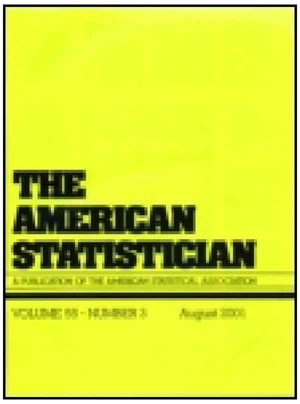

The American Statistician
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utas20

Confidence Intervals for a Discrete Population Median

Denis Larocque[a] & Ronald H Randles[a]

[a] Denis Larocque is Associate Professor, Department of Management Sciences, HEC Montréeal, 3000, chemin de la Côte-Sainte-Catherine, Montréal (Québec), Canada H3T 2A7 . Ronald H. Randles is Professor, Department of Statistics, University of Florida, PO Box 118545, Gainesville, FL 32611-8545 . The research work of Denis Larocque was supported by the Natural Sciences and Engineering Research Council of Canada. The authors are grateful to Dr R. E. Ferner for kindly providing the SIM scores data. The authors would also like to thank the Editor and the Associate Editor whose thought-provoking comments helped to clarify the interpretation of the reported confidence levels and also an anonymous referee whose comments helped to simplify the presentation of the results.
Published online: 01 Jan 2012.

Confidence Intervals for a Discrete Population Median

Denis LAROCQUE and Ronald H. RANDLES

In this article, 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. A 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 software. 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: Confidence level; Discrete distribution; Maximum likelihood; Multinomial distribution; Sign test; Tied observations.

1. INTRODUCTION

Estimation of a population median (M) is an important topic, particularly when the underlying population cannot 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.

Denis Larocque is Associate Professor, Department of Management Sciences, HEC Montréal, 3000, chemin de la Côte-Sainte-Catherine, Montréal (Québec), Canada H3T 2A7 (E-mail: *denis.larocque@hec.ca*). Ronald H. Randles is Professor, Department of Statistics, University of Florida, PO Box 118545, Gainesville, FL 32611-8545 (E-mail: *rrandles@stat.ufl.edu*). The research work of Denis Larocque was supported by the Natural Sciences and Engineering Research Council of Canada. The authors are grateful to Dr R. E. Ferner for kindly providing the SIM scores data. The authors would also like to thank the Editor and the Associate Editor whose thought-provoking comments helped to clarify the interpretation of the reported confidence levels and also an anonymous referee whose comments helped to simplify the presentation of the results.

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 et al. (2005) proposed a scale for measuring whether the Summary of Product Characteristics (SPCs) which come with a nonhaematological drug provide adequate instruction to enable pharmacists to monitor haematologically adverse drug reactions. They called their measurement scale a Systematic Instructions for Monitoring (SIM) score. Five clinicians recorded SIM scores for each of 84 SPCs of nonhaematological drugs. The median score among the five for each of the 84 SPCs are displayed in the upper plot of Figure 4. 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.

  DOI: 10.1198/000313008X269738

This article addresses the problem of finding a confidence interval for a population median when the underlying population is discrete and the possible support values are known. 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 $P(X_1 \le k) = 1/2$, we will define the population median M as the unique integer such that $P(X_1 \le M - 1) < 1/2$ and $P(X_1 \le M) > 1/2$. 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) considered 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 article 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. But because of the discrete nature of the setting considered in this article, only a finite number of associated confidence levels are possible. However, the number of values tied at the endpoints or at adjacent possible values of an interval like (1) contains information that could alter the level of confidence that should be associated with the interval. Hence, more accurate reported confidence levels may result from an appropriate use of this information. This is the motivation for considering Methods 3 to 7 presented in this section. The reported confidence levels of these methods are data dependent in the sense that two different samples may produce different reported confidence levels even though they are based on the same order statistics, depending on the placement of tied values. But even if these methods enlarge the set of possible reported confidence levels, the number of possible levels of confidence still remain finite for a given sample. This is why, if a target level of confidence is given, such as 0.95, 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 $1 - \alpha = 1 - P[B \le d - 1] - P[B \le d]$, 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 $1 - \alpha = 1 - P[B \le r - 1] - P[B \le n - s]$, where B is binomial$(n, \frac{1}{2})$. The integer d is chosen to produce the smallest confidence level that exceeds or equals 0.95. 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.

The next four methods of assigning a confidence level to an interval like (1) are all associated with inverting versions of a two-tailed sign test. 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

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

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 $X_{(n+1-d)}$. 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 $X_{(n+1-d)} + \epsilon < X_{(n+2-d)}$.

Method 4: Motivated by the two-tailed sign test with continuous data, let

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

where $n_+^c =$ (number of X_i's $> c$), $n_-^c =$ (number of X_i's $< c$) and B is binomial$(n, \frac{1}{2})$. Using this p value, the confidence interval $[X_{(d)}, X_{(n+1-d)}]$ is then given a confidence level through (3).

Next consider inverting a two-tailed sign test when X is discrete. In particular, 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:

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

Because there may be one or more observations on $X_{(d)} - 1$ or $X_{(n+1-d)} + 1$, the process of defining a p value will involve inverting 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

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

and

$$n_-^c = (\text{number of } X_i\text{'s} < c).$$

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) provided 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:

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

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

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

Once such a set $(\tilde{p}_+, \tilde{p}_0, \tilde{p}_-)$ is determined, a p value is established using (6) and an associated confidence level is then found using (5). In what follows we describe three different methods for finding a set of estimated parameters $(\tilde{p}_+, \tilde{p}_0, \tilde{p}_-)$ that use information from the data and yet satisfy the null hypothesis condition (7).

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

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

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

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

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 $\tilde{p}_{0\text{mle}}$ and $\tilde{p}_{-\text{mle}}$ in proportion to n_0^c and n_-^c. When $n_+^c = n$, set $\tilde{p}_{+\text{mle}} = \tilde{p}_{-\text{mle}} = \frac{1}{2}$ and $\tilde{p}_{0\text{mle}} = 0$. Likewise, when $n_-^c > \frac{n}{2}$, the estimates look like (8) with the roles of $\tilde{p}_{+\text{mle}}$ and $\tilde{p}_{-\text{mle}}$ 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

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

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

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

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

$$\begin{aligned} \tilde{p}_{+\text{cql}} &= \frac{1}{2}, \\ \tilde{p}_{0\text{cql}} &= \frac{n_0^c}{n} + \frac{1}{2}\left(\frac{n_+^c}{n} - \frac{1}{2}\right), \\ \tilde{p}_{-\text{cql}} &= \frac{n_-^c}{n} + \frac{1}{2}\left(\frac{n_+^c}{n} - \frac{1}{2}\right), \end{aligned} \tag{10}$$

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 $\tilde{p}_{0\text{cql}}$ and $\tilde{p}_{-\text{cql}}$. Likewise, when $n_-^c > \frac{n}{2}$, the estimates look like (10) with the roles of $\tilde{p}_{+\text{cql}}$ and $\tilde{p}_{-\text{cql}}$ 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 $(\tilde{p}_{+\text{mle}}, \tilde{p}_{0\text{mle}}, \tilde{p}_{-\text{mle}}) = (\frac{1}{2}, 0, \frac{1}{2})$ 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 $(\tilde{p}_+, \tilde{p}_0, \tilde{p}_-) = (\tilde{p}_{+\text{cql}}, \tilde{p}_{0\text{cql}}, \tilde{p}_{-\text{cql}})$ 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).

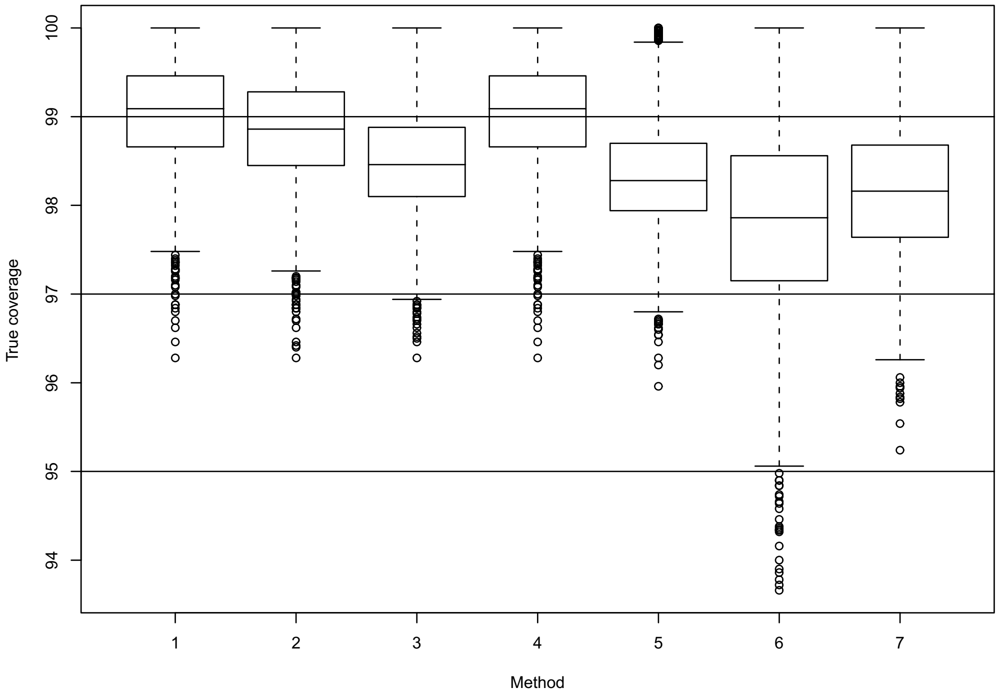

Figure 1. Simulation results with the Poisson and negative binomial distributions. The plot shows the distribution of the true coverage (empirical coverage over the 5,000 samples) over all 1,352 configurations for all seven methods.

Methods 1 and 2 are known and currently implemented in commercial software. While motivated by earlier work on the subject, Methods 3 and 4 have not been proposed before as far as the authors know. Methods 5, 6, and 7 are new methods. However, we can mention already that, based on the results from the simulation study given in the next section, our general recommendation will be to favor Methods 5 and 7 for use with discrete data.

We conclude this section by answering the important question of how one should interpret a reported confidence level. The correct interpretation for a given distribution and sample size is the following: *over repeated sampling, the reported confidence is close on average to the actual coverage probability*. This interpretation is in the spirit of the one advocated by Agresti and Coull (1998), who mentioned in the first paragraph of Section 2 of their paper:

> Most practitioners, however, probably interpret confidence coefficients in terms of "average performance" rather than "worst possible performance." Thus, a possibly more relevant description of performance is the long-run percentage of times that the procedure is correct when it is used repeatedly for a variety of data sets in various problems with possibly different parameter values.

3. SIMULATION STUDY

A 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, 12 different parameterizations 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 1,040 configurations for the Poisson distribution (40 df × 26 sample sizes) and 312 (12 parameterizations × 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, confidence intervals were constructed by choosing the smallest interval that has an associated confidence level of at least 95%. For each of the 1,352 configurations, the quantities of interest were estimated by generating 5,000 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. Does the interval maintain the minimal desired coverage probability (which is 95%)?
2. Is the reported confidence level accurate? That is, is the difference between the reported confidence and the true coverage small?
3. Is the interval short?

Basically, we are looking for short confidence intervals that maintain the minimal desired coverage probability and which

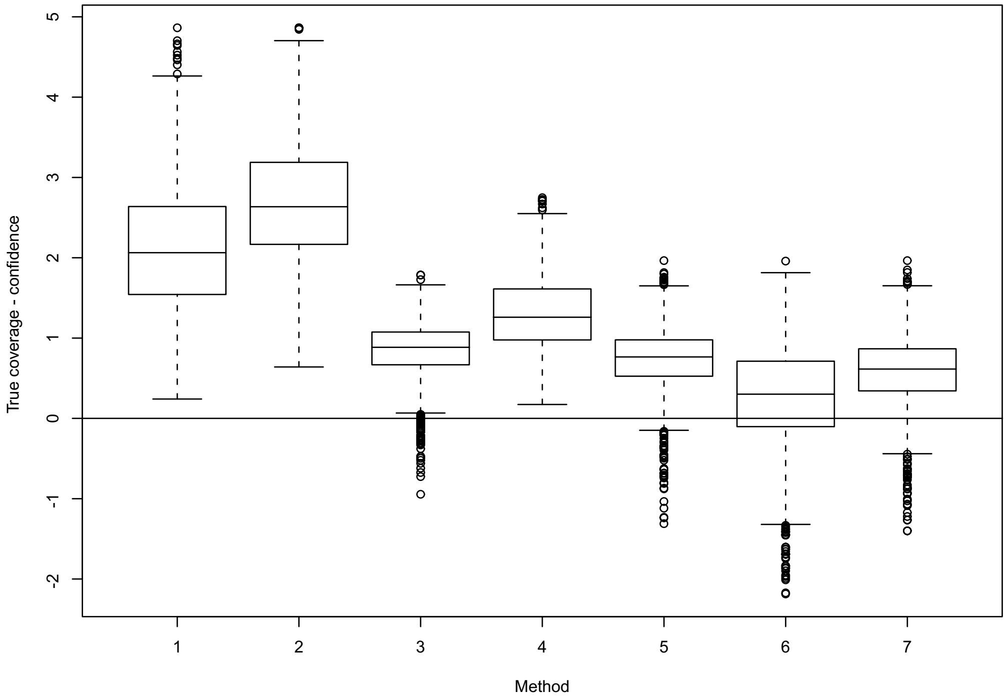

Figure 2. Simulation results with the Poisson and negative binomial distributions. The plot shows the distribution of the true coverage minus the average reported confidence over all 1,352 configurations for all seven methods.

Table 1. Simulation results with the Poisson and negative binomial distributions. The table reports the average ranks (over the 1352 configurations) for each method in the intra-configuration rankings of the absolute difference between average reported confidence and true coverage and the average ranks in the intra-configuration rankings of the interval lengths.

Intra-configuration rankings of absolute difference between reported confidence and true coverage	
Method	Average rank
6	1.78
7	2.15
5	3.13
3	3.45
4	4.55
1	6.27
2	6.67
Intra-configuration rankings of confidence interval length	
Method	Average rank
6	1.56
7	2.61
5	3.00
3	3.94
2	5.11
4	5.88
1	5.88

have associated confidence levels that are accurate. The discussion of the simulation results is built around the three aspects above but complete results from the simulation study are available from the first author.

In what follows, the true coverage for a given configuration and a given method is the empirical coverage over the 5,000 samples for that configuration and method. Figure 1 depicts boxplot summaries of the distribution of the true coverage over all 1,352 configurations for all seven methods. Except for Method 6, all other methods have a true coverage which is above 95% for all configurations. The true coverage of Method 6 falls below 95% for some configuration, 93.66% being the smallest (worst) observed true coverage.

Once the minimal desired coverage probability is maintained, the accuracy of the reported confidence is important. Figure 2 presents the summaries (over the 1,352 configurations) of the true coverage minus the averages (over the 5,000 samples) reported confidence level. A value above 0 indicates that the method underestimates the true coverage for that particular configuration while a value below 0 indicates the opposite. It is clear that Methods 1 and 2 perform very poorly compared to the others. They often severely underestimate the true coverage probability. Methods 3, 5, 6, and 7 are doing a much better job in reporting accurate levels of confidence. The upper part of Table 1 provides another view of the accuracy of the methods. The seven methods were ranked according to the absolute value of true coverage minus average reported confidence levels separately for every single configuration. Hence, 1,352 rankings (each one between 1 to 7) were obtained. The table reports the average rank (over the 1,352 configurations) of each method. Method 6 has the smallest average rank (1.78) among the seven methods. Method 7 and Method 5 come in second and third

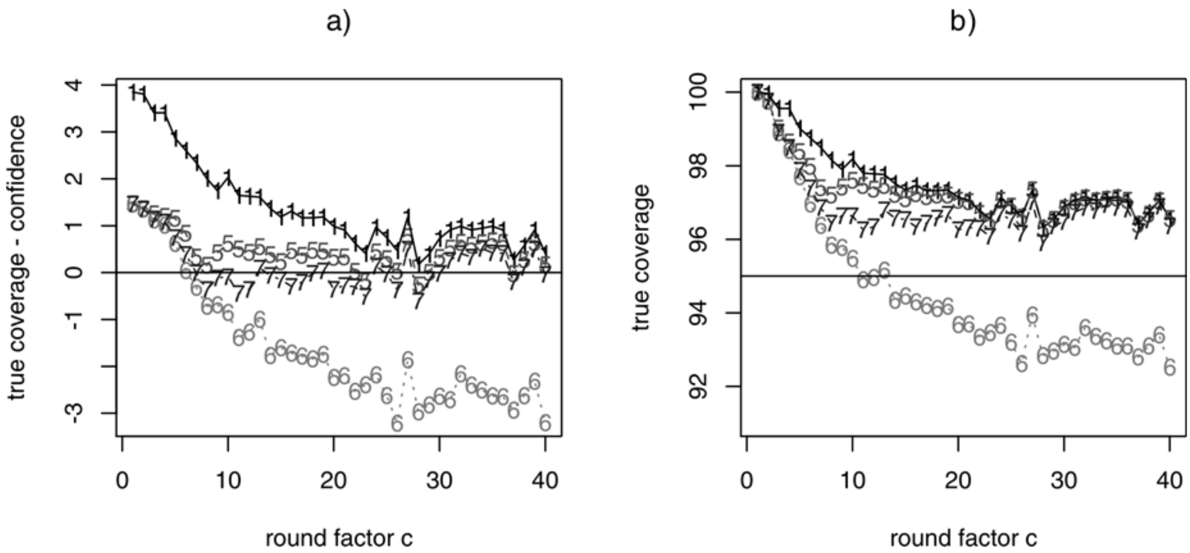

Figure 3. Simulation results for the construction of the 95% confidence interval for the sparse normal(c) distribution with a sample size of 40. The lines are numbered according to the methods they are representing. Plot (a) is the true coverage minus the average reported confidence as a function of the round factor c. Plot (b) is the true coverage.

place, respectively, and once again, Methods 1 and 2 perform very poorly.

The lengths of the intervals is the last important aspect we are interested in. For every single configuration, the seven methods were ranked according to the average length of the confidence interval (over the 5,000 samples). The lower part of Table 1 reports the average rank (over the 1,352 configurations) of each method. Method 6 is again the best one since its average rank (1.56) was the smallest among the seven methods followed once again by Methods 7 and 5. The price that Methods 1, 2, and 4 pay for being unable to report accurate confidences is clearly shown here. Since they tend to underestimate the true coverage more than the other methods, they must propose longer intervals to attain the minimal desired coverage probability.

From these findings, it seems that Methods 5, 6, and 7 would be the methods of choice. Note also that Method 3 is very close to Method 5 although slightly inferior. However, 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 3 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 becomes more sparse as c increases. With $c = 40$, we practically have a continuous normal distribution with standard deviation 40.

Figure 3 displays the results for the sparse normal distribution as a function of the round factor c when the sample size is 40. In order to make the figure easier to read, only Methods 1, 5, 6, and 7 are displayed. As c increases, that is, 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)

Table 2. Confidence interval results for the three datasets

	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

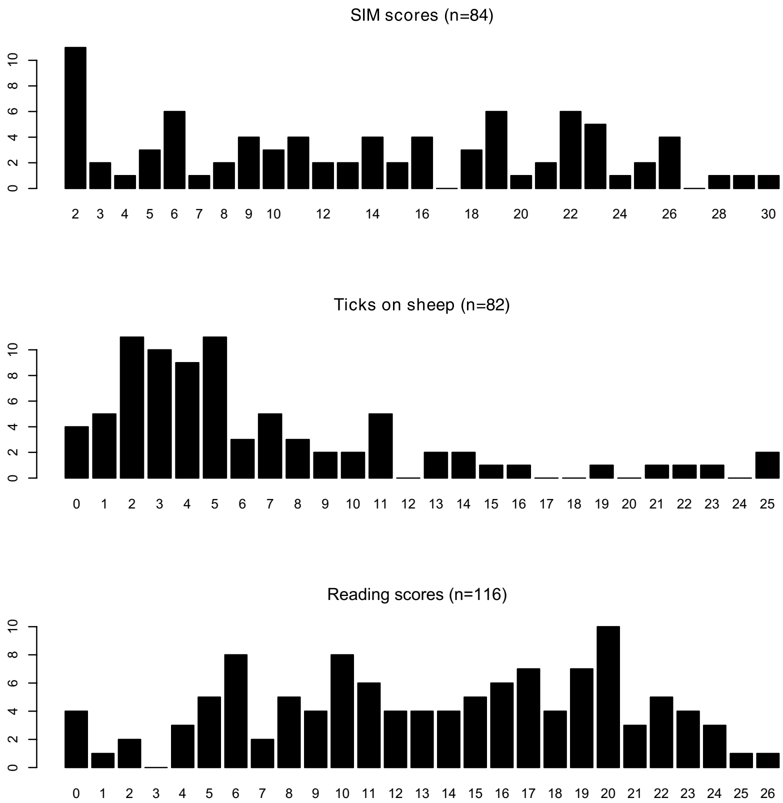

Figure 4. The three datasets used for illustration: SIM scores (upper plot), ticks on sheep (middle plot), and reading scores (lower plot).

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. Consequently, our recommendation is to use Method 5 or Method 7 for routine use.

4. EXAMPLE AND CONCLUDING REMARKS

To illustrate the practical use of the confidence intervals, we first return to the SIM scores example described in Section 1. The left part of Table 2 presents the confidence intervals obtained by all seven methods along with their reported confidence levels. As in the rest of the article, 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 4 display two classic datasets. The middle plot displays counts of ticks on 82 sheep as first reported by Fisher (1941) and included by Hand et al. (1994). The lower plot in Figure 4 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 datasets, the middle and right parts of Table 2 show that Methods 3, 5, 6, and 7 produced the shortest intervals [4, 5] and [12, 16], 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 implemented in MINITAB and SAS, respectively) can be improved upon when dealing with a discrete population. The three methods based on versions of two-tailed sign tests that can handle zeros (Methods 5, 6, and 7) gave the best results in our simulation study. 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. However, Method 6 may encounter problems for distributions that are more sparse and this is why *we recommend Methods 5 and 7 for routine use*.

If we define "conservative" to describe the fact that the reported confidence is on average smaller than the true coverage probability for a given distribution and sample size, then clearly Methods 1, 2, and 4 are the only conservative ones (see Figure 1). However, as advocated by Agresti and Coull (1998), we think that the accuracy of a reported confidence should be the most important criterion. As a matter of fact, statisticians routinely use confidence intervals based on asymptotic results that are not guaranteed to be conservative for finite samples when the assumed model is only slightly misspecified. Hence, Methods 5 and 7, which report more accurate confidences, which maintain the minimal desired coverage probability (95%) and which are still valid for continuous distributions (unlike Method 6) should be preferred unless strict conservativeness is needed.

Borrowing the terminology of Agresti and Coull (1998), we could then say that Methods 5 and 7 are "approximate" while Methods 1 and 2, which are always conservative, are "exact." But the improved accuracy of the reported confidences of Methods 5 and 7 over Methods 1 and 2 is another example that, as in Agresti and Coull (1998), approximate is sometimes better than exact.

[Received July 2007. Revised October 2007.]

REFERENCES

Agresti, A., and Coull, B.A. (1998), "Approximate is Better Than 'Exact' for Interval Estimation of Binomial Proportions," *The American Statistician*, 52, 119–126.

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 ed.), 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.

——— (2005b), "Iterated Smoothed Bootstrap Confidence Intervals for Population Quantiles," *The Annals of Statistics*, 33, 437–462.

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

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

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