Near-consistent robust estimations of moments for unimodal distributions

Tuban Lee

This manuscript was compiled on June 18, 2023

orderliness | invariant | unimodal | adaptive estimation | U-statistics

A. Congruent distribution. In the realm of nonparametric statistics, the relative differences, or orders, of robust estimators are of primary importance. A key implication of this principle is that when there is a shift in the parameters of the underlying distribution, all nonparametric estimates should asymptotically change in the same direction, if they are estimating the same attribute of the distribution. If, on the other hand, the mean suggests an increase in the location of the distribution while the median indicates a decrease, a contradiction arises. It is worth noting that such contradiction is not possible for any LL-statistics in a location-scale distribution, as explained in the previous article on semiparametric robust mean. However, it is possible to construct counterexamples to the aforementioned implication in a shape-scale distribution. In the case of the Weibull distribution, its quantile function is $Q_{Wei}(p) = \lambda(-\ln(1-p))^{1/\alpha}$, where $0 \leq p \leq 1$, $\alpha > 0$, $\lambda > 0$, λ is a scale parameter, α is a shape parameter, ln is the natural logarithm function. Then, $m = \lambda \sqrt[\alpha]{\ln(2)}$, $\mu = \lambda\Gamma\left(1+\frac{1}{\alpha}\right)$, where Γ is the gamma function. When $\alpha = 1$, $m = \lambda\ln(2) \approx 0.693\lambda$, $\mu = \lambda$, when $\alpha = \frac{1}{2}$, $m = \lambda\ln^2(2) \approx 0.480\lambda$, $\mu = 2\lambda$, the mean increases as α changes from 1 to $\frac{1}{2}$, but the median decreases. Previously, the fundamental role of quantile average and its relation to nearly all common nonparametric robust location estimates were demonstrated by using the method of classifying distributions through the signs of derivatives. To avoid such scenarios, this method can also be used. Let the quantile average function of a parametric distribution be denoted as $\mathrm{QA}(\epsilon, \gamma, \alpha_1, \cdots, \alpha_i, \cdots, \alpha_k)$, where α_i represent the parameters of the distribution, then, a distribution is γ-congruent if and only if the sign of $\frac{\partial \mathrm{QA}}{\partial \alpha_i}$ remains the same for all $0 \leq \epsilon \leq \frac{1}{1+\gamma}$. If $\frac{\partial \mathrm{QA}}{\partial \alpha_i}$ is equal to zero or undefined, it can be considered both positive and negative, and thus does not impact the analysis. A distribution is completely γ-congruent if and only if it is γ-congruent and all its central moment kernel distributions are also γ-congruent. Setting $\gamma = 1$ constitutes the definitions of congruence and complete congruence. Replacing the QA with γmHLM gives the definition of γ-U-congruence. Chebyshev's inequality implies that, for any probability distributions with finite second moments, as the parameters change, even if some LL-statistics change in a direction different from that of the population mean, the magnitude of the changes in the LL-statistics remains bounded compared to the changes in the population mean. Furthermore, distributions with infinite moments can be γ-congruent, since the definition is based on the quantile average, not the population mean.

The following theorems show the conditions that a distribution is congruent or γ-congruent.

Theorem A.1. *A γ-symmetric distribution is always γ-congruent and γ-U-congruent.*

Proof. As shown in RSSM I, Theorem .2 and Theorem .18, for any γ-symmetric distribution, all quantile averages and all γmHLMs conincide. The conclusion follows immediately. ☐

Theorem A.2. *A positive definite location-scale distribution is always γ-congruent.*

Proof. As shown in RSSM I, Theorem .2, for a location-scale distribution, any quantile average can be expressed as $\lambda\mathrm{QA}_0(\epsilon,\gamma)+\mu$. Therefore, the derivatives with respect to the parameters λ or μ are always positive. By application of the definition, the desired outcome is obtained. ☐

Theorem A.3. *The second central moment kernal distribution derived from a continuous location-scale unimodal distribution is always γ-congruent.*

Proof. Theorem **??** shows that the central moment kernel distribution generated from a location-scale distribution is also a location-scale distribution. Theorem **??** shows that it is positively definite. Implementing Theorem A.2 yields the desired result. ☐

For the Pareto distribution, $\frac{\partial Q}{\partial \alpha} = \frac{x_m(1-p)^{-1/\alpha}\ln(1-p)}{\alpha^2}$. Since $\ln(1-p) < 0$ for all $0 < p < 1$, $(1-p)^{-1/\alpha} > 0$ for all $0 < p < 1$ and $\alpha > 0$, so $\frac{\partial Q}{\partial \alpha} < 0$, and therefore $\frac{\partial \mathrm{QA}}{\partial \alpha} < 0$, the Pareto distribution is γ-congruent. It is also γ-U-congruent, since γmHLM can also express as a function of $Q(p)$. For the lognormal distribution, $\frac{\partial \mathrm{QA}}{\partial \sigma} = \frac{1}{2}\left(\sqrt{2}\mathrm{erfc}^{-1}(2\gamma\epsilon)\left(-e^{\frac{\sqrt{2}\mu-2\sigma\mathrm{erfc}^{-1}(2\gamma\epsilon)}{\sqrt{2}}}\right)+\left(-\sqrt{2}\right)\mathrm{erfc}^{-1}(2(1-\epsilon))e^{\frac{\sqrt{2}\mu-2\sigma\mathrm{erfc}^{-1}(2(1-\epsilon))}{\sqrt{2}}}\right)$. Since the inverse complementary error function is positive when the input is smaller than 1, and negative when the input is larger than 1, and symmetry around 1, if $0 \leq \gamma \leq 1$, $\mathrm{erfc}^{-1}(2\gamma\epsilon) \geq -\mathrm{erfc}^{-1}(2-2\epsilon)$, $e^{\mu-\sqrt{2}\sigma\mathrm{erfc}^{-1}(2-2\epsilon)} > e^{\mu-\sqrt{2}\sigma\mathrm{erfc}^{-1}(2\gamma\epsilon)}$. Therefore, if $0 \leq \gamma \leq 1$, $\frac{\partial \mathrm{QA}}{\partial \sigma} > 0$, the lognormal distribution is γ-congruent. Theorem A.1 implies that the generalized Gaussian distribution is congruent and U-congruent. For the Weibull distribution, when α changes from 1 to $\frac{1}{2}$, the average probability density on the left side of the median increases, since $\frac{\frac{1}{2}}{\lambda\ln(2)} < \frac{\frac{1}{2}}{\lambda\ln^2(2)}$, but the mean increases, indicating that the distribution is more heavy-tailed, the probability density of large values will also increase. So, the reason for non-congruence of the Weibull distribution lies

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 the simultaneous increase of probability densities on two opposite sides as the shape parameter changes: one approaching the bound zero and the other approaching infinity. Note that the gamma distribution does not have this issue, Numerical results indicate that it is likely to be congruent.

Although some parametric distributions are not congruent, Theorem A.2 establishes that γ-congruence always holds for a positive definite location-scale family distribution and thus for the second central moment kernel distribution generated from a location-scale unimodal distribution as shown in Theorem A.3. Theorem **??** demonstrates that all central moment kernel distributions are unimodal-like with mode and median close to zero, as long as they are generated from unimodal distributions. Assuming finite moments and constant $Q(0) - Q(1)$, increasing the mean of a distribution will result in a generally more heavy-tailed distribution, i.e., the probability density of the values close to $Q(1)$ increases, since the total probability density is 1. In the case of the $\mathbf{k}$th central moment kernel distribution, $\mathbf{k} > 2$, while the total probability density on either side of zero remains generally constant as the median is generally close to zero and much less impacted by increasing the mean, the probability density of the values close to zero decreases as the mean increases. This transformation will increase nearly all symmetric weighted averages, in the general sense. Therefore, except for the median, which is assumed to be zero, nearly all symmetric weighted averages for all central moment kernel distributions derived from unimodal distributions should change in the same direction when the parameters change.

B. A shape-scale distribution as the consistent distribution.

In the last section, the parametric robust estimation is limited to a location-scale distribution, with the location parameter often being omitted for simplicity. For improved fit to observed skewness or kurtosis, shape-scale distributions with shape parameter (α) and scale parameter (λ) are commonly utilized. Weibull, gamma, Pareto, lognormal, and generalized Gaussian distributions (when μ is a constant) are all shape-scale unimodal distributions. Furthermore, if either the shape parameter α or the skewness or kurtosis is constant, the shape-scale distribution is reduced to a location-scale distribution. Let $D(|skewness|, kurtosis, \mathbf{k}, etype, dtype, n) = d_{i\mathbf{k}m}$ denote the function to specify d values, where the first input is the absolute value of the skewness, the second input is the kurtosis, the third is the order of the central moment (if $\mathbf{k} = 1$, the mean), the fourth is the type of estimator, the fifth is the type of consistent distribution, and the sixth input is the sample size. For simplicity, the last three inputs will be omitted in the following discussion. Hold in awareness that since skewness and kurtosis are interrelated, specifying d values for a shape-scale distribution only requires either skewness or kurtosis, while the other may be also omitted. Since many common shape-scale distributions are always right-skewed (if not, only the right-skewed or left-skewed part is used for calibration, while the other part is omitted), the absolute value of the skewness should be the same as the skewness of these distributions. This setting also handles the left-skew scenario well.

For recombined moments up to the fourth ordinal, the object of using a shape-scale distribution as the consistent distribution is to find solutions for the system of equations

$$\begin{cases} rm\left(\mathrm{WL}, \gamma m, D(|rskew|, rkurt, 1)\right) = \mu \\ rvar\left(\mathrm{WHL}var, \gamma mvar, D(|rskew|, rkurt, 2)\right) = \mu_2 \\ rtm\left(\mathrm{WHL}tm, \gamma mtm, D(|rskew|, rkurt, 3)\right) = \mu_3 \\ rfm\left(\mathrm{WHL}fm, \gamma mfm, D(|rskew|, rkurt, 4)\right) = \mu_4 \\ rskew = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} \\ rkurt = \frac{\mu_4}{\mu_2^2} \end{cases},$$

where μ_2, μ_3 and μ_4 are the population second, third and fourth central moments. $|rskew|$ and $rkurt$ should be the invariant points of the functions $\varsigma(|rskew|) = \left|\frac{rtm(\mathrm{WHL}tm, \gamma mtm, D(|rskew|,3))}{rvar(\mathrm{WHL}var, \gamma mvar, D(|rskew|,2))^{\frac{3}{2}}}\right|$ and $\varkappa(rkurt) = \frac{rfm(\mathrm{WHL}fm, \gamma mfm, D(rkurt,4))}{rvar(\mathrm{WHL}var, \gamma mvar, D(rkurt,2))^2}$. Clearly, this is an overdetermined nonlinear system of equations, given that the skewness and kurtosis are interrelated for a shape-scale distribution. Since an overdetermined system constructed with random coefficients is almost always inconsistent, it is natural to optimize them separately using the fixed-point iteration (see Algorithm 1, only $rkurt$ is provided, others are the same).

Algorithm 1 $rkurt$ for a shape-scale distribution

Input: D; WHLvar; WHLfm; $\gamma mvar$; γmfm; $maxit$; δ
Output: $rkurt_{i-1}$

```
    i = 0
 2: rkurt_i ← ϰ(kurtosis_max)   ▷ Using the maximum kurtosis available in D as an initial guess.
    repeat
 4:     i = i + 1
        rkurt_{i-1} ← rkurt_i
 6:     rkurt_i ← ϰ(rkurt_{i-1})
    until i > maxit or |rkurt_i − rkurt_{i-1}| < δ   ▷ maxit is the maximum number of iterations, δ is a small positive number.
```

The following theorem shows the validity of Algorithm 1.

Theorem B.1. *Assuming $\gamma = 1$ and $m\mathbf{k}ms$, where $2 \leq \mathbf{k} \leq 4$, are all equal to zero, $|rskew|$ and $rkurt$, defined as the largest attracting fixed points of the functions $\varsigma(|rskew|)$ and $\varkappa(rkurt)$, are consistent estimators of $\tilde{\mu}_3$ and $\tilde{\mu}_4$ for a shape-scale distribution whose $\mathbf{k}$th central moment kernel distributions are γ-U-congruent, as long as they are within the domain of D, where $\tilde{\mu}_3$ and $\tilde{\mu}_4$ are the population skewness and kurtosis, respectively.*

Proof. Without loss of generality, only $rkurt$ is considered, while the logic for $|rskew|$ is the same. Additionally, the second central moments of the underlying sample distribution and consistent distribution are assumed to be 1, with other cases simply multiplying a constant factor according to Theorem **??**. From the definition of D, $\frac{\varkappa(rkurt_D)}{rkurt_D} = \frac{\frac{fm_D - \mathrm{SWHL}fm_D}{\mathrm{SWHL}fm_D - mfm_D}(\mathrm{SWHL}fm - mfm) + \mathrm{SWHL}fm}{rkurt_D\left(\frac{var_D - \mathrm{SWHL}var_D}{\mathrm{SWHL}var_D - mvar_D}(\mathrm{SWHL}var - mvar) + \mathrm{SWHL}var\right)^2}$, where the subscript D indicates that the estimates are from the central moment kernel distributions generated from the consistent distribution, while other estimates are from the underlying distribution of the sample.

Then, assuming the $m\mathbf{k}ms$ are all equal to zero and $var_D = 1$, $\frac{\varkappa(rkurt_D)}{rkurt_D} = \frac{\frac{fm_D - \mathrm{SWHL}fm_D}{\mathrm{SWHL}fm_D}(\mathrm{SWHL}fm) + \mathrm{SWHL}fm}{rkurt_D\left(\frac{\mathrm{SWHL}var}{\mathrm{SWHL}var_D}\right)^2} =$

$\frac{\left(\frac{fm_D - \mathrm{SWHL}fm_D}{\mathrm{SWHL}fm_D}+1\right)(\mathrm{SWHL}fm)}{fm_D\left(\frac{\mathrm{SWHL}var}{\mathrm{SWHL}var_D}\right)^2} = \frac{\mathrm{SWHL}fm\mathrm{SWHL}var_D^2}{\mathrm{SWHL}fm_D\mathrm{SWHL}var^2} = \frac{\frac{\mathrm{SWHL}fm}{\mathrm{SWHL}var^2}}{\frac{\mathrm{SWHL}fm_D}{\mathrm{SWHL}{var_D}^2}} = \frac{\mathrm{SWHL}kurt}{\mathrm{SWHL}kurt_D}$. Since $\mathrm{SWHL}fm_D$ are from the same fourth central moment kernel distribution as $fm_D = rkurt_D{var_D}^2$, according to the definition of γ-U-congruence, an increase in fm_D will also result in an increase in $\mathrm{SWHL}fm_D$. Combining with Theorem **??**, $\mathrm{SWHL}kurt$ is a measure of kurtosis that is invariant to location and scale, so $\lim_{rkurt_D \to \infty} \frac{\varkappa(rkurt_D)}{rkurt_D} < 1$. As a result, if there is at least one fixed point, let the largest one be fix_{max}, then it is attracting since $|\frac{\partial(\varkappa(rkurt_D))}{\partial(rkurt_D)}| < 1$ for all $rkurt_D \in [fix_{max}, kurtosis_{max}]$, where $kurtosis_{max}$ is the maximum kurtosis available in D.

□

As a result of Theorem B.1, assuming continuity, $m\mathbf{k}m$s are all equal to zero, and γ-U-congruence of the central moment kernel distributions, Algorithm 1 converges surely provided that a fixed point exists within the domain of D. At this stage, D can only be approximated through a Monte Carlo study. The continuity of D can be ensured by using linear interpolation. One common encountered problem is that the domain of D depends on both the consistent distribution and the Monte Carlo study, so the iteration may halt at the boundary if the fixed point is not within the domain. However, by setting a proper maximum number of iterations, the algorithm can return the optimal boundary value.

Data Availability. Data for Table **??** are given in SI Dataset S1. All codes are attached.

ACKNOWLEDGMENTS. I gratefully acknowledge the constructive comments made by the editor which substantially improved the clarity and quality of this paper.