

The Optimally Weighted Median and its Variance

Wolfgang Brehm*

Iserbrooker Weg 65, Hamburg e-mail: wolfgang.brehm@gmail.com

Abstract: The weights that minimize the variance of the weighted median are proportional to the probability density of the samples at the overall weighted median. If the weights are chosen optimally in this sense, the variance of the weighted median estimate will be inverse to four times the sum of probability densities at the weighted median.

MSC2020 subject classifications: Primary 60K35, 60K35; secondary 60K35.

Keywords and phrases: order statistics, weighted median, ll optimization.

Contents

1	Introduction	1
2	Derivation of the optimal weighting and the resulting variance	2
3	Discussion	5
	References	5

1. Introduction

The median is the most robust estimator of central tendency. The weighted median is a straightforward generalization of the median, like the weighted mean is for the mean. They can be generalized to higher dimensions, and computed in near linear time [1, 2].

A different weighting of the samples, aside from the trivial frequency weighting, is preferred, when the samples are known to be not equally reliable. Perhaps because they are on a different scale, measured differently, or not equally significant in some other way. Sometimes unclear, however, is how to choose the weights for the weighted median exactly. A similar problem arises in L^1 optimization with heterogeneous data points, because the median minimizes the L^1 norm. For these cases, this publication will give some guidance. To find the optimal weights, first, the variance of the equally weighted median will be recapitulated. Secondly, the condition for optimality will be set and then the optimal weights will be derived, illustrated by numerical simulations. And lastly, results and implications will be discussed.

The median value of a distribution is the value where the cumulative distribution equals $1/2$, meaning values less than or equal the median are just as likely as values larger than or equal [3, Chap. 2.8]. Just like the mean of a distribution can be estimated with the sample average, the median of a distribution can be estimated with the sample

arXiv: 2010.00000

*This publication is possible due to the support from the Cluster of Excellence “CUI: Advanced Imaging of Matter” of the Deutsche Forschungsgemeinschaft (DFG) – EXC 2056 – project ID 390715994.

median. When there is an even number of samples, all values between and including the two middle most values are median values according to the definition. And like there is the central limit theorem for the median, there is the median theorem for the sample median:

The median $\tilde{\mu}_s$ of a sample of size N drawn from a distribution with density $p(x)$ approaches a Gaussian distribution with mean $\tilde{\mu}$ and variance $\left(4N (p(\tilde{\mu}))^2\right)^{-1}$ for large sample sizes, if the density around the median of the distribution is continuously differentiable and non-zero [4, Supp. mat.] [5].

$$\sigma_{\tilde{\mu}_s}^2 = \frac{1}{4N (p(\tilde{\mu}))^2} \quad (1.1)$$

To estimate the standard deviation of the sample median of samples from an unknown distribution, Woodruff[6] suggests using the following formula:

$$\sigma_{\tilde{\mu}_s} \approx \frac{x_{\lceil \frac{1}{2}(N+\sqrt{N}) \rceil} - x_{\lfloor \frac{1}{2}(N-\sqrt{N}) \rfloor}}{2} \quad (1.2)$$

Because the median, by definition, is the value for which half the values are less than or equal and the other half is greater than or equal, the probability of a value at position i in a ordered list of random samples being closest to the true median follows a binomial distribution with a success probability p of $1/2$. Therefore the values at these positions correspond to $\tilde{\mu} + \sigma_{\tilde{\mu}}$ and $\tilde{\mu} - \sigma_{\tilde{\mu}}$ on average respectively, and their difference, on average, is about two times the standard deviation. Knowing that the probabilities follow a binomial distribution, one can calculate a weighted squared deviation of the random samples to the sample median (analogous to the sample variance of the mean), to improve the accuracy of the variance estimate slightly at the cost of robustness:

$$\sigma_{\tilde{\mu}_s}^2 \approx \sum_{i=0}^{N-1} \binom{N-1}{i} \left(\frac{1}{2}\right)^{1-N} (x_i - \tilde{\mu}_s)^2 \quad (1.3)$$

Because the median of a large random sample approaches a Gaussian distribution, it is sensible to try to minimize its variance for the most accurate estimate. The optimal weights are the ones that minimize the variance of the weighted median.

2. Derivation of the optimal weighting and the resulting variance

We assume all samples x_i can follow a different probability distribution $p_i(x)$, otherwise different weighting would have only very limited application and the non-weighted median would almost always be the preferred choice. Consider a sorted list of weighted samples and the running sum of weights normalized by the sum of all weights. This computes the empirical cumulative distribution of the weighted mixture distribution, let the weighted cumulative distribution be ${}_wP(x)$.

$${}_wP(x) = \frac{\left(\sum_i^N w_i \int_{-\infty}^x p_i(y) dy\right)}{\left(\sum_i^N w_i\right)}$$

Now consider the sum of weights associated with samples less than or equal to the true weighted median ${}_w\tilde{\mu}$, normalized by the sum of all weights, let it be c :

$$c = \frac{\left(\sum_{\{i|x_i \leq {}_w\tilde{\mu}\}} w_i \right)}{\left(\sum_i^N w_i \right)} \quad (2.1)$$

Without knowing the median of each distribution, each weight has probability $1/2$ of being part of this sum or its counterpart. The variance introduced to the nominator by each weight is therefore $1/4$ the weights squared and the variance of c is approximated by the sum of individual variances divided by the normalization factor squared:

$$\sigma_c^2 = \frac{\left(\sum_i^N w_i^2 \right)}{4 \left(\sum_i^N w_i \right)^2}$$

The weighted sample median ${}_w\tilde{\mu}_s$ is the sample where the empirical weighted cumulative distribution reaches $1/2$. The expected value of c is $1/2$ and its variance is the mean squared deviation between the empirical partitioning which determines the weighted sample median and the partitioning which would lead to the value closest to the weighted population median. Therefore the value where the weighted cumulative distribution reaches c has the same variance as the weighted median. Error propagation demands the derivative of ${}_wP^{-1}$ at $1/2$ which is inverse to the derivative of ${}_wP$ at the weighted population median.

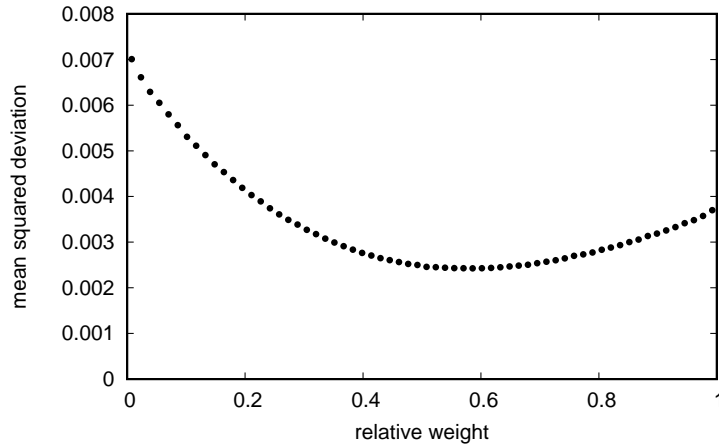


Figure 1: Relative weighting between samples following a Gaussian or uniform distribution with identical variance each. The ratio of the Gaussian density to the uniform density at the median is $\sqrt{6/\pi} \approx 1.38$, this ratio is reached at around 0.58 on the x-axis, coinciding with the minimum of mean squared deviations of the weighted sample median to the actual median.

$$\begin{aligned}
\sigma_{w\tilde{\mu}_s}^2 &= \sigma_c^2 \frac{d_w P^{-1}}{dx} (c) \\
\sigma_{w\tilde{\mu}_s}^2 &= \sigma_c^2 \left(\frac{d_w P}{dx} (w\tilde{\mu}) \right)^{-1} \\
\sigma_{w\tilde{\mu}_s}^2 &= \frac{\left(\sum_i^N w_i^2 \right) \left(\sum_i^N w_i \right)^2}{4 \left(\sum_i^N w_i \right)^2 \left(\sum_i^N w_i p_i (w\tilde{\mu}) \right)^2} \\
\sigma_{w\tilde{\mu}_s}^2 &= \frac{\left(\sum_i^N w_i^2 \right)}{4 \left(\sum_i^N w_i p_i (w\tilde{\mu}) \right)^2}
\end{aligned}$$

This is the variance we seek to minimize and a very similar optimization problem to the optimal weights for the weighted average. The weights that minimize the variance of the weighted median are reciprocal to the probability density of the sample distribution at the median, as can be shown by taking the first and second derivatives with respect to the individual weights. The first derivative can only be zero for $w_i = p_i (w\tilde{\mu})$ and the second derivative is positive, which are the necessary and sufficient conditions for a global minimum:

$$\begin{aligned}
\frac{d\sigma_{w\tilde{\mu}_s}^2}{dw_j} &= \frac{w_j \left(\sum_i^N w_i p_i (w\tilde{\mu}) \right) - p_j (w\tilde{\mu}) \left(\sum_i^N w_i^2 \right)}{2 \left(\sum_i^N w_i p_i (w\tilde{\mu}) \right)^3} \\
\frac{d^2\sigma_{w\tilde{\mu}_s}^2}{d^2w_j} &= \frac{1}{2 \left(\sum_i^N w_i p_i (w\tilde{\mu}) \right)^2} - \frac{2w_j p_j (w\tilde{\mu})}{2 \left(\sum_i^N w_i p_i (w\tilde{\mu}) \right)^3} + \frac{3(p_j (w\tilde{\mu}))^2 \left(\sum_i^N w_i^2 \right)}{2 \left(\sum_i^N w_i p_i (w\tilde{\mu}) \right)^4} \\
\arg \min_{w_i} \sigma_{w\tilde{\mu}_s}^2 &= p_i (w\tilde{\mu}) \tag{2.2}
\end{aligned}$$

Numerical experiments as shown in figure 1 also confirm this result. When the weights are set proportional to $p_i (w\tilde{\mu})$, the variance of the median is equal to $1/4$ the inverse of the sum of probability densities at the median squared:

$$\sigma_{w\tilde{\mu}_s}^2 = \frac{1}{4 \left(\sum (p_i (w\tilde{\mu}))^2 \right)} \tag{2.3}$$

When the weights are proportional, they are still optimal, but estimating the variance from the weights directly with equation 2.3 is not possible any more because the proportionality constant is unknown. The variance can still be estimated analogous to equations 1.2 and 1.3.

$$\sigma_{w\tilde{\mu}_s}^2 \approx \sum_i^N (x_i - w\tilde{\mu}_s)^2 \int_{\sum_j^{i-1} w_i}^{\sum_j^i w_i} \exp \left(-\frac{1}{2} \left(\left(y - \frac{1}{2} \right)^2 \sigma_c^{-2} + \log (2\pi\sigma_c^2) \right) \right) dy \tag{2.4}$$

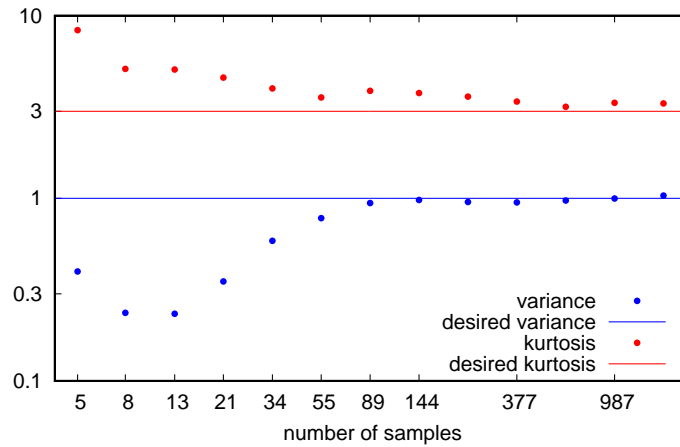


Figure 2: Variance and kurtosis of the weighted sample median divided by its estimated standard deviation (equation 2.4). The samples are drawn from a scaled normal distribution, where the weights are drawn from an exponential distribution and the divisors are set accordingly.

3. Discussion

One possibly unexpected consequence of weights proportional to the probability density at the median is that when the distribution of each sample has the same location and shape, and only differs in scaling, the optimal weights are proportional to the inverse standard deviation, not inverse variances as for the weighted average. This is because a scaling by a factor amounts to a linear decrease of the density at the median, but an increase of the variance by the same factor squared.

As can be seen in figure 2 the distribution of the weighted sample median approaches a normal distribution asymptotically, but not immediately. With fewer samples, the distribution can have significant excess kurtosis, typically being leptokurtic with a kurtosis above three, hinting to heavy tails, and making accurate estimation of a standard deviation difficult or possibly pointless. The average or median of absolute deviations could be a better measure of precision in these cases.

References

- [1] COHEN, M. B., LEE, Y. T., MILLER, G., PACHOCKI, J. and SIDFORD, A. (2016). Geometric Median in Nearly Linear Time. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing. STOC '16* 9–21. Association for Computing Machinery, New York, NY, USA.
- [2] JADHAV, S. and MUKHOPADHYAY, A. (1994). Computing a centerpoint of a finite planar set of points in linear time. *Discrete & Computational Geometry* **12** 291–312.

- [3] KENDALL, M. G. (1945). *Kendall's Advanced Theory of Statistics, Volume I. Kendall's Advanced Theory of Statistics*. Charles Griffin & Company Ltd.
- [4] MILLER, S. J. (2017). *The Probability Lifesaver*. Princeton University Press.
- [5] RIDER, P. R. (1960). Variance of the Median of Small Samples from Several Special Populations. *Journal of the American Statistical Association* **55** 148–150.
- [6] WOODRUFF, R. S. (1952). Confidence Intervals for Medians and Other Position Measures. *Journal of the American Statistical Association* **47** 635–646.