Scientific
Research

Distribution of the Median in Samples from the Laplace Distribution[*]

John Lawrence

US Food and Drug Administration, Silver Spring, USA
Email: John.Lawrence@fda.hhs.gov

ABSTRACT

The Laplace distribution is one of the oldest defined and studied distributions. In the one-parameter model (location parameter only), the sample median is the maximum likelihood estimator and is asymptotically efficient. Approximations for the variance of the sample median for small to moderate sample sizes have been studied, but no exact formula has been published. In this article, we provide an exact formula for the probability density function of the median and an exact formula for the variance of the median.

Keywords: Quantile; Generalized Hypergeometric Function

1. Introduction

Laplace is credited as first discovering the Laplace distribution in 1774 [1]. The discovery of the normal distribution, which holds a central place in many applications of statistics, predates that of the Laplace distribution. It was discovered in 1738 by de Moivre according to [2]. Gauss and Laplace were contemporaries who both made significant discoveries about the normal distribution. Keynes [3] wrote "The popularity of the normal law, with the arithmetic mean and the method of least squares as its corollaries, has been very largely due to its overwhelming advantages, in comparison with all other laws of error, for the purposes of mathematical development and manipulation... So powerful a hold indeed did the normal law obtain on the minds of statisticians, that until quite recent times only a few pioneers have seriously considered the possibility of preferring in certain circumstances other means to the arithmetic and other laws of error to the normal. Laplace's earlier memoir fell, therefore, out of remembrance." Over the past 100 years, the Laplace distribution has enjoyed the resurgence in many applications such as economics, engineering and finance [4]. The reader is also referred to this book for a history of the Laplace distribution, its important properties and generalizations.

Let $f_x(x;\mu) = \exp\left(-\sqrt{2}|x-\mu|\right)/\sqrt{2}$ denote the probability density function of a one-parameter Laplace distribution with variance 1. It is well known that from a sample of odd size $2n + 1$ where n is any positive integer, the sample median, $\tilde{\mu}$, is the maximum likelihood estimate of μ. Moreover, although the derivative does not exist at μ, the Cramer-Rao lower bound for the mean squared error of estimation of μ exists and is $1/(2(2n+1))$. Also, we have $\sqrt{2(2n+1)}(\tilde{\mu}-\mu) \to N(0,1)$ which implies the median is an efficient estimator in the asymptotic sense. However, $\tilde{\mu}$ is not a sufficient statistic and there are more efficient estimators for small samples, for example the average of the middle three observations can be more efficient. Finally, the sample mean is also asymptotically normal and asymptotically the ratio of the variance of the mean to the median is 2.

Previously, [5] investigated approximations to the variance of $\tilde{\mu}$ for small sample sizes. They confirmed that although the variance converges to the Cramer-Rao lower bound, it does so at a slow rate. Consequently, even if the normal approximation was valid for small sample sizes, hypothesis tests and confidence intervals would not have the correct size if the asymptotic variance was used. The former provided the bounds for the ratio of the variance to the Cramer-Rao lower bound as fol-

---

[*]The views expressed are those of the authors and not necessarily those of the US Food and Drug Administration.

lows:

$$B_n \left(1 - \frac{1}{2n+2}\right)^{3/2} \le 2(2n+1)\operatorname{Var}\{\tilde{\mu}\}$$

$$\le 1.51 B_n \left(1 + \frac{1}{2n}\right)^{3/2}$$

where $B_n = \frac{(2n+1)!\sqrt{2\pi}}{n!n!2^{2n+1}\sqrt{2n+1}} \sim 1 + \frac{1}{8n}$. For example, if $n = 499$ and the sample size were 999, the upper bound would be approximately 1.513. This suffices for their purpose, which was mainly to show that this ratio was less than 2, which in turn implies the median is a more efficient estimator than the sample mean. However, it seems like an inadequate upper bound since we know the ratio converges to 1. [6] attempted to tabulate the approximate variance for $n = 1, 2, 3, 4$, and 8. We found that in general his approximations were close to the true values (to two digits), but were not correct in two out of these five cases (see Section 3).

This paper makes two contributions. First, we find an explicit formula for the distribution of the median of a random sample from the Laplace distribution. Next, we find an exact formula for the variance of the sample median. These contributions are important because the median is the essential estimate of location for the Laplace distribution in the same way that the mean is the essential estimate of location for the normal distribution.

2. Exact Distribution of Sample Median

Without loss of generality, we will consider the case where $\mu = 0$. The probability the median is less than x is the probability that at least n + 1 values in the sample are less than x. The distribution of the median is symmetric about 0, so it suffices to consider values of $x < 0$. Then, we have

$$F_{\tilde{\mu},n}(x) \overset{\text{def}}{=} P[\tilde{\mu} < x]$$

$$= \sum_{i=n+1}^{2n+1} \binom{2n+1}{i} \left(\frac{1}{2}e^{\sqrt{2}x}\right)^i \left(1 - \frac{1}{2}e^{\sqrt{2}x}\right)^{2n+1-i}$$

In the Appendix, we show by mathematical induction that the density for $x < 0$ is

$$f_{\tilde{\mu},n}(x) = c_n e^{(n+1)\sqrt{2}x} \left(1 - \frac{1}{2}e^{\sqrt{2}x}\right)^n$$

where the constants are

$$c_n = 2^{n+\frac{1}{2}} \frac{\Gamma\left(n+\frac{3}{2}\right)}{\sqrt{\pi}\Gamma(n+1)} = (n+1)2^{-n-\frac{1}{2}}\binom{2n+1}{n}.$$

Hence, $\frac{1}{2}e^{-\sqrt{2}|\tilde{\mu}|}$ has a Beta distribution truncated on the interval $(0, 1/2)$.

The characteristic function is

$$c_n 2^{n+\frac{3}{2}+\frac{it}{\sqrt{2}}} B\left(\frac{1}{2}; n+1+\frac{t}{\sqrt{2}}i, n+1\right)$$

and the moment generating function is

$$c_n 2^{n+\frac{3}{2}+\frac{t}{\sqrt{2}}} B\left(\frac{1}{2}; n+1+\frac{t}{\sqrt{2}}, n+1\right).$$

In both cases, the formulas involve the incomplete beta function

$$B\left(\frac{1}{2}; a, b\right) = \int_0^{1/2} u^{a-1}(1-u)^{b-1}\,du.$$

3. Variance of the Median

The variance can be calculated exactly using the Binomial expansion as

$$2\int_{-\infty}^0 x^2 c_n e^{(n+1)\sqrt{2}x}\left(1 - \frac{1}{2}e^{\sqrt{2}x}\right)^n dx$$

$$= 2c_n \sum_{i=0}^n \left(-\frac{1}{2}\right)^i \binom{n}{i} \int_{-\infty}^0 x^2 e^{(n+i+1)\sqrt{2}x} dx$$

$$= 2c_n \sum_{i=0}^n \left(-\frac{1}{2}\right)^i \binom{n}{i} \frac{1}{\sqrt{2}(n+i+1)^3}$$

This variance also evaluates to

$$\sqrt{2}c_n \frac{{}_pF_q\left(\{-n, n+1, n+1\}, \{n+2, n+2, n+2\}, \frac{1}{2}\right)}{(n+1)^3}$$

where ${}_pF_q$ is the generalized Hypergeometric function.

For $n = 1, 2, 3, 4$, and 8, the ratios of the variance of the median to the variance of the mean (rounded to 6 digits) are 0.958333, 0.877951, 0.824767, 0.787808, and 0.708761. In comparison to the approximations listed in the bottom row of Siddiqui, these are close in all cases. In the case of $n = 3$, (corresponding to a sample size of $N = 7$ in his notation) he gives the approximate value of 0.81, but it should be 0.82 rounded to 2 digits. Also, in the case of $n = 4$, he gives the approximation 0.78, while it should be 0.79.

As a further example, consider the case $n = 499$, corresponding to a sample size of 999. This would be a rather large dataset by most standards and one might think the asymptotic variance would be accurate. However, the ratio of the actual variance to the asymptotic value is 1.05122, or about 5% larger than the Cramer-Rao lower bound. Although not very close to 1, it is substantially closer than the upper bound of 1.513 provided by the Chu and Hotelling approximation (see Section 1).

For very large values of n, it will still not be possible

to calculate the variance exactly using the formula at the beginning of this Section. One way to estimate the ratio of the true variance compared to the asymptotic variance is

$$\frac{2\int_{-\infty}^{0} x^2 c_n e^{(n+1)\sqrt{2}x}\left(1-\frac{1}{2}e^{\sqrt{2}x}\right)^n dx}{\left(\frac{1}{2(2n+1)}\right)}$$

$$=\frac{\sqrt{2}}{\sqrt{2n+1}}\int_{-\infty}^{0} x^2 c_n e^{(n+1)\frac{x}{\sqrt{2n+1}}}\left(1-\frac{1}{2}e^{\frac{x}{\sqrt{2n+1}}}\right)^n dx$$

$$=\frac{\sqrt{2}}{\sqrt{2n+1}}\int_{-\infty}^{0} x^2 \frac{c_n e^{(n+1)\frac{x}{\sqrt{2n+1}}}\left(1-\frac{1}{2}e^{\frac{x}{\sqrt{2n+1}}}\right)^n}{\phi(x)}d\Phi(x)$$

$$=\frac{E\left\{Z^2 \frac{c_n e^{-(n+1)\frac{|Z|}{\sqrt{2n+1}}}\left(1-\frac{1}{2}e^{\frac{-|Z|}{\sqrt{2n+1}}}\right)^n}{\phi(Z)}\right\}}{\sqrt{2(2n+1)}}$$

where Z is standard normal with distribution function $\Phi(x)$ and density function $\phi(x)$. The numerator can be estimated by Monte Carlo simulation or the integral in the second to the last line can be estimated by numerical integration such as the Riemann sum

$$\frac{1}{N}\sum_{i=1}^{N} h\left(\Phi^{-1}\left(\frac{N-0.5}{2N}\right)\right)$$

where $h(\cdot)$ denotes the integrand and N is a large integer.

It is also possible to sample values having the density $f_{\tilde{\mu}}(x)$ by using acceptance-rejection sampling. To do this, one needs a candidate distribution with density $g(x)$ such that $\frac{f_{\tilde{\mu}}(x)}{g(x)}$ is bounded by a fixed upper bound M. In addition, it is best if M is close to 1 since the proportion of candidate values that will be accepted is $1/M$. The Laplace distribution with variance 2 is a very good candidate distribution to sample from. Note that the ratio

$$\frac{\frac{1}{\sqrt{2(2n+1)}}c_n e^{-(n+1)\frac{|x|}{\sqrt{2n+1}}}\left(1-\frac{1}{2}e^{\frac{-|x|}{\sqrt{2n+1}}}\right)^n}{\frac{e^{-|x|}}{2}}$$

attains its maximum at

$$x=\pm\sqrt{2n+1}\log\left\{\frac{2n+\frac{1}{\sqrt{2n+1}}-\sqrt{2n+1}}{2n}\right\}$$

and that maximum is

$$M=\frac{\sqrt{2}c_n\left(\frac{n}{2n+1}\right)^n\left(\sqrt{2n+1}-1\right)}{n\left(1-\frac{1}{\sqrt{2n+1}}\right)^{\sqrt{2n+1}}}<\sqrt{\frac{2e}{\pi}}.$$

4. Conjecture about the Rate of Convergence

Start with values of n that were the closest integers to $\exp(3),\exp(4),\cdots,\exp(20)$.

For each of those values of n, calculate the ratio given in the previous Section via the Riemann sum described in Section 4 with $N=10$ million. In **Figure 1**, we show the values of $\log(n)$ on the x-axis and log(ratio -1) on the y-axis where the ratio means the ratio of the actual variance to the asymptotic variance computed by numerical integration. We noticed that the points fell very close to the straight line defined by

$$y=\log(1.14)-0.5\log(n).$$

Hence, at least in this range of n, the ratio is approximately $1+1.14/\sqrt{n}$. Note, the largest n here is approximately 500 million, which corresponds to a sample size of about 1 billion.

This leads to a conjecture about the rate of convergence of the variance. In addition, from the formula near the beginning of Section 3, this is also a conjecture about the asymptotic behavior of the Hypergeometric function for large n. Specifically,

$$_pF_q\left(\{-n,n+1,n+1\},\{n+2,n+2,n+2\},\frac{1}{2}\right)$$

$$\sim \frac{\sqrt{\pi}}{2^{n+1}}n^{3/2}\left(1+\frac{1.14}{\sqrt{2n}}\right)$$

using Stirling's approximation $c_n \sim 2^n\sqrt{\frac{2n}{\pi}}$.

5. Summary and Conclusion

We found an exact formula for the density and the variance of the one-parameter Laplace distribution. The moment generating function and characteristic function are also derived. We also describe several methods for approximating moments or other features of the distribution by an algorithm to sample directly from the distribution. This article only deals with the case of odd sample sizes. The median is the $(n+1)^{th}$ order statistic and the same
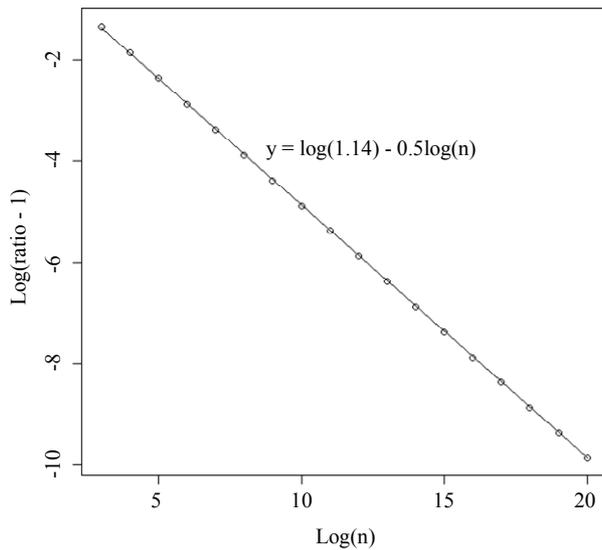
Figure 1. Transformed values of ratio (actual variance of the median divided by asymptotic variance estimated by numerical integration) for different large values of *n*.

approach could be used to find the distribution of any other order statistic. For even sample sizes, the median is the average of the two middle observations, which makes it slightly more complicated to analyze because the joint distribution of two order statistics is needed. However, a similar approach used here may also handle the even sample size case. Furthermore, this approach could also be useful in analyzing the variance of other estimators such as the average of the middle three or middle five and observations. And it could be helpful in finding the optimal estimator for small sample sizes. This exact variance may be useful in constructing approximate confi-

dence intervals or hypothesis tests. But, caution should still be used in using the normal approximation. Exact tests and confidence intervals can be constructed from the exact distribution. Lastly, in the two-parameter Laplace model, not considered here, a further adjustment may be needed in such procedures due to the estimation of the scale parameter.

REFERENCES

[1] P. S. Laplace, "Mémoire sur la Probabilité des Causes par les Évènemens," *Mémoires de Mathematique et de Physique*, *Presentés à l'Académie Royale des Sciences*, *Par Divers Savans & Lûs Dans ses Assemblées*, *Tome Sixième*, 1774, pp. 621-656.

[2] N. L. Johnson, S. Kotz and N. Balakrishnan, "Continuous Univariate Distributions," Vol. 1, Wiley, Hoboken, 1994.

[3] J. M. Keynes, "The Principal Averages and the Laws of Error Which Lead to Them," *Journal of the Royal Statistical Society*, Vol. 74, No. 3, 1911, pp. 322-331. http://dx.doi.org/10.2307/2340444

[4] S. Kotz,, T. J. Kozubowski and K. Podgorski, "The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance (No. 183)," Springer, Berlin, 2001. http://dx.doi.org/10.1007/978-1-4612-0173-1

[5] J. T. Chu and & H. Hotelling, "The Moments of the Sample Median," *The Annals of Mathematical Statistics*, Vol. 26, No. 4, 1955, pp. 593-606. http://dx.doi.org/10.1214/aoms/1177728419

[6] M. M. Siddiqui, "Approximations to the Moments of the Sample Median," *The Annals of Mathematical Statistics*, Vol. 33, No. 1, 1962, pp. 157-168. http://dx.doi.org/10.1214/aoms/1177704720

Appendix

First, note that for any positive integers m and i with $i < m$,

$$\left\{\left(\frac{1}{2}e^{\sqrt{2}x}\right)^i\left(1-\frac{1}{2}e^{\sqrt{2}x}\right)^{m-i}\right\}'$$

$$=\left(\frac{1}{2}e^{\sqrt{2}x}\right)^i\left(1-\frac{1}{2}e^{\sqrt{2}x}\right)^{m-i-1}\sqrt{2}\left(i-m\frac{1}{2}e^{\sqrt{2}x}\right)$$

when $n = 1$,

$$F_{\tilde{\mu},1}(x)=3\left(\frac{1}{2}e^{\sqrt{2}x}\right)^2\left(1-\frac{1}{2}e^{\sqrt{2}x}\right)+\left(\frac{1}{2}e^{\sqrt{2}x}\right)^3$$

and

$$f_{\tilde{\mu},1}(x)=\left\{F_{\tilde{\mu},1}(x)\right\}'=c_1\left(\frac{1}{2}e^{\sqrt{2}x}\right)^2\left(1-\frac{1}{2}e^{\sqrt{2}x}\right)$$

We suppose that the formula for the density is true for some positive integer n.

Consider the random sample of size $2(n+1)+1$ and recalculate the distribution function for the median of this random sample by conditioning on the number of values among the first $2n+1$ elements in the sample that are less than x. If the median of the entire sample is less than x, then there are three possible cases for such numbers: exactly n, exactly $n+1$, or at least $n+2$. We have

$$F_{\tilde{\mu},n+1}(x)=P[\text{exactly }n\text{ among }X_1,\cdots,X_{2n+1}<x]P[X_{2n+2}<x \text{ and }X_{2n+3}<x]$$

$$+P[\text{exactly }n+1\text{ of }X_1,\cdots,X_{2n+1}<x]P[X_{2n+2}<x \text{ or }X_{2n+3}<x]$$

$$+P[n+2\text{ or more of }X_1,\cdots,X_{2n+1}<x]$$

$$=\binom{2n+1}{n}\left(\frac{1}{2}e^{\sqrt{2}x}\right)^n\left(1-\frac{1}{2}e^{\sqrt{2}x}\right)^{n+1}\left(\frac{1}{2}e^{\sqrt{2}x}\right)^2$$

$$+\binom{2n+1}{n+1}\left(\frac{1}{2}e^{\sqrt{2}x}\right)^{n+1}\left(1-\frac{1}{2}e^{\sqrt{2}x}\right)^n\left\{e^{\sqrt{2}x}\left(1-\frac{1}{2}e^{\sqrt{2}x}\right)+\left(\frac{1}{2}e^{\sqrt{2}x}\right)^2\right\}$$

$$+P[n+1\text{ or more of first }2n+1<x]$$

$$-P[\text{exactly }n+1\text{ of first }2n+1<x]$$

$$=\binom{2n+1}{n}\left(\frac{e^{\sqrt{2}x}}{2}\right)^{n+1}\left(1-\frac{e^{\sqrt{2}x}}{2}\right)^{n+1}\left(1-\frac{e^{\sqrt{2}x}}{2}\right)^{n+1}\left(e^{\sqrt{2}x}-1\right)+F_{\tilde{\mu},n}(x)$$

The derivative is

$$-\sqrt{2}\binom{2n+1}{n}\left(\frac{e^{\sqrt{2}x}}{2}\right)^{n+1}\left(1-\frac{e^{\sqrt{2}x}}{2}\right)^n\left\{1+n-(2n+3)e^{\sqrt{2}x}+\frac{2n+3}{2}e^{2\sqrt{2}x}\right\}+f_{\tilde{\mu},n}(x)$$

and by the induction hypothesis, this equals

$$-\sqrt{2}\binom{2n+1}{n}\left(\frac{e^{\sqrt{2}x}}{2}\right)^{n+1}\left(1-\frac{e^{\sqrt{2}x}}{2}\right)^n\left\{1+n-(2n+3)e^{\sqrt{2}x}+\frac{2n+3}{2}e^{2\sqrt{2}x}\right\}+(n+1)2^{-n-\frac{1}{2}}\binom{2n+1}{n}e^{(n+1)\sqrt{2}x}\left(1-\frac{e^{\sqrt{2}x}}{2}\right)^n$$

$$=\binom{2n+1}{n}2^{-n-1/2}\left(1-\frac{e^{\sqrt{2}x}}{2}\right)^n\left(e^{\sqrt{2}x}\right)^{n+1}\left\{(2n+3)e^{\sqrt{2}x}-\frac{2n+3}{2}e^{2\sqrt{2}x}\right\}=(2n+3)\binom{2n+1}{n}2^{-n-\frac{1}{2}}\left(1-\frac{1}{2}e^{\sqrt{2}x}\right)^{n+1}\left(e^{\sqrt{2}x}\right)^{n+2}$$

Finally, notice that

$$(2n+3)\binom{2n+1}{n}2^{-n-\frac{1}{2}}=\frac{(2n+3)(2n+1)!}{n!(n+1)!}2^{-n-\frac{1}{2}}=\frac{(2n+3)(2n+1)!(2n+2)(n+2)}{n!(n+1)!2(n+1)(n+2)}2^{-n-\frac{1}{2}}$$

$$=\frac{(2n+3)!(n+2)}{(n+2)!(n+1)!}2^{-(n+1)-\frac{1}{2}}=\binom{2n+3}{n+1}(n+2)2^{-(n+1)-\frac{1}{2}}=c_{n+1}.$$