Descriptive statistics for semiparametric models. ii. moments

Tuban Lee

This manuscript was compiled on June 12, 2023

A. Robust Estimations of the Central Moments. In 1976, Bickel and Lehmann (1), in their third paper of the landmark series *Descriptive Statistics for Nonparametric Models*, generalized nearly all robust scale estimators of that time as measures of the dispersion of a symmetric distribution around its center of symmetry. In 1979, the same series, they (2) proposed a class of estimators referred to as measures of spread, which consider the pairwise differences of a random variable, irrespective of its symmetry, throughout its distribution, rather than focusing on dispersion relative to a fixed point. While they had already considered one version of the trimmed standard deviation, which is essentially a trimmed second raw moment, in the third paper of that series (1); in the final section of the fourth paper (2), they explored another two versions of the trimmed standard deviation based on symmetric differences and pairwise differences, the latter is modified here for comparison,

$$\left[\binom{n}{2}(1-\epsilon-\gamma\epsilon)\right]^{-\frac{1}{2}}\left[\sum_{i=\binom{n}{2}\gamma\epsilon}^{\binom{n}{2}(1-\epsilon)}\left(X-X'\right)_i^{2}\right]^{\frac{1}{2}}, \qquad [1]$$

where $(X-X')_1 \leq \ldots \leq (X-X')_{\binom{n}{2}}$ are the order statistics of the pairwise differences, $X_{\mathbf{i}}-X_{\mathbf{j}}$, $\mathbf{i}<\mathbf{j}$, provided that $\binom{n}{2}\gamma\epsilon \in \mathbb{N}$ and $\binom{n}{2}(1-\epsilon) \in \mathbb{N}$. They showed that, when $\epsilon=0$, the result obtained using [1] is equal to $\sqrt{2}$ times the sample standard deviation. The paper ended with, “We do not know a fortiori which of the measures is preferable and leave these interesting questions open.”

Two examples of the impacts of that series are as follows. Oja (1981, 1983) (3, 4) provided a more comprehensive and generalized examination of these concepts, and integrated the measures of location, dispersion, and spread as proposed by Bickel and Lehmann (1, 2, 5), along with van Zwet's convex transformation order of skewness and kurtosis (1964) (6) for univariate and multivariate distributions, resulting a greater degree of generality and a broader perspective on these statistical constructs. Rousseeuw and Croux proposed a popular efficient scale estimator based on separate medians of pairwise differences taken over $\mathbf{i}$ and $\mathbf{j}$ (7) in 1993. However the importance of tackling the symmetry assumption has been greatly underestimated, as will be discussed later.

To address their open question (2), the nomenclature used in this paper is introduced as follows:

Nomenclature. Given a robust estimator, $\hat{\theta}$, which has an adjustable breakdown point, ϵ, that can approach zero asymptotically, the name of $\hat{\theta}$ comprises two parts: the first part denotes the type of estimator, and the second part represents the population parameter θ, such that $\hat{\theta} \to \theta$ as $\epsilon \to 0$. The abbreviation of the estimator combines the initial letters of the first part and the second part. If the estimator is symmetric, the upper asymptotic breakdown point, ϵ ($\epsilon_{U_{\mathbf{k}}}$ for LU-statistic), is indicated in the subscript of the abbreviation of the estimator, with the exception of the median. For an asymmetric estimator based on quantile average, the associated γ follows ϵ.

In DSSM I, it was shown that the bias of a robust estimator with an adjustable breakdown point is often monotonic with respect to the breakdown point in a semiparametric distribution. Naturally, the estimator's name should reflect the population parameter that it approaches as $\epsilon \to 0$. If multiplying all pseudo-samples by a factor of $\frac{1}{\sqrt{2}}$, then [1] is the trimmed standard deviation adhering to this nomenclature, since $\psi_2(x_1,x_2)=\frac{1}{2}(x_1-x_2)^2$ is the kernel function of the unbiased estimation of the second central moment by using U-statistic (8). This definition should be preferable, not only because it is the square root of a trimmed U-statistic, which is closely related to the minimum-variance unbiased estimator (MVUE), but also because the second γ-orderliness of the second central moment kernel distribution is ensured by the next exciting theorem.

Theorem A.1. *The second central moment kernel distribution generated from any unimodal distribution is second γ-ordered, provided that $\gamma \geq 0$.*

Proof. In 1954, Hodges and Lehmann established that if X and Y are independently drawn from the same unimodal distribution, $X-Y$ will be a symmetric unimodal distribution peaking at zero (9). Given the constraint in the pairwise differences that $X_{\mathbf{i}}<X_{\mathbf{j}}$, $\mathbf{i}<\mathbf{j}$, it directly follows from Theorem 1 in (9) that the pairwise difference distribution (Ξ_{Δ}) generated from any unimodal distribution is always monotonic increasing with a mode at zero. Since $X-X'$ is a negative variable that is monotonically increasing, applying the squaring transformation, the relationship between the original variable $X-X'$ and its squared counterpart $(X-X')^2$ can be represented as follows: $X-X'<Y-Y' \implies (X-X')^2>(Y-Y')^2$. In other words, as the negative values of $X-X'$ become larger in magnitude (more negative), their squared values $(X-X')^2$ become larger as well, but in a monotonically decreasing manner with a mode at zero. Further multiplication by $\frac{1}{2}$ also does not change the monotonicity and mode, since the mode is zero. Therefore, the transformed pdf becomes monotonically decreasing with a mode at zero. In DSSM I, it was proven that a right-skewed distribution with a monotonic decreasing pdf is always second γ-ordered, which gives the desired result. $\square$

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

In RSSM I, it was shown that any γ-symmetric distribution is νth γ-U-ordered, suggesting that νth γ-U-orderliness does not require unimodality, e.g., a symmetric bimodal distribution is also νth U-ordered. In the SI Text of RSSM I, an analysis of the Weibull distribution showed that unimodality does not assure orderliness. Theorem A.1 uncovers a profound relationship between unimodality, monotonicity, and second γ-orderliness, which is sufficient for γ-trimming inequality and γ-orderliness.

In 1928, Fisher constructed $\mathbf{k}$-statistics as unbiased estimators of cumulants (10). Halmos (1946) proved that a functional θ admits an unbiased estimator if and only if it is a regular statistical functional of degree $\mathbf{k}$ and showed a relation of symmetry, unbiasness and minimum variance (11). Hoeffding, in 1948, generalized U-statistics (12) which enable the derivation of a minimum-variance unbiased estimator from each unbiased estimator of an estimable parameter. In 1984, Serfling pointed out the speciality of Hodges-Lehmann estimator, which is neither a simple L-statistic nor a U-statistic, and considered the generalized L-statistics and trimmed U-statistics (13). Given a kernel function $h_{\mathbf{k}}$ which is a symmetric function of $\mathbf{k}$ variables, the LU-statistic is defined as:

$$LU_{h_{\mathbf{k}},\mathbf{k},k,\epsilon_{U_{\mathbf{k}}},\gamma,n} := LL_{k,\epsilon,\gamma,n}\left(\text{sort}\left(\left(h_{\mathbf{k}}\left(X_{N_1},\ldots,X_{N_{\mathbf{k}}}\right)\right)_{N=1}^{\binom{n}{\mathbf{k}}}\right)\right),$$

where $\epsilon_{U_{\mathbf{k}}} = 1-(1-\epsilon)^{\frac{1}{\mathbf{k}}}$ (proven in Subsection **??**), $X_{N_1},\ldots,X_{N_{\mathbf{k}}}$ are the n choose $\mathbf{k}$ elements from the sample, $LL_{k,\epsilon,\gamma,n}(Y)$ denotes the LL-statistic with the sorted sequence $\text{sort}\left(\left(h_{\mathbf{k}}\left(X_{N_1},\ldots,X_{N_{\mathbf{k}}}\right)\right)_{N=1}^{\binom{n}{\mathbf{k}}}\right)$ serving as an input. In the context of Serfling's work, the term 'trimmed U-statistic' is used when $LL_{k,\epsilon,\gamma,n}$ is $\text{TM}_{\epsilon,\gamma,n}$ (13).

In 1997, Heffernan (8) obtained an unbiased estimator of the $\mathbf{k}$th central moment by using U-statistics and demonstrated that it is the minimum variance unbiased estimator for distributions with the finite first $\mathbf{k}$ moments. The weighted Hodges-Lehmann $\mathbf{k}$th central moment ($2 \leq \mathbf{k} \leq n$) is thus defined as,

$$\text{WHL}km_{k,\epsilon_{U_{\mathbf{k}}},\gamma,n} := LU_{h_{\mathbf{k}}=\psi_{\mathbf{k}},\mathbf{k},k,\epsilon_{U_{\mathbf{k}}},\gamma,n},$$

where $\text{WHLM}_{k,\epsilon,\gamma,n}$ is used as the $LL_{k,\epsilon,\gamma,n}$ in LU, $\psi_{\mathbf{k}}\left(x_1,\ldots,x_{\mathbf{k}}\right) = \sum_{j=0}^{\mathbf{k}-2}(-1)^j\left(\frac{1}{\mathbf{k}-j}\right)\sum\left(x_{i_1}^{\mathbf{k}-j}x_{i_2}\ldots x_{i_{j+1}}\right) + (-1)^{\mathbf{k}-1}(\mathbf{k}-1)x_1\ldots x_{\mathbf{k}}$, the second summation is over $i_1,\ldots,i_{j+1} = 1$ to $\mathbf{k}$ with $i_1 \neq i_2 \neq \ldots \neq i_{j+1}$ and $i_2 < i_3 < \ldots < i_{j+1}$ (8). Despite the complexity, the following theorem offers an approach to infer the general structure of such kernel distributions.

Theorem A.2. *Define a set T comprising all pairs $(\psi_{\mathbf{k}}(\mathbf{v}), f_{X,\ldots,X}(\mathbf{v}))$ such that $\psi_{\mathbf{k}}(\mathbf{v}) = \psi_{\mathbf{k}}\left(Q(p_1),\ldots,Q(p_{\mathbf{k}})\right)$ with $Q(p_1) < \ldots < Q(p_{\mathbf{k}})$ and $f_{X,\ldots,X}(\mathbf{v}) = \mathbf{k}!f(Q(p_1))\ldots f(Q(p_{\mathbf{k}}))$ is the probability density of the $\mathbf{k}$-tuple, $\mathbf{v} = (Q(p_1),\ldots,Q(p_{\mathbf{k}}))$ (a formula drawn after a modification of the Jacobian density theorem). T_Δ is a subset of T, consisting all those pairs for which the corresponding $\mathbf{k}$-tuples satisfy that $Q(p_1) - Q(p_{\mathbf{k}}) = \Delta$. The component quasi-distribution, denoted by ξ_Δ, has a quasi-pdf $f_{\xi_\Delta}(\bar{\Delta}) = \sum_{\substack{(\psi_{\mathbf{k}}(\mathbf{v}), f_{X,\ldots,X}(\mathbf{v}))\in T_\Delta \\ \bar{\Delta}=\psi_{\mathbf{k}}(\mathbf{v})}} f_{X,\ldots,X}(\mathbf{v})$, i.e., sum over all $f_{X,\ldots,X}(\mathbf{v})$ such that the pair $(\psi_{\mathbf{k}}(\mathbf{v}), f_{X,\ldots,X}(\mathbf{v}))$ is in the set T_Δ and the first element of the pair, $\psi_{\mathbf{k}}(\mathbf{v})$, is equal to $\bar{\Delta}$. The $\mathbf{k}$th, where $\mathbf{k} > 2$, central moment kernel distribution, labeled $\Xi_{\mathbf{k}}$, can be seen as a quasi-mixture distribution comprising an infinite number of component quasi-distributions, $\xi_\Delta s$, each corresponding to a different value of Δ, which ranges from $Q(0) - Q(1)$ to 0. Each component quasi-distribution has a support of $\left(-\left(\frac{\mathbf{k}}{\frac{3+(-1)^{\mathbf{k}}}{2}}\right)^{-1}(-\Delta)^{\mathbf{k}}, \frac{1}{\mathbf{k}}(-\Delta)^{\mathbf{k}}\right)$.*

Proof. The support of ξ_Δ is the extrema of the function $\psi_{\mathbf{k}}\left(Q(p_1),\ldots,Q(p_{\mathbf{k}})\right)$ subjected to the constraints, $Q(p_1) < \ldots < Q(p_{\mathbf{k}})$ and $\Delta = Q(p_1) - Q(p_{\mathbf{k}})$. Using the Lagrange multiplier, the only critical point can be determined at $Q(p_1) = \ldots = Q(p_{\mathbf{k}}) = 0$, where $\psi_{\mathbf{k}} = 0$. Other candidates are within the boundaries, i.e., $\psi_{\mathbf{k}}\left(x_1 = x_1, x_2 = x_{\mathbf{k}},\ldots,x_{\mathbf{k}} = x_{\mathbf{k}}\right)$, $\ldots$, $\psi_{\mathbf{k}}\left(x_1 = x_1,\ldots,x_i = x_1, x_{i+1} = x_{\mathbf{k}},\ldots,x_{\mathbf{k}} = x_{\mathbf{k}}\right)$, $\ldots$, $\psi_{\mathbf{k}}\left(x_1 = x_1,\ldots,x_{\mathbf{k}-1} = x_1, x_{\mathbf{k}} = x_{\mathbf{k}}\right)$. $\psi_{\mathbf{k}}\left(x_1 = x_1,\ldots,x_i = x_1, x_{i+1} = x_{\mathbf{k}},\ldots,x_{\mathbf{k}} = x_{\mathbf{k}}\right)$ can be divided into $\mathbf{k}$ groups.

$\square$

1. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models. iii. dispersion in *Selected works of EL Lehmann.* (Springer), pp. 499–518 (2012).
2. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models iv. spread in *Selected Works of EL Lehmann.* (Springer), pp. 519–526 (2012).
3. H Oja, On location, scale, skewness and kurtosis of univariate distributions. *Scand. J. statistics* pp. 154–168 (1981).
4. H Oja, Descriptive statistics for multivariate distributions. *Stat. & Probab. Lett.* **1**, 327–332 (1983).
5. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models ii. location in *selected works of EL Lehmann.* (Springer), pp. 473–497 (2012).
6. W van Zwet, Convex transformations: A new approach to skewness and kurtosis in *Selected Works of Willem van Zwet.* (Springer), pp. 3–11 (2012).
7. PJ Rousseeuw, C Croux, Alternatives to the median absolute deviation. *J. Am. Stat. association* **88**, 1273–1283 (1993).
8. PM Heffernan, Unbiased estimation of central moments by using u-statistics. *J. Royal Stat. Soc. Ser. B (Statistical Methodol.* **59**, 861–863 (1997).
9. J Hodges, E Lehmann, Matching in paired comparisons. *The Annals Math. Stat.* **25**, 787–791 (1954).
10. RA Fisher, Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc.* **2**, 199–238 (1930).
11. PR Halmos, The theory of unbiased estimation. *The Annals Math. Stat.* **17**, 34–43 (1946).
12. W Hoeffding, A class of statistics with asymptotically normal distribution. *The Annals Math. Stat.* **19**, 293–325 (1948).
13. RJ Serfling, Generalized l-, m-, and r-statistics. *The Annals Stat.* **12**, 76–86 (1984).