REVISTA DE MATEMÁTICA: TEORÍA Y APLICACIONES 2018 **25**(2) : 169–184
CIMPA – UCR ISSN: 1409-2433 (PRINT), 2215-3373 (ONLINE)
DOI: https://doi.org/10.15517/rmta.v25i2.33653

LARGE-NUMBERS BEHAVIOR OF THE SAMPLE MEAN AND THE SAMPLE MEDIAN: A COMPARATIVE EMPIRICAL STUDY

EL COMPORTAMIENTO DE LA MEDIA Y LA MEDIANA DE MUESTRAS, SEGÚN CRECE EL TAMAÑO DE LA MUESTRA: UN ESTUDIO EMPÍRICO COMPARATIVO

OSVALDO MARRERO*

Received: 28/Apr/2017; Revised: 13/May/2018;
Accepted: 16/May/2018

---

*Department of Mathematics and Statistics, Villanova University, Villanova, Pennsylvania, U.S.A. E-Mail: Osvaldo.Marrero@villanova.edu

Abstract

Through simulation, we study and compare the large-numbers behavior of the sample mean and the sample median as estimators of their respective corresponding population parameters. We only consider skewed, continuous distributions that do have a mean and a unique median. With respect to the pertinent probabilities, our results throw some light on the speed of convergence, the effect of the skewness, and, intriguinly, also suggest the existence of a couple of changepoints or change-intervals. We propose some questions for further study.

Keywords: change-interval; comparing estimators; convergence rate; simulation; skewness effect.

Resumen

Se presentan los resultados de un estudio de simulación donde hemos comparado el comportamiento de la media y la mediana de muestras como estimadores de sus respectivos parámetros, según crece el tamaño de la muestra. Sólo consideramos distribuciones continuas y sesgadas cuyas medias existen y cuyas medianas son únicas. Con respecto a las probabilidades en cuestión, nuestros resultados esclarecen ambos la rapidez de convergencia y el efecto del sesgo. Además, nos intriga ver que, al parecer, existen puntos de cambio o intervalos de cambio, donde cambia el comportamiento de los estimadores. Para concluir, proponemos varios temas para futuras investigaciones.

Palabras clave: comparación de estimadores; efecto del sesgo; intervalo de cambio; rapidez de convergencia; simulación.

Mathematics Subject Classification: 60F05, 65C05, 65C50, 65C60.

1 Introduction

“There has been a sea change for the mathematics of probability: computer simulation is replacing theorems” [2, p. 222].

In a beginner’s statistics course, students learn about the sample mean and the sample median, and the fact that the sample median is robust or resistant, but the sample mean is not. Consequently, some students tend to see the sample median surrounded by an aura of sanctity. In praise of the sample mean, the students also learn the following nice property: as the sample size increases, the probability that the sample mean will differ from the population mean by less than an arbitrarily prescribed distance or tolerance tends to one. This is a somewhat rough statement of the law of large numbers, which is often illustrated by a graphic such as the one we see in [4, p. 288]. Unfortunately, such graphics might give students the impression that, for sample sizes 5000 and larger, it is almost sure that the sample mean is within a hairline of the population mean. Our results make it clear that, in general, this is not so. The law of large numbers is an intuitive, easy-to-understand statement. Indeed, it is comforting to know that, as the amount of information increases, so do the chances that the sample mean is arbitrarily near the value we are trying to estimate. However, the law of large numbers tells us nothing about the speed of convergence; our study throws some light on this matter.

In this paper, we are only concerned with continuous distributions that do have a mean and a unique median.

Generally, at the undergraduate level, little or nothing is said about the large-numbers behavior of the sample median, even in the standard mathematical-statistics course. An exception is Freund's book, where we find the following result [3, p. 324].

Theorem 1. *Suppose the population density* f_X *is continuous and nonzero at the population median* $\operatorname{med}(X)$*, which satisfies*

$$\int_{-\infty}^{\operatorname{med}(X)} f_X(x)\,dx = \frac{1}{2}.$$

Then, for large n*, the sampling distribution of the sample median for random samples of size* $2n+1$ *is approximately normal, with the mean* $\operatorname{med}(X)$ *and the variance*

$$\frac{1}{8\left\{f_X\left(\operatorname{med}\left(X\right)\right)\right\}^2 n}.$$

□

As Freund notes, the preceding theorem has the following consequence for a normal population f_X. In this case, the mean $E(X)$ and the median $\operatorname{med}(X)$ are equal and

$$f_X\left(E\left(X\right)\right) = f_X\left(\operatorname{med}\left(X\right)\right) = \frac{1}{\sqrt{2\pi\operatorname{Var}(X)}}.$$

So, although they do it differently, for a sample of size n, the sample mean $\bar{X}_n$ and the sample median $\widetilde{X}_n$ are estimating the same number. However, for large samples of size $2n+1$, we find that, approximately, the variances ratio

$$\frac{\operatorname{Var}(\widetilde{X}_{2n+1})}{\operatorname{Var}(\bar{X}_{2n+1})} = \frac{\pi\operatorname{Var}(X)/4n}{\operatorname{Var}(X)/(2n+1)} = \frac{\pi(2n+1)}{4n} \to \frac{\pi}{2}, \quad \text{as} \quad n\to\infty.$$

Thus, for large samples from a normal population, the sample mean is approximately $\pi/2 \approx 1.6$ times less variable —and therefore that much more dependable— than the sample median.

It is natural to ask about a law of large numbers for the sample median. It is also natural to ask, as the sample size n increases, how do the sample mean $\bar{X}_n$ and the sample median $\widetilde{X}_n$ compare as estimators of the respective corresponding parameters $E(X)$ and $\operatorname{med}(X)$. We investigate these matters empirically, through simulation. The notions of a consistent estimator and of convergence in probability are relevant to our study; we give more details in Section 4. We conclude with a few questions for further study.

For notation, we always use $E(X)$ and $\operatorname{med}(X)$ for the population mean and the population median, respectively. We do so because, as is generally done, we use μ and σ for the two lognormal-distribution parameters, respectively the population mean and the population standard deviation of the random variable $\ln(X)$. Thus, in this paper, μ and σ never stand for the population mean and the population standard deviation of a distribution we study.

Additional information about the median and order statistics in general is available in [1].

2 The simulation

In a symmetric distribution, the mean and the median are the same. Therefore, we only considered skewed distributions, all skewed to the right, and with pairwise different skewness coefficients ranging from 1.6 to 414.4. The *skewness coefficient* —*skewness*, for short— of a distribution f_X is defined by

$$\frac{\mu_3}{\{\operatorname{Var}(X)\}^{3/2}},$$

where μ_3 is the third moment about the mean of the random variable X.

The eight distributions we studied are listed in Table 1, along with some properties that are pertinent to our investigation. In a skewed distribution, the longer tail acts as a sort of magnet that strongly pulls the mean in that tail's direction; on the median, such magnetic effect is not so appreciable. Therefore, for each distribution in Table 1, we list the value of $|E(X) - \operatorname{med}(X)|$, which we regard as a rough, quick-and-dirty measure of skewness.

Table 1: Distributions in the study.

distribution	$E(X)$	$\text{med}(X)$	$\|E(X) - \text{med}(X)\|$	skewness
chi-square, df $= 3$	3	2.4	0.6	1.6
gamma, shape $= 1.05$ and rate $= 1$	1.05	0.74	0.31	1.95
gamma, shape $= 0.9$ and rate $= 1$	0.9	0.60	0.30	2.11
F, df1 $= 1$ and df2 $= 7$	1.4	0.5	0.9	14
lognormal, $\mu = 1$ and $\sigma = 1.5$	8.4	2.7	5.7	33.5
lognormal, $\mu = 1$ and $\sigma = 1.8$	13.7	2.7	11.0	136.4
lognormal, $\mu = 1$ and $\sigma = 1.9$	16.5	2.7	13.8	233.7
lognormal, $\mu = 1$ and $\sigma = 2$	20.1	2.7	17.4	414.4

The graphs of the distributions in the study are shown in Figure 1. In all cases, the range of values in the horizontal axis is the same. We do this to help facilitate good visual comparisons of the right tails between the different distributions.

We wanted to investigate the behavior of the estimators $\bar{X}_n$ and $\widetilde{X}_n$ in relation to the respective corresponding parameters $E(X)$ and $\text{med}(X)$, as the sample size increases. That is, we wanted to study the probabilities $P(|\bar{X}_n - E(X)| < \varepsilon)$ and $P(|\widetilde{X}_n - \text{med}(X)| < \varepsilon)$, for $\varepsilon > 0$, as $n \to \infty$. The law of large numbers tells us that, for arbitrary $\varepsilon > 0$, $P(|\bar{X}_n - E(X)| < \varepsilon) \to 1$, as $n \to \infty$. But we wanted actual numerical data; that is, we wished to know what happens along the way as $n \to \infty$. In particular, we wanted to compare $P(|\bar{X}_n - E(X)| < \varepsilon)$ and $P(|\widetilde{X}_n - \text{med}(X)| < \varepsilon)$ for different values of ε and n. We were also curious about how likely it is that the estimators stray toward the other, wrong parameter; therefore, we computed estimates for $P(|\bar{X}_n - \text{med}(X)| < \varepsilon)$ and $P(|\widetilde{X}_n - E(X)| < \varepsilon)$ as well. For the tolerance ε, we chose the values 0.1, 0.01, and 0.001, and for the sample size n we chose 25, 100, 500, 1000, 5000, and $10{,}000$.

The simulation was designed and performed to assure that, with 99% probability, each resulting estimate is within 0.005 of the true value. Thus, to compute each estimate, we used $66{,}564$ samples, the minimum required to satisfy the desired estimation specifications. To determine the number of samples, we assumed that $\frac{1}{2}$ is the true value of the proportion being estimated; in this context, that assumption is the most conservative possible supposition for the unknown parameter value. We did the simulation with the software R.

Figure 1: Distributions in the study.

Table 2: Probability estimates. Distribution: chi-square with 3 degrees of freedom; skewness = 1.6.

ε	n	$P(\lvert\bar{X}_n - E(X)\rvert < \varepsilon)$	$P(\lvert\widetilde{X}_n - \text{med}(X)\rvert < \varepsilon)$	$P(\lvert\bar{X}_n - \text{med}(X)\rvert < \varepsilon)$	$P(\lvert\widetilde{X}_n - E(X)\rvert < \varepsilon)$
0.1	25	0.1591	0.1483	0.0763	0.0700
	100	0.3159	0.2945	0.0097	0.0246
	500	0.6390	0.5990	0	0
	1000	0.8041	0.7613	0	0
	5000	0.9963	0.9921	0	0
	10000	0.99995	0.99980	0	0
0.01	25	0.0159	0.0149	0.0072	0.0073
	100	0.0341	0.0311	0.0007	0.0027
	500	0.0743	0.0687	0	0
	1000	0.1010	0.0928	0	0
	5000	0.2262	0.2110	0	0
	10000	0.3190	0.2950	0	0
0.001	25	0.0016	0.0013	0.0007	0.0007
	100	0.0032	0.0032	0.0001	0.0002
	500	0.0071	0.0069	0	0
	1000	0.0101	0.0101	0	0
	5000	0.0224	0.0214	0	0
	10000	0.0335	0.0288	0	0

Table 3: Probability estimates. Distribution: gamma with shape = 1.05 and rate = 1; skewness = 1.95.

ε	n	$P(\lvert\bar{X}_n - E(X)\rvert < \varepsilon)$	$P(\lvert\widetilde{X}_n - \text{med}(X)\rvert < \varepsilon)$	$P(\lvert\bar{X}_n - \text{med}(X)\rvert < \varepsilon)$	$P(\lvert\widetilde{X}_n - E(X)\rvert < \varepsilon)$
0.1	25	0.3729	0.3719	0.1400	0.1306
	100	0.6713	0.6702	0.0158	0.0308
	500	0.9708	0.9683	0	3.0×10^{-5}
	1000	0.9979	0.9974	0	0
	5000	1	1	0	0
	10000	1	1	0	0
0.01	25	0.0377	0.0379	0.0133	0.0123
	100	0.0773	0.0779	0.0004	0.0019
	500	0.1722	0.1690	0	0
	1000	0.2442	0.2424	0	0
	5000	0.5098	0.5053	0	0
	10000	0.6704	0.6644	0	0
0.001	25	0.0036	0.0035	0.0013	0.0015
	100	0.0083	0.0081	3.0×10^{-5}	0.0002
	500	0.0172	0.0169	0	0
	1000	0.0247	0.0244	0	0
	5000	0.0559	0.0539	0	0
	10000	0.0788	0.0775	0	0

Table 4: Probability estimates. Distribution: gamma with shape $= 0.9$ and rate $= 1$; skewness $= 2.11$.

ε	n	$P(\|\bar{X}_n - E(X)\| < \varepsilon)$	$P(\|\widetilde{X}_n - \text{med}(X)\| < \varepsilon)$	$P(\|\bar{X}_n - \text{med}(X)\| < \varepsilon)$	$P(\|\widetilde{X}_n - E(X)\| < \varepsilon)$
0.1	25	0.3990	0.4194	0.1319	0.1227
	100	0.7051	0.7245	0.0108	0.0232
	500	0.9815	0.9839	0	0
	1000	0.9991	0.9991	0	0
	5000	1	1	0	0
	10000	1	1	0	0
0.01	25	0.0405	0.0429	0.0119	0.0115
	100	0.0829	0.0869	0.0002	0.0012
	500	0.1831	0.1899	0	0
	1000	0.2594	0.2696	0	0
	5000	0.5458	0.5526	0	0
	10000	0.7124	0.7228	0	0
0.001	25	0.0044	0.0042	0.0013	0.0012
	100	0.0078	0.0091	0	0.0001
	500	0.0187	0.0197	0	0
	1000	0.0267	0.0273	0	0
	5000	0.0573	0.0611	0	0
	10000	0.0865	0.0867	0	0

Table 5: Probability estimates. Distribution: F with df $1 = 1$ and df $2 = 7$; skewness $= 14$.

ε	n	$P(\|\bar{X}_n - E(X)\| < \varepsilon)$	$P(\|\widetilde{X}_n - \text{med}(X)\| < \varepsilon)$	$P(\|\bar{X}_n - \text{med}(X)\| < \varepsilon)$	$P(\|\widetilde{X}_n - E(X)\| < \varepsilon)$
0.1	25	0.1560	0.3265	0.0176	0.0093
	100	0.2952	0.5987	0	0
	500	0.5932	0.9353	0	0
	1000	0.7499	0.9899	0	0
	5000	0.9880	1	0	0
	10000	0.9995	1	0	0
0.01	25	0.0158	0.0333	0.0016	0.0009
	100	0.0292	0.0668	0	0
	500	0.0663	0.1465	0	0
	1000	0.0933	0.2044	0	0
	5000	0.2018	0.4384	0	0
	10000	0.2823	0.5848	0	0
0.001	25	0.0016	0.0033	0.0001	9.0×10^{-5}
	100	0.0026	0.0060	0	0
	500	0.0069	0.0142	0	0
	1000	0.0091	0.0209	0	0
	5000	0.0197	0.0454	0	0
	10000	0.0297	0.0643	0	0

Table 6: Probability estimates. Distribution: lognormal with $\mu = 1$ and $\sigma = 1.5$; skewness $= 33.5$.

ε	n	$P(\|\bar{X}_n - E(X)\| < \varepsilon)$	$P(\|\widetilde{X}_n - \mathrm{med}(X)\| < \varepsilon)$	$P(\|\bar{X}_n - \mathrm{med}(X)\| < \varepsilon)$	$P(\|\widetilde{X}_n - E(X)\| < \varepsilon)$
0.1	25	0.0204	0.0788	0.0034	0.0003
	100	0.0387	0.1585	0	0
	500	0.0779	0.3378	0	0
	1000	0.1107	0.4635	0	0
	5000	0.2337	0.8337	0	0
	10000	0.3220	0.9499	0	0
0.01	25	0.0022	0.0078	0.0003	3.0×10^{-5}
	100	0.0036	0.0157	0	0
	500	0.0072	0.0354	0	0
	1000	0.0113	0.0491	0	0
	5000	0.0239	0.1112	0	0
	10000	0.0329	0.1541	0	0
0.001	25	0.0003	0.0008	1.5×10^{-5}	1.5×10^{-5}
	100	0.0004	0.0016	0	0
	500	0.0008	0.0037	0	0
	1000	0.0011	0.0049	0	0
	5000	0.0025	0.0112	0	0
	10000	0.0032	0.0149	0	0

Table 7: Probability estimates. Distribution: lognormal with $\mu = 1$ and $\sigma = 1.8$; skewness $= 136.4$.

ε	n	$P(\|\bar{X}_n - E(X)\| < \varepsilon)$	$P(\|\widetilde{X}_n - \mathrm{med}(X)\| < \varepsilon)$	$P(\|\bar{X}_n - \mathrm{med}(X)\| < \varepsilon)$	$P(\|\widetilde{X}_n - E(X)\| < \varepsilon)$
0.1	25	0.0095	0.0651	0.0012	1.5×10^{-5}
	100	0.0166	0.1319	0	0
	500	0.0314	0.2840	0	0
	1000	0.0440	0.3942	0	0
	5000	0.0901	0.7524	0	0
	10000	0.1234	0.8976	0	0
0.01	25	0.0010	0.0062	7.5×10^{-5}	0
	100	0.0016	0.0130	0	0
	500	0.0032	0.0297	0	0
	1000	0.0042	0.0411	0	0
	5000	0.0093	0.0925	0	0
	10000	0.0116	0.1289	0	0
0.001	25	0.0001	0.0006	0	0
	100	0.0001	0.0014	0	0
	500	0.0004	0.0029	0	0
	1000	0.0004	0.0040	0	0
	5000	0.0009	0.0093	0	0
	10000	0.0010	0.0126	0	0

Table 8: Probability estimates. Distribution: lognormal with $\mu = 1$ and $\sigma = 1.9$; skewness = 233.7

ε	n	$P(\lvert\bar{X}_n - E(X)\rvert < \varepsilon)$	$P(\lvert\widetilde{X}_n - \text{med}(X)\rvert < \varepsilon)$	$P(\lvert\bar{X}_n - \text{med}(X)\rvert < \varepsilon)$	$P(\lvert\widetilde{X}_n - E(X)\rvert < \varepsilon)$
0.1	25	0.0065	0.0618	0.0010	1.5×10^{-5}
	100	0.0113	0.1253	0	0
	500	0.0229	0.2700	0	0
	1000	0.0312	0.3755	0	0
	5000	0.0640	0.7273	0	0
	10000	0.0884	0.8788	0	0
0.01	25	0.0006	0.0059	0.0001	1.5×10^{-5}
	100	0.0011	0.0125	0	0
	500	0.0022	0.0283	0	0
	1000	0.0032	0.0389	0	0
	5000	0.0062	0.0874	0	0
	10000	0.0084	0.1225	0	0
0.001	25	3.0×10^{-5}	0.0006	4.5×10^{-5}	0
	100	9.0×10^{-5}	0.0013	0	0
	500	0.0002	0.0027	0	0
	1000	0.0004	0.0039	0	0
	5000	0.0006	0.0089	0	0
	10000	0.0008	0.0119	0	0

Table 9: Probability estimates. Distribution: lognormal with $\mu = 1$ and $\sigma = 2$; skewness $= 414.4$.

ε	n	$P(\lvert\bar{X}_n - E(X)\rvert < \varepsilon)$	$P(\lvert\widetilde{X}_n - \text{med}(X)\rvert < \varepsilon)$	$P(\lvert\bar{X}_n - \text{med}(X)\rvert < \varepsilon)$	$P(\lvert\widetilde{X}_n - E(X)\rvert < \varepsilon)$
0.1	25	0.0048	0.0590	0.0009	0
	100	0.0078	0.1193	0	0
	500	0.0171	0.2564	0	0
	1000	0.0230	0.3586	0	0
	5000	0.0438	0.7024	0	0
	10000	0.0622	0.8589	0	0
0.01	25	0.0004	0.0056	0.0001	0
	100	0.0009	0.0119	0	0
	500	0.0017	0.0271	0	0
	1000	0.0025	0.0368	0	0
	5000	0.0040	0.0833	0	0
	10000	0.0059	0.1166	0	0
0.001	25	4.5×10^{-5}	0.0006	3.0×10^{-5}	0
	100	7.5×10^{-5}	0.0013	0	0
	500	0.0001	0.0025	0	0
	1000	0.0003	0.0037	0	0
	5000	0.0004	0.0083	0	0
	10000	0.0007	0.0113	0	0

Table 10: Difference of the probability estimates as the skewness increases.

		$P(\lvert\widetilde{X}_n - \mathrm{med}(X)\rvert < \varepsilon) - P(\lvert\bar{X}_n - E(X)\rvert < \varepsilon)$					
ε	skewness	$n = 25$	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$	$n = 10{,}000$
0.1	1.6	-0.0108	-0.0214	-0.0400	-0.0428	-0.0042	-0.00015
	1.95	-0.0010	-0.0011	-0.0025	-0.0005	0	0
	2.11	0.0204	0.0194	0.0024	0	0	0
	14	0.1705	0.3035	0.3421	0.2400	0.0120	0.0005
	33.5	0.0584	0.1198	0.2599	0.3528	0.6000	0.6279
	136.4	0.0556	0.1153	0.2526	0.3502	0.6623	0.7742
	233.7	0.0553	0.1140	0.2471	0.3443	0.6633	0.7904
	414.4	0.0542	0.1115	0.2393	0.3356	0.6586	0.7967
0.01	1.6	-0.0010	-0.0030	-0.0056	-0.0082	-0.0152	-0.0240
	1.95	0.0002	0.0006	-0.0032	-0.0018	-0.0045	-0.0060
	2.11	0.0024	0.0040	0.0068	0.0102	0.0068	0.0104
	14	0.0175	0.0376	0.0802	0.1111	0.2366	0.3025
	33.5	0.0056	0.0121	0.0282	0.0378	0.0873	0.1212
	136.4	0.0052	0.0114	0.0265	0.0369	0.0832	0.1173
	233.7	0.0053	0.0114	0.0261	0.0357	0.0812	0.1141
	414.4	0.0052	0.0110	0.0254	0.0343	0.0793	0.1107
0.001	1.6	-0.0003	0.0000	-0.0002	0.0000	-0.0010	-0.0047
	1.95	-0.0001	-0.0002	-0.0003	-0.0003	-0.0020	-0.0013
	2.11	-0.0002	0.0013	0.0010	0.0006	0.0038	0.0002
	14	0.0017	0.0034	0.0073	0.0118	0.0257	0.0346
	33.5	0.0005	0.0012	0.0029	0.0038	0.0087	0.0117
	136.4	0.0005	0.0013	0.0025	0.0036	0.0084	0.0116
	233.7	0.00057	0.00121	0.0025	0.0035	0.0083	0.0111
	414.4	0.000555	0.001225	0.0024	0.0034	0.0079	0.0106

3 Results and conclusions

The probability estimates resulting from the simulation were rounded to the number of decimal places shown in Tables 2 through 9; however, a table entry is "1" or "0" if and only if that's what `R` returned.

Because of the sample-median's robustness, we were interested in the behavior of the difference $P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon) - P(|\bar{X}_n - E(X)| < \varepsilon)$ as the skewness increases. Therefore, we used the simulation results to investigate this difference; see Table 10, whose entries were computed using the estimates in Tables 2 through 9.

Among all these results, we find that a couple are as anticipated, while others are not a priori evident. From these tables, we note some observations. The following results are numbered merely to facilitate the referencing in this section and in Section 4; in particular, such numbering is not intended to convey rank or importance.

1. As the tolerance ε decreases, and for the same sample size n, the estimated probabilities also decrease. This is an unsurprising, anticipated conclusion.

2. As the sample size increases, and for the same tolerance ε, the estimated probabilities also increase. This is another unsurprising conclusion.

3. The skewness appears to affect and determine the order relation between the probabilities $P(|\bar{X}_n - E(X)| < \varepsilon)$ and $P(|\widetilde{X}_n - \text{med}(X)| < \varepsilon)$.

   - When the skewness is 1.6, the estimates in Table 2 show that, for the same ε and n, it is always true that

     $$P(|\bar{X}_n - E(X)| < \varepsilon) > P(|\widetilde{X}_n - \text{med}(X)| < \varepsilon), \qquad (1)$$

     except in a couple of instances when $\varepsilon = 0.001$. However, we note that the difference between these two probabilities decreases as the sample size increases from 5000 to $10{,}000$.

   - When the skewness is 14 or larger, the estimates in Tables 5–9 show that, for the same ε and n, the order relation in (1) is reversed and then we always have

     $$P(|\bar{X}_n - E(X)| < \varepsilon) < P(|\widetilde{X}_n - \text{med}(X)| < \varepsilon).$$

     In particular, in Table 9, which pertains to the largest skewness we consider, it appears that $P(|\widetilde{X}_n - \text{med}(X)| < \varepsilon)$ is always substantially larger than the corresponding $P(|\bar{X}_n - E(X)| < \varepsilon)$.

   - When the skewness is 1.95, the estimates in Table 3 show that, for the same ε and n,

     $$P(|\bar{X}_n - E(X)| < \varepsilon) > P(|\widetilde{X}_n - \text{med}(X)| < \varepsilon), \qquad (2)$$

     nearly always. However, when the skewness increases to 2.11, the estimates in Table 4 show that the opposite is true, with the order relation in (2) reversed and, generally,

     $$P(|\bar{X}_n - E(X)| < \varepsilon) < P(|\widetilde{X}_n - \text{med}(X)| < \varepsilon).$$

4. The results in the previous item number 3 suggest the existence of a *changepoint* or a *change-interval* near the skewness value of 2. That is, such results suggest the existence of real numbers a and b such that $a < 2 < b$ and:

- if the skewness is less than a, then

$$P(|\bar{X}_n - E(X)| < \varepsilon) > P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon);$$

- if the skewness is between a and b, then we have both

$$\begin{aligned} &P(|\bar{X}_n - E(X)| < \varepsilon) > P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon) \\ \text{and}\quad &P(|\bar{X}_n - E(X)| < \varepsilon) < P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon), \end{aligned}$$

  depending upon ε or n;

- and if the skewness is larger than b, then

$$P(|\bar{X}_n - E(X)| < \varepsilon) < P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon).$$

  Such change-interval is an unanticipated, intriguing, and interesting finding.

5. For $\varepsilon := 0.1$, as $n \to \infty$, the convergence of the probabilities

$$P(|\bar{X}_n - E(X)| < \varepsilon) \to 1 \quad \text{and} \quad P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon) \to 1$$

   is well supported by the estimates in Tables 2–5. However, and also for $\varepsilon := 0.1$, as the skewness increases, the data in Tables 6–9 illustrate the convergence well for $P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon)$, but, in comparison, $P(|\bar{X}_n - E(X)| < \varepsilon)$ appears to converge much more slowly.

6. When $\varepsilon := 0.01$ or $\varepsilon := 0.001$, as $n \to \infty$, the estimates in Tables 2–4 suggest that $P(|\bar{X}_n - E(X)| < \varepsilon)$ and $P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon)$ have nearly equal convergence rates. However, and also for $\varepsilon := 0.01$ or $\varepsilon := 0.001$, as the skewness increases, the data in Tables 5–9 suggest that the convergence rate for $P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon)$ is faster than that of $P(|\bar{X}_n - E(X)| < \varepsilon)$.

7. When $\varepsilon := 0.01$ or $\varepsilon := 0.001$, and the sample size $n = 10{,}000$, both estimated probabilities $P(|\bar{X}_n - E(X)| < \varepsilon)$ and $P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon)$ in Tables 2–9 are often still quite small. This suggests that, as $n \to \infty$, these probabilities converge slowly when $\varepsilon := 0.01$ or smaller.

8. Overall, $\bar{X}_n$ and $\widetilde{X}_n$ appear faithful to their purpose as estimators of the respective parameters $E(X)$ and $\mathrm{med}(X)$. Indeed, the estimated probabilities for $P(|\bar{X}_n - \mathrm{med}(X)| < \varepsilon)$ and $P(|\widetilde{X}_n - E(X)| < \varepsilon)$ that $\bar{X}_n$ and $\widetilde{X}_n$ stray toward the other, wrong parameter are nearly always equal to 0.

9. Table 10 shows the difference

$$P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon) - P(|\bar{X}_n - E(X)| < \varepsilon)$$

of the probability estimates. From that table we note three items.

- A glance at the table reveals that most entries are positive. This means that in our study we generally find that

$$P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon) > P(|\bar{X}_n - E(X)| < \varepsilon),$$

which we attribute to the sample-median's robustness, since we only consider skewed distributions.

- In general, looking across each table row, we find that, as the sample size increases, the entries increase in magnitude. Moreover, separately for each value of ε, such increase is more pronounced when the skewness is 33.5 or larger.
- Looking down the table's columns, and separately for each value of ε and n, we find that the entries stabilize and become nearly constant when the skewness is 33.5 or larger. Moreover, for such skewness values, and separately for each n, the entries decrease proportionally to the decrease in ε from 0.1 to 0.01 to 0.001. Such entries appear to vary regularly enough that they would seem to be predictable with reasonable certainty in terms of ε, n, and the skewness.

The preceding observations suggest the existence of a changepoint or a change-interval (c, d) so that the differences

$$P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon) - P(|\bar{X}_n - E(X)| < \varepsilon)$$

become nearly constant when the skewness is larger than d, and separately for each ε and n. From our results, it seems that the value of d is somewhere between 14 and 33.5. Of course, such change-interval is only interesting when the values of ε, n, and the skewness are within a certain range, to be determined by further research; see Section 4.

This change-interval is another unanticipated, intriguing, and interesting finding.

4 Further study

For completeness, we recall a theorem and two definitions that are relevant to our investigation; these three items are related.

Theorem 2 (Weak law of large numbers). *Let* $(Y_k)_{k=1}^{\infty}$ *be a sequence of independent and identically distributed random variables whose mean* $\theta := E(Y_k)$ *exists. Then, for all* $\varepsilon > 0$*,*

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} Y_k - \theta\right| < \varepsilon\right) = 1. \qquad \square$$

The sequence of random variables $(Y_n)_{n=1}^{\infty}$ *converges in probability* to the random variable Y if, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|Y_n - Y| < \varepsilon) = 1.$$

The sequence of estimators $(Y_n)_{n=1}^{\infty}$ is *consistent* for the parameter θ if, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|Y_n - \theta| < \varepsilon) = 1.$$

As mentioned earlier, we assume that the population mean exists and that the population median is unique. Then the sample mean $\bar{X}_n$ and the sample median $\widetilde{X}_n$ converge in probability to—and are consistent for—the respective parameters $E(X)$ and $\mathrm{med}(X)$.

Following are some questions that we propose for further study.

- In general, does the change-interval (a, b) mentioned in item number 4 of Section 3 exist? Is it possible to express the endpoints a and b as functions of ε, n, and the skewness? If so, how?

  As both $\varepsilon \to 0$ and $n \to 0$, it is reasonable to believe that

  $$P(|\bar{X}_n - E(X)| < \varepsilon) \to 0 \quad \text{and} \quad P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon) \to 0.$$

  Therefore, when both ε and n are very small, we anticipate that

  $$P(|\bar{X}_n - E(X)| < \varepsilon) \approx P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon) \approx 0,$$

  and then such notion of a change-interval is not interesting.

- In items 5, 6, and 7 of Section 3, we made some comments regarding

  $$\lim_{n\to\infty} P(|\bar{X}_n - E(X)| < \varepsilon) = 1 \text{ and } \lim_{n\to\infty} P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon) = 1.$$

  The speed of convergence is always an interesting issue which, however, appears to have been overlooked for the limits we study. Therefore, we ask: Given $\varepsilon > 0$, what can be said about the convergence rate for the limits

$$\lim_{n\to\infty} P(|\bar{X}_n - E(X)| < \varepsilon) = 1 \text{ and } \lim_{n\to\infty} P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon) = 1?$$

Our results suggest that, in general, such convergence rate is slow. In particular, is it possible to express the convergence rate explicitly in terms of the distribution's skewness? And, if so, how?

- Can the stabilized, nearly constant entries in Table 10 be predicted, as mentioned in item number 9 of Section 3? If so, how?
- In general, does the change-interval (c, d) mentioned in item number 9 of Section 3 exist? Is it possible to express the endpoints c and d as functions of ε, n, and the skewness? If so, how?

  Since both $\bar{X}_n$ and $\widetilde{X}_n$ are consistent for their respective parameters $E(X)$ and $\mathrm{med}(X)$, for each $\varepsilon > 0$ and skewness value, there must exist a natural number N such that, for all $n \geq N$,

$$P(|\widetilde{X}_n - \mathrm{med}(X)| < \varepsilon) - P(|\bar{X}_n - E(X)| < \varepsilon) \approx 0.$$

  Therefore, when the sample size n is greater than or equal to such N, the notion of a change-interval is not interesting in this context.

Acknowledgements

The author thanks the referees for their comments.

References

[1] David, H.A. (1981) *Order Statistics*, 2nd Edition. Wiley, New York.

[2] Diaconis, P.; Skyrms, B. (2018) *Ten Great Ideas About Chance*. Princeton University Press, New Jersey.

[3] Freund, J.E. (1992) *Mathematical Statistics*, 5th Edition. Prentice-Hall College Div, New Jersey.

[4] Moore, D.S.; Notz, W.I.; Fligner, M.A. (2013) *The Basic Practice of Statistics*, 6th Edition. W.H. Freeman, New York.