Les Cahiers du GERAD ISSN: 0711–2440

Confidence Intervals for a Discrete Population Median

D. Larocque
R.H. Randles

G–2007–13

March 2007

Confidence Intervals for a Discrete Population Median

Denis Larocque
GERAD and Department of Management Sciences
HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada, H3T 2A7
denis.larocque@hec.ca

Ronald H. Randles
Department of Statistics
University of Florida
P.O. Box 118545
Gainesville, Florida 32611–8545, U.S.A.
rrandles@stat.ufl.edu

March 2007

Les Cahiers du GERAD

G–2007–13

Abstract

In this paper, we consider the problem of constructing confidence intervals for a population median when the underlying population is discrete. We describe seven methods of assigning confidence levels to order statistic based confidence intervals, all of which are easy to implement. An extensive simulation study shows that, with discrete populations, it is possible to obtain consistently more accurate confidence levels and shorter intervals compared to the ones reported by the classical method which is implemented in commercial softwares. More precisely, the best results are obtained by inverting a two-tailed sign test that properly takes into account tied observations. Some real data examples illustrate the use of these confidence intervals.

Key Words: Sign test; Tied observations; Discrete distribution; Multinomial distribution; Confidence level; Maximum likelihood.

Résumé

Dans cet article, nous considérons le problème de la construction d'intervalles de confiance pour une médiane lorsque la population sous-jacente est discrète. Nous décrivons sept méthodes pour assigner un niveau de confiance à un intervalle formé par des statistiques d'ordre. Ces méthodes sont toutes faciles à implémenter. Une étude par simulation démontre que, pour des populations discrètes, il est possible d'obtenir des niveaux de confiance plus précis et des intervalles plus courts comparativement à ceux obtenus des méthodes classiques qui sont implémentées dans les logiciels usuels. Plus précisément, les meilleurs résultats sont obtenus en inversant un test du signe bidirectionnel qui traite adéquatement les ex-æquo. Des exemples avec de vraies données illustrent l'utilisation des intervalles de confiance.

1 Introduction

Estimation of a population median (M) is an important topic, particularly when the underlying population can not be assumed to be symmetric. The sample median ($\hat{M}$) is a natural point estimator of M. It minimizes the sum of the distances to the data points. It also has other desirable properties, including being median unbiased when the underlying population is continuous and asymptotically efficient when the population is double exponential. Order statistic based confidence intervals of the form:

$$[X_{(d)}, X_{(n+1-d)}] \tag{1}$$

for some integer $1 \leq d \leq n/2$ are also used to estimate M, when $X_1, X_2, \ldots, X_n$ denotes a random sample of size n from the underlying population and $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ denote the order statistics of the sample.

When the population is continuous, the interval in (1) has a confidence level

$$1 - \alpha = 1 - 2P[B \leq d - 1], \tag{2}$$

where B is a binomial$(n, \frac{1}{2})$ random variable. This confidence interval has a distribution-free property because

$$P[X_{(d)} \leq M \leq X_{(n+1-d)}]$$

is equal to (2) for any continuous population. The expression (2) results because the confidence interval (1) contains the values M that would not be rejected by the two-tailed α level sign test. An advantage of this interval is that it requires only a binomial$(n, \frac{1}{2})$ distribution to find $1 - \alpha$. However, the choices of $1 - \alpha$ are constrained by the discrete nature of the binomial$(n, \frac{1}{2})$ distribution. To enable construction of, for example, a 95% confidence interval for any sample size n, Hettmansperger and Sheather (1986) proposed an interval created by interpolating between $[X_{(d)}, X_{(n+1-d)}]$ and $[X_{(d+1)}, X_{(n-d)}]$. The resulting interval is now only approximately distribution-free, but it performs well over a variety of continuous distributions. Their method is implemented in MINITAB. A number of authors have described interpolated intervals for use with continuous populations. See Papadatos (1995), Hutson (1999) and Ho and Lee (2005a, 2005b) for recent contributions and additional references. The use of a closed interval in (1) comes from an important result of Scheffé and Tukey (1945) showing that for a closed interval the probability that M is included in the interval is greater than or equal to the expression in (2) even when the population is discrete. Therefore, the confidence level described in (2) extends, possibly conservatively, to any population.

In many applications the underlying population is discrete and takes only a finite (or countably infinite) number of values. For example, Ferner, Coleman, Pirmohamed, Constable and Rouse (2005) proposed a scale for measuring whether the SPC's (Summary of

Product Characteristics) which come with a nonhaematological drug provide adequate instruction to enable pharmacists to monitor haematologically adverse drug reactions. They called their measurement scale a SIM (Systematic Instructions for Monitoring) score. Five clinicians recorded SIM scores for each of 84 SPC's of nonhaematological drugs. The median score among the five for each of the 84 SPC's are displayed in the upper plot of Figure 5. They are naturally integer valued, because SIM scores are integer valued. One of the researchers' objectives was to estimate a population median SIM score. The population is discrete, in fact, integer valued. Thus the population median will be an integer.

This paper addresses the problem of finding a confidence interval for a population median when the underlying population is discrete. For simplicity of description, assume that the population is integer valued. The translation to other discrete settings will follow readily. Since it is a rare discrete population for which there is an integer k satisfying $P(X_1 \leq k) = 1/2$, we will define the population median M as the unique integer such that $P(X_1 \leq M - 1) < 1/2$ and $P(X_1 \leq M) > 1/2$. In discrete population settings it is inappropriate to interpolate between possible values. The median must be one of the possible values. Without particular knowledge about the population, it is still natural to estimate M with the sample median $\hat{M}$. It will be one of the possible values (possibly two when n is an even integer). It is also appropriate to form a confidence interval for M using a closed interval like (1) with order statistics as the endpoints. Scheffé and Tukey's (1945) result shows that this interval is appropriate in discrete cases, though the coverage probability stated in (2) may underestimate the actual coverage probability of the interval. Very little research has been directed toward this setting. Scheffé and Tukey (1945), Emerson and Simon (1979) and Huang (1991) are papers that consider the estimation of M in discrete population settings.

Section 2 discusses the assignment of approximate confidence levels to closed intervals of the form in (1) for discrete populations. Seven different methods of assigning confidence levels are described. The performances of these methods are compared via simulation studies in Section 3. Examples and conclusions are given in Section 4. Throughout the paper the focus is on finding intervals with a confidence level of at least 95%. This is intended to simplify the presentation. The adaptation to other target confidence levels easily follows.

2 Assigning Confidence Levels

In continuous population settings, confidence intervals are commonly constructed by specifying a desired confidence level and then constructing a data based interval that will contain the desired parameter with probability equal to the desired confidence level. In discrete population settings, particularly those involving a broad class of discrete population models, this construction plan can not be followed. A confidence interval (such as described in (1)) is , of course, still data based. But in addition, the level of confidence associated with that interval is also a function of the observed data. The number of values tied at the endpoints or at adjacent possible values will alter the level of confidence that should be associated with the interval. Because of the discrete nature of this setting there are thus only a finite number of associated confidence levels that are possible. If a target level of confidence is given, such as .95, then the interval is chosen which has the smallest associated confidence level exceeding or equal to .95. What follows in this section are descriptions of seven methods of assigning a level of confidence to an interval of the form in (1).

Method 1: Based on the Noether (1967) result, the expression in (2) is clearly a legitimate assignment of a confidence level to interval (1). This is implemented in MINITAB, for example. When $n > 50$, MINITAB approximates the binomial probability with a normal distribution.

Method 2: SAS considers intervals like (1) with confidence levels specified by (2) and compares them with slightly asymmetric (in the order statistics) intervals $[X_{(d+1)}, X_{(n+1-d)}]$ with associated confidence level of $1-\alpha = 1 - P[B \leq d-1] - P[B \leq d]$, where B is binomial$(n, \frac{1}{2})$. The chosen interval is the one among all of these that has the smallest associated confidence level that exceeds or equals 0.95. The level associated with the chosen interval is reported as the confidence level of the interval.

Method 3: Motivated by the descriptions in Scheffé and Tukey (1945) and Noether (1967), consider an interval that is determined symmetrically, but with an assigned confidence level that takes into account the tied values. Consider the interval $[X_{(d)}, X_{(n+1-d)}]$ Let r be the smallest integer for which $X_{(r)} = X_{(d)}$. Let s be the largest integer for which $X_{(s)} = X_{(n+1-d)}$. The interval will be $[X_{(d)}, X_{(n+1-d)}]$, but the confidence level attached to it will be $1 - \alpha = 1 - P[B \leq r-1] - P[B \leq n-s]$, where B is binomial$(n, \frac{1}{2})$. The integer d is chosen to produce the smallest confidence level that exceeds or equals .95. The confidence levels assigned by this method are those associated with the simultaneous inversion of two one-tailed sign tests for the respective directional alternatives.

Methods of assigning a confidence level associated with inverting versions of a two-tailed sign test are considered next. When using an interval $[X_{(d)}, X_{(n+1-d)}]$ and the p-value of a two-tailed test, a confidence level is determined by

$$1 - \alpha = 1 - \max\{\text{p-value}(X_{(d)}-), \text{p-value}(X_{(n+1-d)}+)\}, \quad (3)$$

where p-value(c), denotes the p-value of a two-tailed test of H_0: $M = c$. Here $X_{(d)}-$ denotes the first possible population value below $X_{(d)}$ and $X_{(n+1-d)}+$ denotes the first possible population value above $X_{(n+1-d)}$.

Method 4: When the underlying population is continuous, the first population value below $X_{(d)}$ can be thought to be $X_{(d)} - \epsilon$ where ϵ is a very small positive quantity, such that $X_{(d-1)} < X_{(d)} - \epsilon$. Likewise, the first possible value above $X_{(n+1-d)}$ is considered to be $X_{(n+1-d)} + \epsilon < X_{(n+2-d)}$. The p-value of the sign test for continuous data can be used, namely

$$p - value(c) = 2P[B \leq min\{n_+^c, n_-^c\}], \tag{4}$$

where $n_+^c =$ (number of $X_i's > c$) and $n_-^c =$ (number of $X_i's < c$). The confidence interval $[X_{(d)}, X_{(n+1-d)}]$ is then given a confidence level through (3).

When the underlying population is discrete on the integers the first possible value below $X_{(d)}$ is $X_{(d)} - 1$. There may be zero, one or more observations on $X_{(d)} - 1$. Likewise, the first possible value above $X_{(n+1-d)}$ is $X_{(n+1-d)} + 1$ and it also may or may not be an observed value. So in a discrete population setting (3) becomes:

$$1 - \alpha = 1 - \max(\text{p-value}(X_{(d)} - 1), \text{p-value}(X_{(n+1-d)} + 1)). \tag{5}$$

Consequently, these confidence levels require using p-values of versions of the two-tailed sign test that properly account for the possibility of zeros. When basing p-value(c) on a sign test of H_0: $M = c$ versus H_a: $M \neq c$, let:

$$n_+^c = (\text{number of } X_i's > c), \;\; n_0^c = (\text{number of } X_i's = c) \;\text{ and }\; n_-^c = (\text{number of } X_i's < c).$$

It is possible that $n_0^c > 0$.

The literature on the use of zeros in the sign test is extensive. Coakley and Heise (1996) provide a review of this topic from a testing perspective. Their investigation explored a null hypothesis that the probability of a positive equals the probability of a negative. This makes the number of zeros irrelevant. However, ignoring the number of zeros is not appropriate when the hypothesis concerns the population median. Recent papers by Randles (2001) and Fong, Kwan, Lam and Lam (2003) describe problem settings which are relevant to the population median.

Consider two-tailed sign tests that reject for large values of

$$n_*^c = \max(n_+^c, n_-^c).$$

A p-value could be obtained for such a test via:

$$\text{p-value}(c) = P[N_* \geq n_*^c | \tilde{p}_+, \tilde{p}_0, \tilde{p}_-], \tag{6}$$

where $N_* = max(N_+, N_-)$ and (N_+, N_0, N_-) have a multinomial distribution with parameters n and $(\tilde{p}_+, \tilde{p}_0, \tilde{p}_-)$ satisfying:

$$0 \leq \tilde{p}_+ \leq 1/2 \quad \text{and} \quad 0 \leq \tilde{p}_- \leq 1/2, \tag{7}$$

which is the null hypothesis condition. We suggest three methods of finding appropriate values for $(\tilde{p}_+, \tilde{p}_0, \tilde{p}_-)$.

Method 5 – Maximum Likelihood: Assuming the null hypothesis condition (7), the maximum likelihood estimators of the probabilities are:

$$\tilde{p}_{+mle} = \frac{n_+^c}{n}, \quad \tilde{p}_{0mle} = \frac{n_0^c}{n}, \quad \tilde{p}_{-mle} = \frac{n_-^c}{n},$$

when $n_*^c \leq \frac{n}{2}$, and

$$\tilde{p}_{+mle} = \frac{1}{2}, \quad \tilde{p}_{0mle} = \frac{n_0^c}{2(n - n_+^c)}, \quad \tilde{p}_{-mle} = \frac{n_-^c}{2(n - n_+^c)}, \tag{8}$$

when $n_+^c > \frac{n}{2}$. In this case, it is seen that the excess of n_+^c over $n/2$ is distributed multiplicatively between $\tilde{p}_{0mle}$ and $\tilde{p}_{-mle}$. Likewise, when $n_-^c > \frac{n}{2}$, the estimates look like (8) with the roles of $\tilde{p}_{+mle}$ and $\tilde{p}_{-mle}$ reversed and the roles of n_+^c and n_-^c reversed. The MLE estimates were used in Fong et al. (2003), but there was a slight error in their description in that paper. The estimates given by (8) are substituted into (6) which, in turn, produces a confidence level using (5).

Method 6 – Constrained Quadratic Loss: A second method finds the $(\tilde{p}_+, \tilde{p}_0, \tilde{p}_-)$ that minimizes:

$$\left[(\frac{n_+^c}{n} - p_+)^2 + (\frac{n_0^c}{n} - p_0)^2 + (\frac{n_-^c}{n} - p_-)^2\right], \tag{9}$$

under the null hypothesis condition $0 \leq \tilde{p}_+ \leq 1/2$ and $0 \leq \tilde{p}_- \leq 1/2$. The minimizing probabilities are:

$$\tilde{p}_{+cql} = \frac{n_+^c}{n}, \quad \tilde{p}_{0cql} = \frac{n_0^c}{n}, \quad \tilde{p}_{-cql} = \frac{n_-^c}{n},$$

when $n_*^c \leq \frac{n}{2}$, and

$$\tilde{p}_{+cql} = \frac{1}{2}, \quad \tilde{p}_{0cql} = \frac{n_0^c}{n} + \frac{1}{2}(\frac{n_+^c}{n} - \frac{1}{2}), \quad \tilde{p}_{-cql} = \frac{n_-^c}{n} + \frac{1}{2}(\frac{n_+^c}{n} - \frac{1}{2}), \tag{10}$$

when $n_+^c > \frac{n}{2}$. Note that in contrast with the MLE estimates, the excess of n_+^c over $n/2$ is distributed equally and additively between $\tilde{p}_{0cql}$ and $\tilde{p}_{-cql}$. Likewise, when $n_-^c > \frac{n}{2}$, the estimates look like (10) with the roles of $\tilde{p}_{+cql}$ and $\tilde{p}_{-cql}$ reversed and the roles of n_+^c and n_-^c reversed. The estimates given by (10) are substituted into (6) which, in turn, produces a confidence level using (5).

When $n_0^c = 0$, the Maximum Likelihood Method 5 uses $(\tilde{p}_{+mle}, \tilde{p}_{0mle}, \tilde{p}_{-mle}) = (\frac{1}{2}, 0, \frac{1}{2})$ and the MLE p-value produced by (6) is that of the continuous data two-tailed sign test. The Constrained Quadratic Loss Method 6 does not have this property. In the next section it is shown that this creates some undesirable performance characteristics when the data values are sparse (few ties are observed). The following method is constructed to have improved performance in sparse settings.

Method 7 – Modified Constrained Quadratic Loss: Use $(\tilde{p}_+, \tilde{p}_0, \tilde{p}_-) = (\tilde{p}_{+cql}, \tilde{p}_{0cql}, \tilde{p}_{-cql})$ as described in Method 6 when $n_0^c > 0$ and $(\tilde{p}_+, \tilde{p}_0, \tilde{p}_-) = (\frac{1}{2}, 0, \frac{1}{2})$ when $n_0^c = 0$. The p-value(c) is then described by (6) and the confidence level by (5).

3 Simulation study

An extensive simulation study was performed in order to compare the seven methods for constructing confidence intervals defined in the preceding section. The investigation used the Poisson(λ) and the negative binomial(n,p) distributions. For the Poisson distribution, all integer values of the parameter λ between 1 and 40 were considered. For the negative binomial distribution, twelve different parametrizations were used where (n, p)=(number of successes, probability of success) with parameter set values (1, 0.1), (1, 0.2), (1, 0.3), (1, 0.4), (2, 0.1), (2, 0.2), (2, 0.3), (2, 0.4), (3, 0.1), (3, 0.2), (3, 0.3), and (3, 0.4).

For each of the above distributions, all sample sizes between 15 and 40 were considered. Consequently, there were 1040 configurations for the Poisson distribution (40 df $\times$ 26 sample sizes) and 312 (12 parametrizations $\times$ 26 sample size) for the negative binomial. These configurations cover a large spectrum of nearly symmetric (Poisson) and skewed (negative binomial) distributions and also a broad spectrum of concentrated versus more sparse distributions.

As explained in the last section, conservative 95% confidence intervals were constructed by choosing the smallest interval that has an associated confidence level of at least 95%. For each of the 1352 configurations, the quantities of interest were estimated by generating 5000 samples. The computations were performed with the Ox language version 3.4; Doornik (2002).

We will examine three crucial aspects of a confidence interval:

1. Is the reported confidence level accurate? That is, is the difference between the reported confidence and the true coverage small?
2. Does the interval maintain the desired coverage probability (which is 95%)?
3. Is the interval short?

Basically, we are looking for short confidence intervals that maintain the desired coverage probability and which have associated confidence levels that are accurate. The results for some specific configurations will be presented later, but first the overall picture of the performances for both distributions are described in Tables 1 and 2.

The results for the Poisson distribution are summarized in Table 1. The first part of the table provides the averages (over the 1040 configurations) of the true coverage (empirical coverage of the 5000 samples) minus the average (over the 5000 samples) reported confidence level. The minimums and maximums (over the 1040 configurations) are also presented. We see that Method 6 is the most accurate. On average, the confidence reported by this method underestimates the true coverage by only 0.40%. It is followed by the two other methods based on two-tailed sign tests for tied observations (Method 7 and Method 5) which underestimate the true coverage by 0.67% and 0.81% respectively. Method 1 and Method 2 perform poorly compared to the others.

The second part of the table provides another view of the accuracy of the methods. The seven methods were ranked according to the absolute value of true coverage minus reported confidence levels. The table reports the average ranks (over the 1040 configurations) and also the minimum and maximum ranks. Method 6 has the smallest average rank (1.38) among the seven methods. As before, Method 7 and Method 5 come in second and third place respectively. Once again, Methods 1 and 2 perform very poorly. Indeed, Method 1 was only able to achieve a fifth place during all the 1040 configurations.

The third part of the table gives averages (over the 1040 configurations) of the true coverage (empirical coverage of the 5000 samples). It also reports the minimum and maximum true coverages (over the 1040 configurations). All methods maintained the desired coverage probability (95%) for all configurations. Indeed, all minimums are above 95%.

The last part of the table provides a comparison of the confidence interval lengths across methods. Inside and for each of the 1040 configurations, the seven methods were ranked according to the average length of the confidence interval (over the 5000 samples). The table then reports the average ranks (over the 1040 configurations) and also the minimum and maximum ranks. Method 6 is again the best one since its average rank (1.57) was the smallest among the seven methods. To summarize, Methods 6, 7 and 5 are clearly the best ones for the Poisson distributions considered.

Table 2 presents the same information as Table 1 but summarizing the 312 configurations of the negative binomial distribution. Method 6 is again the most accurate (first part of the table) but with this distribution it overestimates the true coverage by 0.21% on average. Method 7 is almost as accurate by underestimating the true coverage by 0.28%. Methods 7, 3 and 5 are the best ones when we rank the accuracies within configurations (second part of the table). The third part of the table shows that, on average, all seven methods maintain the desired coverage probability over the 312 configurations. But Method 6 can have a true coverage that falls slightly below 95% as indicated by its minimum value of 93.66%. The other six methods always maintain the desired coverage probability since their minimums are all above 95%. Finally, Methods 6, 7 and 5 are respectively in first, second and third place according to the average lengths of their confidence intervals. To summarize, if a conservative confidence interval is desired (i.e., if it is not acceptable that the true coverage could be below 95%), then Method 7 would be the method of choice for the negative binomial distributions considered.

Table 1: Overall results of the simulation study for the Poisson distribution (1040 configurations of df and sample size)

Difference between reported confidence and true coverage (in %)			
Method	Minimum	Maximum	Average
6	-1.16	1.96	0.40
7	-0.35	1.96	0.67
5	-0.05	1.96	0.81
3	0.15	1.79	0.93
4	0.52	2.73	1.37
1	0.71	4.86	2.22
2	1.12	4.86	2.78
Intra-configuration ranking of absolute difference between reported confidence and true coverage			
Method	Minimum	Maximum	Average
6	1.00	6.00	1.38
7	1.00	5.00	2.09
5	1.00	5.00	3.21
3	1.00	5.00	3.68
4	1.00	5.00	4.65
1	5.00	6.50	6.31
2	6.50	7.00	6.69
True coverage (in %)			
Method	Minimum	Maximum	Average
6	95.12	100.00	97.97
7	96.34	100.00	98.27
5	96.88	100.00	98.44
3	97.16	100.00	98.60
2	97.02	100.00	98.93
4	97.16	100.00	99.10
1	97.16	100.00	99.10
Intra-configuration ranking of confidence interval length			
Method	Minimum	Maximum	Average
6	1.00	4.00	1.57
7	1.50	4.50	2.58
5	2.00	4.50	2.96
3	3.00	5.00	4.00
2	2.00	6.00	5.13
4	4.00	6.50	5.88
1	4.00	6.50	5.88

Figures 1, 2 and 3 present some typical results for specific distributions. For these figures, the plot in the upper left corner gives the probability mass function of the underlying distribution. Only support points with a probability greater than 0.01 are included in the

Table 2: Overall results of the simulation study for the negative binomial distribution (312 configurations of parameters and sample size)

Difference between reported confidence and true coverage (in %)			
Method	Minimum	Maximum	Average
6	-2.19	1.81	-0.21
7	-1.40	1.85	0.28
5	-1.31	1.76	0.43
3	-0.95	1.60	0.51
4	0.17	2.75	1.07
1	0.24	4.66	1.86
2	0.64	4.70	2.31
Intra-configuration ranking of absolute difference between reported confidence and true coverage			
Method	Minimum	Maximum	Average
7	1.00	6.00	2.36
3	1.00	5.00	2.68
5	1.00	5.00	2.84
6	1.00	7.00	3.14
4	1.00	5.00	4.24
1	3.00	6.50	6.15
2	4.50	7.00	6.59
True coverage (in %)			
Method	Minimum	Maximum	Average
6	93.66	99.96	97.32
7	95.24	99.96	97.86
5	95.96	99.96	98.03
3	96.28	99.96	98.15
2	96.28	100.00	98.46
4	96.28	100.00	98.74
1	96.28	100.00	98.74
Intra-configuration ranking of confidence interval length			
Method	Minimum	Maximum	Average
6	1.00	4.00	1.55
7	1.50	4.50	2.73
5	2.00	4.50	3.14
3	2.00	5.50	3.75
2	2.00	6.00	5.06
4	4.00	6.50	5.88
1	4.00	6.50	5.88

plot to help visualize the shape of the distribution and give an idea about the number of non-negligible support points. The plot in the upper right corner gives the true coverage (empirical coverage of the 5000 samples) minus the average (over the 5000 samples) reported confidence levels as a function of the sample size. A positive value indicates that the reported confidence under-estimates (on average) the true coverage. Ideally, this value should be close to 0 in absolute value. The plot in the lower left corner gives the true coverage as a function of the sample size. Since we are constructing conservative 95% confidence intervals, these values should ideally be above (but close) to 95%. Finally, the plot in the lower right corner gives the average length of the interval as a function of the sample size. In order to make the figures easier to read, only Methods 1, 5, 6 and 7 are displayed. To get an idea about the other methods, we can say that Method 3 is close to Method 5, that Method 2 is even worse than Method 1 and that Method 4 lies somewhere between Methods 1 and 5.

The results for the Poisson(20) distribution are depicted in Figure 1. Method 6 is clearly the best method for this distribution followed closely by Method 7. Indeed, looking at plot b), we see that the reported confidence of Method 6 is closer to the true coverage for almost all sample sizes. Moreover, all methods are under-estimating the true coverage except for four sample sizes for the Method 6. We also see in plot c) that the true coverage is always greater than the target 95% but Method 6 is closer to the target. Finally, plot d) shows that on average, Method 6 produces shorter intervals.

If the accuracy of the reported confidence is the important criterion, an example where Method 6 is not the best one is the negative binomial(2, 0.1) distribution reported in Figure 2. Subplot b) shows that Method 6 over-estimates the true coverage. Methods 5 and 7 are better at reporting accurate confidence levels. This type of situation was the motivation for the introduction of Method 7. However, the true coverage of the Method 6 is closer to 95% and its intervals are shorter as seen in plots c) and d).

The negative binomial(3, 0.3) depicted in Figure 3 shows that Method 6 can be the best one for a skewed distribution. As with the Poisson(20) distribution, Method 6 over-estimates the true coverage, but the reported confidence levels are more accurate than those of the other six methods. The true coverage of Method 6 is also closer to the target and its intervals are shorter.

A potential problem with Method 6 is that it does not adjust to the sparseness of the distribution. Even for a continuous distribution, the p-value is computed by using a positive value for $\tilde{p}_0$. By defining $\tilde{p}_0 = 0$ when $n_0 = 0$, Method 7 corrects this, thereby adapting to the sparseness of the distribution.

Figure 4 illustrates the benefits of doing that by showing the results of a simulation study for the sparse normal distribution. A sparse normal random variable X with parameter (round factor) c is generated in the following way:

$$X = \text{round}(cZ)$$

where Z is a standard normal variate. All integer values between 1 and 40 were considered for c. With $c = 1$ the distribution is concentrated on a few values and the distribution

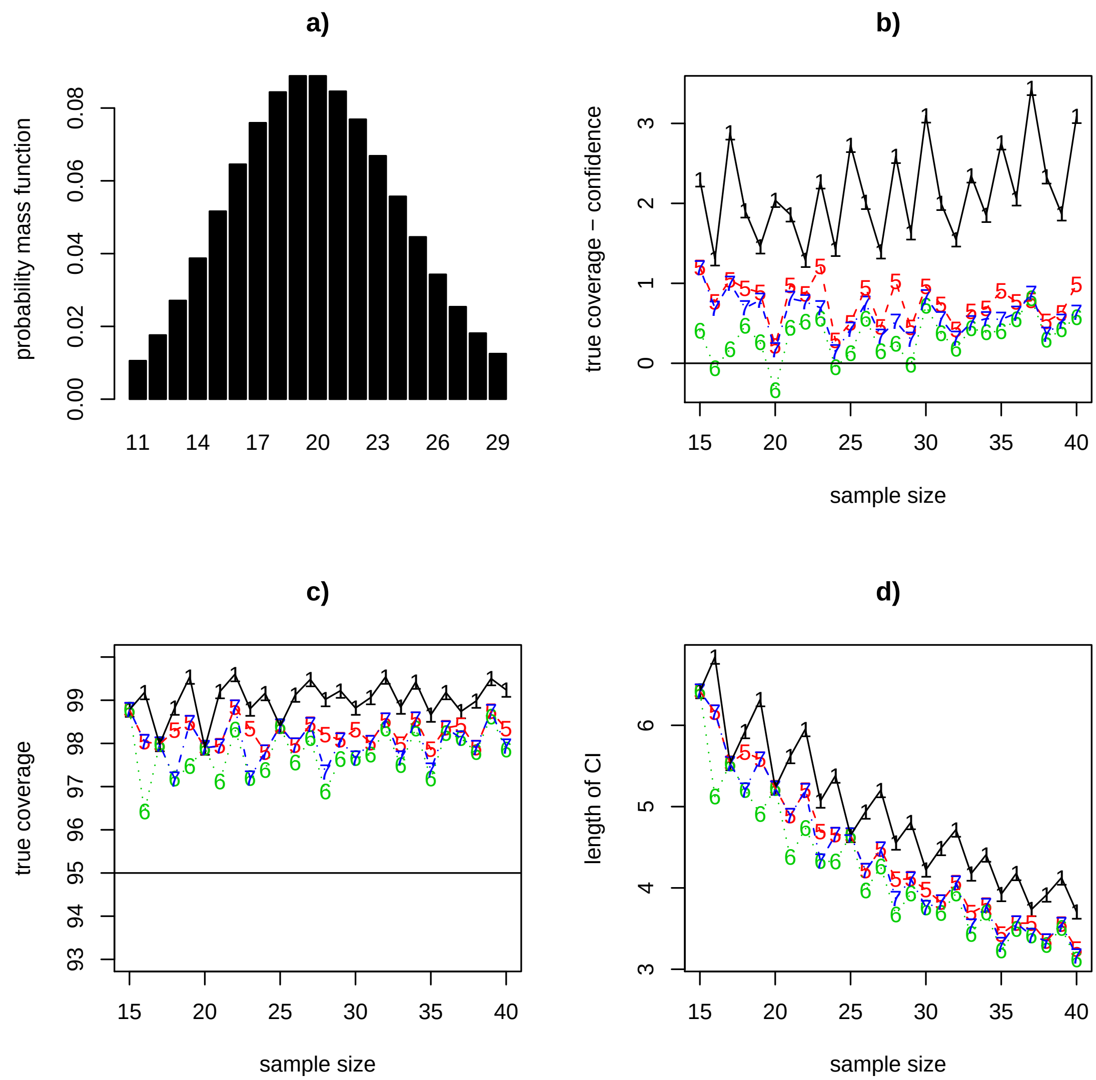

Figure 1: Simulation results for the construction of the conservative 95% confidence interval for the Poisson(20) distribution. The lines are numbered according to the methods they are representing. Plot a) is the probability mass function of the distribution. Plot b) is the empirical true coverage minus the average reported confidence as a function of the sample size. Plot c) is the empirical true coverage. Plot d) is the average length of the confidence interval.

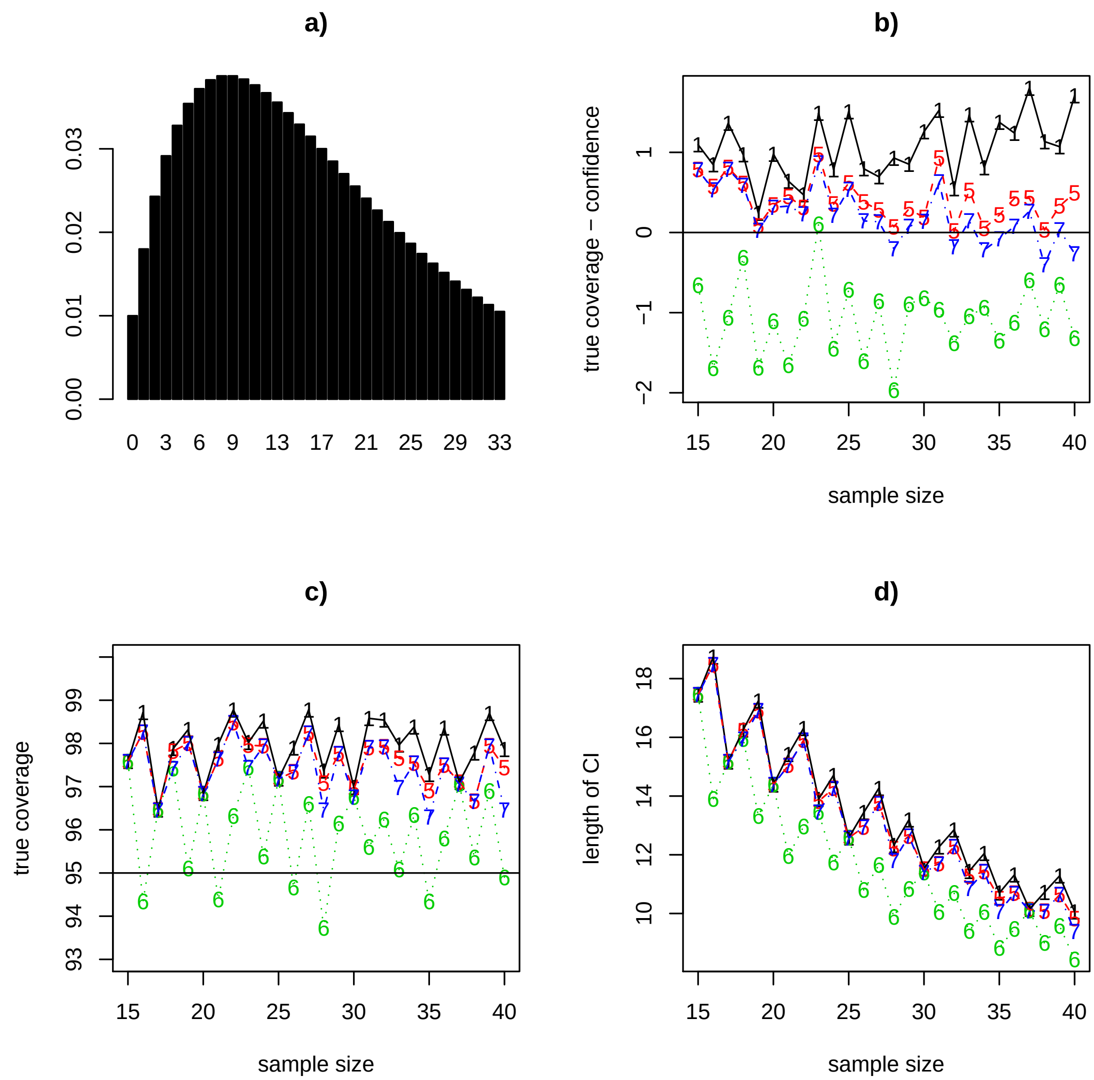

Figure 2: Simulation results for the construction of the conservative 95% confidence interval for the negative binomial(2, 0.1) distribution. The lines are numbered according to the methods they are representing. Plot a) is the probability mass function of the distribution. Plot b) is the empirical true coverage minus the average reported confidence as a function of the sample size. Plot c) is the empirical true coverage. Plot d) is the average length of the confidence interval.

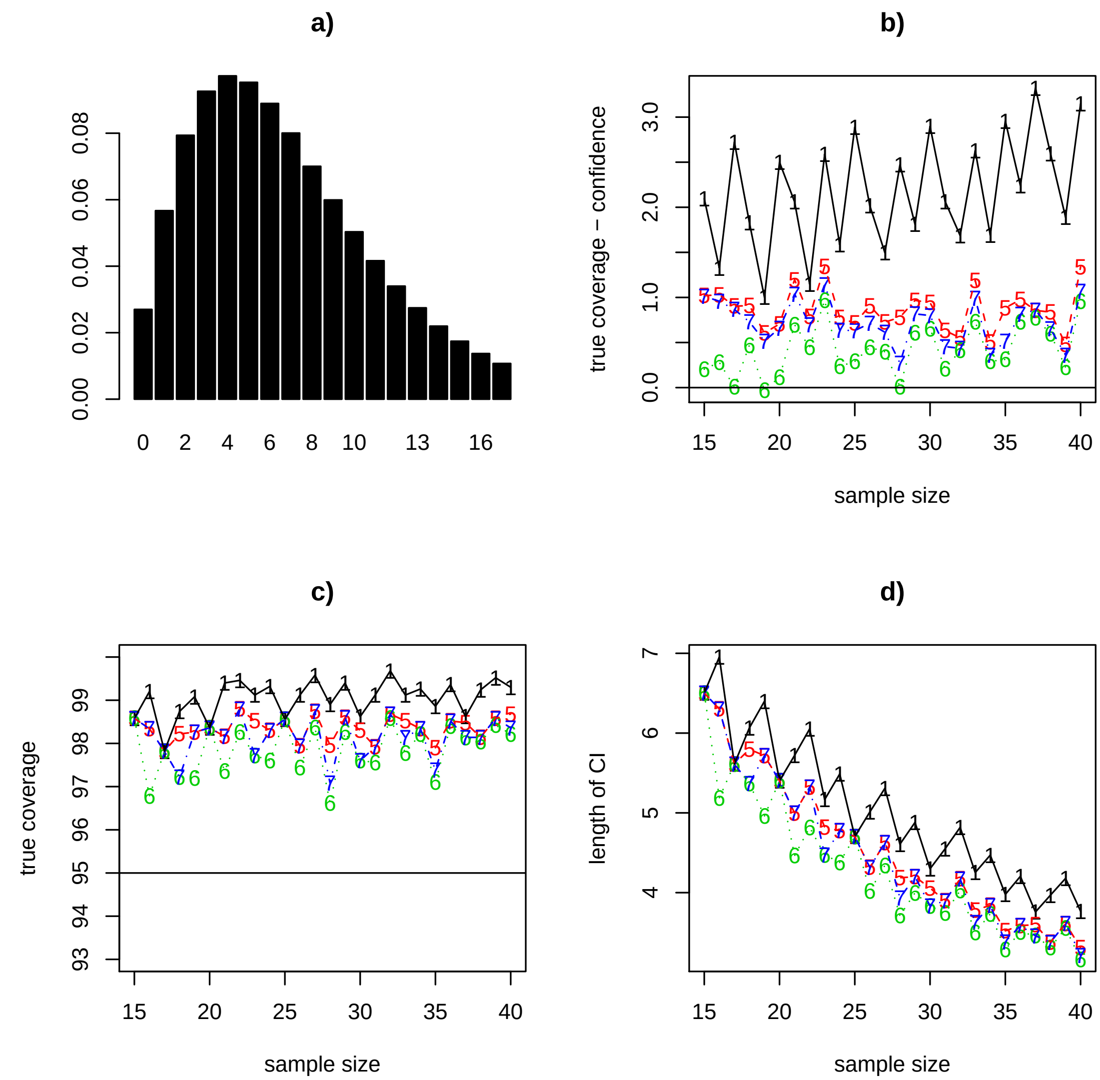

Figure 3: Simulation results for the construction of the conservative 95% confidence interval for the negative binomial(3, 0.3) distribution. The lines are numbered according to the methods they are representing. Plot a) is the probability mass function of the distribution. Plot b) is the empirical true coverage minus the average reported confidence as a function of the sample size. Plot c) is the empirical true coverage. Plot d) is the average length of the confidence interval.

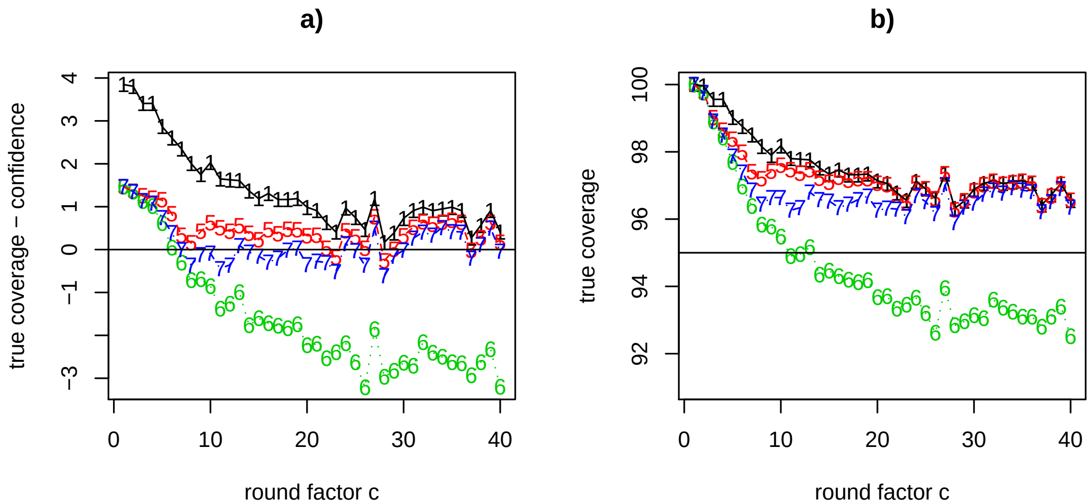

Figure 4: Simulation results for the construction of the conservative 95% confidence interval for the sparse normal(c) distribution with a sample size of 40. The lines are numbered according to the methods they are representing. Plot a) is the empirical true coverage minus the average reported confidence as a function of the round factor c. Plot b) is the empirical true coverage.

becomes more sparse as c increases. With $c = 40$, we practically have a continuous normal distribution with standard deviation 40.

Figure 4 displays the results for the sparse normal distribution as a function of the round factor c when the sample size is 40. As c increases, i.e. as the distribution becomes more and more continuous, Method 6 overestimates the true coverage and its true coverage falls below 95%. Method 7 adjusts to the sparseness of the distribution, provides accurate confidence levels and has a true coverage which is always above 95%. Moreover, as c increases, all methods (except Method 6) become indistinguishable. It's clear that in practice, the analyst could probably judge the appropriateness of using Method 6 depending on the observed data but Method 7 provides a convenient automatic adjustment.

4 Example and concluding remarks

To illustrate the practical use of the confidence intervals, we first return to the SIM scores example described in the Section 1. The left part of Table 3 present the confidence intervals obtained by all seven methods along with their reported confidence levels. As in the rest of the paper, the goal was to construct intervals with a coverage of at least 95%. We see that six methods produced the interval $[10, 18]$. Only Method 6 was able to produce the shorter interval $[11, 16]$ with a reported confidence of 96.63%. Among the six other intervals, Methods 5 and 7 report the highest confidence levels while Methods 1 and 2 report the lowest. Evidence gathered from the simulation study make us believe that the larger values are probably better estimates of the actual probability of coverage.

The middle and lower plots of Figure 5 display two classic data sets. The middle plot displays counts of ticks on 82 sheeps as first reported by Fisher (1941) and included by Hand, Daly, Lunn, McConway and Ostrowski (1994). The lower plot in Figure 5 displays reading scores for 116 persons who dropped out of a Job Corp program. They were reported and analyzed by Taylor (1972) and appear in the text by Daniels (1990). For these two data sets, the middle and right parts of Table 3 show that Methods 3, 5, 6 and 7 produced the shortest intervals $[10, 18]$ and $[4, 5]$ respectively. Among these four methods, Methods 6 and 7 report the highest and probably the most accurate confidence levels.

Based on all the findings from the simulation study and the data examples, it is clear that Methods 1 and 2 (which are respectively implemented in MINITAB and SAS) can be improved when dealing with a discrete population. The three methods based on versions of the sign tests that can handle zeros (Methods 5, 6 and 7) gave the best results. Moreover, these methods do not pose complicated computational challenges compared to the currently implemented methods, as they are simply based on the trinomial distribution instead of the binomial distribution. Our recommendation is that, with a discrete population, Methods 6 and 7 should be the methods of choice. Method 7 could be viewed as a good all around and conservative method while Method 6 should be used more cautiously especially for distributions with many non-negligible support points since it does not perform as well as Method 7 does when the underlying distribution is sparse. But if used appropriately, Method 6 can provide a very accurate analysis in many circumstances as shown in the simulation study and, indeed, it is the method that generally produces the shortest confidence intervals.

Table 3: Confidence interval results for the three data sets

	SIM scores (n=84) Sample median = 14			Ticks on sheep (n=82) Sample median = 5			Reading scores (n=116) Sample median = 14		
Method	lower	upper	confidence	lower	upper	confidence	lower	upper	confidence
1	10	18	96.25	4	6	96.48	11	16	96.77
2	10	18	96.25	4	6	95.25	11	16	95.85
3	10	18	98.84	4	5	96.02	12	16	95.14
4	10	18	98.84	4	6	98.02	11	16	98.01
5	10	18	99.41	4	5	96.70	12	16	96.05
6	11	16	96.63	4	5	96.99	12	16	96.13
7	10	18	99.42	4	5	96.99	12	16	96.13

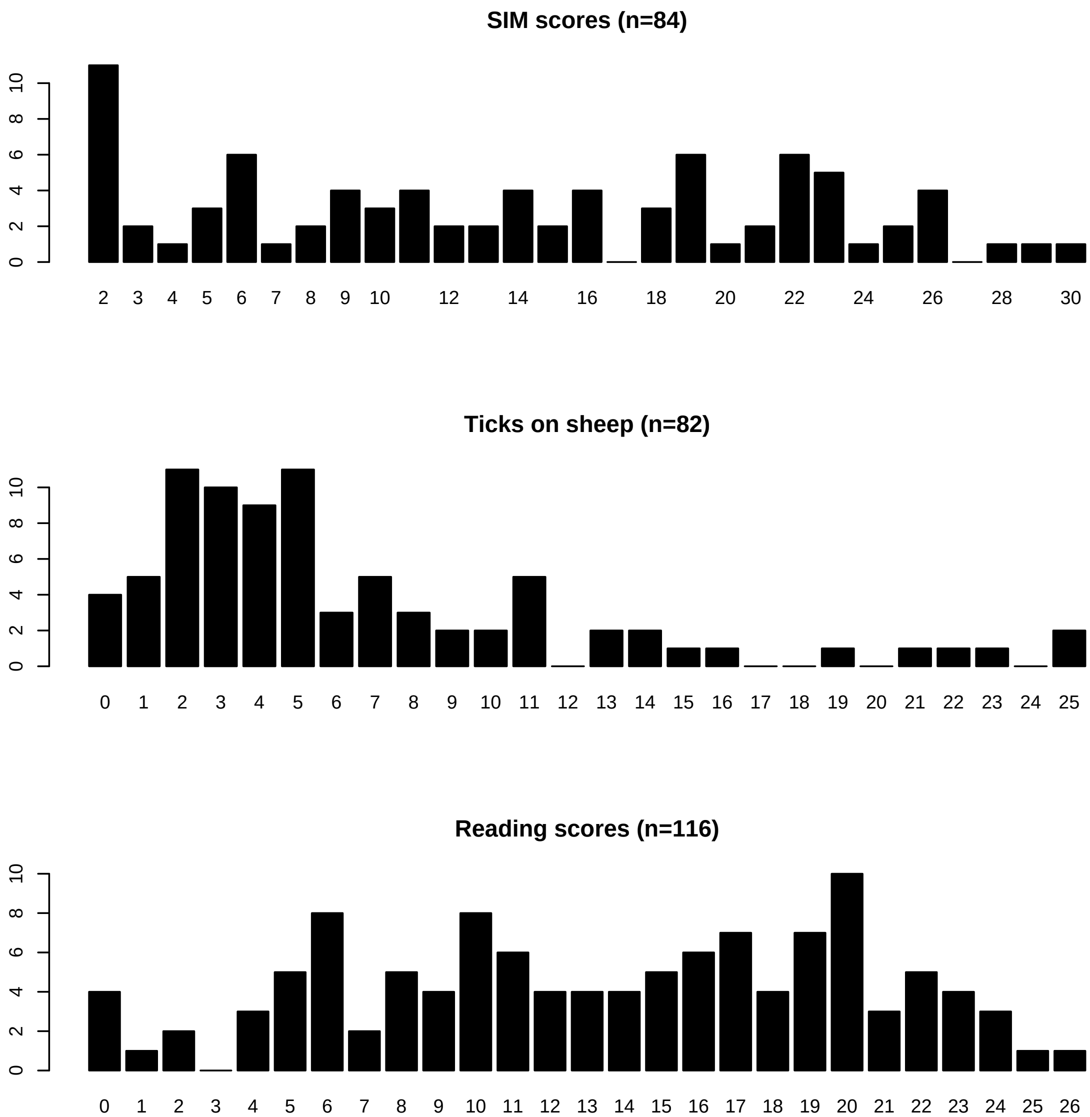

Figure 5: The three data sets used for illustration: SIM scores (upper plot), ticks on sheep (middle plot) and reading scores (lower plot).

References

Coakley, C.W. and Heise, M.A. (1996), "Versions of the Sign Test in the Presence of Ties," *Biometrics*, 52, 1242–1251.

Daniels, W.W. (1990), *Applied Nonparametric Statistics*, 2nd Ed., Boston, MA: Duxbury.

Doornik, J. A. (2002). *Object-Oriented Matrix Programming Using Ox*, 3rd edition. London: Timberlake Consultants Press and Oxford: www.doornik.com.

Emerson, J.D. and Simon, G.A. (1979), "Another Look at the Sign Test when Ties are Present: the Problem of Confidence Intervals," *The American Statistician*, 33, 140–142.

Ferner, R.E., Coleman, J., Pirmohamed, M., Constable, S.A. and Rouse, A. (2005), "The Quality of Information on Monitoring for Haematological Adverse Drug Reactions," *British Journal of Clinical Pharmacology*, 60, 448–451.

Fisher, R.A. (1941), "The Negative Binomial Distribution," *Annals of Eugenics*, XI, 182–187.

Fong, D.Y.T., Kwan, C.W., Lam, K.F. and Lam, K.S.L. (2003), "Use of the Sign Test for the Median in the Presence of Ties," *The American Statistician*, 57, 237–240.

Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J. and Ostrowski, E. (1994), *A Handbook of Small Data Sets*, London: Chapman & Hall.

Hettmansperger, T.P. and Sheather, S.J. (1986), "Confidence Intervals Based on Interpolating Order Statistics," *Statistics and Probability Letters*, 4, 75–79.

Ho, Y.H.S. and Lee, S.M.S. (2005a), "Calibrated Interpolated Confidence Intervals for Population Quantiles," *Biometrika*, 92, 234–241.

Ho, Y.H.S. and Lee, S.M.S. (2005b), "Iterated Smoothed Bootstrap Confidence Intervals for Population Quantiles," *Annals of Statistics*, 33, 437–462.

Hutson, A.D. (1999), "Calculating Nonparametric Confidence Intervals for Quantiles Using Fractional Order Statistics," *Journal of Applied Statistics*, 26, 343–353.

Huang, J.S. (1991), "Estimating the Variance of the Sample Median, Discrete Case," *Statistics and Probability Letters*, 11, 291–298.

Noether, G.E. (1967), *Elements of Nonparametric Statistics*, New York, NY: Wiley.

Papadatos, N. (1995), "Intermediate Order Statistics with Applications to Nonparametric Estimation," *Statistics and Probability Letters*, 22, 231–238.

Randles, R.H. (2001), "On Neutral Responses (Zeros) in the Sign Test and Ties in the Wilcoxon-Mann-Whitney Test," *The American Statistician*, 55, 96–101.

Scheffé, H. and Tukey, J.W. (1945), "Non-Parametric Estimation. I. Validation of Order Statistics," *Annals of Mathematical Statistics*, 16, 187–192.

Taylor, W.H. (1972), "Correlations between Length of Stay in the Job Corp and Reading Ability of the Corpsmen," *Journal of Employment Counseling*, 9, 78–85.