Confidence intervals for the mean of the Poisson distribution

Hannah Pawig

Introduction

Within the last decade, the opioid epidemic has become increasingly known amongst the public in the last few decades, as more adolescents and young adults have been included as fatalities. Some reasons for the opioid epidemic include overprescription and illicitly sourced drugs that are at high risk of containing fatal doses of fentanyl. The University of California Los Angeles (UCLA) Medicine Department published an article that about 22 adolescents aged 14 to 18 died each week in the United States due to illicit drug overdoses in 2022. This would amount to 1200 adolescent deaths per year nationwide, an extremely small percentage of the estimated American children from 14 to 18 years old over 20 million in 2022.

What is even more tragic is the statement, "Adolescent overdoses had more than doubled among this group between 2019 and 2020." Although, what were the estimates in 2019 and 2020? What could be useful in this scenario is a reliable estimate for these years, and all years prior, based on available data on opioid-related deaths in this age group. A way to estimate the number of deaths in a period could be through a short range of possible values or an *interval* of some sort. That way we might have a precise and tangible idea of the change in deaths per year, which could be of use to determine the efficacy of America's opioid epidemic awareness programs or organizations. This is where confidence intervals help.

We will let the random discrete variable X represent the number of U.S. children (ages 14-18) who passed away from an illicit drug overdose in one week. Since X is a *discrete* variable, X can only take on real, integer values. That is whole numbers from 0 and on.

Developed in 1837 by French mathematician Siméon Denis Poisson, the Poisson distribution is a discrete distribution used to represent the number of occurrences for an event within a fixed time interval. The Poisson distribution assumes that all events occur independently, and that the rate of an event is always constant. For Binomial-type scenarios with a very small success rate (*p*), the Poisson distribution can also be used to approximate Binomial distributions (DeGroot & Schervish 2011).

Poisson's only parameter is the mean, usually denoted as the Greek letter λ. Sometimes when dealing with the Poisson distribution, the "rate parameter" usually denoted as θ is given. The relationship between the mean parameter and the rate parameter is that each parameter is the reciprocal of the other (i.e. $1/\theta = \lambda$ and vice versa).

For example, X = number of customers in an hour can be modeled by a Poisson distribution. Thus, we have that X is a Poisson random variable; symbolically, this is X ~ Poisson(λ).

In terms of the mean parameter, the probability mass function (pmf) is:

$$p_X(k) = \frac{e^{-\lambda}\lambda}{k!},\ k = 0,\ 1,\ 2,\ 3,$$

- *k* represents the observed number of events
- Plugging in *k* given an average of λ in a time frame into this function returns a probability of seeing *k* events in that same specified time frame.

The mean of a Poisson distribution is λ and its variance is also λ.

Recall *X*, representing the number of opioid-related deaths in the U.S. per week, is a discrete random variable (RV) that can take on integer values from 0 and on. This RV is a great candidate for being modeled by a Poisson distribution because opioid-related deaths in 14 to 18-year-olds is a rare event (we will use adolescents to reference only the 14 to 18-year-old age group). Given we have observed λ=22 adolescent deaths per week in the U.S. on average (so 52*22 = 1144 deaths in a year), we will compute a Poisson confidence interval (CI) using the Wald method. We'll see more of the Wald method later.

95% Wald CI = (2.81, 31.19)

What follows is a model interpretation for the above CI. "We are 95% confident that in 2022, there were between 2.81 to 31.19 opioid-related deaths in adolescents aged 14 to 18 in the United States each week." This also may be interpreted as, if we repeatedly observed the deaths for many weeks and computed Wald intervals for each of those weeks, 95% of the confidence intervals contain the true average adolescent deaths per week in the United States in 2022. Intervals such as this one could be used to make comparisons to past years and future ones once they have occurred.

It must be cautioned however that the Wald method is one of the worst methods for interval estimation for the Poisson distribution and all other distributions! It is based on large sample theory, but in some cases, data comes from small samples. Confidence intervals and interval estimation have been tricky especially when it comes to discrete distributions. Better interval estimating methods for the infamous Binomial distribution have been found by various statisticians, such as the Blaker method, the Score method, and the LCO method. There is not yet an agreement on the best method for this distribution. The same goes for the Poisson distribution; a few methods have been found but are very inaccessible to the public. Poisson confidence intervals can be helpful for researchers in medicine (much like our example), business, government agencies, education, and more. There is a gap between these newer Poisson confidence interval methods and non-statisticians in other industries. We aim to present and provide that resource with this paper.

Additionally, in a recent article by Doi, Holladay, and Schilling (2023), a new approach to interval estimation with the discrete distribution "Negative Binomial" was discovered and named the *conditional minimal cardinality* (CMC) method. There is not yet a version of the CMC method that has been discovered for the Poisson distribution, which may serve as motivation for the need for Poisson confidence intervals!

Large Sample Methods

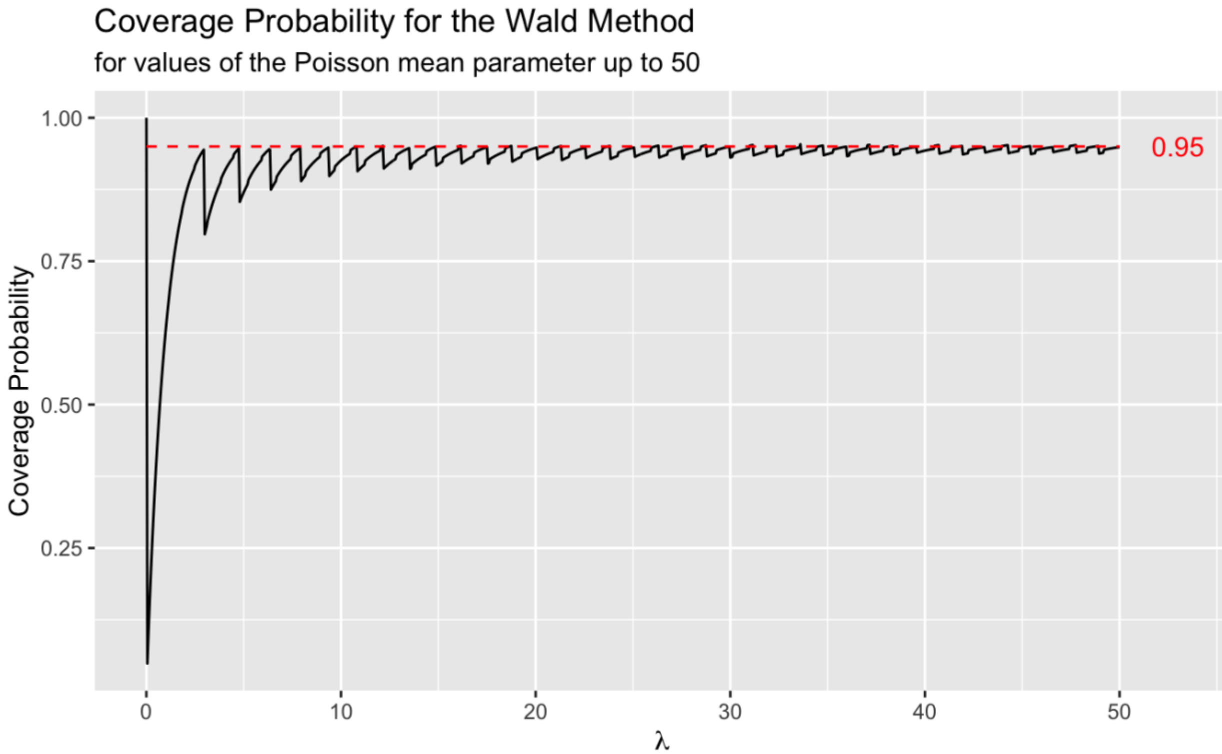
First, we will look at the "large sample methods" that have been used since the 20th century to approximate confidence intervals for discrete distributions—including the Binomial, Poisson, and Negative Binomial. The keyword here is *approximate*, as these intervals make use of large sampling theory, using continuous distributions such as the Standard Normal and Chi-Square distributions to make confidence interval calculations. These methods have become largely used, because of technological limits at the time they were created. Now, these methods can be proven inferior to the "strict" confidence procedures developed more recently with the help of technological advancements.

I. The Wald (W) Method

The Wald method for interval estimation is usually one of the first methods of construction confidence intervals taught in statistics courses. This large sample method provides an approximate confidence interval, relying on the fact that the Normal distribution approximates the Binomial distribution when sample size *n* is large. Additionally, the Poisson distribution is a special case of the Binomial distribution where parameter *p* (rate of success) is small for a potentially infinite sample size. The Wald method is impractical for discrete distributions, as its coverage probability is below the promised confidence level as shown below. Research publications on discrete distributions like that of Holladay (2019) support this.

The Wald Confidence Procedures for Binomial, Poisson, and Negative Binomial distributions can be found below ["Table 1.3" from Holladay (2019, p. 13)].

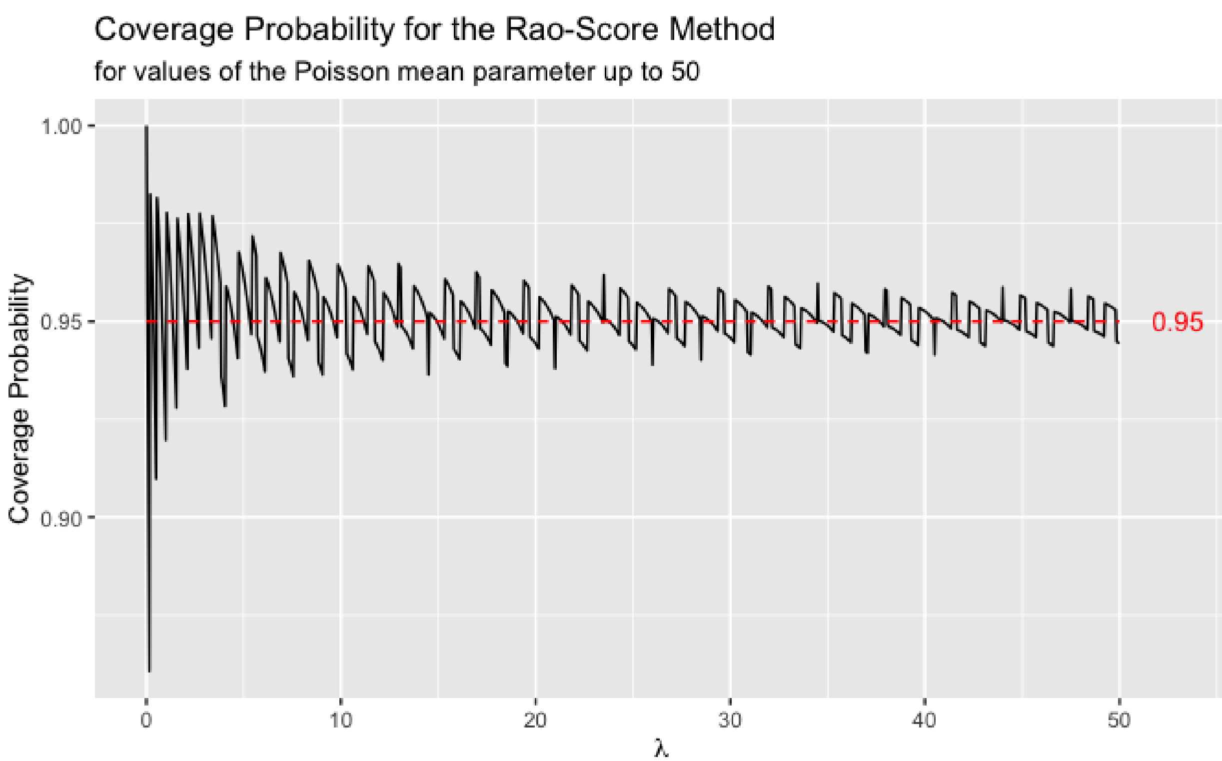
Distribution	$\hat{\theta}$	$[I(\theta)]^{-1}$	Wald Interval
Binomial(n, θ)	$\frac{X}{n}$	$\theta(1-\theta)/n$	$\frac{x}{n} \mp z_{\alpha/2}\sqrt{\left(\frac{x}{n}\right)\left(1-\frac{x}{n}\right)/n}$
Poisson(θ)	X	θ	$x \mp z_{\alpha/2}\sqrt{x}$
NB(r, θ)	$\frac{r}{r+x}$	$\frac{\theta^2(1-\theta)}{r}$	$\frac{r}{r+x} \mp z_{\alpha/2}\sqrt{\frac{rx}{(r+x)^3}}$



Example 95% cpf for the Wald method, for lambda = 0,..., 50.

II. Rao's Score (S) Method

The Score method originated in 1946, from Dr. Calyampudi R. Rao after being tasked with testing simple hypotheses given some information about the parameter is known (Scholarpedia). The Rao-Score confidence interval is Wald's confidence interval but with an actual calculation for the Fisher's Information $I(\theta)$, a special case for Binomial and Poisson distributions [Holladay (2019, pp.14-15)]. This procedure uses the "standardized score" statistic which has an asymptotic standard normal distribution [Casella & Berger (2002, p. 494)]. Below is the graph of the coverage probability function (cpf) when the Score method is applied to the Poisson distribution. The cpf is erratic and not always above or equal to the stated confidence level. It seems more trustworthy than Wald's in reaching desired coverage according to the confidence level, though neither Wald nor Rao's Score is considered superior overall.



Example 95% Cpf for Rao's Score method, for lambda = 0,..., 50.

Rao's Score confidence procedure for the Poisson distribution may be found using the following formula (from Holladay (2019))

$$\{\mathrm{x} + \frac{1}{2} z_{\alpha/2}^2 \mp z_{\alpha/2} \sqrt{x + \frac{1}{4} z_{\alpha/2}^2}\}$$

- Note: $z_{\alpha/2}$ is the $(1 - \alpha/2)^{th}$ quantile of the Standard Normal or Gaussian distribution.

For example, *X* represents the single observation of opioid-related deaths in one year. Then the confidence procedure would construct and interval upon substitution of *X* for the observation.

III. Wilks' Likelihood Ratio (LR) Method

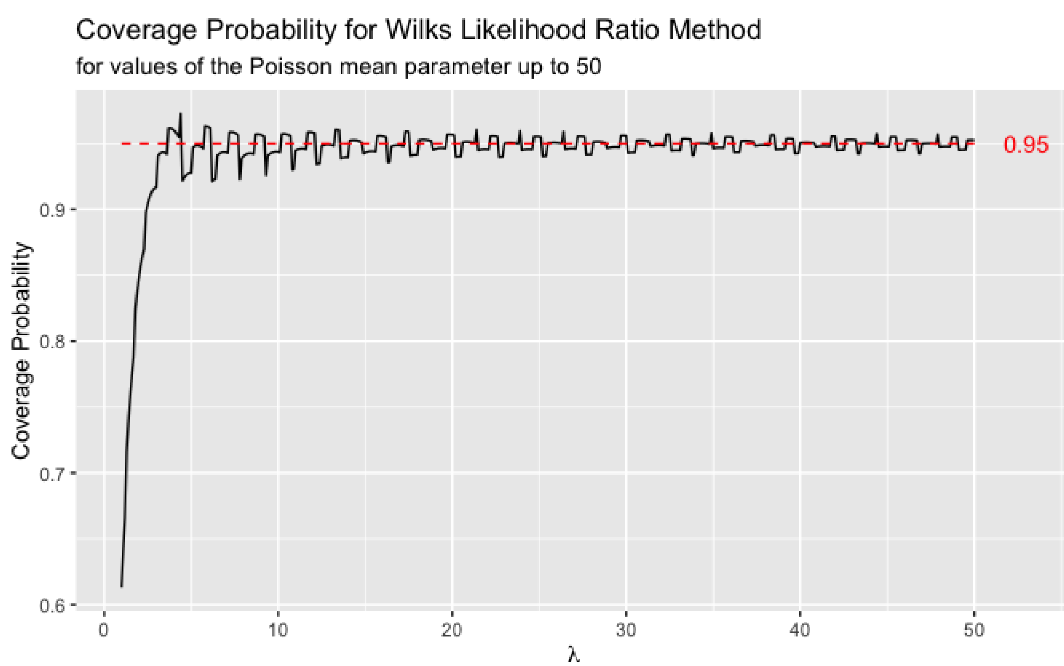
The last of the "Holy Trinity" of confidence procedures is Wilks Likelihood Ratio or Likelihood Ratio. This confidence procedure is part of the test for two-sided alternative hypotheses. The quantity$2\left[\ln \mathrm{L}(\hat{\theta}) - \ln \mathrm{L}(\theta)\right] \approx \chi^2_1$ is given by Wilks' Theorem. This theorem provides the structure for the confidence set of θ for this procedure that is found in Holladay (2019). Like the Wald and Score methods, estimator $\hat{\theta}$ represents the maximum likelihood estimator (MLE) of θ. In the Poisson case, the MLE for θ is the sample mean, and for our example the sample mean is simply the single observation of opioid-related deaths in a fixed time frame. Like the previously discussed procedures, *LR* has erratic coverage probability when applied to discrete distributions. We have the coverage probabilities in the Poisson case of Wilks' Likelihood Ratio confidence procedure displayed here as well. Note here too that the *LR* procedure is not superior to Wald or Rao's Score overall.

To find the approximate CI for the parameter θ given by *LR*, one may use this open-form set given by Holladay (2019):

$$\{\theta: 2\left[lnL(\hat{\theta}) - lnL(\theta)\right] \leq \chi^2_{(1,1-\alpha)}\}$$

- The last term is the $(1-\alpha)^{th}$ quantile of a χ^2 distribution with 1 degree of freedom

We used this form of the confidence procedure to produce the coverage probability plot here:



Strict Confidence Procedures

Strict confidence procedures all satisfy one condition: for any parameter value, the procedure will produce an interval that has a probability of capturing the true parameter that is at least equal to the desired confidence level i.e. 95%. We'll review five strict methods that are applicable to the Poisson distribution.

I. Clopper-Pearson (CP) Method

The strict coverage confidence procedure, Clopper-Pearson, was first proposed as a method only for the binomial distribution. Let *RV* X ~ Binomial(n, θ): *CP* works by finding the test's rejection region with a maximum number of possible *x*-values so both tail probabilities, $\mathrm{P}(X \leq x)$ and $\mathrm{P}(X \geq x)$, are no greater or exactly equal to $\alpha/2$. This process ensures the probability of falsely rejecting under the null hypothesis is at most the chosen α —level. The Binomial and Poisson distributions are known to have relationships to the cumulative distribution functions (CDFs) of the continuous distributions Beta and Chi-Square, respectively (see Holladay (2019, pp. 19-20 for more). Applying *CP* to Poisson, we achieve the form of the lower and upper bounds of *CP*.

CP $(1-\alpha)$ confidence procedure for Poisson:

Lower bound (given x)

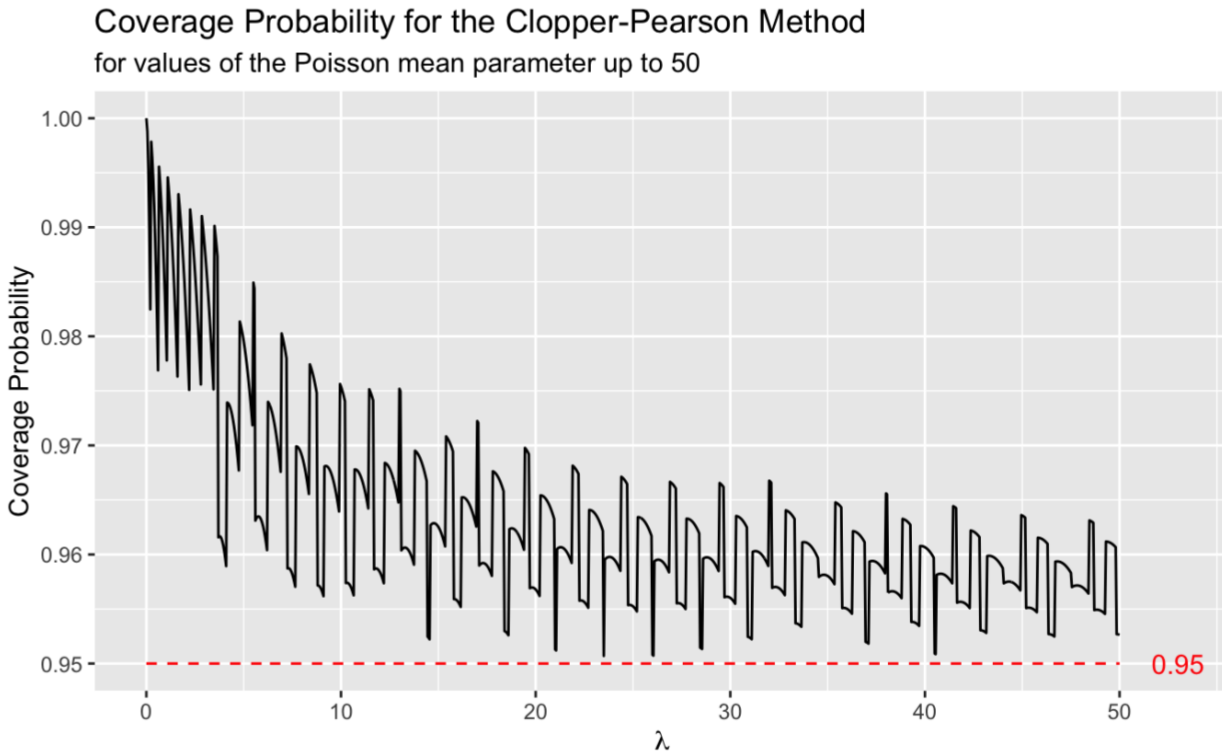
$$\mathrm{l(x)} = 1/2\, \chi^2_{(2\mathrm{x},\alpha/2)}$$

Upper bound (given x)

$$\mathrm{u(x)} = 1/2\, \chi^2_{(2\mathrm{x},1-\alpha/2)}$$

- Note: $\chi^2_{2\mathrm{x},\alpha/2}$ represents the $(\alpha/2)^{\mathrm{th}}$ quantile of the χ^2 distribution with $2x$ degrees of freedom.

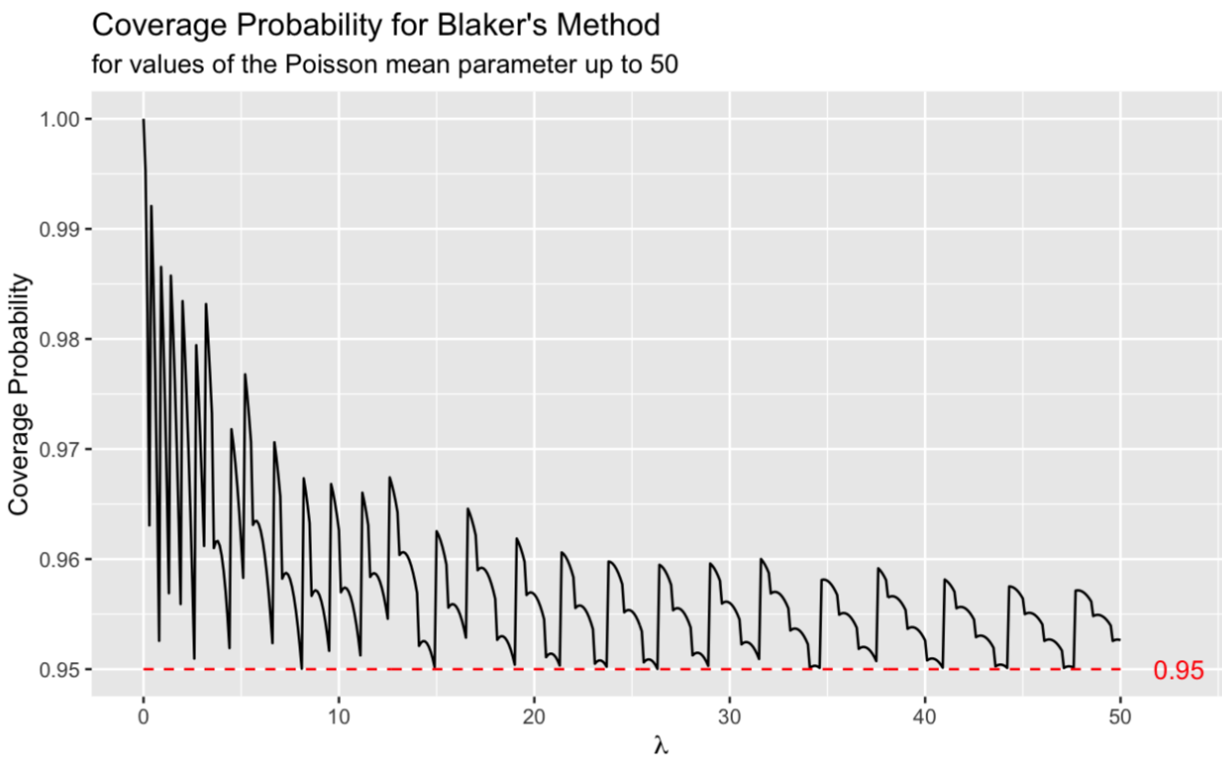
Looking at the coverage probability plot, it is immediately apparent that this method produces conservative confidence intervals:


Coverage Probability for the Clopper-Pearson Method
for values of the Poisson mean parameter up to 50
Coverage Probability
1.00
0.99
0.98
0.97
0.96
0.95
0.95
0
10
20
30
40
50
λ

II. Blaker's Method

In 2000, Helge Blaker introduced the strict confidence procedure (which we may call "Blaker's method") that has applications to all discrete distributions: binomial, hypergeometric, negative binomial, and Poisson. Blaker's procedure has the nesting property, minimizes length, and maximizes coverage. Unlike the methods in the next section, Blaker's method does not always use acceptance curves of minimal cardinality. Blaker notes his method's uniqueness because it does not invert a equal two-tailed test to find the confidence procedure.

Steps to Blaker's method are as follows. Given observed x, find both tail probabilities $\mathrm{P}(X \leq x)$ and $\mathrm{P}(X \geq x)$ and take the smallest of the two, coined as the "min-tail probability" for x. Following this, for a fixed θ, any $\mathrm{x}_0 \in \mathcal{X}$ (including observed x) will be included in the acceptance set of θ (A_θ) if the probability of observing a min-tail probability of any x exceeds that of the chosen α level.



Minimal Cardinality Procedures

Minimal cardinality procedures fall under this category of strict confidence procedures as well. What sets them apart from the Clopper-Pearson or Blaker's procedures is their use of acceptance sets with the smallest number of x's but have corresponding acceptance curves strictly above the desired confidence level. This feature allows them to be named "minimal cardinality" procedures.

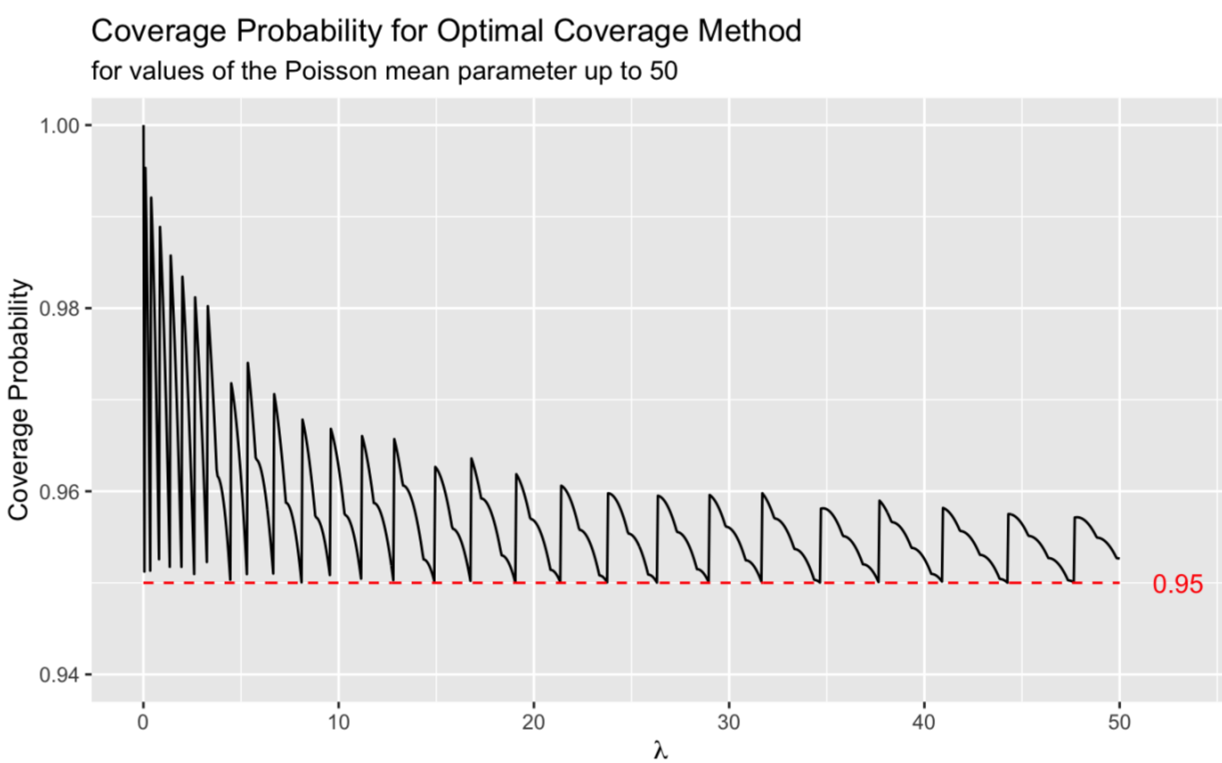
Theodore Sterne first introduced minimal confidence procedures in 1954, with the creation of his minimal cardinality procedure for the Binomial distribution ("Sterne's Method").

Generally, to create such a procedure, start by finding the acceptance sets of cardinality = 1 (the first of these sets is the set containing only 0) that are above the desired confidence level. Keep those acceptance sets that have the smallest cardinality; note that having acceptance sets with tied cardinalities is possible. If the cardinality of the acceptance set is too small to rise above the confidence level, continue to increase the cardinality by one until the acceptance sets produce acceptance curves across the span of the desired maximum lambda value.

Where the minimal cardinality procedures have more than one acceptance curve with minimal cardinality, Sterne's procedure chooses the acceptance curve with the greatest coverage at lambda. However, Sterne's procedure has proven to be faulty since gaps in confidence sets are prone to occur when choosing acceptance curves in this way.

I. Modified Sterne or Optimal Coverage

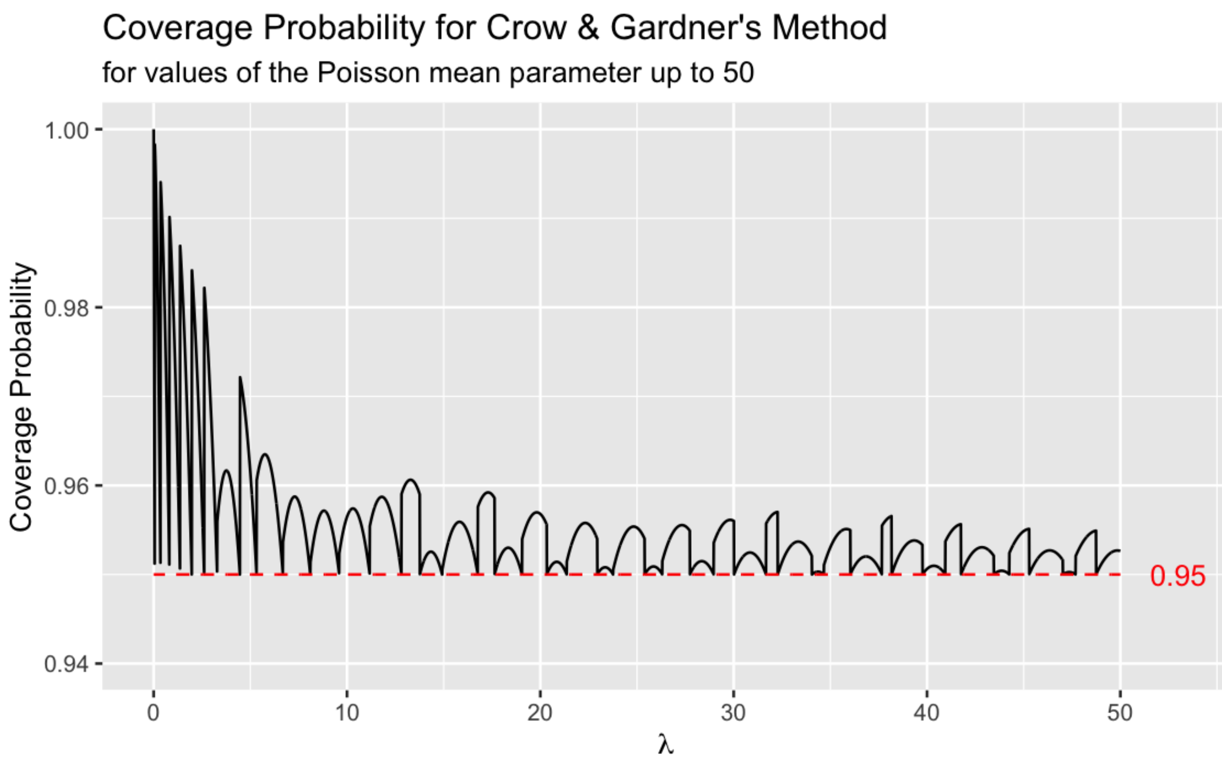
To keep the high coverage that Sterne's procedure provides, a compromise was made to Sterne's procedure to fix the issue with gaps in confidence sets. This is known as the Modified Sterne (MST) or Optimal Coverage (OC) method [from Holladay & Schilling (2017], where gaps are avoided by strategically choosing acceptance curves with non-decreasing cardinalities and non-decreasing {a} sequences. In other words, in the cases where cardinality must be increased, OC instead increases cardinality by only increasing {b} or increasing both {a} and {b} to get an acceptance curve that is above the confidence level, to achieve minimal cardinality without gaps at certain x-values. This prevents gaps from occurring; the original Sterne's method allowed decreases in the {a} sequence as a sacrifice to attain the highest coverage possible.



By way of construction, MST's advantage is that the procedure still has maximal coverage and has the highest average coverage possible (see Holladay (2019) page 40) without interval gaps. Next, we will see another strict minimal cardinality procedure that has similar coverage to MST but it is shorter in length on average (which we touch on more in a later section).

II. Crow & Gardner

Edwin Crow & Robert Gardner revealed their method (abbreviated CG) in their 1959 paper. Instead of defaulting to ACs with the greatest coverage at any given lambda, CG defaults to choosing the AC with the largest {a}, resulting in switching ACs "as soon as possible." Once an AC with a greater {a} than the current AC rises above the confidence level, CG then will use this new AC. As a side note, Byrne and Kabaila introduced their minimal cardinality procedure (abbreviated BK), which does the exact opposite of CG. BK transitions between acceptance curves "as late as possible," riding the current AC until that AC goes below the confidence level then choosing the AC with the next greatest {a}. BK has proven to be the worst of the three minimal cardinality procedures we have discussed regarding length and coverage. Under this, we've provided the CPF of CG—the superior of these two [see Holladay (2019) for discussion on this].

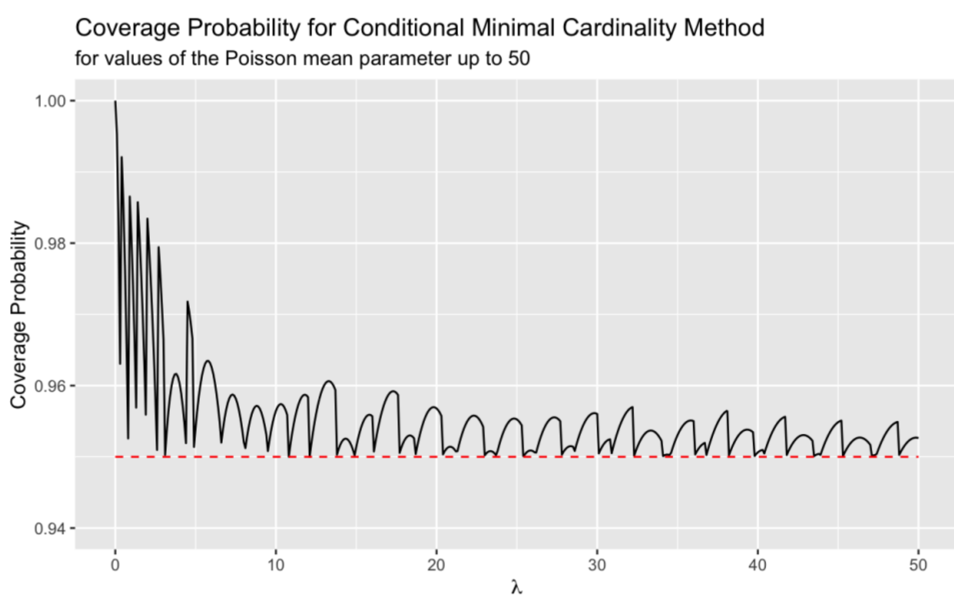


Conditional Minimal Cardinality

The most recent discrete confidence procedure was found in 2023 by Doi, Holladay, and Schilling using the Negative Binomial distribution. Termed "conditional minimal cardinality," (CMC) this procedure makes use of the same minimal cardinality acceptance curves that procedures also use. However, it must be clear to note that CMC is **not** a minimal cardinality procedure; what is different is that the CMC method groups all acceptance curves by their {a} value. Because of their graphical appearance, these groups are called rainbows of {a} (abbreviated as RB(a)). For example, at 95% confidence for the Poisson distribution, RB(1) includes the minimal cardinality curves 1-7, 1-8, and 1-9 (and on). The rainbow's core is defined to be the AC with the smallest {b}; the core for RB(1) is 1-1. This process does not guarantee that the minimal cardinality curve is being used for **all** values of the parameter.

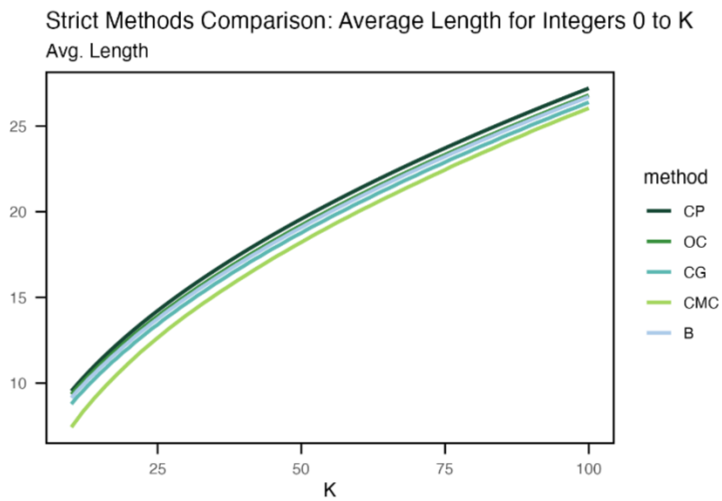
First, the procedure uses the RB's core until it hits the confidence level, then transitions to the next available AC with {b} closest to its own. The transition process within each rainbow continues until the following RB's core rises above the confidence level. In our example, this is until RB(2)'s core AC 2-10 attains the desired confidence level before AC 1-9 goes below the confidence level. The lambda at which AC 2-10's first point of rising above the confidence level is when the procedure moves on to using RB(2). This rainbow's core is used until it goes below the confidence level and then switches to the next eligible AC in the rainbow or until RB(3)'s core comes above the confidence level and then repeats.

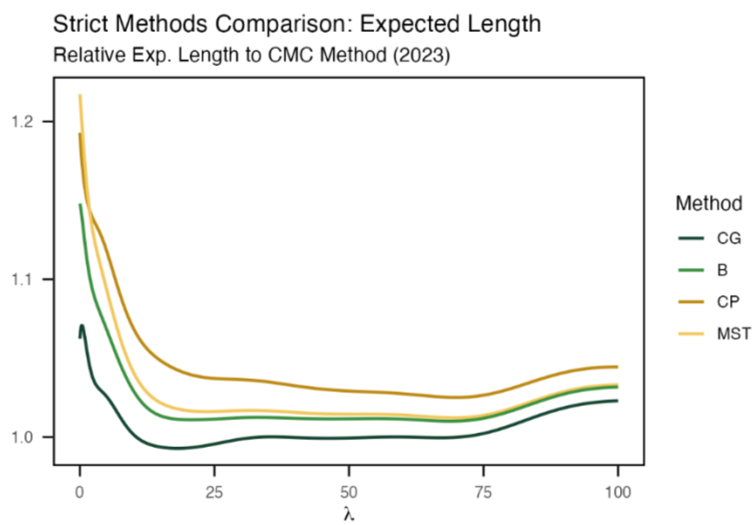
The CMC paper compares the method to other comparable strict procedures like Casella & McCulloch and Blaker methods. CMC was shown to be relatively shorter in various cases of the negative binomial (Schilling et. al p. 135). We'll explore more comparisons to the methods we've seen in the Poisson case instead.



Length and Coverage Comparisons

In Holladay (2019), it is noted that CG has the shortest average length and expected length compared to Blaker, CMC, and MST. The graph below shows two corresponding graphs for average length and expected length as maximum observed x (we call this *K*) increases.





Conclusion

Regarding average and expected length criteria, CMC clearly appears to produce the shortest intervals overall. The plots only compare the strict methods, so we know all methods here guarantee at least 95% coverage when 95% is the desired confidence level. However, we might look into how average coverage compares across these five methods and see if CMC is actually superior in terms of highest coverage *and* smallest interval lengths/widths—the two main criteria for judging confidence procedures as used by the publications we've referenced.

In the app made over this summer project, the available methods for the app user are all the methods discussed in-depth (note: Byrne & Kabaila was not coded, as it performs the worst of all three minimal cardinality procedures thus far). The app is intended to make these methods accessible and usable by researchers who are not in the statistical field or can code these statistical methods. Its use is not limited to this audience, but also can be used as a learning tool in learning the concepts related to discrete distribution confidence procedures at the undergraduate level.

The app intakes the user's desired confidence procedure, confidence level (%), and observed sum. Since the sum of independent and identically distributed (iid) Poisson RV is itself a Poisson distribution with lambda equal to the sum of each iid Poisson's mean parameter, the app can be used for a random sample of Poisson RVs in addition to a single observation. Thus, there is a note in the app that the user may input a total (inherently positive) integer that represents the number of observed successes/occurrences of the event of interest. An important thing to note is that since the sum of iid Poisson rv's is a sufficient statistic for lambda, it is valid for a user to input the sum or total events observed in a sample. To get the confidence limits for the mean parameter, divide the confidence limits by the sample size *n*.

Future Work

One thing that someone may be interested in next is optimizing the algorithms made over the course of this project (Summer 2024). We can also create an R package that includes all confidence procedures that have been established for Poisson. It could even include the method versions for the other discrete distributions (i.e. an all-inclusive R package containing discrete distributions' strict confidence procedures). In the Shiny app, the option to enter an average of an observed sample could be added as well.

References

Blaker, H. (2000), *Confidence curves and improved exact confidence intervals for discrete distributions*. Can J Statistics, 28: 783-798. https://doi.org/10.2307/3315916

Doi, J. A., Holladay, B., & Schilling, M. F. (2023). Improved interval estimation of negative binomial parameters: a coverage probability approach. Journal of Statistical Computation and Simulation, 94(1), 126–141. https://doi.org/10.1080/00949655.2023.2235046

Holladay, B. A. (2019). *Contributions to Interval Estimation for Parameters of Discrete Distributions*. UC Santa Barbara. Retrieved from https://escholarship.org/uc/item/5mz623j1

Schilling, M. F., & Doi, J. A. (2014). *A Coverage Probability Approach to Finding an Optimal Binomial Confidence Procedure*. The American Statistician, 68(3), 133–145. https://doi.org/10.1080/00031305.2014.899274

Schilling, M. F., & Holladay, B. (2016). *Length minimization for Poisson confidence procedures. Communications in Statistics - Theory and Methods*, 46(2), 861–873. https://doi.org/10.1080/03610926.2015.1006782