

Robust estimations of moments for unimodal distributions

Tuban Lee

This manuscript was compiled on June 12, 2023

A. Invariant Moments. All popular robust location estimators, such as the symmetric trimmed mean, symmetric Winsorized mean, Hodges-Lehmann estimator, Huber M -estimator, and median of means, are symmetric. As shown in RSSM I, a γ -weighted Hodges-Lehmann mean ($\text{WHLM}_{k,\epsilon,\gamma}$) can achieve consistency for the population mean in any γ -symmetric distribution with a finite mean. However, it falls considerably short of consistently handling other parametric distributions that are not γ -symmetric. Shifting from semiparametrics to parametrics, consider a robust estimator with a non-sample-dependent breakdown point (defined in Subsection ??) which is consistent simultaneously for both a semiparametric distribution and a parametric distribution that does not belong to that semiparametric distribution, it is named with the prefix ‘invariant’ followed by the name of the population parameter it is consistent with. Here, the recombined I -statistic is defined as

$$\text{RI}_{d,h_{\mathbf{k}},\mathbf{k}_1,\mathbf{k}_2,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,n,LU_1,LU_2} := \lim_{c \rightarrow \infty} \left(\frac{(LU_{1h_{\mathbf{k}},\mathbf{k}_1,k_1,\epsilon_1,\gamma_1,n} + c)^{d+1}}{(LU_{2h_{\mathbf{k}},\mathbf{k}_2,k_2,\epsilon_2,\gamma_2,n} + c)^d} - c \right),$$

where d is the key factor for bias correction, $LU_{h_{\mathbf{k}},\mathbf{k},k,\epsilon,\gamma,n}$ is the LU -statistic, \mathbf{k} is the degree of the U -statistic, k is the degree of the LL -statistic, ϵ is the upper asymptotic breakdown point of the LU -statistic. It is assumed in this series that in the subscript of an estimator, if \mathbf{k} , k and γ are omitted, $\mathbf{k} = 1$, $k = 1$, $\gamma = 1$ are assumed, if just one γ is indicated, $\gamma_1 = \gamma_2$, if n is omitted, only the asymptotic behavior is considered, in the absence of subscripts, no assumptions are made. The subsequent theorem shows the significance of a recombined I -statistic.

Theorem A.1. Define the recombined mean as $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,n,WL_1,WL_2} := \text{RI}_{d,h_{\mathbf{k}},\mathbf{k}_1,k_1,\epsilon_1,\gamma_1,n,WL_1,WL_2}$. Assuming finite means, $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,n,WL_1,WL_2} = \frac{\mu - WL_{1k_1,\epsilon_1,\gamma_1}}{WL_{1k_1,\epsilon_1,\gamma_1} - WL_{2k_2,\epsilon_2,\gamma_2}}$, $k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,n$ is a consistent mean estimator for a location-scale distribution, where μ , $WL_{1k_1,\epsilon_1,\gamma_1}$, and $WL_{2k_2,\epsilon_2,\gamma_2}$ are different location parameters from that location-scale distribution. If $\gamma_1 = \gamma_2$, $WL = \text{WHLM}$, rm is also consistent for any γ -symmetric distributions.

Proof. Finding d that make $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,WL_1,WL_2}$ a consistent mean estimator is equivalent to finding the solution of $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,WL_1,WL_2} = \mu$. First consider the location-scale distribution. Since $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,WL_1,WL_2} = \frac{\mu - WL_{1k_1,\epsilon_1,\gamma_1}}{WL_{1k_1,\epsilon_1,\gamma_1} - WL_{2k_2,\epsilon_2,\gamma_2}}$, $k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,n$ is a consistent mean estimator for a location-scale distribution, where μ , $WL_{1k_1,\epsilon_1,\gamma_1}$, and $WL_{2k_2,\epsilon_2,\gamma_2}$ are different location parameters from that location-scale distribution. If $\gamma_1 = \gamma_2$, $WL = \text{WHLM}$, rm is also consistent for any γ -symmetric distributions.

Proof. Finding d that make $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,WL_1,WL_2}$ a consistent mean estimator is equivalent to finding the solution of $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,WL_1,WL_2} = \mu$. First consider the location-scale distribution. Since $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,WL_1,WL_2} = \frac{\mu - WL_{1k_1,\epsilon_1,\gamma_1}}{WL_{1k_1,\epsilon_1,\gamma_1} - WL_{2k_2,\epsilon_2,\gamma_2}}$, $k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,n$ is a consistent mean estimator for a location-scale distribution, where μ , $WL_{1k_1,\epsilon_1,\gamma_1}$, and $WL_{2k_2,\epsilon_2,\gamma_2}$ are different location parameters from that location-scale distribution. If $\gamma_1 = \gamma_2$, $WL = \text{WHLM}$, rm is also consistent for any γ -symmetric distributions.

$dWL_{2k_2,\epsilon_2,\gamma} = \mu$. So, $d = \frac{\mu - WL_{1k_1,\epsilon_1,\gamma_1}}{WL_{1k_1,\epsilon_1,\gamma_1} - WL_{2k_2,\epsilon_2,\gamma_2}}$. In RSSM I, it was established that any $WL(k,\epsilon,\gamma)$ can be expressed as $\lambda WL_0(k,\epsilon,\gamma) + \mu$ for a location-scale distribution parameterized by a location parameter μ and a scale parameter λ , where $WL_0(k,\epsilon,\gamma)$ is a function of $Q_0(p)$, the quantile function of a standard distribution without any shifts or scaling, according to the definition of the weighted L -statistic. The simultaneous cancellation of μ and λ in $\frac{(\lambda\mu_0 + \mu) - (\lambda WL_{10}(k_1,\epsilon_1,\gamma_1) + \mu)}{(\lambda WL_{10}(k_1,\epsilon_1,\gamma_1) + \mu) - (\lambda WL_{20}(k_2,\epsilon_2,\gamma_2) + \mu)}$ assures that the d in rm is always a constant for a location-scale distribution. The proof of the second assertion follows directly from the coincidence property. According to Theorem 18 in RSSM I, for any γ -symmetric distribution with a finite mean, $\text{WHLM}_{1k_1,\epsilon_1,\gamma} = \text{WHLM}_{2k_2,\epsilon_2,\gamma} = \mu$. Then $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma,WL_1,WL_2} = \lim_{c \rightarrow \infty} \left(\frac{(\mu + c)^{d+1}}{(\mu + c)^d} - c \right) = \mu$. This completes the demonstration. \square

For example, the Pareto distribution has a quantile function $Q_{Par}(p) = x_m(1-p)^{-\frac{1}{\alpha}}$, where x_m is the minimum possible value that a random variable following the Pareto distribution can take, serving a scale parameter, α is a shape parameter. The mean of the Pareto distribution is given by $\frac{\alpha x_m}{\alpha - 1}$. As $WL(k,\epsilon,\gamma)$ can be expressed as a function of $Q(p)$, one can set the two $WL_{k,\epsilon,\gamma}$ s in the d value as two arbitrary quantiles $Q_{Par}(p_1)$ and $Q_{Par}(p_2)$. For the Pareto distribution, $d_{Per} = \frac{\mu_{Per} - Q_{Par}(p_1)}{Q_{Par}(p_1) - Q_{Par}(p_2)} = \frac{\frac{\alpha x_m}{\alpha - 1} - x_m(1-p_1)^{-\frac{1}{\alpha}}}{x_m(1-p_1)^{-\frac{1}{\alpha}} - x_m(1-p_2)^{-\frac{1}{\alpha}}} \cdot x_m$ can be canceled out. Intriguingly, the quantile function of exponential distribution is $Q_{exp}(p) = \ln\left(\frac{1}{1-p}\right)\lambda$, $\lambda \geq 0$. $\mu_{exp} = \lambda$. Then, $d_{exp} = \frac{\mu_{exp} - Q_{exp}(p_1)}{Q_{exp}(p_1) - Q_{exp}(p_2)} = \frac{\lambda - \ln\left(\frac{1}{1-p_1}\right)\lambda}{\ln\left(\frac{1}{1-p_1}\right)\lambda - \ln\left(\frac{1}{1-p_2}\right)\lambda} = -\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)}$. Since $\lim_{\alpha \rightarrow \infty} \frac{\frac{\alpha}{\alpha-1} - (1-p_1)^{-1/\alpha}}{(1-p_1)^{-1/\alpha} - (1-p_2)^{-1/\alpha}} = -\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)}$, the d value for the Pareto distribution approaches that of the exponential distribution, as $\alpha \rightarrow \infty$, regardless of the type of weighted L -statistic used. That means, for the Weibull, gamma,

Significance Statement

Bias, variance, and contamination are the three main errors in statistics. Consistent robust estimation is unattainable without parametric assumptions. In this article, invariant moments are proposed as a means of achieving near-consistent and robust estimations of moments, even in scenarios where moderate violations of distributional assumptions occur, while the variances are sometimes smaller than those of the sample moments.

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

The author declares no competing interest.

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

58 Pareto, lognormal and generalized Gaussian distribution,

$$59 \quad rm_{d=\frac{\mu - \text{WHLM}_{1k_1, \epsilon_1, \gamma}}{\text{WHLM}_{1k_1, \epsilon_1, \gamma} - \text{WHLM}_{2k_2, \epsilon_2, \gamma}}, k_1, k_2, \epsilon_1, \epsilon_2, \gamma, \text{WHLM}_1, \text{WHLM}_2}$$

60 is consistent for at least one particular case, where
61 μ , $\text{WHLM}_{1k_1, \epsilon_1, \gamma}$, and $\text{WHLM}_{2k_2, \epsilon_2, \gamma}$ are differ-
62 ent location parameters from an exponential dis-
63 tribution. Let $\text{WHLM}_{1k_1, \epsilon_1, \gamma} = \text{BM}_{\nu=3, \epsilon=\frac{1}{24}}$,

64 $\text{WHLM}_{2k_2, \epsilon_2, \gamma} = m$, then $\mu = \lambda$, $m = Q\left(\frac{1}{2}\right) = \ln 2\lambda$,

$$65 \quad \text{BM}_{\nu=3, \epsilon=\frac{1}{24}} = \lambda \left(1 + \ln \left(\frac{26068394603446272 \sqrt[6]{\frac{\gamma}{247}} \sqrt[3]{11}}{391^{5/6} 101898752449325 \sqrt{5}} \right) \right),$$

66 the detailed formula is given in the SI Text. So, $d =$

$$67 \quad \frac{\mu - \text{BM}_{\nu=3, \epsilon=\frac{1}{24}}}{\text{BM}_{\nu=3, \epsilon=\frac{1}{24}} - m} = \frac{\lambda - \lambda \left(1 + \ln \left(\frac{26068394603446272 \sqrt[6]{\frac{\gamma}{247}} \sqrt[3]{11}}{391^{5/6} 101898752449325 \sqrt{5}} \right) \right)}{\lambda \left(1 + \ln \left(\frac{26068394603446272 \sqrt[6]{\frac{\gamma}{247}} \sqrt[3]{11}}{391^{5/6} 101898752449325 \sqrt{5}} \right) \right) - \ln 2\lambda} =$$

$$68 \quad - \frac{\ln \left(\frac{26068394603446272 \sqrt[6]{\frac{\gamma}{247}} \sqrt[3]{11}}{391^{5/6} 101898752449325 \sqrt{5}} \right)}{1 - \ln(2) + \ln \left(\frac{26068394603446272 \sqrt[6]{\frac{\gamma}{247}} \sqrt[3]{11}}{391^{5/6} 101898752449325 \sqrt{5}} \right)} \approx 0.103. \text{ The biases of}$$

69 $rm_{d \approx 0.103, \nu=3, \epsilon_1=\frac{1}{24}, \epsilon_2=\frac{1}{2}, \text{BM}, m}$ for distributions with skewness
70 between those of the exponential and symmetric distributions
71 are tiny (SI Dataset S1). $rm_{d \approx 0.103, \nu=3, \epsilon_1=\frac{1}{24}, \epsilon_2=\frac{1}{2}, \text{BM}, m}$
72 exhibits excellent performance for all these common unimodal
73 distributions (SI Dataset S1).

74 The recombined mean is an recombined I -statistic.
75 Consider an I -statistic whose LEs are percentiles of a distri-
76 bution obtained by plugging LU -statistics into a cumulative
77 distribution function, I is defined with arithmetic operations,
78 constants and quantile functions, such an estimator is classified
79 as a quantile I -statistic. One version of the quantile I -statistic
80 can be defined as $\text{QI}_{d, h_{\mathbf{k}}, \mathbf{k}_1, \mathbf{k}_2, k_1, k_2, \epsilon_1, \epsilon_2, \gamma_1, \gamma_2, n, LU_1, LU_2} :=$

$$81 \quad \begin{cases} \hat{Q}_{n, h_{\mathbf{k}}} \left(\left(\hat{F}_{n, h_{\mathbf{k}}} (LU) - \frac{\gamma}{1+\gamma} \right) d + \hat{F}_{n, h_{\mathbf{k}}} (LU) \right) & \hat{F}_{n, h_{\mathbf{k}}} (LU) \geq \frac{\gamma}{1+\gamma} \\ \hat{Q}_{n, h_{\mathbf{k}}} \left(\hat{F}_{n, h_{\mathbf{k}}} (LU) - \left(\frac{\gamma}{1+\gamma} - \hat{F}_{n, h_{\mathbf{k}}} (LU) \right) d \right) & \hat{F}_{n, h_{\mathbf{k}}} (LU) < \frac{\gamma}{1+\gamma} \end{cases}$$

82 where LU is $LU_{\mathbf{k}, k, \epsilon, \gamma, n}$, $\hat{F}_{n, h_{\mathbf{k}}} (x)$ is the empirical cumulative
83 distribution function of the $h_{\mathbf{k}}$ kernel distribution, $\hat{Q}_{n, h_{\mathbf{k}}}$ is
84 the quantile function of the $h_{\mathbf{k}}$ kernel distribution.

85 Similarly, the quantile mean can be defined as
86 $qm_{d, k, \epsilon, \gamma, n, \text{WL}} := \text{QI}_{d, h_{\mathbf{k}}=x, \mathbf{k}=1, k, \epsilon, \gamma, n, LU=\text{WL}}$. Moreover, in
87 extreme right-skewed heavy-tailed distributions, if the calcu-
88 lated percentile exceeds $1 - \epsilon$, it will be adjusted to $1 - \epsilon$.
89 In a left-skewed distribution, if the obtained percentile is
90 smaller than $\gamma\epsilon$, it will also be adjusted to $\gamma\epsilon$. Without loss
91 of generality, in the following discussion, only the case where
92 $\hat{F}_n (\text{WL}_{k, \epsilon, \gamma, n}) \geq \frac{\gamma}{1+\gamma}$ is considered. A widely used method
93 for calculating the sample quantile function involves employ-
94 ing linear interpolation of modes corresponding to the order
95 statistics of the uniform distribution on the interval $[0, 1]$, i.e.,
96 $\hat{Q}_n (p) = X_{[h]} + (h - [h]) (X_{\lceil h \rceil} - X_{\lfloor h \rfloor})$, $h = (n - 1)p + 1$.