Another Look at the Sign Test When Ties Are Present: The Problem of Confidence Intervals

JOHN D. EMERSON AND GARY A. SIMON*

This article discusses several methods by which ties are treated in the usual sign test. The process of inverting a test to obtain a confidence interval shows that the different methods may lead to unusual intervals.

KEY WORDS: Nonparametric confidence intervals; Sign test; Sign test confidence intervals.

1. INTRODUCTION

Let $X_1, X_2, \ldots, X_n$ represent a random sample from some population. Consider testing the null hypothesis

$$H_0\text{: median} = \theta_0 \quad \text{versus} \quad H_1\text{: median} \neq \theta_0$$

regarding the median of the X population. The conventional way to do this is to compute $N_+ = \sum_{i=1}^{n} I(X_i > \theta_0)$, the number of observations greater than θ_0. If the X's come from a continuous population, it follows that N_+ is binomial $(n, \frac{1}{2})$ under H_0; that is, $P(N_+ = k) = \binom{n}{k} 2^{-n}$. The related procedure, called the sign test, rejects when either $N_+ \leq a$ or $N_+ \geq b$, where a and b are determined from

$$\Pr(\text{binomial } (n, \tfrac{1}{2}) \leq a) = \tfrac{1}{2}\alpha = \Pr(\text{binomial } (n, \tfrac{1}{2}) \geq b).$$

In reality, one's variables are always discrete and their values occur with nonzero probability, making ties possible. We show here that the different methods for handling ties lead to various problems, especially in the construction of confidence intervals. These methods may be described in terms of N_+ and two related symbols, $N_- = \sum_{i=1}^{n} I(X_i < \theta_0)$ and $N_0 = \sum_{i=1}^{n} I(X_i = \theta_0)$. Note that $N_- + N_0 + N_+ = n$. We also let X represent a generic member of the population, and let $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ be the sorted values from the sample.

Consider the following four ways for handling ties at θ_0:

1. Discard the N_0 ties. Refer N_+ to the binomial distribution with parameters $(n - N_0, \frac{1}{2})$. This procedure is equivalent to imagining that the N_0 observations that tie θ_0 never occurred.

2. Compute the statistic $T = N_+ + \frac{1}{2}N_0$ and refer T to the binomial distribution $(n, \frac{1}{2})$. This approach treats half the ties as values greater than θ_0. Of course, T can sometimes take half-integer values. If $n = 10$, for instance, calculate $\Pr(T \geq 8\frac{1}{2})$ as

$$\Pr(\text{binomial } (10, \tfrac{1}{2}) \geq 9) + \tfrac{1}{2} \Pr(\text{binomial } (10, \tfrac{1}{2}) = 8) = .0107 + \tfrac{1}{2}(.0440) = .0327.$$

3. Flip a coin N_0 times to break the ties, and then use the ordinary sign test.

4. Reject if either $N_- \geq n_-^*$ or $N_+ \geq n_+^*$, where n_-^* and n_+^* are determined from

$$\Pr(\text{binomial } (n, \tfrac{1}{2}) \geq n_-^*) + \Pr(\text{binomial } (n, \tfrac{1}{2}) \geq n_+^*) = \alpha,$$

where α is the desired level of significance. Usually one will choose $n_-^* = n_+^*$. These methods for handling ties are among those enumerated by Bradley (1968, pp. 49–53).

Noether (1967, pp. 44–45) points out that Procedure 1 is really testing

$$H_0'\text{: } P(X < \theta_0) = P(X > \theta_0)$$

versus

$$H_1'\text{: } P(X < \theta_0) \neq P(X > \theta_0).$$

Procedures 2 and 3 are also testing H_0'. Of the given methods, only Procedure 4 tests H_0. (To see that H_0 and H_0' are not the same, consider a discrete population assigning probabilities .2, .4, .4 to the values 1, 2, 3.)

Hollander and Wolfe (1973, p. 40) recommend Procedure 1, although they say nothing about distinguishing H_0 from H_0'. Noether (1975, pp. 44–45) recommends Procedure 1, and he points out that it is a test of H_0'. Lehmann (1975, p. 164) recommends Procedure 1 for the continuous case and discusses Procedures 2 and 3 for the case in which $\Pr[X = \theta_0] > 0$. Dixon and Mood (1946) recommend Procedure 2. Gibbons (1971, p. 102) mentions Procedures 1, 2, and 3, and she suggests a one-sided version of Procedure 4 (accept H_0 versus H_1^*: median $> \theta_0$ when $N_- + N_0$ is small). Conover (1970, pp. 104–110) recommends Procedure 4, although in the context of the paired–difference sign test (p. 123) he resorts to Procedure 1. Bradley (1968, p. 54) gives a critique of the different methods.

Putter (1955) shows that, as a test of H_0', Procedure 1 is superior to Procedures 2 and 3. Walsh (1951) considers the actual alpha level when the X's are independently selected from different populations that may have atoms of probability at the medians. Scheffé and Tukey (1945) investigate open or closed confidence intervals having order statistics as endpoints. Conover (1971, pp. 110–115) gives confidence intervals for general quantiles, and his work is closely related to the problems discussed in this article. Lehmann (1975, pp. 93–95) shows how to handle confidence intervals using the sign test when the discreteness has resulted from rounding to the nearest multiple of ϵ.

* John D. Emerson is Assistant Professor, Department of Mathematics, Middlebury College, Middlebury, VT 05753. Gary A. Simon is Associate Professor, Department of Applied Mathematics and Statistics, State University of New York, Stony Brook, NY 11794. This work was partially supported by National Science Foundation Grant MCS77–03649.

2. DIFFICULTIES POSED BY DISCRETENESS

A delicate confusion involving all the procedures arises in inverting the test to produce a confidence interval for the median. The usual point of view (Noether 1967, p. 56) is that the confidence set consists of all θ_0 that are acceptable in a corresponding null hypothesis. Technically only Procedure 4 can be used to furnish an interval for the median, since only Procedure 4 tests H_0. Procedures 1, 2, and 3 are testing H_0', and the resulting confidence intervals are for the parameter θ_0, for which $\Pr(X < \theta_0) = \Pr(X > \theta_0)$, even though such a parameter need not exist.

Consider first the case in which no ties exist among the X values. Let $x_{(1)} < x_{(2)} < \ldots < x_{(10)}$ be the sorted values on 10 observations. With $\alpha = .0214$, Procedure 1 accepts H_0' for $2 \leq N_+ \leq 8$. If θ_0 is in the open interval $(x_{(2)}, x_{(9)})$, then θ_0 is an acceptable value since the data split is no worse than 8–2. Suppose that one tests a value of θ_0 exactly equal to $x_{(2)}$. Procedure 1 sees a split of 8–1 with $n - N_0 = 9$. Since Pr (8–1 split or 9–0 split) = .0390 and Pr (9–0 split) = .0040, performing the test at level .0214 would require randomization. Without the randomization, when $\theta_0 = x_{(2)}$ the test accepts H_0', and thus $x_{(2)}$ must also be included in the confidence interval. The example supports Noether's advice (1967, pp. 26–27) that the conservative closed interval $[x_{(2)}, x_{(9)}]$ should be used. Procedure 2 views the situation with $\theta_0 = x_{(2)}$ as a split of 8½–1½. Since the corresponding P value is .0654 (see original description of Procedure 2), H_0' is acceptable at level .0214, again giving a closed interval. Procedure 3 sometimes accepts and sometimes rejects $\theta_0 = x_{(2)}$, and it sometimes accepts and sometimes rejects $\theta_0 = x_{(9)}$. It gives confidence intervals that may be open or closed on either end. Procedure 4 views $\theta_0 = x_{(2)}$ as creating a 1–1–8 split, and the value of $N_+ = 8$ is not sufficient to reject. Procedure 4 puts both $x_{(2)}$ and $x_{(9)}$ in the confidence set, giving a closed interval.

3. COMPARISON OF PROCEDURES

The table gives some examples of how Procedures 1, 2, 4 handle confidence intervals when ties exist among the X's. Since Procedure 3 utilizes a secondary randomization process, its outcome is ambiguous and is therefore not illustrated. In each situation the alpha level is fixed in consideration of the original n. It is assumed that there are no ties aside from those listed. Several comments may be made about these results.

Procedure 1 can produce closed, open, half-open, singleton, or empty intervals. The empty interval for problem E is not so distressing, since the data suggest that no value of θ_0 satisfies H_0'.

Procedure 2 can produce closed, open, half-open, or singleton intervals. Empty intervals are impossible, since the sample median will always be accepted at usual levels of significance.

Procedure 4 produces an interval that is closed or a singleton. Open intervals are impossible because of the way the test procedure attaches signs to values tied at θ_0. (See Bradley 1968, pp. 186–189.)

Examples of Confidence Intervals Created by Procedures 1, 2, and 4

Problem	Procedure 1	Procedure 2	Procedure 4
A: $n = 10$, $\alpha = .0214$, no ties	$[x_{(2)}, x_{(9)}]$	$[x_{(2)}, x_{(9)}]$	$[x_{(2)}, x_{(9)}]$
B: $n = 10$, $\alpha = .0214$, $x_{(2)} = x_{(3)}$	$[x_{(2)}, x_{(9)}]$	$[x_{(2)}, x_{(9)}]$	$[x_{(2)}, x_{(9)}]$
C: $n = 10$, $\alpha = .0214$, $x_{(1)} = x_{(2)}$	$[x_{(2)}, x_{(9)}]$	$[x_{(2)}, x_{(9)}]$	$[x_{(2)}, x_{(9)}]$
D: $n = 10$, $\alpha = .0214$, $x_{(1)} = x_{(2)}$, $x_{(9)} = x_{(10)}$	$(x_{(2)}, x_{(9)})$	$(x_{(2)}, x_{(9)})$	$[x_{(2)}, x_{(9)}]$
E: $n = 20$, $\alpha = .1154$, $x_{(1)} = \ldots = x_{(14)}$	empty	$\{x_{(1)}\}$	$\{x_{(1)}\}$
F: $n = 20$, $\alpha = .1154$, $x_{(3)} = \ldots = x_{(14)}$	$\{x_{(3)}\}$	$\{x_{(3)}\}$	$\{x_{(3)}\}$
G: $n = 20$, $\alpha = .0004$, $x_{(1)} = \ldots = x_{(6)}$	$[x_{(1)}, x_{(18)}]$	$[x_{(1)}, x_{(18)}]$	$[x_{(1)}, x_{(18)}]$

The Procedure 1 interval is a subset of the Procedure 2 interval, which is in turn a subset of the Procedure 4 interval.

4. CONCLUSIONS

In the discrete case with

$$H_0': \Pr(X < \theta_0) = \Pr(X > \theta_0)$$

versus

$$H_1': \Pr(X < \theta_0) \neq \Pr(X > \theta_0),$$

one should use Procedure 1 for handling ties at θ_0. Confidence sets found by inverting this test may be exotic.

Procedure 4 tests:

$$H_0: \text{median} = \theta_0 \quad \text{versus } H_1: \quad \text{median} \neq \theta_0.$$

This procedure automatically accepts at those points at which uncertainty exists, a characteristic not shared by the first three procedures for handling ties.

When inverting hypothesis tests to produce confidence intervals, special consideration should be given to what test procedure to use. The different tests will give different intervals, although differences are only at the endpoints.

[*Received October 1977. Revised December 1978.*]

REFERENCES

Bradley, James V. (1968), *Distribution-Free Statistical Tests*, Englewood Cliffs, N.J.: Prentice-Hall.

Conover, W.J. (1971), *Practical Nonparametric Statistics*, New York: John Wiley & Sons.

Dixon, W.J., and Mood, A.M. (1946), "The Statistical Sign Test," *Journal of the American Statistical Association*, 41, 557–566.

Gibbons, Jean D. (1971), *Nonparametric Statistical Inference*, New York: McGraw-Hill Book Co.

Hollander, Myles, and Wolfe, Douglas A. (1973), *Nonparametric Statistical Methods*, New York: John Wiley & Sons.

Lehmann, E.L. (1975), *Nonparametrics*, San Francisco: Holden-Day.

Noether, Gottfried E. (1967), *Elements of Nonparametric Statistics*, New York: John Wiley & Sons.

Putter, Joseph (1955), "The Treatment of Ties in Some Nonparametric Tests," *Annals of Mathematical Statistics*, 26, 368–386.

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

Walsh, John E. (1951), "Some Bounded Significance Level Properties of the Equal-Tail Sign Test," *Annals of Mathematical Statistics*, 22, 408–417.

A Note on the Derivation of Fisher's Transformation of the Correlation Coefficient

ALAN WINTERBOTTOM*

Fisher's transformation of the bivariate-normal correlation coefficient is usually derived as a variance-stabilizing transformation and its normalizing property is then demonstrated by the reduced skewness of the distribution resulting from the transformation. In this note the transformation is derived as a normalizing transformation that incorporates variance stabilization. Some additional remarks are made on the transformation and its uses.

KEY WORDS: Correlation coefficient; Cumulants; Normalizing transformations; Variance-stabilizing transformations.

1. INTRODUCTION

Much has been written concerning Fisher's inverse hyperbolic tangent transformation for the distribution of the bivariate-normal sample correlation coefficient. In a major work by Hotelling (1953), the normalizing properties of the transformation were thoroughly investigated and it was noted that it could be derived as a variance-stabilizing transformation. Because for variance-stabilizing transformations the variance becomes approximately independent of the mean, there is often an accompanying normalization. In the case of Fisher's transformation the normalization is remarkably effective, and this suggests that it might be possible to derive the transformation primarily as a normalizing transformation with the variance-stabilizing property following from it. In Section 2 the transformation is derived as one that reduces skewness, and this reduction, together with a corresponding correction for bias, gives the normalization. The variance-stabilization property follows immediately.

2. DERIVATION OF THE TRANSFORMATION

2.1 Normalization

Let r and ρ denote the bivariate-normal sample and distribution correlation coefficients, respectively. The following moment expansions, given by Hotelling, will be required:

$$E(r - \rho) = -\rho(1 - \rho^2)/(2n) + O(n^{-2}),$$

$$E(r - \rho)^2 = (1 - \rho^2)^2/n + O(n^{-2}),$$

$$E(r - \rho)^3 = -15\rho(1 - \rho^2)^3/(2n^2) + O(n^{-3}),$$

and

$$E(r - \rho)^4 = 3(1 - \rho^2)^4/n^2 + O(n^{-3}),$$

where n is the sample size minus one.

Let $v(\rho) = \lim_{n\to\infty} \{n^{1/2} \operatorname{var}(r)\} = (1 - \rho^2)^2$. Then $n^{1/2}(r - \rho)/v^{1/2}(\rho)$ is asymptotically standard normal and has cumulant generating function $-t^2/2 + O(n^{-1/2})$. For large values of $|\rho|$, in $-1 < \rho < 1$, the distribution of r is markedly skew and the use of the asymptotic formula is unsatisfactory unless sample sizes are very large, say $n > 500$.

In order to accelerate convergence to normality, consider a transformation of form

$$T(r,\rho) = \{\phi(r) - \phi(\rho)\}/\{\phi'(\rho)v^{1/2}(\rho)\},$$

where the prime denotes differentiation with respect to ρ. The particular form chosen for $T(r,\rho)$ ensures that

$$\lim_{n\to\infty} E\{n^{1/2}T(r,\rho)\} = 0 \quad \text{and} \quad \lim_{n\to\infty} \operatorname{var}\{n^{1/2}T(r,\rho)\} = 1.$$

We now show that the function ϕ can be determined so that, for increasing n, the rate of convergence to standard normality of $n^{1/2}T(r,\rho)$, with an appropriate correction for bias, is more rapid than for $n^{1/2}(r - \rho)/v^{1/2}(\rho)$.

Assuming that $T(r,\rho)$ satisfies differentiability conditions that admit a stochastic Taylor's series expansion, we have

$$T(r,\rho) = (r - \rho)/(1 - \rho^2) + (r - \rho)^2\phi''(\rho)/2(1 - \rho^2)\phi'(\rho) + \ldots . \quad (2.1)$$

Using the moment expansions and (2), the first four cumulants of $n^{1/2}T(r,\rho)$ are

* Alan Winterbottom is a Senior Lecturer in the Department of Mathematics at The City University, Northampton Square, London EC1V OHB, England.