ψ_k by Ψ_k, then combing all terms with $\lambda^{k-h+1} \mu^{h-1} x_1^{k-h-l+2} x_2 \cdots x_l$, $x_1^{k-h-l+2} \neq x_1$, the summed coefficient is $S1_l=\sum_{g=l}^{h+l-1}(-1)^{g+1} \frac{1}{k-g+1}\binom{k-g+1}{k-h-l+2}\binom{k-l}{g-l}=\sum_{g=l}^{h+l-1}(-1)^{g+1} \frac{(k-l)!}{(h+l-g-1)!(k-h-l+2)!(g-l)!}=0$, 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 ψ_k has $(k-h)\binom{h-1}{g-k+h-1}$ terms having the form $(-1)^{g+1} \frac{1}{k-g+1} x_1 x_2 \ldots x_j^{k-g+1} \ldots x_{k-h+1} x_{i_1} \ldots x_{i_{g-k+h-1}}$, provided that $k \neq g$, $2 \leq j \leq k-h+1$, where $x_1, \ldots, x_{k-h+1}$ are fixed, x_j^{k-g+1} and $x_{i_1}, \cdots, x_{i_{g-k+h-1}}$ are selected. Transforming these terms by $\Psi_k\left(x_1, x_2, \ldots, x_j, \ldots, x_{k-h+1}, x_{i_1}, \ldots, x_{i_{g-k+h-1}}\right)= \left(\lambda x_1+\mu\right)\left(\lambda x_2+\mu\right) \cdots\left(\lambda x_j+\mu\right)^{k-g+1} \cdots\left(\lambda x_{k-h+1}+\mu\right)\left(\lambda x_{i_1}+\mu\right) \cdots\left(\lambda x_{i_{g-k+h-1}}+\mu\right)$, then, there are $k-g+1$ terms having the form $\lambda^{k-h+1} \mu^{h-1} x_1 x_2 \ldots x_{k-h+1}$. So, the combined result is $(-1)^{g+1}(k-h)\binom{h-1}{g-k+h-1} \lambda^{k-h+1} \mu^{h-1} x_1 x_2 \ldots x_{k-h+1}$. Transforming the final kth group of ψ_k by Ψ_k, then, there is one term having the form $(-1)^{k-1}(k-1) \lambda^{k-h+1} \mu^{h-1} x_1 x_2 \ldots x_{k-h+1}$. Another possible combination is that the gth group of ψ_k contains $(g-k+h-1)\binom{h-1}{g-k+h-1}$ terms having the form

$(-1)^{g+1} \frac{1}{k-g+1} x_1 x_2 \ldots x_{k-h+1} x_{i_1} \ldots x_{i_j}^{k-g+1} \ldots x_{i_{g-k+h-1}}$, there is only one term having the form $\lambda^{k-h+1} \mu^{h-1} x_1 x_2 \ldots x_{k-h+1}$.

The above summation $S1_l$ should also be included, i.e., $x_1^{k-h-l+2}=x_1$, $k=h+l-1$, so, combing all terms with $\lambda^{k-h+1} \mu^{h-1} x_1 x_2 \ldots x_{k-h+1}$, according to the binomial theorem, the summed coefficient is

$S2_l=\sum_{g=k-h+1}^{k-1}(-1)^{g+1}\binom{h-1}{g-k+h-1}\left(k-h+1+\frac{g-k+h-1}{k-g+1}\right)+(-1)^{k-1}(k-1)=(-1)^k+(-1)^k(k-h)+(h-2)(-1)^k+(-1)^{k-1}(k-1)=0$. The result is the same if replacing x_1 with x_i, where i is from 2 to 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 $\lambda^k \psi_k\left(x_1, \cdots, x_k\right)$. ☐

Methods

A. d value calibration. Asymptotic d values for the invariant central moments for the exponential distribution were approximated by a quasi-Monte Carlo study (1, 2) based on generating a large quasi-random sample with sample size 1.8 million from the exponential distribution and quasi-subsampling the sample $1.8k$ million times to approximate the distributions of the kernels of the corresponding U-statistics, then computing the binomial kth central moment (Bkm), median kth central moment (mkm), and corresponding quantiles, finally obtained by the formula $d_{rkm}=\frac{km_{bs}-\mathrm{B}km_{bs}}{\mathrm{B}km_{bs}-mkm_{bs}}$ and $d_{qkm}=\frac{pkm_{bs}-p\mathrm{B}km_{bs}}{p\mathrm{B}km_{bs}-\frac{1}{2}}$, where $p\mathrm{B}km=\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq\left(\mathrm{B}km_{bs}\right)}$, $pkm=\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq\left(km_{bs}\right)}$, bs indicates the bootstrap moments. The accuracy of the estimates was verified by comparing the bootstrap central moments to their asymptotic values (errors), yielding errors of ≈ 0.0003, ≈ 0.001, and ≈ 0.03 for the second, third, and fourth central moments, respectively. The sample standard deviation of the kernel distributions for these moments are 2.234, 9.627, and 60.064, respectively, resulting in standardized errors for the biases that were all smaller than 0.001, thus ensuring the accuracy implied in the number of significant digits of the values in Table 1. The calculations of invariant moments using other symmetric weighted averages, $\mathrm{SQM}_{\frac{1}{8}}$, $\mathrm{BM}_{\nu=2, \epsilon=\frac{1}{8}}$, $\mathrm{WM}_{\frac{1}{8}}$, $\mathrm{BWM}_{\frac{1}{8}}$, and $\mathrm{TM}_{\frac{1}{8}}$, were performed analogously, by substituting of the appropriate values.

For a finite sample size of $n=5400$, the d values were estimated using 1000 pseudorandom samples with a bootstrap size of 18000. To estimate the errors of d value estimations under finite sample size, 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 $\sigma_d^2 \approx\left|\frac{\partial d}{\partial x_1}\right|^2 \sigma_{x_1}^2+\left|\frac{\partial d}{\partial x_2}\right|^2 \sigma_{x_2}^2+\left|\frac{\partial d}{\partial x_3}\right|^2 \sigma_{x_3}^2+2\left|\frac{\partial d}{\partial x_1}\right|\left|\frac{\partial d}{\partial x_2}\right| Cov\left(X_1, X_2\right)+2\left|\frac{\partial d}{\partial x_1}\right|\left|\frac{\partial d}{\partial x_3}\right| Cov\left(X_1, X_3\right)+ 2\left|\frac{\partial d}{\partial x_2}\right|\left|\frac{\partial d}{\partial x_3}\right| Cov\left(X_2, X_3\right)=\left|\frac{1}{x_2-x_3}\right|^2 \sigma_{x_1}^2+\left|-\frac{x_1-x_2}{\left(x_2-x_3\right)^2}-\frac{1}{x_2-x_3}\right|^2 \sigma_{x_2}^2+\left|\frac{x_1-x_2}{\left(x_2-x_3\right)^2}\right|^2 \sigma_{x_3}^2+2\left|\frac{1}{x_2-x_3}\right|\left|-\frac{x_1-x_2}{\left(x_2-x_3\right)^2}-\frac{1}{x_2-x_3}\right| Cov\left(X_1, X_2\right)+2\left|\frac{1}{x_2-x_3}\right|\left|\frac{x_1-x_2}{\left(x_2-x_3\right)^2}\right| Cov\left(X_1, X_3\right)+2\left|-\frac{x_1-x_2}{\left(x_2-x_3\right)^2}-\frac{1}{x_2-x_3}\right|\left|\frac{x_1-x_2}{\left(x_2-x_3\right)^2}\right| Cov\left(X_2, X_3\right)$. Since for the recombined mean, $\sigma_{x_1}^2=0$, so, $\sigma_{d_{rm}}^2 \approx\left|-\frac{x_1-x_2}{\left(x_2-x_3\right)^2}-\frac{1}{x_2-x_3}\right|^2 \sigma_{x_2}^2+\left|\frac{x_1-x_2}{\left(x_2-x_3\right)^2}\right|^2 \sigma_{x_3}^2+2\left(-\frac{x_1-x_2}{\left(x_2-x_3\right)^2}-\frac{1}{x_2-x_3}\right)\left(\frac{x_1-x_2}{\left(x_2-x_3\right)^2}\right) Cov\left(X_2, X_3\right)$, where x_1 is the expected value, x_2 is the weighted average used, x_3 is the median. For quantile mean, since $\sigma_{x_3}^2=0$, $\sigma_{d_{qm}}^2 \approx\left|\frac{1}{x_2-x_3}\right|^2 \sigma_{x_1}^2+\left|-\frac{x_1-x_2}{\left(x_2-x_3\right)^2}-\frac{1}{x_2-x_3}\right|^2 \sigma_{x_2}^2+2\left(\frac{1}{x_2-x_3}\right)\left(-\frac{x_1-x_2}{\left(x_2-x_3\right)^2}-\frac{1}{x_2-x_3}\right) Cov\left(X_1, X_2\right)$, where x_1 is the percentile of the expected value, x_2 is the percentile of the weighted average used, x_3 is the percentile of median, $\frac{1}{2}$. Finally, the errors were estimated by the corresponding sample statistics.

B. AAB, AB, and SSE. The computations of ABs and AABs for invariant central moments were just the same as described in the main text. The SSE was computed by approximating the sampling distribution with 1000 pseudorandom samples for $n=5400$

and 30 pseudorandom samples for $n = 1.8 \times 10^6$. 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 AB 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 (3). The computational methods used for two-parameter distributions were identical. The computations of invariant moments were described in the SI Text.

C. Comparisons to M-estimator and Percentile estimator. Within the same kurtosis range and five two-parametric distributions as the above, the percentile estimator for the Weibull distribution were computed using the method proposed by Marks (2005) (4) and the parameter setting proposed by Boudt, Caliskan, and Croux (2011) (5). The results were then transformed to the first four moments to compute AABs. The robust M-estimator for the Weibull distribution were also computed in the same way using the method proposed by He and Fung (1999) (6). Bisection is used to find the solution of the key equation in (6), while the results from the percentile estimator were used as initial values (-0.3 and +0.3).

D. The impact of bootstrap size on variance. The study of the impact of the bootstrap size on the variance for the exponential distribution was done the same as above, just changing the bootstrap size from $n = 1.8 \times 10^2$ to $n = 1.8 \times 10^4$.

E. Maximum asymptotic biases. For simplicity, a brute force approach is used to estimate the maximum biases of SWAs and SWkms for five unimodal distributions. A wide range is 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 is broken to 100 parts, combining with the above results for AAB estimations, the maximum of both is determined to be very close to the true maximum. Pseudo-maximum bias is the same as 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 SWAs approach zero.

For example, for the Perato distribution,

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

$\lim_{\alpha\to 2} B_{\mathrm{SQA}}(\epsilon,\alpha) = \lim_{\alpha\to 2} \frac{\frac{1}{2}\left(x_m(1-\epsilon)^{-\frac{1}{\alpha}} + x_m\epsilon^{-\frac{1}{\alpha}}\right)}{\sqrt{\frac{\alpha {x_m}^2}{(1-\alpha)^2(\alpha-2)}}} - \frac{\frac{\alpha x_m}{\alpha-1}}{\sqrt{\frac{\alpha {x_m}^2}{(1-\alpha)^2(\alpha-2)}}} = \lim_{\alpha\to 2} \frac{\frac{1}{2}\left(\frac{1}{\sqrt{\epsilon}} + \frac{1}{\sqrt{1-\epsilon}}\right)}{\sqrt{\frac{\alpha}{(1-\alpha)^2(\alpha-2)}}} + \lim_{\alpha\to 2} \frac{-\alpha}{\sqrt{\frac{\alpha}{(\alpha-2)}}} = 0.$

In the previous article, it is proven that when the kurtoses of the distributions approach infinity, all distributions will be ordered, that means the SWAs based on the orderliness will follow the mean-SWA-median inequality, 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 $\lim_{\alpha\to 0} \frac{\Gamma\left(1+\frac{1}{\alpha}\right)}{\sqrt{\Gamma\left(\frac{\alpha+2}{\alpha}\right)}} = \lim_{\alpha\to 0} \frac{\left(1+\frac{1}{\alpha}-1\right)!}{\sqrt{\left(\frac{\alpha+2}{\alpha}-1\right)!}} = \lim_{\alpha\to 0} \frac{\left(\frac{1}{\alpha}\right)!}{\sqrt{\left(2\times\frac{1}{\alpha}\right)!}} = 0$, the ratio of m and σ is $\lim_{\alpha\to 0} \frac{\sqrt[\alpha]{\ln(2)}}{\sqrt{\Gamma\left(\frac{\alpha+2}{\alpha}\right)}} = \lim_{\alpha\to 0} \frac{0}{\sqrt{\Gamma\left(\frac{\alpha+2}{\alpha}\right)}} = 0$.

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 $\lim_{\alpha\to 0} \frac{P^{-1}\left(\alpha,\frac{1}{2}\right)}{\sqrt{\alpha}} = 0$ (7).

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.

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. CR Rao, *Linear statistical inference and its applications.* (John Wiley & Sons Inc), (1965).
4. NB Marks, Estimation of weibull parameters from common percentiles. *J. applied Stat.* **32**, 17–24 (2005).
5. K Boudt, D Caliskan, C Croux, Robust explicit estimators of weibull parameters. *Metrika* **73**, 187–209 (2011).
6. X He, WK Fung, Method of medians for lifetime data with weibull models. *Stat. medicine* **18**, 1993–2009 (1999).
7. MA Chaudhry, SM Zubair, *On a class of incomplete gamma functions with applications.* (Chapman and Hall/CRC), (2001).