Statistica Sinica **22** (2012), 655-679
doi:http://dx.doi.org/10.5705/ss.2010.155

ON THE DERIVATIVES OF THE TRIMMED MEAN

Subhra Sankar Dhar and Probal Chaudhuri

Indian Statistical Institute

Abstract: The trimmed mean is well-known for being more robust and for having better mean square error than the mean when data arise from non-Gaussian distributions with heavy tails. In this paper, we consider the derivatives of the trimmed mean with respect to the trimming proportion, and investigate some statistical applications of those derivatives. We develop a diagnostic tool based on the first derivative of the trimmed mean to determine whether the data is generated from a symmetric distribution or not. We also propose a test of symmetry of the distribution based on the first derivative, and demonstrate by theoretical and simulation studies that it performs better than several other well-known tests of symmetry. Further we introduce an estimate, based on the second derivative of the trimmed mean, of the contamination proportion $\beta \in (0, 1/2)$ in the contamination model $F(x) = (1-\beta)H(x) + \beta G(x)$, where H and G are two distributions such that G is stochastically larger than H. In addition to some theoretical studies, we carry out a detailed numerical study to show that, in many situations, our proposed estimate of the contamination proportion outperforms other estimates that are based on the idea of maximum likelihood estimation in mixture models.

Key words and phrases: Contamination model, Karhunen-Loeve expansion, Pitman efficacy, proportion of contamination, test of symmetry, weak convergence of processes.

1. Introduction

For a random sample $x_1, \ldots, x_n$, Tukey (1948) introduced the sample α-trimmed mean

$$\bar{x}_\alpha = \frac{1}{n - 2[n\alpha]} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} x_{(i)},$$

where $\alpha \in (0, 1/2)$, and $x_{(i)}$ is the ith order statistic of the sample. Tukey and McLaughlin (1963) proposed a trimmed version of the t-statistic using the α-trimmed mean. In a review paper on adaptive robust procedures, Hogg (1967) discussed some practical reasons for using the α-trimmed mean. Bickel (1965) derived the asymptotic distribution of the α-trimmed mean under appropriate regularity conditions, and Stigler (1973) obtained the same under a weaker setup.

Jaeckel (1971) proposed an estimate $\hat{\sigma}^2_\alpha$ for the asymptotic variance of the α-trimmed mean and used the value of α that minimizes $\hat{\sigma}^2_\alpha$ to construct an adaptive version of trimmed mean that was subsequently studied by Hall (1981). Recently, Dhar and Chaudhuri (2009) has shown that, for a large class of symmetric distributions with exponential and polynomial tails, the trimmed mean is more efficient than the least trimmed squares estimate, which is a robust estimate of location based on an alternative trimming procedure.

If we assume that the random sample consists of i.i.d. observations from an absolutely continuous distribution F having density f, the population analogue of the α-trimmed mean is

$$\theta(\alpha) = \frac{1}{(1-2\alpha)} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x f(x) dx$$

(see, e.g., Serfling (1980, p.236)). If F happens to be a symmetric distribution with the center of symmetry μ, it is obvious that $\theta(\alpha) = \mu$ for all α. When $\theta(\alpha)$ is a continuously differentiable function of $\alpha \in (0, 1/2)$,

$$\begin{aligned} \frac{d}{d\alpha}\theta(\alpha) &= \frac{2}{(1-2\alpha)^2} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x f(x) dx - \frac{1}{(1-2\alpha)} \{F^{-1}(\alpha) + F^{-1}(1-\alpha)\} = 0 \\ \Leftrightarrow \theta'(\alpha) &:= \frac{1}{(1-2\alpha)} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x f(x) dx - \frac{\{F^{-1}(\alpha) + F^{-1}(1-\alpha)\}}{2} = 0 \end{aligned}$$

for all $\alpha \in (0, 1/2)$. Therefore, one can develop diagnostic tools and tests for the presence or the absence of symmetry in the distribution based on estimates of $\theta'(\alpha)$. Such estimates are expected to be close to zero when the assumption of symmetry holds.

Several tests of symmetry have been proposed and studied in the literature. Mira (1999) proposed and investigated a test for $(Mean_F - Median_F) = 0$ based on the sample mean and sample median. Though Mira's test is easy to implement, the derivation of its asymptotic distribution requires finiteness of the second moment of the underlying distribution. Csorgo and Heathcote (1987) proposed a test based on the estimate of the characteristic function of F. Ahmad and Li (1997) proposed a test that compares kernel-based estimates of $f(x)$ and $f(-x)$. Schuster and Barker (1987) developed a test based on the Kolmogorov-Smirnov distance between the empirical distribution function $\hat{F}_n$ and its symmetrized version, and implemented the test using bootstrap techniques. In the next section, we develop a graphical tool based on the derivative of the α-trimmed mean to determine whether the data is generated from a symmetric distribution or not, and propose a new test of symmetry.

In Section 3, we propose and investigate an estimate of the contamination proportion β in the model $F(x) = (1-\beta)H(x) + \beta G(x)$ based on the second derivative of the α-trimmed mean. We show that our estimate, which is nonparametric in nature, outperforms well-known maximum likelihood type estimates for the contamination proportion in several examples.

2. Detection of Asymmetry Based on α-trimmed Mean

A common way to assess the presence or absence of symmetry in data is based on the histogram or some other density estimates of the data. However, the use of those density estimates is justified only when one has a reasonably large sample. It is not possible to form the class intervals and the frequency distribution in a sensible way in a small sample, the basic ingredients needed for construction. Density estimates are often statistically unreliable in small samples due to their high variability and slow convergence rates. In addition, the histogram or any other density estimate involves the choice of a smoothing parameter.

An assessment of symmetry or asymmetry in the data can be done using $\theta'(\alpha)$. Note that a natural estimate of $\theta'(\alpha)$ is $T_n(\alpha) := \bar{x}_\alpha - (\{\hat{F}_n^{-1}(\alpha) + \hat{F}_n^{-1}(1-\alpha)\})/2$, where $\hat{F}_n$ is the empirical distribution function of F. The following theorem describes the asymptotic behavior of the process $T_n(\alpha)$ as α varies over the interval $(0, 1/2)$.

Theorem 1. *Suppose the x_i's are i.i.d. with a continuous and positive density on the entire real line. Then, for any $0 < b_1 < b_2 < 1/2$, the process $\sqrt{n}\{T_n(\alpha) - \theta'(\alpha)\}$, where α varies over the interval $[b_1, b_2]$, converges weakly to a Gaussian process with zero mean and covariance function $\lim_{n\to\infty} \sqrt{n}E\{T_n(\alpha_1) - \theta'(\alpha_1)\}\{T_n(\alpha_2) - \theta'(\alpha_2)\} = k(\alpha_1, \alpha_2)$. For the form of $k(\alpha_1, \alpha_2)$, see Lemma 1 in Section 5.*

Corollary 1. *Under the conditions of Theorem 1, for any $0 < b_1 < b_2 < 1/2$ we have $\sup_{\alpha\in[b_1,b_2]} |T_n(\alpha) - \theta'(\alpha)| = O_P(n^{-1/2})$.*

It follows from Corollary 1 that if the graph of $T_n(\alpha)$ is roughly constant around zero, one may conclude that the data is obtained from a symmetric distribution. Also, using the asymptotic normality of $\sqrt{n}\{T_n(\alpha) - \theta'(\alpha)\}$, for each fixed $\alpha \in (0, 1/2)$, one can calculate the asymptotic p-value for the testing problem with the null hypothesis $H_{0,\alpha} : \theta'(\alpha) = 0$ against the alternative hypothesis $H_{1,\alpha} : \theta'(\alpha) \neq 0$ for each $0 < \alpha < 1/2$. This is illustrated in Figures 1 and 2 for some specific distributions. The symmetry of the generated data is quite visible in Figure 1, while the asymmetry in the data is clearly indicated in each of the diagrams in Figure 2.

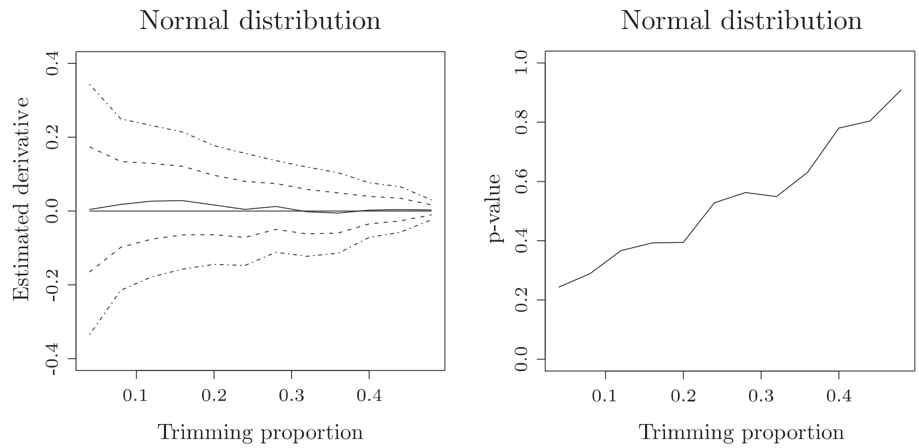

Figure 1. The graphs of the averages of $T_n(\alpha)$ and corresponding p-values (the *solid* curves) obtained from 50 Monte-Carlo replications of 25 i.i.d. standard normal observations. The *dashed* and the *dotted dashed* curves represent average $\pm$ (std. dev.) and average $\pm$ 2(std. dev.) limits, respectively.

2.1. Positive and negative skewness and their detection

Note that in Figure 2, the graphs of $T_n(\alpha)$ lie below the x-axis for almost all α. This indicates a monotonically decreasing nature of the α-trimmed mean, when data are generated from the gamma and the mixture normal distributions (see Figure 3). We now state a result that establishes the monotonically decreasing nature of α-trimmed means for distributions that are asymmetric in some appropriate sense.

Proposition 1. Suppose that F is a distribution function having continuous density f with $f(F^{-1}(\alpha)) \geq f(F^{-1}(1-\alpha))$ for all α. Then the α-trimmed mean $\theta(\alpha) = (1-2\alpha)^{-1} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} xf(x)dx$ associated with F is a decreasing function of α, and $\theta'(\alpha) \leq 0$ for all α.

The condition $f(F^{-1}(\alpha)) \geq f(F^{-1}(1-\alpha))$ for all α can be used as a definition of positively skewed distributions (see, e.g., van Zwet (1979)). If the graph of $T_n(\alpha)$ lies below the x-axis, one may conclude that the data are obtained from a positively skewed distribution.

2.2. A Statistical test for symmetry

We have already seen that $\theta'(\alpha) = 0$ for all $\alpha \in (0, 1/2)$ for a symmetric distribution, and hence one can formulate the problem of testing the hypothesis of symmetry as follows. For any b_1 and b_2 such that $0 < b_1 < b_2 < 1/2$, consider the hypothesis $H_0 : \theta'(\alpha) = 0$ for all $\alpha \in [b_1, b_2]$ against $H_1 : \theta'(\alpha) \neq 0$ for some

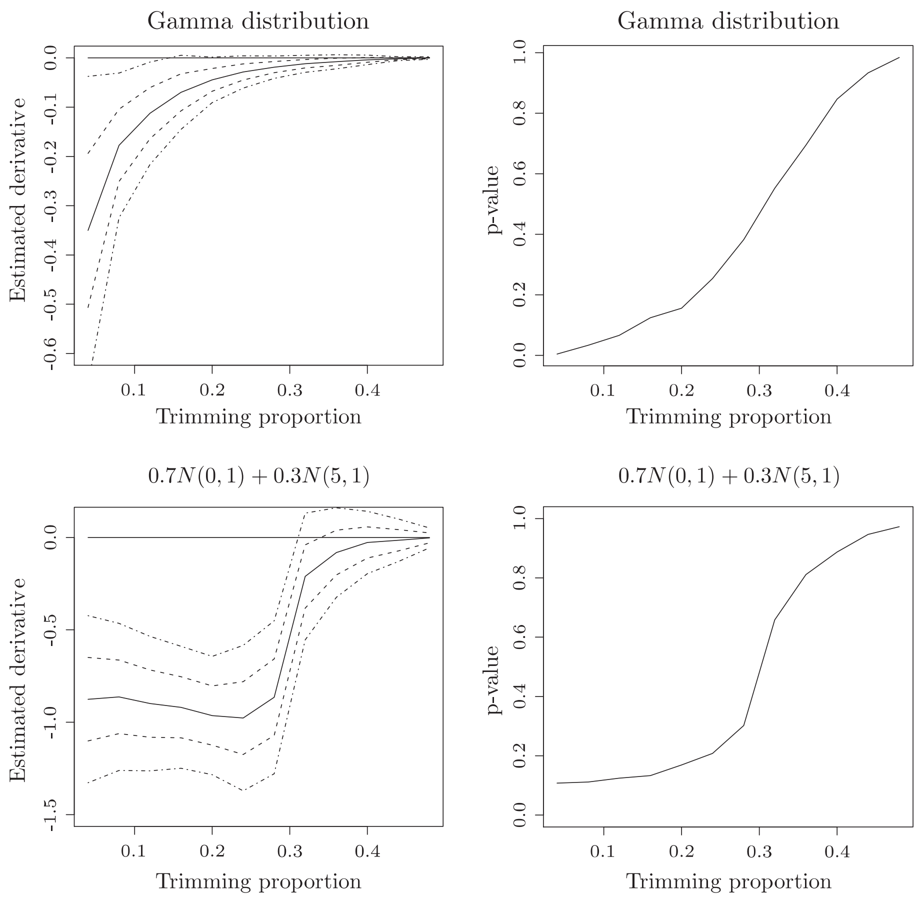

Figure 2. The graphs of the averages of $T_n(\alpha)$ and corresponding p-values (the *solid* curves) obtained from 50 Monte-Carlo replications of 25 i.i.d. observations from the gamma distribution with shape parameter = 0.15 and scale parameter = 1, and the $0.7N(0,1)+0.3N(5,1)$ distribution. The *dashed* and the *dotted dashed* curves represent average $\pm$ (std. dev.) and average $\pm$ 2(std. dev.) limits, respectively.

$\alpha \in [b_1, b_2]$. To test H_0 against H_1, consider $V_n := M[\sqrt{n}\{T_n(\cdot)\}]$, where $M : D([b_1, b_2]) \to [0, \infty)$ is a continuous map on $D[b_1, b_2]$, the space of real functions on $[b_1, b_2]$ that are right continuous and have left-hand limits. Here M is chosen to satisfy $M(\omega) = 0 \Leftrightarrow \omega(\alpha) = 0$ for all $\alpha \in [b_1, b_2]$, and for some $p \in [1, \infty]$, $M(\omega) \to \infty$ whenever $||\omega||_{[b_1,b_2],p} \to \infty$. We let $||\omega||_{[b_1,b_2],p} = \{\int_{b_1}^{b_2} |w(\alpha)|^p d\alpha\}^{1/p}$ if $p \in [1, \infty)$, and $||\omega||_{[b_1,b_2],\infty} = \sup_{\alpha\in[b_1,b_2]} |w(\alpha)|$.

Theorem 2. *Under H_0 and the conditions of Theorem 1, for any $0 < b_1 < b_2 < 1/2$ and $\alpha \in [b_1, b_2]$, $V_n = M\,[\sqrt{n}\{T_n(\cdot)\}]$ converges weakly to $M[Z(\cdot)]$,*

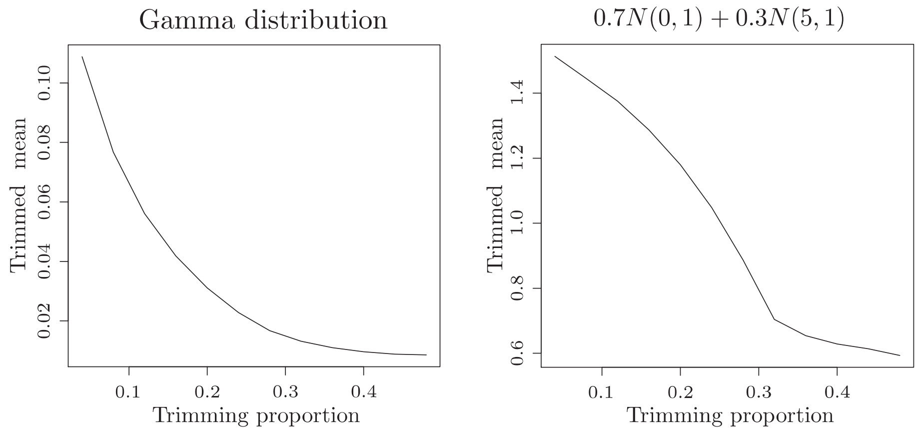

Figure 3. The graphs of the averages of $\bar{x}_\alpha$ obtained from 50 Monte-Carlo replications of 25 i.i.d. observations from gamma and $0.7N(0,1)+0.3N(5,1)$ distributions.

where $\{Z(\alpha) : \alpha \in [b_1, b_2]\}$ *is a Gaussian process with zero mean and covariance function* $k(\alpha_1, \alpha_2)$*,* $(\alpha_1, \alpha_2 \in [b_1, b_2])$*, which is as in Theorem* 1 *and Lemma 1. A test that rejects the null hypothesis* H_0 *when* $V_n > \xi_\rho$*, where* ξ_ρ *is the* $(1-\rho)$*th quantile of the distribution of* $M[Z(\cdot)]$*, has asymptotic size* ρ $(0 < \rho < 1)$*. Further, such a test is a consistent test in the sense that, under the alternative hypothesis, the asymptotic power of the test is 1 for appropriate choices of* b_1 *and* b_2*, depending on the asymmetric distribution.*

In our implementation of the test, we have used $M[\sqrt{n}\{T_n(\cdot)\}] = \int_{b_1}^{b_2} n\{T_n(\alpha)\}^2 \, d\alpha := \tau_n$. This choice of M satisfies the conditions stated before Theorem 2 if we take $p = 2$.

Theorem 3. *Under* H_0 *and the conditions of Theorem* 1*,* τ_n *converges weakly to* $\sum_{i=1}^{\infty} \lambda_i W_i$*, where the* W_i*'s are independent chi-square variables each with one degree of freedom. Here the* λ_i*'s are the eigen-values of the covariance function* $k(\alpha_1, \alpha_2)$ $(\alpha_1, \alpha_2 \in [b_1, b_2])$*, which is as in Theorem* 1 *and Lemma* 1*.*

In practice, τ_n can be approximated by $\sum_{i=[nb_1]+1}^{[nb_2]-1} \{T_n(i/n)\}^2$, and we have shown in Lemma 3 in Section 5 that $\tau_n - \sum_{i=[nb_1]+1}^{[nb_2]-1} \{T_n(i/n)\}^2 \xrightarrow{P} 0$ as $n \to \infty$. In order to implement the test, we first estimate the eigen-values λ_i of the covariance kernel $k(\alpha_1, \alpha_2)$ $(\alpha_1, \alpha_2 \in [b_1, b_2])$. We have used kernel density estimate based on standard Gaussian kernel and the adaptive choice of the bandwidth given as a default in the "ks" package in the statistical software R for f that appears in the expression of $k(\alpha_1, \alpha_2)$. Using estimated eigen-values $\hat{\lambda}_i$, one can generate a finite approximation to the sum $\sum_{i=1}^{\infty} \lambda_i W_i$. Finally, one can generate several Monte-Carlo replications of that finite sum and, depending on the specified level of the test, choose the critical value of the test as a specific quantile of the empirical distribution of that sum. In our implementations, we have taken $b_1 = 0.001$, $b_2 = 0.499$; these choices are based on our experience with computational costs and the performance of the test.

2.3. Finite sample study

We have carried out a simulation study to compare our proposed test with some others in terms of their powers and sizes. We considered tests with nominal levels 5% and 1% and, in order to estimate the power and the level of a test, we used the proportion of rejection of the null hypothesis of symmetry in several Monte-Carlo replications. We used 1,000 Monte-Carlo replications with each replication consisting of a sample of size 50 or 100, and computed the ratio between the power of our test based on τ_n (numerator) and that of another test (denominator). We considered different mixture distributions of the form $(1-\beta)H + \beta G$, as mentioned in Section 2.1, where H and G are two distribution functions, and β lies in the closed interval $[0, 1/2]$. In our investigation, we took H and G to be $N(\mu, 1)$, Cauchy $C(\mu, 1)$, and exponential $E(\mu)$ distributions with different location parameters μ. This ensured the inclusion of a wide variety of symmetric and asymmetric distributions.

In the case of normal mixture distributions, for different values of β, the range of the ratios between the powers of our test and the powers of the tests considered by Mira (1999), Csorgo and Heathcote (1987), Schuster and Barker (1987), and Ahmad and Li (1997) turns out to be 4.53–6.51. For a Cauchy mixture, a normal and exponential mixture, and a Cauchy and exponential mixture, the ranges of ratios of powers were 1.02–1.41, 1.01–1.19, and 1.15–1.35, respectively. More details about the powers of different tests are presented in Figure 6.1 in the supplementary material.

2.4. Asymptotic efficiency study under contiguous alternatives

In this section, we study the asymptotic power properties of different tests by deriving their Pitman efficacies under contiguous alternatives. We consider the contamination model $(1 - \delta/\sqrt{n})H(x) + \delta/\sqrt{n}G(x)$ to form the sequence of alternative hypotheses, where H is a symmetric distribution and G is any distribution function that is stochastically larger than H. Thus we test $H_0 : F(x) = H(x)$ vs. $H_n : F(x) := F_n(x) = (1 - \delta/\sqrt{n})H(x) + (\delta/\sqrt{n})G(x)$ for a fixed $\delta > 0$.

Theorem 4. *Assume that H and G have continuous and positive densities h and g, respectively and $E_h\{g(x)/h(x) - 1\}^2 < \infty$. Then the sequence of alternatives H_n form a contiguous sequence. Let $x_1, \ldots, x_n$ be i.i.d. observations from F, and $\psi(x,t)$ be a real valued function of x and t, each varying on some interval on the line such that $E_h\{\psi(x_i,t)\}^2 < \infty$. Assume that the process $\sqrt{n}((1/n)\sum_{i=1}^n \psi(x_i,t) - E_h\{\psi(x_i,t)\})$ converges weakly to a Gaussian process with zero mean and covariance function $k_1(t_1,t_2) = E_h[\psi(x_i,t_1) - E_h\{\psi(x_i,t_1)\}, \psi(x_i,t_2) - E_h\{\psi(x_i,t_2)\}]$ under H_0 as $n \to \infty$. Then, under the sequence H_n, this process converges weakly to a Gaussian process with the same covariance function but with mean function $m(t,\delta) = \delta[E_g\{\psi(x,t)\} - E_h\{\psi(x,t)\}]$.*

Under H_0, the asymptotic distributions of our test statistic τ_n, the test statistic $T_n^{(1)} = \sup_t |(1/n)\sum_{i=1}^n (1_{\{x_i \le t\}} + 1_{\{x_i \le 2\eta - t\}} - 1)/2|$ of Schuster and Barker (1987), and the test statistic $T_n^{(2)}$ of Csorgo and Heathcote (1987) follow from Theorem 1, and results of Arcones and Gine (1991), and Csorgo and Heathcote (1987), respectively. Under contiguous alternatives H_n, using Theorem 4, one can obtain the asymptotic distributions of $T_n^{(1)}$ and $T_n^{(2)}$ by taking $\psi(x,t) = (1_{\{x \le t\}} + 1_{\{x \le 2\eta - t\}} - 1)/2$ and $\psi(x,t) = [(U(t)\{\sin t(x)\} - V(t)\{\cos t(x)\})/(t\{U^2(t) + V^2(t)\})]$, respectively. Further, one can derive the asymptotic distribution of τ_n under H_n by using $\psi(x,\alpha) = \{H^{-1}(\alpha)1_{\{x \le H^{-1}(\alpha)\}} + x1_{\{H^{-1}(\alpha) \le x \le H^{-1}(1-\alpha)\}} + H^{-1}(1-\alpha)1_{\{x \ge H^{-1}(1-\alpha)\}}\}/(1 - 2\alpha) - (1/2)[(\alpha - 1_{\{x \le H^{-1}(\alpha)\}})/\{h(H^{-1}(\alpha))\} + ((1-\alpha) - 1_{\{x \le H^{-1}(1-\alpha)\}})/\{h(H^{-1}(1-\alpha))\}]$. More generally, for the statistic V_n defined before Theorem 2, the asymptotic distribution is the same as the distribution of $M[Z_1(\cdot)]$, where $\{Z_1(\alpha) : \alpha \in [b_1, b_2]\}$ is the Gaussian process of Table 1. On the other hand, the asymptotic distribution of Mira's (1999) test statistic $T_n^{(3)} = 2(mean - median)$, under H_0 as well as under the sequence H_n, can be established using straightforward applications of Bahadur's asymptotic expansion of the sample median (see, e.g., Serfling (1980)) and Lecam's third lemma in Hajek and Sidak (1967). These asymptotic distributions of different test-statistics are summarized in Table 1.

In order to evaluate Pitman efficacies of different tests, we have chosen H to be $N(0,1)$, Laplace $L(0,1)$ and Cauchy $C(0,1)$ distributions, and G to be exponential with mean 1, $N(\mu,1)$, $L(\mu,1)$, and $C(\mu,1)$ with $\mu = 1/2, 1$ and 2. We have taken the asymptotic size of all the tests as 0.05, and their asymptotic powers were chosen to be $0.1, 0.2, 0.3, \ldots, 0.9$. The Pitman efficacy (see, e.g., Serfling (1980) and Lehmann and Romano (2005)) of our test relative to another test for varying choices of asymptotic power determined by δ is $(\delta'/\delta)^2$, where δ and δ' are such that the asymptotic power of our test under contiguous alternative $(1-\delta/\sqrt{n})H(x) + (\delta/\sqrt{n})G(x)$ is the same as the asymptotic power of the other test under the alternative $(1-\delta'/\sqrt{n})H(x) + (\delta'/\sqrt{n})G(x)$.

Table 1. Asymptotic distributions of test statistics for different tests. η is the center of symmetry of H, $k(\alpha_1, \alpha_2)$ is as defined in Theorem 1, $\sigma_M^2 = 4\sigma_H^2 + \left\{h(H^{-1}(1/2))\right\}^{-2} - (4/h(H^{-1}(1/2))) \times \{E_H x - 2\int_{-\infty}^{H^{-1}(1/2)} x dH(x)\}$, $\sigma_H^2 = Var_H(x)$ and s_0, t_0, σ, U and V are as defined in Csorgo and Heathcote (1987).

Test	Asymptotic distribution of test statistic under contiguous alternatives
Our test based on τ_n	The distribution of $\int_{b_1}^{b_2}\{Z_1(\alpha)\}^2 d\alpha$, where $Z_1(\alpha)$ is a Gaussian process with mean function $\delta E_G\{\psi(x,\alpha)\}$ and covariance kernel $k(\alpha_1,\alpha_2)$.
Schuster and Barker's (1987) test	The distribution of $\sup_{t\in R}\|Y(t)\|$, where $Y(t)$ is a Gaussian process with mean function $\frac{\delta}{2}\{G(t)+G(2\eta-t)-1\}$ and covariance kernel $k_2(s,t)=\frac{1}{4}[2+2H(\min(s,t))]$.
Mira's (1999) test	Gaussian distribution with mean $2\delta E_G\left[\left\{x - \frac{1/2 - 1_{\{x\le H^{-1}(1/2)\}}}{h(H^{-1}(1/2))}\right\}\right] - 2\delta E_H x$ and variance σ_M^2.
Csorgo and Heathcot's (1987) test	Gaussian distribution with mean $\delta\left[\frac{U(t_0)E_G\{\sin t_0(x)\} - V(t_0)E_G\{\cos t_0(x)\}}{t_0\{U^2(t_0)+V^2(t_0)\}\{\sigma^2(s_0)+\sigma^2(t_0)-2\sigma(s_0,t_0)\}}\right] - $ $-\delta\left[\frac{U(s_0)E_G\{\sin s_0(x)\} - V(s_0)E_G\{\cos s_0(x)\}}{s_0\{U^2(s_0)+V^2(s_0)\}\{\sigma^2(s_0)+\sigma^2(t_0)-2\sigma(s_0,t_0)\}}\right]$ and variance 1.

For different values of asymptotic power and μ, the ranges of Pitman efficacies of our test relative to tests studied by Mira (1999), Schuster and Barker (1987), and Csorgo and Heathcote (1987) are $12.39 - 15.68$, $1.99 - 3.20$ and $2.28 - 3.46$, respectively, when $H = N(0,1)$ and $G = N(\mu,1)$. When $H = L(0,1)$ and $G = L(\mu,1)$, those ranges turn out to be $16.08 - 20.16$, $3.80 - 9.48$ and $4.97 - 8.94$, respectively. When $H = N(0,1)$, and G is chosen as exponential distribution with mean 1, the ranges are $2.99 - 4.75$, $1.46 - 3.31$ and $1.61 - 3.96$, respectively. For $H = C(0,1)$ and $G = C(\eta,1)$, the ranges for tests considered by Schuster and Barker (1987) and Csorgo and Heathcote (1987) are $9.05 - 17.16$ and $7.82 - 17.60$, respectively. In this last case, we did not consider Mira's (1999) test because the sample mean does not have finite moments under the Cauchy distribution. Thus our test asymptotically outperforms all other tests considered here under the chosen sequences of contiguous alternatives. Further details about Pitman efficacies of different tests are presented in Figure 6.2 in the supplementary material.

2.5. Analysis of data

In this section, we have analyzed three data sets. Detailed information of the iris and the yeast data is available in `http://archive.ics.uci.edu/ml`, and the diabetes data can be obtained from the "mclust" package in the software R.

At 5% level, the tests considered in Section 2.3 rejected the hypothesis of symmtry for petal widths of *Iris setosa* and *Iris versicolor*, for the variable s.s.p.g.

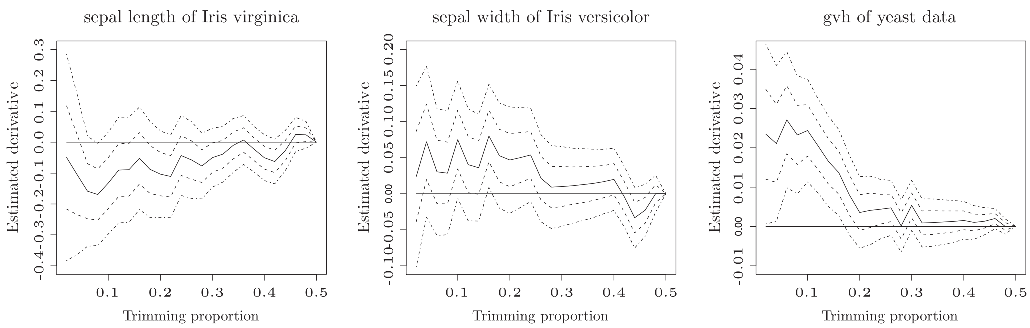

Figure 4. Graphs of $T_n(\alpha)$ for the sepal length of *Iris virginica*, the sepal width of *Iris versicolor*, and the variable gvh of the yeast data. The *dashed* and the *dotted dashed* curves represent average $\pm$ (std. dev.) and average $\pm$ 2(std. dev.), respectively.

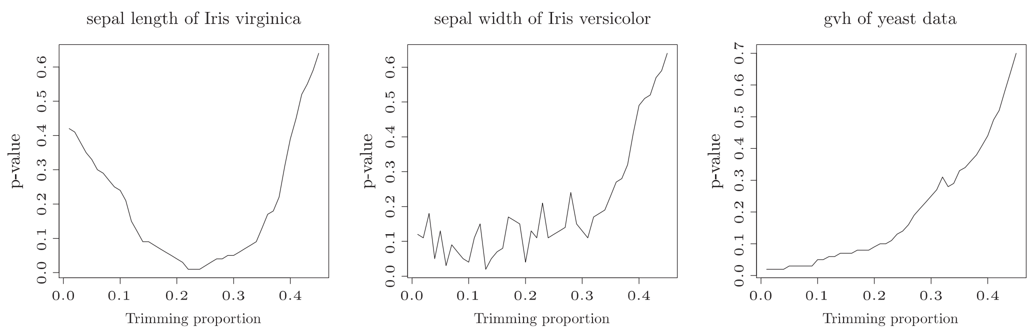

Figure 5. Graphs of p-values for different values of α of the sepal length of *Iris virginica*, the sepal width of *Iris versicolor*, and the variable gvh of the yeast data.

in the case of normal individuals, and for all three variables in the case of overt and chemical diabetic individuals. Also, at the 5% level, all tests accepted the hypothesis of symmetry for sepal length, sepal width, and petal length in the case of *Iris setosa*, for sepal width in the case of *Iris verginica*, for petal length and petal width in the case of *Iris verginica*, for sepal length and petal length in the case of *Iris versicolor*, and for glucose level and insulin area in the case of normal individuals. In the case of the yeast data, all tests accepted the null hypothesis of symmetry at 5% level for all the variables except gvh.

However, at the 5% level, our test rejected the null hypothesis of symmetry for sepal length in the case of *Iris virginica*, for the sepal width in the case of *Iris versicolor*, and for the variable gvh in the case of the yeast data, while the other four tests accepted the null hypothesis of symmetry for these three variables. The graphs of $T_n(\alpha)$ for $0 < \alpha < 1/2$ and the corresponding p-values for the sepal length of *Iris virginica*, the sepal width of *Iris versicolor*, and the variable gvh in the yeast data are in Figures 4 and 5. It is evident there that the distributions of these three variables are asymmetric; our test rejects the null hypothesis for these three variables, while the other tests fail to detect the asymmetry.

3. The Second Derivative of α-trimmed Mean and Estimation of Contamination Proportion

Consider the contamination model $F(x) = (1-\beta)H(x) + \beta G(x)$, where $\beta \in (0, 1/2)$ and H and G are such that G is stochastically larger than H. For estimating the location parameter, Huber (1981, pp.74-75) showed that median achieves the smallest bias among all translation invariant functionals. However, the bias associated with median is strictly nonzero and depends on the contamination proportion β. Huber also showed that the maximal asymptotic bias and variance of the α-trimmed mean for data following a β-contaminated asymmetric normal distribution is finite when $\alpha \geq \beta$, and is infinite when $\alpha < \beta$. This indicates the importance of assessing the extent of contamination in the data.

Note the sharp curvature in the graph of the average of α-trimmed mean when α is close to the contamination proportion 0.3, in Figure 3. Also, in Figure 2, for α close to 0.3, we see a sharp change in the graphs that plot the averages over several Monte-Carlo simulations of $T_n(\alpha)$ and the p-value. This behavior of the α-trimmed mean and its derivative motivated us to investigate the behavior of the second derivative of α-trimmed mean when data are generated from a contamination model. It is given by

$$\theta''(\alpha) = \frac{2}{1-2\alpha}\Big[\theta(\alpha) - \frac{1}{2}\{F^{-1}(\alpha) + F^{-1}(1-\alpha)\} \\ -\frac{1}{2}\Big\{\frac{1}{f(F^{-1}(\alpha))} - \frac{1}{f(F^{-1}(1-\alpha))}\Big\}\Big].$$

One can use

$$S_n(\alpha) := \frac{2}{1-2\alpha}\Big[\bar{x}_\alpha - \frac{1}{2}\{\hat{F}_n^{-1}(\alpha) + \hat{F}_n^{-1}(1-\alpha)\} \\ -\frac{1}{2}\Big\{\frac{1}{\hat{f}_n(\hat{F}_n^{-1}(\alpha))} - \frac{1}{\hat{f}_n(\hat{F}_n^{-1}(1-\alpha))}\Big\}\Big]$$

as a natural estimate of $\theta''(\alpha)$. Here $\bar{x}_\alpha$ is the sample α-trimmed mean, $\hat{F}_n$ is the empirical distribution function, and $\hat{f}_n$ is some suitable estimate of the density f as before. In our numerical work, we have estimated $\hat{f}_n$ as in the last paragraph in Section 2.2.

In Figure 6, we have plotted the average of the values of $S_n(\alpha)$, where observations are obtained from the mixture normal distribution $0.7N(0,1)+0.3N(5,1)$. There the maxima of $S_n(\alpha)$ is close to $\beta = 0.3$. This motivated us to investigate

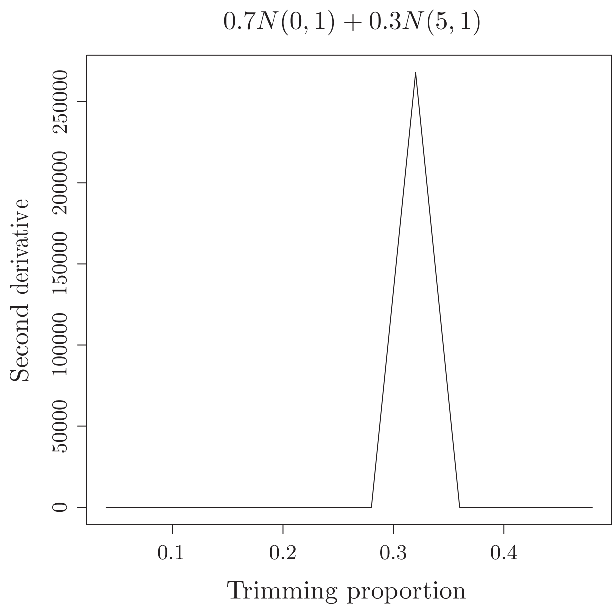

Figure 6. The graph of the average of $S_n(\alpha)$ obtained from 50 Monte-Carlo replications each consisting of 25 i.i.d. observations from the $0.7N(0,1) + 0.3N(5,1)$ distribution.

the behavior of the maxima of $\theta''(\alpha)$. In the discussion following Theorem 5, we note that the behavior of the maxima of $\theta''(\alpha)$ depends on the overlap between the distributions H and G, where G is a stochastically larger than H. A quantitative measure of the overlap between H and G is the common probability mass $\Delta_{H,G}$ between H and G, $\Delta_{H,G} = \int_{-\infty}^{\kappa} g(x)dx + \int_{\kappa}^{\infty} h(x)dx$. Here κ is the unique intersecting point between h and g, the density functions of H and G, respectively, if we assume that such a unique point of intersection exists. On the other hand, if H and G have disjoint supports, we set $\Delta_{H,G} = 0$.

One can find a closed form expression for Δ_{H,H_ϕ} in the case of the model $F(x) = (1-\beta)H(x) + \beta H_\phi(x)$, where H has a symmetric unimodal density, ϕ is a location shift, and $H_\phi(x) = H(x-\phi)$. It follows that

$$\Delta_{H,H_\phi} = \int_{-\infty}^{\phi/2} h(x-\phi)dx + \int_{\phi/2}^{\infty} h(x)dx = H(-\frac{\phi}{2}) + 1 - H(\frac{\phi}{2}) = 2\left\{1 - H(\frac{\phi}{2})\right\}.$$

In this case, $\phi/2$ is the intersecting point of the densities of $H(x)$ and $H(x-\phi)$.

Theorem 5. *Consider the model $F(x) = (1-\beta)H(x) + \beta G(x)$ where $\beta \in (0, 1/2)$, H is any distribution function having a continuous density h, and G is a stochastically larger than H with continuous density g. Suppose that h and g are positive on any compact subinterval strictly inside the supports of h*

and g, and $\lim_{\alpha\to 0+} h(H^{-1}(\alpha)) = \lim_{\alpha\to 1-} h(H^{-1}(\alpha)) = \lim_{\alpha\to 0+} g(G^{-1}(\alpha)) = \lim_{\alpha\to 1-} g(G^{-1}(\alpha)) = 0$. *Assume that* $f(F^{-1}(\alpha)) > f(F^{-1}(1-\alpha))$ *for all* α, *where* $f = (1-\beta)h + \beta g$ *is the density function of* F. *When* h *is supported on a compact interval and its support is disjoint from that of* g, *we have* $\sup_{\alpha\in[\beta-\gamma,\beta+\gamma],\alpha\neq\beta} \theta''(\alpha) = \infty$ *and* $\sup_{\alpha\notin[\beta-\gamma,\beta+\gamma]} \theta''(\alpha) < \infty$ *for any* $\gamma > 0$. *For any given* H *with its density* h *supported on the entire real line, if* G *varies in such a way that* $\inf_{q\in(0,1)} \left\{G^{-1}(q) - H^{-1}(q)\right\} \to \infty$, *we have* $\theta''(\beta) \to \infty$ *while* $\sup_{\alpha\notin[\beta-\gamma,\beta+\gamma]} \theta''(\alpha)$ *remains bounded above for any* $\gamma > 0$.

Note that the preceding theorem implies that when $\Delta_{H,G}$ is zero or close to zero and α lies in a small neighborhood of β, $\theta''(\alpha)$ assumes very large values. On the other hand, $\theta''(\alpha)$ takes relatively smaller values for α lying outside that neighborhood. Hence, if we obtain a maximizer of an appropriate estimate of $\theta''(\alpha)$ with respect to α, that maximizer will be close to β for small values of $\Delta_{H,G}$. This behavior of the maxima of $\theta''(\alpha)$ motivated us to propose the estimate

$$\hat{\beta} = arg \max_{\alpha\in[b_1,b_2]} S_n(\alpha), 0 < b_1 < b_2 < \frac{1}{2}.$$

Overall, in view of the preceding discussion, for any contamination model of the form $(1-\beta)H + \beta G$ satisfying the conditions stated in Theorem 5, it is expected that the performance of $\hat{\beta}$ will be good when $\Delta_{H,G}$ is small. In the following subsection, we investigate the behavior of $\hat{\beta}$ in different location contamination models with varying choices of $\Delta_{H,G}$.

3.1. A comparison with maximum likelihood estimators

The contamination model described in this section can be viewed as a special case of mixture models. The estimation of mixing proportion in mixture models is thoroughly discussed in Everitt and Hand (1981) using maximum likelihood and related techniques. We have compared the performance of our estimate with some other estimates of β based on the idea of maximum likelihood, with $\Delta_{H,G} = 10\%$, 15%, and 20% for location contamination models involving normal, Cauchy, and Laplace distributions. Using the relation $\Delta_{H,H_\phi} = 2\left\{1 - H(\phi/2)\right\}$, we have considered appropriate values of the location shift ϕ and varying choices of $\beta \in (0, 1/2)$ in the simulation study.

We simulated $m = 1{,}000$ samples from each distribution with sample sizes 100 and 1,000, and calculated mean square error (m.s.e.) $= (1/m)\sum_{i=1}^{m}(b_i - \beta)^2$ for different estimates. Here b_i is the estimate of β for the ith sample. We computed the efficiency of our estimate relative to other estimates, where the efficiency of an estimate E_1 relative to another estimate E_2 is $m.s.e.(E_2)/m.s.e.(E_1)$.

In the case of mixtures of normal distributions, for different values of contamination proportion and $\Delta_{H,G}$, the ranges of efficiencies of our estimate relative to the estimate based on the E.M. algorithm and the Newton-Raphson method are 1.25–5.01 and 1.31–5.25, respectively. In the case of mixtures of Cauchy distributions, the EM algorithm did not converge in our numerical studies. Here the range of efficiency of our estimate relative to the estimate based on the Newton-Raphson method is 0.97–2.52 for different values of contamination proportion and $\Delta_{H,G}$. For the mixtures of Laplace distributions, the Newton-Raphson method is not feasible as the density functions involved are not differentiable. However, the EM algorithm can be carried out in this case, and the range of efficiency of our estimate relative to the estimate based on the E.M. algorithm is 0.96–2.87 for different values of contamination proportion and $\Delta_{H,G}$. The computation of estimates based on the Newton-Raphson method were done by the "micEcon" package in the Statistical software R, and we used the R codes given in Horton, Brown, and Qian (2004, p.353) to compute estimates based on the E.M. algorithm. In all our computations, we used the true values of the parameters to start the iterations. Detailed results obtained from these finite sample efficiency studies are in Section 6.1 of the supplementary material.

4. Some Concluding Remarks

The Pitman efficacy of Ahmad and Li's (1997) test of symmetry based on kernel density estimate is not mentioned in Section 2.4 because the kernel density estimates converge at rates slower than the $n^{-1/2}$, and consequently, the Pitman efficacy of that test relative to ours is zero.

Our nonparametric estimate of contamination proportion, in Section 3, does not require iterative computation while other competing estimates available in the literature are based on the iterative procedures, and the performance of those estimates depend on the chosen initial values. Moreover, second order differentiability of the model is required for computing estimates by the Newton-Raphson method, whereas we need only the existence of the density function.

When an appropriate parametric model holds for the data, the maximum likelihood estimate of the contamination proportion is $\sqrt{n}$-consistent, asymptotically normal, and its asymptotic variance coincides with the Cramer-Rao lower bound. So, a natural question is how does the m.s.e. of $\hat{\beta}$ compare with the true Cramer-Rao lower bound for a specified parametric model. We have computed the efficiency of $\hat{\beta}$ relative to the Cramer-Rao lower bound when data follow mixtures of normal and Cauchy distributions. The range of those efficiency values was 0.453–0.871 for sample size 100, and 0.193–0.454 for sample size 1,000.

We close by pointing out that the power and size of our test, and the M.S.E. of $\hat{\beta}$, depend on the choice of the bandwidth. We used different choices. Following the suggestion in Ghosh, Chaudhuri, and Sengupta (2006), we took $l_{0.05}/3$ and $l_{0.95}$ as the lower and the upper limits of the bandwidths, respectively. Here

l_q is the $q \in [0,1]$th quantile of the pair-wise differences of the observations in a given data set. For different choices of bandwidth, the differences between the maximum power and the power of our test based on an adaptive choice of the bandwidth was at most 2%. Further details about the powers of the tests for varying choices of bandwidths can be found in Figure 6.1 in the supplementary material. We also found that the ranges of efficiencies of our estimate corresponding to the minimum M.S.E. over different choices of bandwidths relative to the estimate based on adaptive choice of bandwidth were $1.03 - 1.07$ and $1.01 - 1.04$ for sample sizes 100 and 1,000, respectively. It appears that the adaptive choice of the bandwidth used in the construction of $\hat{\beta}$ works quite well in this case.

5. Proofs of the Theorems

For Theorem 1, we first need to prove following two lemmas.

Lemma 1. *If observations are from an absolutely continuous distribution function F with a positive and continuous density f on the entire real line, then for any $\alpha_1, \alpha_2, \ldots, \alpha_k \in (0,1/2)$, $k > 1$, the asymptotic distribution of $\sqrt{n}(T_n(\alpha_1) - \theta'(\alpha_1), T_n(\alpha_2) - \theta'(\alpha_2), \ldots, T_n(\alpha_k) - \theta'(\alpha_k))$ is k-variate normal with zero mean and a variance-covariance matrix in which the (i,j)th entry $(1 \le i \le j \le k)$ is*

$$
\begin{aligned}
&k(\alpha_i,\alpha_j)\\
&= \frac{2\int_0^{F^{-1}(1-\alpha_j)} x^2 f(x)dx + 2F^{-1}(\alpha_j)\int_{F^{-1}(\alpha_i)}^{F^{-1}(\alpha_j)} xf(x)dx + 2\alpha_i F^{-1}(\alpha_i)F^{-1}(\alpha_j)}{(1-2\alpha_i)(1-2\alpha_j)}\\
&\quad + \frac{\alpha_i F^{-1}(\alpha_i)}{2(1-2\alpha_j)f(F^{-1}(\alpha_i))} + \frac{\alpha_i F^{-1}(\alpha_j)}{2(1-2\alpha_j)f(F^{-1}(1-\alpha_i))}\\
&\quad + \frac{\alpha_i F^{-1}(\alpha_i) + \int_{F^{-1}(\alpha_i)}^{F^{-1}(\alpha_j)} xf(x)dx}{2(1-2\alpha_i)f(F^{-1}(\alpha_j))} + \frac{\alpha_i(1-\alpha_j)}{4f(F^{-1}(\alpha_i))f(F^{-1}(\alpha_j))}\\
&\quad + \frac{\alpha_i\alpha_j}{4f(F^{-1}(\alpha_i))f(F^{-1}(1-\alpha_j))} + \frac{\alpha_i F^{-1}(\alpha_i) + \int_{F^{-1}(\alpha_i)}^{F^{-1}(1-\alpha_j)} xf(x)dx}{2(1-2\alpha_i)f(F^{-1}(1-\alpha_j))}\\
&\quad + \frac{\alpha_i\alpha_j}{4f(F^{-1}(1-\alpha_i))f(F^{-1}(\alpha_j))} + \frac{\alpha_i(1-\alpha_j)}{4f(F^{-1}(1-\alpha_i))f(F^{-1}(1-\alpha_j))}.
\end{aligned}
$$

Proof of Lemma 1. Recall that $T_n(\alpha) = \bar{x}_\alpha - \{\hat{F}_n^{-1}(\alpha) + \hat{F}_n^{-1}(1-\alpha)\}/2$, where $\alpha \in (0,1/2)$. As given in DasGupta (2008), we have

$$
\begin{aligned}
&\bar{x}_\alpha - \frac{1}{(1-2\alpha)}\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} xf(x)dx\\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{F^{-1}(\alpha)1_{\{x_i \le F^{-1}(\alpha)\}} + x_i 1_{\{F^{-1}(\alpha) \le x_i \le F^{-1}(1-\alpha)\}} + F^{-1}(1-\alpha)1_{\{x_i \ge F^{-1}(1-\alpha)\}}}{(1-2\alpha)}
\end{aligned}
$$

$$+o_P\left(\frac{1}{\sqrt{n}}\right).$$

From Serfling (1980), we have

$$\hat{F}_n^{-1}(\alpha) - F^{-1}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \frac{\alpha - 1_{\{x_i \le F^{-1}(\alpha)\}}}{f(F^{-1}(\alpha))} + R_n$$

where, as $n \to \infty$, $R_n = O(n^{-3/4}(\log n)^{3/4})$ with probability 1. Using these linear expansions of α-trimmed mean and quantiles, it is straightforward to see that $\sqrt{n}(T_n(\alpha_1) - \theta'(\alpha_1), \ldots, T_n(\alpha_k) - \theta'(\alpha_k))$ can be written as a sum of i.i.d. k-variate random vectors along with a remainder term that goes to zero in probability. By an application of the C.L.T. and Slutsky's Theorem, we have the asymptotic normality of $\sqrt{n}(T_n(\alpha_1) - \theta'(\alpha_1), \ldots, T_n(\alpha_k) - \theta'(\alpha_k))$ with zero mean. The asymptotic variance-covariance matrix as given in the statement of the lemma can be obtained by a direct algebraic computation.

Lemma 2. *Under the assumptions stated in Lemma* 1 *and for* $\alpha \in [b_1, b_2]$*, the stochastic process* $\sqrt{n}\{T_n(\alpha) - \theta'(\alpha)\}$ *is tight, where* $0 < b_1 < b_2 < 1/2$.

Proof of Lemma 2. First, for any distribution function F and $\alpha \in (0, 1/2)$, we set $F^{-1}(\alpha) = \inf_x\{x : F(x) \ge \alpha\}$ and $F^{-1}(1-\alpha) = \sup_x\{x : F(x) < (1-\alpha)\}$ so that $F^{-1}(\alpha)$ and $F^{-1}(1-\alpha)$ both become right continuous function of $\alpha \in (0, 1/2)$ and have left-hand limits. It follows from the definitions of $\bar{x}_\alpha$, $F^{-1}(\alpha)$, and $F^{-1}(1-\alpha)$ that the process $T_n(\alpha)$ for $\alpha \in [b_1, b_2]$ lies in $D[b_1, b_2]$ for each $n \ge 1$, where $D[b_1, b_2]$ is the space of real functions on $[b_1, b_2]$ that are right continuous and have left-hand limits (see Billingsley (1999, p.121)). In order to prove the tightness of $T_n(\alpha)$, one needs to verify the two conditions stated in Theorem 13.2 in Billingsley (1999, p.139). It follows from Theorem 3.1 in Bickel (1967) that for $\alpha \in [b_1, b_2]$, where $0 < b_1 < b_2 < 1/2$, and under the conditions stated in Lemma 1, the standardized quantile process converges weakly to a Gaussian process. This implies that the processes $\sqrt{n}\{\hat{F}_n^{-1}(\alpha) - F^{-1}(\alpha)\}$ and $\sqrt{n}\{\hat{F}_n^{-1}(1-\alpha) - F^{-1}(1-\alpha)\}$ are tight for $\alpha \in [b_1, b_2]$ and, consequently, the quantile processes that we need to deal with here satisfy Conditions 1 and 2 in Theorem 13.2 in Billingsley (1999, p.139). Next, we try to establish the tightness of the α-trimmed mean process. For the α-trimmed mean process, Condition 2 in Theorem 13.2, related to the oscillation of the stochastic process, follows from Theorem A.1 in Leger and Romano (1990, pp.311-312) considering $\hat{F}_n$ and F instead of $\hat{G}_n$ and F_n, respectively. Condition 1 in Theorem 13.2, related to the uniform boundedness of the process, holds for the α-trimmed mean process because the α-trimmed mean is the average of certain quantiles. Note that we are using the fact that quantile processes satisfy the condition of uniform boundedness for $\alpha \in [b_1, b_2]$. Consequently, the α-trimmed mean process is tight, and the process $T_n(\alpha)$ is also tight for $\alpha \in [b_1, b_2]$ and $0 < b_1 < b_2 < 1/2$. This completes the proof of the lemma.

Proof of Theorem 1. It follows from Lemmas 1 and 2 that any finite-dimensional distribution of the stochastic process $\sqrt{n}\{T_n(\alpha) - \theta'(\alpha)\}$ is multivariate normal, and the process satisfies the tightness condition. Therefore, $\sqrt{n}\{T_n(\alpha) - \theta'(\alpha)\}$ converges weakly to a Gaussian process in view of Theorem 13.1 in Billingsley (1999, p.139).

Proof of Corollary 1. Since $\sqrt{n}\{T_n(\alpha) - \theta'(\alpha)\}$ is tight in the space $D[b_1, b_2]$ equipped with the supremum norm, it follows from Condition 1 in Theorem 13.2 that, for every $0 < \eta < 1$ and for any $0 < b_1 < b_2 < 1/2$, there exists a constant $M(\eta, b_1, b_2) > 0$ such that $P[\sup_{\alpha\in[b_1,b_2]} \sqrt{n}|T_n(\alpha) - \theta'(\alpha)| \leq M(\eta, b_1, b_2)] > 1 - \eta$. This implies that $\sup_{\alpha\in[b_1,b_2]} |T_n(\alpha) - \theta'(\alpha)| = O_P(n^{-1/2})$ and completes the proof.

Proof of Proposition 1. The α-trimmed mean $\theta(\alpha) = (1-2\alpha)^{-1} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} xf(x)dx$ is a decreasing function of α if

$$\begin{aligned}
\frac{d}{d\alpha}\theta(\alpha) \leq 0 &\Leftrightarrow \frac{1}{(1-2\alpha)} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} xf(x)dx \leq \frac{F^{-1}(\alpha) + F^{-1}(1-\alpha)}{2} \\
&\Leftrightarrow \lim_{N\to\infty} \sum_{j=1}^{j=N} \frac{1}{N} \frac{F^{-1}(\alpha_j) + F^{1}(1-\alpha_j)}{2} \leq \frac{F^{-1}(\alpha) + F^{-1}(1-\alpha)}{2}, \quad (5.1)
\end{aligned}$$

where $(\alpha_1, \alpha_2, \ldots, \alpha_N)$ is an equally spaced partition of $(\alpha, 1-\alpha)$. The last implication follows from the convergence of Riemann sum to the Riemann integral. In order to prove (5.1), it is enough to show that for any $j = 1, \ldots, N$, $(F^{-1}(\alpha_j) + F^{-1}(1-\alpha_j))/2$ is smaller than $(F^{-1}(\alpha) + F^{-1}(1-\alpha))/2$. So, it is sufficient to prove that $(F^{-1}(\alpha) + F^{-1}(1-\alpha))/2$ is a decreasing function of α, i.e.,

$$\frac{d}{d\alpha} \frac{F^{-1}(\alpha) + F^{-1}(1-\alpha)}{2} \leq 0 \Leftrightarrow f(F^{-1}(\alpha)) \geq f(F^{-1}(1-\alpha)).$$

This completes the proof of the proposition.

Proof of Theorem 2. In view of the weak convergence of the process $\sqrt{n}\{T_n(\alpha) - \theta'(\alpha)\}$ for $\alpha \in [b_1, b_2]$ and the continuity of M, it is obvious that the test has asymptotic size ρ.

Recall now from Theorem 1 that for $\alpha \in [b_1, b_2]$, the process $\sqrt{n}\{T_n(\alpha) - \theta'(\alpha)\}$ converges weakly to a Gaussian process. Consequently, $||\sqrt{n}\{T_n(\alpha) - \theta'(\alpha)\}||_{[b_1,b_2],p} = O_P(1)$ for any b_1, $b_2 \in (0, 1/2)$ and $p \in [1, \infty]$. Further, in view of the continuity θ', there exist b_1, $b_2 \in (0, 1/2)$ under H_1 such that $||\theta'||_{[b_1,b_2],p} > 0$ for any $p \in [1, \infty]$. These facts imply that for every $C_1 > 0$, $\lim_{n\to\infty} P_{H_1}[||\sqrt{n}\{T_n(\alpha)\}||_{[b_1,b_2],p} > C_1] = 1$. Hence, we have $\lim_{n\to\infty} P_{H_1}[M[\sqrt{n}\{T_n(\alpha)\}] > \xi_\rho] = 1$ in view of the condition imposed on M.

Proof of Theorem 3. It follows from Theorem 1 that for $\alpha \in [b_1, b_2]$, $\sqrt{n}\{T_n(\alpha) -\theta'(\alpha)\}$ converges weakly to a Gaussian process with zero mean and covariance function $k(\alpha_1, \alpha_2)$. Here $k(\alpha_1, \alpha_2)$ is a non-negative kernel and, from its expression in Lemma 1, that it is a continuous function of α_1 and α_2. Let the $e_i(\alpha)$'s and the λ_i's be the eigen-functions and the eigen-values of the kernel $k(\alpha_1, \alpha_2)$, respectively. From the Karhunen-Loeve expansion (see, e.g., Loeve (1978)) of the weak limit of the process $\sqrt{n}\{T_n(\alpha) - \theta'(\alpha)\}$, this process converges in distribution to the process $\sum\limits_{i=1}^{\infty} Z_i e_i(\alpha)$, where the Z_i's are independent random variables such that Z_i has $N(0, \lambda_i)$ distribution. Using the continuity of the integral functional on $D[b_1, b_2]$ equipped with the supremum norm and the orthonormality of the $e_i(\alpha)$'s, under H_0, when $\theta'(\alpha) = 0$ for all $\alpha \in [b_1, b_2]$, $\tau_n = \int_{b_1}^{b_2} n\{T_n(\alpha)\}^2$ converges weakly to $\sum\limits_{i=1}^{\infty} \lambda_i W_i$, where W_i is as defined in the statement of the theorem. This completes the proof.

Lemma 3. *Let $\tau_n^* = \sum_{i=[nb_1]+1}^{[nb_2]-1} \{T_n(i/n)\}^2$. Then, $\tau_n - \tau_n^*$ converges to zero in probability.*

Proof of Lemma 3. Note that

$$\int_{b_1}^{b_2} n\{T_n(\alpha)\}^2 d\alpha = \int_{b_1}^{b_2} n\left\{\bar{x}_\alpha - \frac{\hat{F}_n^{-1}(\alpha) + \hat{F}_n^{-1}(1-\alpha)}{2}\right\}^2 d\alpha.$$

In other words, it is enough to show that

$$n\int_{b_1}^{b_2} \{\hat{F}_n^{-1}(\alpha)\}^2 d\alpha - \sum_{i=[nb_1]+1}^{[nb_2]-1} \left\{\hat{F}_n^{-1}(\frac{i}{n})\right\}^2 \xrightarrow{P} 0,$$

$$n\int_{b_1}^{b_2} \{\hat{F}_n^{-1}(1-\alpha)\}^2 d\alpha - \sum_{i=[nb_1]+1}^{[nb_2]-1} \left\{\hat{F}_n^{-1}(1-\frac{i}{n})\right\}^2 \xrightarrow{P} 0$$

and

$$\int_{b_1}^{b_2} \{\bar{x}_\alpha\}^2 d\alpha - \sum_{i=[nb_1]+1}^{[nb_2]-1} \bar{x}_{i/n}^2 \xrightarrow{P} 0.$$

Note that

$$\sum_{i=[nb_1]+1}^{[nb_2]-1} \left\{\hat{F}_n^{-1}(\frac{i}{n})\right\}^2 = \left[x^2_{([nb_1]+1)} + \ldots + x^2_{([nb_2]-1)}\right],$$

and

$$n\int_{b_1}^{b_2}\{\hat{F}_n^{-1}(\alpha)\}^2 d\alpha = n\left[\int_{b_1}^{([nb_1]+1)/n}\{\hat{F}_n^{-1}(\alpha)\}^2 d\alpha + \int_{([nb_1]+1)/n}^{([nb_1]+2)/n}\{\hat{F}_n^{-1}(\alpha)\}^2 d\alpha \right.$$
$$\left. +\cdots+\int_{([nb_2]-1)/n}^{[nb_2]/n}\{\hat{F}_n^{-1}(\alpha)\}^2 d\alpha + \int_{([nb_2]-1)/n}^{b_2}\{\hat{F}_n^{-1}(\alpha)\}^2 d\alpha\right]$$
$$= \left[x^2_{[nb_1]+1} + \ldots + x^2_{([nb_2]-1)}\right] + x^2_{([nb_2])}\{b_2 - \frac{[nb_2]}{n}\}.$$

Hence,

$$n\int_{b_1}^{b_2}\{\hat{F}_n^{-1}(\alpha)\}^2 d\alpha - \sum_{i=[nb_1]+1}^{[nb_2]-1}\left\{\hat{F}_n^{-1}(\frac{i}{n})\right\}^2 = x^2_{([nb_2])}\left(b_2 - \frac{[nb_2]}{n}\right) \xrightarrow{P} 0$$

using the facts that $(b_2 - [nb_2]/n) \to 0$ and $x_{([nb_2])} = O_P(1)$ as $n \to \infty$. The proofs of

$$n\int_{b_1}^{b_2}\{\hat{F}_n^{-1}(1-\alpha)\}^2 d\alpha - \sum_{i=[nb_1]+1}^{[nb_2]-1}\left\{\hat{F}_n^{-1}(1-\frac{i}{n})\right\}^2 \xrightarrow{P} 0$$

and

$$\int_{b_1}^{b_2}\{\bar{x}_\alpha\}^2 d\alpha - \sum_{i=[nb_1]+1}^{[nb_2]-1}\bar{x}^2_{i/n} \xrightarrow{P} 0$$

follow from algebraic computations. This completes the proof of the lemma.

Proof of Theorem 4. In order to establish contiguity of the sequence of densities associated with H_n, it is enough to show that the log likelihood ratio L_n is asymptotically normal with mean $-\sigma^2/2$ and variance σ^2 (see Hajek and Sidak (1967, p.204)). Suppose that $f_n(x_i)$ is the density function of $F_n(x_i)$. The likelihood ratio for testing H_0 against H_n is

$$L_n = \sum_{i=1}^n \log\frac{f_n(x_i)}{h(x_i)} = \sum_{i=1}^n \log\frac{(1-\delta/\sqrt{n})h(x_i) + \delta/\sqrt{n}g(x_i)}{h(x_i)}$$
$$= \sum_{i=1}^n \log\left[1 + \frac{\delta}{\sqrt{n}}\left\{\frac{g(x_i)}{h(x_i)} - 1\right\}\right]$$
$$= \frac{\delta}{\sqrt{n}}\sum_{i=1}^n\left\{\frac{g(x_i)}{h(x_i)} - 1\right\} - \frac{\delta^2}{2n}\sum_{i=1}^n\left\{\frac{g(x_i)}{h(x_i)} - 1\right\}^2 + R_n$$
$$= \delta\sqrt{n}\frac{1}{n}\sum_{i=1}^n k_i - \frac{\delta^2}{2}\times\frac{1}{n}\sum_{i=1}^n k_i^2 + R_n \text{ (here } k_i = \tfrac{g(x_i)}{h(x_i)} - 1). \qquad (5.2)$$

Since $E_h\{g(x)/h(x)-1\}^2 < \infty$, we have $R_n \xrightarrow{p} 0$ as $n \to \infty$. Further, using straightforward applications of the C.L.T. and W.L.L.N., it follows that the first term in (5.2) is asymptotically normal with mean zero and variance $\delta^2 var(k_i) = \delta^2\sigma^2$, and the second term in (5.2) converges in probability to $-(\delta^2/2)\sigma^2$. So, by Slutsky's Theorem, L_n is asymptotically normally distributed with mean $-(\delta^2/2)\sigma^2$ and variance $\delta^2\sigma^2$. This ensures the contiguity of the sequence of densities associated with H_n using Hajek and Sidak (1967, p.204).

We consider $t_1, \ldots, t_k \in R$. Under H_0, one can establish that the joint distribution of $\sqrt{n}((1/n)\sum_{i=1}^n \psi(x_i,t_1) - E_h\{\psi(x_i,t_1)\}, \ldots, (1/n)\sum_{i=1}^n \psi(x_i,t_k) - E_h\{\psi(x_i,t_k)\}, L_n)$ is asymptotically multivariate Gaussian using the C.L.T. and the expansion of log likelihood ratio L_n (given in (5.2) above). Note that for any $p = 1, \ldots, k$, the asymptotic covariance between $\sqrt{n}((1/n)\sum_{i=1}^n \psi(x_i,t_p) - E_h\{\psi(x_i,t_p)\})$ and $\sqrt{n}L_n$ is

$$\begin{aligned} m(t_p,\delta) &= E_h\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(x_i,t_p) \times \frac{\delta}{\sqrt{n}}\sum_{i=1}^n \left\{\frac{g(x_i)}{h(x_i)} - 1\right\}\right] \\ &= \frac{\delta}{n}E_h\left[\sum_{i=1}^n \psi(x_i,t_p) \times \left\{\frac{g(x_i)}{h(x_i)} - 1\right\}\right] \ \left(\text{since } E_h\left\{\tfrac{g(x)}{h(x)} - 1\right\} = 0\right) \\ &= \delta[E_g\{\psi(x,t_p)\} - E_h\{\psi(x,t_p)\}]. \end{aligned}$$

Now, by a straightforward application of Hajek and Sidak (1967, p.208) one can establish that, under contiguous alternatives, $\sqrt{n}((1/n)\sum_{i=1}^n \psi(x_i,t_1) - E_h\{\psi(x_i,t_1)\}, \ldots, (1/n)\sum_{i=1}^n \psi(x_i,t_k) - E_h\{\psi(x_i,t_k)\})$ is asymptotically k-dimensional multivariate normal with mean having pth component $m(t_p,\delta)$ and $k \times k$-dimensional covariance matrix $k_1(t_1,t_2)$, which is as defined in the statement of the theorem. Further, the process $\sqrt{n}((1/n)\sum_{i=1}^n \psi(x_i,t) - E_h\{\psi(x_i,t)\})$ satisfies tightness condition under contiguous alternatives in view of the fact that it is tight under H_0. The tightness under H_0 follows from the fact that the process $\sqrt{n}((1/n)\sum_{i=1}^n \psi(x_i,t) - E_h\{\psi(x_i,t)\})$ converges weakly to a Gaussian process under H_0, which is assumed in the theorem. So, under H_n, $\sqrt{n}((1/n)\sum_{i=1}^n \psi(x_i,t) - E_h\{\psi(x_i,t)\})$ converges to a Gaussian process with mean function $m(t,\delta)$ and covariance kernel $k_1(t_1,t_2)$. This completes the proof.

Proof of Theorem 5. We first consider the case, when H is supported on a compact interval and H and G have disjoint supports. Recall that

$$\begin{aligned} \theta''(\alpha) = \frac{2}{1-2\alpha}\Bigg[&\frac{1}{1-2\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} xf(x)dx - \frac{1}{2}\{F^{-1}(\alpha) + F^{-1}(1-\alpha)\} \\ &- \frac{1}{2}\left\{\frac{1}{f(F^{-1}(\alpha))} - \frac{1}{f(F^{-1}(1-\alpha))}\right\}\Bigg]. \end{aligned}$$

It follows from the proof of Proposition 1 that, under the skewness condition $f(F^{-1}(\alpha)) > f(F^{-1}(1-\alpha))$,

$$\left[\frac{1}{1-2\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} xf(x)dx - \frac{1}{2}\{F^{-1}(\alpha)+F^{-1}(1-\alpha)\}\right]$$

is bounded above for all $\alpha \in [b_1, b_2]$. So, in order to prove Theorem 5 it is enough to investigate the behavior of

$$N(\alpha) := -\frac{1}{2}\left[\frac{1}{f(F^{-1}(\alpha))} - \frac{1}{f(F^{-1}(1-\alpha))}\right],$$

where $\alpha \in [b_1, b_2]$ and $0 < b_1 < b_2 < 1/2$.

If $y_1 = F^{-1}(\alpha)$ and $\alpha < (1-\beta)$, then y_1 is located inside the support of H as G is stochastically larger than H and they have disjoint supports. In other words,

$$(1-\beta)H(y_1) = \alpha \Leftrightarrow y_1 = H^{-1}\left(\frac{\alpha}{(1-\beta)}\right).$$

If $\alpha > (1-\beta)$, y_1 is located inside the support of G and, in that case,

$$\begin{aligned} F^{-1}(\alpha) = y_1 &\Leftrightarrow (1-\beta)H(y_1) + \beta G(y_1) = \alpha \\ &\Leftrightarrow (1-\beta) + \beta G(y_1) = \alpha \text{ (since } H(y_1) = 1) \\ &\Leftrightarrow y_1 = G^{-1}\left(\frac{\alpha-(1-\beta)}{\beta}\right). \end{aligned}$$

If $\alpha = (1-\beta)$, then $F^{-1}(\alpha)$ can be defined as any point that lies between $H^{-1}(1)$ and $G^{-1}(0)$.

Next, when H and G have disjoint supports, we show that $\sup_{\alpha\in[\beta-\gamma,\beta+\gamma],\alpha\neq\beta} N(\alpha) = \infty$ and $\sup_{\alpha\notin[\beta-\gamma,\beta+\gamma]} N(\alpha) < \infty$ for any $\gamma > 0$. We have

$$\begin{aligned} \lim_{\alpha\to\beta+} N(\alpha) = \lim_{\alpha\to\beta+} -\frac{1}{2}\Bigg[&\frac{1}{(1-\beta)h(F^{-1}(\alpha)) + \beta g(F^{-1}(\alpha))} \\ &-\frac{1}{(1-\beta)h(F^{-1}(1-\alpha)) + \beta g(F^{-1}(1-\alpha))}\Bigg] = \infty \end{aligned}$$

(since $F^{-1}(\alpha) = H^{-1}(\alpha/(1-\beta))$ if $\alpha < (1-\beta)$, $h(H^{-1}(x)) \to 0$ and $g(H^{-1}(x)) \to 0$ as $x \to 1-$). In the same way,

$$\begin{aligned} \lim_{\alpha\to\beta-} N(\alpha) = \lim_{\alpha\to\beta-} -\frac{1}{2}\Bigg[&\frac{1}{(1-\beta)h(F^{-1}(\alpha)) + \beta g(F^{-1}(\alpha))} \\ &-\frac{1}{(1-\beta)h(F^{-1}(1-\alpha)) + \beta g(F^{-1}(1-\alpha))}\Bigg] = \infty \end{aligned}$$

(since $F^{-1}(\alpha) = G^{-1}((\alpha - (1-\beta))/\beta)$ if $\alpha > (1-\beta)$, $h(G^{-1}(x)) \to 0$ and $g(G^{-1}(x)) = 0$ as $x \to 0+$). Hence, $\lim_{\alpha\to\beta} N(\alpha) = \infty$. This implies that $\sup_{\beta-\gamma\leq\alpha\leq\beta+\gamma,\alpha\neq\beta} N(\alpha) = \infty$.

Next, we try to investigate $N(\alpha)$ when $\alpha < \beta-\gamma$ or $\alpha > \beta+\gamma$ for any $\gamma > 0$. If $\alpha > \beta+\gamma$, we have

$$N(\alpha) = -\frac{1}{2}\Bigg[\frac{1}{(1-\beta)h(H^{-1}(\alpha/(1-\beta))) + \beta g(H^{-1}(\alpha/(1-\beta)))} - \frac{1}{(1-\beta)h(H^{-1}((1-\alpha)/(1-\beta))) + \beta g(H^{-1}((1-\alpha)/(1-\beta)))}\Bigg].$$

The last expression is bounded as h and g are continuous and positive on any compact subinterval strictly inside the supports of h and g. For $\alpha < \beta-\gamma$,

$$N(\alpha) = -\frac{1}{2}\Bigg[\frac{1}{(1-\beta)h(H^{-1}(\alpha/(1-\beta))) + \beta g(H^{-1}(\alpha/(1-\beta)))} - \frac{1}{(1-\beta)h(G^{-1}((\beta-\alpha)/\beta)) + \beta g(G^{-1}((\beta-\alpha)/\beta))}\Bigg].$$

Again, the last expression is bounded as h and g are continuous and positive on any compact subinterval strictly inside the supports of h and g. This completes the proof in the case when H is supported on compact interval, and H and G have disjoint supports.

Next we consider the case in which H is supported on the entire real line and G varies in such a way that $G >_{st} H$ and $\inf_{q\in(0,1)}\{G^{-1}(q) - H^{-1}(q)\} \to \infty$. As in the case of disjoint supports for H and G, it is enough to investigate the term $N(\alpha) = -[1/f(F^{-1}(\alpha)) - 1/f(F^{-1}(1-\alpha))]/2$ that appears in the expression of $\theta''(\alpha)$.

Suppose that $\sup_{\alpha>\beta+\delta}\theta''(\alpha)$ is not bounded above as G varies satisfying the conditions stated above. There must exist a sequence of distributions $G_n >_{st} H$ and a sequence of positive real numbers $1/2 > \alpha_n > \beta+\delta$ such that $\inf_{q\in(0,1)}\{G_n^{-1}(q) - H^{-1}(q)\} \to \infty$ and $\theta_n''(\alpha_n) \to \infty$ as $n \to \infty$. Let y_n be the α_nth quantile of $F_n = (1-\beta)H + \beta G_n$, i.e., $y_n = F_n^{-1}(\alpha_n)$. We have

$$\begin{aligned} F_n^{-1}(\alpha_n) = y_n &\Leftrightarrow (1-\beta)H(y_n) + \beta G_n(y_n) = \alpha_n \\ &\Leftrightarrow (1-\beta)H(y_n) + \beta H(y_n - z_{y_n}) = \alpha_n \ (z_{y_n} > 0 \text{ as } G_n >_{st} H). \quad (5.3)\end{aligned}$$

Note that $z_{y_n} \to \infty$ as $n \to \infty$ for any y_n since $\inf_{q\in(0,1)}\{G_n^{-1}(q) - H^{-1}(q)\} \to \infty$ as $n \to \infty$. If $\theta_n''(\alpha_n) \to \infty$ (and hence, $N_n(\alpha) = -[1/f_n(F_n^{-1}(\alpha)) - 1/f_n(F_n^{-1}(1-\alpha))]/2 \to \infty$) as $n \to \infty$, then we must have either $y_n = F_n^{-1}(\alpha_n) \to \pm\infty$ and $y_n - z_{y_n} \to \pm\infty$, or $u_n := F_n^{-1}(1-\alpha_n) \to \pm\infty$ and $u_n - z_{u_n} \to \pm\infty$ as $n \to \infty$. This follows from the expression

$$N_n(\alpha_n) = -\frac{1}{2}\left[\frac{1}{(1-\beta)h(y_n) + \beta h(y_n - z_{y_n})} - \frac{1}{(1-\beta)h(u_n) + \beta h(u_n - z_{u_n})}\right],$$

and the fact that h is positive and continuous on any compact subinterval within its support. If $y_n \to -\infty$ and $z_{y_n} \to \infty$, we have $y_n - z_{y_n} \to -\infty$ because $z_{y_n} \to \infty$. Now, using $y_n \to -\infty$ and $y_n - z_{y_n} \to -\infty$ in (5.3), we have $\alpha_n \to 0$ as $n \to \infty$, which contradicts the fact that $\alpha_n > \beta + \delta$ for all n. If $y_n \to \infty$ and $z_{y_n} \to \infty$, then $y_n - z_{y_n}$ may be bounded or unbounded. In that case, we again need to consider several distinct possibilities, as either $y_n - z_{y_n}$ remains bounded as $n \to \infty$, or *we can extract a subsequence along which* $y_n - z_{y_n} \to \pm\infty$ as $n \to \infty$. Using $y_n \to \infty$ and $y_n - z_{y_n} \to \infty$ in (5.3), we have $\alpha_n \to 1$, which contradicts $\alpha_n < 1/2$ for all n. Lastly, if $y_n \to \infty$ while $y_n - z_{y_n}$ either remains bounded or tend to $-\infty$ as $n \to \infty$, then it follows from (5.3) that for all n sufficiently large, $\alpha_n > 1/2$, in view of the fact that $(1-\beta) > 1/2$, and this is again a contradiction. Combining all these, it follows that the sequence y_n must remain bounded as $n \to \infty$. In a similar way, one can show that u_n remains bounded as $n \to \infty$ using the equation

$$(1-\beta)H(u_n) + \beta H(u_n - z_{u_n}) = (1-\alpha_n), \tag{5.4}$$

where z_{u_n} satisfies $G_n(u_n) = H(u_n - z_{u_n})$, and hence $z_{u_n} \to \infty$ as $n \to \infty$ in view of the condition $\inf_{q\in(0,1)}\{G_n^{-1}(q) - H^{-1}(q)\} \to \infty$ as $n \to \infty$. This completes the proof of boundedness of $\sup_{\alpha>\beta+\gamma} N_n(\alpha)$ as $n \to \infty$.

Next, assume that $\sup_{\alpha<\beta-\gamma} \theta_n''(\alpha)$ is not bounded above. As before, there must exist a sequence of distributions $G_n >_{st} H$ and a sequence of positive real numbers $1/2 > \alpha_n > \beta + \gamma$ satisfying the conditions as stated in the preceding paragraph. Once again, if $\theta''(\alpha_n) \to \infty$ as $n \to \infty$, it follows from the preceding expression of $N_n(\alpha_n)$ that, as $n \to \infty$, either $y_n \to \pm\infty$ and $y_n - z_{y_n} \to \pm\infty$, or we have $u_n \to \pm\infty$ and $u_n - z_{u_n} \to \pm\infty$. Now, arguing as before, and using (4), one can show that $u_n - z_{u_n}$ remains bounded as $n \to \infty$. Since h is positive and continuous on any compact subinterval of its support, we can conclude that $\sup_{\alpha<\beta-\gamma} N_n(\alpha)$ remains bounded as $n \to \infty$.

We look to the case $\alpha = \beta$. Here also, let G_n be a sequence of distributions satisfying $G_n >_{st} H$ and $\inf_{q\in(0,1)}\{G_n^{-1}(q) - H^{-1}(q)\} \to \infty$ as $n \to \infty$. Let $v_n := F_n^{-1}(1-\beta)$ and $w_n := F_n^{-1}(\beta)$. Hence, we have

$$(1-\beta)H(v_n) + \beta H(v_n - z_{v_n}) = (1-\beta), \tag{5.5}$$

where z_{v_n} satisfies $G_n(v_n) = H(v_n - z_{v_n})$, and consequently, $z_{v_n} \to \infty$ as $n \to \infty$ in view of the condition $\inf_{q\in(0,1)}\{G_n^{-1}(q) - H^{-1}(q)\} \to \infty$ as $n \to \infty$. Now, we try to show that (5.5) is satisfied only when $v_n \to \infty$ and $v_n - z_{v_n} \to -\infty$ as $n \to \infty$. If $v_n \to -\infty$ and $v_n - z_{v_n} \to -\infty$ as $n \to \infty$ in (5.5), then we have $\beta = 1$, which contradicts $\beta < 1/2$. On the other hand, if v_n is bounded and $v_n - z_{v_n} \to -\infty$ as $n \to \infty$, then (5.5) along with the fact that H is supported on the entire real line implies that the left hand side of (5.5) is strictly smaller than $(1-\beta)$ for all n sufficiently large, which is also a contradiction. If both v_n and $v_n - z_{v_n}$ tend to ∞ as $n \to \infty$, it follows from (5.5) that $\beta = 0$, which is not possible. Further, if we consider the possibility that $v_n \to \infty$ and $v_n - z_{v_n}$ remains bounded as $n \to \infty$ in (5.5), then $\{(1-\beta) - (1-\beta)H(v_n)\}$ tends to zero as $n \to \infty$ while $\beta H(v_n - z_{v_n})$ remains bounded away from zero as $n \to \infty$, which leads to a contradiction in view of (5.5). Combining these facts, (5.5) is satisfied only if $v_n \to \infty$ and $v_n - z_{v_n} \to -\infty$ as $n \to \infty$. Consequently, we must have

$$N_n(\beta) = \left[\frac{1}{(1-\beta)h(w_n) + \beta h(w_n - z_{w_n})} - \frac{1}{(1-\beta)h(v_n) + \beta h(v_n - z_{v_n})}\right] \to \infty$$

as $n \to \infty$, since $h(x) \to 0$ as $x \to \pm\infty$. This completes the proof of the theorem.

Acknowledgement

The research of the first author is partially supported by a grant from Council of Scientific and Industrial Research (CSIR), Government of India. The authors are thankful to two anonymous referees for several useful comments.

References

Ahmad, I. A. and Li, Q. (1997). Testing symmetry of an unknown density function by kernel method. *J. Nonparametric Statist.* **7**, 279-293.

Arcones, M. A. and Gine, E. (1991). Some bootstrap tests of symmetry for univariate continuous distributions. *Ann. Statist.* **19**, 1496-1511.

Bickel, P. J. (1965). On some robust estimates of location. *Ann. Statist.* **36**, 847-858.

Bickel, P. J. (1967). Some contributions to the theory of order statistics. *Proceedings of the Fifth Berkeley Symposium Mathematical Statistics and Probability* **1**, 575-591.

Billingsley, P. (1999). *Convergence of Probability Measures.* Wiley, New York.

Csorgo, S. and Heathcote, C. R. (1987). Testing for symmetry. *Biometrika* **74**, 177-184.

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*, Springer Verlag, New York.

Dhar, S. S. and Chaudhuri, P. (2009). A comparison of robust estimators based on two types of trimming. *Adv. Statist. Anal.* **93**, 151-158.

Everitt, B. S. and Hand, D. J. (1981). *Finite Mixture Distributions.* Chapman and Hall, London.

Ghosh, A. K., Chaudhuri, P. and Sengupta, D. (2006). Classification using kernel density estimates: multiscale analysis and visualization. *Technometrics* **48**, 120-132.

Hajek, J. and Sidak, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

Hall, P. (1981). Large sample properties of Jackel's adaptive trimmed mean. *Ann. Inst. Statist. Math.* **33**, 449-462.

Hogg, R. V. (1967). A partial review and some suggestions for future applications and theory. *J. Amer. Statist. Assoc.* **69**, 909-923.

Horton, N. J., Brown, E. R. and Qian, L. (2004). Use of *R* as a toolbox for mathematical statistics exploration. *Amer. Statist.* **58**, 343-357.

Huber, P. J. (1981). *Robust Statistics*. Wiley, New York.

Jaeckel, L. A. (1971). Some flexible estimates of location. *Ann. Math. Statist.* **42**, 1540-1552.

Leger, C. and Romano, J. P. (1990). Bootstrap adaptive estimation: The trimmed-mean example. *Canad. J. Statist.* **18**, 297-314.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer International Edition.

Loeve, M. (1978). *Probability Theory II*. Springer-Verlag.

Mira, A. (1999). Distribution-free test for symmetry based on Bonferroni measure. *J. Appl. Statist.* **26**, 959-972.

Schuster, E. F. and Barker, R. C. (1987). Using the bootstrap in testing symmetry versus asymmetry. *Comm. Statist. Simulation Comput.* **16**, 69-84.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Stigler, S. M. (1973). The asymptotic distribution of the trimmed mean. *Ann. Statist.* **1**, 472-477.

Tukey, J. W. (1948). Some elementary problems of importance to small sample practice. *Human Biology* **20**, 205-214.

Tukey, J. W. and McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample:trimming/Winsorization 1. *Sankhyā* **25**, 331-352.

van Zwet, W. R. (1979). Mean, Median, Mode, II. *Statist. Neerl.* **33**, 1-5.

Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK.

E-mail: dsubhra@gmail.com

203 B. T. Road, Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Calcutta, India.

E-mail: probal@isical.ac.in

(Received July 2010; accepted May 2011)