Semiparametric robust mean estimations based on the orderliness of quantile averages

Tuban Lee

This manuscript was compiled on June 18, 2023

semiparametric | mean-median-mode inequality | asymptotic | unimodal | Hodges–Lehmann estimator

Hodges–Lehmann inequality and γ-U-orderliness

The Hodges–Lehmann estimator stands out as a unique robust location estimator due to its definition being substantially dissimilar from conventional L-estimators, R-estimators, and M-estimators. In their landmark paper, *Estimates of location based on rank tests*, Hodges and Lehmann (1) proposed two methods for computing the H-L estimator: the Wilcoxon score R-estimator and the median of pairwise means. The Wilcoxon score R-estimator is a location estimator based on signed-rank test, or R-estimator, (1) and was later independently discovered by Sen (1963) (2, 3). However, the median of pairwise means is a generalized L-statistic and a trimmed U-statistic, as classified by Serfling in his novel conceptualized study in 1984 (4). Serfling further advanced the understanding by generalizing the H-L kernel as $hl_k(x_1,\ldots,x_k)=\frac{1}{k}\sum_{i=1}^{k}x_i$, where $k\in\mathbb{N}$ (4). Here, the weighted H-L kernel is defined as $whl_k(x_1,\ldots,x_k)=\frac{\sum_{i=1}^{k}x_i\mathbf{w}_i}{\sum_{i=1}^{k}\mathbf{w}_i}$, where $\mathbf{w}_i$s are the weights applied to each element.

By using the weighted H-L kernel and the L-estimator, it is now clear that the Hodges-Lehmann estimator is an LL-statistic, the definition of which is provided as follows:

$$LL_{k,\epsilon,\gamma,n}:=L_{\epsilon_0,\gamma,n}\left(\mathrm{sort}\left(\left(whl_k\left(X_{N_1},\cdots,X_{N_k}\right)\right)_{N=1}^{\binom{n}{k}}\right)\right),$$

where $L_{\epsilon_0,\gamma,n}(Y)$ represents the ϵ_0,γ-L-estimator that uses the sorted sequence, $\mathrm{sort}\left(\left(whl_k\left(X_{N_1},\cdots,X_{N_k}\right)\right)_{N=1}^{\binom{n}{k}}\right)$, as input. The upper asymptotic breakdown point of $LL_{k,\epsilon,\gamma}$ is $\epsilon=1-(1-\epsilon_0)^{\frac{1}{k}}$, as proven in DSSM II. There are two ways to adjust the breakdown point: either by setting k as a constant and adjusting ϵ_0, or by setting ϵ_0 as a constant and adjusting k. In the above definition, k is discrete, but the bootstrap method can be applied to ensure the continuity of k, also making the breakdown point continuous. Specifically, if $k\in\mathbb{R}$, let the bootstrap size be denoted by b, then first sampling the original sample $(1-k+\lfloor k\rfloor)b$ times with each sample size of $\lfloor k\rfloor$, and then subsequently sampling $(1-\lceil k\rceil+k)b$ times with each sample size of $\lceil k\rceil$, $(1-k+\lfloor k\rfloor)b\in\mathbb{N}$, $(1-\lceil k\rceil+k)b\in\mathbb{N}$. The corresponding kernels are computed separately, and the pooled sorted sequence is used as the input for the L-estimator. Let $\mathbf{S}_k$ represent the sorted sequence. Indeed, for any finite sample, X, when $k=n$, $\mathbf{S}_k$ becomes a single point, $whl_{k=n}(X_1,\ldots,X_n)$. When $\mathbf{w}_i=1$, the minimum of $\mathbf{S}_k$ is $\frac{1}{k}\sum_{i=1}^{k}X_i$, due to the property of order statistics. The maximum of $\mathbf{S}_k$ is $\frac{1}{k}\sum_{i=1}^{k}X_{n-i+1}$. The monotonicity of the order statistics implies the monotonicity of the extrema with respect to k, i.e., the support of $\mathbf{S}_k$ shrinks monotonically. For unequal $\mathbf{w}_i$s, the shrinkage of the support of $\mathbf{S}_k$ might not be strictly monotonic, but the general trend remains, since all LL-statistics converge to the same point, as $k\to n$. Therefore, if $\frac{\sum_{i=1}^{n}X_i\mathbf{w}_i}{\sum_{i=1}^{n}\mathbf{w}_i}$ approaches the population mean when $n\to\infty$, all LL-statistics based on such consistent kernel function approach the population mean as $k\to\infty$. For example, if $whl_k=\mathrm{BM}_{\nu,\epsilon_k,n=k}$, $\nu\ll\epsilon_k^{-1}$, $\epsilon_k\to 0$, such kernel function is consistent. These cases are termed the LL-mean ($\mathrm{LLM}_{k,\epsilon,\gamma,n}$). By substituting the $\mathrm{WA}_{\epsilon_0,\gamma,n}$ for the $L_{\epsilon_0,\gamma,n}$ in LL-statistic, the resulting statistic is referred to as the weighted L-statistic ($\mathrm{WL}_{k,\epsilon,\gamma,n}$). The case having a consistent kernel function is termed as the weighted L-mean ($\mathrm{WLM}_{k,\epsilon,\gamma,n}$). The $w_i=1$ case of $\mathrm{WLM}_{k,\epsilon,\gamma,n}$ is termed the weighted Hodges-Lehmann mean ($\mathrm{WHLM}_{k,\epsilon,\gamma,n}$). The $\mathrm{WHLM}_{k=1,\epsilon,\gamma,n}$ is the weighted average. If $k\geq 2$ and the WA in WHLM is set as TM_{ϵ_0}, it is called the trimmed H-L mean (Figure **??**, $k=2$, $\epsilon_0=\frac{15}{64}$). The $\mathrm{THLM}_{k=2,\epsilon,\gamma=1,n}$ appears similar to the Wilcoxon's one-sample statistic investigated by Saleh in 1976 (5), which involves first censoring the sample, and then computing the mean of the number of events that the pairwise mean is greater than zero. The $\mathrm{THLM}_{k=2,\epsilon=1-\left(1-\frac{1}{2}\right)^{\frac{1}{2}},\gamma=1,n}$ is the Hodges-Lehmann estimator, or more generally, a special case of the median Hodges-Lehmann mean ($m\mathrm{HLM}_{k,n}$). $m\mathrm{HLM}_{k,n}$ is asymptotically equivalent to the $\mathrm{MoM}_{k,b=\frac{n}{k}}$ as discussed previously, Therefore, it is possible to define a series of location estimators, analogous to the WHLM, based on MoM. For example, the γ-median of means, $\gamma m\mathrm{oM}_{k,b=\frac{n}{k},n}$, is defined by replacing the median in $\mathrm{MoM}_{k,b=\frac{n}{k},n}$ with the γ-median.

The hl_k kernel distribution, denoted as F_{hl_k}, can be defined as the probability distribution of the sorted sequence $\mathrm{sort}\left(\left(hl_k\left(X_{N_1},\cdots,X_{N_k}\right)\right)_{N=1}^{\binom{n}{k}}\right)$. For any real value y, the cdf of the hl_k kernel distribution is given by: $F_{h_{\mathbf{k}}}(y)=\Pr(Y_i\leq y)$, where Y_i represents an individual element from the sorted sequence. The overall hl_k kernel distributions possess a two-dimensional structure, encompassing n kernel distributions with varying k values, from 1 to n, where one dimension is inherent to each individual kernel distribution, while the other is formed by the alignment of the same percentiles across all kernel distributions. As k increases, all percentiles converge to $\bar{X}$, leading to the concept of γ-U-orderliness:

$$\left(\forall k_2\geq k_1\geq 1, \gamma m\mathrm{HLM}_{k_2,\epsilon=1-\left(\frac{\gamma}{1+\gamma}\right)^{\frac{1}{k_2}},\gamma}\geq \gamma m\mathrm{HLM}_{k_1,\epsilon=1-\left(\frac{\gamma}{1+\gamma}\right)^{\frac{1}{k_1}},\gamma}\right)\vee$$
$$\left(\forall k_2\geq k_1\geq 1, \gamma m\mathrm{HLM}_{k_2,\epsilon=1-\left(\frac{\gamma}{1+\gamma}\right)^{\frac{1}{k_2}},\gamma}\leq \gamma m\mathrm{HLM}_{k_1,\epsilon=1-\left(\frac{\gamma}{1+\gamma}\right)^{\frac{1}{k_1}},\gamma}\right),$$

where $\gamma m\mathrm{HLM}_k$ sets the WA in WHLM as γ-median, with γ being constant. The direction of the inequality depends

T.L. designed research, performed research, analyzed data, and wrote the paper.

The author declares no competing interest.

[1] To whom correspondence should be addressed. E-mail: tl@biomathematics.org

on the relative magnitudes of $\gamma m\mathrm{HLM}_{k=1,\epsilon,\gamma} = \gamma m$ and $\gamma m\mathrm{HLM}_{k=\infty,\epsilon,\gamma} = \mu$. The Hodges-Lehmann inequality can be defined as a special case of the γ-U-orderliness when $\gamma = 1$. When $\gamma \in \{0, \infty\}$, the γ-U-orderliness is valid for any distribution as previously shown. If $\gamma \notin \{0, \infty\}$, analytically proving the validity of the γ-U-orderliness for a parametric distribution is pretty challenging. As an example, the hl_2 kernel distribution has a probability density function $f_{hl_2}(x) = \int_0^{2x} 2f(t)\,f(2x-t)\,dt$ (a result after the transformation of variables); the support of the original distribution is assumed to be $[0, \infty)$ for simplicity. The expected value of the H-L estimator is the positive solution of $\int_0^{\text{H-L}} (f_{hl_2}(s))\,ds = \frac{1}{2}$. For the exponential distribution, $f_{hl_2,exp}(x) = 4\lambda^{-2}xe^{-2\lambda^{-1}x}$, λ is a scale parameter, $E[\text{H-L}] = \frac{-W_{-1}\left(-\frac{1}{2e}\right)-1}{2}\lambda \approx 0.839\lambda$, where W_{-1} is a branch of the Lambert W function which cannot be expressed in terms of elementary functions. However, the violation of the γ-U-orderliness is bounded under certain assumptions, as shown below.

Theorem .1. *For any distribution with a finite second central moment, σ^2, the following concentration bound can be established for the γ-median of means,*

$$\mathbb{P}\left(\gamma moM_{k,b=\frac{n}{k},n} - \mu > \frac{t\sigma}{\sqrt{k}}\right) \leq e^{-\frac{2n}{k}\left(\frac{1}{1+\gamma}-\frac{1}{k+t^2}\right)^2}.$$

Proof. Denote the mean of each block as $\widehat{\mu_i}$, $1 \leq i \leq b$. Observe that the event $\left\{\gamma mo\mathrm{M}_{k,b=\frac{n}{k},n} - \mu > \frac{t\sigma}{\sqrt{k}}\right\}$ necessitates the condition that there are at least $b(1-\frac{\gamma}{1+\gamma})$ of $\widehat{\mu_i}$s larger than μ by more than $\frac{t\sigma}{\sqrt{k}}$, i.e., $\left\{\gamma mo\mathrm{M}_{k,b=\frac{n}{k},n} - \mu > \frac{t\sigma}{\sqrt{k}}\right\} \subset \left\{\sum_{i=1}^{b} \mathbf{1}_{\left(\widehat{\mu_i}-\mu\right)>\frac{t\sigma}{\sqrt{k}}} \geq b\left(1-\frac{\gamma}{1+\gamma}\right)\right\}$, where $\mathbf{1}_A$ is the indicator of event A. Assuming a finite second central moment, σ^2, it follows from one-sided Chebeshev's inequality that $\mathbb{E}\left(\mathbf{1}_{\left(\widehat{\mu_i}-\mu\right)>\frac{t\sigma}{\sqrt{k}}}\right) = \mathbb{P}\left((\widehat{\mu_i}-\mu) > \frac{t\sigma}{\sqrt{k}}\right) \leq \frac{\sigma^2}{k\sigma^2+t^2\sigma^2}$.

Given that $\mathbf{1}_{\left(\widehat{\mu_i}-\mu\right)>\frac{t\sigma}{\sqrt{k}}} \in [0,1]$ are independent and identically distributed random variables, according to the aforementioned inclusion relation, the one-sided Chebeshev's inequality and the one-sided Hoeffding's inequality, $\mathbb{P}\left(\gamma mo\mathrm{M}_{k,b=\frac{n}{k},n} - \mu > \frac{t\sigma}{\sqrt{k}}\right) \leq \mathbb{P}\left(\sum_{i=1}^{b} \mathbf{1}_{\left(\widehat{\mu_i}-\mu\right)>\frac{t\sigma}{\sqrt{k}}} \geq b\left(1-\frac{\gamma}{1+\gamma}\right)\right) = \mathbb{P}\left(\frac{1}{b}\sum_{i=1}^{b}\left(\mathbf{1}_{\left(\widehat{\mu_i}-\mu\right)>\frac{t\sigma}{\sqrt{k}}} - \mathbb{E}\left(\mathbf{1}_{\left(\widehat{\mu_i}-\mu\right)>\frac{t\sigma}{\sqrt{k}}}\right)\right) \geq \left(1-\frac{\gamma}{1+\gamma}\right) - \mathbb{E}\left(\mathbf{1}_{\left(\widehat{\mu_i}-\mu\right)>\frac{t\sigma}{\sqrt{k}}}\right)\right) \leq e^{-2b\left(\left(1-\frac{\gamma}{1+\gamma}\right)-\mathbb{E}\left(\mathbf{1}_{\left(\widehat{\mu_i}-\mu\right)>\frac{t\sigma}{\sqrt{k}}}\right)\right)^2} \leq e^{-2b\left(1-\frac{\gamma}{1+\gamma}-\frac{\sigma^2}{k\sigma^2+t^2\sigma^2}\right)^2} = e^{-2b\left(\frac{1}{1+\gamma}-\frac{1}{k+t^2}\right)^2}$. □

Theorem .2. *Let $B(k,\gamma,t,n) = e^{-\frac{2n}{k}\left(\frac{1}{1+\gamma}-\frac{1}{k+t^2}\right)^2}$. If $n \in \mathbb{N}$, $\gamma \geq 0$, $0 \leq t^2 < \gamma+1$, and $\gamma - t^2 + 1 \leq k \leq \frac{1}{2}\sqrt{9\gamma^2+18\gamma-8\gamma t^2-8t^2+9}+\frac{1}{2}\left(3\gamma-2t^2+3\right)$, B is monotonic decreasing with respect to k.*

Proof. Since $\frac{\partial B}{\partial k} = \left(\frac{2n\left(\frac{1}{\gamma+1}-\frac{1}{k+t^2}\right)^2}{k^2} - \frac{4n\left(\frac{1}{\gamma+1}-\frac{1}{k+t^2}\right)}{k\left(k+t^2\right)^2}\right) e^{-\frac{2n\left(\frac{1}{\gamma+1}-\frac{1}{k+t^2}\right)^2}{k}}$ and $n \in \mathbb{N}$, $\frac{\partial B}{\partial k} \leq 0 \Leftrightarrow \frac{2n\left(\frac{1}{\gamma+1}-\frac{1}{k+t^2}\right)^2}{k^2} - \frac{4n\left(\frac{1}{\gamma+1}-\frac{1}{k+t^2}\right)}{k\left(k+t^2\right)^2} \leq 0 \Leftrightarrow \frac{2n\left(-\gamma+k+t^2-1\right)\left(k^2-3(\gamma+1)k+2kt^2+t^2\left(-\gamma+t^2-1\right)\right)}{(\gamma+1)^2k^2\left(k+t^2\right)^3} \leq 0 \Leftrightarrow \left(-\gamma+k+t^2-1\right)\left(k^2-3(\gamma+1)k+2kt^2+t^2\left(-\gamma+t^2-1\right)\right) \leq 0$. When the factors are expanded, it yields a cubic inequality in terms of k: $k^3+k^2\left(3t^2-4(\gamma+1)\right)+3k\left(\gamma-t^2+1\right)^2+t^2\left(\gamma-t^2+1\right)^2 \leq 0$. Assuming $0 \leq t^2 < \gamma+1$ and $\gamma \geq 0$, using the factored form and subsequently applying the quadratic formula, the inequality is valid if $\gamma - t^2 + 1 \leq k \leq \frac{1}{2}\sqrt{9\gamma^2+18\gamma-8\gamma t^2-8t^2+9}+\frac{1}{2}\left(3\gamma-2t^2+3\right)$. □

Let X be a random variable and $\bar{Y} = \frac{1}{k}(Y_1+\cdots+Y_k)$ be the average of k independent, identically distributed copies of X. Applying the variance operation gives: $\mathrm{Var}(\bar{Y}) = \mathrm{Var}\left(\frac{1}{k}(Y_1+\cdots+Y_k)\right) = \frac{1}{k^2}(\mathrm{Var}(Y_1)+\cdots+\mathrm{Var}(Y_k)) = \frac{1}{k^2}(k\sigma^2) = \frac{\sigma^2}{k}$, since the variance operation is a linear operator for independent variables, and the variance of a scaled random variable is the square of the scale times the variance of the variable, i.e., $\mathrm{Var}(cX) = E[(cX-E[cX])^2] = E[(cX-cE[X])^2] = E[c^2(X-E[X])^2] = c^2E[((X)-E[X])^2] = c^2\mathrm{Var}(X)$. Thus, the standard deviation of the hl_k kernel distribution, asymptotically, is $\frac{\sigma}{\sqrt{k}}$. By utilizing the asymptotic bias bound of any quantile for any continuous distribution with a finite second central moment, σ^2,(6), a conservative asymptotic bias bound of $\gamma mo\mathrm{M}_{k,b=\frac{n}{k}}$ can be established as $\gamma mo\mathrm{M}_{k,b=\frac{n}{k}} - \mu \leq \sqrt{\frac{\frac{\gamma}{1+\gamma}}{1-\frac{\gamma}{1+\gamma}}}\sigma_{hl_k} = \sqrt{\frac{\gamma}{k}}\sigma$. That implies in Theorem .1, $t < \sqrt{\gamma}$, so when $\gamma = 1$, the upper bound of k, subject to the monotonic decreasing constraint, is $2+\sqrt{5} < \frac{1}{2}\sqrt{9+18-8t^2-8t^2+9}+\frac{1}{2}\left(3-2t^2+3\right) \leq 6$, the lower bound is $1 < 2-t^2 \leq 2$. These analyses elucidate a surprising result: although the conservative asymptotic bound of $\mathrm{MoM}_{k,b=\frac{n}{k}}$ is monotonic with respect to k, its concentration bound is optimal when $k \in (2+\sqrt{5}, 6]$.

Then consider the structure within each individual hl_k kernel distribution. The sorted sequence $\mathbf{S}_k$, when $k = n-1$, has n elements and the corresponding hl_k kernel distribution can be seen as a location-scale transformation of the original distribution, so it is νth γ-ordered if and only if the original distribution is νth γ-ordered according to Theorem **??**.

Data Availability. Data for Figure **??** are given in SI Dataset S1. All codes have been deposited in GitHub.

ACKNOWLEDGMENTS. I sincerely acknowledge the insightful comments from the editor which considerably elevated the lucidity and merit of this paper.

1. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. *The Annals Math. Stat.* **34**, 598–611 (1963).
2. PK Sen, On the estimation of relative potency in dilution (-direct) assays by distribution-free methods. *Biometrics* pp. 532–552 (1963).
3. M Ghosh, MJ Schell, PK Sen, A conversation with pranab kumar sen. *Stat. Sci.* pp. 548–564 (2008).
4. RJ Serfling, Generalized l-, m-, and r-statistics. *The Annals Stat.* **12**, 76–86 (1984).
5. A Ehsanes Saleh, Hodges-lehmann estimate of the location parameter in censored samples. *Annals Inst. Stat. Math.* **28**, 235–247 (1976).
6. L Li, H Shao, R Wang, J Yang, Worst-case range value-at-risk with partial information. *SIAM J. on Financial Math.* **9**, 190–218 (2018).