CONSERVATIVE CONFIDENCE INTERVALS FOR A SINGLE PARAMETER

By Mark Finkelstein, Howard G. Tucker, and Jerry Alan Veeh

University of California Irvine, University of California Irvine, and Auburn University

March 10, 1999

Abstract. The method of constructing confidence intervals from hypothesis tests is studied in the case in which there is a single unknown parameter and is proved to provide confidence intervals with coverage probability that is at least the nominal level. The confidence intervals obtained by the method in several different contexts are seen to compare favorably with confidence intervals obtained by traditional methods. The traditional intervals are seen to have coverage probability less than the nominal level in several instances.

Introduction

Traditional methods of obtaining confidence intervals for a single unknown parameter often rest, at least in part, on the asymptotic distribution of a pivotal statistic as the sample size tends to infinity. The coverage probability provided by a confidence interval obtained in this way is therefore unknown, since the actual sample size is necessarily finite. In this paper the method of obtaining confidence intervals from tests of hypothesis is studied. The advantage of using this method to obtain a confidence interval is that the coverage probability is provably at least the nominal level regardless of the size of the sample. This property is the content of our first theorem. For this reason the method of obtaining confidence intervals from tests of hypothesis will be referred to as the conservative method. The conservative method is then applied in several different settings. In those settings for which traditional methods of obtaining a confidence interval are known, two comparisons are made. First, the coverage probability for the traditional and the conservative intervals are computed in the context of a representative example. The traditional methods are seen to provide a coverage probability which often falls substantially below the nominal level. Second, the lengths of the traditional and the conservative intervals are compared. In those instances in which the traditional and conservative intervals provide nearly equal coverage

1991 *Mathematics Subject Classification.* 62XXX.

Key words and phrases. confidence intervals.

probability the length of the conservative interval is seen to be quite competitive with the length of the traditional interval. Thus the cost of maintaining a known level of confidence by using the conservative method is not high. Further, the theoretical underpinning of the conservative method is simple and intuitively appealing.

General Theory

The general theory of constructing confidence intervals from tests of hypothesis will now be developed in the case in which there is a single unknown parameter. The test statistic M used to construct the confidence interval will be assumed to be stochastically monotone in the unknown parameter. This property will be seen to hold in the examples presented later, and makes the presentation of the theory fairly straightforward.

Suppose $\Theta \subset \mathbf{R}$ is the set of possible values of the unknown parameter θ. Denote $\theta_{\min} = \inf \Theta$ and $\theta_{\max} = \sup \Theta$ the minimum and maximum values of the unknown parameter. One or both of $\theta_{\min}$ and $\theta_{\max}$ may be infinite.

The objective is to construct a conservative $100(1-\alpha)\%$ confidence interval for θ. This means that random variables $\mathcal{L}$ and $\mathcal{U}$ are sought having the property that $P_\theta[\mathcal{L} \leq \theta \leq \mathcal{U}] \geq 1-\alpha$ for all $\theta \in \Theta$. Here and throughout whenever E is an event $P_\theta[E]$ is the probability of the event E when θ is the true value of the parameter. The random variables $\mathcal{L}$ and $\mathcal{U}$ may depend on α but not on θ.

To motivate the method, suppose that M is a test statistic suitable for testing the null hypothesis $H_0 : \theta = \theta_0$ against the altenative $H_1 : \theta > \theta_0$. Assume that the statistic M is stochastically increasing in θ. Intuitively this means that larger values of θ make M probabilistically larger. Technically this means that for any number m, $P_\theta[M \geq m]$ is a nondecreasing function of $\theta \in \Theta$. Under this assumption the null hypothesis is rejected if the observed value of M is too large. If m is the observed value of the statistic M the null hypothesis is then rejected if the p-value $P_{\theta_0}[M \geq m] \leq \alpha$. Based on the observed value m of M, a reasonable choice for the lower endpoint $\mathcal{L}(m)$ of a confidence interval for θ would be the smallest value of θ_0 for which $P_{\theta_0}[M \geq m] > \alpha$. Parallel reasoning can be used when searching for the upper endpoint of the confidence interval.

This discussion suggests the following method for constructing the random variables $\mathcal{L}$ and $\mathcal{U}$. For each real number r define $\mathcal{L}(r) = \inf\{\theta \in \Theta : P_\theta[M \geq r] > \alpha\}$, and, if this set is empty set $\mathcal{L}(r) = \theta_{\max}$. Similarly define $\mathcal{U}(r) = \sup\{\theta \in \Theta : P_\theta[M \leq r] > \alpha\}$ and set $\mathcal{U}(r) = \theta_{\min}$ if this set is empty.

Theorem 1. *Suppose M is a statistic which is stochastically monotone increasing in θ. For any $0 < \alpha < 1$ and any $\theta \in \Theta$, $P_\theta[\mathcal{L}(M) \leq \theta] \geq 1-\alpha$, $P_\theta[\theta \leq \mathcal{U}(M)] \geq 1-\alpha$, and $P_\theta[\mathcal{L}(M) \leq \theta \leq \mathcal{U}(M)] \geq 1-2\alpha$.*

When the statistic M is stochastically decreasing in θ the endpoints are defined using the formulas $\mathcal{L}(r) = \sup\{\theta \in \Theta : P_\theta[M \leq r] > \alpha\}$, and, if this set is empty, $\mathcal{L}(r) = \theta_{\min}$. Similarly $\mathcal{U}(r) = \inf\{\theta \in \Theta : P_\theta[M \geq r] > \alpha\}$ with $\mathcal{U}(r) = \theta_{\max}$ if this set is empty. Theorem 1 is then proved similarly for stochastically decreasing M.

Applications and Comparisons

The general theory will now be illustrated in several different contexts. Comparison with traditional methods will be provided as appropriate.

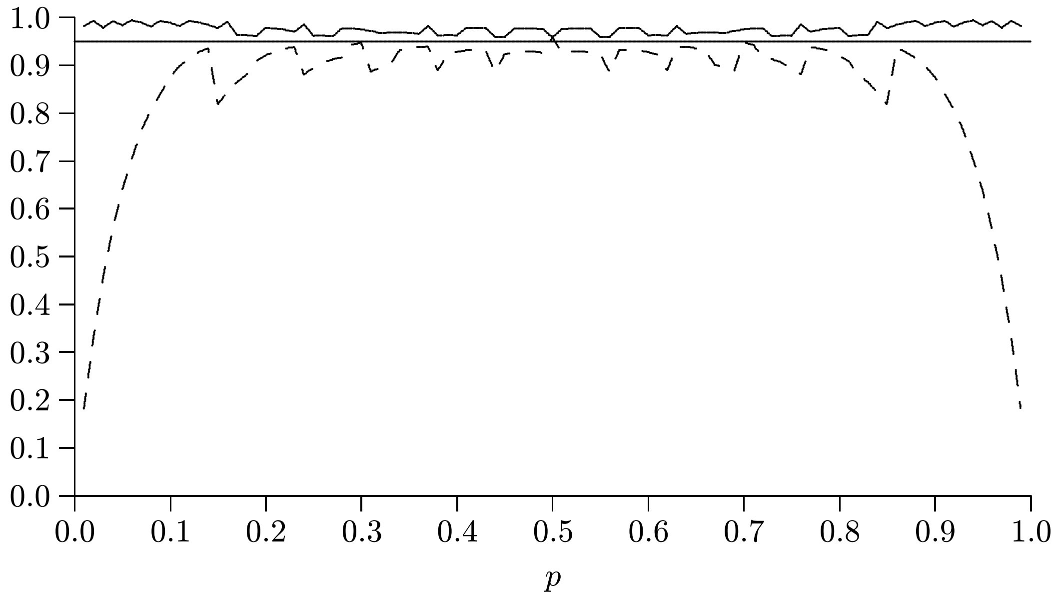
Binomial Success Probability. Perhaps the first application of the conservative method historically was made in the context of finding a confidence interval for the success probability p in a sequence of independent Bernoulli trials. Bickel and Doksum (1977) and Kendall and Stuart (1991) are two of many references to the use of this method in this setting.

Let M denote the number of successes in a known number n of independent Bernoulli trials each with unknown success probability p. The statistic M is stochastically increasing in p, and since the tail probabilities of M are continuous functions of p, $\mathcal{L}(m)$ and $\mathcal{U}(m)$ are the solutions of the equations $P_p[M \geq m] = \alpha$ and $P_p[M \leq m] = \alpha$ respectively, provided only that $m \neq 0, n$. These equations are easily solved numerically by making use of the connection between binomial tail probabilities and the incomplete beta function.

In order to access the reasonableness of the method, a comparison was made with traditional normal theory confidence intervals for p. For concreteness, two sided 95% confidence intervals were found, so that $\alpha = 0.025$. The traditional normal theory interval for p has endpoints $\mathcal{L}'(m) = m/n - 1.96\sqrt{(m/n)(1 - m/n)/n}$ and $\mathcal{U}'(m) = m/n + 1.96\sqrt{(m/n)(1 - m/n)/n}$.

For the first comparison the coverage probability for intervals obtained by the two methods were computed in the case in which $n = 20$ and the true value of p varied from 0.01 to 0.99 by increments of 0.01. The coverage probability is defined by $P_p[\mathcal{L}(M) \leq p \leq \mathcal{U}(M)]$ and $P_p[\mathcal{L}'(M) \leq p \leq \mathcal{U}'(M)]$ respectively.

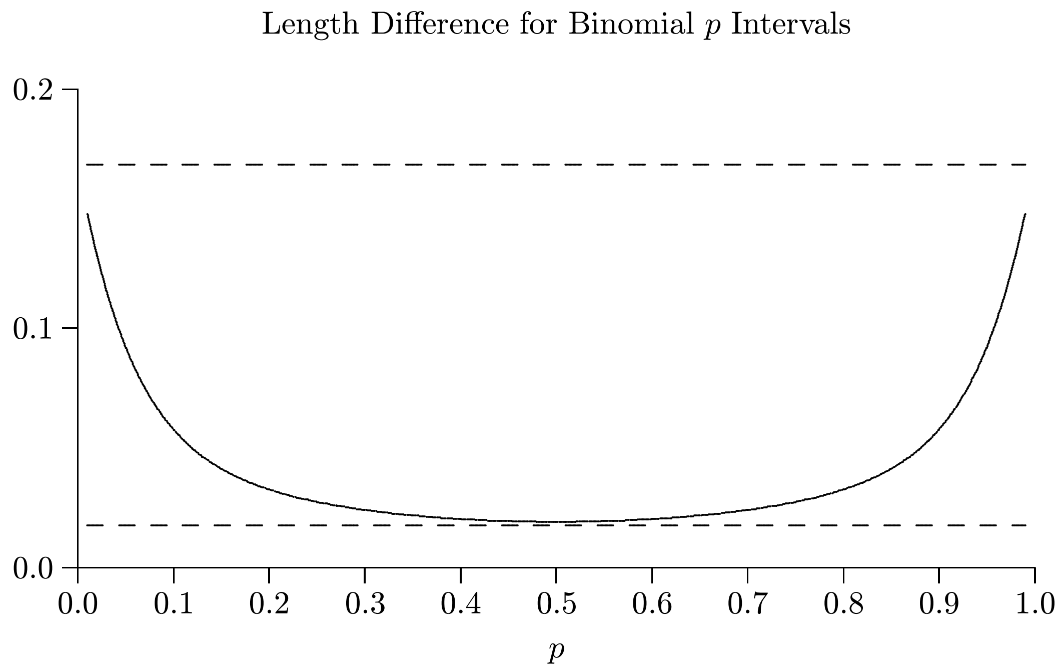
Coverage Probability for Binomial p ($n = 20$)



Conservative Method (solid)
Normal Method (dashed)

As expected, the coverage probability for the normal theory intervals is quite poor for both small and large values of p. What is somewhat unexpected is the fact that the coverage probability for the normal theory intervals is often near 90% even for values of p near 0.5. The conservative method always has coverage probability of at least 95%, and is quite conservative for both small and large values of p.

A high level of confidence can always be achieved by choosing an extremely wide confidence interval. To see whether the intervals obtained from the conservative method were extraordinarily wide, three quantities were computed: the minimum of the difference in length $\min\{\mathcal{U}(m) - \mathcal{L}(m) - (\mathcal{U}'(m) - \mathcal{L}'(m)) : 0 \leq m \leq n\}$, the maximum of the difference in length $\max\{\mathcal{U}(m) - \mathcal{L}(m) - (\mathcal{U}'(m) - \mathcal{L}'(m)) : 0 \leq m \leq n\}$, and the expected difference in length $E_p[\mathcal{U}(M) - \mathcal{L}(M) - (\mathcal{U}'(M) - \mathcal{L}'(M))]$. (Here E_p denotes the expectation when p is the true value of the parameter.) This comparison is again for $n = 20$ and $\alpha = 0.025$.



Expected Difference (solid)
Maximum and Minimum Difference (dashed)

The minimum difference was 0.0178 for all values of p; the maximum difference was a constant 0.1684. The expected difference in length was about 0.02 for the central values of p for which the two methods give comparable coverage probabilities. The expected ratio of the lengths (not shown) indicates that the conservative intervals are about 5% longer than the normal theory intervals over the central values of p. The conservative intervals are very competitive with the traditional normal theory intervals.

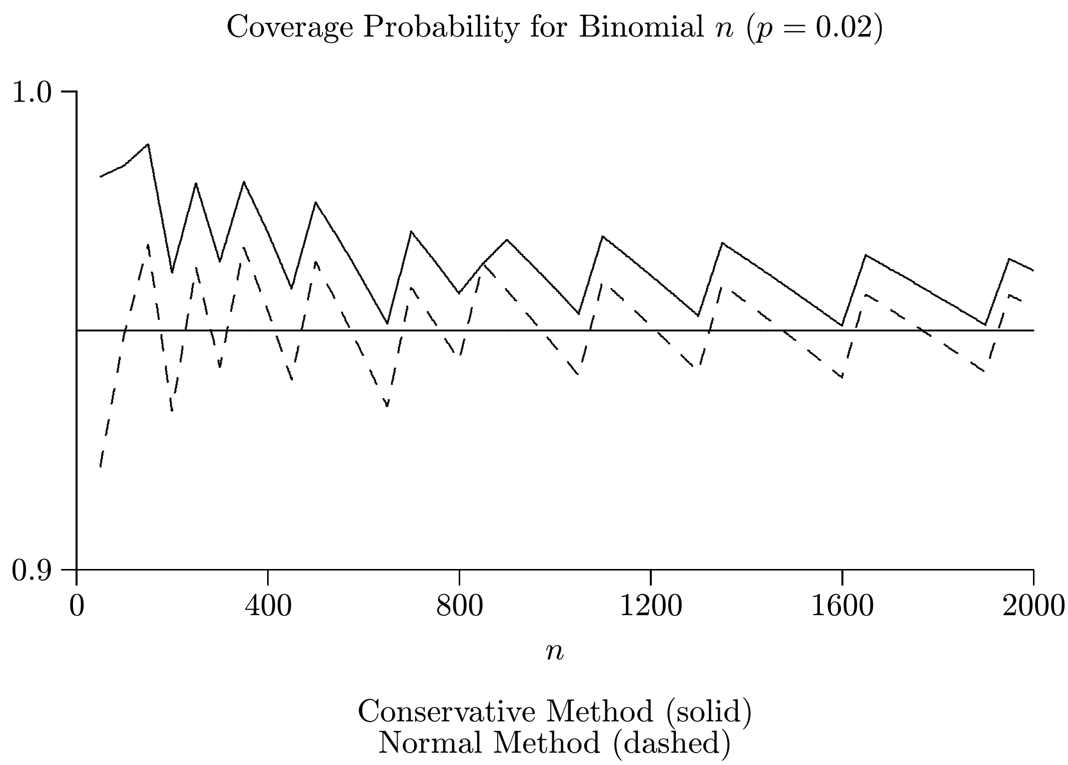
Suppose that instead of a single observation M on the binomial distribution a random sample $M_1, \dots, M_r$ on the binomial distribution with parameters n and p is given. The conservative method can then be applied using the statistic $M_1 + \cdots + M_r$ in place of M, and making use of the fact that the sum has the binomial distribution with parameters nr and p.

Binomial Number of Trials. A second application of the method in the binomial setting occurs when the success probability p is known but the number n of trials is unknown. Such is the case in transect sampling of wildlife populations when the effective width of the transect is known. See XXX (YYY). The observed number M of successes is stochastically increasing in n (for p fixed) and the method is easily applied numerically.

In this setting normal theory intervals can again be obtained by using the fact that $(M - np)/\sqrt{np(1-p)}$ is approximately normal.

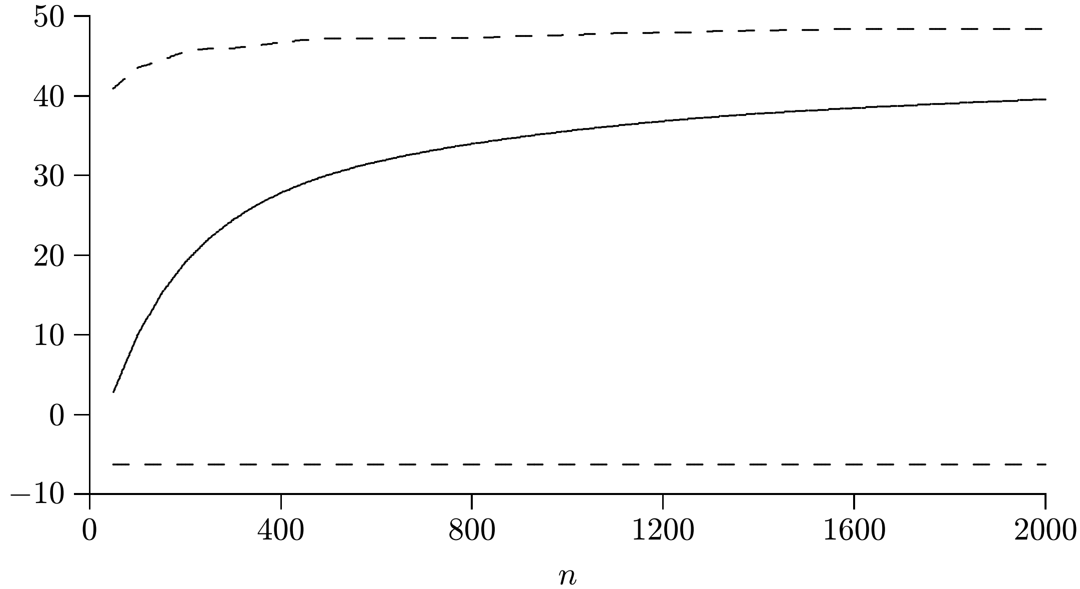
The same two comparisons as before were made in this setting. Here $p = 0.02$ was used as the known value of p. Values of n were taken as multiples of 50, up to $n = 2000$.

The coverage probability for the normal theory intervals (lower curve) is again seen to fall slighty below the nominal level.



The length comparison shows that the expected difference in length is around 40 for most values of N. This is not unexpected, since both methods are sensitive to changes in the nominal coverage probability.

Length Difference for Binomial n Intervals



Expected Difference (solid)
Maximum and Minimum Difference (dashed)

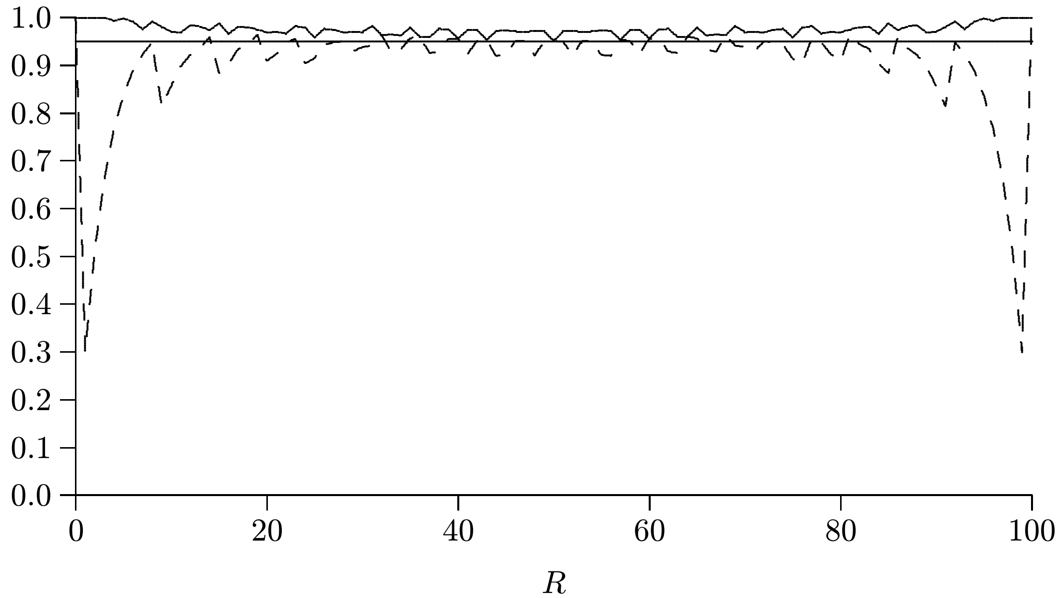
Suppose that instead of a single observation M on the binomial distribution a random sample $M_1, \ldots, M_r$ on the binomial distribution with parameters n and p is given. The conservative method can then be applied using the statistic $M_1 + \cdots + M_r$ in place of M, and making use of the fact that the sum has the binomial distribution with parameters nr and p. In this case, since r and p are known, the endpoints of the conservative confidence interval are given by the formulas $\mathcal{L}(m) = \inf\{k \geq 0 : P_{kr}[M_1 + \cdots + M_r \geq m] > \alpha\}$ and $\mathcal{U}(m) = \sup\{k \geq 0 : P_{kr}[M_1 + \cdots + M_r \leq m] > \alpha\}$ respectively. Here m is the observed value of $M_1 + \cdots + M_r$.

Hypergeometric Success Size. Suppose a simple random sample without replacement of size n is taken from a population of known size N. Suppose further that the population consists of two types of objects: successes, of which there are R (unknown), and failures, of which there are $N - R$. The objective is to find a confidence interval for R. Such a model was used by one of the authors to estimate the unknown number of defective homes in a housing tract based on the results of a sample. As is shown in Theorem 2, the number M of successes observed in the sample is stochastically increasing in R. This makes the conservative method easy to apply.

For the comparison of the two methods, $N = 100$, $n = 30$, and $\alpha = 0.05$ were used. The normal theory intervals were constructed as in Tucker (1998).

As in the binomial p case, the coverage probability for the normal theory intervals falls quite far below the nominal level even for moderate values of R.

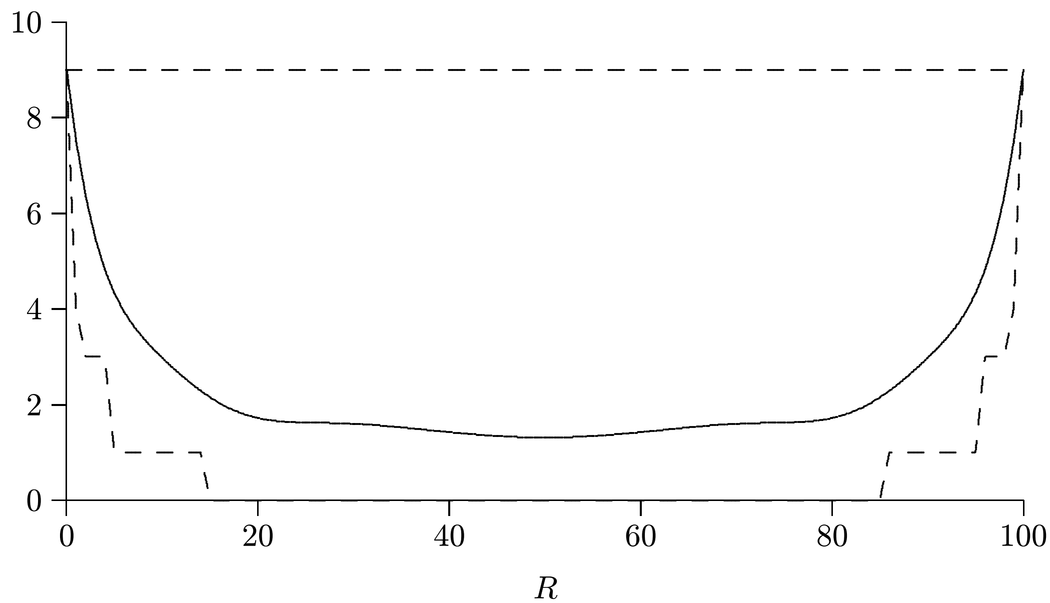
Coverage Probability for Hypergeometric R ($N = 100$, $n = 30$)



Conservative Method (solid)
Normal Method (dashed)

Comparison of the lengths of the intervals shows that the conservative method is very competitive, since the expected difference in length was less than 2 for most values of R.

Length Difference for Hypergeometric R Intervals



Expected Difference (solid)
Maximum and Minimum Difference (dashed)

Hypergeometric Population Size. In the hypergeometric setting the case in which R is known while N is unknown is also of interest. This situation arises in the use of the capture/recapture method of estimating the size of a population. Theorem 2 shows that the number M of successes is stochastically increasing in N for fixed R. The conservative method is then easily applied.

Poisson Parameter. Suppose M has a Poisson distribution with parameter $\lambda > 0$. The fact that M is stochastically increasing in λ can be seen as follows. Suppose $X(t)$, $t \geq 0$, is a Poisson process with intensity 1. Then $X(t)$ is a non-decreasing function of t, with probability 1. Also $X(t)$ has the Poisson distribution with parameter t. Thus M and $X(\lambda)$ have the same distribution. Since $X(t)$ is stochastically increasing in t, M is stochastically increasing in λ. The conservative method can now be easily applied to find confidence intervals for λ.

Suppose that instead of a single observation M on the Poisson distribution a random sample $M_1, \ldots, M_r$ on the Poisson distribution with parameter λ is given. The conservative method can then be applied using the statistic $M_1 + \cdots + M_r$ in place of M, and making use of the fact that the sum has the Poisson distribution with parameter $r\lambda$. In this case, since r is known, the endpoints of the conservative confidence interval are given by the formulas $\mathcal{L}(m) = \inf\{\lambda > 0 : P_{r\lambda}[M_1 + \cdots + M_r \geq m] > \alpha\}$ and $\mathcal{U}(m) = \sup\{\lambda > 0 : P_{r\lambda}[M_1 + \cdots + M_r \leq m] > \alpha\}$ respectively. Here m is the observed value of $M_1 + \cdots + M_r$.

Exponential Parameter. Suppose $M_1, \ldots, M_r$ is a random sample on the exponential distribution with expectation $1/\lambda$. The statistic $M_1 + \cdots + M_r$ is stochastically decreasing in λ and the conservative method can be applied. The traditional method uses the fact that $2\lambda(M_1 + \cdots + M_r)$ has a chi-squared distribution. The traditional and the conservative method are easily seen to give the same confidence interval in this case.

Number of Unseen Types. In Finkelstein, Tucker, and Veeh (1998), the conservative method was applied to find a confidence interval for the number c of coupons in the coupon collectors problem. This problem has applications to estimating wildlife abundance, among others. The conservative method was seen to be competitive with traditional normal theory methods.

Proofs

Proof of Theorem 1. The proof will be given only for $\mathcal{U}$ since the proof for $\mathcal{L}$ is similar. First observe that if $r < r'$ then $\mathcal{U}(r) \leq \mathcal{U}(r')$. To see this, the monotonicity of probability provides the inclusion $\{\theta \in \Theta : P_\theta[M \leq r] > \alpha\} \subset \{\theta \in \Theta : P_\theta[M \leq r'] > \alpha\}$, from which $\mathcal{U}(r) = \sup\{\theta \in \Theta : P_\theta[M \leq r] > \alpha\} \leq \sup\{\theta \in \Theta : P_\theta[M \leq r'] > \alpha\} = \mathcal{U}(r')$. Hence $\{r \in \mathbf{R} : \mathcal{U}(r) < \theta\}$ is an interval, say, $(-\infty, r_0]$ or $(-\infty, r_0)$. In the first case $\mathcal{U}(r_0) < \theta$. Using this, the definition of $\mathcal{U}(r_0)$, and the fact that $P_\theta[M \leq r_0]$ is a decreasing function of θ, shows that $P_\theta[M \leq r_0] \leq \alpha$. Then $P_\theta[\mathcal{U}(M) < \theta] = P_\theta[M \leq r_0] \leq \alpha$, which is the desired conclusion. In the second case the interval is open, and for any integer $n \geq 1$, $\mathcal{U}(r_0 - 1/n) < \theta$. Arguing as before this implies $P_\theta[M \leq r_0 - 1/n] \leq \alpha$. Since this holds for all n, passing to the limit as $n \to \infty$ and using the Lebesgue Dominated Convergence Theorem gives $P_\theta[M < t_0] \leq \alpha$. Again, after equating events, $P_\theta[\mathcal{U}(M) < \theta] \leq \alpha$. The one sided part of the theorem is therefore proved. For the final conclusion simply note that

$P_\theta[[\mathcal{L}(M) \le \theta \le \mathcal{U}(M)]^c] \le P_\theta[\theta < \mathcal{L}(M)] + P_\theta[\mathcal{U}(M) < \theta] \le 2\alpha$ by the first part of the theorem. □

The following theorem justifies the use of the method in the case of a hypergeometric sample.

Theorem 2. *Suppose M is the observed number of successes in a simple random sample of size n taken without replacement from a population of total size N of which R items are successes. For each fixed $N > 0$, M is stochastically increasing in R. For each fixed R, M is stochastically increasing in N.*

proof. Suppose $X_1, \dots, X_{n+1}$ is a simple random sample without replacement from $\{1, \dots, N+1\}$. Here the integers $\{1, \dots, R\}$ are considered the successes and the remaining integers are considered as failures. Define n dimensional random vectors by

$$X^{N+1} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

and

$$X^N = \begin{pmatrix} X_1 1_{[X_1 < N+1]} + X_{n+1} 1_{[X_1 = N+1]} \\ X_2 1_{[X_2 < N+1]} + X_{n+1} 1_{[X_2 = N+1]} \\ \vdots \\ X_n 1_{[X_n < N+1]} + X_{n+1} 1_{[X_n = N+1]} \end{pmatrix}.$$

Define a function $g : \mathbf{R}^n \to \mathbf{R}$ by

$$g(y_1, \dots, y_n) = \sum_{i=1}^{n} 1_{[0,R]}(y_i).$$

First observe that g is a function which is coordinatewise decreasing in each y_i. Second note that X^{N+1} is surely coordinatewise larger than X^N. Thus $g(X^{N+1}) \le g(X^N)$ surely. Hence $P[g(X^{N+1}) \ge s] \le P[g(X^N) \ge s]$ for all real s.

Observe now that when the total population size is $N+1$, $M = g(X^{N+1})$, since X^{N+1} has the same joint distribution as a simple random sample without replacement of size n from $\{1, \dots, N+1\}$. Hence $P[g(X^{N+1}) \ge s] = P_{N+1}[M \ge s]$. Thus it suffices to show that $P[g(X^N) \ge s] = P_N[M \ge s]$. This will follow if we show that X^N has the same joint distribution as a simple random sample without replacement of size n from $\{1, \dots, N\}$. To do this, observe that the set of possible values of X^N is the same as the set of outcomes for a simple random sample without replacement. The result then follows from the fact that the event $[X^N = x^N]$ consists of $N+1$ elementary outcomes of the random vector $(X_1, \dots, X_{n+1})$: the $N+1-n$ elementary outcomes in which $(X_1, \dots, X_n) = x^N$ and X_{n+1} is otherwise arbitrary, together with the n elementary outcomes in which one of $X_1, \dots, X_n$ is $N+1$ while the other X_i's and X_{n+1} take on the required values in x^N. Since the elementary outcomes for $(X_1, \dots, X_{n+1})$ are equally likely, the result is proved.

The case in which the total number N of balls is known while the number R of successes is the parameter can be analyzed as follows. Define $X = (X_1, \ldots, X_n)$, with the X's as before and set $g_R(y_1, \ldots, y_n) = \sum_{j=1}^{n} 1_{[0,R]}(y_j)$. Then g_R is increasing in R for each fixed y so $g_R(X) \leq g_{R+1}(X)$ surely. Hence $P_R[M \geq s] = P[g_R(X) \geq s] \leq P[g_{R+1}(X) \geq s] = P_{R+1}[M \geq s]$, proving the desired monotonicity. $\square$

References

P. J. Bickel and K. A. Doksum, *Mathematical Statistics*, Holden-Day, San Francisco, 1977.

M. Finkelstein, H. Tucker, and J. Veeh, *Confidence intervals for the number of unseen types*, Statistics and Probability Letters **37** (1998), 423–430.

M. G. Kendall and A. Stuart, *Advanced Theory of Statistics*, volume 2 fifth edition, Edward Arnold, London, 1991.

Howard G. Tucker, *Mathematical Methods in Sample Surveys*, World Scientific Publishing, Singapore, 1998.

Department of Mathematics, University of California, Irvine, 92697-3875 and Department of Mathematics, Auburn University, 36849-5310