Weighted Geometric Mean and its Properties

Ievgen Turchyn

Department of Mechanics and Mathematics
Oles Honchar Dnipro National University
72, Gagarin av.
Dnipro, 49010, Ukraine
evgturchyn@gmail.com

Abstract

Various means (the arithmetic mean, the geometric mean, the harmonic mean, the power means) are often used as central tendency statistics. A new statistic of such type is offered for a sample from a distribution on the positive semi-axis, the γ-weighted geometric mean. This statistic is a certain weighted geometric mean with adaptive weights. Monte Carlo simulations showed that the γ-weighted geometric mean possesses low variance: smaller than the variance of the 0.20-trimmed mean for the Lomax distribution. The bias of the new statistic was also studied. We studied the bias in terms of nonparametric confidence intervals for the quantiles which correspond of our statistic for the case of the Lomax distribution. Deviation from the median for the γ-weighted geometric mean was measured in terms of the MSE for the log-logistic distribution and the Nakagami distribution (the MSE for the γ-weighted geometric mean was comparable or smaller than the MSE for the sample median).

Keywords: Central tendency; Weighted geometric mean.

MSC 2010 No.: 62G05, 62G30

1 Introduction

Such measures of central tendency as the median, the trimmed mean (and variants of the trimmed mean), the winsorized mean, the trimean, the Gastwirth's estimator, etc., have been studied extensively, see, for example, [17] and surveys in [2], [12]. In particular, there are some recent articles devoted to study of the trimmed mean and trimmed mean analogs — for instance, [3], [5], [6].

Other similar "means" exist for distributions on the positive semi-axis: the geometric mean, the harmonic mean, their generalizations and analogues. A generalized notion of the geometric mean for a non-negative random variable was introduced in [4]. An excellent review of properties of the geometric mean and the sample geometric mean (which includes examples of the geometric mean usage in various applied spheres) was given by [16]. It is worth saying that the geometric mean was successfully employed as a statistic of central tendency in such areas as scientometrics ([15]), seismology ([1]) and veterinary medicine ([14]), to name but a few.

Two natural generalizations of the geometric mean are the weighted geometric mean and the power mean. However, statistical properties of such means did not receive much attention (this is highlighted in [10] for the case of power means). But [10] showed that the power mean of suitable order has good robustness. It is also natural to expect robustness from a weighted geometric mean in which lower weights are assigned to observations that are closer to the extreme values in a sample.

We introduce a new measure of central tendency for a sample from a distribution on the positive semi-axis: a weighted geometric mean with observation-dependent weights. As far as the author knows there was virtually no research devoted to this type of means as a statistical estimator.

We will demonstrate by Monte Carlo simulations that our weighted geometric mean has smaller sample variance than the 0.20-trimmed mean for the Lomax distribution. Other strengths of the new statistic are existence of an analytical expression and simplicity of calculation. The choice of the Lomax distribution (see [7]) is justified by the fact that Pareto-like distributions can be used quite successfully for modeling many real-life phenomena (this was shown, for instance, by [11]).

2 γ-weighted geometric mean

Let us introduce a new statistic.

Definition 1. *Let* $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$ *be a sample from a probability distribution* F *on* $(0; +\infty)$*. The* γ*-weighted geometric mean (*γ*-wgm) of the sample* $\boldsymbol{\xi}$ *is defined as*

$$\tilde{\xi} = \prod_{i=1}^{n} \xi_i^{w'_i},$$

where

$$w'_i = \frac{w_i}{\sum_{j=1}^{n} w_j}, \quad w_i = \frac{1}{|\xi_i - \text{med}(\boldsymbol{\xi})|^{\gamma} + 1},$$

$\gamma > 0$ *is a constant,* $\text{med}(\boldsymbol{\xi})$ *is the sample median of* $\boldsymbol{\xi}$*.*

Remark 1. *The motivation for such choice of weights* w_i *is that we penalize strongly those observations* ξ_k *which are far from the sample median.*

Remark 2. *It is worth noting that the* γ*-wgm is a certain exponentiated weighted arithmetic mean.*

Remark 3. *The* γ*-wgm possesses substantial robustness against the pollution of a sample with "large-valued noise": i.e., if we replace a certain percentage of observations* ξ_i *by large positive numbers the relative change of the* γ*-wgm value will be only limited. The author conducted a simulation experiment with samples from the Lomax distribution. Samples* $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$ *of size* $n = 200$ *from the Lomax distribution with given parameters were generated repeatedly. The* γ*-wgm values* $\tilde{\xi}$ *and* $\widetilde{\xi^{(1)}}$ *were calculated for* $\boldsymbol{\xi}$ *and* $\boldsymbol{\xi}^{(1)}$ *correspondingly, where*

$$\boldsymbol{\xi}^{(1)} = (\xi_1, \xi_2, \dots, \xi_{180}, \eta_1, \eta_2, \dots, \eta_{20}),$$

$(\eta_1, \eta_2, \dots, \eta_{20})$ *was a sample from the uniform distribution on* $[a; b]$ *(*a *is relatively large). The relative change was not too big, namely,*

$$\frac{|\widetilde{\xi^{(1)}} - \tilde{\xi}|}{\tilde{\xi}} \leq 0.55,$$

for all the parameters of the Lomax distribution which were considered.

We will show that the γ-weighted geometric mean is a measure of central tendency with low sample variance for the Lomax distribution. We will use the following parametrization of the two-parameter Lomax (or Pareto type II) distribution: its density is

$$f(x; q, b) = \frac{q}{b} \left(1 + \frac{x}{b}\right)^{-q-1}, \ x > 0.$$

Let us formulate a property of the γ-wgm.

Lemma 1. *Suppose that* $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)$ *is a sample from a probability distribution* F *on* $(0; +\infty)$, $m_k(F)$ *is the* k*-th moment of* F, $\tilde{\xi}$ *is the* γ*-wgm of* $\boldsymbol{\xi}$.

If $m_k(F) < \infty$, *then*

$$\mathsf{E}\tilde{\xi}^k < \infty.$$

Proof. Suppose that $m_k(F) < +\infty$. Then we have

$$\mathsf{E}\tilde{\xi}^k = \mathsf{E}\prod_{i=1}^{n} \xi_i^{kw_i'} \leq \sum_{i=1}^{n} \mathsf{E}(w_i'\xi_i^k)$$

using the weighted AM-GM inequality. But the right-hand side is finite since

$$\mathsf{E}(w_i'\xi_i^k) \leq \mathsf{E}(\xi_i^k) < +\infty$$

for all i. So $\mathsf{E}\tilde{\xi}^k < +\infty$. □

3 Variance of the γ-wgm

Let us describe the comparison of the average values of the γ-wgm variance and the sample trimmed mean variance. But first we need to recall the notion of the trimmed mean.

Definition 2. *The sample* α*-trimmed mean for a sample* $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ *is defined as*

$$T_\alpha = \frac{1}{n-2g} \sum_{i=g+1}^{n-g} \xi_{i;n},$$

where $g = [\alpha n]$, $\xi_{i;n}$ *are the order statistics of* $\boldsymbol{\xi}$.

Example 1. *Let us consider the uniform distribution* $U([0;\theta])$ *on* $[0;\theta]$. *We generated a sample of size* $n = 13$ *from* $U([0;\theta])$ *(*$\theta = 1$ *for this sample):*

$$\boldsymbol{\xi} = (0.915, 0.763, 0.806, 0.734, 0.682, 0.012, 0.758, 0.499, 0.466, 0.901, 0.942, 0.320, 0.793).$$

The sample 0.25*-trimmed mean for this sample is*

$$T_{0.25} = \frac{1}{7}(0.499 + 0.682 + 0.734 + 0.758 + 0.763 + 0.793 + 0.806) \approx 0.719.$$

Definition 3. *The population* α*-trimmed mean for a distribution* F *is defined as*

$$\mu_\alpha = \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x dF(x).$$

Example 2. *Let us consider a distribution with the c.d.f.*

$$F(x) = \begin{cases} 0, & x < 0; \\ x^m, & 0 \leq x \leq 1; \\ 1, & x > 1, \end{cases}$$

where $m > 1$. *Its p.d.f. is* $p(x) = mx^{m-1}$, $x \in [0;1]$. *The population* α*-trimmed mean for this distribution is*

$$\mu_\alpha = \frac{1}{1-2\alpha} \int_{\alpha^{1/m}}^{(1-\alpha)^{1/m}} xp(x)dx = \frac{m}{(1-2\alpha)(m+1)} \left((1-\alpha)^{(m+1)/m} - \alpha^{(m+1)/m}\right).$$

Let us formulate the Yuen test (see [18]) which is used for testing the null hypothesis about equality of the population trimmed means for two samples.

The Yuen test

Suppose that $(\xi_1^{(1)}, \xi_2^{(1)}, \ldots, \xi_{n_1}^{(1)})$ and $(\xi_1^{(2)}, \xi_2^{(2)}, \ldots, \xi_{n_2}^{(2)})$ are ordered samples from unknown distributions F_1 and F_2 correspondingly, $\beta > 0$ is the amount of trimming and denote $g_i = [\beta n_i]$. Let

$$\bar{\xi}_{tg}^{(i)} = \frac{1}{n_i - 2g_i}(\xi_{g_i+1}^{(i)} + \xi_{g_i+2}^{(i)} + \ldots + \xi_{n_i-g_i}^{(i)}),$$

$$\bar{\xi}_{wg}^{(i)} = \frac{1}{n_i}((g_i+1)\xi_{g_i+1}^{(i)} + \xi_{g_i+2}^{(i)} + \ldots + \xi_{n_i-g_i-1}^{(i)} + (g_i+1)\xi_{n_i-g_i}^{(i)}),$$

$$\mathrm{SSD}_{i,wg} = (g_i+1)(\xi_{g_i+1}^{(i)} - \bar{\xi}_{wg}^{(i)})^2 + (\xi_{g_i+2}^{(i)} - \bar{\xi}_{wg}^{(i)})^2 + \ldots + (\xi_{n_i-g_i-1}^{(i)} - \bar{\xi}_{wg}^{(i)})^2 + (g_i+1)(\xi_{n_i-g_i}^{(i)} - \bar{\xi}_{wg}^{(i)})^2,$$

$s_{i,wg}^2 = \mathrm{SSD}_{i,wg}/(h_i - 1)$, where $h_i = n_i - 2g_i$, $i = 1, 2$.

Let the H_0 be $\mu_{1,\beta} = \mu_{2,\beta}$ (the population β-trimmed means are equal). The Yuen test statistic t is defined by

$$t = \frac{\bar{\xi}_{tg}^{(1)} - \bar{\xi}_{tg}^{(2)}}{(s_{1,wg}^2/h_1 + s_{2,wg}^2/h_2)^{1/2}}.$$

The approximate distribution of this statistic for true H_0 is the t distribution with f degrees of freedom, where f can be found from

$$\frac{1}{f} = \frac{c^2}{h_1 - 1} + \frac{(1-c)^2}{h_2 - 1},$$

$$c = \frac{s_{1,wg}^2/h_1}{s_{1,wg}^2/h_1 + s_{2,wg}^2/h_2}.$$

We will delineate now the procedure of comparing the average values of variances of the sample γ-wgm and the sample trimmed mean.

Let $\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \ldots \boldsymbol{\xi}^{(M)}$ be M independent samples, each of size N, from a given parent distribution F. We will use the Lomax distribution with various (q, b) parameters as F. Notations W and T_α will be used correspondingly for the γ-wgm and the sample α-trimmed mean.

We denote by $s_{(W);M}^2$ and $s_{(T_\alpha),M}^2$ the sample variances of $(W(\boldsymbol{\xi}^{(1)}), W(\boldsymbol{\xi}^{(2)}), \ldots, W(\boldsymbol{\xi}^{(M)}))$ and $(T_\alpha(\boldsymbol{\xi}^{(1)}), T_\alpha(\boldsymbol{\xi}^{(2)}), \ldots, T_\alpha(\boldsymbol{\xi}^{(M)}))$ correspondingly. The notations $\breve{\sigma}_{(W),M}^2$ and $\breve{\sigma}_{(T_\alpha),M}^2$ are used for the population β-trimmed means of the distributions of $s_{(W);M}^2$ and $s_{(T_\alpha),M}^2$ correspondingly.

We performed the following for all the chosen values of (q, b):

1) samples $\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \ldots, \boldsymbol{\xi}^{(M)}$ from F were simulated;

2) the hypothesis $H_0 : \breve{\sigma}_{(W),M}^2 = \breve{\sigma}_{(T_\alpha),M}^2$ was tested versus the alternative $H_1 : \breve{\sigma}_{(W),M}^2 < \breve{\sigma}_{(T_\alpha),M}^2$ using the Yuen test.

The following values of the parameters were taken: $M = 200$, $\alpha = 0.20$. The value of γ we chose was $\gamma_0 = 0.4$. We set the trimming amount $\beta = 0.10$.

Results of the Yuen test for sample sizes $N = 50$, 200 and 500 are given correspondingly in Tables 1, 2 and 3. So in all these cases we rejected H_0 in favor of H_1: "the 0.10-trimmed mean of the sample variance of W is less than the 0.10-trimmed mean of the sample variance of $T_{0.20}$" at the 0.01 significance level. And we interpret this as relative smallness of the variance of the γ-wgm.

The value $\gamma_0 = 0.4$ seems to be more or less close to the "optimal" γ value for the Lomax distribution: the γ-wgm had bigger variance for several other values of $\gamma \in (0; 1)$.

The γ-wgm does not outperform the 0.20-trimmed mean with respect to the variance for several other distribution families. For instance, this is true for such families as the Weibull distribution, the Fréchet distribution and the gamma distribution.

The γ-wgm (for $\gamma = 0.4$) for these three families has smaller variance than the 0.20-trimmed mean for certain values of the distribution parameters but it is has bigger variance than the 0.20-trimmed mean for some other values of the parameters. The Yuan test results for testing the hypothesis $H_0 : \breve{\sigma}_{(W),M}^2 = \breve{\sigma}_{(T_\alpha),M}^2$ against the alternative $H_1 : \breve{\sigma}_{(W),M}^2 < \breve{\sigma}_{(T_\alpha),M}^2$ are given in Tables 4, 5 and 6 correspondingly for the Fréchet distribution, gamma distribution and the Weibull distribution.

We set $M = 400$, $N = 200$. We used the parametrizations

$$f_F(x; a, \beta) = a\beta^a x^{-a-1} \exp\{-(x/\beta)^{-a}\}, \ x > 0,$$

$$f_G(x; p, \beta) = \frac{1}{\beta^p \Gamma(p)} x^{p-1} \exp\{-x/\beta\}, \ x > 0,$$

$$f_W(x; a, \beta) = \frac{a}{\beta}(x/\beta)^{a-1} \exp\{-(x/\beta)^a\}, \ x > 0,$$

correspondingly for the Fréchet density, the gamma density and the Weibull density.

Table 1: Testing $H_0 : \sigma^2_{(W),M} = \sigma^2_{(T_\alpha),M}$, $N = 50$ (Lomax distribution)

q	b	p-value	q	b	p-value	q	b	p-value
5	100	< 0.001	10	100	< 0.001	200	100	< 0.001
5	200	< 0.001	50	10	< 0.001	500	20	< 0.001
5	1000	< 0.001	50	100	< 0.001	500	100	< 0.001
10	20	< 0.001	200	10	< 0.001	500	200	< 0.001
10	50	< 0.001	200	20	< 0.001			

Table 2: Testing $H_0 : \sigma^2_{(W),M} = \sigma^2_{(T_\alpha),M}$, $N = 200$ (Lomax distribution)

q	b	p-value	q	b	p-value
5	100	< 0.001	50	100	< 0.001
5	200	< 0.001	200	10	< 0.001
5	1000	< 0.001	200	20	< 0.001
10	20	< 0.001	200	100	< 0.001
10	50	< 0.001	500	20	< 0.001
10	100	< 0.001	500	100	< 0.001
50	10	< 0.001	500	200	< 0.001
50	20	< 0.001			

Table 3: Testing $H_0 : \sigma^2_{(W),M} = \sigma^2_{(T_\alpha),M}$, $N = 500$ (Lomax distribution)

q	b	p-value	q	b	p-value	q	b	p-value
5	50	< 0.001	50	10	< 0.001	200	100	< 0.001
5	100	< 0.001	50	20	< 0.001	500	20	< 0.001
5	1000	< 0.001	50	100	< 0.001	500	100	< 0.001
10	20	< 0.001	200	10	< 0.001	500	200	< 0.001
10	50	< 0.001	200	20	< 0.001			

Table 4: Testing $H_0 : \sigma^2_{(W),M} = \sigma^2_{(T_\alpha),M}$ (Fréchet distribution, $N = 200$)

a	β	p-value	a	β	p-value
50	30	< 0.0001	80	8	0.0084
50	40	< 0.0001	80	10	0.0019
50	50	< 0.0001	100	5	0.0239
70	5	0.0209	100	8	0.0158
70	8	0.0024	100	10	0.0040
70	10	0.0010	140	5	0.0664
70	30	< 0.0001	140	8	0.0452
70	40	< 0.0001	160	5	0.1853
70	50	< 0.0001	160	8	0.0483
80	5	0.0143			

Table 5: Testing $H_0 : \sigma^2_{(W),M} = \sigma^2_{(T_\alpha),M}$ (gamma distribution, $N = 200$)

p	β	p-value	p	β	p-value
10	5	< 0.0001	20	10	< 0.0001
10	10	< 0.0001	20	15	0.0001
10	15	< 0.0001	20	30	0.0009
10	30	< 0.0001	20	40	0.0608
10	40	< 0.0001	20	50	0.0277
10	50	0.0001	50	40	0.1877
15	5	< 0.0001	50	45	0.2553
15	10	< 0.0001	50	50	0.4720
15	15	< 0.0001	50	55	0.3457
15	30	0.0005	50	60	0.5783
15	40	0.0044	60	40	0.2107
15	50	0.0004	60	45	0.3676
20	5	< 0.0001	60	50	0.4926

Table 6: Testing $H_0 : \sigma^2_{(W),M} = \sigma^2_{(T_\alpha),M}$ (Weibull distribution, $N = 200$)

a	β	p-value	a	β	p-value
10	7	0.1402	15	10	0.0023
10	10	0.2453	15	15	0.0549
10	15	0.1640	20	2	0.0136
10	100	0.9303	20	3	0.0583
10	200	0.9996	20	5	0.0327
15	2	0.0554	20	7	0.0182
15	3	0.0099	20	10	0.0128
15	5	0.0279	20	15	0.0445
15	7	0.0210			

4 Smallness of bias and MSE

It turned out that the γ-wgm is a median estimator with small MSE for certain distribution families. This is true for the log-logistic distribution (see [7]) and the Nakagami distribution (see [8]). We checked this fact by means of simulation. These distributions have the following densities: the density of the log-logistic distribution with parameters a, b is

$$f_{LL}(x; a, b) = \frac{ax^{a-1}}{b^a(1 + (x/b)^a)^2}, \;\; x > 0;$$

the density of the Nakagami distribution with parameters m, Ω is

$$f_{Nak}(x; m, \Omega) = \frac{2m^m}{\Gamma(m)\Omega^m} x^{2m-1} \exp\left\{-\frac{m}{\Omega}x^2\right\}, \;\; x > 0.$$

We calculated for simulated samples the mean square error for the γ-wgm and the sample median as the median estimators and compared these errors (see Tables 7, 8).

Table 7: MSE values for log-logistic distribution

a	10	10	10	15	15	20	20	30
b	50	70	100	50	80	5	10	5
$\mathsf{MSE}_{\mathrm{med}}$	0.504	0.945	2.000	0.217	0.577	$1.270 \cdot 10^{-3}$	$4.986 \cdot 10^{-3}$	$5.651 \cdot 10^{-4}$
$\mathsf{MSE}_{\mathrm{WGM}}$	0.412	0.773	1.677	0.170	0.459	$1.014 \cdot 10^{-3}$	$3.936 \cdot 10^{-3}$	$4.374 \cdot 10^{-4}$

Table 8: MSE values for Nakagami distribution

m	5	10	20	70	80	70
Ω	100	300	300	200	200	500
$\mathsf{MSE}_{\mathrm{med}}$	$3.840 \cdot 10^{-2}$	$5.636 \cdot 10^{-2}$	$2.957 \cdot 10^{-2}$	$5.623 \cdot 10^{-3}$	$4.960 \cdot 10^{-3}$	$1.376 \cdot 10^{-2}$
$\mathsf{MSE}_{\mathrm{WGM}}$	$5.006 \cdot 10^{-2}$	$5.628 \cdot 10^{-2}$	$2.388 \cdot 10^{-2}$	$3.979 \cdot 10^{-3}$	$3.455 \cdot 10^{-3}$	$9.909 \cdot 10^{-3}$

The numeric value $\gamma = 0.4$ was left unchanged. There were generated $M = 5000$ samples of size $N = 200$ from the Nakagami distribution and the log-logistic distribution. The MSE for the γ-wgm was smaller than the MSE for the sample median for most of the parameter values for the Nakagami distribution and for all the parameter values for the log-logistic distribution.

We will demonstrate (using another technique) that the γ-wgm can also be used as a median estimator with small "deviation" for the Lomax distribution. Namely, it will be shown that the quantile which corresponds to the γ-wgm is close to 0.5.

Let us describe the methodology. We simulated M samples $\boldsymbol{\xi}^{(j)}$ $(1 \le j \le M)$ of size N from a Lomax distribution with some fixed parameters (denote this distribution by F) and the γ-wgm value w_j was calculated for each such sample $\boldsymbol{\xi}^{(j)}$. So we obtained a sample $\mathbf{w} = (w_1, w_2, \ldots, w_M)$ of size M of γ-wgm values for the Lomax distribution F. Let us consider the vector $\boldsymbol{\zeta} = (F(w_1), F(w_2), \ldots, F(w_M))$. We calculated the distribution-free confidence intervals (based on the normal approximation, see [13]) with the 0.95 confidence level for $q_{0.1}, q_{0.5}$ and $q_{0.9}$ — correspondingly the 0.1-quantile, 0.5-quantile and 0.9-quantile of the distribution of $F(w_1)$ (i.e. the sample distribution of the γ-wgm) using $\boldsymbol{\zeta}$. We set $\gamma = 0.4$, $N = 200, M = 5000$. These confidence intervals and quantile estimates $\hat{q}_{0.1}$, $\hat{q}_{0.5}$ and $\hat{q}_{0.9}$ for various Lomax distribution parameters q, b are given in Table 9.

Denote the confidence interval for $q_{0.1}$ by $(q^{(l)}_{0.1}; q^{(r)}_{0.1})$ and the confidence interval for $q_{0.9}$ by $(q^{(l)}_{0.9}; q^{(r)}_{0.9})$. We see from Table 9 that for all these parameters $(q; b)$

$$|q^{(l)}_{0.1} - 0.5| \le 0.13,$$

$$|q^{(r)}_{0.9} - 0.5| \le 0.05.$$

We interpret this as follows: the γ-wgm distribution is concentrated mostly "not too far" from the median of F.

Table 9: Estimates and confidence intervals for $q_{0.1}, q_{0.5}$ and $q_{0.9}$ ($N = 200, M = 5000$)

q	b	$\hat{q}_{0.1}$	CI for $q_{0.1}$	$\hat{q}_{0.5}$	CI for $q_{0.5}$	$\hat{q}_{0.9}$	CI for $q_{0.9}$
5	100	0.3759	(0.3745; 0.3781)	0.4155	(0.4144; 0.4166)	0.4561	(0.4545; 0.4576)
5	200	0.3749	(0.3734; 0.3765)	0.4150	(0.4139; 0.4160)	0.4559	0.4544; 0.4575)
5	1000	0.3770	(0.3750; 0.3782)	0.4155	(0.4143; 0.4166)	0.4564	(0.4550; 0.4580)
10	20	0.3766	(0.3754; 0.3778)	0.4159	(0.4151; 0.4169)	0.4569	(0.4557; 0.4583)
10	50	0.3766	(0.3750; 0.3778)	0.4154	(0.4143; 0.4166)	0.4547	(0.4535; 0.4562)
10	100	0.3741	(0.3727; 0.3757)	0.4143	(0.4132; 0.4156)	0.4534	(0.4521; 0.4551)
50	10	0.3810	(0.3793; 0.3826)	0.4191	(0.4180; 0.4201)	0.4575	(0.4566; 0.4586)
50	20	0.3789	(0.3775; 0.3802)	0.4168	(0.4157; 0.4178)	0.4548	(0.4535; 0.4564)
50	100	0.3769	(0.3754; 0.3780)	0.4136	(0.4126; 0.4146)	0.4531	(0.4519; 0.4542)
200	10	0.3846	(0.3834; 0.3860)	0.4218	(0.4205; 0.4228)	0.4582	(0.4568; 0.4597)
200	20	0.3839	(0.3828; 0.3849)	0.4208	(0.4198; 0.4220)	0.4581	(0.4569; 0.4596)
200	100	0.3791	(0.3773; 0.3805)	0.4168	(0.4158; 0.4177)	0.4558	(0.4544; 0.4574)
500	20	0.3844	(0.3833; 0.3861)	0.4220	(0.4211; 0.4232)	0.4603	(0.4587; 0.4620)
500	100	0.3816	(0.3799; 0.3830)	0.4186	(0.4176; 0.4197)	0.4579	(0.4566; 0.4592)
500	200	0.3793	(0.3779; 0.3805)	0.4172	(0.4159; 0.4182)	0.4552	(0.4537; 0.4567)

5 Conclusion

It is well known that such a classical measure of central tendency as the sample mean is not very suitable for heavy-tailed distributions due to absence of robustness. The sample median is robust but has high variance. Therefore, a statistic of central tendency which is both more robust than the sample mean and has relatively low variance is of significant value. In particular new optimal and suboptimal median estimators certainly deserve study.

A new statistic, the γ-weighted geometric mean, was devised. Simulations showed that the γ-wgm has small variance for the Lomax distribution: smaller than that of the 0.20-trimmed mean. Besides, the γ-wgm possesses acceptable “bias” as a median estimator (we measured this “bias” in terms of closeness of quantiles).

The γ-wgm can also serve as a median estimator with relatively small “deviation” for some other distributions. Simulations demonstrated that the MSE of the γ-wgm (as a median estimator) is quite small for the Nakagami distribution and the log-logistic distribution.

The γ-weighted geometric mean can be used as an alternative to the sample median in various spheres due to the fact that Pareto-like distributions model many real datasets quite well. Such areas include biology, econometrics, hydrology, pharmaceutics, scientometrics, seismology, etc.

References

[1] Abyani, M., Asgarian, B. and Zarrin, M. (2019). Sample geometric mean versus sample median in closed form framework of seismic reliability evaluation: a case study comparison, Earthq. Eng. Eng. Vib., Vol. 18, No. 1, pp. 187–201.

[2] Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. and Tukey, J.W. (1972). *Robust Estimates of Location: Survey and Advances*, Princeton University Press, Princeton.

[3] Bieniek, M. (2016). Comparison of the bias of trimmed and winsorized means, Comm. Statist. Theory Methods, Vol. 45, No. 22, pp. 6641–6650.

[4] Feng, C., Wang, H., Zhang, Y., Han, Y., Liang, Y. and Tu, X.M. (2017). Generalized definition of the geometric mean of a non negative random variable, Comm. Statist. Theory Methods, Vol. 46, No. 7, pp. 3614–3620.

[5] Figueiredo, F. and Gomes, M.I. (2016). The total median statistic to monitor contaminated normal data, Qual. Technol. Quant. M., Vol. 13, No. 1, pp. 78–87.

[6] García-Pérez, A. (2016). A von Mises approximation to the small sample distribution of the trimmed mean, Metrika, Vol. 79, No. 4, pp. 369–388.

[7] Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley and Sons, New Jersey.

[8] Kumar, K., Garg, R. and Krishna, H. (2017). Nakagami distribution as a reliability model under progressive censoring, Int. J. Syst. Assur. Eng. Manag., Vol. 8, No. 1, pp. 109–122.

[9] Martinez, M.N. and Bartholomew, M.J. (2017). What does it "mean"? A review of interpreting and calculating different types of means and standard deviations, Pharm., Vol. 9, No. 2(14).

[10] Mukhopadhyay, S., Das, A.J., Basu, A., Chatterjee, A. and Bhattacharya S. (2019). Does the generalized mean have the potential to control outliers?, Comm. Statist. Theory Methods, DOI: 10.1080/03610926.2019.1652320

[11] Newman, M.E.J. (2005). Power laws, Pareto distributions and Zipf's law, Contemp. Phys., Vol. 46, No. 5, pp. 323–351.

[12] Reed III, J.F. and Stark, D.B. (1994). Robustness estimators of location: a classification of linear and adaptive estimators, J. Appl. Stat., Vol. 21, No. 3, pp. 95–124.

[13] Serfling, R.J. (2009). *Approximation Theorems of Mathematical Statistics*, Wiley and Sons, New Jersey.

[14] Smothers, C.D., Sun, F. and Dayton, A.D. (1999). Comparison of arithmetic and geometric means as measures of a central tendency in cattle nematode populations, Vet. Parasitol., Vol. 81, No. 3, pp. 211–224.

[15] Thelwall, M. (2016). The precision of the arithmetic mean, geometric mean and percentiles for citation data: an experimental simulation modelling approach, J. Informetr., Vol. 10, No. 1, pp. 110–123.

[16] Vogel, R.M. (2020). The geometric mean?, Comm. Statist. Theory Methods, DOI: 10.1080/03610926.2020.1743313

[17] Wilcox, R.R. (2016). *Introduction to Robust Estimation and Hypothesis Testing*, Academic Press, New York.

[18] Yuen, K.K. (1974). The two-sample trimmed t for unequal population variances, Biometrika, Vol. 61, No. 1, pp. 165–170.