COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2024.2437503

Short confidence intervals for the count parameters of the binomial, negative binomial, and hypergeometric distributions

Bret A. Holladay[a] and Mark F. Schilling[b]

[a]Statistics Department, California Polytechnic State University San Luis Obispo, San Luis Obispo, CA, USA; [b]Department of Mathematics, California State University Northridge, Northridge, CA, USA

ABSTRACT

We study interval estimation of the sample size parameter of the binomial and hypergeometric distributions and of the count parameter of the negative binomial distribution. We compare the performance of the analogs of some of the most notable confidence procedures in the literature, adapting these procedures to our setting. We consider only methods that maintain a coverage probability at or above the nominal confidence level, and in all cases analyzed we identify the procedure(s) attaining the shortest confidence interval length, as judged by cumulative average length and expected length. A link to a Shiny web app is provided for computing the recommended confidence intervals in practice.

ARTICLE HISTORY
Received 17 July 2023
Accepted 28 November 2024

KEYWORDS
Confidence interval; length minimizing; coverage probability

1. Introduction

Due to advances in computing and the ubiquitousness of statistical software such as R, statisticians have begun to revisit the classic problem of interval estimation for parameters of discrete distributions. Because practitioners are more open than ever to using applets or specialized software packages for statistical analysis, there is decreasing dependence on large sample confidence intervals based on distributional approximations, such as Wald or score intervals. Instead, using software, high-performing confidence procedures can be determined numerically from the actual distribution at hand based on selected criteria for optimality, such as minimal overall confidence interval length. Such exact procedures are particularly valuable since one does not always have a large sample.

In an article from a recent issue of *Significance* magazine, Stapleton (2023) notes that "confidence procedures are just algorithms, and we can adjust the algorithm to suit our needs." Since precision is desirable, shortness is often high on the list of needs. Length minimizing algorithms have already been developed for what are often considered the preeminent discrete statistical models, namely the binomial, Poisson, hypergeometric, and negative binomial distributions (see Sterne 1954; Crow 1956; Blyth and Still 1983; Casella 1986; Schilling and Doi 2014; Schilling and Holladay 2017; Schilling and Stanley 2022; Schilling, Holladay, and Doi 2023). However, the focus in the literature has been

CONTACT Bret A. Holladay bahollad@calpoly.edu Statistics Department, California Polytechnic State University San Luis Obispo, San Luis Obispo, CA, USA

Table 1. Parameterizations of the binomial, negative binomial (NB), and hypergeometric (HG) distributions.

Distribution	PMF	Support	Parameter restrictions
Binomial $(\boldsymbol{n}, p)$	$\binom{n}{x} p^x (1-p)^{n-x}$	$x \in \{0, 1, 2, \ldots, n\}$	$p \in [0, 1], n \in \mathbb{Z} : n \geq 0$
NB $(\boldsymbol{r}, p)$; x observed	$\binom{x+r-1}{x} p^r (1-p)^x$	$x \in \{0, 1, 2, \ldots\}$	$p \in [0, 1], r \in \mathbb{Z} : r \geq 1$
NB $(\boldsymbol{r}, p)$; N observed	$\binom{N-1}{r-1} p^r (1-p)^{N-r}$	$N \in \{r, r+1, r+2, \ldots\}$	$p \in [0, 1], r \in \mathbb{Z} : r \geq 1$
HG $(N, M, \boldsymbol{n})$	$\dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$\max\{0, n+M-N\} \leq x \leq \min\{M, n\}$	$N, M, n \in \mathbb{Z} : 0 \leq M, n \leq N$

The parameters given in bold are those that we wish to estimate. The remaining parameters of each distribution are assumed to be known.

on the parameters of these distributions that are most commonly unknown, such as the binomial success probability p. In this paper we determine short confidence intervals for the reverse cases in which those parameters are known, while the remaining parameters of these distributions that are typically known in practice are unknown. In particular, we discuss estimation of the sample size parameter n in both binomial and hypergeometric (HG) experiments and the count parameter r for the negative binomial (NB) distribution (see Table 1).

We treat a binomial random variable in the usual way and define a negative binomial random variable to be either the number of failures x in successive independent Bernoulli trials before the rth success, or the number of trials $N = r + x$ until the rth success. The hypergeometric distribution gives the probability of drawing x units possessing a particular attribute ("successes") when we randomly sample n units without replacement from a finite population of size N; M represents the total number of units in the population possessing the attribute of interest.

Although inference for the count parameters of these distributions is less common than for their other parameters, many researchers have in fact examined such problems. For the binomial model, Feldman and Fox (1968), Draper and Guttman (1971), Blumenthal and Dahiya (1981), Johnson (1981), Tang and Sindler (1987), Iliopoulos (2003), Bayoud (2011), De and Zacks (2016), and Cheng, Eck, and Crawford (2020) each treat the case when inference about n is desired when p is known.

In the prior research represented by the above references, several observations from a common binomial distribution are typically assumed. However, there are occasions in which inference about n is desired when only a single binomial count is available—the situation we consider below. Examples are provided in Section 2. Tang and Sindler (1987) and Iliopoulos (2003) provide some results for this situation.

Previous work on estimation of the parameter r of the negative binomial distribution is extensive. Generally, r has been treated as a dispersion parameter for the fitting of the negative binomial distribution to data to model overdispersion relative to a Poisson model, whereas in our setting it is a count parameter, thereby restricting interval estimates to integer values. A treatment of the latter situation, with p assumed known, has been given by Ganji, Eghbali, and Azimian (2013).

For the hypergeometric distribution, Schilling and Stanley (2022) developed procedures that produce short confidence intervals for both the parameter N (useful for capture-recapture problems) and the parameter M. This leaves the question of how to estimate the remaining hypergeometric parameter, n. However, since we can redefine the items sampled to represent the successes and let M correspond to the number of items sampled, the length minimizing procedure developed by Schilling and Stanley (2022) for estimating M can be extended to estimation of n by simply swapping the roles of n and M in the procedure. This reduces our investigations below to the binomial and negative binomial cases only.

Since we are interested in estimating integer-valued parameters, we consider countable confidence sets over integers, rather than uncountable intervals over all real values. Therefore, an interval $[l, u]$ denotes the set $\{l, l+1, ..., u-1, u\}$. We seek intervals that are both (i) *strict* in that they maintain coverage at or above the nominal level, and (ii) length minimizing.

We take a *coverage probability perspective*, extending the work of Schilling and Doi (2014), Schilling and Holladay (2017), Schilling and Stanley (2022), and Schilling, Holladay, and Doi (2023). Their approach can be summarized as follows: Let X be the observed random variable, having distribution dependent on a parameter θ which we wish to estimate. For each θ, determine the family $\mathcal{A}_\theta$ of sets S of consecutive possible values of X such that $P_\theta(X \in S) \geq 1-\alpha$. Then, for every θ, choose $S_\theta \in \mathcal{A}_\theta$ that satisfies a specific set of criteria. We call the chosen S_θ's *acceptance sets*. By the duality between confidence sets and families of hypothesis tests, the chosen S_θ's lead to a confidence set

$$C(X) = \{\theta : X \in S_\theta\}$$

for the true parameter value θ. If any of these choices of S_θ's results in a confidence set that is not an interval of consecutive values, an alternative set is substituted to resolve this anomaly.

It will be convenient to think of the *coverage probability* $CP(\theta)$ of each confidence set to be given by $P_\theta(X \in S_\theta)$, since

$$P_\theta(\theta \in C(X)) = P_\theta(X \in S_\theta).$$

The function that maps each θ to its corresponding coverage probability is called the *coverage probability function* (*cpf*). Confidence procedures can be constructed and compared through their cpf's.

Our paper's structure is as follows: In Section 2, we present a few motivating examples of situations where the confidence procedures we provide might be desired in practice. Section 3 introduces the analogs of several existing procedures. In Section 4 we compare the average and expected confidence interval length of our procedures with these previously developed procedures and identify the procedure(s) in each case that tend to produce the shortest confidence intervals.

2. Motivating examples

Johnson (1981) presents an example of estimation of n in the binomial case where p is known. The goal is to estimate the number n of submarine pingos (a type of large underwater mound) in the southern Beaufort Sea by means of transect sampling, searching along equally spaced parallel tracks. Based on known size information for the pingos, the probability that a transect intersects a pingo is known. An additional example due to Drago et al. (2015) involves network flows that are involved in cloud services. Flow data is sampled at a prescribed proportion p, and the total numbers of connections n and "healthy" connections n_h are to be estimated.

Here are two additional illustrations of when inference about the binomial n with known p may be desired:

1. Consider a community's exposure to a certain disease (for example Lyme disease). Suppose the proportion of exposed individuals requiring treatment at a hospital is known from past studies. Based on the number of such treatments given in a certain community in the past 6 months, we wish to estimate the total number n of people in the community that have been exposed during that time.
2. A factory is running an assembly line in which the time to produce each item varies. It randomly withdraws each item with equal probability and sets aside those withdrawn for later inspection. At the end of the inspection period, the selected items are carefully examined. Based on the number of items that have been set aside, the factory wishes to estimate the total number n of items produced in that run of the assembly line.

As an example for estimating the number of successes in a negative binomial situation, consider a woodpecker foraging for insects by making tiny holes in the bark of a tree. Suppose the probability p that any particular hole will contain an insect is known from prior studies. A woodpecker will continue searching for bugs until satiated. Once the woodpecker's search is over an ecologist counts the number of holes drilled and wishes to estimate the total number of insects r that the woodpecker consumed to reach satiation.

3. Summary of existing procedures

We discuss here some notable strict confidence procedures. One of the earliest such procedures in the literature is due to Clopper and Pearson (1934) for estimating p in a binomial experiment. Their method is derived by inverting the equal tailed two-sided level α test of $H_0 : p = p_0$. The analog of the Clopper and Pearson intervals can be obtained for estimation of parameters of other distributions by inverting the corresponding two-sided level α test of $H_0 : \theta = \theta_0$. Clopper and Pearson's procedure tends to be conservative due to the inflexibility of the rejection regions of an equal-tailed test (i.e., a test where the probability content in each tail does not exceed $\alpha/2$). As a result, although the Clopper and Pearson method has sometimes been referred to as the "gold standard", it tends to produce excessively wide confidence intervals when compared with other high-performing strict methods. We illustrate this point by providing comparative results for the Clopper and Pearson analogs in Section 4 below.

Blaker (2000) developed the following innovative approach to constructing confidence intervals: For each value of θ, determine the tail probability $T_\theta\ (x) = \min(P(X \leq x), P(X \geq x))$; then θ is included in the confidence interval for x if and only if $P(X\colon T_\theta\ (X)$

$\leq T_\theta(x)) > \alpha$. A primary advantage of Blaker's method is that two confidence intervals, obtained from the same data but having different confidence levels, are guaranteed to be nested; i.e., the interval at the higher confidence level will contain the one at the lower confidence level. Blaker's method can be seen as an inversion of a two-tailed test that does not restrict each tail area to be $\leq \alpha/2$, consequently the resulting confidence intervals tend to be reasonably competitive with those of length minimizing methods.

Though instances are typically quite rare, violations of nesting may occur for alternative methods that produce intervals that are shorter than Blaker's. However, one could argue that a few instances of nonnested intervals are acceptable when a significant reduction in net confidence interval length is attained.

In the negative binomial case, when the number of failures x is observed and the only unknown parameter is r, Blaker's method will sometimes produce confidence sets that contain gaps—i.e., sets not comprised of a sequence of consecutive integers. To see this, it is constructive to look at Figure 1, which shows the 90% cpf of Blaker's method for the NB $(r, p = 0.90)$ case. Each point on the graph indicates the coverage probability of r, $CP(r)$, which represents the probability that Blaker's procedure will cover r, assuming r is the true value of the parameter. We find it visually useful to connect the values in play $(P_\theta\ (X \in S) \geq 1 - \alpha)$ for each acceptance set, forming a collection of *acceptance curves* (each curve being associated with a different acceptance set).

We use the notation *a-b* to denote an acceptance set and to refer to the corresponding acceptance set probability function $P_r(a \leq X \leq b)$. To better understand the interplay between a confidence procedure and its cpf, please refer to Section 2 of Schilling and Doi (2014). The red crosses in Figure 1 indicate a region that results in a gap. Notice that the cpf transitions from 0-6, to 1-6, and back to 0-6 again; thus, the confidence set for $x = 0$ is $\{1, 2, 3, \ldots, 26, 28\}$, which omits the value $x = 27$. This is because the probability of observing a value of X with a tail probability as small as that of $x = 0$ is greater than 0.1 when $r = 26$ or 28, but not when $r = 27$.

Many existing procedures employ only acceptance sets of minimal cardinality, which tends to produce narrow intervals. Since for each value of the parameter, there may be several acceptance sets of minimal cardinality, such procedures comprise an entire class.

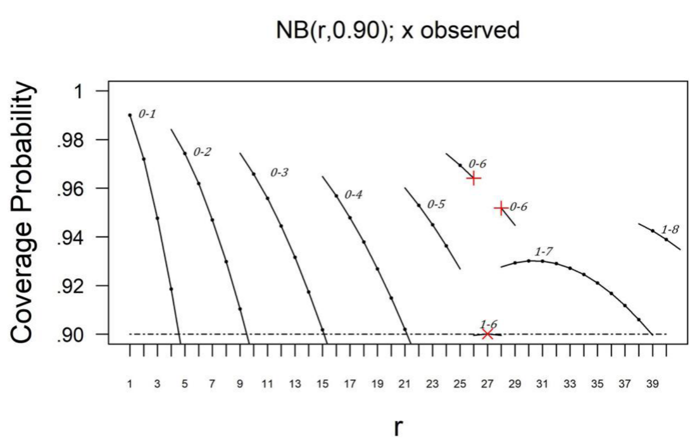

Figure 1. Illustration of a gap in Blaker's 90% cpf for the NB $(r, 0.90)$ distribution when the number of failures is observed. The acceptance set probability functions $\{P_r(a \leq X \leq b)\}$ are labeled by *a-b*. The red crosses highlight the region that corresponds to a gap.

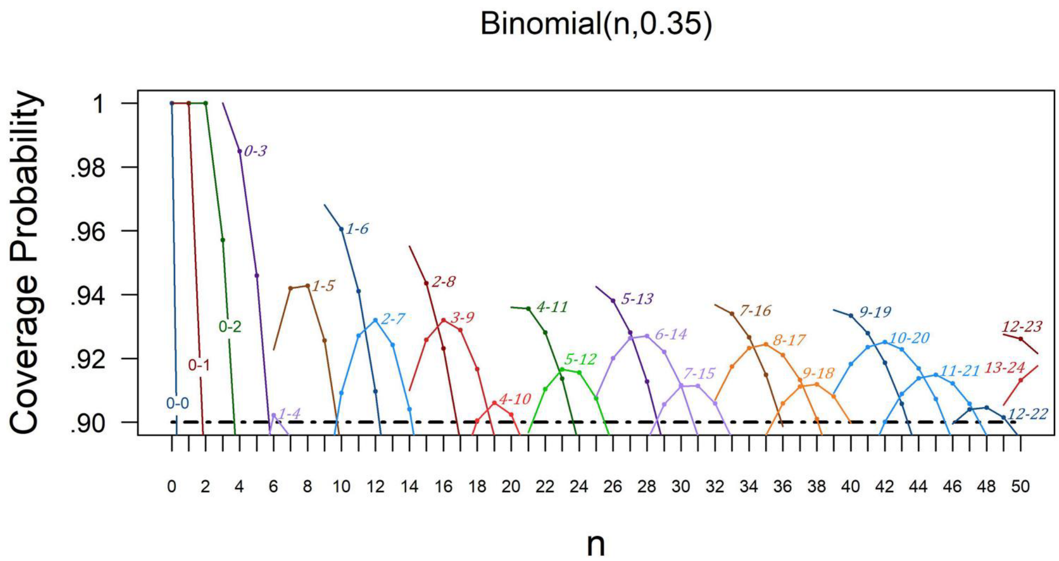

Figure 2. All 90% minimal cardinality acceptance curves in the binomial $(n, 0.35)$ case, excluding those that would cause gaps. We label acceptance curves $P_n(a \leq X \leq b)$ by *a*-*b*. Colors are used to better distinguish nearby curves. Nearby curves that are of different shades of the same color indicate that the curves have equal cardinality.

Members of this class of procedures include the analogs of the methods of Crow and Gardner (1959) and Byrne and Kabaila (2005), which were developed for estimating the parameter of the Poisson distribution. Additionally, Schilling and Doi (2014), Schilling and Holladay (2017), and Schilling and Stanley (2022) developed minimal cardinality procedures for the usual parameters of the binomial, Poisson, and hypergeometric distributions that are length minimizing.

Plotting the coverage probabilities for all the candidate minimal cardinality acceptance sets is an invaluable way to explore and contrast minimal cardinality procedures. For example, Figure 2 shows all 90% ($P_\theta\ (X \in S) \geq 0.90$) minimal cardinality acceptance curves in the binomial $(n, 0.35)$ case, excluding those that would cause gaps. The analogs of the methods of Crow and Gardner (CG), Byrne and Kabaila (BK), and Schilling and Holladay (see their Optimal Coverage procedure (OC)) choose acceptance curves in the following way: For each n, selection of acceptance curves is first reduced to those sets of minimal span curves $\{P_n(a \leq X \leq b)\}$ that would keep the sequences of $\{a\}$ and $\{b\}$ values monotonic in the parameter. These are the curves shown in Figure 2. Then, for each n, whenever there are multiple curves $\{P_n(a \leq X \leq b)\}$ available, CG chooses the curve with largest value of a, BK chooses the one with smallest value of a, and OC chooses the one with highest coverage.

Figure 3 shows all 90% minimal cardinality acceptance curves in the NB $(r, 0.84)$ case when the number of failures x is observed, except those that would cause gaps. The analogs of CG, BK, and OC choose acceptance curves in an identical fashion to the binomial case above. Other confidence levels and choices of p result in similar plots.

When the likelihood function is strongly skewed right (e.g., $L(p)$ for the negative binomial distribution), minimal cardinality procedures may not be length minimizing. Schilling, Holladay, and Doi (2023) proposed a new method that deals more effectively with such situations and applied it to estimation of p for the negative binomial distribution. Rather than being restricted to minimal cardinality sets, the procedure steps through

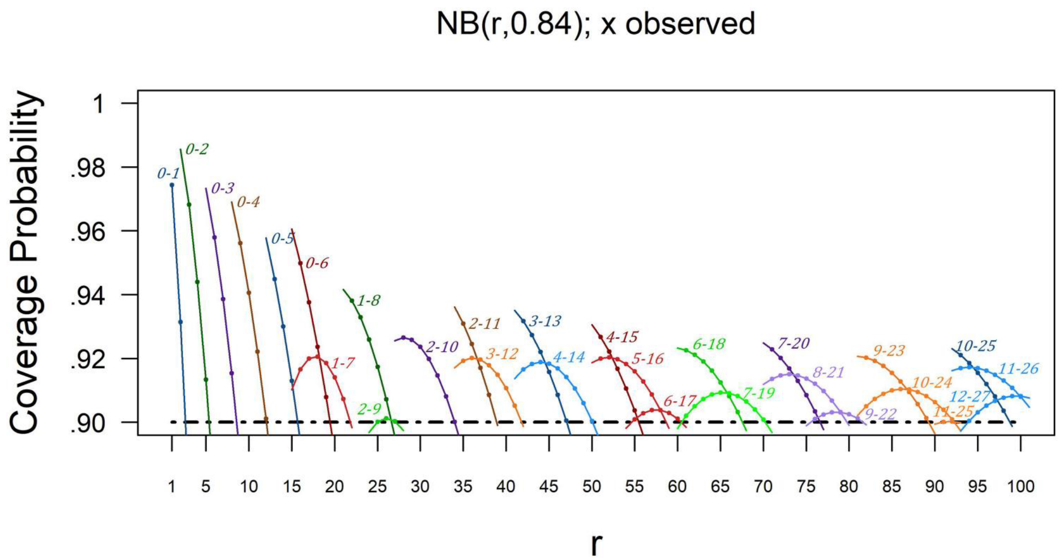

Figure 3. All 90% minimal cardinality acceptance curves in the NB $(r, 0.84)$ case when the number of failures is observed, excluding those that would cause gaps.

the values of a (beginning at $a = 0$) as the parameter to be estimated increases, transitioning from eligible acceptance sets $\{a\text{–}b\colon b \geq a\}$ to those starting at $a + 1$ at the smallest value of the parameter for which b achieves its minimum among the collection of eligible acceptance sets $\{(a+1)\text{–}b : b \geq a+1\}$. The resulting process is referred to as a *conditional* minimal cardinality (CMC) procedure, since for each given $a + 1$ the acceptance sets $(a + 1)\text{–}b$ used have minimal cardinality among all choices of b. The acceptance sets used in this procedure do not always have overall minimal cardinality, as a transition from a to $a + 1$ may lead to using an acceptance set with a larger cardinality than one that is available with a. However, significantly narrower confidence intervals are obtained than result from any (unconditional) minimal cardinality procedure.

4. Results

All the methods presented have coverage at or above the nominal level. To compare confidence interval length, we consider both cumulative average length,

$$\frac{1}{(x \leq K)} \sum_{x \leq K} (u_x - l_x + 1),$$

considered as a function of the nonnegative integer K, where $(x \leq K)$ is the number of x's less than or equal K, and expected length,

$$\sum_{all\ \ x} (u_x - l_x + 1) P_\theta(X = x),$$

considered as a function of θ. Note that confidence interval length refers to the number of elements in the confidence set, $u_x - l_x + 1$. Note that with access to both the complete set of confidence intervals for each procedure and the corresponding probability mass function, we can compute exact values for cumulative average length and expected length, without needing simulations or additional steps.

4.1. Estimation of binomial n

Figure 4a-d displays the cumulative average length of each procedure (Byrne and Kabaila (BK), Clopper and Pearson (CP), Optimal Coverage (OC), Blaker (B), and Conditional Minimal Cardinality (CMC)) relative to Crow and Gardner's method (CG) at 95% confidence for $p = 0.35$, $0.50, 0.65$ and 0.90 and $0 \leq K \leq 200$ in the binomial case.

In all four cases CG and CMC attain smaller average length than the other methods, with CG and CMC performing similarly for $p = 0.35$, 0.50, 0.65 and producing identical intervals for $0 \leq x \leq 200$ when $p = 0.90$. The results given here (at the 95% level) and throughout the remaining plots below are similar to those at confidence levels 90% and 99%. In addition, since there tends to be only subtle differences in the plots between successive values of p, the four values of p we selected ($p =$ 0.35, 0.50, 0.65 and 0.90) provide a reasonable representation of the behavior that would be seen for other values. We checked additional values of p throughout the range (0,1) and found similar results for these cases as well.

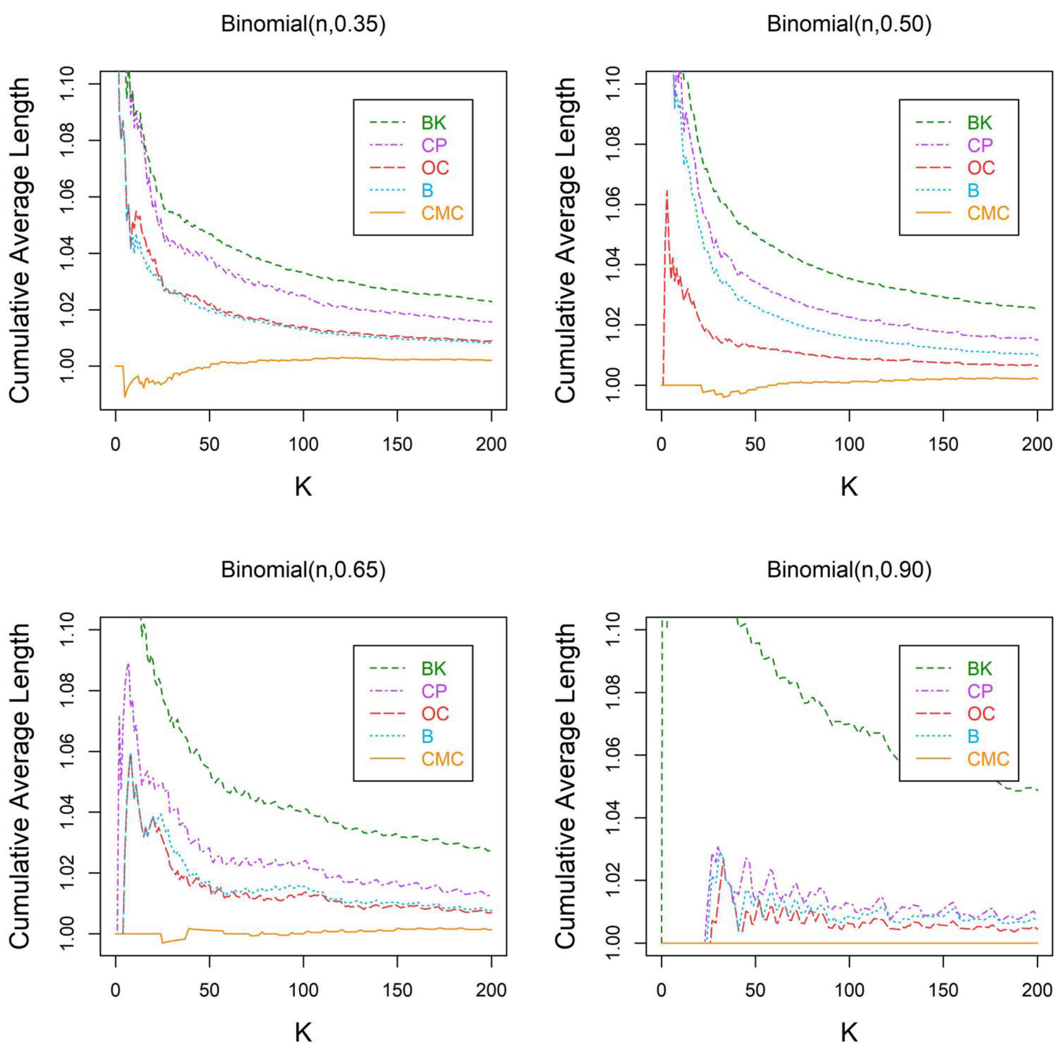

Figure 4. (a–d) 95% cumulative average length for $p = 0.35$, 0.50, 0.65 and 0.90 relative to CG in the binomial case. For each K, the average length is calculated for all x values up to and including K.

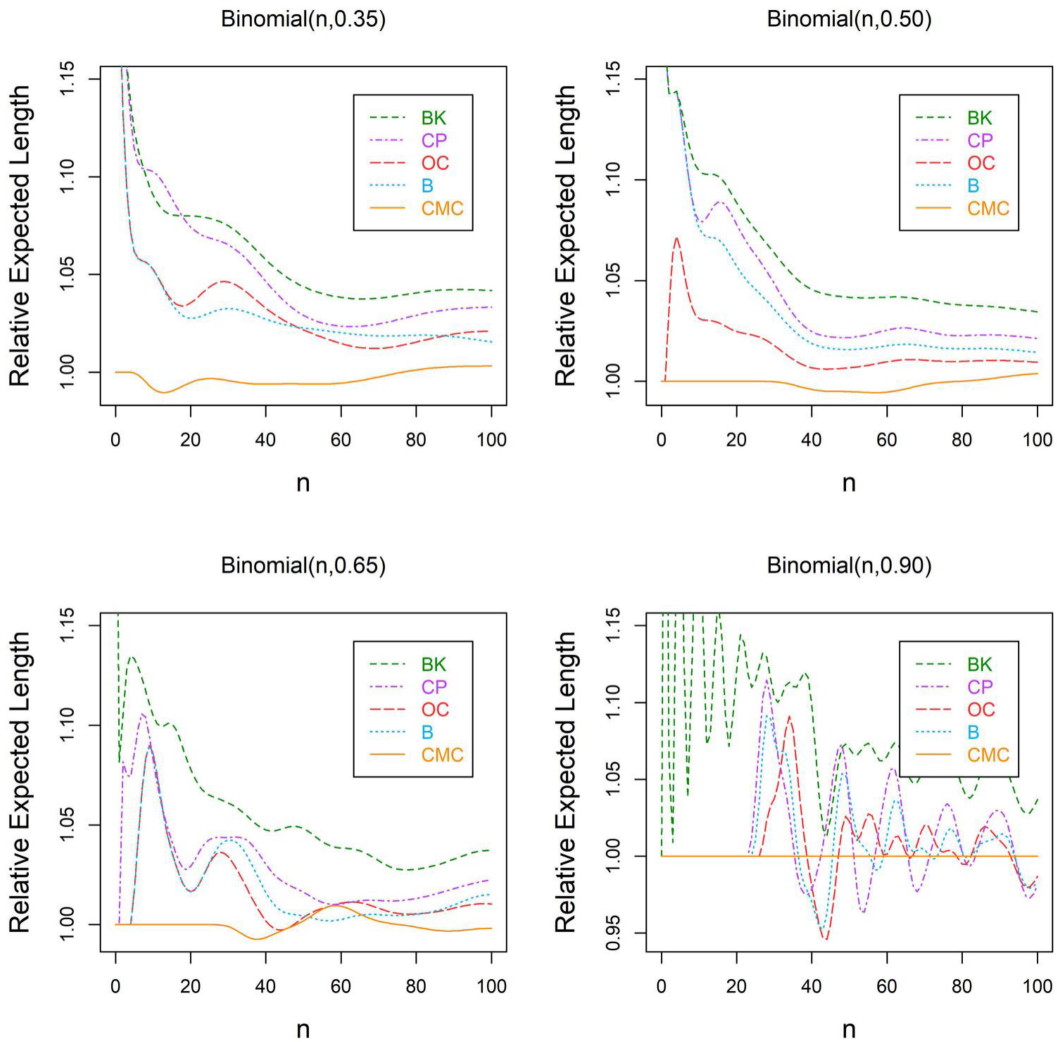

Figure 5. (a–d). 95% expected length for $p = 0.35$, 0.50, 0.65 and 0.90 relative to CG in the binomial case.

Figure 5(a–d) displays the expected length of each procedure relative to CG at 95% confidence for $p = 0.35$, 0.50, 0.65 and 0.90 and $0 \leq n \leq 100$ in the binomial case.

Here a similar hierarchy can be seen to that found for cumulative average length, with CG and CMC possessing the shortest expected lengths of all methods considered.

4.2. Estimation of negative binomial r (x observed)

Figure 6a–d shows the cumulative average length of each procedure relative to CG at 95% confidence for $p = 0.35$, 0.50, 0.65 and 0.90 and $0 \leq K \leq 200$ in the negative binomial case when the number of failures is observed. CG and CMC again have the shortest average length with CMC attaining a shorter average length for K roughly less than 100 when $p = 0.35$ and for K less than about 50 when $p = 0.50$, 0.65 and 0.90; CG attains the shortest average length elsewhere.

Figure 7(a–d) displays the expected length of each method relative to CG at 95% confidence for $p = 0.35$, 0.50, 0.65 and 0.90 and $0 \leq r \leq 100$ in the negative binomial case when the number of failures is observed. CMC tends to attain a smaller expected width

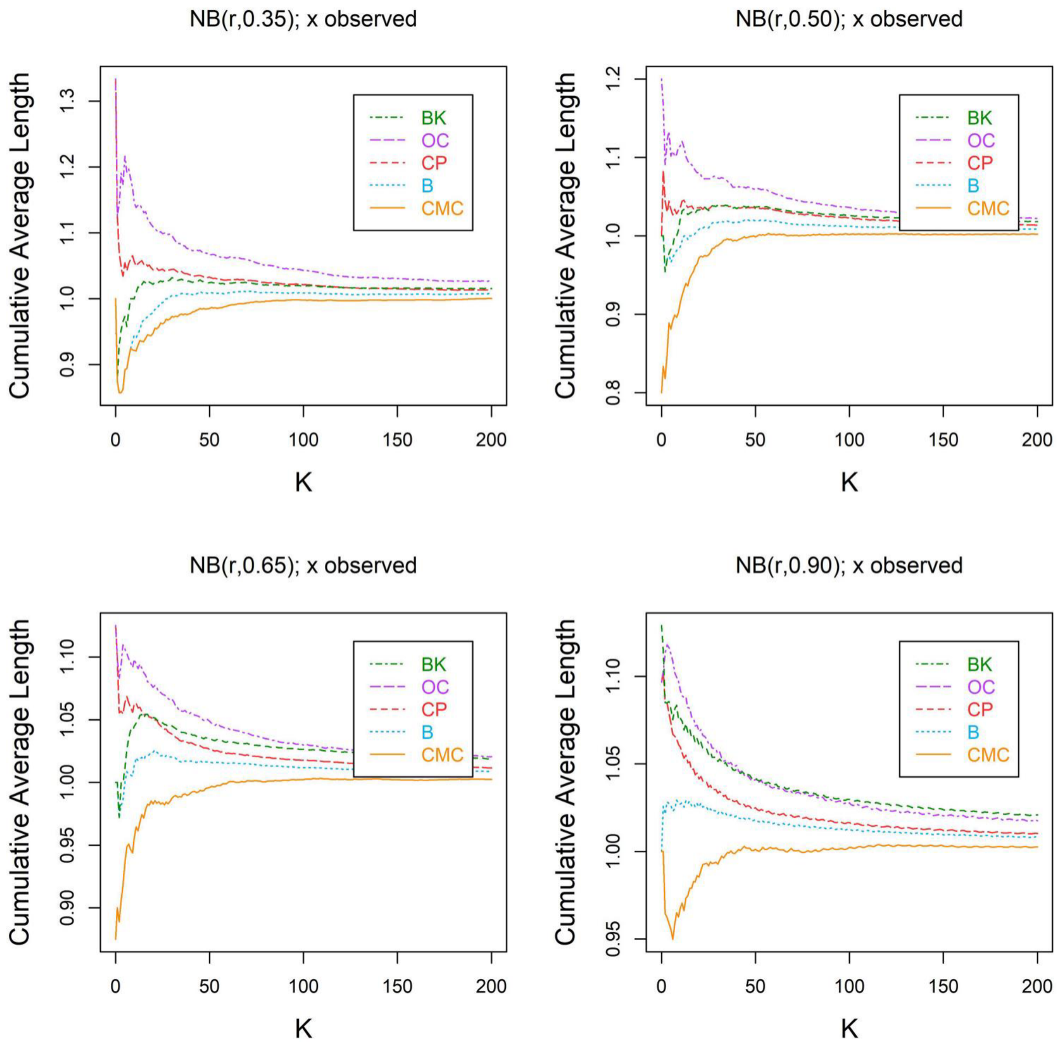

Figure 6. (a–d) 95% cumulative average length for $p = 0.35$, 0.50, 0.65 and 0.90 relative to CG in the negative binomial case when the number of failures is observed. For each K, the average length is calculated for all x values up to and including K.

than CG for all r less than some cutoff, with CG having smaller expected width elsewhere. When $p = 0.90$ CMC maintains a smaller expected width throughout the entire range provided, $0 \le r \le 100$.

4.3. Estimation of negative binomial r (N observed)

Figure 8(a–d) shows the cumulative average length relative to CG at 95% confidence for $p =$0.35, 0.50, 0.65 and 0.90 and $0 \le K \le 200$ in the negative binomial case when the total number of trials is observed. CG and CMC again have the shortest average length with CMC attaining a shorter average length for K roughly less than 100 when $p = 0.35$ and 0.90 and for K less than about 40 when $p =$ 0.50 and 0.65; CG attains the shortest average length elsewhere.

Figure 9(a–d) shows the expected length relative to CG at 95% confidence for $p =$ 0.35, 0.50, 0.65 and 0.90 and $0 \le r \le 100$ in the negative binomial case when the total

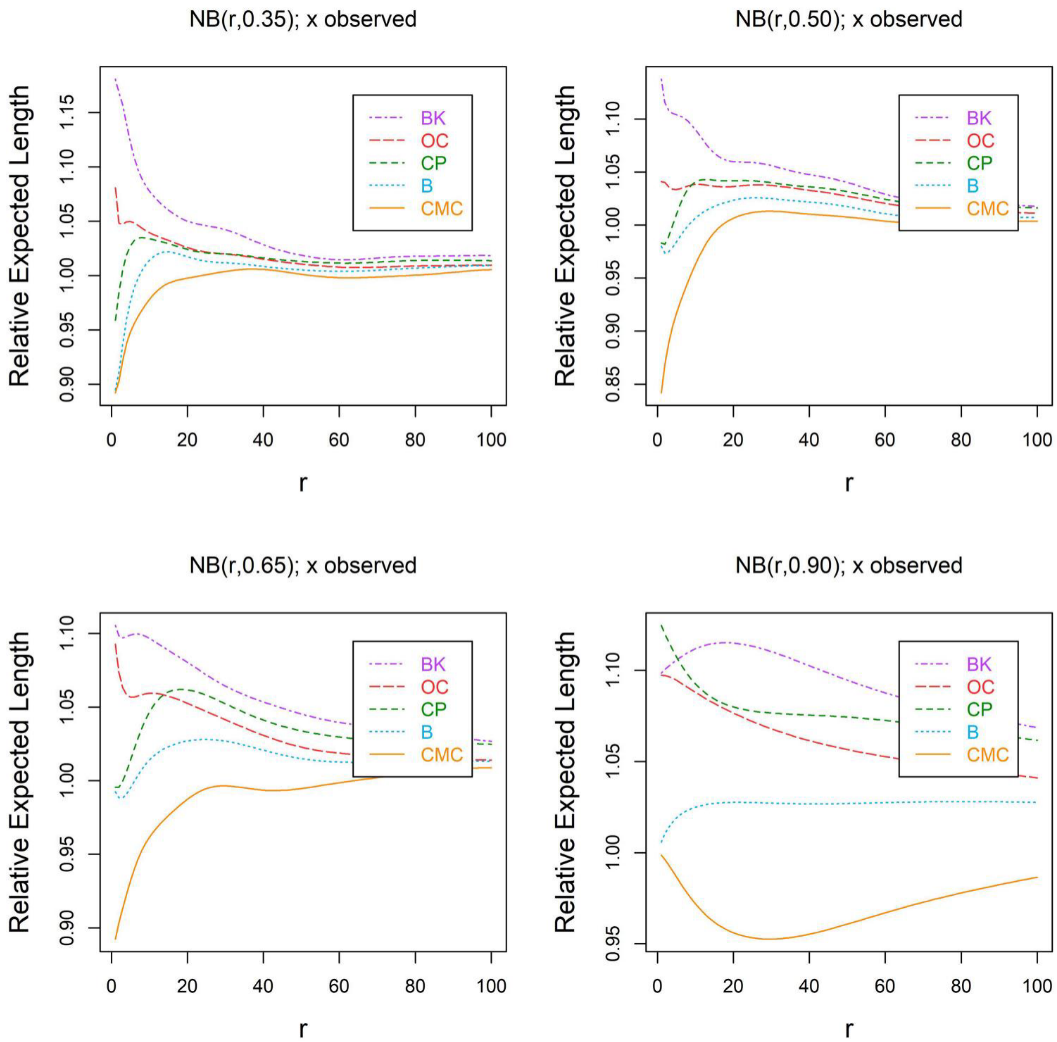

Figure 7. (a–d) 95% expected length for $p = 0.35$, 0.50, 0.65 and 0.90 relative to CG in the negative binomial case when the number of failures is observed.

number of trials is observed. In this case CG has the shortest expected length except for small values of r, and OC and B do almost just as well as CMC.

5. Conclusion

When estimating the sample size parameter of the binomial or hypergeometric distributions, or the count parameter of the negative binomial distribution, we desire a procedure that is strict and minimizes confidence interval length. Obtaining an interval estimate of the sample size parameter n of the hypergeometric distribution is immediate since the preexisting procedures that have been developed for estimating M (e.g., Schilling and Stanley 2022) can be applied directly for estimation of n by simply swapping the roles of n and M in the procedure. On the other hand, to handle the binomial and negative binomial distributions, we extended the most competitive strict procedures in the literature to these new cases. This includes the Byrne and Kabaila method (BK), the Clopper and Pearson method (CP), Schilling and Doi's Optimal

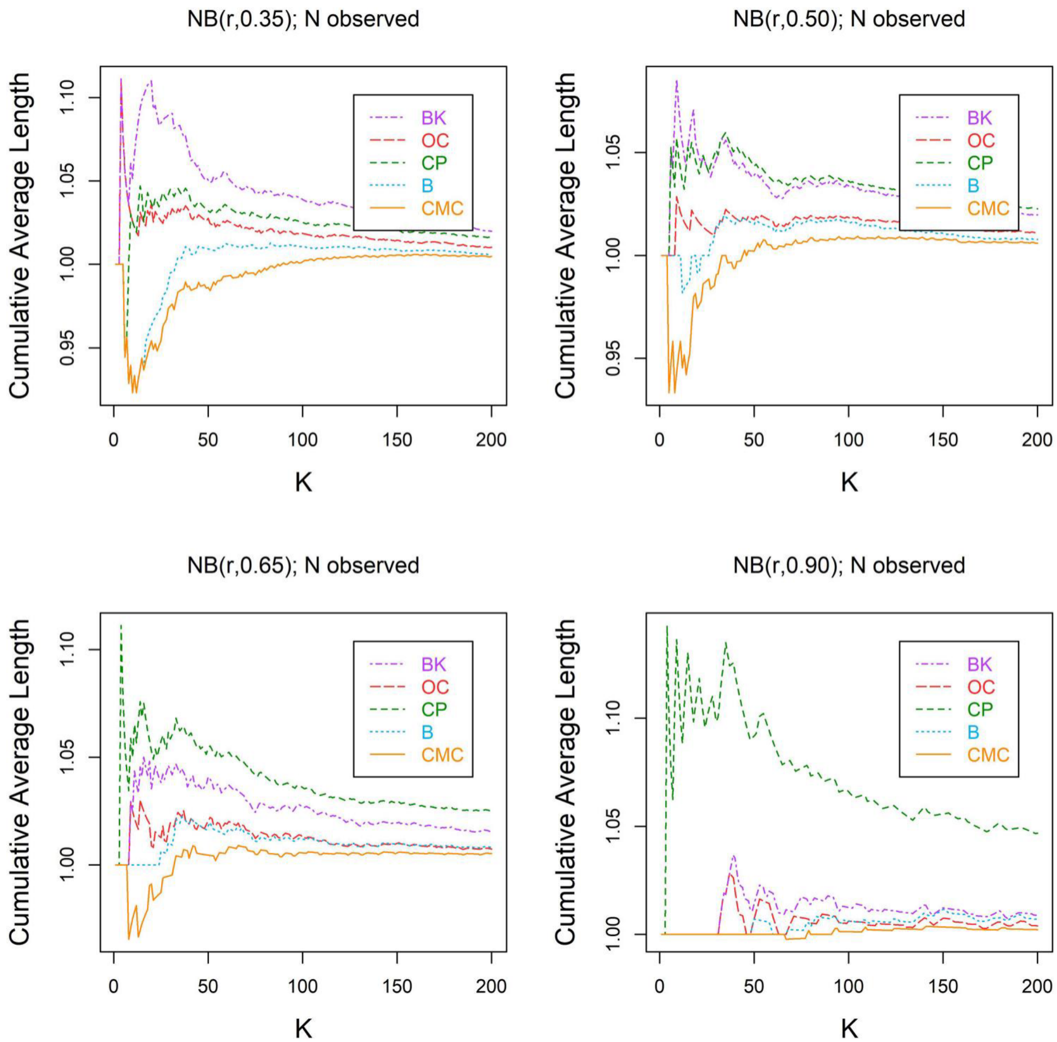

Figure 8. (a–d) 95% cumulative average length for $p = 0.35$, 0.50, 0.65 and 0.90 relative to CG in the negative binomial case when the total number of trials is observed. For each K, the average length is calculated for all N up to and including K.

Coverage method (OC), Blaker's method (B), the Crow and Gardener method (CG), and the Conditional Minimal Cardinality procedure (CMC) of Schilling, Holladay and Doi.

Upon comparing both the cumulative average width and the expected length of these procedures, we found that the Crow and Gardener method (CG) and the conditional minimal cardinality approach (CMC) perform the best in nearly every case. For the binomial case, both of these methods perform almost equally as well. For the negative binomial case, when the number of failures is observed, if it is suspected that r might be fairly small, then CMC is preferable to CG (and to all other methods considered) as judged by expected length. Of course, if there is more than weak prior information about what values of r are plausible, then a Bayesian approach (perhaps using credible intervals) may be preferable. Note that credible intervals come with their own set of issues due to the posterior distributions themselves being discrete. This includes the fact that it will not usually be possible to find an interval with a specified credible probability.

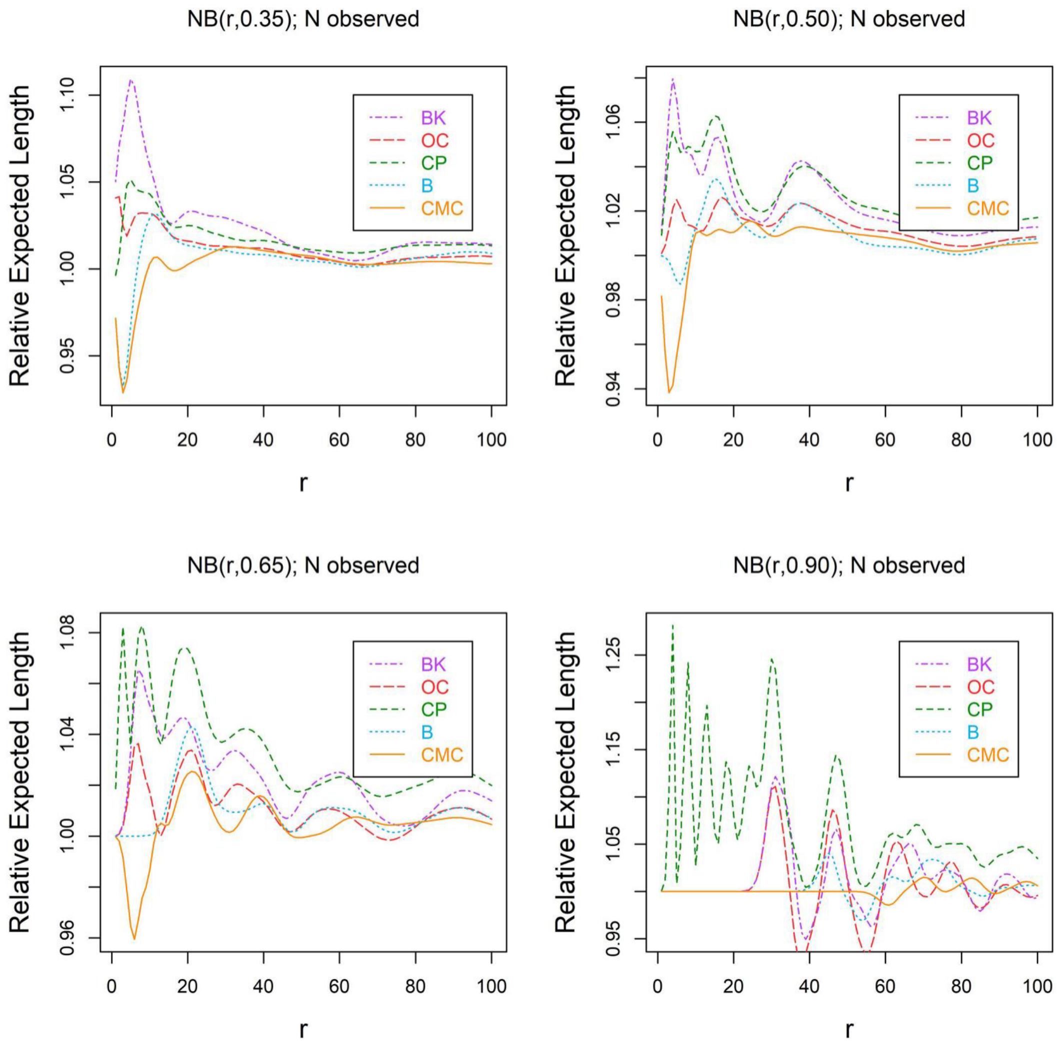

Figure 9. (a–d) 95% expected length for $p = 0.35$, 0.50, 0.65 and 0.90 relative to CG in the negative binomial case when the total number of trials is observed.

In the locations where CG had shorter expected length, (at least through $r = 100$) CMC performed as well or nearly as well on this measure as CG in every case investigated. Results for cumulative average length showed that for small x (which is consistent with a small value of r), confidence intervals produced by CMC tend to be shortest, with CMC only doing slightly worse than CG for larger x. In the negative binomial case, when the total number of trials N is observed, CG is preferable to CMC except at small values of r, where CMC shines.

6. Technical notes

1. All computations were performed in the statistical software R. A Shiny web app for determining confidence intervals for the count parameters of the binomial and negative binomial distributions is available at https://discrete-ci.shinyapps.io/count_ci/. Confidence intervals for the sample size parameter n of the hypergeometric distribution can be obtained from Schilling and Stanley (2022) shiny app for estimating the success count M (https://github.com/mfschilling/HGCIs), by swapping the roles of M and n.

2. All confidence interval computations are relatively fast. For instance, for the binomial and both versions of the negative binomial, determining the first 100 intervals at a 90% confidence level with $p = 0.5$ for both CG and CMC takes no more than a couple of seconds on a standard home desktop computer.
3. At each value of the parameter, the Optimal Coverage method (OC) uses the candidate minimal cardinality acceptance curve with the highest coverage. When two or more acceptance curves $\{P_\theta(a \leq X \leq b)\}$ are tied for the highest coverage, we define OC to be the one using the larger value of a.
4. In the negative binomial case, when p is sufficiently small (e.g., p less than about 0.04), the confidence intervals for small x (or small N if the number of trials is observed), will sometimes be the empty set for each of CP, B, and CMC. This is a result of the procedures using an acceptance set a-b with $a > 0$ (or $a > 1$ when the number of trials is observed) when $r = 1$. This causes the confidence intervals for $x = 0, 1, \ldots, a-1$ (or $N = 1, 2, \ldots, a-1$) to be empty. We recommend the use of CG in such cases as it avoids this issue by using acceptance sets of the form 0-b (or 1-b if the number of trials is observed) when $r = 1$.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Bayoud, H. A. 2011. Bayes and empirical bayes estimation of the parameter n in a binomial distribution. *Communications in Statistics - Simulation and Computation* 40 (9):1422–33. doi: 10.1080/03610918.2010.526737.

Blaker, H. 2000. Confidence curves and improved exact confidence intervals for discrete distributions. *Canadian Journal of Statistics* 28 (4):783–98. doi: 10.2307/3315916.

Blumenthal, S., and R. C. Dahiya. 1981. Estimating the binomial parameter n. *Journal of the American Statistical Association* 76 (376):903–9. doi: 10.2307/2287586.

Blyth, C. R., and H. A. Still. 1983. Binomial confidence intervals. *Journal of the American Statistical Association* 78 (381):108–16. doi: 10.1080/01621459.1983.10477938.

Byrne, J., and P. Kabaila. 2005. Comparison of Poisson confidence intervals. *Communications in Statistics - Theory and Methods* 34 (3):545–56. doi: 10.1081/STA-200052109.

Casella, G. 1986. Refining binomial confidence intervals. *Canadian Journal of Statistics* 14 (2):113–29. doi: 10.2307/3314658.

Cheng, S., D. J. Eck, and F. W. Crawford. 2020. Estimating the size of a hidden finite set: Large-sample behavior of estimators. *Statistics Surveys* 14:1–31. doi: 10.1214/19-SS127.

Clopper, C. J., and E. S. Pearson. 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26 (4):404–13. doi: 10.1093/biomet/26.4.404.

Crow, E. L. 1956. Confidence intervals for a proportion. *Biometrika* 43 (3-4):423–35. doi: 10.1093/biomet/43.3-4.423.

Crow, E. L., and R. S. Gardner. 1959. Confidence intervals for the expectation of a Poisson variable. *Biometrika* 46 (3-4):441–53. doi: 10.1093/biomet/46.3-4.441.

De, S. K., and S. Zacks. 2016. Two-stage and sequential estimation of parameter N of binomial distribution when p is known. *Sequential Analysis* 35 (4):440–52. doi: 10.1080/07474946.2016.1238255.

Drago, I., R. Hofstede, R. Sadre, A. Sperotto, and A. Pras. 2015. Measuring cloud service health using NetFlow/IPFIX: The WikiLeaks case. *Journal of Network and Systems Management* 23 (1):58–88. doi: 10.1007/s10922-013-9278-0.

Draper, N., and I. Guttman. 1971. Bayesian estimation of the binomial parameter. *Technometrics* 13 (3):667–73. doi: 10.1080/00401706.1971.10488827.

Feldman, D., and M. Fox. 1968. Estimation of the parameter n in the binomial distribution. *Journal of the American Statistical Association* 63 (321):150–8. doi: 10.2307/2283836.

Ganji, M., N. Eghbali, and M. Azimian. 2013. Bayes and empirical bayes estimation of parameter k in negative binomial distribution. *Journal of Hyperstructures* 2 (2):185–200.

Iliopoulos, G. 2003. Some new estimators of the binomial parameter n. *Communications in Statistics - Theory and Methods* 32 (7):1361–72. doi: 10.1081/STA-120021563.

Johnson, B. R. 1981. The search for submarine pingos in the Beaufort sea: An unusual application of the binomial distribution. *Canadian Journal of Statistics* 9 (2):209–14. doi: 10.2307/3314614.

Schilling, M. F., and A. Stanley. 2022. A new approach to precise interval estimation for the parameters of the hypergeometric distribution. *Communications in Statistics - Theory and Methods* 51 (1):29–50. doi: 10.1080/03610926.2020.1737879.

Schilling, M. F., and B. Holladay. 2017. Length minimization for Poisson confidence procedures. *Communications in Statistics - Theory and Methods* 46 (2):861–73. doi: 10.1080/03610926.2015.1006782.

Schilling, M. F., and J. A. Doi. 2014. A coverage probability approach to finding an optimal binomial confidence procedure. *The American Statistician* 68 (3):133–45. doi: 10.1080/00031305.2014.899274.

Schilling, M. F., B. Holladay, and J. A. Doi. 2023. Improved interval estimation of negative binomial parameters: A coverage probability approach. *Journal of Statistical Computation and Simulation* 94 (1):126–41.

Stapleton, M. 2023. The beauty of binomial coverage curves. *Significance* 20 (2):26–7. doi: 10.1093/jrssig/qmad028.

Sterne, T. E. 1954. Some remarks on confidence of fiducial limits. *Biometrika* 41 (1-2):275–8.

Tang, V. K. T., and R. B. Sindler. 1987. Confidence interval for parameter n in a binomial distribution. Tech. Rep. CRM, 86–265. Center for Naval Analyses. https://apps.dtic.mil/sti/citations/ADA178325