12-16-2014

# A More Efficient Nonparametric Test of Symmetry Based on Overlapping Coefficient

Hani M. Samawi
*Georgia Southern University*, hsamawi@georgiasouthern.edu

Robert L. Vogel
*Georgia Southern University*, rvogel@georgiasouthern.edu

## Recommended Citation

# A More Efficient Nonparametric Test of Symmetry Based on Overlapping Coefficient

### Abstract

In this paper we provide a more efficient nonparametric test of symmetry based on the empirical overlap coefficient using kernel density estimation applied to an extreme order statistics, namely extreme ranked set sampling. Our simulation investigation reveals that our proposed test of symmetry is at least as powerful as currently available tests of symmetry. Intensive simulation is conducted to examine the power of the proposed test. An illustration is provided using cardiac output and body weight of neonates in a neonatal intensive care unit.

### Keywords

Test of symmetry; Power of the test; Bootstrap method; Overlap coefficients; Weitzman's measure; Extreme ranked set sample; Kernel Density estimation; AMS: 62G10

**Hani M Samawi\* and Robert Vogel**

*Jiann-Ping Hsu College of Public Health, Georgia Southern University, USA*

**\*Corresponding author:** Hani M Samawi, Jiann-Ping Hsu College of Public Health, Georgia Southern University, USA, Tel: 912 478 1345; Email: hsamawi@georgiasouthern.edu

## Abbreviations

ERSS: Extreme Ranked Set Sample; SRS: Simple Random Sample; RSS: Ranked Set Sampling

## Introduction

Parametric and some nonparametric statistical inferences and modeling are valid only under certain assumptions. One of most common assumptions in the literature is that of symmetry of the underlying distribution. If the underlying distribution is not symmetric the question becomes how to define the appropriate location and scale measures. Thus to choose the appropriate statistical analysis, we need to check for underlying assumptions, including symmetry. Most tests of symmetry available in the literature typically have low statistical power and fail to detect a small but meaningful asymmetry in the population. Examples of those tests have been suggested by Butler [1], Rothman and Woodroofe [2], Hill and Rao [3], McWilliams [4] and Ozturk [5]. McWilliams's [4] runs test of symmetry is more powerful than those provided by Butler [1], Rothman and Woodroofe [2], Hill and Rao [3] against various asymmetric alternatives. Tajuddin [6], proposed a test for symmetry based on the Wilcoxon two-sample test and found his test to be more powerful than the runs test.

Baklizi [7] suggested a runs test of symmetry based on the conditional distribution and demonstrated that it performed slightly better than the unconditional test by McWilliams [4]. Baklizi's test is also very robust for misspecification of the median. Modarres and Gastwirth [8-9] provided a modification to McWilliams [4] runs test based on Wilcoxon scores to weigh the runs. Their procedure improved the power for testing symmetry when the center of the distribution is known. However, their test did not perform well when asymmetry is focused on regions close to the median. Samawi [10] investigated the use of extreme ranked set sample (ERSS). Samawi et al. [11] used (ERSS) to provide a more powerful runs test of symmetry. Finally, Samawi et al. [12] used the overlap coefficient to test for symmetry and
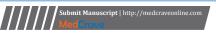
showed that their test procedure is competitive with the other available tests of symmetry. This paper uses ERSS to provide a more powerful overlap coefficient test of symmetry.

The overlap measure $\Delta$ (OVL) is defined as the area of intersection of the graphs of two probability density functions. It measures the similarity, which is the agreement or the closeness of the two probability distributions. The OVL measure $\Delta$ was originally introduced by Weitzman [13]. Recently, several authors including Bradley and Piantadosi [14], Inman and Bradley [15], Clemons [16], Reiser and Faraggi [17], Clemons and Bradley [18], Mulekar and Mishra [19], Al-Saidy, et al. [20], Schmid and Schmidt [21], Al-Saleh and Samawi [22], and Samawi and Al-Saleh [23] considered this measure. The sampling behavior of a nonparametric estimator of $\Delta$ using naive kernel density estimation was examined by Clemons and Bradley [18], using Monte Carlo and bootstrap techniques.

Let $f_1(x)$ and $f_2(x)$ be two probability density functions. Assume samples of observations are drawn from continuous distributions. The overlap measure used in the literature is defined by Weitzman's Measure (1970) as $\Delta = \int \min\{f_1(x), f_2(x)\}\, dx$.

The overlap measure of two densities assumes values between 0 and 1. An overlap value close to 0 indicates extreme inequality of the two density functions, and an overlap value of 1 indicates exact equality. In most statistical applications the data used is assumed to consist of a simple random sample (SRS). Cost savings of quantifying sampling units can be achieved by using ranked set sampling (RSS) methods as described by McIntyre [24] to estimate the population mean. The procedure introduced by McIntyre was later called RSS. As a variation of RSS, an extreme ranked set sample (ERSS) is introduced and investigated by Samawi et al [10].

First we describe the RSS procedure as follows: First, randomly sample a group of sampling units, of size $r^2$ from the target population. Randomly partition the group into disjoint subsets each having a pre-assigned size *r*. In most practical

situations, the size *r* will be two, three or four. Rank the elements in each subset by a suitable method of ranking such as prior information, visual inspection or by the subject-matter experimenter himself. The ith order statistic from the i-th subset, $X_{i(i)}$, *i = 1, ...,* *r*, will be quantified (actual measurement). Finally, $X_{1(1)}, X_{2(2)}, ..., X_{r(r)}$ constitutes the RSS. This will represent one complete cycle. The procedure can be repeated *m*-times as needed, to get a RSS of size *n=Mr*. A detailed explanation of uni variate RSS and its variations may be found in Kaur et al. [25], Patil et al. [26], Kaur et al. [27] and Sinha [28]. However, an extreme ranked set sample (ERSS) of size *n=2m,* is described by Samawi et al. [10], is similar to RSS procedure except that we quantify only the minima or the maxima from each subject to get an ERSS as $X_{1(1)}, X_{2(1)}, ..., X_{m(1)}; X_{1(r)}, X_{2(r)}, ..., X_{m(r)}$.

Now consider testing the null hypothesis of symmetry for an underlying absolutely continuous distribution $F(.)$ with density denoted by $f(.)$ : $H_0 : f(x) = f(-x)$ versus $H_a : f(x) \neq f(-x)$; for some *x*. It is clear that under the null hypothesis of symmetry, if we let $f_1(x) = f(x)$ and $f_2(x) = f(-x)$ then the overlap measure is equal to one $(\Delta = 1)$, which will be our focus in this paper.

Samawi et al. [12] used the overlap measure $\Delta$ to develop a new test of symmetry based on kernel density estimation of $\Delta$. The availability of kernel density estimation in some of statistical software also, motivated us to use $\Delta$ when the sample in hand is ERSS. This paper will introduce a powerful test of symmetry based on ERSS overlap measure. The overlap test of symmetry using ERSS and its asymptotic properties are introduced is Section 2. A simulation study is given in Section 3. Illustrations using cardiac output data from neonates along with final comments are given in Section 4.

## Test of Symmetry Based on ERSS Overlap Measure

Let $X_{1(1)}, X_{2(1)}, ..., X_{m(1)}; X_{1(r)}, X_{2(r)}, ..., X_{m(r)}$ be an ERSS random sample from an absolutely continuous and differentiable distribution $F(.)$ having known median. Without loss of generality, we will assume the median to be zero. When the median or the center of the distribution is unknown, the data can be centered by a consistent estimate of the median. The implications on the asymptotic properties resulting from centering the data on a consistent estimator of the median are not intuitively clear. Therefore, further investigations are needed to study the robustness of the proposed test of symmetry and compare it with other available tests of symmetry in case of an unknown median. In this paper we will discuss only the case when the median of the underlying distribution is assumed known.

Consider the test for symmetry $H_0 : f(x) = f(-x)$ versus $H_a : f(x) \neq f(-x)$; for some *x*. Under the assumption of symmetry, $F(x) = 1 - F(-x)$. Let $f_{(1)}(x)$ be the density function of the first order statistics $X_{(1)}$ and $f_{(r)}(x)$ be the density function of the rth order statistics $X_{(r)}$9 from random samples of size *r* respectively. Under the assumption of symmetry, it can be shown that $f_{(1)}(-x) = f_{(r)}(x)$. If we let $f_1(x) = f_{(1)}(-x)$ and $f_2(x) = f_{(r)}(x)$, the null hypothesis of symmetry is equivalent to $H_o : f_{(1)}(-x) = f_{(r)}(x)$ and under the null hypothesis $\Delta = 1$. Therefore, an equivalent hypothesis for testing the symmetry is $H_0 : \Delta = 1$ versus $H_a : \Delta < 1$. we propose using $\hat{\Delta}_{ERSS}$ as it is a consistent nonparametric estimator of $\Delta$ using ERSS. Under the null hypothesis of symmetry and some mild regularity assumptions, which will be discussed later in this paper, we will derive the asymptotic distribution $\hat{\Delta}_{ERSS}$ to use it as a test of symmetry, say:

$$z_0 \sim N(0,1) , \tag{1}$$

for large *n=2m*. An asymptotic significant test procedure at level $\alpha$ is to reject $H_0$ if $z_0 < -z_\alpha$, where $z_\alpha$ is the upper $\alpha$ percentile of the standard normal distribution.

## Kernel estimation of Δusing ERSS

Based on the results of Schmid and Schmidt [21] and Anderson et al. [29] we will study the asymptotic properties using ERSS. Using one of the several available nonparametric density estimation procedures, see for example Wegman [30-31], Van Kerm [32] and Chen and Kelton [33], one can use the overlap coefficient estimators for inferential purposes.

Let $X_{1(1)}, X_{2(1)}, ..., X_{m(1)}; X_{1(r)}, X_{2(r)}, ..., X_{m(r)}$ be ERSS random sample from a differentiable distribution $F(.)$ having known median. Without a loss of generality assume the median to be zero. Let $X_{1(1)}, X_{2(1)}, ..., X_{m(1)}$ denote a random sample of minimums in ERSS and $X_{1(r)}, X_{2(r)}, ..., X_{m(r)}$ denote a random sample of maximums in ERSS, where $n = 2m$. We will use a kernel function *K* that satisfies the condition

$$\int_{-\infty}^{\infty} K(x)dx = 1 . \tag{2}$$

The kernel *K* is normally considered as a symmetric density function with mean 0 and finite variance; an example is the standard normal density. The kernel estimators of $f_{(1)}(-w_i)$ and $f_{(r)}(w_i)$, $i = 1, 2, ..., C$, respectively are

$$\hat{f}_{(1)}(-w_i) = \frac{1}{mh_-} \sum_{j=1}^{m} K\left(\frac{-w_i - X_{j(1)}}{h_-}\right) \tag{3}$$

and

$$\hat{f}_{(r)}(w_i) = \frac{1}{mh_+} \sum_{j=1}^{m} K\left(\frac{w_i - X_{j(r)}}{h_+}\right), \tag{4}$$

where $C$ is the number of bins that depends on the sample size. In practice, we suggest to take $C = $ integer of $\sqrt{m}$ (or the default setting of the software). Also, $h_-$ and $h_+$ are the bandwidths of the kernel estimators satisfying the conditions that $h_-, h_+ > 0, h_-, h_+ \to 0$ and $(mh_- \to \infty, mh_+ \to \infty)$ as $m \to \infty$. There are many choices of the bandwidths $(h_-, h_+)$, however, in our procedure we use [34] Silverman's [33] suggestion as follows: Using the normal distribution as the parametric family, the bandwidths of the kernel estimators are

$$h_- = 0.9 A_-(m)^{-1/5} \text{ and } h_+ = 0.9 A_+(m)^{-1/5}, \tag{5}$$

where $A_- = \min\{$standard deviation of $(X_{1(1)}, X_{2(1)}, ..., X_{m(1)})$, interquantile range of $(X_{1(1)}, X_{2(1)}, ..., X_{m(1)}) / 1.349\}$ and $A_+ = \min\{$standard deviation of $(X_{1(r)}, X_{2(r)}, ..., X_{m(r)})$, interquantile range of $((X_{1(r)}, X_{2(r)}, ..., X_{m(r)}) / 1.349\}$. These were found to be adequate choices of the bandwidth for minimizing the integrated mean squared error (*IMSE*),

$$IMSE = \int E[\hat{f}_{(i)}(x) - f_{(i)}(x)]^2 dx, \text{ where } i = 1, r. \tag{6}$$

The bins used are as follows: Let $R_1 = Max(R_{(1)}, R_{(r)})$, where $R_{(1)} = range(X_{1(1)}, X_{2(1)}, ..., X_{m(1)})$ and $R_{(r)} = range(X_{1(r)}, X_{2(r)}, ..., X_{m(r)})$. The bins will be selected as $w_i = w_{i-1} + \delta_x$, where $i = 2, ..., C$, $w_1$ is an initial value chosen based on the minimum value used in $R's$ calculation and $\delta_x = \dfrac{R_1}{C}$.

Using the aforementioned kernel estimator the nonparametric kernel estimator of $\Delta$ is given by

$$\hat{\Delta}_{ERSS} = \int_{\mathbb{R}} \min(\hat{f}_{(1)}(-w_i), \hat{f}_{(r)}(w_i)) dw, \tag{7}$$

which can be approximated by a trapezoidal rule, resulting in

$$\hat{\Delta}_{ERSS} \approx \sum_{i=1}^{C} \frac{\delta_x}{2} \left[ \min(\hat{f}_{(1)}(-w_i), \hat{f}_{(r)}(w_i)) + \min(\hat{f}_{(1)}(-w_{i-1}), \hat{f}_{(r)}(w_{i-1})) \right].$$

## Asymptotic properties of $\hat{\Delta}_{ERSS}$

The nonparametric kernel estimator of $\Delta$ $(\hat{\Delta}_{ERSS})$ is based on the univariate kernel for density estimation, $K : \mathbb{R} \to \mathbb{R}$. Some of the necessary regularity conditions imposed on the univariate kernel for density estimation sees for example Silverman [34], Wand and Jones [35] and Schmid and Schmidt [21] are stated below:

$$\int_{\mathbb{R}} K(z) dz = 1.9$$

$$\int_{\mathbb{R}} z^\beta K(z) dz = 0 \text{ for any } \beta = 1, ..., d-1, \text{ and } \int_{\mathbb{R}} |z|^d K(z) dz < \infty, 1 < d \in \mathbb{N}.$$

$$R = \int_{\mathbb{R}} K^2(z) dz < \infty.$$

$$h_-, h_+ > 0, h_-, h_+ \to 0, (mh_- \to \infty, mh_+ \to \infty) \text{ and } (\frac{mh_-}{\log m} \to \infty, \frac{mh_+}{\log m} \to \infty).$$

To show consistency of $\hat{\Delta}_{ERSS}$, we use some of the kernel density asymptotic properties from Silverman [33] and Wand and Jones [34] under simple random sample (SRS). Under the assumptions 1-4 and assuming that the density $f : \mathbb{R} \to \mathbb{R}$ is continuous at each $w_i$, i=1, 2, ..., $C$, and $F(x)$ is absolutely continuous. Note that $f_{(1)}(x) = r[1 - F(x)]^{r-1} f(x)$ and $f_{(r)}(x) = r[F(x)]^{r-1} f(x)$ are also absolutely continuous, therefore, the following apply:

$$Bias(\hat{f}_{(1)}(-w_i)) = o(1)_- \text{ and } Bias(\hat{f}_{(r)}(w_i)) = o(1)_+ \tag{8}$$

$$Var(\hat{f}_{(1)}(-w_i)) = \frac{f_{(1)}(-w_i)}{mh_-} \int_{\mathbb{R}} K^2(z) dz + o(\frac{1}{mh_-}) \text{ and}$$

$$Var(\hat{f}_{(r)}(w_i)) = \frac{f_{(r)}(w_i)}{mh_+} \int_{\mathbb{R}} K^2(z) dz + o(\frac{1}{mh_+}). \tag{9}$$

Since both variances converge to zero and for $h_-, h_+ > 0, h_-, h_+ \to 0$ and $(mh_- \to \infty, mh_+ \to \infty)$ as $m \to \infty$, then

$$\hat{f}_{(1)}(-w_i) \to^P f_{(1)}(-w_i) \text{ and } \hat{f}_{(r)}(w_i) \to^P f_{(r)}(w_i), \text{ for all continuous points, } w_i \text{ of } f_{(1)} \text{ and } f_{(r)}. \text{ Also, if } f_{(i)}(.), i = 1, r \text{ are}$$

uniformly continuous, then the kernel densities estimate are strongly consistent.

As in Samawi et al. [12] we can redefine $\Delta$ as follows: for any two numbers $a$ and $b$ $\min(a, b) = \dfrac{a+b}{2} - \dfrac{|a-b|}{2}$, then

$\Delta = 1 - \dfrac{1}{2} \int |f_{(1)}(-x) - f_{(r)}(x)| dx$. Thus $\hat{\Delta}_{ERSS} \approx \sum_{i=1}^{C} \dfrac{\delta_x}{2} \left[ \min(\hat{f}_{(1)}(-w_i), \hat{f}_{(r)}(w_i)) + \min(\hat{f}_{(1)}(-w_{i-1}), \hat{f}_{(r)}(w_{i-1})) \right]$, can be written as

$$\hat{\Delta}_{ERSS} = 1 - \frac{1}{2}\int |\hat{f}_{K(1)}(-w) - \hat{f}_{K(r)}(w)| \, dw$$

$$\approx 1 - \frac{1}{2}\sum_{i=1}^{C}\frac{\delta_x}{2}\Big[|\hat{f}_{(1)}(-w_i) - \hat{f}_{(r)}(w_i)| + |\hat{f}_{(1)}(-w_{i-1}) - \hat{f}_{(r)}(w_{i-1})| \; \Big].$$

Using the above results and under the null hypothesis of symmetry, i.e. $f_{(1)}(-w_i) = f_{(r)}(w_i)$, then,

$$|\hat{f}_{(1)}(-w) - \hat{f}_{(r)}(w)| = |(\hat{f}_{(1)}(-w) - f_{(1)}(-w)) + (f_{(r)}(w) - \hat{f}_{(r)}(w))| \tag{10}$$
$$< |\hat{f}_{(1)}(-w) - f_{(1)}(-w)| + |f_{(r)}(w) - \hat{f}_{(r)}(w)| \to^P 0,$$

To prove that $\hat{\Delta}_{ERSS} \xrightarrow{P} \Delta$, where $\Delta = 1$ under the null hypothesis, we need to show that $\left|\hat{\Delta}_{ERSS} - \Delta\right| \xrightarrow{p} 0$. Now

$$\left|\hat{\Delta}_{ERSS} - \Delta\right| = \frac{1}{2}\left|\int |\hat{f}_{(1)}(-w) - \hat{f}_{(r)}(w)| \, dw - \int |f_{(1)}(-w) - f_{(r)}(w)| \, dw\right|$$

$$\leq \frac{1}{2}\int\left[\Big||\hat{f}_{(1)}(-w) - \hat{f}_{(r)}(w)| - |f_{(1)}(-w) - f_{(r)}(w)|\Big|\right] dw$$

$$\leq \frac{1}{2}\int\Big|[\hat{f}_{(1)}(-w) - \hat{f}_{(r)}(w)] - [f_{(1)}(-w) - f_{(r)}(w)]\Big| dw$$

$$\leq \frac{1}{2}\int |\hat{f}_{(1)}(-w) - f_{(1)}(-w)| \, dw + \frac{1}{2}\int |f_{(r)}(w) - \hat{f}_{(r)}(w)| \, dw.$$

Since, $\hat{f}_{(1)}(-w) \xrightarrow{p} f_{(1)}(-w)$ and $\hat{f}_{(r)}(w) \xrightarrow{p} f_{(r)}(w)$ for all continuous points $w$ of $f_{(1)}$ and $f_{(r)}$, we have

$$\lim_{m\to\infty}\left|\hat{\Delta}_{ERSS} - \Delta\right| \leq \frac{1}{2}\lim_{m\to\infty}\int |\hat{f}_{(1)}(-w) - f_{(1)}(-w)| \, dw + \frac{1}{2}\lim_{m\to\infty}\int |f_{(r)}(w) - \hat{f}_{(r)}(w)| \, dw.$$

Now for a given $\varepsilon > 0$, let $A = \{w : |\hat{f}_{(1)}(-w) - f_{(1)}(-w)| > \varepsilon/2\}$ and $A_1 = \{w : |f_{(r)}(w) - \hat{f}_{(r)}(w)| > \varepsilon/2\}$, we have

$$\lim_{m\to\infty}\left|\hat{\Delta}_{ERSS} - \Delta\right| \leq \frac{1}{2}\lim_{m\to\infty}\int |\hat{f}_{(1)}(-w) - f_{(1)}(-w)| \, dw + \frac{1}{2}\lim_{m\to\infty}\int |f_{(r)}(w) - \hat{f}_{(r)}(w)| \, dw$$

$$= \frac{1}{2}\lim_{m\to\infty}\Big\{\int_A |\hat{f}_{(1)}(-w) - f_{(1)}(-w)| \, dw + \int_{A^c} |\hat{f}_{(1)}(-w) - f_{(1)}(-w)| \, dw$$

$$+ \int_{A_1} |f_{(r)}(w) - \hat{f}_{(r)}(w)| \, dw + \int_{A_1^c} |f_{(r)}(w) - \hat{f}_{(r)}(w)| \, dw\Big\},$$

then

$$\lim_{m\to\infty}\left|\hat{\Delta}_{ERSS} - \Delta\right| \leq \frac{1}{2}\lim_{m\to\infty}\{P(\{w : |\hat{f}_{(1)}(-w) - f_{(1)}(-w)| > \varepsilon/2\}) + \varepsilon/2$$

$$+ P(\{w : |f_{(r)}(w) - \hat{f}_{(r)}(w)| > \varepsilon/2\}) + \varepsilon/2\}.$$

Clearly $\lim_{m\to\infty}\left|\hat{\Delta}_{ERSS} - \Delta\right| \to 0$ in probability, i.e. $\left|\hat{\Delta}_{ERSS} - \Delta\right| \xrightarrow{p} 0$. Hence $\hat{\Delta}_{ERSS} \xrightarrow{p} \Delta$.

The asymptotic distribution of $\hat{\Delta}_{ERSS}$ under the null hypothesis, using the results derived by Anderson et al. [29] is as follows: Let $f_1(x) = f_{(1)}(-x)$ and $f_2(x) = f_{(r)}(x)$,

$$C_{f_1, f_2} = \{x \in \mathbb{R} : f_1(x) = f_2(x) > 0\}, \quad C_{f_1} = \{x \in \mathbb{R} : f_1(x) < f_2(x) > 0\} \text{ and } \quad C_{f_2} = \{x \in \mathbb{R} : f_1(x) > f_2(x) > 0\}. \text{ Let } n_1 = n_2 = m, \quad h_- = h_+ = h$$

$p_0 = P(X \in C_{f_1, f_2})$, $p_1 = P(X \in C_{f_1})$ and $p_2 = P(X \in C_{f_2})$. Under the above assumption, we have the following asymptotic result:

$$\sqrt{m}(\hat{\Delta} - \Delta) - a_m \Rightarrow N(0, v), \text{ where, } v = p_0\sigma_0^2 + \sigma_1^2, \; \sigma_1^2 = p_1(1 - p_1) + p_2(1 - p_2), \; a_m = \sqrt{\frac{R}{h}}\int_{C_{f_1, f_2}} f_{(r)}^{1/2}(x) \, dx \cdot E(\min\{Z_1, Z_2\}),$$

$$\sigma_0^2 = R\int_{T_0}\Big(\text{cov}(\min\{Z_1, Z_2\}, \min\{\rho(t)Z_1 + \sqrt{1 - \rho(t)^2}Z_3, \rho(t)Z_2 + \sqrt{1 - \rho(t)^2}Z_4\}\Big) dt,$$

$$\rho(t) = \frac{1}{R}\int_{\mathbb{R}} K(u)K(u + t) \, du, \; R = \int_{\mathbb{R}} K^2(u) \, du, \; Z_1, Z_2, Z_3 \text{ and } Z_4 \text{ are independent standard normal variables and } T_0 = \{t \in \mathbb{R} : |t| < 1\}.$$

However, under the null hypothesis that $f_1(x) = f_{(1)}(-x) = f_2(x) = f_{(r)}(x)$ then the above result is reduced to

$$\sqrt{m}(\hat{\Delta}_{ERSS} - \Delta)_{H_0} - a_m \Rightarrow N(0, \sigma_0^2).$$

## Simulation Study

To get some insights about the performance of our new test of symmetry based on $\hat{\Delta}_{ERSS}$ we conducted the following simulation. We compared our proposed test of symmetry with its counterpart used by McWilliams [4], Modarres and Gastwirth [8] and Samawi et al. [12] (using overlap measure) tests of symmetry.

McWilliams [4] runs test is described as follows: For any random sample of size n, let $Y_{(1)}, Y_{(2),...}, Y_{(n)}$ denote the sample values ordered from the smallest to largest according to their absolute value (signs are retained), and $S_1, S_{2,...}, S_n$ denote indicator variables designating the sign of the $Y_{(j)}$ values [ $S_j = 1$ if $Y_{(j)}$ is nonnegative, 0 otherwise ]. Thus, the test statistic used for testing symmetry is $R^*$ = the number of runs in $S_1, S_{2,...}, S_n$ sequence= $1 + \sum_{j=2}^{n} I_j$, where $I_j = \begin{cases} 0 & \text{if } S_j = S_{j-1} \\ 1 & \text{if } S_j \neq S_{j-1} \end{cases}$.

The test is to reject the null hypothesis if $R^*$ is smaller than a critical value $(c_\alpha)$ at level of significant $\alpha$. However, the Modarres and Gastwirth [8] test is $M_p = 1 + \sum_{j=2+n_p}^{n} \phi(j) I_j$, where,

$\phi(j) = \begin{cases} j - n_p & \text{if } j > n_p \\ 0 & \text{otherwise,} \end{cases}$ and $n_p$ is an interger.

If p=0, $\phi(j)$ terms are Wilcoxon scores. Otherwise, they are percentile-modified scores. The Modarres and Gastwirth [9] test is the hybrid test of sign test in the first stage and a percentile-modified two-sample Wilcoxon test in the second stage $(W_{0.80})$.

In this simulation, SAS version 9.2 {proc kde; method=srot} is used. The generalized lambda distribution see, Ramberg and Schmeiser [36] is used in our simulation with the following set of parameters:

1- $\lambda_1 = 0, \lambda_2 = 0.197454, \lambda_3 = 0.134915, \lambda_4 = 0.134915,$ (Symmetric)

2- $\lambda_1 = 0, \lambda_2 = 1, \lambda_3 = 1.4, \lambda_4 = 0.25,$

3- $\lambda_1 = 0, \lambda_2 = 1, \lambda_3 = 0.00007, \lambda_4 = 0.1,$

4- $\lambda_1 = 3.586508, \lambda_2 = 0.04306, \lambda_3 = 0.025213, \lambda_4 = 0.094029,$

5- $\lambda_1 = 0, \lambda_2 = -1, \lambda_3 = -0.0075, \lambda_4 = -0.03,$

6- $\lambda_1 = -0.116734, \lambda_2 = -0.351663, \lambda_3 = -0.13, \lambda_4 = -0.16,$

7- $\lambda_1 = 0, \lambda_2 = -1, \lambda_3 = -0.1, \lambda_4 = -0.18,$

8- $\lambda_1 = 0, \lambda_2 = -1, \lambda_3 = -0.001, \lambda_4 = -0.13,$

9- $\lambda_1 = 0, \lambda_2 = -1, \lambda_3 = -0.0001, \lambda_4 = -0.17.$

To generate the observations we used $x_i = \lambda_1 + \frac{1}{\lambda_2}(u_i^{\lambda_3} - (1-u_i)^{\lambda_4}, i = 1,...,m$, where $u_i$ a uniform random number is. The significance level considered is $\alpha = 0.05$, with sample sizes n=30, 50, and 100. Our simulation is based on 1,000 simulated samples. It is clear that 95% and 99% confidence intervals of the true probability of type I error under the null hypothesis with $\alpha = 0.05$ are (0.0457, 0.0543) and (0.0435, 0.0575) respectively. Note that in the below tables, the values of skewnees $(\alpha_3)$ and kurtosis $(\alpha_4)$ are from McWilliams [4].

Table 1a first showcases the estimated probability of type I error. Our test is an asymptotic test with a slight bias in $\Delta$ estimation and in the variance estimation for a small sample size. For sample sizes more than 30, the test seems to have an estimated probability of type I error close to the nominal value 0.05. Table 1a and Table 1b show that using $\hat{\Delta}_{ERSS}$ based test is more powerful than McWilliams [4] and Baklizi [7]. In all cases, our proposed procedure is even more efficient than the tests of symmetry proposed by Modarres and Gastwirth [8], Modarres and Gastwirth [9] and Samawi et al. [12] In all of the tests within the comparison, the power of all tests of symmetry increases as the sample size increases. Finally the power of $\hat{\Delta}_{ERSS}$ based tests increases as the set size r increases

## Illustration using Noninvasive Measurement of Cardiac Output by Electrical Velocimetry in Neonate Data

The samples selected in our illustration is from a study designed to evaluate the effectiveness of a new technology, Electrical Velocimetry (E.V.) for a non-invasive cardiac output (CO) and stroke volume (SV) in neonates [37]. One of the research questions is whether the CO measure is the same for low birth-weight infants and non-low birth-weight infants. Low birth weight is defined as less than 1.5 kg. Thus we compared CO for neonates with birth-weight less than 1.5kg to neonates with birth-weight greater or equal to 1.5kg.

As it is frequently the case for this type of study, the underlying distribution is assumed "normal", or at least symmetric. In either

**Table 1a:** Probability of Type I Error under the Null Hypothesis ($\alpha = 0.05$).

| Distribution | $n$ | $R^*$ | $\hat{\Delta}_{SRS}$ Samawi et al. [12]. | $M_{0.25}$ | $W_{0.80}$ | $\hat{\Delta}_{ERSS}\ r=2$ | $\hat{\Delta}_{ERSS}\ r=3$ | $\hat{\Delta}_{ERSS}\ r=4$ | $\hat{\Delta}_{ERSS}\ r=5$ |
|---|---|---|---|---|---|---|---|---|---|
| (1) | 30 | 0.047 | 0.069 | 0.053 | 0.055 | 0.046 | 0.056 | 0.056 | 0.044 |
| $\lambda_1 = 0, \lambda_2 = 0.197454,$ | 50 | 0.050 | 0.054 | 0.048 | 0.048 | 0.054 | 0.056 | 0.049 | 0.049 |
| $\lambda_3 = 0.134915,$ $\lambda_4 = 0.134915,$ $\alpha_3 = 0, \alpha_4 = 3.0$ **Null** | 100 | 0.064 | 0.053 | 0.050 | 0.052 | 0.048 | 0.053 | 0.046 | 0.051 |
| (2) | 30 | 0.297 | 0.495 | 0.583 | 0.656 | 0.751 | 0.906 | 0.973 | 0.999 |
| $\lambda_1 = 0, \lambda_2 = 1,$ | 50 | 0.476 | 0.836 | 0.846 | 0.949 | 0.997 | 1.000 | 1.000 | 1.000 |
| $\lambda_3 = 1.4, \lambda_4 = 0.25$ $\alpha_3 = 0.5, \alpha_4 = 2.2$ | 100 | 0.776 | 0.999 | 0.990 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| (3) | 30 | 0.438 | 0.852 | 0.761 | 0.762 | 0.960 | 0.975 | 0.999 | 0.999 |
| $\lambda_1 = 0, \lambda_2 = 1,$ | 50 | 0.683 | 0.966 | 0.950 | 0.992 | 1.000 | 1.000 | 1.000 | 1.000 |
| $\lambda_3 = 0.00007,$ $\lambda_4 = 0.1,$ $\alpha_3 = 1.5, \alpha_4 = 5.8$ | 100 | 0.927 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| (4) | 30 | 0.117 | 0.375 | 0.172 | 0.280 | 0.384 | 0.398 | 0.413 | 0.482 |
| $\lambda_1 = 3.586508,$ | 50 | 0.131 | 0.512 | 0.251 | 0.544 | 0.689 | 0.706 | 0.766 | 0.808 |
| $\lambda_2 = 0.04306,$ $\lambda_3 = 0.025213,$ $\lambda_4 = 0.094029$ $\alpha_3 = 0.9, \alpha_4 = 4.2$ | 100 | 0.223 | 0.767 | 0.414 | 0.883 | 0.929 | 0.940 | 0.958 | 0.985 |
| (5) | 30 | 0.145 | 0.459 | 0.234 | 0.407 | 0.484 | 0.569 | 0.616 | 0.716 |
| $\lambda_1 = 0, \lambda_2 = -1,$ | 50 | 0.192 | 0.580 | 0.356 | 0.736 | 0.832 | 0.846 | 0.889 | 0.921 |
| $\lambda_3 = -0.0075,$ $\lambda_4 = -0.03,$ $\alpha_3 = 1.5, \alpha_4 = 7.5$ | 100 | 0.338 | 0.846 | 0.588 | 0.972 | 0.983 | 0.985 | 0.991 | 0.997 |

case, a test of symmetry is almost never considered in determining how to proceed in the analysis. Based on the conclusions of a test of symmetry, the analyst can choose the most powerful test for location. However, before deciding on the test procedure, we need to check the assumption of symmetry of underlying distribution of CO for the premature and term infants, with birth weight less than 1.5 gk and with birth weight greater or equal to 1.5 kg.

Table 2a consists of two selected SRS samples of CO for neonates with birth weight less than 1.5 gk and neonates with birth weight more than 1.5 kg. Also, Table 2a consists of two selected ERSS samples of CO for both neonates with birth weight less than 1.5 kg and neonates with birth weight more than or equal to1.5 kg. Since CO measure and birth weight are positively correlated, the ranking was performed based on birth weights. ERSS samples in Table 2a consist of first half as the first order statistics and the second half as the third order statistics (the maximum).

Table 2b has the results of the runs and overlap tests of symmetry for the underlying distribution for CO patients. From all samples, we reject the assumption that the underlying distribution is symmetric. Table 2b shows the results of the Mann-Whitney test for two-independent samples. Table 2c shows that there is a significant difference on average in the CO measures between the low birth weight neonates (less than 1.5kg) and the non-low-birth-weight neonates (greater than or equal to 1.5kg).

Based on our simulation and real data example, the proposed test of symmetry based on ERSS sample overlap measure, appears to outperform the other tests of symmetry in the literature in terms of power. Our test is more sensitive to detect a slight asymmetry in the underlying distribution than other tests proposed in the literature. Drawing an ERSS is easier than the ordinary RSS and other RSS variations. Also, the kernel density estimation literature is very rich and many of the proposed and the improved methods are available on statistical software, such as SAS™, S-plus, Stata and R. Since overlap measures can be used in multivariate cases as well as in univariate cases, our proposed test of symmetry can be extended to multivariate cases for diagonal symmetry, conditional symmetry and other types of symmetry. In addition, our test procedure and kernel density estimation are valid under large sample size (n>30)

**Table 1b:** (continue) Power of Overlap based test and Run Tests under Alternative Hypotheses ($\alpha = 0.05$).

| Case # | $n$ | $R^*$ | $\hat{\Delta}_{SRS}$ | $M_{0.25}$ | $W_{0.80}$ | $\hat{\Delta}_{ERSS}$ r=2 | $\hat{\Delta}_{ERSS}$ r=3 | $\hat{\Delta}_{ERSS}$ r=4 | $\hat{\Delta}_{ERSS}$ r=5 |
|---|---|---|---|---|---|---|---|---|---|
| (6) $\lambda_1 = 0, \lambda_2 = 0.197454,$ $\lambda_3 = 0.134915,$ $\lambda_4 = 0.134915,$ $\alpha_3 = 0, \alpha_4 = 3.0$ | 30 | 0.050 | 0.155 | 0.055 | 0.068 | 0.253 | 0.304 | 0.327 | 0.431 |
| | 50 | 0.056 | 0.166 | 0.060 | 0.077 | 0.500 | 0.717 | 0.739 | 0.747 |
| | 100 | 0.051 | 0.207 | 0.068 | 0.130 | 0.793 | 0.875 | 0.911 | 0.942 |
| (7) $\lambda_1 = 0, \lambda_2 = 1,$ $\lambda_3 = 1.4, \lambda_4 = 0.25$ $\alpha_3 = 0.5, \alpha_4 = 2.2$ | 30 | 0.090 | 0.196 | 0.096 | 0.166 | 0.325 | 0.410 | 0.458 | 0.520 |
| | 50 | 0.097 | 0.236 | 0.125 | 0.284 | 0.656 | 0.667 | 0.758 | 0.835 |
| | 100 | 0.124 | 0.354 | 0.176 | 0.589 | 0.891 | 0.946 | 0.963 | 0.974 |
| (8) $\lambda_1 = 0, \lambda_2 = -1,$ $\lambda_3 = -0.001,$ $\lambda_4 = -0.13,$ $\alpha_3 = 3.16, \alpha_4 = 23.8$ | 30 | 0.534 | 1.000 | 0.830 | 0.806 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 50 | 0.744 | 1.000 | 0.972 | 0.995 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 100 | 0.972 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| (9) $\lambda_1 = 3.586508,$ $\lambda_2 = 0.04306,$ $\lambda_3 = 0.025213,$ $\lambda_4 = 0.094029$ $\alpha_3 = 0.9, \alpha_4 = 4.2$ | 30 | 0.560 | 1.000 | 0.865 | 0.808 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 50 | 0.816 | 1.000 | 0.985 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 100 | 0.976 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

*Results are taken from Modarres and Gastwirth (1996) and Modarres and Gastwirth (1998) respectively

**Table 2a:** Selected samples of CO data.

| SRS | | ERSS | |
|---|---|---|---|
| Birth-weight <1.5 | Birth-weight ≥1.5 | Birth-weight <1.5 | Birth-weight ≥1.5 |
| 0.05 | 0.21 | 0.050 | 0.100 |
| 0.06 | 0.16 | 0.050 | 0.120 |
| 0.06 | 0.12 | 0.050 | 0.132 |
| 0.05 | 0.20 | 0.060 | 0.220 |
| 0.05 | 0.13 | 0.060 | 0.150 |
| 0.06 | 0.22 | 0.060 | 0.228 |
| 0.06 | 0.22 | 0.090 | 0.178 |
| 0.08 | 0.15 | 0.080 | 0.182 |
| 0.10 | 0.22 | 0.080 | 0.215 |
| 0.08 | 0.20 | 0.080 | 0.158 |
| 0.08 | 0.23 | 0.070 | 0.155 |
| 0.08 | 0.25 | 0.070 | 0.218 |
| 0.07 | 0.23 | 0.080 | 0.208 |
| 0.11 | 0.16 | 0.128 | 0.220 |
| 0.08 | 0.22 | 0.128 | 0.350 |
| 0.13 | 0.16 | 0.208 | 0.350 |
| 0.16 | 0.22 | 0.165 | 0.360 |
| 0.12 | 0.33 | 0.162 | 0.420 |
| 0.18 | 0.26 | 0.188 | 0.270 |
| 0.16 | 0.31 | 0.188 | 0.440 |
| 0.16 | 0.29 | 0.262 | 0.510 |
| 0.19 | 0.27 | 0.372 | 0.510 |
| 0.37 | 0.51 | 0.358 | 0.520 |
| 0.16 | 0.51 | 0.222 | 0.540 |
| 0.18 | 0.52 | 0.202 | 0.520 |
| 0.11 | 0.52 | 0.182 | 0.520 |
| | 0.53 | | 0.530 |
| | 0.73 | | 0.730 |

**Table 2b:** Runs test of symmetry with summary statistics.

|  | CO Measure | N | Runs test | P-value | OVL test | P-value |
|---|---|---|---|---|---|---|
| **SRS** | **Birth-weight <1.5** | 26 | 6 | 0.003 | -6.414 | <0.00001 |
|  | **Birth-weight ≥1.5** | 28 | 4 | <0.00001 | -4.356 | <0.00001 |
| **ERSS** | **Birth-weight <1.5** | 26 | 2 | <0.00001 | -14.540 | <0.00001 |
|  | **Birth-weight ≥1.5** | 28 | 2 | <0.00001 | -4.729 | <0.00001 |

Table 2b has the results of the runs and overlap tests of symmetry for the underlying distribution for CO patients. From all samples, we reject the assumption that the underlying distribution is symmetric

**Table 2c:** Mann-Whitney test for two-Independent Samples.

| | Mann-Whitney U test (difference of medians of CO between<1.5 kg and ≥1.5kg weight ) | |
|---|---|---|
| **Sample Type** | **Test** | **P-value** |
| **SRS** | 59.5 | <0.00001 |
| **ERSS** | 118 | <0.00001 |

Table 2c: Shows that there is a significant difference on average in the CO measures between the low birth weight neonates (less than 1.5kg) and the non-low-birth-weight neonates (greater than or equal to 1.5kg)

and some regular conditions such as light tail underlying distribution functions. However, our simulation indicates that our procedure is still perform better than other test even for a sample size n=30 or larger and different underlying distributions.

## Acknowledgment

## Conflict of Interest

No conflict of interest

## References

1. Butler CC (1969) A test for symmetry using sample distribution function. The Annals of Mathematical Statistics 40: 2211-2214.

2. Rothman ED, Woodroofe M (1972) A Cramer-Von Mises type statistic for testing symmetry. The Annals of Mathematical Statistics 43(6): 2035-2038.

3. Hill DL, Rao PV (1977) Test of Symmmetry based on Cramer-Von Mises statistics. Biometrika 64(3): 489-494.

4. McWilliams TP (1990) A distribution–free test of symmetry based on a runs statistic. Journal of the American Statistical Association 85 (412): 1130-1133.

5. Ozturk O (2001) A nonparametric test of symmetry versus asymmetry for ranked-set samples. Communication in Statistics- Theory and Methods 30(10): 2117-2133.

6. Tajuddin IH (1994) Distribution-Free test for symmetry based on Wilcoxon two-sample test. Journal of Applied Statistics 21(5): 409-415.

7. Baklizi A (2003) A conditional distribution free runs test for symmetry. Nonparametric Statistics 15(6): 713-718.

8. Modarres R, Gastwirth JL (1996) A modified runs test of symmetry. Statistics & Probability Letters 31: 107-112.

9. Modarres R, Gastwirth JL (1998) Hybrid test for the hypothesis of symmetry. Journal of Applied Statistics 25(6): 777-783.

10. Samawi HM, Ahmed MS, Abu-Dayyeh W (1996) Estimating the population mean using extreme ranked set sampling. Biometrical Journal 38(5): 577- 586.

11. Samawi HM, Vogel R, Senkowski CK (2010) On Distribution-Free Runs Test for Symmetry Using Extreme Ranked Set Sampling With an Application Involving Base Deficit Score. Journal of Statistical Theory and Practice 4(2): 289-301.

12. Samawi HM, Helu A, Vogel R (2011) A Nonparametric Test of Symmetry Based on the Overlapping Coefficient. Journal of Applied Statistics 38(5): 885-898.

13. Weitzman, M. S. (1970). Measures of overlap of income distributions of white and Negro families in the United States. Technical paper No. 22, Department of Commerce, Bureau of Census, Washington. USA.

14. Bradley, E.L., Piantadosi, S. (1982). The overlapping coefficient as a measure of Agreement between distributions. Technical Report, Department of Biostatisticsand Biomathematics, University of Alabama at Birmingham, Birmingham, USA.

15. Inman HF, Bradley EL (1989). The Overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. Communications in Statistics - Theory and Methods 18(10): 3851-3874.

16. Clemons TE (1996) The overlapping coefficient for two normal probability functions with unequal variances. Unpublished Thesis, Department of Biostatistics, University of Alabama at Birmingham, Birmingham, USA.

17. Reiser B, Faraggi D (1999) Confidence intervals for the overlapping coefficient: the normal equal variance case. The statistician 48(3): 413-418.

18. Clemons TE, Bradley EL (2000) A nonparametric measure of the overlapping coefficient. Computational Statistics & Data Analysis 34(1): 51-61.

19. Mulekar MS, Mishra SN (2000) Confidence interval estimation of overlap: Equal means case. Computational Statistics & Data Analysis 34(2): 121-137.

20. Al-Saidy O, Samawi HM, Al-Saleh MF (2005) Inference on Overlapping Coefficients under the Wei bull Distribution: Equal Shape Parameter. ESAIM: Probability and Statistics (P&S) 9: p206-219.

21. Schmid F, Schmidt A (2006) Nonparametric estimation of the coefficient of overlapping-theory and empirical application. Computational Statistics & Data Analysis 50 (6): 1583-1596.

22. Al-Saleh MF, Samawi HM (2007) Inference on Overlapping Coefficients in Two Exponential Populations. Journal of Modern Statistical Science 6(2): 503-516.

23. Samawi HM, Al-Saleh MF (2008) Inference of Overlapping Coefficients in two Exponential Populations using Ranked set Sample. Communication of Korean of Statistical Society 15 (2): 147-159.

24. McIntyre GA (1952) A method for unbiased selective sampling, using ranked sets. Australian Journal of Agriculture Research 39(4): 385-90.

25. Kaur, A, Patil GP, Sinha AK, Taillie, C (1995) Ranked set sampling: Anannotated bibliography. Environmental Ecological Statistics 2: 25-54.

26. Patil GP, Sinha Ak, Tillie C (1999) Ranked set sampling: a bibliography. Environmental Ecological Statistics 6: 91-98.

27. Kaur A, Patil GP, Taillie C (2000) Optimal allocation for symmetric distributions in ranked set sampling. Ann Inst Statist Math 52(2): 239-254.

28. Sinha, Arun K (2005) On some recent developments in ranked set sampling. Bulletin of Informatics and Cybernetics 37: 137-160.

29. Anderson G, Linton O, Whang Y (2010) Nonparametric estimation of a polarization measure. University of Toronto, The London School of Economics and Seoul National University. Discussion Paper.

30. Wegman EJ (1972) Nonparametric probability density estimation: a summary of available methods. Technometrics 14(3): 533-546.

31. Wegman EJ (1982) Density estimation. Encyclopedia of Statistical Sciences. (Vol. 2), Wiley, New York, Pp. 309-315.

32. Van Kerm P (2003) Adaptive kernel density estimation. The Stata Journal 3(2): 148-156.

33. Chen EJ, Kelton WD (2006) Empirical Evaluation of Data-Based Density Estimation. Proceedings of the Winter Simulation Conference: 333-341.

34. Silverman BW (1986) Density estimation for statistics and data analysis. Chapman & Hall, London.

35. Wand MP, Jones MC (1995) Kernel Smoothing, Chapman & Hall, London.

36. Ramberg JS, Schmeiser BW (1974) An approximate method for generating Asymmetric random variables. Communications of the ACM 17: 78-82.

37. Norozi K, C Beck1 C, Osthaus WA, Wille I, Wessel A, Bertram H (2008) Electrical velocimetry for measuring cardiac output in children with congenital heart disease. Oxford Journals Medicine & Health BJA 100(1): 88-94.