LOCATION ESTIMATION WITH NONPARAMETRIC CORRELATION COEFFICIENTS

JOHN BRUDER, LI-CHIOU LEE, MIKE THIEL, RUDY A. GIDEON

Abstract. This paper provides a location estimator from an equation for continuous data using any nonparametric correlation coefficient. Specific results will be given for GDCC, the Greatest Deviation Correlation Coefficient (Gideon and Hollister, 1987); the robustness of the estimator is studied. The location estimator based on Kendall's correlation coefficient is closely related to the signed rank test estimator for location which uses the Walsh averages. This paper completes the development of basic statistics for location, scale, and regression areas based on CCs which is independent of classical statistics.

1. Introduction

Let $\{x_i\}$, $i = 1, 2, \ldots, n$ be a random sample from an absolutely continuous random variable with distribution function F which has a point of symmetry. Let x represent the data in vector notation, and let x^0 represent the vector of order statistics, $x^0 = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})\prime$. For any correlation coefficient r, let $r(x, y)$ be its value on the bivariate data (x, y) which is an $n \times 2$ matrix. Define $e' = (1, 2, \ldots, n)$, $\quad 1' = (1, 1, \ldots, 1)$ and let the following equation be examined for the location estimator θ:

$$r(e, |x^0 - \theta 1|) = 0. \tag{1}$$

The absolute value notation denotes the vector of absolute values. It is claimed that the quantity θ is an estimator whose properties depend on the chosen correlation coefficient. Before specializing to the Greatest Deviation Correlation Coefficient (see Section 6 for an example of the computation of GDCC and Section 2 for its definition), it can be shown that equation (1) results in an estimator θ that satisfies some essential properties of a location estimator. Even though this paper mainly examines only NPCCs, any CC could be studied; some comparisons are done with Pearson's CC. This work is part of a system of estimation based on correlation coefficients called CES or Correlation Estimation System.

2. Scale changes, location shifts, and symmetry

The estimator θ must behave properly with respect to scale and location changes on the data. First, let the data x be shifted by an amount h : $\quad y = x + h1$ so that $y^0 = x^0 + h1$. If θ satisfies equation (1) and if equation (1) is a location equation then $\theta + h$ must satisfy equation (1) when the shifted data y are used. For any nonparametric correlation coefficient, which only depends on ranks,

$$r(e, |y^0 - \theta^* 1|) = r(e, |x^0 + h1 - \theta^* 1|) = r(e, |x^0 - \theta 1|) = 0, \text{where} \quad \theta^* = \theta + h.$$

Thus estimator θ is shifted by the correct amount. Now let $y^0 = hx^0$, where $h > 0$ is the scale change, and if θ^* is the location estimate for y^0, it must be shown that $\theta^* = h\theta$. Now $r(e, |y^0 - \theta^* 1|) = r(e, |hx^0 - \theta^* 1|)$ and this is obviously zero if $\theta^* = \theta h$ since, $r(e, h|x^0 - \theta 1|) = 0$ by the scale invariance property of the CC.

If μ is a point of symmetry of a data set, then any location estimator should give this point as the estimate; thus, for x a data set symmetric about μ, it must be shown that $r(e, |x^0 - \mu 1|) = 0$. Consider the case $n = 2k + 1$, where the point of symmetry is $\mu = x_{(k+1)}$, the middle order statistic. Then

$$\begin{aligned} |x_{(1)} - \mu| &= x_{(n)} - \mu \\ |x_{(2)} - \mu| &= x_{(n-1)} - \mu \\ &\vdots \\ |x_{(k)} - \mu| &= x_{(k+2)} - \mu \\ x_{(k+1)} - \mu &= 0 \end{aligned}$$

For $y_i = x_{(i)} - \mu, \quad i = 1, 2, \ldots, n$, we have

$$\begin{aligned} -y_1 &= y_n \\ -y_2 &= y_{n-1} \\ &\vdots \\ -y_k &= y_{k+2} \\ y_{k+1} &= 0 \end{aligned}$$

and if y is the vector $|x^0 - \mu 1|$, then $y' = (-y_1, -y_2, \ldots, -y_k, 0, y_{k+2}, \ldots, y_n)$ where all components are positive. Let $rev(y)$ be the vector where components of y' are reversed, so $rev(y) = (y_n, y_{n-1}, \ldots, y_{k+2}, 0, -y_k, \ldots, -y_1) = y$, and $rev(e) = (n, n-1, \ldots, 2, 1)$. Clearly $r(e, y) = r(rev(e), rev(y))$ but since $y = rev(y)$, $r(e, y) = r(rev(e), y)$.

Now for any nonparametric correlation coefficient $r(rev(e), y) = -r(e, y)$; this is easy to show for any rank based CC. Thus, $r(e, y) = r(rev(e), y) = -r(e, y)$, but this is impossible unless $r(e, y) = 0$. This proof assumes that a symmetric rank adjustment method is used with ties such as the mid-rank procedure or the maximum-minimum method proposed in Gideon and Hollister (1987).

For reference, the definition of GDCC or r_{gd} is given here:

$$r_{gd}(x, y) = \left(\max_{1 \le i \le n} (d_i^-) - \max_{1 \le i \le n} (d_i^+) \right) / \left[n/2 \right],$$

where $d_i^+ = \sum_{j=1}^{i} I(u_j > i)$, $d_i^- = \sum_{j=1}^{i} I(n+1-u_j > i)$, I is the indicator function, and the brackets denote the greatest integer function. The vector u consists of the ranks of y once x has been sorted.

3. Kendall's τ: location estimator and its relationship to the Signed Rank statistic

In this section the location estimator will be developed from equation (1) using Kendall's τ. The development closely parallels the location estimator of the Wilcoxon signed rank test using Walsh averages as explained in the textbook by Hettmansperger (1984), section 2.3.

Let $W_{ij} = (x_i + x_j)/2$ for $i < j$ over the $n(n-1)/2$ distinct pairs, and let $V = \{x_i | i = 1, 2, \ldots, n\}$. Then the Walsh averages are the set of numbers $\bigcup_{i<j} W_{ij} \bigcup V$, and the signed rank location estimator is the median of this set. It will be shown that the location estimator using Kendall's τ in equation (1) is the median of the set $\{W_{ij} | i < j\}$ and they are asymptotically equivalent. Intuitively, since the number of elements in W_{ij} is growing as n^2 while the number of elements in the set V is only growing as n, for large n, the estimation based on these two statistics will become very close, and the asymptotic distribution the same.

Theorem 1. The solution to $\tau(e, |x^0 - \theta 1|) = 0$ is given by the median of the set of numbers $W_{ij} = (x_i + x_j)/2$ for $i < j$; that is, $\hat{\theta} = med(W_{ij})$ is the CES location estimator with Kendall's τ.

Proof: If a location null hypothesis is $H_0 : \mu = \mu_0$, then the value of the signed rank statistic has $(n(n-1)/2+n)+1 = n(n+1)/2+1$ distinct values depending on where μ_0 lies. The Walsh averages partition the axis into $n(n+1)/2+1$ disjoint sets upon which the signed rank statistic takes its values. If μ_0 were moved continuously from less than $x_{(1)}$ to beyond $x_{(n)}$, the value of the signed rank statistic would change at each Walsh average in a monotonic fashion from one extreme to the other. In the same manner the value of $\tau(e, |x^0 - \theta 1|)$ goes from $+1$ when $\theta < (x_{(1)} + x_{(2)})/2$ to -1 when $\theta > (x_{(n)} + x_{(n-1)})/2$ and its value changes in a monotonic fashion at each point in the set $\bigcup_{i<j} W_{ij}$. The essential difference is that the ranks of $|x^0 - \theta 1|$ do not change at points in the set V whereas the signs of the ranks do change at these points. The details of the monotonic change are now shown. As θ approaches the average $(x_{(i)} + x_{(j)})/2, i < j$ from the left, $|x_{(i)} - \theta| < |x_{(j)} - \theta|$, and once this average is crossed, $|x_{(i)} - \theta| > |x_{(j)} - \theta|$. Thus, in $\tau(e, |x^0 - \theta 1|)$, a transposition in the ranks of $|x^0 - \theta 1|$ occurs at the point W_{ij}. Let n_c and n_d be the number of concordances and discordances in the calculation of τ. Every transposition must change $n_c - n_d$ by at least two because if a concordance is lost, a discordance is gained. Now $\tau = (n_c - n_d)/\binom{n}{2}$ and $n_c - n_d$ equals $n(n-1)/2$ for $\theta < (x_{(1)} + x_{(2)})/2$ and $-n(n-1)/2$ for $\theta > (x_{(n)} + x_{(n-1)})/2$.

A transposition occurs $n(n-1)/2$ times since there are that many points in $\bigcup_{i<j} W_{ij}$. Then $n_c - n_d$ changes by at least twice $n(n-1)/2$, i.e. by $n(n-1)$. Since $n(n-1)/2 - n(n-1) = -n(n-1)/2$, it must be true that $n_c - n_d$ changes its value by exactly two at each transposition. Since $n_c - n_d$ is a nonincreasing step function at the points in the set $\bigcup_{i<j} W_{ij}$, it follows that for Kendall's τ, equation (1) will be satisfied at $\hat{\theta} = med(W_{ij})$, since at this point $n_c - n_d = 0$. •

Kendall's τ location estimate and the signed rank statistic will now be shown to have the same asymptotic distributions except for a location and scale change. By a similar type of argument that appears later for r_{gd} or the Wilcoxon signed rank statistic (see exercise 2.10.2 in Hettmansperger, 1984), the distribution of $\tau(e, |X^0 - \mu 1|)$ is symmetric about μ.

To develop the asymptotic theory for the estimate of location from the signed rank statistics, the asymptotic distribution of T, the number of positive Walsh averages is first found. The following notation follows similarly to that of Hettmansperger (1984) on pages 48, 54, 76, and 202. For these Walsh averages, define

$$T_{ij} = \begin{cases} 1, & \text{if } W_{ij} > 0, i < j; \\ 0, & \text{otherwise.} \end{cases}$$

$$T_i = \begin{cases} 1, & \text{if } x_{(i)} > 0, i = 1, 2, \ldots, n\ ; \\ 0, & \text{otherwise.} \end{cases}$$

Then $T = \sum_{i<j} T_{ij} + \sum_{k=1}^{n} T_k$, and
$\binom{n}{2}\tau(e, |x^0|) = \sum_{i<j} sgn(j-i)sgn(|x_{(j)}| - |x_{(i)}|) = \sum_{i<j} sgn(|x_{(j)}| - |x_{(i)}|)$.
Since $x_{(i)} < x_{(j)}$ for $i < j$, $|x_{(j)}| - |x_{(i)}| < 0$ implies $W_{ij} < 0$, and $|x_{(j)}| - |x_{(i)} > 0$ implies $W_{ij} > 0$. Thus $\binom{n}{2}\tau(e, |x^0|) = \sum_{i<j} T_{ij} - \sum_{i<j} T^*_{ij}$, where

$$T^*_{ij} = \begin{cases} 1, & \text{if } W_{ij} < 0; \\ 0, & \text{otherwise.} \end{cases}$$

Since $\sum_{i<j} T_{ij} = n_c$, the number of concordances, $\sum_{i<j} T^*_{ij} = n_d$, the number of discordances, and $n_c + n_d = \binom{n}{2}$. In the no tie case, $\binom{n}{2}\tau(e, |x^0|) = 2\sum_{i<j} T_{ij} - \binom{n}{2}$. Finally, $\tau(e, |x^0|) = -1 + (2\sum_{i<j} T_{ij})/\binom{n}{2}$.

Now, to abbreviate notation, use τ in place of $\tau(e, |X^0|)$ and $T_K = \sum_{i<j} T_{ij}$; K denotes Kendall. The relationship between τ and T_K is developed by modifying work in Hettmansperger (1984). Assuming that the null hypothesis is true and using the semicircle discussion in Hettmansperger, it is easily shown that $T_K = \sum_{j=1}^{n}(j-1)W_j$, so $E(T_K) = \sum_{j=1}^{n}(j-1)E(W_j) = \frac{n(n-1)}{4}$. In the proceeding, as in Hettmansperger, W_j is 1 if $|x|_{(j)}$ corresponds to a positive observation and 0 otherwise. Then $T = \sum_{j=1}^{n} jW_j$ and the W_i are independent with $P(W_i = 0) = P(W_i = 1) = 1/2$, for all values of i.

Now $V(T_K) = \sum_{j=1}^{n}(j-1)^2 V(W_j) = \frac{1}{4}\sum_{k=1}^{n-1} k^2 = \frac{n(n-1)(2n-1)}{24}$, since $E(W_j) = 1/2$ and $V(W_j) = 1/4$. Hence, $E(\binom{n}{2}\frac{\tau+1}{2}) = E(T_K) = \frac{n(n-1)}{2}\frac{1}{2} = \binom{n}{2}\frac{1}{2}$. From this $E(\frac{\tau+1}{2}) = 1/2$ or $E(\tau) = 0$. Because $V(\binom{n}{2}\frac{\tau+1}{2}) = V(T_K) = \frac{n(n-1)(2n-1)}{24} = \frac{2n-1}{12}\binom{n}{2}$, $V(\tau) = \frac{2(2n-1)}{3n(n-1)}$.

Finally, the asymptotic null distribution of T_K is $N(\frac{n(n-1)}{4}, \frac{n(n-1)(2n-1)}{24})$ and of τ is $N(0, \frac{2(2n-1)}{3n(n-1)})$.

Thus the asymptotics of T, T_K and τ, for Hodges-Lehmann location estimate, are very similar. The main thrust of this section is to show that (1) is a reasonable criteria for a location estimate as it connects to the signed rank test when the correlation coefficient is τ. In the next section, correlation coefficient r_{gd} is used in (1).

4. The Greatest Deviation Correlation Coefficient (GDCC), r_{gd}

Theorem 2. Let $x^0 = (x_{(1)}, x_{(2)}, \ldots x_{(n)})'$ be the order statistics from a random sample and let $\hat{\theta}$ be the solution to the equation $r_{gd}(e, |x^0 - \theta 1|) = 0$ where again $1' = (1, 1, \ldots 1)$. Because of discreetness, the solution can be an interval; its midpoint

is taken to define a unique solution. Then $4\hat{\theta} = x_{\left(\left[\frac{n+1}{3}\right]\right)} + x_{\left(\left[\frac{n+3}{3}\right]\right)} + x_{\left(\left[\frac{2n+2}{3}\right]\right)} + x_{\left(\left[\frac{2n+4}{3}\right]\right)}$, where [ ] is the greatest integer notation, and $\hat{\theta}$ gives the r_{gd} location estimator for $n > 1$.

Proof: There are three cases which are proved separately but similarly,

(1) $n = 3k$, so that $4\hat{\theta} = x_{(k)} + x_{(k+1)} + x_{(2k)} + x_{(2k+1)}$
(2) $n = 3k + 1$, so that $4\hat{\theta} = x_{(k)} + x_{(k+1)} + x_{(2k+1)} + x_{(2k+2)}$
(3) $n = 3k + 2$, so that $4\hat{\theta} = 2x_{(k+1)} + 2x_{(2k+2)}$.

r_{gd} is a rank correlation coefficient and as a function of θ in the defining equation, is a decreasing step function. For $\theta < (x_{(1)} + x_{(2)})/2$, the ranks of the elements of $|x^0 - \theta 1|$ are in numerical order and hence, $r_{gd}(e, |x^0 - \theta 1|) = +1$. Only when θ passes over the average of a pair of the order statistics, can the value of r_{gd} decrease.

For $\theta > (x_{(n-1)} + x_{(n)})/2$, the elements of $|x^0 - \theta 1|$ are in exact reverse numerical order and hence, $r_{gd}(e, |x^0 - \theta 1|) = -1$. Unlike Kendall's τ, r_{gd} does not always change its value at these averages.

The solution set of equation (1) is defined by left and right end points which are given by the average of a pair of the order statistics. The value of θ is taken to be the average of the left and right end points of the solution set. For each case, the left and right end points are now listed.

	Left	Right
(1)	$(x_{(k)} + x_{(2k)})/2$	$(x_{(k+1)} + x_{(2k+1)})/2$
(2)	$(x_{(k)} + x_{(2k+1)})/2$	$(x_{(k+1)} + x_{(2k+2)})/2$
(3)	$(x_{(k+1)} + x_{(2k+2)})/2$	$(x_{(k+1)} + x_{(2k+2)})/2$

In general, for tied values in any data set, r_{gd} is defined to be $(r^+_{gd} + r^-_{gd})/2$ where r^+_{gd} is the maximum value and r^-_{gd} the minimum value under all possible permutations of ranks within the tied value sets. Thus, for this last case, r_{gd} jumps over zero at this point, and by the definition of r_{gd} at a jump point, as the average of the two values, r_{gd} is exactly zero at this one point.

For case (1), let $x_l = (x_{(k)} + x_{(2k)})/2$. It is to be shown that for $\theta = x_l$,

$$\lim_{\theta \to x_l^-} r_{gd}(e, |x^0 - \theta 1|) = r^+_{gd}(e, |x^0 - \theta 1|) \equiv r^+_{gd} = 1/[n/2]$$

and that

$$\lim_{\theta \to x_l^+} r_{gd}(e, |x^0 - \theta 1|) = r^-_{gd}(e, |x^0 - \theta 1|) \equiv r^-_{gd} = 0$$

From this we have $r_{gd} = (r^+_{gd} + r^-_{gd})/2 = 1/(2[n/2])$, and hence x_l is the left end point of the solution set. In Table 1, let e and $|x^0 - \theta 1|$ be listed in a column with the elements of e denoted by i and those of $|x^0 - \theta 1|$ by r_i, the rank of the i^{th} one. Note that r_i is a function of θ. For $\theta = x_l$, the inequalities of Table 1 hold for the r_i. The standard layout (Gideon and Hollister, 1987) for the calculation of r_{gd} is found in Table 1 and the inequalities on the ranks r_i are explained as follows.

Inequalities on the ranks are sufficient to calculate r_{gd}. These inequalities come from examining the ranks of $|x^0 - \theta 1|$ where θ is chosen to be x_l. All the ranks of r_{k+1} to r_{2k-1} are less than or equal to $k - 1$ because all the $x_{(i)}$ in this range are closer to x_l than any $x_{(i)}$ for which $i < k+1$ or $i > 2k-1$. The values of $|x_{(k)} - x_l 1|$ and $|x_{(2k)} - x_l 1|$ are the same so that $r_k = r_{2k}$. The inequality $2k \leq r_1 \leq n$ is true

because the rank of $|x_{(1)} - x_l|$ is greater than the rank of $|x_{(i)} - x_l|, i = 2, 3, \ldots, 2k$. Likewise, $2k - 1 \leq r_2 \leq n$ is true because the rank of $|x_{(2)} - x_l|$ is greater than the rank of $|x_{(i)} - x_l|, i = 3, 4, \ldots, 2k$. In a similar fashion the inequalities proceed to $k + 2 \leq r_{k-1} \leq n$. By a similar examination the ranks r_i for $2k + 1 \leq i \leq n$ are as given in the table; however there are k terms not $k - 1$. Now $r_k < r_{k-1}$ and $r_{2k} < r_{2k-1}$ and $r_k = r_{2k}$ so both r_k and r_{2k} are less than or equal to $k+1$. But r_k and r_{2k} are also greater than $k-1$, so that r_k and r_{2k} are k or $k+1$ if differentiated, but actually tied at $(2k + 1)/2$.

Table 1: Inequalities for x_l

$rank(e)$	$rank(\lvert x^0 - \theta 1\rvert)$	
i	r_i	
1	r_1	$2k \leq r_1 \leq n$
2	r_2	$2k - 1 \leq r_2 \leq n$
$\vdots$	$\vdots$	$\vdots$
$k - 1$	r_{k-1}	$k + 2 \leq r_{k-1} \leq n$
k	r_k	$r_k = k$ or $k + 1$
$k + 1$	r_{k+1}	$0 \leq r_{k+1} \leq k - 1$
$k + 2$	r_{k+2}	$0 \leq r_{k+2} \leq k - 1$
$\vdots$	$\vdots$	$\vdots$
$2k - 1$	r_{2k-1}	$0 \leq r_{2k-1} \leq k - 1$
$2k$	r_{2k}	$r_{2k} = k$ or $k + 1$
$2k + 1$	r_{2k+1}	$k + 2 \leq r_{2k+1} \leq n$
$2k + 2$	r_{2k+2}	$k + 3 \leq r_{2k+2} \leq n$
$\vdots$	$\vdots$	$\vdots$
n	r_n	$2k \leq r_n \leq n$

In Table 2 r_{gd}^+ and r_{gd}^- are computed. The column headed by (1) is $\sum_{k=1}^{i} I(i < r_k)$, and column (2) is $\sum_{k=1}^{i} I(i < n + 1 - r_k)$, where I is the indicator function which is one if the inequality is true or else zero. Pay particular attention to the difference in ranks in the calculation of r_{gd}^+ and r_{gd}^- at $i = k$ and $i = 2k$.

Table 2: Calculation of GDCC for Location Estimate, $n = 3k$

i	$r_{gd}^{+} : r_i$	(1)	$n+1-r_i$	(2)	$r_{gd}^{-} : r_i$	(1)	$n+1-r_i$	(2)
1	$\geq 2k$	1	$\leq k+1$	≤ 1	$\geq 2k$	1	$\leq k+1$	≤ 1
2	$\geq 2k-1$	2	$\leq k+2$	≤ 2	$\geq 2k-1$	2	$\leq k+2$	≤ 2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$k-1$	$\geq k+2$	$k-1$	$\leq 2k-1$	$\leq k-1$	$\geq k+2$	$k-1$	$\leq 2k-1$	$\leq k-1$
k	k	$k-1$	$=2k+1$	$\leq k$	$k+1$	k	$2k$	$\leq k$
$k+1$	$\leq k-1$	$\leq k-1$	$\geq 2k+2$	$\leq k$	$\leq k-1$	$\leq k-1$	$\geq 2k+2$	$\leq k$
$k+2$	$\leq k-1$	$\leq k-1$	$\geq 2k+2$	$\leq k$	$\leq k-1$	$\leq k-1$	$\geq 2k+2$	$\leq k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2k-1$	$\leq k-1$	$\leq k-1$	$\geq 2k+2$	$=k$	$\leq k-1$	$\leq k-1$	$\geq 2k+2$	$\leq k$
$2k$	$k+1$	$\leq k-1$	$=2k$	$\leq k$	k	$\leq k-1$	$2k+1$	$=k$
$2k+1$	$\geq k+2$	$\leq k-1$	$\leq 2k-1$	$\leq k-1$	$\geq k+2$	$\leq k-1$	$\leq 2k-1$	$\leq k-1$
$2k+2$	$\geq k+3$	$\leq k-2$	$\leq 2k-2$	$\leq k-2$	$\geq k+3$	$\leq k-2$	$\leq 2k-2$	$\leq k-2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	$\geq 2k$	≤ 0	$\leq k+1$	≤ 0	$\geq 2k$	≤ 0	$\leq k+1$	≤ 0
		max		max		max		max
		$=k-1$		$=k$		$=k$		$=k$

In the computation of r_{gd}, there are only k terms for i from $2k+1$ to $n = 3k$ and the computation in columns (1) and (2) must decrease to zero which forces the entries on these columns to be decreasing as stated. Thus $r_{gd}^{+} = (k-(k-1))/\left[n/2\right] = 1/\left[n/2\right]$ and $r_{gd}^{-} = (k-k)/\left[n/2\right] = 0$; so $r_{gd} = (r_{gd}^{+} + r_{gd}^{-})/2 = 1/(2[n/2])$.

In a similar fashion if $x_r = (x_{(k+1)} + x_{(2k+1)})/2$, the right end point, and $\theta = x_r$ in $r_{gd}(e, |x^0 - \theta 1|)$, then $r_{gd}^{+} = (k-k)/\left[n/2\right] = 0$ and $r_{gd}^{-} = ((k-1)-k)/\left[n/2\right] = -1/\left[n/2\right]$. Thus, x_r is the right end point of the solution set and at this point r_{gd} is $(r_{gd}^{+} + r_{gd}^{-})/2 = -1/(2\left[n/2\right])$. The formula for case (1) comes from the average of x_r and x_l.

The details of cases (2) and (3) are done in a similar fashion and only the results are given. For case (2), at $\theta = (x_{(k)} + x_{(2k+1)})/2$, $r_{gd} = (r_{gd}^{+} + r_{gd}^{-})/2 = (((k+1)-k)+(k-k))/(2\left[n/2\right]) = 1/(2\left[n/2\right])$ at the left end of the solution set. At $\theta = (x_{(k+1)} + x_{(2k+2)})/2$, $r_{gd} = (r_{gd}^{+} + r_{gd}^{-})/2 = ((k-k)+(k-(k+1)))/(2\left[n/2\right]) = -1/(2\left[n/2\right])$ at the right end of the solution set. Again the average of x_r and x_l gives the case (2) result.

For case (3), if $\theta = (x_{(k+1)}+x_{(2k+2)})/2$, and $\lim_{c\to\theta^-} r_{gd}(e, |x^0-c1|) = r_{gd}^{+}(e, |x^0-\theta 1|) = 1/\left[n/2\right]$ and $\lim_{c\to\theta^+} r_{gd}(e, |x^0 - c1|) = r_{gd}^{-}(e, |x^0 - \theta 1|) = -1/[n/2]$, so that $r_{gd}^{+} = 1/\left[n/2\right]$ and $r_{gd}^{-} = -1/\left[n/2\right]$ which gives $r_{gd} = 0$ and, as stated, there is a unique point for the solution set. •

Theorem 3. If μ is the point of symmetry of F, the sampling distribution of $r_{gd}(e, |X^0 - \mu 1|)$ is distribution free and symmetric.

Proof: Let A be a class of distribution functions that are symmetric and absolutely continuous. Let μ_F represent the point of symmetry of distribution function $F \in A$, and let, as before, x^0 be the order statistics of a random sample of F. Then $r_{gd}(e, |X^0 - \mu_F 1|)$, is distribution-free within the class A. The proof of this depends on the transformation $X \to F(X) \equiv U$ where X is the notation for a random variable. First, $F(\mu_F) = 1/2$, $X - \mu_F$ is symmetric about zero and $F(X)$ is a $U(0,1)$ random variable. Clearly, $rank(x_{(i)} - \mu_F) = rank(F(x_{(i)}) - 1/2)$.

Thus, the set of ranks of $|x^0 - \mu_F 1|$ remain unchanged under the transformation to the ranks of $|u^0 - (1/2)1|$ where $u^0 = (F(x_{(1)}), F(x_{(2)}), \ldots, F(x_{(n)}))'$. Now, let $F_u(x) = x$ over the interval $(0,1)$ be the distribution function of a $U(0,1)$ random variable. F_u is in the class A. The random variable r_{gd} assumes only values $i/\left[n/2\right], \quad i = -\left[n/2\right], \ldots, \left[n/2\right]$.

Let $E_i = \{x|\ \ r_{gd}(e, |x^0 - \mu_F 1|) = i/\left[n/2\right]\}$. It will be shown that $P(r_{gd}(e, |X^0 - \mu_F 1|) = i/\left[n/2\right] = P(E_i|F \in A) = P(U_i|F_u)$ where U_i is contained in the n-fold region $(0,1) \times (0,1) \times \ldots (0,1)$ where a random sample of the random variable $F(X) = U(0,1)$ assumes the value $i/\left[n/2\right]$ for $r_{gd}(e, |U^0 - (1/2)1|)$.

Since $\{x|\ \ r_{gd}(e, |x^0 - \mu_F 1|) = i/\left[n/2\right]\} = \{u|\ \ r_{gd}(e, |u^0 - (1/2)1|) = i/\left[n/2\right]\}$, because of the equivalence of the ranks of the arguments, $P(E_i|F) = P(U_i|F_u)$. Thus the distribution of $r_{gd}(e, |X^0 - \mu_F 1|)$ is unchanged for all $F \in A$ and hence nonparametric distribution-free.

For the symmetry part, we use the equal in distribution technique that is summarized in Randles and Wolfe (1979). The distribution of $r_{gd}(e, |X^0 - \mu_F 1|)$ is symmetric about zero. It is true that $x - \mu_F 1 \stackrel{d}{=} \mu_F 1 - x \Rightarrow (x - \mu_F 1)^0 = x^0 - \mu_F 1 \stackrel{d}{=} (\mu_F 1 - x)^0 = \mu_F 1 - x^0 = -rev(x^0 - \mu_F 1)$. Therefore, $ranks(|x^0 - \mu_F 1|) \stackrel{d}{=} ranks(|-rev(x^0 - \mu_F 1)|) = ranks(|rev(x^0) - \mu_F 1|)$. Let the vector $M' = (m_1, m_2, \ldots, m_n) = ranks(|x^0 - \mu_F 1|)$ so that $r_{gd}(e, |x^0 - \mu_F 1|) = r_{gd}(e, M)$ and $r_{gd}(e, rev(M)) = -r_{gd}(e, M)$ by result (d) on page 658 in Gideon and Hollister (1987). In order to show the symmetry, Theorem 1.3.16 in Randles and Wolfe (1979) is utilized with $g(x^0 - \mu_F 1) = -rev(x^0 - \mu_F 1)$, $u((x - \mu_F 1)^0) = r_{gd}(e, |x^0 - \mu_F 1|)$ and $ranks(|rev(x^0) - \mu_F 1|) = rev(M)$. Thus, the conditions of Theorem 1.3.16 are satisfied and the distribution of $r_{gd}(e, |X^0 - \mu_F 1|)$ is symmetric about zero.•

5. Asymptotics and Relative Efficiencies

In this section the asymptotic relative efficiency of the r_{gd} location estimator relative to the mean, median, and a weighted quartile estimate are compared for four sampling distributions, first developing the large sample estimate for r_{gd}. The following notations are used: converges with probability one is $\stackrel{wp1}{\longrightarrow}$ and converges in distribution is $\stackrel{d}{\rightarrow}$.

Theorem 4. The limiting distribution of $\widehat{\theta}$ as $n \rightarrow \infty$, is $\widehat{\theta} \stackrel{wp1}{\longrightarrow} \frac{1}{2}(\widehat{\xi}_{1/3} + \widehat{\xi}_{2/3})$ where $\widehat{\xi}_p$ is the p^{th} quantile of the sample.

Proof: In the following, k_n represents the sequence of integer subscripts on the order statistics. Serfling (1980) gave conditions on k_n/n that are sufficient for the following asymptotic results. These conditions are shown for the case $k_n = \left[\frac{n+3}{3}\right]$. Substituting $n = 3k$, $3k+1$, and $3k+2$, in

$$4\widehat{\theta} = x_{([\frac{n+1}{3}])} + x_{([\frac{n+3}{3}])} + x_{([\frac{2(n+1)}{3}])} + x_{([\frac{2(n+2)}{3}])}$$

from Theorem 2 yields

1. $n = 3k, \quad \frac{[(3k+3)/3]}{n} = \frac{[k+1]}{3k} = \frac{k+1}{3k} = \frac{1}{3} + \frac{1}{n},$
2. $n = 3k+1, \quad \frac{[(3k+1+3)/3]}{n} = \frac{k+1}{3k+1} = \frac{1}{3} + \frac{2}{3n},$
3. $n = 3k+2, \quad \frac{[(3k+2+3)/3]}{n} = \frac{k+1}{3k+2} = \frac{1}{3} + \frac{1}{3n}.$

The case $k_n = \left[\frac{n+1}{3}\right]$ leads to $k_n/n = 1/3 + o(1/\sqrt{n})$. The cases $k_n = \left[\frac{2(n+1)}{3}\right]$ and $\left[\frac{2(n+2)}{3}\right]$ lead to $k_n/n = 2/3 + o(1/\sqrt{n})$. Thus, the results in Serfling (1980) can be used with $p = 1/3$ or $p = 2/3$.

The asymptotic variances in the cases $n = 3k$ and $n = 3k + 1$ lead to $\hat{\theta} = 0.25(2\widehat{\xi}_{1/3} + 2\widehat{\xi}_{2/3})$ which is the same as the case $n = 3k + 2$ where directly $\hat{\theta} = 0.5(\widehat{\xi}_{1/3} + \widehat{\xi}_{2/3})$. So the asymptotic variance of the estimate is developed.

The estimator $\hat{\theta}$ can be compared to other estimators in the sense of asymptotic relative efficiency using the criterion of the asymptotic variance in the normal approximation. It will be compared to the mean ($\bar{x}$), median ($\widehat{\xi}_{1/2}$), and quartile $\widehat{T} = (\widehat{\xi}_{1/4} + 2\widehat{\xi}_{1/2} + \widehat{\xi}_{3/4})/4$ for the symmetric distributions normal, uniform, Cauchy, and double exponential.

In the results to follow, the fact that $\hat{\theta}$ is asymptotically the same as $0.5(\widehat{\xi}_{1/3} + \widehat{\xi}_{2/3})$ is key.

The asymptotic normal distribution of $\hat{\theta}$ is given by Theorem B on page 80 of Serfling (1980). $(\widehat{\xi}_{1/3}, \widehat{\xi}_{2/3})$ is asymptotically $N\left(\begin{pmatrix}\xi_{1/3}\\ \xi_{2/3}\end{pmatrix}, \Sigma\right)$ where

$\Sigma = \frac{1}{9n}\begin{pmatrix} \frac{2}{f^2(\xi_{1/3})} & \frac{1}{f(\xi_{1/3})f(\xi_{2/3})} \\ \frac{1}{f(\xi_{1/3})f(\xi_{2/3})} & \frac{2}{f^2(\xi_{2/3})} \end{pmatrix}$ and $\xi_m = F^{-1}(m), m = 1/3$ or $2/3$.

From this result, the asymptotic distribution of $\hat{\theta}$ is

$$N\left(\frac{\xi_{1/3} + \xi_{2/3}}{2}, \frac{1}{18n}\left(\frac{1}{f^2(\xi_{1/3})} + \frac{1}{f(\xi_{1/3})f(\xi_{2/3})} + \frac{1}{f^2(\xi_{2/3})}\right)\right).\bullet$$

The specific results for the four distributions are given now for $\hat{\theta}$ and the relative efficiencies comparing other estimations follow.

(a) $N(\mu, \sigma^2)$: $\hat{\theta} \sim N(\mu, 1.2607\sigma^2/n)$.

(b) $U(\alpha, \beta)$: $\hat{\theta} \sim N((\alpha + \beta)/2, (\beta - \alpha)^2/(6n))$.

(c) Cauchy: $f(x) = \frac{1}{\pi(1+x^2)}, -\infty < x < +\infty$, $\hat{\theta} \sim N(0, 2.9243/n)$.

(d) double exponential: $f(x) = exp(-|x|/2), -\infty < x < +\infty$, $\hat{\theta} \sim N(0, 1.5000/n)$.

From the same theorem B of Serfling, the analogous results for the other three estimators can be derived, but listed here will be the asymptotic relative efficiencies, $are(-,-)$, where $-$ will be one of the four estimates of centrality.

Table 3: Asymptotic Relative Efficiencies

distribution	$are(\hat{\theta}, \bar{x})$	$are(\hat{\theta}, \widehat{\xi}_{1/2})$	$are(\hat{\theta}, T)$	$are(\widehat{\xi}_{1/2}, \bar{x})$	$are(T, \bar{x})$
$N(0,1)$	0.7932	1.2460	0.9476	0.6366	0.8370
$U(\alpha, \beta)$	0.5000	1.5000	0.9375	0.3333	0.5333
$Cauchy$	∞	0.8437	1.0546	∞	∞
$exp(-\|x\|/2)$	1.3333	0.6667	0.8333	1.9999	1.5999

For example, $are(\hat{\theta}, \widehat{\xi}_{1/2})$, the ratio of asymptotic variances of $\widehat{\xi}_{1/2}$ to $\hat{\theta}$ for the normal is 1.2460, showing that $\hat{\theta}$ is better than the median. It can be seen from

this listing that no one estimator is best. For the Cauchy entries, $\hat{\theta}$ is better than T and infinitely better than $\bar{x}$.

In summary, it has been shown how to use a correlation coefficient to produce a location estimator. Specific results were developed for Kendall's τ and GDCC for which the estimator depends essentially on the 1/3 and 2/3 quantiles, and hence, in terms of robustness, the "breakdown point" would be 1/3.

6. Estimation of Location after Scale

Using the results of previous work (Gideon and Rothan, 2010), the CES estimate of σ is used in (1) to estimate μ. This is done by estimating σ directly by the CES method with the original data and then the estimate of μ is the solution to equation (1) with the residuals, $x^0 - \hat{\sigma}q$, in place of x^0. The quantity q is the vector of appropriate theoretical quantiles and is elaborated below. This extended procedure has been tested with GDCC using computer simulations and seems to give an unbiased estimate of μ with a smaller variance than the direct method discussed in Section 4. Pearson's CC is also used to solve the location equation (1) and, in addition, to estimate μ after the CES estimate of σ. This is done to connect CES to classical estimates via Pearson's CC which uses least squares methods.

In Section 8 the CES location methods are used with censored data. It is especially easy to use in the type II censored data setting as defined in Gupta's paper (1952). Under normality assumptions this new method gives similar results to the examples given in Gupta, but for the exponential model the CES method gives a somewhat different result. The main results are based on computer simulations except for samples of sizes 2 and 3 and the unbiasedness proof.

In the censored data setting with any set of size h of known order statistics, σ and μ can be estimated in almost the same manner as for the full data set. Let x_h represent the vector of h known order statistics and k_h the corresponding vector of expectations of the standardized order statistics. Then the same procedure as for the full sample is done but with x_h and k_h. The estimate will remain unbiased and some examples are given later. The motivation will be done for a symmetric random variable although it is applicable to asymmetic distributions also.

This paragraph explains the CES two step estimate of μ by calculating the result for a perfect random sample (no error) or, in other words, what happens in the limit. Let $q_i = \Phi^{-1}(i/(n+1)), \quad i = 1, 2, \ldots, n$ where Φ is the cumulative distribution function of Z, a $N(0,1)$ random variable. Asymptotically q_i approaches $E(Z_{(i)})$, the expectation of the i^{th} order statistic of Z. So if F is the cdf of a $N(\mu, \sigma)$ random variable, then $F(x) = \Phi(\frac{x-\mu}{\sigma})$. Since $x_{(i)} = \mu + \sigma z_{(i)}$, $x_{(i)} = \mu + \sigma q_i + \epsilon_i$ and this relationship is used to estimate first σ and then μ. The CES scale equation for the normal distribution is $r(q, x^0 - sq) = 0$, where q is the ordered vector of q_i and s is the estimate of σ. If $\epsilon_i = 0$ for all i, then $x_{(i)} = \mu + \sigma q_i$ and the scale equation becomes $r(q, \mu 1 + \sigma q - sq) = 0$. For $s = \sigma$, the equation is $r(q, \mu 1) = 0$, (by the min-max tied value procedure) so σ is the solution. Then the location equation (1) is $r(e, |x^0 - \theta 1|) = 0$ and so $r(e, |\mu 1 + \sigma q - \theta 1|) = 0$. Now it is shown that if $\theta = \mu$, this equation is satisfied. Continuing, $r(e, |\sigma q|) = r(e, \sigma |q|) = r(e, |q|)$. Geometrically, the graph of e and $|q|$ looks like a perfectly symmetric U-shape centered at zero on the horizontal axis.

Two cases are taken, $n = 10$ and $n = 11$, to show the pattern that gives the zero solution using GDCC and the CES tied value procedure. For $n = 10$ the ranks of

$|q|$ are 9.5, 7.5, 5.5, 3.5, 1.5, 1.5, 3.5, 5.5, 7.5, 9.5. The ranks of $|q_1|$ and $|q_{10}|$ are tied at 9 and 10, the ranks of $|q_2|$ and $|q_9|$ are tied at 7 and 8, and so forth until the ranks of $|q_5|$ and $|q_6|$ are tied at 1 and 2. See Tables 4 and 5, where the min-max procedure converts the ranks of $|q|$ into unique vectors.

To compute r_{gd} for $r(e, |q|)$, one must compute r_{gd}^+ and r_{gd}^-, by using the permutations of $|q|$ which give the most positive and most negative correlations. To this end a table is constructed. The correlation data are in columns 1 and 2, most positive, and columns 6 and 7, most negative. Column 4 is $(n + 1 - col2)$ and column 9 is $(n+1-col7)$. As an example, the 6th row elements of columns 3, 5, 8, 10 are 2, 3, 2, 3 because: column 2 has two numbers greater than 6 that are at or above row 6; column 4 has three numbers greater than 6 that are at or above row 6; column 7 has two numbers greater than 6 that are at or above row 6; column 9 has three numbers greater than 6 that are at or above row 6. The other rows follow in the same counting manner. The denominator is the greatest integer in $n/2$, or $[10/2] = [11/2] = 5$.

Table 4: Tied Value Procedure for GDCC, $n = 10$

			r_{gd}^+					r_{gd}^-		
cols	1	2	3	4	5	6	7	8	9	10
	1	9	1	2	1	1	10	1	1	0
	2	7	2	4	1	2	8	2	3	1
	3	5	3	6	2	3	6	3	5	1
	4	3	3	8	2	4	4	3	7	2
	5	1	2	10	3	5	2	3	9	2
	6	2	2	9	3	6	1	2	10	3
	7	4	1	7	3	7	3	2	8	3
	8	6	1	5	2	8	5	1	6	2
	9	8	0	3	1	9	7	1	4	1
	10	10	0	1	0	10	9	0	2	0
	max		3		3			3		3

Consequently, $r_{gd}^+ = (3-3)/5 = 0$ and $r_{gd}^- = (3-3)/5 = 0$ so $r_{gd} = (r_{gd}^+ + r_{gd}^-)/2 = 0$. The pattern is clear so that for all even n, $r_{gd}(e, |q|) = 0$

Take $n = 11$ for an odd sample size example. The ranks of $|q|$ are tied in pairs with a middle observation at zero having rank 1. The calculation of r_{gd}^+ and r_{gd}^- in Table 5 should be clear from the $n = 10$ case.

Table 5: Tied Value Procedure for GDCC, $n = 11$

			r^+_{gd}					r^-_{gd}		
cols	1	2	3	4	5	6	7	8	9	10
	1	10	1	2	1	1	11	1	1	0
	2	8	2	4	1	2	9	2	3	1
	3	6	3	6	2	3	7	3	5	1
	4	4	3	8	2	4	5	4	7	2
	5	2	3	10	3	5	3	3	9	2
	6	1	2	11	3	6	1	3	11	3
	7	3	2	9	4	7	2	2	10	3
	8	5	1	7	3	8	4	2	8	3
	9	7	1	5	2	9	6	1	6	2
	10	9	1	3	1	10	8	1	4	1
	11	11	0	1	0	11	10	0	2	0
	max		3		4			4		3

Consequently, $r^+_{gd} = (4-3)/5 = 1/5$ and $r^-_{gd} = (3-4)/5 = -1/5$ so $r_{gd} = (r^+_{gd} + r^-_{gd})/2 = (1/5 + (-1/5))/2 = 0$. The pattern is clear so that for all odd n, $r_{gd}(e, |q|) = 0$.

When there is random variation, $|\epsilon_i| > 0$ for all i, the scale equation is solved for s and then the location equation, $r(e, |(x - sq)^0 - \theta 1|) = 0$, is solved for θ.

In general, the location estimate on the data with non-zero error gives good results as shown in Table 6. The CES method is a general procedure and any CC can be used and any F as long as the expectations of the standardized order statistics or their approximations, as herein, are available.

Table 6: Averages of 500 Random Samples

	Estimation Method (no outliers)					
	Classical	Pces	Pces2	GDCC	GDCC2	true value
μ	5.019	5.019	5.020	5.033	5.027	5
$\sigma/\sqrt{n}$	0.432	0.445	0.436	0.463	0.431	$3/7 = 0.429$
σ	2.979	3.088	—	2.988	—	3
	Estimation Method (outliers)					
μ	4.828	4.749	4.929	4.930	4.910	5
$\sigma/\sqrt{n}$	0.527	0.633	0.808	0.493	0.471	$3/7 = 0.429$
σ	3.605	3.464	—	3.283	—	3

As an illustration, a sample size of 49 was used to demonstrate typical results. The normal distribution was used with mean 5 and standard deviation 3 and some runs had random outliers (with mean 3, SD 7) introduced for 5 of the 49 observations. Pearson's CC was used with the CES method, equation (1), to compare it to classical results. GDCC was used to demonstrate its robust properties and to show how close it is to classical methods. Thus, there are two runs on Pearson's CC — with and without outliers — and likewise for GDCC. One hundred random samples were run several times with summary statistics to find differences or trends. Once trends were identified, in order to increase accuracy, five hundred random samples

produced the results in Table 6. Five means were computed: the classical mean, the Pces mean (the solution to equation (1) with Pearson's CC), Pces2 (Pces after σ was estimated), the CES mean with GDCC, and the CES mean with GDCC after σ was estimated (labeled with 2 in Table 6). An alternate way to describe the means labeled with a 2 is the CES mean of $(x^0 - \hat{\sigma}q)$ where $\hat{\sigma}$ is the CES estimate of σ. Note that if the classical mean of $(x^0 - \hat{\sigma}q)$ is taken, one gets $\overline{x}$ because q is symmetric about zero.

The estimate of the standard deviation of $\bar{x}$, $\sigma/\sqrt{n}$, from 500 random samples is 0.432 and the true value is $3/7 \cong 0.429$. From classical distribution theory the standard deviation of this estimate is 0.0116 which is greater than $(0.432 - 0.429) = 0.003$. Thus, the simulations are producing the expected results, and so the GDCC2 estimate, 0.431, is remarkable. This result has been verified with numerous simulations so the GDCC location estimate based on the $1/3-2/3$ quantile average is as good as the classical $\bar{x}$ estimate using normal data. In the second half of Table 6 the classical $\bar{x}$ using five possible outliers has a theoretical standard deviation of 0.517 and the observed sample standard deviation, 0.527, with 500 random samples, is within a standard deviation. Note however that both the GDCC (0.498) and GDCC2 (0.471) estimates are much smaller. This, of course, is due to the robustness of GDCC. Also note that for pure normal data, Pces and classical results are roughly equivalent. With respect to the mean, Pces2 has some robustness (the average of the μ estimate is 4.929) but the variation is significantly larger, 0.808.

The two-step procedure for a location estimate can also be used on the median to improve its variability without biasing it for symmetric distributions. GDCC is again used to estimate σ and then the median is taken on the residuals, $x^0 - \hat{\sigma}q$. Several sets of simulations again were used to determine the results and then a set of 500 simulations gave the values to follow. For the normal data the asymptotic efficiency of the mean to the median is $2/\pi$, the variance ratio; the standard deviation ratio is $\sqrt{2/\pi} \cong 0.7979$. For the same setup using $N(5, 3^2)$ and n=49, for 500 simulations the standard deviation ratio of the regular median was $0.4129/0.5244 = 0.7833$, fairly close to 0.7979. However, the mean of the medians of the residuals gave an average of 5.045 and $\hat{\sigma}_{mean}/\hat{\sigma}_{median} = 0.4129/0.4276 = 0.9656$. Thus, the two-step procedure makes the median almost as efficient as the mean.

Just as in the GDCC location estimate, the two-step process is robust for the median. For the same sample setup with five outliers as described above, the mean of the 500 random samples was 4.7723 and the mean of the medians was 4.8680 while the mean of the medians of the residuals was 4.8574. The corresponding standard deviations were 0.5329, 0.5709, and 0.4970. So the standard deviation ratios are $0.5329/0.5709 = 0.9334$ and $0.5327/0.4970 = 1.0722$. This means the median in terms of variability is superior.

Other papers such as Gideon (2012), have already demonstrated the strength of CES in regression analysis.

7. Small Sample Results for $\hat{\theta}$

Consider two equations:

(a) $r_{gd}(k, x^0 - sk) = 0$ and

(b) $r_{gd}(e, |(x^0 - sk)^0 - \theta 1|) = 0$.

To estimate μ after σ, solve equation (a) for s. Denote this estimate of σ by $GDCC(s)$. Now solve equation (b) for θ using this value of s. Denote this estimate

of μ by $GDCC(\theta|s)$. Finally, denote by $GDCC(\theta|x)$ the estimate of μ using the raw data.

For $n = 2$ it can be shown that $GDCC(s) = \frac{x_{(2)}-x_{(1)}}{k_2-k_1}$ and that $GDCC(\theta|x) = GDCC(\theta|s) = \frac{x_{(2)}+x_{(1)}}{2}$. Because $k_i = E(Z_{(i)})$ and $X_{(i)} = \mu + \sigma Z_{(i)}$, the unbiasedness of each estimator for its respective parameter is easy to prove:

$E(GDCC(s)) = E\left(\frac{X_{(2)}-X_{(1)}}{k_2-k_1}\right) = \sigma$ and $E\left(\frac{X_{(1)}+X_{(2)}}{2}\right) = E(\overline{X}) = \mu$.

For $n = 3$ and symmetric F, it can be shown that $GDCC(s) = 0.5\left(\frac{x_{(3)}-x_{(1)}}{k_3-k_1} + \frac{x_{(2)}-x_{(1)}}{k_2-k_1}\right) = \frac{x_{(3)}+2x_{(2)}-3x_{(1)}}{4k_3}$, and $GDCC(\theta|x) = \frac{x_{(1)}+2x_{(2)}+x_{(3)}}{4}$, because $k_2 = 0$ and $k_3 = -k_1$.

For $GDCC(\theta|s)$ the residual vector must be examined for its six possible orderings in detail and if F is symmetric about μ, which implies symmetry in the k_i (e.g. $k_1 = -k_3$), then, after some tedious work and letting $s = GDCC(s)$,
$GDCC(\theta|s) = \frac{\left((x_{(2)}-k_2 s)+2(x_{(1)}-k_1 s)+(x_{(3)}-k_3 s)\right)}{4} = \frac{x_{(2)}+2x_{(1)}+x_{(3)}}{4} - \frac{s(k_2+2k_1+k_3)}{4} = \frac{x_{(2)}+2x_{(1)}+x_{(3)}}{4} + \frac{k_3 s}{4} = \frac{1}{16}(5x_{(1)} + 6x_{(2)} + 5x_{(3)})$.

It is again easy to show that all estimates are unbiased. It can also be shown that if F is a $U(0,1)$ distribution function, $Var(GDCC(\theta|x)) = 0.0313$ which is more than $Var(GDCC(\theta|s)) = 0.0285$. Hence, the residual method of estimating μ after σ has the smaller variance. Although no closed forms have been found for the cases $n > 3$, computer simulations in the next section and Table 6 confirm that the residual method is best. It is conjectured that this would also be the case for other nonparametric correlation coefficients. After showing the unbiasedness of the residual method, the next section gives the results of some computer simulations and several examples.

8. Proof of the Unbiasedness of the Residual (Two-Step) Method

This work uses the equal in distribution technique that is summarized in Randles and Wolfe (1979), employing Theorem 1.3.2 which states that random variable X is symmetric about a point μ if and only if $X - \mu \stackrel{d}{=} \mu - X$, and Theorem 1.3.7 which states that if $X \stackrel{d}{=} Y$ and U is a measureable function defined on the common support of X and Y then $U(X) \stackrel{d}{=} U(Y)$. In what follows, these are applied for random sampling and order statistics.

Let $(X_1, X_2, \ldots, X_n)$ be a random sample of X where $X - \mu \stackrel{d}{=} \mu - X$. Then $(X_1 - \mu, X_2 - \mu, \ldots, X_n - \mu) \stackrel{d}{=} (\mu - X_1, \mu - X_2, \ldots, \mu - X_n)$, and $(X_{(1)} - \mu, X_{(2)} - \mu, \ldots, X_{(n)} - \mu) \stackrel{d}{=} (\mu - X_{(n)}, \mu - X_{(n-1)}, \ldots, \mu - X_{(1)})$. This latter equal in distribution statement implies that $X_{(i)} - \mu \stackrel{d}{=} \mu - X_{(n+1-i)}$. We first apply these statements to show that the r_{gd} estimate of location is unbiased for μ. Using r_{gd} and X in equation (1), $r_{gd}(e, |x^0 - \theta 1|) = 0$, for $n = 3k$, k a positive integer, gives the solution $\widehat{\theta} = 0.25(x_{(k)} + x_{(k+1)} + x_{(2k)} + x_{(2k+1)})$. Now $k + (2k+1) = n+1$ and $k + 1 + 2k = n + 1$ so that $X_{(k)} - \mu \stackrel{d}{=} \mu - X_{(2k+1)}$ and $X_{(k+1)} - \mu \stackrel{d}{=} \mu - X_{(2k)}$. We take expectations to obtain $E(X_{(k)}) + E(X_{(2k+1)}) = E(X_{(k+1)}) + E(X_{(2k)}) = 2\mu$. Therefore $E(\hat{\theta}) = \frac{4\mu}{4} = \mu$.

For the residual method of estimation, start with the statement about the centered order statistics being equal in distribution and apply the same operation to

both sides; i.e. subtract $E(Z_{(i)})s = k_i s$ from the corresponding terms on both sides to obtain

$$(X_{(1)} - \mu - k_1 s, X_{(2)} - \mu - k_2 s, \ldots, X_{(n)} - \mu - k_n s) \stackrel{d}{=}$$
$$(\mu - X_{(n)} - k_1 s, \mu - X_{(n-1)} - k_2 s, \ldots, \mu - X_{(1)} - k_n s)$$

but by the symmetry of the distribution, $k_i = -k_{n+1-i}, \quad i = 1, 2, \ldots, n$ so that the right hand side becomes $(\mu - (X_{(n)} - k_n s), \mu - (X_{(n-1)} - k_{n-1} s), \ldots, \mu - (X_{(1)} - k_1 s))$.

Now the residuals are $V_i = X_{(i)} - k_i s$ so that $(V_1 - \mu, V_2 - \mu, \ldots, V_n - \mu) \stackrel{d}{=} (\mu - V_n, \mu - V_{n-1}, \ldots, \mu - V_1)$ and so $V_i - \mu \stackrel{d}{=} \mu - V_{n+1-i}, \quad i = 1, 2, \ldots, n$. Thus, $E(V_i - \mu) = E(\mu - V_{n+1-i}), \quad i = 1, 2, \ldots, n$ or $E(V_i) + E(V_{n+1-i}) = 2\mu, \quad i = 1, 2, \ldots, n$. Now for $i = k, n + 1 - i = 2k + 1$ and for $i = k + 1, n + 1 - i = 2k$. Thus, $E(V_{(k)} + V_{(k+1)} + V_{(2k)} + V_{(2k+1)}) = 4\mu$. So solving equation (1) with V for the case $n = 3k$ again gives $E(\hat{\theta}) = \frac{E(V_{(k)} + V_{(k+1)} + V_{(2k)} + V_{(2k+1)})}{4} = \mu$.

For all simulation examples, $n = 20$ and the standardized variables are $N(0, 1)$ for the normal case, and $U\left(-\frac{\sqrt{12}}{2}, \frac{\sqrt{12}}{2}\right)$ for the uniform case. Simultaneous estimation for μ and σ is done for the full data set and for a reduced data set with the number of remaining points being $h = 6, 11, 15$. Table 7 gives the results for both the normal and uniform distributions. Row 1 gives the average value of the $GDCC$ estimate of μ (no residuals used) for the available data set, $h = 6, 11, 15, 20$. The sample standard deviation appears in parentheses to the right of each average. Row 2 gives the average value of $GDCC(\theta|s)$, the residual method for estimating μ. Row 3 is the same as row 2 except the true σ is used and not the estimate s. Note that, as expected, the standard deviation is less for row 3 than for row 2.

Table 7: Estimation of μ and σ with the GDCC Method for Full and Reduced Data Sets

sampling	distribution	$N(30, 25)$		
h	6	11	15	20
$GDCC(\theta\|x)$	24.53(1.50)	26.70(1.35)	28.15(1.11)	29.76(1.14)
$GDCC(\theta\|s)$	29.74(2.09)	29.67(1.22)	29.74(1.05)	29.69(1.01)
$GDCC(\theta\|\sigma)$	29.89(1.43)	29.83(1.15)	29.76(1.06)	29.68(1.01)
sampling	distribution	$U(5, 15)$		
h	6	11	15	20
$GDCC(\theta\|x)$	6.70(0.77)	8.09(0.87)	8.99(0.87)	10.21(0.84)
$GDCC(\theta\|s)$	10.10(1.63)	10.15(1.13)	10.17(0.79)	10.20(0.62)
$GDCC(\theta\|\sigma)$	10.22(0.75)	10.21(0.78)	10.22(0.71)	10.17(0.59)

Each table entry is the mean of 50 samples followed by the sample SD of these samples. An important point to observe from this table is that when all the data are used all methods are unbiased, but the residual method has the smaller variation. Both distributions confirm the same thing and although not shown, other computer runs substantiated this.

For $h < 20$ it is clear that as the sample gets increasingly truncated, the ordinary nonresidual method is biased whereas the residual method remains unbiased. As expected, as h decreases the standard deviation of the simultaneous estimate of

μ increases. Rows 2 and 3 indicate that knowing σ helps in the estimation of μ by decreasing the standard deviation. Note that this exactly parallels the classical estimation of μ and σ but with the the order of estimation reversed.

In order to show the generality of this censored data estimation technique, a run was made of 50 samples of $N(30, 25)$ in which $h = 11$ but the censored data was spread throughout the $n = 20$ original observations. The GDCC residual method gave an average of 30.13 with a standard deviation of 1.26. This is very close to the value in Table 7 for the normal case with $h = 11$.

9. Gupta's (1952) Normal and Exponential Examples

Gupta (1952) has two normal distribution examples which we will compare to the CES method with GDCC. In his first example only 119 of 300 observations are available on the lifetime of electric lamps. Grouped data are used and the midpoints of the intervals are used as the data. Two GDCC comparisons were used: (1) midpoints are used for the data so that there are many tied values and (2) each data point is made unique by randomly selecting the correct number of data points within each grouped set. Gupta denotes his estimates by stars and his notation is retained. The following two tables give his result and the corresponding GDCC estimates.

Table 8: Gupta's First Example

statistic	Gupta	GDCC random	GDCC midpoint
μ^*	1502	1507	1498
σ^*	202	212	200

It is apparent from the table that the tie breaking procedure of evaluating GDCC allows this procedure to work on data that has a large number of ties. It is also apparent that in this example the robust GDCC method is very similar to Gupta's result.

In Gupta's second example only the first 7 of 10 number of days to death after an inoculation are recorded for some mice in an experiment. Because it is a small data set, Gupta uses three of his methods, as labeled below in Table 9; he also used the log_{10} transformation on the number of days to death to achieve better normality. The retransformed estimate appears in parentheses (number of days).

Table 9: Gupta's Second Example

	Gupta's			CES
statistic	best linear	alternative linear	MLE	GDCC
μ^*	1.746(55.7)	1.784(56.0)	1.742(55.2)	1.751(56.3)
σ^*	0.101	0.094	0.072	0.100

Again it is clear from this table that the robust GDCC procedure gives comparable results. It should be noted that, when a different sampling distribution is assumed, if the expected values of a standardized random variable are available,

then the GDCC method can be used to obtain estimates whereas Gupta's method is only for the normal distribution. Our next example is exponential.

10. Censored Data Estimation for the Exponential Distribution

The two parameter exponential distribution with density function $f(x) = \frac{1}{\sigma} exp\left(\frac{x-\theta}{\sigma}\right), \quad x > \theta$ is used. The standard exponential is obtained by taking $\theta = 0, \sigma = 1$. The parameter θ can be thought of as a time delay until the exponential variate starts. The overall expectation is $\mu = \theta + \sigma$. Since σ estimates both a location and a scale parameter, the simulation results are presented differently than the earlier simulation tables. Simulations were run separately for $h = 6, 11, 15$ with $\theta = 0$ and $\sigma = 10$. The quantity h is the number of observations remaining after censoring the $n - h$ right-most observations. For each run, the censored estimation ($h < 20$) and the full sample ($n = h = 20$) estimation are done and given in one of the columns. In each case, the number in parentheses is the sample standard deviation for the 50 simulations. Row 1 of Table 10 gives the average of the GDCC location estimator which for the exponential would be the average of the 1/3 and 2/3 quantiles of the exponential ($\theta = 0, \sigma = 10$) which is 7.52. The 2-step method estimates σ and then θ, and since $\theta = 0$, row 2 gives the average of 50 GDCC location estimates for the residuals ($x^0 - sq$) and this should be close to zero, because a study of the distribution of these residuals shows it to be nearly symmetric. Since s also estimates location, its average from the estimates in row 2 are given in row 3; the averages are reasonably close to the true value of 10. Row 5 gives the average of 50 GDCC location estimates for the censored data; they behave as expected, increasing as the number of observaions increases. Row 6 is like row 2 except that the censored data are used; note that the table entries are close to zero ($\theta = 0$). Row 7 is like row 3 except the censored data are used and it gives the average of s for 50 simulations that were used in row 6; again the values are close to 10. Rows 4 and 8 give the GDCC estimate of μ for the full data set (row 4) and the censored data (row 8). In Gupta's second example, the overall estimate would come from a single estimate as in row 8. These results clearly show that GDCC and in general the CES 2-step method of estimation should be further studied because of its potential value in improving estimators.

Table 10: Simultaneous Estimation of θ, σ, and $\mu = \theta + \sigma$ for an Exponential Variate $x > \theta, \theta = 0, \sigma = 10$, with $n = 20$

Full	1. GDCC Loc Est	7.49(2.18)	7.74(2.24)	7.31(1.96)
Data	2. GDCC2 Loc Est	0.014(0.92)	0.31(1.11)	-0.11(1.08)
	3. Mean of s	9.62(2.73)	9.24(2.33)	9.26(2.70)
	4. Mean of $\hat{\theta} + s$	9.63(2.62)	9.54(2.45)	9.25(2.25)
			h	
		6	11	15
Cen-	5. GDCC Loc Est	1.78(0.81)	3.63(1.46)	4.80(1.32)
sored	6. GDCC2 Loc Est	0.036(0.85)	0.26(0.95)	-0.014(0.93)
Data	7. Mean of s	9.20(4.31)	9.28(3.67)	9.26(2.90)
	8. Mean of $\hat{\theta} + s$	9.23(3.91)	9.54(3.46)	9.24(2.50)

If all the data had been shifted, $\theta > 0$, then the estimate of θ would be estimating the point at which the exponential process starts. In Gupta's second example, it is easy to obtain the GDCC estimate assuming an exponential model. Assuming the exponential model above $\hat{\theta} = 38.86$ and $s = 23.86$ so that $\hat{\mu} = \hat{\theta} + s = 62.45$. Note that this estimate is higher than the estimate given by the log_{10} model assumption, and hence, the choice of the model is probably more important than the elaborate error analysis of any particular model. A simplistic Kolmogorov test of fit was performed on the data by randomly completing the sample size to $n = 10$ under the assumed model and using the estimated parameters of the null hypothesis. The test statistic was 0.1756 for the exponential model and 0.1301 for the log_{10} transformation. Neither of these is close to any significance and so the choice of the model would be subjective. In this example, the choice of the model involves more than just an evaluation of tests of fit because deciding $\theta > 0$ in the exponential model is assuming an exponentially delayed action characteristic.

11. Conclusion

The solution of equation (1) gives remarkable location estimates. Only a few correlation coefficients — Kendall, GDCC, and Pearson — have been examined using equation (1) and all provided very good location estimators. The most surprising result may be seen in Table 6, where the two-step CES method of estimating μ with GDCC is just as good as $\bar{x}$ when the data are normal, but with just a few possible outliers, is better. Also the GDCC two-step method reduces the variability of the median and is robust so that this method should be used in all statistical analysis. Because it is distribution free, the GDCC location estimator is valid and surprisingly accurate for the Cauchy distribution. Sections 8 and 9 show that there are many possible worthwhile extensions of CES methods into a wide variety of areas. Future work should formulate a theoretical reason for why the residual or two-step method improves the estimation. This location work extends to the location estimation part of regression analysis. CES results in regression are more general than least squares and CES methods in variation show they are just as accurate for truly normal data while they are considerably better when there is any deviation from normality.

12. References

Gideon, R.A. (2012). Obtaining Estimators from Correlation Coefficients: The Correlation Estimation System and R, *Journal of Data Science*, **10**, 597-617.

Gideon, R.A. and Hollister, R.A. (1987). A Rank Correlation Coefficient Resistant to Outliers, *Journal of the American Statistical Association*, **82**, 656-666.

Gideon, R.A. and Rothan, A.M., CSJ (2010). Location and Scale Estimation with Correlation Coefficients, *Communications in Statistics-Theory and Methods*, **40**, 1561-1572.

Gupta, A.K. (1952). Estimation of the Mean and Standard Deviation of a Normal Population from a Censored Sample, *Biometrika*, **39**, 260-273.

Hettsmansperger, T.P. (1984). Statistical Inference Based on Ranks, *John Wiley & Sons*, N.Y.

Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics, *John Wiley & Sons*, N.Y.