JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
2024, VOL. 94, NO. 1, 126–141
https://doi.org/10.1080/00949655.2023.2235046

Taylor & Francis
Taylor & Francis Group

Improved interval estimation of negative binomial parameters: a coverage probability approach

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

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

ABSTRACT

We present a numeric search algorithm to determine confidence intervals for the parameters of the negative binomial distribution that are substantially more precise (shorter) than those of any existing method. We derive confidence intervals for the mean μ for any specified number of successes k by means of a conditional minimal cardinality approach that efficiently accommodates the relationship of the variability of the negative binomial distribution to its mean. Confidence intervals for the success parameter p are easily obtained from the inverse relationship between μ and p.

ARTICLE HISTORY
Received 23 December 2022
Accepted 5 July 2023

KEYWORDS
Confidence interval;
Coverage probability;
Prevalence rate

MSC 2010
Primary 62F25

1. Introduction

In this paper we present a new approach to interval estimation that we apply to obtain short confidence intervals based on a sample from a negative binomial distribution. Our method is based on an efficient search through the candidate acceptance sets that a negative binomial confidence procedure can have. The method is *strict* in the sense of satisfying the coverage condition $\inf_\theta P(\theta \in CI) \geq 1-\alpha$, and results in confidence intervals that tend to be shorter than those of existing methods. (Note: The term *exact* is often used rather than *strict*; however this may lead to confusion since the term *exact* can also refer to the fact that the confidence procedure is derived from the actual distribution involved – here, the negative binomial – rather than from, say, a normal approximation.)

Much attention has been given to the problem of determining short confidence intervals for the parameters of the binomial and Poisson distributions, and length minimizing procedures have been obtained for both cases (see Sterne [1]; Crow [2]; Crow and Gardner [3]; Casella [4]; Schilling and Doi [5]; Schilling and Holladay [6]). Recent work (Wang [7]; Schilling and Stanley [8]) has produced short confidence intervals for the parameters of the hypergeometric distribution. No such confidence procedures have been developed for the parameters of the negative binomial distribution, however.

We take the negative binomial random variable X to be the number of failures before the kth success, with success probability p. In medical applications p is known as the *prevalence*

CONTACT Mark F. Schilling mark.schilling@csun.edu Department of Mathematics, California State University Northridge, Northridge, CA 91330, USA

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

rate. Note that X can also represent the sufficient statistic $\sum_{i=1}^{n} X_i$ obtained from a random sample of i.i.d. geometric or negative binomial random variables, since this sum is also negative binomially distributed.

Our primary objective is to develop a strict $1-\alpha$ level confidence procedure for the mean $\mu = E(X) = k(1–p)/p$; that is, an infinite set of intervals (l_x, u_x), $x = 0, 1, \ldots$ for which $\inf_\mu P(\mu \in [l_X, u_X)) \geq 1-\alpha$. (We use half-open intervals to avoid certain technical difficulties that arise with fully open or fully closed intervals.) Confidence intervals for p are readily obtained from the inverse relationship $p = 1/(1+\mu/k)$.

To see how one can approach the problem at hand, it is instructive to look at Figure 1, in which estimation of the parameter λ of the Poisson distribution at 95% confidence is the goal. This situation is similar to estimation of the negative binomial parameter μ in that the parameter space $(0, \infty)$ is the same.

Each curve labelled a-b in Figure 1 is a portion of the graph of the function $P_{ab}(\lambda) = P(a \leq X \leq b)$, where X has a Poisson(λ) distribution. The set of curves shown are those that have *minimal cardinality* for each λ; that is, those for which $b-a+1$ is as small as possible among curves that rise above the confidence level. We refer to the set of values a, $a+1$, $a+2, \ldots$, b–1, b as the *acceptance set* $AS(a\text{-}b)$. This terminology derives from the duality between confidence procedures and hypothesis tests; see for example Theorem 9.2.2 of Casella and Berger [9]. The *coverage probability function* (*cpf*) of a confidence procedure is comprised of the *acceptance curve* segments $AC(a\text{-}b) = P_{ab}(\lambda)$ associated with the particular acceptance sets used for each λ, and this cpf *determines* the corresponding confidence procedure (Schilling and Doi [5]). The property that the a and b sequences each be monotone is essential in order that none of the resulting confidence intervals contain gaps.

For small values of the parameter in both the Poisson and negative binomial cases, X is also likely to be small. Thus the acceptance sets for these parameter values consist of small nonnegative integers. As the parameter increases, the values of a and b are each monotonically nondecreasing, with the a value increasing by one at any transition from one acceptance curve to another. As the cardinalities of the acceptance sets gradually increase with the parameter, the b value sometimes increases by more than one at such a transition.

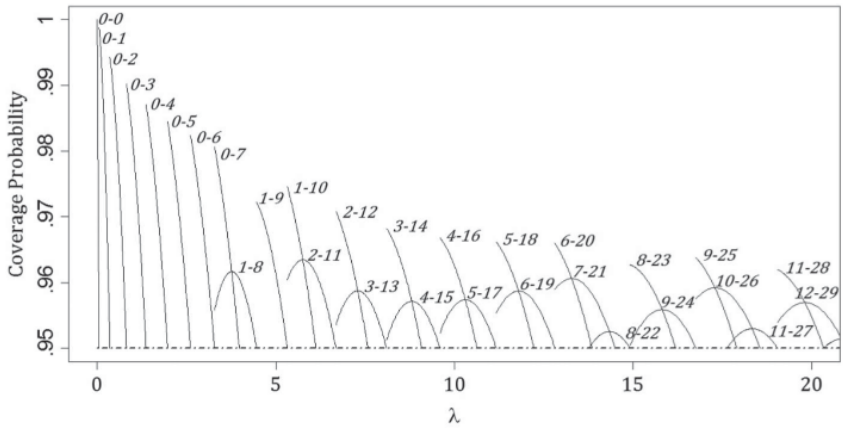

Figure 1. Acceptance sets for a strict Poisson 95% confidence procedure.

These observations form the basis for our search for a short negative binomial confidence procedure. In the binomial and Poisson cases, choosing appropriate acceptance curve segments corresponding to minimal cardinality acceptance sets produces confidence procedures whose confidence intervals minimize overall interval length in a suitably defined sense (see Schilling and Doi [5]; Schilling and Holladay [6]). In the case of the negative binomial distribution, however, this approach does not lead to short confidence intervals for μ unless k is large. We illustrate the reason for this in Section 6.

We show below that our procedure generates much shorter confidence intervals than any existing negative binomial confidence procedure.

2. An approach to finding a negative binomial confidence procedure for μ yielding short confidence intervals

Our strategy for finding a negative binomial confidence procedure that yields short intervals follows the method described by Bain and Engelhardt [10] and others. We construct two nondecreasing, right continuous, integer-valued functions $a(\mu)$, $b(\mu)$ defined on $\mu > 0$ satisfying.

$$P_\mu(a(\mu) \leq X \leq b(\mu)) \geq 1\text{–}\alpha \quad \forall \mu > 0,$$

with $l_x = \inf\mu\colon b(\mu) \geq x$, $u_x = \sup\mu\colon a(\mu) \leq x$, and the goal of making the resulting confidence intervals $[l_x, u_x)$, $x = 0, 1, 2, \ldots$ as small as possible. The above condition ensures that the confidence procedure is strict.

To implement our approach for given k and confidence level $1\text{–}\alpha$, we use a grid search of values of μ beginning at 0 with spacing Δ, constructing an array M whose rows are indexed by the values of μ in $\Delta, 2\Delta, 3\Delta, \ldots$ and whose columns are indexed by the values $a = 0, 1, 2, \ldots$. Each entry $M(\mu, a)$ of M is the smallest value of $b \in 0, 1, 2, \ldots$ for which $P_{ab}(\mu) = P(a \leq X \leq b) \geq 1\text{–}\alpha$. If no such value exists the entry is left empty.

The entries in the $a = 0$ column are nondecreasing as μ increases, while the situation is a bit more complicated for $a > 0$. Typically the initial values of each column with $a > 0$ will be empty. This is because when μ is small, no value of b results in $P_{ab}(\mu) \geq 1\text{–}\alpha$ since too much (more than α) of the probability content of the distribution of X is concentrated on the values of $X < a$. At the first value of μ which allows $P_{ab}(\mu) \geq 1\text{–}\alpha$, the value of $b = M(\mu, a)$ is necessarily very large since $P(X < a)$ is nearly α. From this value of μ on, the column values of $M(\mu, a)$ are nonincreasing until reaching a minimum value $m(a)$, then are nondecreasing thereafter. It is also clear that for fixed μ the row values of M are nondecreasing as a increases.

As an illustration of the column behaviour of M described above, consider Figure 2, which shows the negative binomial probability mass function (pmf) for $k = 10$ and $a = 4$. Suppose $\alpha = 0.05$. When μ is sufficiently small a significant portion of the mass ($> \alpha$) is concentrated on lower values of X, which makes it impossible for $M(\mu, a)$ to be defined. Figure 2(a), which displays the pmf for $\mu = 4.50$, illustrates this point. It can be helpful to think of the pmf form as a wave, where the wave's crest's lateral position increases with μ. When μ is small, the proximity of the crest to a makes $P(X < a)$ large. In Figure 2(a), $P(X < a) = 0.3889$; hence $P(X \geq a) = 0.6111$ and thus there is no b such that $P(a \leq X \leq b) \geq 1\text{–}\alpha = 0.95$.

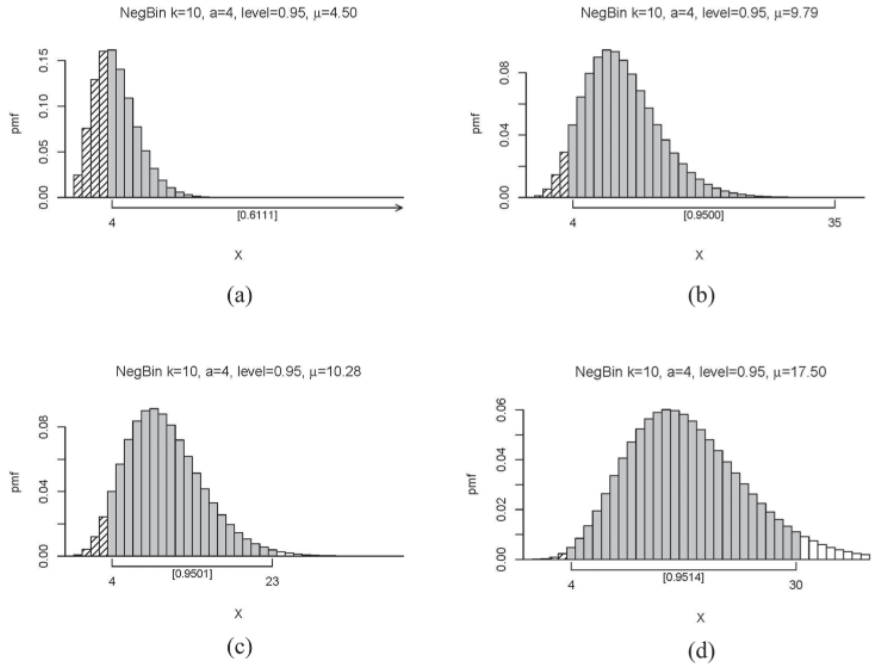

Figure 2. Negative binomial probability mass functions for $k = 10$, $a = 4$ and (a) $\mu = 4.50$, (b) $\mu = 9.79$, (c) $\mu = 10.28$, (d) $\mu = 17.50$. Crosshatched bars correspond to $P(X < a)$, grey bars correspond to $P(a \leq X \leq M(\mu, a))$, and white bars correspond to $P(X > M(\mu, a))$.

In Figure 2(b) μ is increased to 9.79, and we can see a shift in the pmf, with the crest of the wave moving sufficiently to the right that $P(X < a)$ is now slightly under 0.05. Due to the fact that $P(X \geq a)$ is now barely above 0.95, the smallest b such that $P(a \leq X \leq b) \geq 0.95$ is large. In this case $b = M(\mu, a) = 35$. As μ becomes slightly larger, the corresponding pmf looks similar to that shown in Figure 2(b); however the increase in μ results in a decrease in $P(X < a)$, transferring more of the mass towards the central portion of the pmf; therefore $M(\mu, a)$ need not be as large as before. At some particular μ the wave is located at a sweet spot where $M(\mu, a)$ achieves a minimum. In the present example, $M(\mu, a)$ reaches a minimum $m(a) = 23$ when $\mu = 10.28$. This is shown in Figure 2(c).

When μ is greater than this optimal value, the right tail enlarges and the pmf has sufficient mass for larger values of X to cause $M(\mu, a)$ to increase. As the wave continues to shift to the right, the probability content at smaller X values diminishes and it is necessary to reach further into the upper tail region in order to find the smallest b where $P(a \leq X \leq b) \geq 0.95$. See Figure 2(d).

The proposed confidence procedure is obtained from M as follows: Begin in the $a = 0$ column. Specifically, beginning at $\mu = \Delta$ and moving down through M, choose for each value of μ the acceptance set $AS(a\text{-}b)$ indicated by the current column a and the given value of b in M. Increment μ until reaching $\mu_a =$ the smallest value of μ for which the value of b in column $a + 1$ achieves its minimum, $m(a + 1)$. At that point switch to the acceptance set $AS((a + 1)\text{-}b)$ and proceed in the same fashion as before indefinitely.

Consider the rationale for this process: For small μ no acceptance sets other than those of the form $AS(0\text{-}b)$ are possible, and using for any μ a larger value of b than that listed in M would lead to unnecessarily large confidence intervals, since if $b' > M(\mu, a)$ then using $AS(0\text{-}b')$ rather than $AS(0\text{-}M(\mu, a))$ puts the lower confidence limits of the confidence intervals for $x = b+1, b+2, \ldots, b'$ at μ instead of at higher values. Transitioning from any a to $a+1$ where $AS((a+1)\text{-}b)$ uses a minimal value of b has a similar benefit by delaying the start (i.e. increasing the lower confidence limits) of the confidence intervals for $b+1, b+2, \ldots$. The logic behind the subsequent transitions is similar. Although we do not restrict to minimal cardinality acceptance sets overall, for each fixed a we do indeed choose the acceptance set with minimal cardinality at each μ. Consequently, we refer to our method as a *conditional minimal cardinality* (CMC) confidence procedure.

Now consider further what happens when there is a transition from one column to the next. Switching from column a to column $a+1$ at μ_a marks μ_a as the upper confidence limit for $x = a$, and also establishes μ_a as the lower confidence limit for all x between $M(\mu_a, a)+1$ and $m(a+1)$. Points where the upper confidence limit for one x is the lower limit for another x are called *coincidental endpoints* (Casella [4]).

Transitioning earlier than where $b = m(a+1)$ would lead to gaps in the confidence intervals for values of b in the new column that exceed $m(a+1)$. Consider also the option of transitioning at a somewhat later value of μ where b is still equal to $m(a+1)$. At first glance this might appear to be a better strategy. In many cases the value $m(a+1)$ would be more than one larger than the value of b, say b_a, used for the μ value immediately before the transition. This would result in larger lower confidence limits and therefore potentially shorter confidence intervals for $x = b_a+1, b_a+2, \ldots, m(a+1)$ at the expense of a larger upper confidence limit and therefore a longer confidence interval for only $x = a$. However, delaying the transition in this way actually gives worse results because the confidence intervals for the values of $x > b_a$ above will also have larger upper limits when the transition is made from their columns, and the increases in those upper confidence limits exceeds the increases in their lower confidence limits resulting from the earlier transition. Near the end of Section 6 we provide an illustration that clarifies this phenomenon.

3. A CPF perspective

We provide here a description of our procedure from a cpf perspective. Given a, let $B(a)$ be the set containing all $b \geq a$ such that $P_\mu(a \leq X \leq b) \geq 1-\alpha$. The distinct non-empty entries of M in the a column then correspond to the set of curves obtained from the functions $RB(a) = \{P_\mu(a \leq X \leq b) : b \in B(a)\}$. Plotting this collection of arcs makes a rainbow-like picture as shown in Figure 3 for $k = 10$, $1-\alpha = 0.95$ and $a = 0, 3, 15$.

Recall that for fixed column a, $m(a)$ is the smallest b such that $\max_\mu\{P_\mu(a \leq X \leq b)\} \geq 1-\alpha$. Since $P_\mu(a \leq X \leq b) < P_\mu(a \leq X \leq b+1)$ for all μ, the function $P_\mu(a \leq X \leq m(a))$ will be the *core* (interior arc) of each rainbow. See for example the middle plot in Figure 3. The elements of $RB(3)$ shown, in order from the top arc to the core, are the functions $P_\mu(3 \leq X \leq b)$ for $b = 23, 22, 21, 20$, and 19 (the core). Thus as μ increases through the range of values shown in the plot, $M(\mu, 3)$ decreases from 23 to 19 and then increases to 23 again.

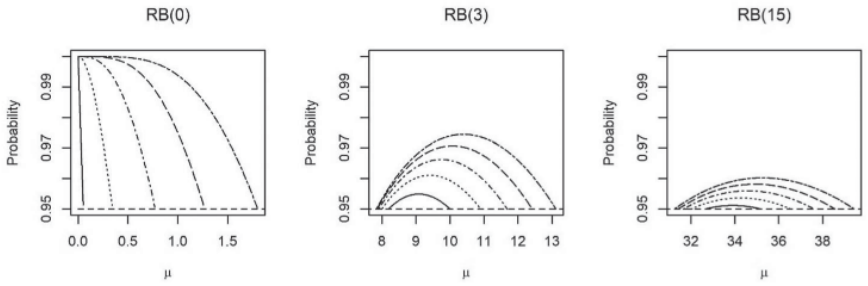

Figure 3. Sample rainbow plots for $k = 10$, $1 - \alpha = 0.95$ and $a = 0$, 3, and 15.

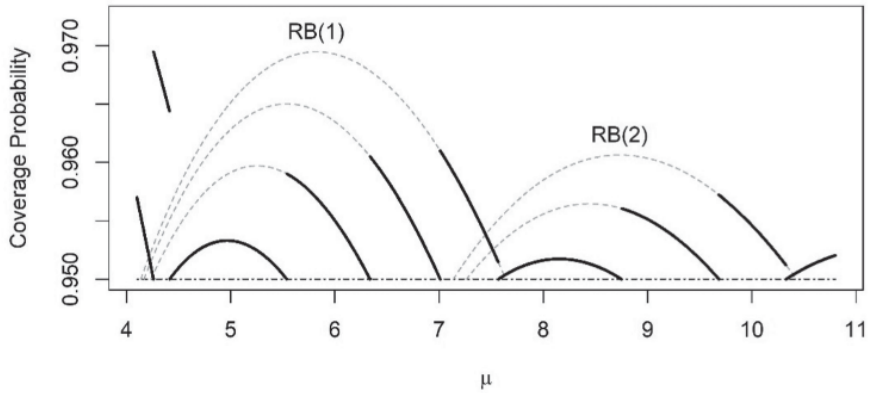

Figure 4. Illustration of how the cpf of the CMC method transitions between rainbows when $k = 5$ and $1-\alpha = 0.95$. RB(1) is used from $\mu = 4.41$ to $\mu = 7.57$, at which point the cpf transitions to RB(2), which is used until $\mu = 10.33$. Overlaid solid lines represent the portion of the rainbows belonging to the cpf.

We can describe our CMC procedure from a CPF perspective as follows: For each μ in $\{\Delta, 2\Delta, \ldots\}$ use only acceptance curves from $RB(0)$, until switching to the core of $RB(1)$ when it first rises above the confidence level. While each$RB(a)$ is in use, choose for each successive μ the lowest curve in that rainbow having $P_\mu(a \leq X \leq b) \geq 1 - \alpha$ until switching to the core of $RB(a + 1)$ once it rises above the confidence level, continuing this process indefinitely. See Figure 4. A detailed description of the algorithm is provided in Appendix.

4. Problems with minimal cardinality for negative binomial estimation

For other discrete distributions such as the binomial, Poisson and hypergeometric, optimal interval estimation is realized by using (unconditional) minimal cardinality acceptance curves throughout the entire parameter space. Schilling and Holladay [6] showed, for example, that for estimation of the parameter of a Poisson distribution, the method of Crow and Gardner [3] is the unique length minimizing confidence procedure when using a running average interval length criterion. Crow and Gardner's (CG) approach is to use only minimal cardinality acceptance curves. Whenever there are multiple curves $AC(a\text{-}b)$ of minimal cardinality available, CG uses the one with largest values of a and b.

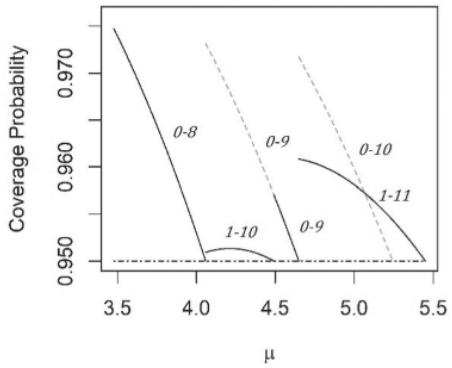

Figure 5. The Crow-Gardner method yields a gap in the confidence interval for $x = 0$ when $k = 9$.

Applying this approach to estimation of the negative binomial mean, however, regularly produces gaps in the resulting confidence intervals. Figure 5 illustrates a specific instance where the cpf of CG for $k = 9$ and $1–\alpha = 0.95$ produces a gap. The figure shows all minimal cardinality curves in the given range. Solid black lines represent CG's cpf. Notice that the cpf transitions from $AC(0\text{-}8)$ to $AC(1\text{-}10)$, and then to $AC(0\text{-}9)$. The result is that the confidence interval for $x = 0$ has a gap corresponding to the interval where the cpf follows $AC(1\text{-}10)$.

CG's choice of acceptance curves will often result in a transition from curve $AC(a\text{-}b)$ to $AC((a–1)\text{-}(b–1))$. This causes a gap in the confidence interval for $x = a–1$, since the previously generated confidence interval for $a–1$ will have the subinterval μ: $cpf(\mu) = AC((a–1)\text{-}(b–1))$ appended to it.

To resolve the gap problem while preserving minimal cardinality acceptance regions we can require that an acceptance curve $AC(a\text{-}b)$ only be used if all other curves $AC(a'\text{-}b')$ of the same cardinality with $b' < b$ and $a' < a$ fall below the confidence level before $AC(a\text{-}b)$ does. This requirement would make $AC(1\text{-}10)$ ineligible in the example shown. The resulting modified procedure uses $AC(0\text{-}9)$ in place of $AC(1\text{-}10)$, resolving the gap. We refer to this approach as the *Adjusted Crow-Gardner method* (ACG).

In the case $k = 1$ (the geometric distribution), however, each acceptance curve $AC(0\text{-}b)$ remains above $AC(1\text{-}(b + 1))$ for all μ. As a result, the cpf for ACG will ipso facto be made up entirely of acceptance curves from $AC(0\text{-}b)$, $b \geq 0$; consequently, all of the resulting confidence intervals have no upper limit. This may be a suitable result if one is interested in a lower confidence bound for μ, but produces confidence intervals of infinite length rather than short ones. Furthermore, it is shown in what follows that ACG is universally inferior in length performance to our CMC procedure, often by a substantial amount.

5. Conditional minimal cardinality: some results

Figure 6 shows an example of the coverage probability function of the CMC confidence procedure for the case when $k = 5$ and $1–\alpha = 0.95$ for $\mu \leq 40$. Note that the cpf stays close to the confidence level and in fact equals it frequently. This property implies that the resulting confidence procedure generates short confidence intervals.

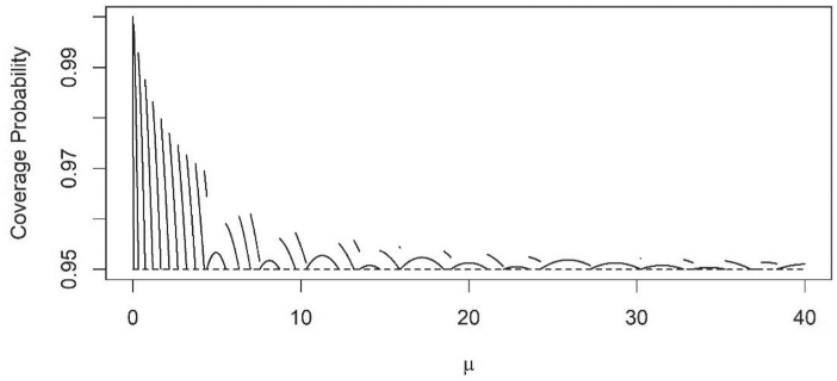

Figure 6. Coverage probability function for the CMC method when $k = 5$ and $1 - \alpha = 0.95$.

Table 1. Runtimes for the CMC algorithm.

k	Runtime (seconds)
1	29.68
2	2.14
3	0.90
4	0.61
5	0.49
6	0.45
7	0.37
8	0.36
9	0.37
10	0.31

To provide the reader with a rough idea of the runtime for our algorithm, we ran our code on a home desktop computer to produce 95% confidence intervals for all $0 \leq x \leq 100$ for varying values of k. Table 1 shows the resulting runtimes for $1 \leq k \leq 10$, while Figure 7 shows the resulting runtimes for $2 \leq k \leq 50$. With the exception of $k = 1$ and 2 all runtimes were a fraction of a second. Runtimes are longer for the smallest values of k because the upper confidence limits are very large, necessitating evaluation of the M array for large values of μ.

6. Comparison to existing procedures

Casella and McCulloch [11] developed the first explicit strict procedure for estimating a parameter of a negative binomial distribution; also determined by Lui [12]. This procedure is the natural counterpart to the Clopper–Pearson method for the binomial distribution, obtained by inverting the equal tailed level α test of H_0: $p = p_0$ to yield a confidence interval for p. Using the relationship between p and μ given earlier allows conversion of their confidence intervals into confidence intervals for μ.

Blaker [13] introduced a new approach for interval estimation for discrete distributions that can be applied to the negative binomial distribution as follows. For any particular value of the parameter to be estimated (μ or p), determine the tail probability generated by the

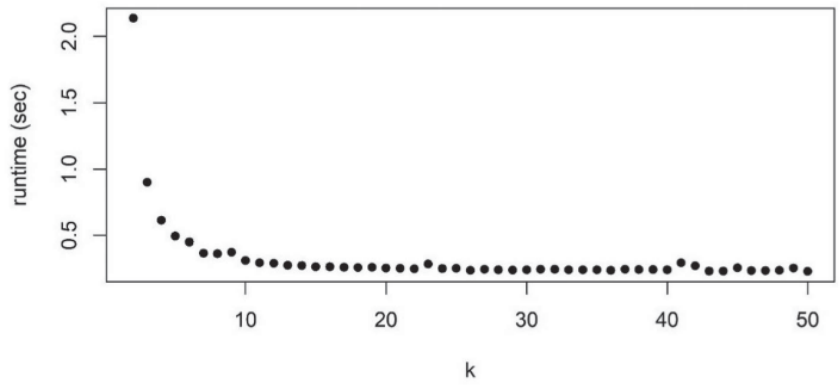

Figure 7. Runtimes to compute 95% confidence intervals for μ for all $0 \leq x \leq 100$.

observed data value x, $T = \min(P(X \leq x), P(X \geq x))$. Add this probability to the largest tail probability in the opposite tail (if any) not exceeding T. If the resulting sum is greater than α, then the given parameter value is included in the confidence interval for x.

Byrne and Kabaila [14] also developed a method for interval estimation for a parameter of a discrete distribution. Their method, in brief, is analogous to Crow and Gardner's method in that it also restricts to minimal cardinality sets, but transitions as late as possible while keeping the sequences of $\{a\}$ and $\{b\}$ each nondecreasing.

Blyth and Still [15] proposed a procedure for the binomial case that transitions at points intermediate to those where the Crow-Gardner and Byrne-Kabaila procedures transition. Schilling and Doi [5] developed a gapless modification of a method proposed by Sterne [1]. Their LCO (Length/Coverage Optimal) procedure transitions at points similar to where the Blyth and Still procedure does and generally gives similar confidence intervals, but has the property of maximizing coverage among all possible methods that minimize average confidence interval length.

We examined the performance of the various procedures described above and found that all of the methods that rely on minimal cardinality acceptance sets perform similarly, and all produce confidence intervals that are significantly longer than those obtained from the methods of Casella/McCulloch (CM), Blaker (B), and the CMC procedure.

Figure 8(a)–(d) displays the lengths of the intervals obtained from CM, B and ACG relative to those of the CMC procedure for $k = 1$, 2, 5 and 10 at 95% confidence for $0 \leq x \leq 100$. Results for Adjusted Crow/Gardner (ACG) are shown for $k = 5$ and 10 only, as for $k = 1$ ACG interval lengths are infinite and for $k = 2$ they are not feasible to include in the same graph since they are generally more than seven times as long as those of the CMC method. Figure 9(a)–(d) displays corresponding results for expected 95% confidence interval length. We also examined the relative performance of these methods for the same values of k for confidence levels 90% and 99% and found the results to be very similar to those for 95%.

Figures 8(a)–(d) and 9(a)–(d) demonstrate that the CMC method produces substantially shorter confidence intervals than the methods of either Casella and McCulloch and Blaker, and are therefore much shorter than those of any existing method. For instance for

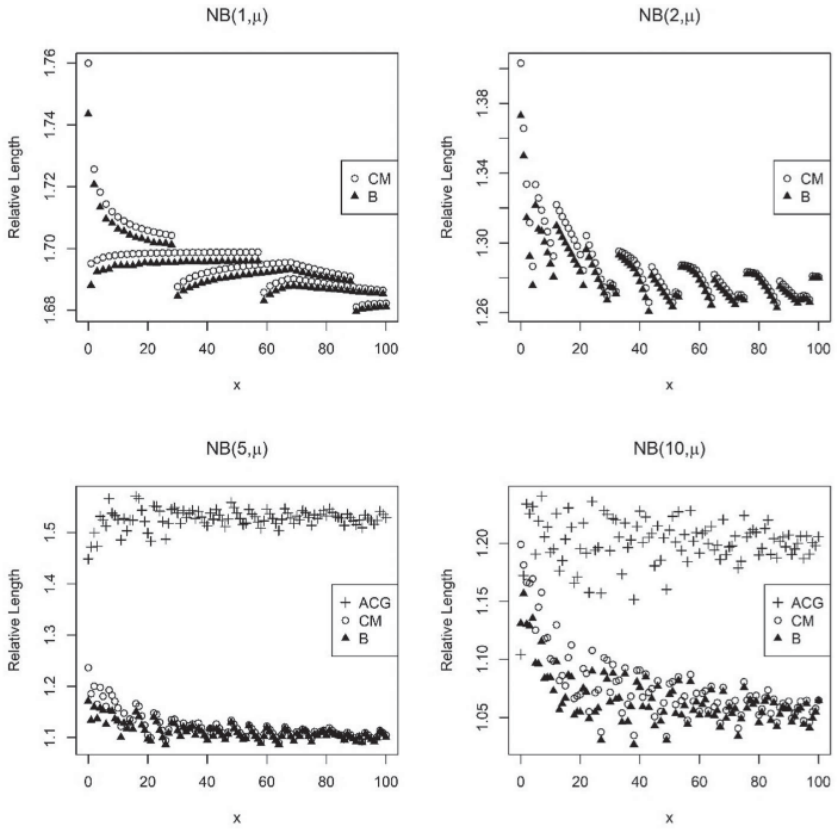

Figure 8. (a)–(d) 95% confidence interval lengths of competing methods for estimation of μ relative to those of obtained from the CMC confidence procedure for $k = 1, 2, 5$ and 10.

$k = 5$ both CM and B have an expected 95% confidence interval length more than 10% longer than that of the CMC method for all $\mu \in (0,100)$; for smaller k the length savings of the CMC method are even more dramatic as the best competitors produce intervals on the order of 70% longer.

Odds of Failure: Since the mean of the negative binomial distribution is given by $\mu = k(1-p)/p$, confidence limits for μ lead trivially to confidence limits for the *odds of failure* $(1-p)/p$, which is also the expected number of failures Y until the next success at any stage of the inverse sampling process, since the distribution of Y is memoryless.

We conclude this section with a further explanation of why the CMC approach is superior for estimating the mean of a negative binomial distribution to procedures that use (unconditional) minimal cardinality acceptance sets exclusively. Figure 10 shows the 95% confidence intervals for μ in the case $k = 10$ for $0 \leq x \leq 50$ for our CMC procedure and for the Adjusted Crow/Gardner procedure, which uses minimal cardinality sets exclusively, plotted against the values of μ where the transitions between acceptance sets occur. At each such value of μ, the acceptance sets of each method are represented by the horizontal span between the two points plotted for that method. It is evident from the graph that these spans

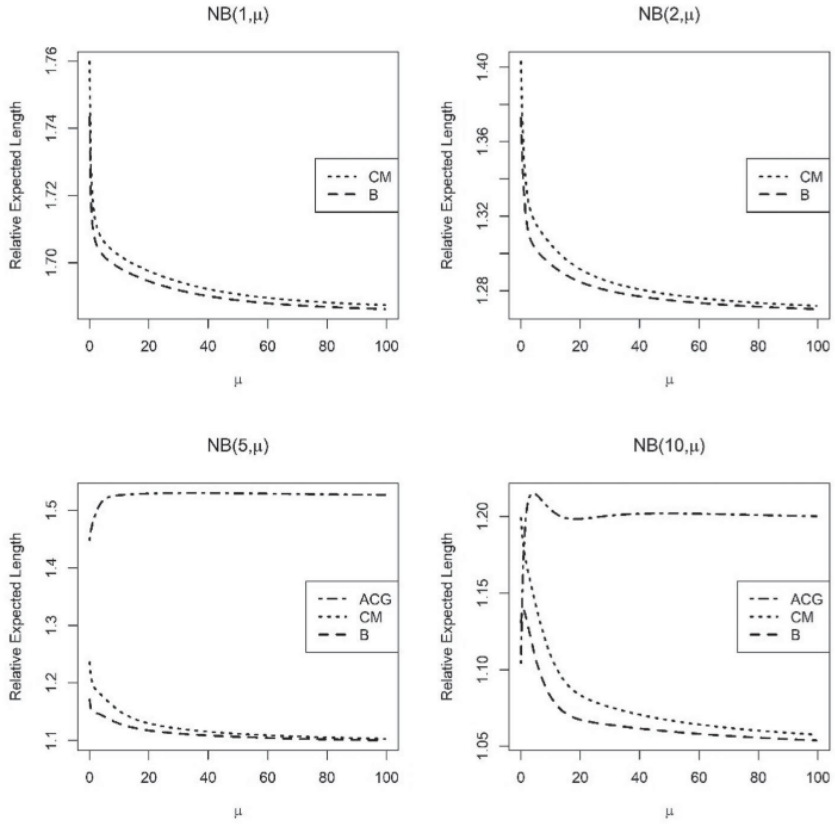

Figure 9. (a)–(d) Expected 95% confidence interval lengths of competing methods for estimation of μ relative to those ~~of~~ obtained from the CMC confidence procedure for $k = 1, 2, 5$ and 10.

are narrower for ACG than for our procedure, as they must be since the former spans have minimal cardinality and the latter typically do not. However, confidence interval lengths are shown by the *vertical* spans. Looking at Figure 10 we can see that, at any x, the vertical span for our method is less than that of ACG; therefore our confidence interval lengths are smaller.

The phenomenon exhibited in the case shown above applies in general regardless of the confidence level or value of k. The effectiveness of the CMC confidence procedure lies in how it pulls down the upper confidence limits, with minimal decreases in the lower confidence limits.

The reason this phenomenon occurs in the negative binomial case and not for other discrete distributions such as the binomial, Poisson and hypergeometric, is that the variability of the negative binomial distribution (as measured for example by the standard deviation) is of the same order of magnitude as the mean itself: $\sigma_X = \sqrt{\frac{\mu(\mu+k)}{k}} \sim O(\mu)$, whereas for other distributions the variability of the observed random variable grows much slower than

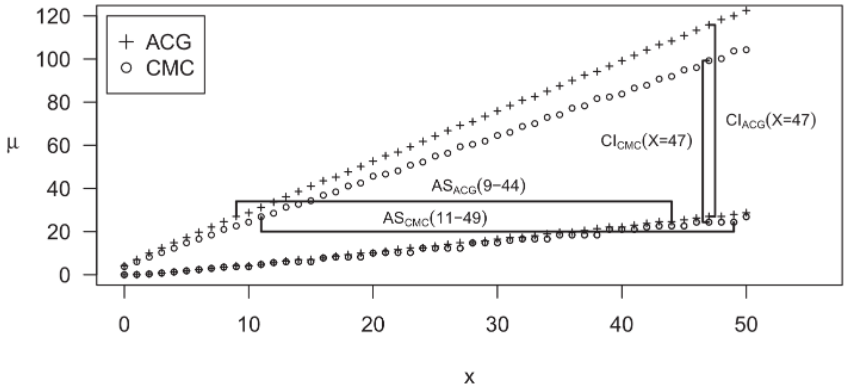

Figure 10. Acceptance sets and confidence intervals for the CMC confidence procedure and for ACG.

the parameter, and a graph analogous to Figure 10 shows roughly parallel sequences for the upper and lower confidence limits.

7. Estimation of *p*

Confidence intervals for p are directly obtainable from those for μ via the relationship $p = 1/(1 + \mu/k)$. The efficacy of CMC here is somewhat different than what occurs when estimating μ, however the conditional minimal cardinality method still retains a marked advantage in estimation accuracy over all of the competing methods for the majority of the parameter space. Figure 11(a)–(d) shows expected 95% confidence interval length for $k = 1, 2, 5$ and 10 at 95% confidence for Casella and McCulloch, Blaker and Adjusted Crow/Gardner relative to the CMC method. Results for the Sterne/LCO, Byrne/Kabaila and Blyth/Still procedures are very similar to those of ACG and are therefore omitted. We also mention here that Byrne [16] developed a confidence procedure for the geometric case that produces short intervals for the probability of failure $1 - p$; Byrne included a table of 95% confidence limits for $x = 0, 1, 2, \ldots, 10$. Making the trivial change to obtain CMC confidence intervals for $1 - p$ from those for p, we find that the CMC intervals are slightly shorter in nearly every case than Byrne's.

Figure 11 demonstrates that, except for small values of p, each of the competing confidence procedures yields substantially greater expected confidence interval length for the values of k investigated than does the proposed procedure.

8. Adjustment for tied endpoints

One of the natural properties a confidence procedure can possess is monotonicity in its endpoints. Specifically, whenever x is increased by one we would expect that both limits of the confidence interval would also increase. Strict confidence procedures sometimes produce instances of tied endpoints, however. For the CMC procedure, there are cases where consecutive lower limits for μ are the same. In order to achieve the desired monotonicity, our algorithm makes slight modifications of those limits by replacing them with an increasing sequence of equally spaced values, ending with the previously tied one and beginning

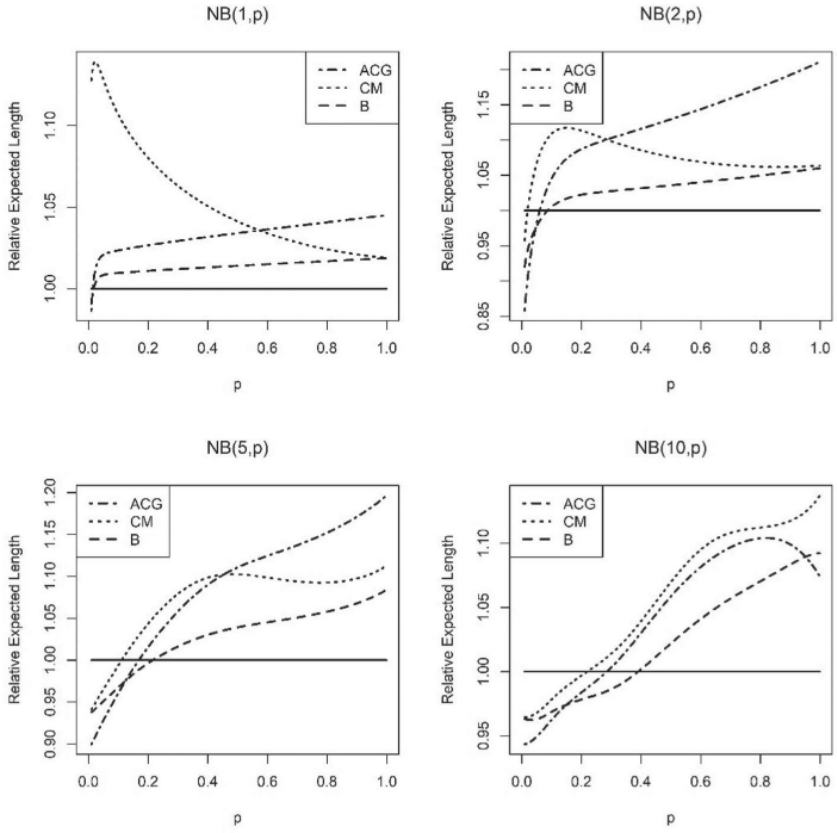

Figure 11. Expected 95% confidence interval length of competing methods for estimation of p relative to that of the CMC method for $k = 1, 2, 5$ and 10.

either (i) immediately after the value below the tied one or (ii) at the value that increases the length of the resulting interval by 1%, whichever is larger. Corresponding adjustments to the upper limits for p are obtained from the relationship $p = 1/(1+\mu/k)$. These adjustments have a negligible effect on expected interval length in both cases. See Schilling and Holladay [6] for further details on this approach to breaking ties.

9. Inverse sampling is more precise than sampling with fixed sample size

It turns out that for estimation of the success parameter p from Bernoulli trials, the precision of estimation obtained from the CMC procedure nearly always exceeds that of binomial sampling that produces the same outcome. We compared interval lengths for a fixed number of successes k and a random number x of failures to those obtained from the LCO binomial confidence intervals of Schilling and Doi [5] for a random number of successes k in $n = k+x$ trials, for $n = 100$ and various k. The latter intervals are from a procedure that yields minimum possible average length for given n of the binomial confidence intervals for $x = 0, 1, \ldots, n$.

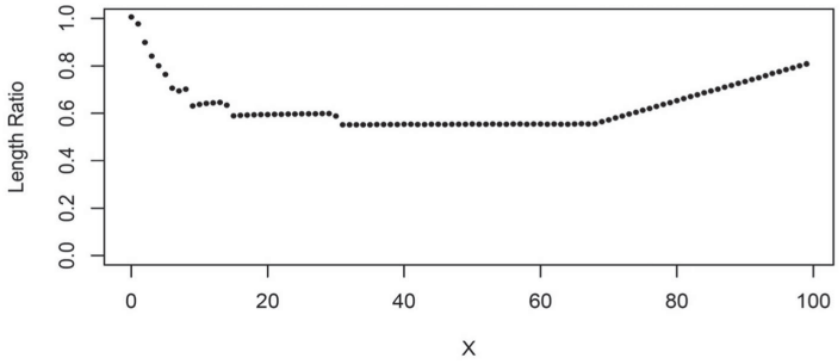

Figure 12. Ratios of 95% geometric confidence interval length obtained from the CMC procedure to the length of the corresponding LCO binomial confidence interval.

Figure 12 shows ratios of minimal cardinality procedure 95% confidence interval length to binomial 95% confidence interval length for the geometric case ($k = 1$). On average, the first 100 geometric 95% confidence intervals are approximately 31% shorter than the corresponding binomial confidence intervals; 90% confidence intervals are approximately 37% shorter, and 99% confidence intervals are approximately 32% shorter. Differences for $k > 1$ are less dramatic, but negative binomial confidence intervals produced by our method are still noticeably shorter than the corresponding binomial intervals in most cases.

10. Conclusion

The conditional minimal cardinality confidence procedure provides substantially more precise interval estimates of μ (as well as the odds of failure) than any existing method, using a computationally efficient algorithm. The corresponding confidence procedure for estimating p has shorter expected length throughout the great majority of the parameter space. We therefore recommend our procedure for general use in situations calling for negative binomial estimation. Researchers may also wish to consider employing inverse sampling rather than binomial sampling when possible, as using the CMC method for the former is likely to produce shorter confidence intervals than are obtained for the latter.

Technical Note

All computations were performed using the R software. A Shiny web app (available from the authors) for determining confidence intervals for μ and for p is available at https://discrete-ci.shinyapps.io/nb_ci/.

Acknowledgement

The authors are grateful for several important suggestions from the referee.

Disclosure statement

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

References

[1] Sterne TE. *Some remarks on confidence of fiducial limits*. Biometrika. 1954;41:275–278. doi:10.1093/biomet/41.1-2.275

[2] Crow EL. *Confidence intervals for a proportion*. Biometrika. 1956;43(3/4):423–435. doi:10.1093/biomet/43.3-4.423

[3] Crow EL, Gardner RS. *Confidence intervals for the expectation of a poisson variable*. Biometrika. 1959;46:441–453. doi:10.1093/biomet/46.3-4.441

[4] Casella G. *Refining binomial confidence intervals*, Canad. J. Statist. 1986;14:113–129. doi:10.2307/3314658

[5] Schilling MF, Doi JA. *A coverage probability approach to finding an optimal binomial confidence procedure*, Amer. Statist. 2014;68(3):133–145. doi:10.1080/00031305.2014.899274

[6] Schilling MF, Holladay B. *Length minimization for poisson confidence procedures*. Commun. Stat. Theor. Method. 2017;46(2):861–873. doi:10.1080/03610926.2015.1006782

[7] Wang W. *Exact optimal confidence intervals for hypergeometric parameters*. J. Am. Stat. Assoc. 2015;110:1491–1499. doi:10.1080/01621459.2014.966191

[8] M. F. Schilling and A. Stanley, *A new approach to precise interval estimation for the parameters of the hypergeometric distribution*. Commun. Stat. Theor. Method. 2022;51(1):29-50. doi:10.1080/03610926.2020.173
7879

[9] Casella G, Berger RL. Statistical inference. 2nd ed. Duxbury: Thomson Learning Inc.

[10] Bain LJ, Engelhardt M. Introduction to probability and mathematical statistics. 2nd ed. Boston: PWS-KENT.

[11] Casella G, McCulloch CE. Confidence intervals for discrete distributions; 1984. (Biometrics Unit Technical Reports: Number BU-839-M).

[12] Lui Y. *Confidence limits for the population prevalence rate based on the negative binomial distribution*. Stat Med. 1995;14:1471–1477. doi:10.1002/sim.4780141307

[13] Blaker H. *Confidence curves and improved exact confidence intervals for discrete distributions*, Canad. J. Statist. 2000;28:783–798. doi:10.2307/3315916

[14] Byrne J, Kabaila P. *Comparison of poisson confidence intervals*, Commun. Statist.—Theory and Methods. 2005;34(3):545–556. doi:10.1081/STA-200052109

[15] Blyth CR, Still HA. *Binomial confidence intervals*. J Am Stat Assoc. 1983;78:108–116. doi:10.1080/01621459.1983.10477938

[16] Byrne J. *A new short exact geometric confidence interval*. Aust. N Z J. Statist. 2005;47:563–569. doi:10.1111/j.1467-842X.2005.00417.x

Appendix

We provide here a detailed discussion of how our algorithm for finding confidence limits works. First note that $RB(a)$ is non-empty for all a because we can always increase μ enough that $P_\mu(X \geq a) > 1-\alpha$, thus we can find a large enough b such that $P_\mu(a \leq X \leq b) \geq 1-\alpha$. Second, much of our algorithm hinges on the fact that a graph of the curves in each $RB(a)$ is actually 'rainbow-like' in shape as depicted in Figures 3 and 4. This follows from that fact that acceptance curves are unimodal and satisfy $P_\mu(a \leq X \leq b) < P_\mu(a \leq X \leq (b+1))$ for all μ, and non-negative integers $a \leq b$.

Determining the interior arc $P_\mu(a \leq X \leq m(a))$ of each rainbow is straightforward since the location of $\max_\mu\{P_\mu(a \leq X \leq b)\}$ can be shown to have a closed form solution:

$$\mu_{\max}(a,b) = k(1-p_{\max})/p_{\max},$$

where

$$p_{\max}(a,b) = 1 - \left[\frac{(a+k-1)(a+k-2)\cdots(a)}{(b+k)(b+k-1)\cdots(b+1)}\right]^{\frac{1}{b-a+1}}.$$

Thus a simple *while* loop can easily find $m(a)$ by looking for the smallest value of b satisfying $\max_\mu\{P_\mu(a \le X \le b)\} \ge 1-\alpha$. In particular we can start the loop at $b = m(a-1)+1$ and increment b by one until $\max_\mu\{P_\mu(a \le X \le b)\}$ first exceeds the confidence level. We can also quickly compute the locations where each acceptance curve $P_\mu(a \le X \le b)$ that exceeds the confidence level intersects it using R's *uniroot* function or any similar efficient algorithm. Let $\text{root1}(a,b)$ and $\text{root2}(a,b)$ be the smaller and larger of these two roots, respectively, where $\text{root1}(a,b) = \text{DNE}$ when $a = 0$. All upper and lower endpoints of the procedure correspond to these roots.

We begin the algorithm with $a = 0$. Starting with $\mu = 0$, use $AS(a - m(a))$ until $\mu = \text{root2}(a, m(a))$, at which point use the subsequent curve $AS(a-(m(a)+1))$. Then use $AS(a-(m(a)+1))$ until $\mu = \text{root2}(a, m(a)+1)$, followed by the subsequent curve $AS(a-(m(a)+2))$. Continue in a similar fashion until $\mu = \text{root1}(a+1, m(a+1))$, at which point transition to the next rainbow, $RB(a+1)$, and move to the curve $AS((a+1)-m(a+1))$. Repeat as before using the subsequent curves of the rainbow whenever the current curve drops below $1-\alpha$, continuing until the core of the next rainbow first rises above the confidence level (at which point we change to that rainbow).

A transition that occurs between curves from the same rainbow, $P_\mu(a \le X \le b)$ to $P_\mu(a \le X \le b+1)$, determines the lower endpoint for $b+1$, $l(b+1) = \text{root2}(a,b)$. When transitioning between different rainbows $RB(a)$ to $RB(a+1)$, we transition between curves $P_\mu(a \le X \le b)$ and $P_\mu(a+1 \le X \le m(a+1))$, for $m(a+1) \ge m(a)+1$. The location of this transition occurs at $\text{root1}(a+1, m(a+1))$ and determines the upper endpoint for a, $u(a) = \text{root1}(a+1, m(a+1))$, and the lower endpoints for $b+1, \ldots, m(a+1)$, $l(b+1) = \cdots = l(m(a+1)) = \text{root1}(a+1, m(a+1))$.

To determine confidence intervals for all x up to some value n, the algorithm continues until $l(n)$ is determined. The remaining values of $u(x)$ can then be separately determined by $u(x) = \text{root1}(x+1, m(x+1))$.

Note: To obtain a confidence interval only for one specific value of x, a shortcut can be used to determine the lower endpoint for x: Start the search at the first transition to $RB(a)$ (at $\mu = \text{root1}(a, m(a))$) where a is the largest value such $m(a) < x$ (i.e no acceptance curves have involved x quite yet). Then $l(x)$ will correspond to one of the next few acceptance curve transitions, including possibly the transition to the next rainbow, $RB(a+1)$ (but, no later than that, otherwise $m(a+1) < x$, leading to a contradiction). $u(x)$ is determined as before by $u(x) = \text{root1}(x+1, m(x+1))$.