CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

OPTIMAL CONFIDENCE INTERVALS FOR THE
EXPECTATION OF A POISSON RANDOM VARIABLE

A thesis submitted in partial fulfillment of the requirements
For the degree of Master of Science in Mathematics

by

Bret A. Holladay

May 2014
The thesis of Bret A. Holladay is approved:

____________________________________  ___________________
Majid Mojirsheibani, Ph.D.  Date

____________________________________  ___________________
Ann Watkins, Ph.D.  Date

____________________________________  ___________________
Mark Schilling, Ph.D., Chair  Date

California State University, Northridge
Acknowledgements

I would like to express my appreciation to all those who have helped me along my journey towards a Master of Science in Mathematics. My family, friends, classmates, professors, and colleagues all provided me with unending support. They encouraged me when I was feeling overwhelmed and celebrated my successes.

A special thanks to Professor Mark Schilling for all of the feedback and assistance on this thesis. You are the spine to these pages – keeping everything together, structured and ordered. You have been a great role model through your dedication to your work, your devotion to your students, and your gift for teaching. You cultivated my interest in statistics and inspired me to pursue a doctoral degree in the field.

Heather Holladay and Matt Heitner, thank you for all the advice and assistance on my Ph.D. applications.

I especially want to thank my parents for their unconditional love and support during my (ongoing) education. I could not have done it without you!

And finally, Redilyn Mesa, thank you for your love and encouragement. But most of all, thank you for your patience when you did not hear back from me because I was deep into my thesis and thank you for your understanding when I was forced to cancel our weekend plans because of homework or an upcoming exam. I know it isn't always easy dating a full-time graduate student.
Dedication

For Sadie, my best friend for more than fifteen years, because of her endless love. You will forever be in my heart.
# Table of Contents

Signature Page ii  
Acknowledgements iii  
Dedication iv  
List of Figures vii  
List of Tables viii  
Abstract ix  

Chapter 1: Introduction 1  
1.1 Confidence Intervals 1  
1.2 Poisson Process 1  
1.3 Poisson Distribution 2  

Chapter 2: Examples and Preliminaries 3  
2.1 Formula-based Methods for Constructing Confidence Intervals for $\lambda$ 3  
2.2 Illustrative Examples 5  
2.3 Coverage Probability 8  
2.4 Confidence Interval Performance 8  
2.5 Preliminaries 11  

Chapter 3: The Modified Sterne’s Method 14  
3.1 MS 14  
3.2 Group the Curves 15  
3.3 Necklaces 16  
3.4 Choosing the Ideal Coverage Probability Function 18  
3.5 Constructing Confidence Intervals 20  
3.6 Gaps and the Transition Between Successive Necklaces 22  
3.7 Pseudocode for MS 27  

Chapter 4: Comparing Methods 29  
4.1 Comparison to Prevailing Formula-based Methods 29  
4.2 Optimality 31  
4.3 Notable Methods 34  
4.4 Comparison of the Least Cardinality Procedures 40  
4.5 The Least Cardinality Percentage Method(LC-$p$) 47  
4.6 Further Discussion of Length 49
4.7 Coverage
4.8 The Modified Crow and Gardner (MCG) Method
Chapter 5: Conclusion
  5.1 Opportunities for Further Research
  5.2 Summary
References
Appendix A: Tables
Appendix B: Proofs
List of Figures

Figure 1: An illustration of Propositions 5 and 6. 12
Figure 2: "Sea of curves." 14
Figure 3: Curves $P_{j,j+9}(\lambda)$ which contain ten values of $x$. 16
Figure 4a-b: Necklaces. 17
Figure 5: Necklaces at or above the .95 line. 18
Figure 6: Pieces chosen for the ideal coverage probability function. 19
Figure 7: The 95% coverage probability function for MS for $\lambda \in [0,.50]$. 20
Figure 8: Illustration of relationship between the cpf and the resulting CIs for MS. 21
Figure 9: Illustration of an issue that can cause gaps in confidence intervals. 22
Figure 10a-b: Solution 1 to the gap problem. 24
Figure 11a-b: Solution 2 to the gap problem. 25
Figure 12a-d: Shows the 95% cpf's for W, S, G, and MS respectively. 29
Figure 13: The 95% cpf's for G and MS for $\lambda \in [0,15]$. 30
Figure 14a-d: Illustrates a coincidental endpoint. 32
Figure 15: Shows two possible paths for a length optimal confidence procedure. 33
Figure 16: The 95% coverage probability function for B for $\lambda \in [0,.50]$. 37
Figure 17: Shows B violates the Least Cardinality property. 37
Figure 18: The 95% coverage probability function for CG for $\lambda \in [0,.50]$. 38
Figure 19: The 95% coverage probability function for KB for $\lambda \in [0,.50]$. 39
Figure 20a-b: A cpf comparison of MS with CG and KB respectively. 42
Figure 21a-d: Segments that CG, MS and KB have in common for $\lambda \in [0,25]$. 45
Figure 22: Illustrates how LC-p chooses coincidental endpoints. 48
Figure 23: The 95% coverage probability function for LC-15 for $\lambda \in [0,.50]$. 49
Figure 24: Confidence interval length rankings. 51
Figure 25a-b: Shows running averages and expected widths. 52
Figure 26: Average coverage as a function of $\lambda$ on $[0,30]$. 54
Figure 27: Average width vs. average coverage. 55
Figure 28: Average width vs. average coverage for LC-p for all integer values of $p$. 55
Figure 29: Confidence intervals for $x = 18$ for all methods. 56
List of Tables

Table 1: Observations for the amount of foot traffic each day for 14 days. 7
Table 2: Confidence intervals for λ. 7
Table 3: CIs for average number of people passing through the outlet on a single day. 7
Table 4: Confidence intervals for x = 23 from three methods. 26
Table 5: Table of the occurrence of gaps for λ ∈ [0,50]. 27
Table 6: Intervals for x = 23 and x = 47 for ST and MS. 36
Table 7: Average Coverage on [0,15]. 42
Table 8: Table of coincidental endpoints. 46
Table 9: 95% confidence intervals for G, S, & W for x = 0, 1, 2, ..., 50. 62
Table 10: 95% confidence intervals for CG, MS, and KB for x = 0, 1, 2, ..., 50. 63
Table 11: 95% confidence intervals for ST, B, and LC-15 for x = 0, 1, 2, ..., 50. 64
ABSTRACT

OPTIMAL CONFIDENCE INTERVALS FOR THE
EXPECTATION OF A POISSON RANDOM VARIABLE

By

Bret A. Holladay

Master of Science in Mathematics

Confidence intervals are a very useful tool for making inferences on the expectation, \( \lambda \), of a Poisson random variable. When making inferences about \( \lambda \) using a confidence interval, one would hope the process used to create the interval has the level of confidence it promises and also produces confidence intervals that are as short as possible so as to hone in on the true value of \( \lambda \). Many confidence procedures have been developed in an attempt to achieve these two goals; that is, to be the "shortest" strict confidence procedure. I discuss several of these methods through the perspective of coverage probability functions. I also introduce three new methods which are optimal according to Kabaila and Byrne's Inability to be Shortened property. One method is derived by first creating a specialized coverage probability function through an exhaustive graphical examination of all Poisson probability functions for a set of consecutive values. Subsequently, confidence intervals for \( \lambda \) with any desired confidence level can then be formed for all possible values of the observed Poisson random variable. The process for creating the coverage probability function of a Poisson confidence procedure and the resulting confidence intervals will be described in detail. Then from the insight gained from deriving this method two other high-performance methods are produced. Finally, the collection of intervals derived from these three methods will be compared with those derived from existing methods with respect to coverage and several different measures of length.
Chapter 1: Introduction

1.1 Confidence Intervals

Confidence intervals, also sometimes called interval estimators, are ways to estimate a target parameter \( \theta \) with an indication of how "confident" we are in the result. A confidence interval is an interval \((l, u)\) determined from sample data that comes with a confidence level. The interval estimates the location of \( \theta \); that is, with the interval we can conclude it is likely that \( l < \theta < u \). Suppose the confidence level is chosen to be 95%, then this means in the long run 95% of the intervals that could be constructed will contain the true value of \( \theta \). In other words, given a specific confidence interval obtained from some sample data, we can say that the interval was calculated by a method which gives correct bounds on \( \theta \) in 95% of all possible samples.

1.2 Poisson Process

A particular example of a parameter we might want to estimate using confidence intervals, and the subject of this thesis, is the expectation \( \lambda \) of a Poisson random variable. This problem will be explored in great detail; however, before discussing confidence intervals for \( \lambda \), we will establish a foundation for the Poisson distribution. Birnbaum (1954), Haight (1967), and Pitman (1993) were consulted for the following information on the Poisson distribution.

**Definition 1:** A **counting process**, \( N(t) \), is a representation of the total number of events up to time \( t \) with the following properties:

1. \( N(t) \geq 0 \ \forall t \)
2. \( N(t) \in \mathbb{Z} \)
3. \( s < t \Rightarrow N(s) \leq N(t) \) (\( N(t) \) is non-decreasing)
4. if \( s < t \), then \( N(t) - N(s) \) is the number of events occurring in \((s, t]\)

**Definition 2:** \( N(t) \) is said to have **independent increments** if \( N(t_1) - N(s_1) \) and \( N(t_2) - N(s_2) \) are independent random variables for any non-overlapping intervals \( (s_1, t_1] \) and \( (s_2, t_2] \).

**Definition 3:** \( N(t) \) is said to have **stationary increments** if the probability distribution of \( N(t) - N(s) \) depends only on the length of the interval \((s, t]\) and not on the location.

**Definition 4:** A **Poisson process** is a counting process, \( N(t) \), with the following properties:

1. \( N(0) = 0 \).
2. \( N(t) \) has independent increments.
3. \( N(t) \) has stationary increments.
4. No counted occurrences are simultaneous.
5. For any $t > 0$, $0 < P[N(t) > 0] < 1$.

For example a Poisson process $N$ might be one of the following counting variables:

1. The number of telephone calls into a 24 hour help line during an interval of time.
2. The number of times a Geiger counter is hit by radioactive particles in an interval of time.
3. The number of cars arriving at a toll booth in an interval of time.

Often a Poisson process, $N(t)$, is observed only at a particular value $t = T$; that is, $N(T)$ is the representation of the total number of events in a specified time interval of length $T$. In this case, the random variable $X = N(T)$ has the Poisson distribution. That is to say, the probability distribution for the number of events that occur in a Poisson process with a fixed time interval is the Poisson distribution. It should be noted that the Poisson distribution can also model the number of events in specified spatial regions having length, area, or volume. In this case, a Poisson process would be defined with analogous properties to the ones mentioned above. As a result, $N$ could also represent something like the number of blood cells distributed over an area on a slide.

### 1.3 Poisson Distribution

The Poisson distribution with parameter $\lambda$ is defined by the formula

$$p_\lambda(x) = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0,1,2, \ldots,$$

which represents the probability that a Poisson random variable $X$ takes on the value $x$; that is,$$
p_\lambda(x) = P(X = x | \lambda).

We will commonly deal with the sum $\sum_{i=x}^{y} p_\lambda(i)$, which represents $P(x \leq X \leq y | \lambda)$. So for convenience define

$$P_{x,y}(\lambda) := P(x \leq X \leq y | \lambda) = \sum_{i=x}^{y} p_\lambda(i).$$

**Theorem 1:** If a random variable $X$ has the Poisson distribution then $E(X) = \lambda$ and $Var(X) = \lambda$.

The proof of Theorem 1 can be found in Appendix B.
2.1 Formula-based Methods for Constructing Confidence Intervals for $\lambda$

Here we will discuss different closed form methods for constructing a $100(1 - \alpha)$% confidence interval, $(l_x, u_x)$ for $\lambda$ given $X = x$ observed events. There are quite a few closed form methods (see Barker (2002), Kabaila and Byrne (2005), Swift (2009) or Patil and Kulkarni (2012)); however, we will narrow our focus to only three.


The Wald method (W) is the "classic" approach to constructing confidence intervals. The Wald method is introduced in a large proportion of elementary statistics textbooks partly due to its simplicity; but as we will see, the method should not be used in practice because of its poor performance. The basic goal of any method is to find an interval $(l_x, u_x)$ for each $x$ that satisfies $P(l_x \leq \lambda \leq u_x) = 1 - \alpha$. The idea of W is to use a normal approximation. Because of the Central Limit Theorem, we know that $X \sim \text{Poisson}(\lambda)$ can be approximated by $X \approx \text{Normal}(\lambda, \lambda)$. This approximation is satisfactory when $\lambda$ is sufficiently large. We can then transform $X$ into a standard normal random variable, $Z = \frac{X - \lambda}{\sqrt{\lambda}}$. It follows then that $\frac{X - \lambda}{\sqrt{\lambda}} \sim \text{Normal}(0,1)$. Therefore,

$$P\left(-z_{\alpha/2} \leq \frac{X - \lambda}{\sqrt{\lambda}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

where $z_{\alpha/2}$ is the $100\left(\frac{\alpha}{2}\right)$th percentile of the standard normal distribution. The value of $z_{\alpha/2}$ can easily be obtained from a standard normal $Z$-table. Hence, after replacing $\sqrt{\lambda}$ by its maximum likelihood estimate $\sqrt{X}$ and isolating $\lambda$ in the center of the inequality we see that

$$P\left(X - z_{\alpha/2}\sqrt{X} \leq \lambda \leq X + z_{\alpha/2}\sqrt{X}\right) = 1 - \alpha.$$ 

Thus, we obtain the Wald (W) confidence limits,

$$l_x = x - z_{\alpha/2}\sqrt{x},$$

$$u_x = x + z_{\alpha/2}\sqrt{x}.$$ 

For a 95% confidence interval $z_{0.025} = z_{0.025} \approx 1.96$ and hence the confidence limits become
Method 2: Garwood (G) \cite{1,7,15}

A very commonly used method for constructing confidence intervals for the Poisson parameter $\lambda$ is the Garwood (1936) method (G). This is the method of choice for the popular statistical software StatXact. In addition, the commonly used R package exactci has the Garwood method as one of three options for Poisson confidence intervals. A detailed description of the derivation of G will not be described here. But to summarize it, because of the relationship,

$$\sum_{k=0}^{x} \frac{e^{-\lambda} \lambda^k}{k!} = P(\chi^2_v > 2\lambda)$$

where $\chi^2_v$ is a Chi-square random variable with $v = 2(1 + x)$ degrees of freedom we can solve the equations,

$$P(\chi^2_{2x} > 2l_x) = 1 - \alpha/2,$$
$$P(\chi^2_{2(x+1)} > 2u_x) = \alpha/2$$

to get the Garwood (G) confidence limits:

$$l_x = \frac{1}{2} \chi^2_{2x, \alpha/2},$$
$$u_x = \frac{1}{2} \chi^2_{2(x+1), 1-\alpha/2}.$$

Method 3: Scores (S) \cite{1,4,18}

The Scores method (S) is another method that uses a normal approximation. The interval is derived from isolating $\lambda$ in the equation

$$P\left(-z_{\alpha/2} \leq \frac{X - \lambda}{\sqrt{\lambda}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

as was done for W but, this time without replacing $\sqrt{\lambda}$ by its maximum likelihood estimate. After isolating $\lambda$ we obtain the Scores (S) limits,
At a 95% confidence level, these equations become,

\[ l_x = x + \frac{1}{2} \left( z_{\alpha/2} \right)^2 - z_{\alpha/2} \sqrt{ x + \frac{1}{4} \left( z_{\alpha/2} \right)^2 }, \]

\[ u_x = x + \frac{1}{2} \left( z_{\alpha/2} \right)^2 + z_{\alpha/2} \sqrt{ x + \frac{1}{4} \left( z_{\alpha/2} \right)^2 }. \]

2.2 Illustrative Examples

**Example 1: Lung Cancer**

The following example was modified from a similar example in Sahai and Khurshid (1993):

A researcher followed a large group of coal miners for several years and found that 18 died of lung cancer. For the general population we would expect 11 deaths in a group this size in this time period.

The appropriate model of the number of deaths from lung cancer is the Poisson distribution because the events, lung cancer cases, occur independently of each other and at an overall rate that we assume doesn’t change with time. \( \lambda \) represents the expected number of deaths among a group of coal miners this size for the given time period. We want to determine if the value of \( \lambda \) is larger than 11, which is the expected number of deaths from lung cancer for the general population.

We can compute a 95% confidence interval using W with the following formulas:

\[ l_x = x - z_{\alpha/2} \sqrt{x}, \]

\[ u_x = x + z_{\alpha/2} \sqrt{x}. \]
In our example, \( x = 18 \) observed deaths and \( 100(1 - \alpha)\% = 95\% \). At a 95% confidence level \( z_{\alpha/2} = z_{0.025} \approx 1.96 \).

So,

\[
\begin{align*}
l_{18} &= 18 - 1.96\sqrt{18} = 9.68, \\
u_{18} &= 18 + 1.96\sqrt{18} = 26.34.
\end{align*}
\]

Thus, the \( W \) interval is

\[
(l_{18}, u_{18}) = (9.68, 26.34)
\]

This says we are 95% confident that the expected number of the deaths from lung cancer, for a group of coal miners this size, is between 9.68 and 26.34. Since 11 falls in the confidence interval, we fail to conclude from this method that the excess deaths among the coal miners are due to anything more than just chance variation.

**Remark 1:** Later we will see that not all methods produce an interval that contains 11 and so the method a researcher chooses to use can drastically affect what they may ultimately conclude from a study.

For the next example we will find the following theorem useful:

**Theorem 2:** If \( N_1, N_2, ..., N_j \) are independent Poisson random variables with parameters \( \lambda_1, \lambda_2, ..., \lambda_j \) respectively, then \( N = \sum_{j=1}^{j} N_j = N_1 + N_2 + ... + N_j \) is a Poisson random variable with parameter \( \lambda = \sum_{j=1}^{j} \lambda_j = \lambda_1 + \lambda_2 + ... + \lambda_j \).

In summary, this theorem gives us that the sums of independent Poisson variables are Poisson. The proof of this theorem can be found in Appendix B.

**Example 2: Foot Traffic**

The following example was taken from Byrne (2006). Suppose a prospective purchaser of a retail outlet wants to estimate the foot traffic passing through the retail outlet to help make an informed decision on whether or not to buy. So, the purchaser counts the number of people passing through the retail outlet every day for 14 days. The purchaser's results are shown below in

Table 1.

Let \( X_i \) be the number of people passing through the retail outlet on day \( i \). Assume \( X_i \sim Poisson(\lambda_i) \). The goal is to estimate \( \lambda \), the average number of people passing through the outlet in a two week period. By Theorem 2,
\[ X = \sum_{i=1}^{14} X_i \sim \text{Poisson}(\lambda = \sum_{i=1}^{14} \lambda_i). \]

**Table 1:** Observations for the amount of foot traffic each day for 14 days.

<table>
<thead>
<tr>
<th>Day</th>
<th>Traffic</th>
<th>Day</th>
<th>Traffic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>86</td>
<td>8</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>96</td>
<td>9</td>
<td>132</td>
</tr>
<tr>
<td>3</td>
<td>129</td>
<td>10</td>
<td>149</td>
</tr>
<tr>
<td>4</td>
<td>146</td>
<td>11</td>
<td>118</td>
</tr>
<tr>
<td>5</td>
<td>107</td>
<td>12</td>
<td>143</td>
</tr>
<tr>
<td>6</td>
<td>138</td>
<td>13</td>
<td>125</td>
</tr>
<tr>
<td>7</td>
<td>95</td>
<td>14</td>
<td>100</td>
</tr>
</tbody>
</table>

The purchaser then estimates a confidence interval for the expected value, \( \lambda \), of the number of people passing through the outlet in 14 days using the W, S, and G confidence intervals for \( \lambda \). \( X \) is found to be 1654 and the three confidence intervals for \( \lambda \) are displayed in Table 2.

**Table 2:** Confidence intervals for \( \lambda \).

<table>
<thead>
<tr>
<th>Method</th>
<th>Lower limit</th>
<th>Upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garwood (G)</td>
<td>1575.241</td>
<td>1735.678</td>
</tr>
<tr>
<td>Wald (W)</td>
<td>1574.289</td>
<td>1733.711</td>
</tr>
<tr>
<td>Scores (S)</td>
<td>1576.187</td>
<td>1735.654</td>
</tr>
</tbody>
</table>

Dividing both sides of the intervals by \( n = 14 \), we can get confidence intervals for the average number of people passing through the outlet on a single day. These results are displayed in Table 3.

**Table 3:** CIs for average number of people passing through the outlet on a single day.

<table>
<thead>
<tr>
<th>Method</th>
<th>Lower limit</th>
<th>Upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garwood (G)</td>
<td>112.52</td>
<td>123.98</td>
</tr>
<tr>
<td>Wald (W)</td>
<td>112.45</td>
<td>123.84</td>
</tr>
<tr>
<td>Scores (S)</td>
<td>112.58</td>
<td>123.98</td>
</tr>
</tbody>
</table>

This example illustrates that for large \( X \) the G, W, and S intervals all give similar results.
2.3 Coverage Probability

From this point on, as done by Casella and Robert (1989), we will use half open confidence intervals of the form \([l, u]\) to avoid certain technical difficulties that arise when we use fully open or fully closed intervals. This change will not have any effect on any theoretical properties in regard to length or coverage. It allows computations to run more smoothly and also allows the possibility that \(0 \in [l_0, u_0]\) when \(l_0 = 0\). Casella and Robert (1989) make a great reference for the information in the remainder of this section.

**Definition 5:** Define a **confidence procedure** \(C\) to be an infinite collection of intervals \(C = \{[l_x, u_x], x = 0, 1, 2, \ldots\}\), where the observed value \(X = x\) results in the confidence interval \([l_x, u_x]\).

Given a confidence procedure \(C\), the coverage probability for a particular value of \(\lambda\) is the probability that you observe an \(x\) that gives a confidence interval that contains \(\lambda\). This probability can be found by summing the probabilities of all \(x\) values that achieve this. So the coverage probability of \(\lambda\) is given by

\[
\text{cov}(\lambda) = \sum_{x=0}^{\infty} I_{[l_x, u_x]}(\lambda) p_x(x),
\]

where \(I_{[l_x, u_x]}(\lambda)\) is an indicator function defined by

\[
I_{[l_x, u_x]}(\lambda) = \begin{cases} 
1 & \text{if } \lambda \in [l_x, u_x) \\
0 & \text{if } \lambda \notin [l_x, u_x) 
\end{cases}
\]

**Definition 6:** \(\text{cov}(\lambda)\) is said to be the **coverage probability function** (cpf) of \(C\).

\(\text{cov}(\lambda)\) can also be referred to as the coverage probability function of a particular method for constructing confidence intervals. This just means that \(\text{cov}(\lambda)\) is the coverage probability function for the confidence procedure that is produced by that method. There is a one-to-one correspondence between a \(1 - \alpha\) confidence procedure and its cpf. Moreover, examining Poisson confidence procedures through their cpfs is essential to understanding their behavior and performance.

2.4 Confidence Interval Performance

When making inferences using a confidence interval, one would hope the confidence interval came from a confidence procedure having the following two properties:
**Property 1:** *(Strict Coverage)* We would like the confidence procedure to produce confidence intervals that have a level of confidence greater than or equal to the one promised; that is, for all \( \lambda \) the chance that a confidence interval will contain \( \lambda \) should be at least as high as the chosen confidence level. This means for a confidence level of \( 100(1 - \alpha)\% \), we want \( \text{cov}(\lambda) \geq 1 - \alpha \) for all \( \lambda \).

**Property 2:** *(Length Optimality)* We would like the confidence procedure to have intervals of the shortest possible length so that we can hone in on the true value of \( \lambda \).

A confidence procedure possessing Property 1 is called strict. For a distribution with a finite number of possible values such as the binomial, length optimality can be assessed using average confidence interval length. In the case of the Poisson, however, length optimality is difficult to assess in this manner since the total number of intervals is infinite. In addition, comparing lengths of intervals from two confidence procedures is tricky since the confidence intervals of one confidence procedure may be shorter for some \( x \)'s but longer for others. Casella and Robert (1989) and Kabaila and Byrne (2005) have each come up with their own criteria for judging how well a confidence procedure performs in regards to the length of its intervals.

Casella and Robert consider an asymptotic criterion for confidence interval length through use of partial sums \( \sum_{x=0}^{K}(u_x - l_x) \). A strict confidence procedure \( C' = \{[l'_x, u'_x), x = 0, 1, 2, \ldots \} \) is said to dominate another strict confidence procedure \( C = \{[l_x, u_x), x = 0, 1, 2, \ldots \} \) if there exists an integer \( K_0 \) such that for all \( K > K_0 \), either

(i) \( \sum_{x=0}^{K}(u'_x - l'_x) < \sum_{x=0}^{K}(u_x - l_x) \)

or

(ii) \( \sum_{x=0}^{K}(u'_x - l'_x) = \sum_{x=0}^{K}(u_x - l_x) \) and \( \text{cov}_{C'}(\lambda) \geq \text{cov}_C(\lambda) \), with strict inequality for at least one \( \lambda \).

Note that this allows the possibility that \( C' \) has wider intervals than \( C \) for small values of \( x \).

On the other hand, Kabaila and Byrne have come up with the Inability to be Shortened property. A strict confidence procedure has the Inability to be Shortened property if increasing any lower endpoint or decreasing any upper endpoint causes the coverage probability to fall below \( 1 - \alpha \) for at least one \( \lambda \). In other words a confidence procedure has the Inability to be Shortened property if creating a new confidence procedure \( C' = \{[l'_x, u'_x), x = 0, 1, 2, \ldots \} \) with \( l'_x \geq l_x \) and \( u'_x \leq u_x \) for all \( x \) and with either \( l'_x > l_x \) or \( u'_x < u_x \) for at least one \( x \) causes \( \text{cov}(\lambda) < 1 - \alpha \) for at least one \( \lambda \).
Another way to assess the overall confidence interval length of a confidence procedure is to consider for each $\lambda$ the number of confidence intervals containing $\lambda$. This indirectly evaluates the performance of a confidence procedure in regard to length. Define an acceptance set of a confidence procedure to be the set $A_\lambda = \{x: l_x \leq \lambda < u_x\}$, for $\lambda \in [0, \infty)$. We assume that all acceptance sets consist of consecutive $x$-values. An explanation of why this is intuitive is given in section 3.1.

Define an acceptance curve (or acceptance region) associated with $A_\lambda$ as $P_{a,b}(\lambda) = P(X \in A_\lambda)$ considered as a function of $\lambda$, where $a = \min\{x: x \in A_\lambda\}$ and $b = \max\{x: x \in A_\lambda\}$. The cpf of a Poisson confidence procedure is a piecewise function formed by the collection of its acceptance curve segments.

**Definition 7:** We will say that a strict confidence procedure has the Least Cardinality property if all acceptance curves have minimal cardinality for all $\lambda$; i.e., for all $\lambda$ the acceptance curves $P_{a,b}(\lambda)$ of the cpf minimize the quantity $b - a + 1$.

If a confidence procedure has the Least Cardinality property then every $\lambda$ will be contained in the fewest possible number of confidence intervals and hence a confidence procedure having the Least Cardinality property tends to lead to short confidence intervals.

**Theorem 3:** If confidence procedure $C$ has the Least Cardinality property then it has the Inability to be Shortened property.

The proof of Theorem 3 can be found in Appendix B. Several papers: Casella and Robert (1989), Kabaila and Byrne (2001,2005), and Byrne (2006), favor either the Least Cardinality property or the Inability to be Shortened property, which by Theorem 3 is a consequence of the Least Cardinality property. However, we will see later that there are infinitely many confidence procedures which have this property and that a confidence procedure having this property may still be inferior to other methods when length is assessed interval by interval. Because of this fact, we later introduce several other criteria to compare the length performance of Poisson confidence procedures.

Another property that we would like a confidence procedure to have is the following:

**Property 3:** (Strictly Increasing Endpoints) A confidence procedure $C$ has strictly increasing endpoints if $l_x < l_{x+1}$ and $u_x < u_{x+1}$ for all $x$.

If we allow for equality in one or both of the above inequalities then we say that $C$ has nondecreasing endpoints.

**Theorem 4:** Assume that $x \in [l_x, u_x] \forall x$. Then if for some integer $k \geq 0$, the confidence procedure $C$ has $l_k > l_{k+1}$ (or $u_k > u_{k+1}$), then we can modify $C$ so that
\( l_k < l_{k+1} \) (or \( u_k < u_{k+1} \)) without affecting overall interval length but improving coverage.

Theorem 4 was modified from a proposition found in Casella and Robert (1989). This Theorem allows us conclude that any procedure which does not have increasing endpoints is inadmissible since we can put the endpoints in proper order without affecting net interval length and additionally create an increase in coverage in some locations. The term inadmissible is used to describe a procedure that is not worth using because it contains some major flaw or because it can be easily improved without causing a decrease in coverage or an increase in interval width. The overall interval length remains unchanged because in modifying the confidence procedure, whenever an increase is made in the length of one interval it is followed by a decrease of the same amount in another interval. In the statement of Theorem 4 it is assumed that \( x \in [l_x, u_x] \) \( \forall x \). This condition forces the MLE, which is \( x \) itself for the Poisson distribution, to be in the interval constructed. In all confidence procedures we investigated this condition holds and although not proven we suspect that all procedures that violate this condition are inadmissible.

Lastly we would like our confidence procedure to have the property below:

**Property 4:** (Gapless intervals) A confidence procedure \( C \) is gapless or has gapless intervals if all confidence sets are single intervals; i.e., there no confidence sets in \( C \) of the form \( [a, b] \cup [c, d] \) with gap \( [b, c] \).

The new method, which will be introduced shortly, produces a strict gapless confidence procedure that is length optimal according to both Kabaila and Byrne's length criterion and the Least Cardinality property and additionally has strictly increasing endpoints.

### 2.5 Preliminaries

Kabaila and Byrne (2001) and Crow and Gardner (1959) state several useful facts about Poisson probability functions which are included along with several other additional results in Propositions 5-8. Haight (1967) also published several of these results. These propositions will be useful for deriving the new method that will be discussed in the next section and for understanding how the cpf's of various methods look and behave.

**Proposition 5:** The graph of \( P_{x,y}(\lambda) \) attains its maximum at \( \lambda = \lambda_{max}(x, y) \), where

\[
\lambda_{max}(x, y) = \left[ (x)(x + 1) \cdots (y) \right]^{1/(y-x+1)}.
\]

Furthermore, \( P_{x,y}(\lambda) \) is strictly increasing on \( \lambda \in (0, \lambda_{max}(x, y)) \) and is strictly decreasing on \( \lambda \in (\lambda_{max}(x, y), \infty) \).
**Proposition 6:** For $\lambda \in (0, \infty)$, the graph of $P_{x,y}(\lambda)$ intersects the graph of $P_{x+1,y+1}(\lambda)$ at a single point, which is the maximum of the graph $P_{x+1,y+1}(\lambda)$.

![Figure 1: An illustration of Propositions 5 and 6. The graph of $P_{x,y}(\lambda)$ (dotted line) intersects the maximum of the graph of $P_{x+1,y+1}(\lambda)$ (solid line).](image)

**Proposition 7:**

a) $P_{x,y}(\lambda) > P_{x+1,y+1}(\lambda)$ for all $\lambda \in (0, \lambda_{max}(x + 1, y + 1))$

b) $P_{x,y}(\lambda) < P_{x+1,y+1}(\lambda)$ for all $\lambda \in (\lambda_{max}(x + 1, y + 1), \infty)$

c) $\lambda_{max}(x,y) < \lambda_{max}(x + 1, y + 1)$

d) $\max_{\lambda}\{P_{x,y}(\lambda)\} > \max_{\lambda}\{P_{x+1,y+1}(\lambda)\}$

**Proposition 8:** Suppose $x, y, \bar{x}, \bar{y}$ are nonnegative integers such that $\bar{x} \leq x$ and $y \leq \bar{y}$ where at least one of these inequalities is strict. Then $P_{x,y}(\lambda) < P_{\bar{x},\bar{y}}(\lambda)$ for all $\lambda \in (0, \infty)$.

The proofs of Propositions 5-8 can be found in the Appendix B. We will make extensive use of the above propositions in the derivation of the new method we develop below. We will call this new method the Modified Sterne's method (MS) since as we will see later...
its cpf is nearly identical to the cpf for Sterne's (1954) method but with some modifications. For the remainder of the paper we will choose our level of confidence to be 95%, but it should be noted that MS performs just as well at all other levels.
Chapter 3: The Modified Sterne's Method

3.1 MS

The approach of MS is to construct a confidence interval, for each possible number of observed events \(x\), by first creating a coverage probability function (cpf). This is the opposite from what would usually be done. In most circumstances a method is created and its performance can then be evaluated by examining the cpf produced by the method. But why not choose the ideal coverage probability function first, and then find the confidence intervals that would be given by this function? This approach to confidence intervals is modeled after work done by Schilling and Doi (2015) on confidence intervals for the one sample binomial problem. Their resulting method called LCO, short for Length/Coverage Optimal produced an optimal strict length minimizing binomial confidence procedure.

The coverage probability function will be a piecewise function strategically constructed so that we obtain a strict confidence procedure having the Least Cardinality property. Figure 2 is a plot of many of the possible curves from which the ideal coverage probability function can be constructed; i.e., it shows curves from which we can choose our acceptance regions. So it is the plot of functions from the infinite set,

\[ \{P_{i,j}(\lambda): i < j, i = 0, 1, 2 \ldots, j = 0, 1, 2 \ldots \}. \]

![Figure 2: This "sea of curves" is a plot of many of the functions from the set \(P_{i,j}(\lambda)\). The red and blue curves are the graphs of \(P_{0.5}(\lambda)\) and \(P_{3.4}(\lambda)\) respectively.](image)


For example, the red curve in Figure 2 is the graph of $P_{0.5}(\lambda)$; it shows the probability, as a function of $\lambda$, that $x$ falls between 0 and 5. The basic idea is if we were to let the confidence intervals for these $x$’s: 0, 1, 2, 3, 4, 5, all include the $\lambda$’s that correspond to the places where this curve is above .95, then the coverage probability for those $\lambda$’s will be above 95%.

**Remark 2:** Although Figure 2 is a plot of a finite number of curves, it should be noted that there are actually an infinite number curves. Additionally the curves continue on past $\lambda = 20$ since $\lambda \in [0, \infty)$.

**Remark 3:** The coverage probability function must be constructed using segments from the set $\{P_{L,j}(\lambda)\}$ as a direct consequence of the definition of coverage probability. The $x$’s must be consecutive since otherwise ourcpf could use pieces from curves like $P(1 \leq X \leq 2, 4 \leq X \leq 5| \lambda)$. It would then turn out that some values of $\lambda$ would be included in the confidence intervals for $x = 2$ and $x = 4$ yet would not be included in the confidence interval for $x = 3$, which would not make intuitive sense. Another possible problem is that the confidence intervals for some $x$’s could be the empty set.

This “sea of curves" seems nearly impossible to deal with because of the overwhelming amount of curves to choose from for any particular value of $\lambda$; to continue on, we must think of a practical way to reduce the number of curves under consideration.

### 3.2 Group the Curves

One idea to fix the problem of overcrowding mentioned above is to look separately at each group of curves that involve the same number of $x$ values. The group of all the curves that involve $k$ values of $x$ will make up the set,

$$\{P_{j,j+k-1}(\lambda) : j = 0, 1, 2, \ldots\}.$$  

For example, Figure 3 shows many of the curves that involve ten values of $x$. These curves make up the set

$$\{P_{j,j+9}(\lambda) : j = 0, 1, 2, \ldots\} = \{P_{0,9}(\lambda), P_{1,10}(\lambda), P_{2,11}(\lambda), \ldots\}$$

In order from largest to smallest by the height of their maximums, the curves in Figure 3 represent the functions $P_{0,9}(\lambda), P_{1,10}(\lambda), P_{2,11}(\lambda)$, .... The red and blue curves are the graphs of $P_{0,9}(\lambda)$ and $P_{11,20}(\lambda)$ respectively. With the exception of $P_{0,9}(\lambda)$, all the curves shown form a nonsymmetrical bell shape similar to the shape of $P_{1,1,20}(\lambda)$. In general, the graphs of $P_{j,j+k-1}(\lambda)$ will form a nonsymmetrical bell shape for all $j \neq 0$. When $j = 0$ the acceptance curve will
decrease similarly to $P_{0.9}(\lambda)$. We can visually verify the results of Proposition 6 by noting that $P_{0.9}(\lambda)$ intersects the maximum of $P_{1.10}(\lambda)$, $P_{1.10}(\lambda)$ intersects the maximum of $P_{2.11}(\lambda)$, etc.

![Figure 3: Curves $P_{j,j+9}(\lambda)$ which contain ten values of $x$. The dashed line crosses .95.](image)

Grouping the curves this way turns out to be very helpful as will be discussed in Remark 8 of Section 3.5 since the number of confidence intervals containing a fixed $\lambda$ is equal to the number of $x$'s involved in the segment used by the coverage probability function with respect to that $\lambda$. The idea now is that since each curve in a fixed group represent the same number of $x$'s they will work equally well for satisfying the Least Cardinality property. So when choosing among such curves for our coverage probability function, we should ideally choose the one with the highest coverage. This will give us greater confidence in the results of our confidence intervals while still maintaining the Least Cardinality property. What is desired of a confidence interval is that it is as precise as possible in trapping $\lambda$. If we can increase coverage without increasing "overall" confidence interval length, this will provide an additional benefit to the procedure. So, it is helpful to create a new plot from each of these groups of curves, one that, for each value of $\lambda$, takes the maximum of all the curves in the group. The rationale for this will be made more clear in Section 4.2.

### 3.3 Necklaces

We call this new collection of curves necklaces. Figure 4a is an example of one such necklace, a necklace created from the group of curves shown in Figure 3. We call this a
necklace of cardinality 10, denoted \( N_{10}(\lambda) \), since each segment of the curve involves ten \( x \)-values. The inset in Figure 4a gives a closer look at a small part of the necklace.

![Graph of \( N_{10}(\lambda) \) and necklaces of different cardinalities](image)

**Figure 4a-b**: Figure 4a(left) is a graph of the \( N_{10}(\lambda) \). Figure 4b(right) is a graph of necklaces of cardinalities 0, 1, 2, 3, ... The red curve in Figure 4b is the graph of \( N_{10}(\lambda) \).

In general, we will denote a necklace of cardinality \( k \) by \( N_k(\lambda) \). Since \( N_k(\lambda) \) is defined as

\[
N_k(\lambda) := \max_j \{ P_{j,j+k-1}(\lambda) : j = 0,1,2,\ldots \},
\]

by Proposition 7

\[
N_k(\lambda) = \begin{cases} 
P_{0,k-1}(\lambda) & \lambda \in [0, \lambda_{\max}(1,k)) \\
\vdots & \vdots \\
P_{2,k+1}(\lambda) & \lambda \in [\lambda_{\max}(2,k+1), \lambda_{\max}(3,k+2))
\end{cases}
\]

For the necklace of cardinality 10 depicted in Figure 4a this function becomes,

\[
N_{10}(\lambda) = \begin{cases} 
P_{0,9}(\lambda) & \lambda \in [0, \lambda_{\max}(1,10)) \\
P_{1,10}(\lambda) & \lambda \in [\lambda_{\max}(1,10), \lambda_{\max}(2,11)) \\
P_{2,11}(\lambda) & \lambda \in [\lambda_{\max}(2,11), \lambda_{\max}(3,12)) \\
\vdots & \vdots 
\end{cases}
\]

A careful comparison of Figure 3 and Figure 4a shows why \( N_{10}(\lambda) \) equals the above piecewise function.

**Remark 4**: As a consequence of the definition of \( N_k(\lambda) \), Proposition 8 implies that

\[
N_k(\lambda) \leq N_{k+1}(\lambda) \quad \forall \lambda \in (0, \infty).
\]
Remark 5: For a better understanding of the motivation for necklaces see Section 4.2.

If we choose pieces of the curves in Figure 4b below the .95 line for our cpf then there will be places where the coverage probability is below 95%; therefore, we can omit all the pieces of the graph below the .95 line from our selection. Figure 5 is a zoomed in version of Figure 4b with only the pieces at or above the .95 line. It is from this graph that we can now start to choose our acceptance regions and create our cpf. Because we are only looking at pieces of the curves above 95% we can already be certain that we are maintaining a 95% level of confidence; as a result, we are now just concerned with making sure that our resulting confidence procedure will possess the Least Cardinality property.

![Figure 5: Necklaces of cardinalities 0, 1, 2, 3, ... at or above the .95 line. The red curve is the graph of $N_{10}(\lambda)$.](image)

3.4 Choosing the Ideal Coverage Probability Function

To put it simply, to choose the ideal coverage probability function, one that possesses the Least Cardinality property and also maximizes coverage, we initially choose for each $\lambda$, the necklace with the smallest cardinality. Intuitively it makes sense that we want our coverage probability function to use the lowest possible curves, while still staying at or above .95. This is because naturally the higher the coverage the wider we would expect our intervals to be. For instance, 100% coverage would require intervals that cover every possible value of $\lambda$, creating intervals of infinite length. So, Remark 4 implies for each $\lambda$ our coverage probability function must use, of the necklaces still at or above .95, the one with least cardinality. Figure 6 shows in bold the pieces we need to choose to do this.
Again note that there are actually infinitely many curves and they continue beyond $\lambda = 50$; however, the pieces are chosen similarly everywhere.

**Figure 6:** Pieces chosen for the ideal coverage probability function. The red curve is the graph of $N_{10}(\lambda)$.

**Remark 6:** The vertical lines in Figure 6 (and in the cpf plots throughout the rest of the paper) are just a consequence of using graphical software and are not a part of the actual cpf. Additionally most the literature on this topic leave the lines in and so it almost customary to do so.

The following is a more detailed description of how to construct the ideal coverage probability function. Starting from $\lambda = 0$ we choose the values of our coverage probability function to take on the values of $N_1(\lambda)$, the necklace with minimum cardinality, until $N_1(\lambda)$ is no longer above .95. At the $\lambda$ for which $N_1(\lambda) = .95$, we must choose the values of our coverage probability function from the subsequent necklace, $N_2(\lambda)$ until it is no longer above .95; at this point we switch to $N_3(\lambda)$. Continuing in a similar fashion determines the ideal coverage probability function for $\lambda \in [0, \infty)$. The following piecewise function summarizes the above.

$$
cov_{MS}(\lambda) = \begin{cases} 
N_1(\lambda) & \lambda \in [0, \lambda_1) \\
N_2(\lambda) & \lambda \in [\lambda_1, \lambda_2) \\
N_3(\lambda) & \lambda \in [\lambda_2, \lambda_3) \\
\vdots & \vdots 
\end{cases}
$$

where $\lambda_i$ is the value of $\lambda$ satisfying $N_i(\lambda) = .95$. Figure 7 is the resulting cpf for $\lambda \in [0,50]$ without the clutter of the unused segments.
Remark 7: In most circumstances the value of $\lambda$ for which we switch from $N_k(\lambda)$ to the successive necklace, $N_{k+1}(\lambda)$ will be the unique value of $\lambda$ for which $N_k(\lambda) = .95$ as discussed above. However, in rare situations if we transition to $N_{k+1}(\lambda)$ when $N_k(\lambda) = .95$ we can cause some of the confidence intervals given by the cpf to have gaps. For example, we might produce a confidence interval of the form $[a, b) \cup [c, d)$ with gap $[b, c)$ where $a < b < c < d$. This is a highly undesirable property for a confidence interval. More information about gaps and how handle them, can be found in Section 3.6. But first, it will be helpful to understand how to construct a confidence interval given $X = x$ using our coverage probability function.

3.5 Constructing Confidence Intervals

This section covers how to use our coverage probability function to create confidence intervals for all values of $x$. Suppose the number of events a Poisson process produces is $x = 9$. The first step to creating a 95% confidence interval for $\lambda$, given $x = 9$, is to find all the segments from the coverage probability function that include $x = 9$. In other words, all the acceptance curves from the cpf of the form $P_{a,b}(\lambda)$ such that $a \leq 9 \leq b$.

The blue curves in Figure 8 are all the acceptance curves from the 95% coverage probability function that involve $x = 9$. These curves start at 4.46 and terminate at 17.3. Thus, our 95% confidence interval for $\lambda$ when $x = 9$ is $[4.46, 17.3)$, as shown. This is because by the definition of coverage probability if a piece of the cpf involves $x = 9$, then all the $\lambda$’s corresponding to this piece must be in the confidence interval.
Repeating this process we can obtain the confidence intervals for $\lambda$ for all the possible numbers of observed events $x$. Table 10 in Appendix A shows the 95% confidence intervals for $\lambda$ for $0 \leq x \leq 50$ given by MS.

Remark 8: At this point it should be clear why the best way to handle the "sea of curves" shown in Figure 2 was to group the curves by cardinality. This is because for any interval of $\lambda$'s if the cpf uses pieces with small cardinality then the number of intervals containing those $\lambda$'s will also be small. For example consider a relatively small interval $[a, b]$. If the pieces of the cpf corresponding to the $\lambda$'s from $[a, b]$ involve $k$ $x$-values then each $\lambda$ from interval $[a, b]$ will be contained in $k$ different confidence intervals. However if the pieces of the cpf corresponding to the $\lambda$'s from $[a, b]$ have cardinality, say, $k + 1$, then each $\lambda$ will be contained in $k + 1$ confidence intervals. Hence, as a general rule, using pieces with higher cardinality for the cpf will result in wider confidence intervals overall. Grouping the curves by cardinality makes it easy to determine for each $\lambda$ which available curves involve the fewest number of $x$-values. Thus, we can effortlessly construct the ideal cpf, one which produces a confidence procedure possessing the Least Cardinality property. Furthermore, during the construction of the ideal cpf we only use pieces at or above the .95 line, which produces a strict confidence procedure. Also, after grouping the curves by cardinality, for each $\lambda$ we used the highest
of all the curves in a fixed group, which maximized coverage. Now we proceed to discuss the gap problem.

### 3.6 Gaps and the Transition Between Successive Necklaces

Our rule for transitioning from one part of our cpf to the next is to transition from $N_k(\lambda)$ to the successive necklace, $N_{k+1}(\lambda)$ at the unique value of $\lambda$ satisfying $N_k(\lambda) = 1 - \alpha$. However, there are rare occasions in which this causes some of the confidence intervals given by the cpf to have gaps. For instance, for a 95% confidence procedure we encounter one of these instances during the transition from necklaces $N_{23}(\lambda)$ to $N_{24}(\lambda)$. Figure 9 helps illustrate why transitioning from $N_{23}(\lambda)$ to $N_{24}(\lambda)$ at the $\lambda$ satisfying $N_{23}(\lambda) = .95$ causes a gap in one of the confidence intervals.

**Figure 9**: Illustration of an issue that can cause gaps in confidence intervals.

Figure 9 shows the graph of the coverage probability function when the switch from $N_{23}(\lambda)$ to $N_{24}(\lambda)$ takes place at the $\lambda$ satisfying $N_{23}(\lambda) = .95$. The inset in Figure 9 gives a closer look at the intersection of $P_{23,46}(\lambda)$ and $P_{24,47}(\lambda)$. Calculating the 95% confidence interval for $x = 23$ from the cpf shown in Figure 9 we get $[14.92, 34.36) \cup [34.67, 34.81)$. This is because the segments of the coverage probability function involving $x = 23$ start at $\lambda = 14.92$, stop at $\lambda = 34.36 (\lambda : P_{23,45}(\lambda) = P_{24,46}(\lambda))$, start up again at $\lambda = 34.67 (\lambda : P_{24,46}(\lambda) = .95)$, and then finally end when
In general, the confidence interval for $\lambda$ corresponding to $x$ will have a gap if as $\lambda$ increases the cpf temporarily discontinues using pieces that involve $x$ to once again begin to use pieces which involve $x$ shortly after. Or in our case the interval corresponding to $x$ will have a gap if we switch from one necklace that no longer involves $x$ to another necklace that still does. To be more precise, when transitioning from $N_k(\lambda)$ to $N_{k+1}(\lambda)$ we depart from $N_k(\lambda)$ at the $\lambda$ satisfying $N_k(\lambda) = .95$. At this $\lambda$ we will leave a segment from a function of the form $P_{x,x+k-1}(\lambda)$ located on $N_k(\lambda)$ and switch to another segment from a function of the form $P_{y,y+k}(\lambda)$ located on $N_{k+1}(\lambda)$. If $x > y$ then the confidence interval we obtain for $x$ will have a gap and if $x \leq y$ then it will not. Graphically, if the transition from the function $P_{x,x+k-1}(\lambda)$ to the next necklace takes place to the left of the maximum of $P_{x,x+k}(\lambda)$ we will obtain a confidence interval for $x$ with a gap.

Below I present two solutions to the gap problem, only one of which will allow us to sustain a confidence procedure satisfying the Least Cardinality property. Notice in Figure 9 that a modification of the transition in question can only affect the confidence intervals for 23, 46 and 47. This is because the location of this transition determines where the cpf no longer uses pieces that involve $x = 23$ and begins to use pieces that involve $x = 46$ and $x = 47$; it will not affect any other interval.

**Solution 1: Filling the Gap**

One possible solution to the gap problem is to simply fill the gap. This amounts to transitioning to the next necklace sooner. In particular, if we transition to $N_{24}(\lambda)$ from a segment of $N_{23}(\lambda)$ that still involves 23, then we can fill the gap. $P_{23,45}(\lambda)$ is closest segment lying on $N_{23}(\lambda)$ that still involves 23 and transitioning from $P_{23,45}(\lambda)$ directly to $P_{23,46}(\lambda)$ at $\lambda = \lambda_{\text{max}}(24,46)$ will fill the gap. Figure 10a shows this plan in action. The transition from $N_{23}(\lambda)$ to $N_{24}(\lambda)$ takes place at the $\lambda$ satisfying $P_{23,45}(\lambda) = P_{24,46}(\lambda)$. Constructing the confidence interval for 23 from the coverage probability function in Figure 10a, we obtain $[14.92, 34.81]$, a gapless interval. This is precisely the interval obtained from filling in the gap of the interval $[14.92, 34.36) \cup [34.67, 34.81]$. Here the confidence intervals for 46 and 47 remain unchanged.
Figure 10a-b: Figure 10a(left) is an illustration of Solution 1 to the gap problem. Figure 10b(right) shows the cpf's from Figure 9 and Figure 10a in black and dotted orange respectively.

Figure 10b shows the difference Solution 1 makes on the cpf by comparing the cpf's from Figure 9 and Figure 10a on the same plot. In summary, when transitioning from \( N_k(\lambda) \) to \( N_{k+1}(\lambda) \) at the \( \lambda \) satisfying \( N_k(\lambda) = .95 \) causes an interval to have a gap, to fill in the gap we can transition to \( N_{k+1}(\lambda) \) from the end of the segment lying on \( N_k(\lambda) \) which precedes the segment on \( N_k(\lambda) \) that intersected the .95 line.

Merely filling the gap increases the total length of the intervals by the length of the gap since it uses a segment that is not of minimal cardinality. As a result, Solution 1 will not create a confidence procedure satisfying the Least Cardinality property. I propose a second solution, one which fixes the gap problem while still minimizing interval length.

**Solution 2: Relocating the Gap**

In order to preserve minimal confidence interval length we must stay on \( N_{23}(\lambda) \) until \( N_{23}(\lambda) = .95 \); otherwise, there will be a place where we are not using curves that involve the fewest number of \( x \)'s. Now, instead of jumping straight to \( N_{24}(\lambda) \) and landing on \( P_{23,46}(\lambda) \) as was done originally in Figure 9, we can instead jump to \( P_{24,47}(\lambda) \) and use this curve until it merges with \( N_{24}(\lambda) \); that is, instead of using \( P_{23,46}(\lambda) \) we can start using \( P_{24,47}(\lambda) \) sooner. We want to transition to \( P_{24,47}(\lambda) \) because it is the closest segment to \( P_{23,46}(\lambda) \) lying on \( N_{24}(\lambda) \) that no longer involves 23. Note that the cpf will temporarily be using a
segment that is not a part of any necklace and so we will temporarily sacrifice maximal coverage to eliminate the gap. This piece that lies on $P_{24,47}(\lambda)$, will only be used for a short distance until it merges with $N_{24}(\lambda)$ at $\lambda = \lambda_{\text{max}}(24,47)$; from here the cpf will continue to use the same pieces used in the original cpf. Figure 11a illustrates the modifications made by Solution 2 to the original cpf. Figure 11b compares the cpf from Figure 9(black) and the cpf from Figure 11a (dotted orange) on the same graph.

Table 4 below compares the confidence sets obtained from the original approach and Solutions 1 and 2.

Remark 9: Solutions 1 and 2 did not affect the interval for 46 and so only the intervals for 23 and 47 are different from those constructed from original cpf. However, there are solutions that fix the gap for $x = 23$ and also affect the interval for 46. For example, by using $P_{23,45}(\lambda)$ until $P_{23,45}(\lambda) = .95$ instead of using $P_{24,46}(\lambda)$ and then jumping to $P_{23,46}(\lambda)$ we fill the gap and also increase the lower endpoint for 46. This solution however ends up having longer intervals in totality for 23, 46, and 47 than the ones produced by Solution 2 and was only mentioned to demonstrate that the interval for 46 can be affected.
Table 4: Confidence intervals for \( x = 23 \) from three methods.

<table>
<thead>
<tr>
<th></th>
<th>( x = 23 )</th>
<th>( x = 47 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>[14.92, 34.36) ( \cup ) [34.67, 34.81)</td>
<td>[34.81, 62.36)</td>
</tr>
<tr>
<td>Solution 1</td>
<td>[14.92, 34.81)</td>
<td>[34.81, 62.36)</td>
</tr>
<tr>
<td>Solution 2</td>
<td>[14.92, 34.36)</td>
<td>[34.67, 62.36)</td>
</tr>
</tbody>
</table>

We can see from Table 4 that Solution 1 filled in the gap for the interval for \( x = 23 \) without affecting the length of any other intervals; for this reason, Solution 1 increased the total sum of the lengths of the intervals for \( x = 23 \) and \( x = 47 \) by 
\[
(34.67 - 34.36) + 0 = .31
\]. On the other hand, Solution 2 removed the extra piece [34.67, 34.81) from the interval for \( x = 23 \) that was causing the gap and adjoined it with the interval for \( x = 47 \). Therefore, Solution 2 decreased the length of the interval for \( x = 23 \) by 34.81 - 34.67 = .14, and then increased the length of the interval for \( x = 47 \) by the same amount. As a result of this relocation, Solution 2 successfully eliminated the gap for the interval for \( x = 23 \) while producing no net increase in confidence interval length. In the end, the total sum of the lengths of the intervals for \( x = 23 \) and \( x = 47 \) turned out to be the same for both the original cpf and for the cpf produced by Solution 2.

What we ultimately gain from Solution 2 is a confidence procedure which contains no gaps and satisfies the Least Cardinality property. And what we lose is a miniscule amount of coverage since \( P_{24,47}(\lambda) < P_{23,46}(\lambda) \) in the small interval in which we use \( P_{24,47}(\lambda) \) instead of \( P_{23,46}(\lambda) \) as used in the original cpf.

To summarize the approach of Solution 2, when transiting from \( N_k(\lambda) \) to \( N_{k+1}(\lambda) \) we usually depart from \( N_k(\lambda) \) when \( N_k(\lambda) = .95 \). Here we normally leave a segment \( P_{x,x+k-1}(\lambda) \) located on \( N_k(\lambda) \) and switch to another segment \( P_{y,y+k}(\lambda) \) located on \( N_{k+1}(\lambda) \). Whenever \( x > y \) we must instead transition from \( N_k(\lambda) \) directly to \( P_{x,x+k}(\lambda) \) and then use this segment until it merges with \( N_{k+1}(\lambda) \).

MS avoids the undesirable property of having confidence intervals with gaps by always choosing to take this approach when there is a situation where a gap would otherwise occur. However, the incidence of gaps is rare; for \( \lambda \in [0,50] \) for a 95\% confidence procedure there are only four places where gaps occur. Table 5 below gives a list of these four places. The instance discussed above was the second occurrence.
Table 5: Table of the occurrence of gaps for $\lambda \in [0,50]$.

<table>
<thead>
<tr>
<th>Gap #</th>
<th>Transition</th>
<th>Segments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$N_{19}(\lambda)$ to $N_{20}(\lambda)$</td>
<td>$P_{15,33}(\lambda)$ to $P_{14,33}(\lambda)$</td>
</tr>
<tr>
<td>2</td>
<td>$N_{23}(\lambda)$ to $N_{24}(\lambda)$</td>
<td>$P_{24,46}(\lambda)$ to $P_{23,46}(\lambda)$</td>
</tr>
<tr>
<td>3</td>
<td>$N_{26}(\lambda)$ to $N_{27}(\lambda)$</td>
<td>$P_{32,57}(\lambda)$ to $P_{31,57}(\lambda)$</td>
</tr>
<tr>
<td>4</td>
<td>$N_{27}(\lambda)$ to $N_{28}(\lambda)$</td>
<td>$P_{35,61}(\lambda)$ to $P_{34,61}(\lambda)$</td>
</tr>
</tbody>
</table>

From the above illustration we can also conclude that any confidence procedure with gaps is length inadmissible since we can easily relocate any gap without increasing overall net length.

3.7 Pseudocode for MS

For completeness the derivation of the cpf for the MS was long and thorough; however, there is actually a very simple algorithm for finding all the intervals for the $1 - \alpha$ MS confidence procedure $C = \{[l_x, u_x], x = 0, 1, 2, \ldots \}$.

Recall, $P_{a,b}(\lambda) = \sum_{l=a}^{b} p_{\lambda}(l)$ and $\lambda_{\text{max}}(a, b) = [(a)(a + 1) \cdots (b)]^{1/(b-a+1)}$. Define $\text{root}(a, b)$ to be the largest value of $\lambda$ for which $P_{a,b}(\lambda) = 1 - \alpha$. Note when $a = 0$ there is a unique solution to $P_{a,b}(\lambda) = 1 - \alpha$, but when $a > 0$ there are two solutions because of the bell shape of $P_{a,b}(\lambda)$. In the latter case $\text{root}(a, b)$ is the larger of the two roots. Then to find the $1 - \alpha$ MS confidence intervals use the three-step pseudocode below:

0. Set $l_0 = 0$ and let $a = 0, b = 0$.

1. If $\max[P_{a+1,b+1}(\lambda)] < 1 - \alpha$, set $l_{b+1} = \text{root}(a, b)$, let $b = b + 1$ and repeat this step.

2. Else let $a = a + 1, b = b + 1$ and set $l_b = \lambda_{\text{max}}(a, b), u_{a-1} = \lambda_{\text{max}}(a, b)$. Then go to 1.

Suppose all confidence intervals up to some $x = K$ are desired. Then one would simply need to run through the loop created by steps 1 and 2 until $a = K + 1$. Note that at this stopping point, in addition to having both the lower and upper limits of all the intervals for $x = 0,1,\ldots,K$, one will also have the lower limits of several intervals for $x$’s larger than $K$ whose upper limits were not yet calculated. This a direct result of the fact that
every time an upper limit $u_x$ is calculated so is a lower limit $l_y = u_x$ for some $y > x$. Such endpoints are called coincidental; that is, together the two equal limits, $u_x$ and $l_y$, make a pair of coincidental endpoints.

The information in this chapter does more than just introduce MS. This chapter can also be viewed as guidelines for designing one's own unique confidence procedure by first strategically constructing a Poisson cpf. As a result, one can create a confidence procedure having the properties desired.
Chapter 4: Comparing Methods

4.1 Comparison to Prevailing Formula-based Methods

To really evaluate the performance of MS a comparison to some of the widely used formula based methods is helpful.

Figure 12a-d shows the graphs of the 95% coverage probability functions for W, S, G and MS respectively.

Figure 12a-d: Shows the 95% cpf's for W, S, G, and MS respectively.
Of the four cpf's depicted, as can be seen, only G and MS produce strict confidence procedures. W and S do not satisfy the minimum coverage probability requirement of .95. In fact, the cpf for W in particular is rarely above .95. The alleged 95% confidence intervals given by W shockingly have coverage way below what was promised. On the other hand, S seems to almost average .95 over \( \lambda \in [0,50] \), but is still not in play if we adhere to the Strict Coverage property of a confidence procedure.

After the first round of eliminations, we find ourselves left with MS and G. Next we will compare the lengths of the confidence intervals for MS and G for \( x \)'s from 0 to 50. The lengths of the confidence intervals produced by MS are shorter than the lengths of the confidence intervals produced by G for 49 of these 51 \( x \) values. This can be attributed to the fact that the cpf for G is much more conservative then the cpf for MS. Figure 13 illustrates how much higher the coverage is for G as compared to MS, which results in the excess length of the confidence intervals produced by G. Additionally G fails to satisfy both the Inability to be Shortened property and the Least Cardinality property. All in all, MS outperforms the other three contenders discussed. Only G and MS are strict, but MS produces shorter intervals due to the fact that it has a cpf that is closer to the nominal confidence level while still always remaining at or above it.

![Figure 13: The 95% cpf's for G and MS for \( \lambda \in [0,15] \).](image)

**Remark 10:** If we relax the .95 minimum coverage requirement and allow coverage to be on average approximately equal to the confidence level, then S performs quite well. For \( x \)'s from 0 to 50 S has shorter intervals than both MS and G. However, we could also modify MS to have a similar averaging property, that is, to have average coverage approximately equal to .95. This will result in intervals having even shorter average length then S.
4.2 Optimality

Casella and Robert (1989) identified the entire collection of minimal cardinality confidence procedures that maintain the nominal confidence level. Casella and Robert argue that certain endpoints are forced and any least cardinality procedure must agree on these endpoints but other endpoints are not unique and have a whole range of possible values. They call the latter coincidental endpoints. Coincidental endpoints were first introduced in chapter 3; they are endpoints that are both an upper endpoint for one $x$ and lower endpoint for another. Shifting each of these points within a certain range has no effect on the overall length of a confidence procedure's intervals since this will just decrease the length of one interval by the same amount it increases the length of another. MS produces a confidence procedure that not only belongs to this collection of strict least cardinality confidence procedures, but also has the highest coverage among the entire collection. This is due to the fact that MS always uses the highest available curves for its cpf.

One example of a coincidental endpoint for MS is the point that represents the upper limit for $x = 7$ and the lower limit for $x = 22$. This sort of point is a coincidental endpoint because transitioning from the curve $P_{7,21}(\lambda)$ to the curve $P_{8,22}(\lambda)$ forces the confidence intervals for 7 and 22 to share an endpoint. This is because the $\lambda$ for which we switch from $P_{7,21}(\lambda)$ to $P_{8,22}(\lambda)$ is the place where cpf stops using curves that involve $x = 7$ and begins to use curves which involve $x = 22$. Hence this $\lambda$ will be the upper limit of the confidence interval for $x = 7$ and the lower limit for the confidence interval for $x = 22$.

Now since the choice of $\lambda$ for which we switch from $P_{7,21}(\lambda)$ to $P_{8,22}(\lambda)$ only affects the lengths of the confidence intervals for 7 and 22, and since a decrease of the length of one of these intervals increases the length of other by the same amount, the transition can take place at a $\lambda$ from a range of different values. Figure 14a shows in purple all the possible places that the transition between $P_{7,21}(\lambda)$ and $P_{8,22}(\lambda)$ can take place. The red and blue dots represent the location of the smallest solution for $\lambda$ of $P_{8,22}(\lambda) = .95$ and largest solution for $\lambda$ of $P_{7,21}(\lambda) = .95$ respectively. If we were to switch from $P_{7,21}(\lambda)$ to $P_{8,22}(\lambda)$ before the red dot or after the blue dot this would cause the cpf to take on values below .95. Hence, the earliest we can transition is at $\lambda = 13.76$ (red dot) and the latest we can transition is at $\lambda = 14.49$ (blue dot). Figure 14b and Figure 14c show what the corresponding cpf's would look like if we transition at these two extreme values.
Since at any coincidental endpoint we have many choices of where to transition while still maintaining the least cardinality property and the nominal confidence level, we can optimize coverage for "free" by choosing to transition at the place that yields the maximum coverage over the interval of choice. Thus, in this example for each $\lambda$, we should choose the cpf to take on the maximum value of the two curves $P_{7,21}(\lambda)$ and $P_{8,22}(\lambda)$ at each point where these curves overlap. By Proposition 7 this just amounts to using $P_{7,21}(\lambda)$ up to $\lambda_{\text{max}}(8, 22) = 14.34$ (green dot) and then after using $P_{8,22}(\lambda)$. But, this is exactly how we defined necklaces. In fact, this is the reason why necklaces were used in the construction of the cpf for MS. Figure 14d shows what the cpf looks like when we transition at $\lambda_{\text{max}}(8, 22)$. Note that in regards to the Least Cardinality property we can transition anywhere on the purple line segment since both $P_{7,21}(\lambda)$ and $P_{8,22}(\lambda)$ have cardinality 15.

**Remark 11:** It is possible for a minimal cardinality procedure to avoid using the segment $P_{7,21}(\lambda)$ all together. For example, see Figure 15 below. The cpf could skip $P_{7,21}(\lambda)$ and transition directly from $P_{6,20}(\lambda)$ to $P_{8,22}(\lambda)$. This transition can take place for any $\lambda$ on the purple line segment in Figure 15. However, even in this case the upper limit for $x = 7$ and the lower limit for $x = 22$ are still
coincidental. This choice however would not only decrease coverage but it would also cause the upper and lower limits for 6 and 7 of the confidence procedure to be the same, which violates The Strictly Increasing Endpoints property. We will investigate this property more later on. Also notice, the cpf must at some point use $P_{8,22}(\lambda)$ since no other segment of cardinality 15 takes on values above the .95 line for $\lambda \in [14.49, 14.92]$. Once $P_{7,21}(\lambda)$ falls below .95 at $\lambda = 14.49$, $P_{8,22}(\lambda)$ is the last remaining segment of cardinality 15 above .95 and must be used so that we are using pieces of minimum cardinality for as long as possible. In fact, for increasing $\lambda$ the last piece remaining above the confidence level in any group of fixed cardinality will always be used and must be used until it falls below the confidence level in order to satisfy the Least Cardinality property.

![Figure 15](image)

**Figure 15:** Shows two possible paths that a length optimal confidence procedure can take to $P_{8,22}(\lambda)$.

Coincidental endpoints are not at all rare; in fact, the majority of confidence interval endpoints are coincidental endpoints. The following discussion will clarify when and why an interval endpoint will be coincidental, and thereby open to choice within a certain range, rather than fixed.

The cpf's of any strict length minimizing confidence procedure involve only two types of transitions:

1. **Type I:** A transition from a segment of cardinality $k$ to a segment of cardinality $k + 1$.
2. **Type II:** A transition between two segments of equal cardinality.

For a length optimal confidence procedure all Type I transitions correspond to a fixed endpoint and all Type II transitions correspond to a coincidental endpoint.
Type I transitions take place between two pieces of the form $P_{x,x+k-1}(\lambda)$ and $P_{y,y+k}(\lambda)$. If $x > y$ the confidence interval for $x - 1$ will have a gap. When gaps occur we modify the transition so that $x = y$. So assume $x \leq y$. Then for each $k$ all length minimizing confidence procedures transition from a segment of cardinality $k$ to a segment of cardinality $k + 1$ at the unique location where the last remaining piece of cardinality $k$ that is still above the confidence level decreases to the confidence level. This is because we have to use the segments of the smallest cardinality that are still at or above the confidence level until the last possible moment in order to satisfy the Least Cardinality property. The location of this transition determines the value of $l_{y+k}$ which in turn is a forced endpoint.

Type II transitions take place between two pieces of the form $P_{x,y}(\lambda)$ and $P_{x+k,y+k}(\lambda)$ for $k \geq 1$. For illustration purposes assume $k = 1$ and consider the transition between $P_{x,y}(\lambda)$ and $P_{x+1,y+1}(\lambda)$. This transition corresponds to the location of a coincidental endpoint since this is the place where we stop using the pieces that involve $x$ and begin to use the pieces that involve $y + 1$. This forces $u_x = l_{y+1}$ to be a coincidental endpoint. Similar conclusions can be made for $k > 1$. The location of this transition can take place anywhere between the smallest solution for $\lambda$ of $P_{x+k,y+k}(\lambda) = 1 - \alpha$ and the largest solution for $\lambda$ of $P_{x,y}(\lambda) = 1 - \alpha$; any choice made outside this range causes the coverage probability to fall below $1 - \alpha$.

The coverage probability function for MS has all Type I transitions between $N_k(\lambda)$ and $N_{k+1}(\lambda)$ at the $\lambda$ satisfying $N_k(\lambda) = .95$. All Type II transitions for MS are between pieces of the form $P_{x,y}(\lambda)$ and $P_{x+1,y+1}(\lambda)$; that is, the transition between two pieces of equal cardinality is only done between consecutive pieces. The transition occurs at $\lambda = \lambda_{max}(x + 1, y + 1)$ which results in maximal coverage. Therefore, MS satisfies the aforementioned criteria necessary to be a minimal cardinality procedure and additionally maximizes coverage by making the optimal choice for coincidental endpoints. Hence, MS is a member of the collection of minimal cardinality confidence procedures, but also has the additional property of having the maximum coverage at each $\lambda$ of the whole collection.

4.3 Notable Methods

Methods based on closed form formulas such as G, S, and W are not often strict. The ones that are strict, such as G, do not satisfy the Least Cardinality property and thus are length inadmissible as they produce intervals that can be shortened without causing their
cpfs to drop below the confidence level. So we will turn our focus to non-closed form methods.

One way to define a $100(1 - \alpha)\%$ confidence procedure is given $\lambda$, for each $x$ include $\lambda$ in the confidence interval for $x$ if $x$ is not "rare" for that $\lambda$. This means that the coverage probability for $\lambda$ will be $\text{cov}(\lambda) = \sum_x p_{\lambda}(x)$ where the sum is taken over all $x$'s which are considered not "rare" with respect to the fixed $\lambda$. Looking at a confidence procedure in this light will be useful in defining and comparing Sterne's (1954) and Blaker's (2000) methods.

**Method 4: Sterne (ST)**

For Sterne's (1954) method (ST) fix $\lambda$ and sum the Poisson probabilities from smallest to largest until the sum is as large as possible without exceeding $\alpha$. The “rare” values of $x$ are precisely those contributing to this sum. So any $x$ that did not contribute to the sum includes $\lambda$ in its confidence interval. Alternatively one can go the other direction and find directly which $x$'s are not rare; that is, sum the Poisson probabilities from largest to smallest until the first time the total probability equals or exceeds the required $100(1 - \alpha)\%$ confidence level. The confidence intervals for the $x$'s included in this sum all contain $\lambda$. Schilling and Doi (2015) and Swift (2009) are good references for ST.

Sterne's method turns out to be nearly identical to the Modified Sterne's method (MS). This is because for a given $\lambda$ by summing the Poisson probabilities from largest to smallest we always make it to $1 - \alpha$ using the fewest amount of $x$'s (Least Cardinality property). Additionally summing this way causes the sum to be as large as possible for this fixed number of $x$'s (maximal coverage). In other words, for each $\lambda$, ST always uses the necklace of least cardinality at or above $1 - \alpha$. However, as seen in Section 3.6 this causes some intervals to have gaps. In contrast, when there is a situation where a gap could occur, MS will temporarily use a curve not on any necklace to eliminate the gap. So despite the large difference in the way the two methods were explained and derived, the methods will only differ on the small fraction of intervals affected by the occurrence of gaps, but will agree for all other intervals.

For instance, consider the gap discussed in Section 3.6. In that example the graph of the cpf referred to as the "original cpf" was actually the cpf for ST. The two intervals affected by this gap are shown below in Table 6. Recall that MS removed the extra piece $[34.67, 34.81]$ from the interval for 23 that was causing the gap in the "original cpf" and adjoined it with the interval for 47. Now since the intervals for both $x = 23$ and $x = 47$ were affected by the gap, MS and ST differ with respect to these two intervals. Similar differences between the two methods occur for the small proportion of other intervals that are affected by the occurrence of gaps.
Table 11 in Appendix A shows the 95% confidence intervals for ST for \( x = 0, 1, \ldots, 50 \). If a confidence interval has a gap (i.e., is a union of two intervals) then the upper and lower limits of the first interval are listed in the "LL" column and the upper and lower limits of the second interval are listed in the "UL" column.

**Table 6: Intervals for \( x = 23 \) and \( x = 47 \) for ST and MS.**

<table>
<thead>
<tr>
<th>Method</th>
<th>( x = 23 )</th>
<th>( x = 47 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST</td>
<td>[14.92, 34.36) ( \cup ) [34.67, 34.81)</td>
<td>[34.81, 62.36)</td>
</tr>
<tr>
<td>MS</td>
<td>[14.92, 34.36)</td>
<td>[34.67, 62.36)</td>
</tr>
</tbody>
</table>

To help explain why ST obtains confidence intervals with gaps, consider the confidence interval corresponding to \( x = 23 \). This \( x \) value is rare in the sense defined above for all \( \lambda \leq 14.92 \) and for all \( \lambda \geq 34.81 \); additionally, it can be checked that \( x = 23 \) is rare with respect to every \( \lambda \) in the interval [34.36, 34.67]. However, \( x = 23 \) is not rare for any \( \lambda \) in [14.92, 34.36) \( \cup \) [34.67, 34.81]. Because the Modified Sterne's (MS) method improves upon ST by eliminating gaps and not affecting overall confidence interval length, ST is length inadmissible.

**Method 5: Blaker (B)**

Define the tail probability of an observed value \( x \) to be the minimum of \( P(X \leq x|\lambda) \) and \( P(X \geq x|\lambda) \). Then for a fixed \( \lambda \), the Blaker (2000) method (B) defines \( x \) as rare if the probability of observing something with a tail probability as small as that of \( x \) is does not exceed \( \alpha \). In other words, given \( \lambda \) and \( x \) determine all values that have tail probabilities less than or equal to the tail probability of \( x \) (\( x \) is one of these values). If the sum of the probabilities for these values is less than or equal to \( \alpha \) then \( x \) is rare for \( \lambda \). Hence, \( \lambda \) is only included in the confidence intervals for the \( x \)'s that do not have this property. Figure 16 shows the 95% coverage probability function for B.
Figure 16: The 95% coverage probability function for B for \( \lambda \in [0, 50] \).

B violates both the Least Cardinality property and the Inability to be Shortened property. Figure 17 shows a portion of the 95% cpf for B where both properties are violated.

Figure 17: Illustration showing that B violates the Least Cardinality property and the Inability to be Shortened property.

Clearly since the cpf transitions from \( P_{0.6}(\lambda) \) to \( P_{0.7}(\lambda) \) before \( P_{0.6}(\lambda) \) decreases to .95 it violates the Least Cardinality property. Now to see how the Inability to be Shortened property is violated consider the lower endpoint for \( x = 7 \). This endpoint corresponds to the location of the transition between \( P_{0.6}(\lambda) \) and \( P_{0.7}(\lambda) \), which occurs at 3.162. We can
increase the value of \( l_7 \) to any value in the interval \((3.162, 3.285)\), thus shortening the confidence interval for \( x = 7 \), without causing coverage to fall below .95. Hence, B does not satisfy the Inability to be Shortened property. Note that 3.285 is the value of \( \lambda \) for which \( P_{0.6}(\lambda) \) decreases to .95 and increasing the value of \( l_7 \) to any value in the interval \((3.162, 3.285)\) corresponds to transitioning from the dotted line in Figure 17 to \( P_{0.7}(\lambda) \) at \( \lambda = l_7 \).

**Method 6: Crow and Gardner (CG)**

Crow and Gardner (1959) developed another modified version of Sterne's method. For the Crow and Gardner method (CG), for every \( \lambda \) look at the corresponding segment \( P_{x,y} \) used in the cpf for ST and if there exist other segments of the same cardinality \( P_{x+k,y+k} \) for \( k \geq 1 \) such that \( P_{x+k,y+k}(\lambda) \geq 1 - \alpha \) then use the one with the maximum value of \( k \). CG is strict and satisfies the Least Cardinality property (see Crow and Gardner (1959) or Casella and Robert (1989)). Figure 18 shows the 95% coverage probability function for CG. We will describe the cpf for CG in more detail during the comparison of MS, CG, and KB (yet to be introduced) in the next section.

![Figure 18: The 95% coverage probability function for CG for \( \lambda \in [0,50] \).](image)

**Method 7: Kabaila and Byrne (KB)**
Kabaila and Byrne (2001) give an algorithm for calculating a $100(1 - \alpha)\%$ confidence procedure, which goes as follows:

For every $x \geq 1$ define

$$r(x) := \min_s \left\{ s \in \mathbb{Z}^+ : \max_\lambda \left\{ P_{x-s,x-1}(\lambda) \right\} > 1 - \alpha \right\}$$

And for each $x \geq 0$ define

$$p(x) := \min_q \left\{ q \in \mathbb{Z}^+ : \max_\lambda \left\{ P_{x+1,x+q}(\lambda) \right\} > 1 - \alpha \right\}$$

In other words, of all the functions $\{P_{x-s,x-1}(\lambda) : s \in \mathbb{Z}^+\}$ that are above $1 - \alpha$, $r(x)$ is the positive integer $s$ that gives the one of least cardinality. Similarly of all the functions $\{P_{x+1,x+q}(\lambda) : q \in \mathbb{Z}^+\}$ that are above $1 - \alpha$, $p(x)$ is the positive integer $q$ that gives the one of least cardinality.

Next, set $l_0 = 0$. Then for each $x \geq 1$ let $l_x$ be the largest solution for $\lambda$ of $P_{x-r(x),x-1}(\lambda) = 1 - \alpha$ and for each for each $x \geq 0$ let $u_x$ be the largest solution for $\lambda$ of $P_{x+1,x+p(x)-1}(\lambda) = 1 - \alpha$.

Kabaila and Byrne (2001) prove that their method is strict and that it has the Inability to be Shortened property. Additionally, from the definitions of $r(x)$ and $p(x)$ it follows that KB has the Least Cardinality property and thus belongs to the collection of strict length-minimizing confidence procedures. Figure 19 shows the 95% coverage probability function for KB.

**Figure 19**: The 95% coverage probability function for KB for $\lambda \in [0, 50]$. 

![Figure 19: The 95% coverage probability function for KB for $\lambda \in [0, 50]$.](image)
4.4 Comparison of the Least Cardinality Procedures

We now investigate how the cpf's for the Kabaila and Byrne (KB) and Crow and Gardner (CG) methods compare with that of MS. Figure 21a-d (page 45) will be helpful in discussing the similarities and differences between the three methods. Figure 21a shows all the segments which the three methods use. They are color coded by cardinality where different shades of one color represent different pieces of equal cardinality. Respectively Figure 21b-d shows how the cpf's for CG, MS, and KB use the pieces in Figure 21a.

Since all three method satisfy the Least Cardinality property Type I transitions from segments of cardinality \( k \) to a segment of cardinality \( k + 1 \) are forced to occur in the same locations for all three methods. These transitions occur between two pieces of the form \( P_{x,x+k-1}(\lambda) \) and \( P_{y,y+k}(\lambda) \) where \( P_{x,x+k-1}(\lambda) \) is the last remaining segment of cardinality \( k \) still above \( 1 - \alpha \). The transitions occur when \( P_{x,x+k-1}(\lambda) \) decreases to \( 1 - \alpha \). MS and KB both transition to a segment \( P_{y,y+k}(\lambda) \) having \( y = x \), but CG will transition to the lowest available segment of cardinality \( k + 1 \) which is often of the form \( P_{y,y+k}(\lambda) \) for \( y > x \). Since KB does Type I transitions between the same pieces at the identical locations as MS, near these locations the cpf's for both methods are identical. On the other hand, CG is typically quite different near Type I transitions.

Type II transitions for all three methods only occur between consecutive pieces of the form \( P_{x,y}(\lambda) \) and \( P_{x+1,y+1}(\lambda) \). Through the use of necklaces MS transitions at \( \lambda = \lambda_{\text{max}}(x + 1, y + 1) \). For CG this Type II transition occurs when \( P_{x+1,y+1}(\lambda) \) first attains \( 1 - \alpha \); that is, it takes place at the smallest solution for \( \lambda \) of \( P_{x+1,y+1}(\lambda) = 1 - \alpha \). For KB the transition occurs at the largest solution for \( \lambda \) of \( P_{x,y}(\lambda) = 1 - \alpha \) which is a direct consequence of how \( l_x \) and \( u_x \) are defined. All Type II transitions for KB occur between the same pieces as MS. And since MS and KB are identical near Type I transitions it is only the location of Type II transitions that distinguish the cpf's for the two methods.

For a visual comparison of Type I transitions for the three methods see Figure 21a-d and examine the Type I transition between segments of cardinality 10 and 11. The last remaining segment of cardinality 10 at or above the .95 line is \( P_{2,11}(\lambda) \); as a result, all three methods transition from this segment when \( P_{2,11}(\lambda) \) decreases to .95. KB and MS then both transition to \( P_{2,12}(\lambda) \), while CG transitions to the lowest available segment of cardinality 11, namely \( P_{3,13}(\lambda) \).

For a visual example of how the cpf's for the three methods differ with respect to Type II transitions see Figure 14a-d (page 32). Figure 14a-d illustrates all the possible places for
which a 95% minimal cardinality confidence procedure can transition between the two segments $P_{7,21}(\lambda)$ and $P_{8,22}(\lambda)$. MS transitions as shown in Figure 14d at $\lambda = \lambda_{max}(8,22)$. On the other hand, CG transitions as shown in Figure 14b at the smallest solution for $\lambda$ of $P_{8,22}(\lambda) = .95$, and KB transitions as shown in Figure 14c at the largest solution for $\lambda$ of $P_{7,21}(\lambda) = .95$. For CG coincidental endpoints are always the minimum of all permissible values; whereas, coincidental endpoints for KB are always the maximum of all allowable values. Thus, as discussed in Section 4.2, neither KB nor CG maximize coverage.

Figure 20a and Figure 20b compare the cpf's for CG and KB respectively with the cpf for MS. The segments which are used by MS, but not by the featured method in each graph are shown with dotted lines. The gray regions in Figure 20a and Figure 20b show the extra coverage that MS consistently obtains by always transitioning as shown in Figure 14d in contrast to transitioning as early as possible (CG) or as late as possible (KB). The corresponding insets in Figure 20a and Figure 20b give a closer look at the last two shaded regions before $\lambda = 15$ for each graph. These insets show the same pieces that were involved in the coincidental endpoint illustration in Figure 14a-d. Additionally the average coverage for all three methods on the interval [0,15] is listed in Table 7. All three methods have the Least cardinality property, but both KB and CG fail to obtain coverage as high as that of MS.

In fact, not only does CG not maximize coverage, it does not have strictly increasing endpoints. For example, $l_7 = l_8 = 3.285$. Figure 21b shows graphically why $l_7 = l_8$ for CG. The first time the cpf uses a segment which involves $x = 7$ is when it begins to use $P_{1,8}(\lambda)$ at the $\lambda$ satisfying $P_{0,6}(\lambda) = .95$. But, this is also the location for which the cpf first begins to use segments involving $x = 8$ which forces the lower limits for 7 and 8 to be identical. In contrast, both MS and KB avoid violating the strictly increasing endpoints property by first transitioning to $P_{0,7}(\lambda)$ at the $\lambda$ satisfying $P_{0,6}(\lambda) = .95$ and using this segment until they transition between $P_{0,7}(\lambda)$ and $P_{1,8}(\lambda)$. In this way both MS and KB start using pieces which involve $x = 7$ and $x = 8$ for different values of $\lambda$, ensuring that $l_7 \neq l_8$. 
Figure 20a-b: A comparison of the 95% cpf of MS with the 95% cpf's of CG and KB respectively. In each graph the segments used by MS, but not by the featured method are shown with dotted lines.

Table 7: Average Coverage on [0,15].

<table>
<thead>
<tr>
<th>Method</th>
<th>Average Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>0.96255</td>
</tr>
<tr>
<td>CG</td>
<td>0.96032</td>
</tr>
<tr>
<td>KB</td>
<td>0.96186</td>
</tr>
</tbody>
</table>
In general, a confidence procedure having the least cardinality property will violate the strictly increasing endpoints property if a Type I transition occurs between two pieces \( P_{x,x+k-1}(\lambda) \) and \( P_{y,y+k}(\lambda) \) for \( y > x \). This is because the cpf will start to use pieces which involve \( x + k, x + k + 1, \ldots, y + k \) concurrently which forces \( l_{x+k} = l_{x+k+1} = \cdots = l_{y+k} \). Similarly if a Type II transition occurs between two segments of the form \( P_{x,y}(\lambda) \) and \( P_{x+k,y+k}(\lambda) \) for \( k > 1 \) then the strictly increasing endpoints property is violated since this forces \( l_{y+1} = l_{y+i} \) for \( 1 \leq i \leq k \) and \( u_x = u_{x+i} \) for \( 0 \leq i \leq k - 1 \). CG only violates this property during Type I transitions. In fact from Figure 21b it’s clear that this property is regularly violated by CG; for example the transitions from \( P_{0,6}(\lambda) \) to \( P_{1,8}(\lambda) \) and \( P_{11,27}(\lambda) \) to \( P_{12,29}(\lambda) \) cause \( l_7 = l_8 \) and \( l_{28} = l_{29} \) respectively. Kabaila and Byrne (2001) provide a proof that their method satisfies has strictly increasing endpoints and it’s clear that by construction MS also possesses this property.

**Remark 12:** Although a confidence procedure with the Least Cardinality property may not having strictly increasing endpoints (e.g. CG), it will always have nondecreasing endpoints. This follows from the fact that all confidence procedures with the least cardinality property have acceptance sets with consecutive \( x \)-values. To see this suppose that the lower endpoints of some Least Cardinality procedure are not nondecreasing, and suppose the first time we have \( l_x > l_{x+1} \) is when \( x = x_0 \). Then we have \( l_{x_0} > l_{x_0-1} \) and \( l_{x_0} > l_{x_0+1} \). Thus, for any \( \lambda_0 \in (l_{x_0-1}, l_{x_0}) \cap (l_{x_0+1}, l_{x_0}) \) we have \( x_0 - 1, x_0 + 1 \in A_{\lambda_0} = \{ x : l_x \leq \lambda_0 < u_x \} \), and \( x_0 \notin A_{\lambda_0} \), contradicting that acceptance sets consist of consecutive \( x \)-values. A similar contradiction occurs if we assume that the upper limits are not nondecreasing.

To further compare these three methods and to further investigate fixed and coincidental endpoints see Figure 21a-d and Table 8 (page 46). Figure 21a shows the set of all the minimal cardinality segments. They are color coded by cardinality where different shades of one color represent pieces of equal cardinality. Figure 21b-d each show how the pieces in Figure 21a are used in cpf’s for CG, MS, and KB respectively. Table 8 displays all possible values for endpoints of a strict length optimal confidence procedure for \( x = 0, 1, 2, \ldots, 25 \). For each \( x \) the first row represents the lower endpoint and the second row represents the upper endpoint. If an endpoint is coincidental then two values are given. They represent the smallest and largest possible value for the endpoint. Furthermore, the other \( x \)-value that shares this endpoint is listed in the Coincidental Points column. The last three columns of the table show the corresponding upper and lower limits for CG, MS, and KB respectively.

In Figure 21a-d first notice that in the leftmost portion of the figure for each \( \lambda \) the acceptance curve of minimal cardinality is unique and hence all three methods must use the same curves here; that is, all three methods must use \( P_{0,0}(\lambda) \).
$P_{0.1}(\lambda), P_{0.2}(\lambda), P_{0.3}(\lambda), P_{0.4}(\lambda), P_{0.5}(\lambda),$ and $P_{0.6}(\lambda)$. All other segments of cardinalities one through seven are below .95. Additionally all consecutive transitions between these curves are of Type I which means in order to satisfy the Least Cardinality property the endpoints corresponding to these transitions are fixed.

Table 8 shows that the aforementioned lower limits for $x = 1, 2, 3, 4, 5,$ and 6 are all forced and that all three methods share this fixed value. Also notice the first time the cpf's for the three methods differ is during the Type I transition between segments of cardinality 7 and 8. The location of this transition corresponds to the fixed lower limit for $x = 7$. Both KB and MS transition from $P_{0.6}(\lambda)$ to $P_{0.7}(\lambda)$, but CG transitions to $P_{1,8}(\lambda)$ causing the upper limit for $x = 0$ to be determined by this transition. The upper limit for 0 is also the first coincidental endpoint (coincidental with 8). The smallest value for this coincidental endpoint corresponds to transitioning from $P_{0.6}(\lambda)$ to $P_{1,8}(\lambda)$ as done by CG. The largest value corresponds to transitioning from $P_{0.6}(\lambda)$ to $P_{0.7}(\lambda)$ and using $P_{0.7}(\lambda)$ for as long as possible before finally using $P_{1,8}(\lambda)$ as done by KB. The smallest and largest values for this coincidental endpoint are listed in the table as the range of possible values for the upper limit for $x = 0$ or the lower limit for $x = 8$. Notice that the smallest and largest values match the corresponding limits for CG and KB respectively. In fact, scrolling through the table you'll notice that CG always matches the smallest possible value for coincidental endpoints, KB matches the largest, and MS will have some value in between.
Figure 21a-d: Illustrates the segments that CG, MS and KB have in common for $\lambda \in [0,25]$. 
### Coincidental Endpoints

<table>
<thead>
<tr>
<th>( x )</th>
<th>Endpoints</th>
<th>Coinc. Pts.</th>
<th>CG</th>
<th>MS</th>
<th>KB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.000</td>
<td>--</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>0.051</td>
<td>--</td>
<td>0.051</td>
<td>0.051</td>
<td>0.051</td>
</tr>
<tr>
<td>5.333</td>
<td>6.121</td>
<td>11</td>
<td>5.333</td>
<td>5.756</td>
<td>6.121</td>
</tr>
<tr>
<td>2</td>
<td>0.355</td>
<td>--</td>
<td>0.355</td>
<td>0.355</td>
<td>0.355</td>
</tr>
<tr>
<td>3</td>
<td>0.818</td>
<td>--</td>
<td>0.818</td>
<td>0.818</td>
<td>0.818</td>
</tr>
<tr>
<td>8.102</td>
<td>9.081</td>
<td>15</td>
<td>8.102</td>
<td>8.808</td>
<td>9.081</td>
</tr>
<tr>
<td>4</td>
<td>1.366</td>
<td>--</td>
<td>1.366</td>
<td>1.366</td>
<td>1.366</td>
</tr>
<tr>
<td>5</td>
<td>1.970</td>
<td>--</td>
<td>1.970</td>
<td>1.970</td>
<td>1.970</td>
</tr>
<tr>
<td>11.177</td>
<td>12.226</td>
<td>19</td>
<td>11.177</td>
<td>11.799</td>
<td>12.226</td>
</tr>
<tr>
<td>6</td>
<td>2.613</td>
<td>--</td>
<td>2.613</td>
<td>2.613</td>
<td>2.613</td>
</tr>
<tr>
<td>9</td>
<td>4.460</td>
<td>--</td>
<td>4.460</td>
<td>4.460</td>
<td>4.460</td>
</tr>
<tr>
<td>16.768</td>
<td>17.912</td>
<td>26</td>
<td>16.768</td>
<td>17.298</td>
<td>17.912</td>
</tr>
<tr>
<td>10</td>
<td>5.333</td>
<td>--</td>
<td>5.333</td>
<td>5.333</td>
<td>5.333</td>
</tr>
<tr>
<td>11</td>
<td>5.333</td>
<td>6.121</td>
<td>1</td>
<td>5.333</td>
<td>5.756</td>
</tr>
<tr>
<td>12</td>
<td>6.666</td>
<td>--</td>
<td>6.666</td>
<td>6.666</td>
<td>6.666</td>
</tr>
<tr>
<td>14</td>
<td>8.102</td>
<td>--</td>
<td>8.102</td>
<td>8.102</td>
<td>8.102</td>
</tr>
<tr>
<td>22.945</td>
<td>23.433</td>
<td>33</td>
<td>22.945</td>
<td>23.352</td>
<td>23.433</td>
</tr>
<tr>
<td>15</td>
<td>8.102</td>
<td>9.081</td>
<td>3</td>
<td>8.102</td>
<td>8.808</td>
</tr>
<tr>
<td>25.400</td>
<td>25.950</td>
<td>36</td>
<td>25.400</td>
<td>25.852</td>
<td>25.950</td>
</tr>
<tr>
<td>26.306</td>
<td>27.885</td>
<td>38</td>
<td>26.306</td>
<td>27.324</td>
<td>27.885</td>
</tr>
<tr>
<td>18</td>
<td>11.177</td>
<td>--</td>
<td>11.177</td>
<td>11.799</td>
<td>11.177</td>
</tr>
<tr>
<td>27.735</td>
<td>28.524</td>
<td>39</td>
<td>27.735</td>
<td>28.349</td>
<td>28.524</td>
</tr>
<tr>
<td>19</td>
<td>11.177</td>
<td>12.226</td>
<td>5</td>
<td>11.177</td>
<td>11.799</td>
</tr>
<tr>
<td>28.966</td>
<td>30.479</td>
<td>41</td>
<td>28.966</td>
<td>29.821</td>
<td>30.479</td>
</tr>
<tr>
<td>20</td>
<td>12.817</td>
<td>--</td>
<td>12.817</td>
<td>12.817</td>
<td>12.817</td>
</tr>
<tr>
<td>30.017</td>
<td>31.143</td>
<td>42</td>
<td>30.017</td>
<td>30.843</td>
<td>31.143</td>
</tr>
<tr>
<td>31.675</td>
<td>33.103</td>
<td>44</td>
<td>31.675</td>
<td>32.315</td>
<td>33.103</td>
</tr>
<tr>
<td>32.277</td>
<td>33.798</td>
<td>45</td>
<td>32.277</td>
<td>33.336</td>
<td>33.798</td>
</tr>
<tr>
<td>34.048</td>
<td>34.403</td>
<td>46</td>
<td>34.048</td>
<td>34.356</td>
<td>34.403</td>
</tr>
<tr>
<td>34.665</td>
<td>36.477</td>
<td>48</td>
<td>34.665</td>
<td>35.828</td>
<td>36.477</td>
</tr>
<tr>
<td>36.030</td>
<td>37.139</td>
<td>49</td>
<td>36.030</td>
<td>36.846</td>
<td>37.139</td>
</tr>
</tbody>
</table>

**Table 8:** Coincidental endpoint table. For each \( x \) value the first row is the lower limit and the second row is the upper limit. If an endpoint is coincidental then the smallest and largest possible values are given.
4.5 The Least Cardinality Percentage Method (LC-p)

With Table 8 one has the potential to create a unique confidence procedure that satisfies the Least Cardinality property by simply choosing forced endpoints to agree with the fixed values given in the table and choosing coincidental endpoints to be any value within the allowable range. However, since it's possible for the allowable ranges of two consecutive coincidental endpoints to overlap, when choosing values for two consecutive lower coincidental endpoints $l_x$ and $l_{x+1}$ one needs make the restriction that $l_x \leq l_{x+1}$ so that the resulting confidence procedure has nondecreasing endpoints. It would then follow that all the upper endpoints would also be nondecreasing. Using Table 8 to create a minimal cardinality procedure gives motivation for a method that chooses coincidental endpoints in the following systematic way. First pick a percentage $p$ such that $0 \leq p \leq 100$. Now suppose for a particular coincidental endpoint the allowable range is the interval $[a, b]$. Then for this method make the coincidental endpoint equal to $a + \frac{p}{100}(b - a)$; that is, choose the endpoint as the $p$th percentile of the interval $[a, b]$.

Since we know that for any coincidental endpoint CG and KB choose their endpoints to be the smallest and largest values of $[a, b]$, respectively, we can also define the method as follows. For any $x$ if $(l_{CG}, u_{CG})$ and $(l_{KB}, u_{KB})$ are the corresponding confidence intervals for CG and KB respectively then define

$$\begin{align*}
l &= l_{CG} + \frac{p}{100}(l_{KB} - l_{CG}) \\
u &= u_{CG} + \frac{p}{100}(u_{KB} - u_{CG})
\end{align*}$$

to be the upper and lower confidence limits for this method for that $x$. The method with limits defined above will be called LC-p where LC-p is short for Least Cardinality percentage method since the method satisfies the Least Cardinality property by construction and since its coincidental endpoints are chosen at a selected percentage of the distance between the smallest and largest allowable values. Notice if a lower endpoint was a forced endpoint then $l_{KB} = l_{CG}$ which would imply that $l = l_{CG}$ and thus LC-p would have its forced endpoints equal to the same fixed value as all least cardinality procedures. Also, if $p = 0$ or $100$ then LC-p would be identical to CG or KB respectively. Since all coincidental endpoints are being increased a fixed percentage of the way past CG towards MS and since CG has nondecreasing endpoints and MS has strictly increasing endpoints it follows that LC-p will have nondecreasing endpoints.

To illustrate how LC-p works let's choose $p = 15$ and look at the interval for $x = 7$. Note in practice, the same $p$ is chosen for all $x$; that is, first one would choose $p$, which
establishes a particular confidence procedure. Since the lower limit for this interval is forced we let \( l = 3.29 \), the fixed value of this endpoint. However, since the upper limit for this interval is a coincidental endpoint we pick the upper limit, \( u \), to be 15\% of the distance between \( u_{KB} \) and \( u_{CG} \) past \( u_{CG} \); that is,

\[
u = u_{CG} + \frac{p}{100} (u_{KB} - u_{CG}) = 13.77 + .15(14.5 - 13.77) = 13.88
\]

where all endpoints were rounded to two decimal places.

Figure 22 compares the confidence interval for \( x = 7 \) of LC-15 with the intervals for CG, MS and KB. Since the lower limit for \( x = 7 \) has a fixed value of 3.29, all four methods agree on this value. The upper limit, on the other hand, is coincidental and so CG and KB take on the smallest and largest values respectively. MS chooses its upper limit so that coverage is maximized and LC-15 chooses the endpoint so that it is 15\% of the way between the two points \( u_{CG} = 13.77 \) and \( u_{KB} = 14.5 \).

Figure 22: Illustrates how LC-p chooses coincidental endpoints.

Figure 23 shows the 95\% cpf for LC-15. The cpf for LC-15 looks very similar to the cpf’s of MS and KB since it uses the identical pieces as both methods but, with earlier Type II transitions. For this same reason LC-15 will have strictly increasing endpoints. The choice of \( p = 15 \) was "almost" arbitrary. It was chosen small enough so that it's interval length falls close to that of CG, but large enough so that it comfortably has strictly increasing endpoints. However, the user could use any method for selecting a value for the percentage \( p \).
4.6 Further Discussion of Length

We have talked a lot about the Least Cardinality property and Kabaila and Byrne's Inability to be Shortened property but there are several more ways to evaluate the performance of a confidence procedure in regards to confidence interval length. One way is through Casella and Roberts' asymptotic length criterion. However before discussing their length criterion the following two results will useful.

**Lemma 9:** If \( C \) has the Least Cardinality property then as \( \lambda \) increases the cardinality of the acceptance sets for \( C \) never decrease.

**Theorem 10:** If \( C = \{(l_x, u_x), x = 0, 1, 2, \ldots\} \) has the Least Cardinality property then every \( u_x \) is coincidental with some \( l_y \) for \( y > x \).

Recall Casella and Robert compare two confidence procedures asymptotically through the use of partial sums. A strict confidence procedure \( C' = \{(l'_x, u'_x), x = 0, 1, 2, \ldots\} \) dominates another strict confidence procedure \( C = \{(l_x, u_x), x = 0, 1, 2, \ldots\} \) if for sufficiently large \( K \),

\[
\sum_{x=0}^{K}(u'_x - l'_x) < \sum_{x=0}^{K}(u_x - l_x).
\]

If a tie occurs for all \( K \) past some cut off \( K_0 \) then coverage is used as a tie breaker.

Suppose we try to refine CG as suggested by Casella and Robert (see their Proposition 2.2) by increasing coverage without increasing "net" interval length; that is, without changing the value of \( \sum_{x=0}^{K}(u_x - l_x) \) for sufficiently large \( K \). To do this we can increase
all of the CG coincidental endpoints one by one to the place which yields the most gain in coverage. So we first increase \( u_0 = l_8 \) from 3.285 to the 3.764, the transition value used for MS. This creates an increase in coverage for all \( \lambda \in [3.285, 3.764] \). Moreover, this move increases the interval length of \( x = 0 \) by \( 3.764 - 3.285 = .479 \) and decreases the interval for \( x = 8 \) by the same amount. So after refining the first coincidental endpoint we have increased the sum of interval lengths, \( \sum_{x=0}^{K}(u_x - l_x) \) for \( K < 8 \) but, attained equality in the sum for all \( K \geq 8 \). So the new refined confidence procedure dominates CG since we improved overall coverage without increasing "net" length.

Suppose we do this again for the next coincidental endpoint, increasing \( u_1 = l_{11} = 5.323 \) to the MS value, 5.756. These two moves together again increase coverage without increasing "net" length since they lead to an increase in \( \sum_{x=0}^{K}(u_x - l_x) \) for \( K < 11 \) but to equality in the sum for \( K \geq 11 \). Now since by Theorem 10 for all \( x, u_x \) is coincidental with \( l_y \) for some \( y > x \) we can continue this refinement indefinitely. In the limit, this yields MS. But as a result of this refinement, \( \sum_{x=0}^{K}(u_x - l_x) \) will be larger than CG for all \( K \). Thus, comparing two strict confidence procedure's "net" interval length using Casella and Robert's criterion creates a paradox -- an infinite sequence of improvements to CG leads to MS, which is inferior to CG with respect to asymptotic \( (K \to \infty) \) confidence interval length. In fact, we have the following result:

**Theorem 11:** Let \( C' = \{(l'_x, u'_x), x = 0, 1, 2, \ldots \} \) be the confidence procedure produced by CG and let \( C = \{(l_x, u_x), x = 0, 1, 2, \ldots \} \) be any other confidence procedure which has the Least Cardinality property. Then

\[
\sum_{x=0}^{K}(u'_x - l'_x) \leq \sum_{x=0}^{K}(u_x - l_x) \ \forall k.
\]

**Remark 13:** A way to get around this paradox is to instead use a refinement only for the first, say, 50 coincidental endpoints. This would create a new confidence procedure which dominates CG since it would increase coverage without changing overall "net" length. Or we could go even further and refine the coincidental endpoints for the first 500 or 1000 \( x \)'s. Again the new hybrid confidence procedure, part MS and part CG would dominate the original CG confidence procedure. In practice we could refine all the coincidental endpoints of CG for \( x = 0, 1, 2, \ldots K_0 \), where \( K_0 \) represents the largest possible value that we might reasonably expect to observe.

Instead of looking at the asymptotic behavior of confidence interval widths we can also compare strict confidence procedures interval by interval for small \( x \) since the confidence intervals for large \( x \) of most strict confidence procedures yield similar results and even a normal approximation can be satisfactory. Figure 24 below is a graph which displays the rankings of the confidence intervals of each method for each \( x \) from shortest to longest. For instance for \( x = 0 \) the methods are ranked CG, LC-15, B, G, MS, and KB, in order of their interval length for \( x = 0 \).
Figure 24: Methods ranked by confidence interval length for each $x$ from 0 to 50 for KB, MS, G, B, LC-15, and CG.

From this figure it is clear that CG has the shortest intervals of all strict confidence procedures being compared here. In fact, it is only the intervals corresponding to $x = 41$, 44, and 48 for which CG does not have the shortest interval. LC-15 is second best, followed by B and MS.

We can also use a running average formula,

$$\frac{1}{K} \sum_{i=0}^{K} (u_i - l_i)$$

to compare the "average" interval widths of different confidence procedures for small $K$. Additionally, we can contrast their expected widths for small values of $\lambda$. The expected width of a confidence procedure is

$$E_{\lambda}(W) = \sum_{x=0}^{\infty} (u_x - l_x) \frac{e^{-\lambda x}}{x!};$$

see Barker (2002) or Swift (2009). Note that, as opposed to average width, expected width is a function of the parameter $\lambda$. Since the tails of the sum in the above equation go to zero rapidly it is sufficient to calculate

$$E_{\lambda}(W) = \sum_{x=0}^{M_{\lambda}} (u_x - l_x) \frac{e^{-\lambda x}}{x!},$$

for sufficiently large $M_{\lambda}$. For $\lambda \in [0,50]$, $M_{\lambda} = \lceil \lambda \rceil + 30$ is adequate, where $\lceil \cdot \rceil$ is the ceiling function which outputs the smallest integer greater than or equal to its' input.

Figure 25a is the graph of the average confidence interval widths for $K = 0,1,2,\ldots,50$ for KB, MS, G, B, LC-15 and CG relative to G. For each method average widths were calculated for each $K$ and then divided by the average width for G for that fixed $K$. Values below 1.00 imply that the average width is less than that of G. Figure 25b shows expected widths for $\lambda \in [0,50]$ relative to G.
These figures show that both LC-15 and MS are competitive with the best known existing methods on two fundamental measures of confidence interval width. Of the six methods in Figure 25a-b CG has both the smallest average confidence interval widths and narrowest expected widths. LC-15 is second, followed by B and MS. For KB average and expected widths are worse than even G for small values of $K$ and $\lambda$ respectively.

As discussed CG will always have smaller partial sums, $\sum_{x=0}^{\infty} (u_x - l_x)$, than MS because an infinite choice of smaller coincidental endpoints never allows MS to "catch up" to CG with respect to partial sums of interval widths. This can also be seen from Figure 24 since CGs has shorter intervals for nearly every $x$ from 0 to 50. In fact, by Theorem 10 we know this is true for all methods possessing the Least Cardinality property. As a result, CG will always have the smallest expected widths and average interval widths of any method satisfying the Least Cardinality property. This reveals a significant limitation of the Least Cardinality property. Using the Least Cardinality property as the only length criterion would imply that MS is superior to the Crow Gardner method (and all other methods satisfying the Least Cardinality property) since it has maximal coverage and strictly increasing endpoints. But, further investigation of length shows that when comparing the procedures interval by interval, through partial sums of interval lengths, by average length, or by expected length, CG is better on length. In fact, MS is only superior to CG in coverage and in satisfying the strictly increasing endpoints property, two criteria which have nothing to do with length. As a result, the use of only the Least Cardinality property and/or the Inability to be Shortened property is insufficient in the
determining which methods are best on length and so we need to also compare
confidence procedures interval by interval for small x and look at their average and
expected interval widths.

Moreover, we have seen that Casella Roberts asymptotic criterion for assessing the length
performance of Poisson confidence procedure is problematic in that not only does it
create a paradox, but it does not address how two confidence procedure compare over
confidence intervals for small x.

Hence, we propose an alternative criterion by which to compare the length performance
of Poisson confidence procedures which satisfy the Least Cardinality property. We shall
say that \( C' = \{(l'_x, u'_x), x = 0, 1, 2, \ldots \} \) is superior to \( C = \{(l_x, u_x), x = 0, 1, 2, \ldots \} \) if
\[
\sum_{x=0}^{K}(u'_x - l'_x) \leq \sum_{x=0}^{K}(u_x - l_x)
\]
for all \( K \), with \( \sum_{x=0}^{K}(u'_x - l'_x) < \sum_{x=0}^{K}(u_x - l_x) \) for some \( K \). Using this measure, Theorem 11 implies CG is best on length, when being
compared to any other method with the Least Cardinality property.

All in all, if we ignore maximal coverage and the monotonicity of interval endpoints and
contrast strict confidence procedures solely on the basis of length then CG performs best.
But, if we then require that a confidence procedure has strictly increasing endpoints then
LC-\( p \), for select choices of \( p \), performs the best of the procedures we have considered so
far.

4.7 Coverage

When comparing strict confidence procedures, length is probably the most important
factor to examine; however, coverage also plays a vital role. Figure 26 is a graph of the
average coverage as a function of \( \lambda \) for G, B, MS, KB, LC-15, and CG, the six strict
methods that we have discussed. For each \( \lambda \) average coverage is calculated by taking the
total coverage of the respective cpf's for the six methods on the interval \( [0, \lambda] \) and
dividing by \( \lambda \). In other words for each \( \lambda \)

\[
\text{Average Coverage} = \frac{1}{\lambda} \int_{0}^{\lambda} \text{cov}(\mu) \, d\mu .
\]
Overall G is best on coverage, followed by B, MS, KB, LC-15, and CG. Since the cpf's for B, MS, KB, LC-15, and CG all have identical cpf's between 0 and about 4 they all have equal coverage here. Hence, the graphs of their average coverage coincide in this region. Figure 27 shows a graph of average interval width for \( x = 0, 1, 2, \ldots, 50 \) versus average coverage on [0,50]. For each method average interval width and average coverage were calculated with the formulas,

\[
\frac{1}{50} \sum_{i=0}^{50} (u_i - l_i) \quad \& \quad \frac{1}{50} \int_{0}^{50} \text{cov}(\mu) \, d\mu
\]

respectively. The ideal method would be located in the upper left hand corner. CG is the extreme case of small interval width and low coverage; whereas, G is the extreme case of high coverage with wide intervals. On the hand, B and MS provide a good balance of both factors.
To see how different choices of $p$ for LC-$p$ affect average interval width and average coverage see Figure 28 below, which shows the average width and average coverage for LC-$p$ for all possible integer values of $p$. If $p = 0$ or $p = 100$ then LC-$p$ has coverage and interval widths equal to those of CG and KB respectively. If $p$ is between 60 and 65 then LC-$p$ has average length and average coverage very close to that of MS.

Now that we have added a few more methods to our arsenal it will beneficial to revisit a previous example.

**Example 1: Lung Cancer (Revisited)**

Recall from Example 1 located in Section 2.2 that a researcher studying a large group of coal miners for several years found that 18 died of lung cancer. For the general population, 11 deaths is expected for a group this size during this time period. Using
For the number of observed deaths we can calculate and compare the 95% confidence intervals determined from all the methods we have considered thus far. Figure 29 below shows the resulting intervals graphically.

Since the expected number of deaths, 11, falls in the confidence intervals for W and G, any researcher using either of these two methods would have failed to conclude that the excess deaths among the coal miners was due to anything more than just chance variation. However a researcher using any one of S, B, KB, CG, LC-15, or MS would have come to the opposite conclusion since 11 does not fall in the intervals for any of these methods. The researcher would have concluded that the deaths from lung cancer for the coal miners is "significantly" higher than what should be expected from a group the same size from the general population. This demonstrates how different confidence procedures can produce different results and so it is important to use an efficient method with both narrow intervals and valid coverage.

4.8 The Modified Crow and Gardner (MCG) Method

Although Crow and Gardner’s method performs best on measures of length, it has one major flaw—it fails to have strictly increasing endpoints. This is a serious drawback, as it provides seemingly contradictory results in practice when an increase or decrease in the observed count does not always result in a change in both confidence limits. However, we can adopt an idea similar to the one used for LC-\(p\) to easily modify CG and create a new confidence procedure with strictly increasing endpoints while still largely preserving
the length advantages that CG possesses. Below is a rough sketch of how one might produce such a procedure.

Choosing coincidental endpoints to be the smallest allowable value, as done by CG, sometimes causes $l_x = l_{x+1} = \cdots = l_{x+k}$. To fix these sets of tied endpoints we can create a new procedure as follows. Starting from $x = 0$ scan the lower endpoints of CG and look for sets of tied lower endpoints. Each set you find can be treated in a similar fashion. The first member in the set will always be a fixed endpoint and so it cannot be increased, but all other members are coincidental and need to be increased so that the resulting procedure has strictly increasing endpoints. Starting from the last member of the set "adequately" increase its value making sure the endpoint not only stays in the coincidental endpoint region of allowable values but that it also does not get increased past the following lower endpoint of the procedure. Next working backwards increase all the endpoints in the set, except the first, in a similar manner.

Note that each time we modify a lower coincidental endpoint we are also changing the corresponding upper coincidental endpoint to the same value. When increasing each lower coincidental endpoint one can increase the endpoint a certain percentage of way past CG towards the maximum possible value for that endpoint, or increase the endpoint by some other preferred method. The values allowable for each coincidental endpoint are those between the original (CG) value for that endpoint and the minimum of the corresponding lower limit for KB and the value of the following lower endpoint of the procedure. We will call the resulting confidence procedure the Modified Crow and Gardner method (MCG). Note that MCG is really a class procedures where each procedure in the class differs depending on method used to increase CG's tied endpoints. MCG will be almost as short as CG, but never equal since we increase a large subset of coincidental endpoints. However, unlike CG it will now have strictly increasing endpoints. Regardless of how much you increase the tied coincidental endpoints, MCG will have good length performance; however, keeping these coincidental endpoint values relatively close to CG produces a shorter procedure.
Chapter 5: Conclusion

5.1 Opportunities for Further Research

Research on this topic is far from over. For example, the approach of MS could be extended to the problem of estimating the difference between two Poisson parameters based on independent data from different populations. It can also be extended to other discrete distributions. Moreover, further investigation can be done at different confidence levels such as 90 and 99 percent. Also, as discussed in Remark 10 MS could be modified so that, rather than choosing pieces for the cpf that make the coverage strict, pieces can be chosen so that the method has coverage which on average falls at or near the chosen confidence level. This modification could also be made for other methods as well. One could also work on choosing the "best" value of $p$ for LC-$p$ or determine the "best" way for MCG to increase CG's tied endpoints where the term "best" would depend on a carefully chosen criteria.

5.2 Summary

Unlike published methods, the Modified Sterne's method (MS) creates a confidence procedure with any desired confidence level by first creating a specialized coverage probability function; this in turn determines confidence intervals for $\lambda$ for all possible number of events observed. This approach to confidence intervals gives a unique outlook on the relationship between a confidence procedure and its coverage probability function. We constructed the "ideal" cpf by first grouping all Poisson probability functions for a set of consecutive values by the number of $x$-values that were involved in each function. This is because, of the available curves, the coverage probability function for MS needed to use those that came from the set with least cardinality in order to insure that the Least Cardinality property was satisfied. Next, for each value of $\lambda$, whenever choosing between curves having equal cardinality, we used the highest of all the curves in that group. As was discussed in Section 4.2, this made it possible to maintain the highest possible coverage during Type II transitions while still managing to satisfy the Least Cardinality property. This choice of curves forms a graph comprised of an infinite number of necklaces. We disregarded all sections of these necklaces below the .95 line since any use of these sections for the cpf would create a confidence procedure with coverage below 95%. Consequently, for each $\lambda$ we defined our cpf to take on values from the lowest available necklace while still staying above 95%. And lastly, in the rare occasions that choosing the cpf in this way caused gaps we used the technique demonstrated in
Solution 2 of Section 3.6 to eliminate the gaps. For illustrative purposes the derivation of MS was long and comprehensive; however, we introduced a short and simple 3-step algorithm for calculating all interval endpoints.

As a result, MS creates a gapless confidence procedure that possesses the Least Cardinality property, maintains the nominal confidence level, and has interval endpoints that can be calculated rather easily. In fact, it belongs to the collection strict confidence procedures with the Least Cardinality property, but has the highest coverage among the entire collection. Both the Kabaila and Byrne (KB), and Crow and Gardner (CG) methods also belong to this collection, but they are inferior to MS in this regard since they often have lower coverage due to where Type II transitions occur for two methods. Analyzing MS and comparing it to the other known methods that satisfy the Least cardinality property gave motivation to create LC-\(p\), another minimal cardinality procedure that could create a desired balance of high coverage, close to that of MS, and short interval lengths, close or equal to those of the Crow and Gardner method, by making different choices for the value of \(p\). Choosing \(p = 15\), giving interval lengths close to those of the Crow and Gardner method but comfortably satisfying the strictly increasing endpoints property produced a particular method that is quite competitive with the best known existing methods.

Although Crow and Gardner’s method performs best on measures of length, its failure to have strictly increasing endpoints is a serious drawback. In addition, the Crow and Gardner method has much lower coverage than the alternative methods considered herein. The Modified Crow and Gardner method (MCG), on the other hand, can perform almost as well on measures of length and by design has strictly increasing endpoints, thus making it an excellent alternative to the Crow and Gardner method. To summarize, LC-\(p\) (for select choices of \(p\)), MCG, and MS perform well overall but MS has particular strength in coverage while LC-\(p\) (for select choices of \(p\)) and MCG excel on measures of length.

We have seen that no procedure is superior to all of the others with respect to all of the desirable properties that have been presented. The ultimate determination of what is the “best” method to use in a practical situation must therefore be left to the user. However, based on their high performance qualities in coverage and length and their achievement of monotonicity in their intervals endpoints as established in this thesis, the proposed procedures: MS, LC-\(p\), and MCG deserve serious consideration.
References


### Appendix A: Tables

**Table 9**: 95% confidence intervals for G, S, & W for \( x = 0, 1, 2, \ldots, 50 \).

<table>
<thead>
<tr>
<th>( x )</th>
<th><strong>LL</strong></th>
<th><strong>UL</strong></th>
<th><strong>LL</strong></th>
<th><strong>UL</strong></th>
<th><strong>LL</strong></th>
<th><strong>UL</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.000</td>
<td>3.689</td>
<td>0.000</td>
<td>3.341</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>0.025</td>
<td>5.572</td>
<td>0.177</td>
<td>5.666</td>
<td>0.000</td>
<td>2.960</td>
</tr>
<tr>
<td>2</td>
<td>0.242</td>
<td>7.225</td>
<td>0.548</td>
<td>7.293</td>
<td>0.000</td>
<td>4.772</td>
</tr>
<tr>
<td>3</td>
<td>0.619</td>
<td>8.767</td>
<td>1.020</td>
<td>8.821</td>
<td>0.000</td>
<td>6.395</td>
</tr>
<tr>
<td>4</td>
<td>1.030</td>
<td>10.242</td>
<td>1.556</td>
<td>10.286</td>
<td>0.080</td>
<td>7.920</td>
</tr>
<tr>
<td>5</td>
<td>1.623</td>
<td>11.668</td>
<td>2.136</td>
<td>11.706</td>
<td>0.617</td>
<td>9.383</td>
</tr>
<tr>
<td>8</td>
<td>3.454</td>
<td>15.763</td>
<td>4.054</td>
<td>15.788</td>
<td>2.456</td>
<td>15.544</td>
</tr>
<tr>
<td>9</td>
<td>4.115</td>
<td>17.085</td>
<td>4.735</td>
<td>17.106</td>
<td>3.220</td>
<td>18.880</td>
</tr>
<tr>
<td>10</td>
<td>4.795</td>
<td>18.390</td>
<td>5.432</td>
<td>18.409</td>
<td>3.802</td>
<td>18.996</td>
</tr>
<tr>
<td>12</td>
<td>6.201</td>
<td>20.962</td>
<td>6.865</td>
<td>20.977</td>
<td>5.210</td>
<td>18.790</td>
</tr>
<tr>
<td>13</td>
<td>6.922</td>
<td>22.230</td>
<td>7.598</td>
<td>22.244</td>
<td>5.933</td>
<td>20.067</td>
</tr>
<tr>
<td>17</td>
<td>9.903</td>
<td>27.219</td>
<td>10.614</td>
<td>27.227</td>
<td>8.919</td>
<td>25.081</td>
</tr>
<tr>
<td>21</td>
<td>12.999</td>
<td>32.101</td>
<td>13.736</td>
<td>32.105</td>
<td>12.018</td>
<td>29.982</td>
</tr>
<tr>
<td>22</td>
<td>13.787</td>
<td>33.308</td>
<td>14.529</td>
<td>33.312</td>
<td>12.807</td>
<td>31.193</td>
</tr>
<tr>
<td>23</td>
<td>14.580</td>
<td>34.511</td>
<td>15.327</td>
<td>34.515</td>
<td>13.600</td>
<td>32.400</td>
</tr>
<tr>
<td>25</td>
<td>16.179</td>
<td>36.905</td>
<td>16.934</td>
<td>36.907</td>
<td>15.200</td>
<td>34.800</td>
</tr>
<tr>
<td>26</td>
<td>16.964</td>
<td>38.096</td>
<td>17.744</td>
<td>38.098</td>
<td>16.006</td>
<td>35.994</td>
</tr>
<tr>
<td>28</td>
<td>18.606</td>
<td>40.468</td>
<td>19.373</td>
<td>40.468</td>
<td>17.629</td>
<td>38.371</td>
</tr>
<tr>
<td>30</td>
<td>20.241</td>
<td>42.827</td>
<td>21.015</td>
<td>42.826</td>
<td>19.265</td>
<td>40.735</td>
</tr>
<tr>
<td>31</td>
<td>21.063</td>
<td>44.002</td>
<td>21.840</td>
<td>44.001</td>
<td>20.087</td>
<td>41.913</td>
</tr>
<tr>
<td>32</td>
<td>21.888</td>
<td>45.174</td>
<td>22.668</td>
<td>45.173</td>
<td>20.913</td>
<td>43.087</td>
</tr>
<tr>
<td>33</td>
<td>22.716</td>
<td>46.344</td>
<td>23.499</td>
<td>46.343</td>
<td>21.741</td>
<td>44.259</td>
</tr>
<tr>
<td>34</td>
<td>23.546</td>
<td>47.512</td>
<td>24.332</td>
<td>47.509</td>
<td>22.572</td>
<td>45.428</td>
</tr>
<tr>
<td>35</td>
<td>24.379</td>
<td>48.677</td>
<td>25.167</td>
<td>48.674</td>
<td>23.405</td>
<td>46.595</td>
</tr>
<tr>
<td>36</td>
<td>25.214</td>
<td>49.839</td>
<td>26.005</td>
<td>49.835</td>
<td>24.240</td>
<td>47.760</td>
</tr>
<tr>
<td>37</td>
<td>26.051</td>
<td>51.000</td>
<td>26.845</td>
<td>50.996</td>
<td>25.078</td>
<td>48.922</td>
</tr>
<tr>
<td>38</td>
<td>26.881</td>
<td>52.158</td>
<td>27.687</td>
<td>52.154</td>
<td>25.918</td>
<td>50.082</td>
</tr>
<tr>
<td>39</td>
<td>27.733</td>
<td>53.314</td>
<td>28.531</td>
<td>53.310</td>
<td>26.760</td>
<td>51.240</td>
</tr>
<tr>
<td>40</td>
<td>28.577</td>
<td>54.469</td>
<td>29.377</td>
<td>54.465</td>
<td>27.604</td>
<td>52.396</td>
</tr>
<tr>
<td>41</td>
<td>29.422</td>
<td>55.621</td>
<td>30.225</td>
<td>55.617</td>
<td>28.450</td>
<td>53.560</td>
</tr>
<tr>
<td>42</td>
<td>30.270</td>
<td>56.772</td>
<td>31.074</td>
<td>56.767</td>
<td>29.298</td>
<td>54.720</td>
</tr>
<tr>
<td>43</td>
<td>31.119</td>
<td>57.921</td>
<td>31.926</td>
<td>57.916</td>
<td>30.148</td>
<td>55.852</td>
</tr>
<tr>
<td>44</td>
<td>31.970</td>
<td>59.068</td>
<td>32.779</td>
<td>59.063</td>
<td>30.999</td>
<td>57.001</td>
</tr>
<tr>
<td>45</td>
<td>32.823</td>
<td>60.214</td>
<td>33.633</td>
<td>60.208</td>
<td>31.852</td>
<td>58.148</td>
</tr>
<tr>
<td>46</td>
<td>33.678</td>
<td>61.358</td>
<td>34.490</td>
<td>61.352</td>
<td>32.707</td>
<td>59.293</td>
</tr>
<tr>
<td>47</td>
<td>34.534</td>
<td>62.500</td>
<td>35.347</td>
<td>62.494</td>
<td>33.563</td>
<td>60.437</td>
</tr>
<tr>
<td>48</td>
<td>35.381</td>
<td>63.641</td>
<td>36.207</td>
<td>63.635</td>
<td>34.421</td>
<td>61.579</td>
</tr>
<tr>
<td>49</td>
<td>36.250</td>
<td>64.781</td>
<td>37.067</td>
<td>64.774</td>
<td>35.280</td>
<td>62.720</td>
</tr>
<tr>
<td>50</td>
<td>37.111</td>
<td>65.919</td>
<td>37.929</td>
<td>65.912</td>
<td>36.141</td>
<td>63.859</td>
</tr>
<tr>
<td>x</td>
<td>Crow &amp; Gardner</td>
<td>Modified Sterne's</td>
<td>Kabaila &amp; Byrne</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LL</td>
<td>UL</td>
<td>LL</td>
<td>UL</td>
<td>LL</td>
<td>UL</td>
</tr>
<tr>
<td>0</td>
<td>0.000</td>
<td>3.285</td>
<td>0.000</td>
<td>3.764</td>
<td>0.000</td>
<td>3.981</td>
</tr>
<tr>
<td>1</td>
<td>0.051</td>
<td>5.323</td>
<td>0.051</td>
<td>5.756</td>
<td>0.051</td>
<td>6.121</td>
</tr>
<tr>
<td>2</td>
<td>0.355</td>
<td>6.686</td>
<td>0.355</td>
<td>7.295</td>
<td>0.355</td>
<td>7.579</td>
</tr>
<tr>
<td>3</td>
<td>0.818</td>
<td>8.102</td>
<td>0.818</td>
<td>8.808</td>
<td>0.818</td>
<td>9.081</td>
</tr>
<tr>
<td>5</td>
<td>1.970</td>
<td>11.177</td>
<td>1.970</td>
<td>11.799</td>
<td>1.970</td>
<td>12.226</td>
</tr>
<tr>
<td>6</td>
<td>2.613</td>
<td>12.817</td>
<td>2.613</td>
<td>13.286</td>
<td>2.613</td>
<td>13.856</td>
</tr>
<tr>
<td>9</td>
<td>4.460</td>
<td>16.768</td>
<td>4.460</td>
<td>17.298</td>
<td>4.460</td>
<td>17.912</td>
</tr>
<tr>
<td>10</td>
<td>5.323</td>
<td>17.633</td>
<td>5.323</td>
<td>18.339</td>
<td>5.323</td>
<td>18.565</td>
</tr>
<tr>
<td>14</td>
<td>8.102</td>
<td>22.945</td>
<td>8.102</td>
<td>23.352</td>
<td>8.102</td>
<td>23.433</td>
</tr>
<tr>
<td>15</td>
<td>8.102</td>
<td>23.762</td>
<td>8.808</td>
<td>24.825</td>
<td>9.081</td>
<td>25.330</td>
</tr>
<tr>
<td>18</td>
<td>11.177</td>
<td>27.735</td>
<td>11.177</td>
<td>28.349</td>
<td>11.177</td>
<td>28.524</td>
</tr>
<tr>
<td>20</td>
<td>12.817</td>
<td>30.017</td>
<td>12.817</td>
<td>30.843</td>
<td>12.817</td>
<td>31.413</td>
</tr>
<tr>
<td>21</td>
<td>12.817</td>
<td>31.675</td>
<td>13.286</td>
<td>32.135</td>
<td>13.856</td>
<td>33.103</td>
</tr>
<tr>
<td>23</td>
<td>14.921</td>
<td>34.048</td>
<td>14.921</td>
<td>34.356</td>
<td>14.921</td>
<td>34.403</td>
</tr>
<tr>
<td>24</td>
<td>14.921</td>
<td>34.665</td>
<td>15.820</td>
<td>35.828</td>
<td>16.197</td>
<td>36.477</td>
</tr>
<tr>
<td>26</td>
<td>16.768</td>
<td>37.667</td>
<td>17.298</td>
<td>38.318</td>
<td>17.912</td>
<td>39.174</td>
</tr>
<tr>
<td>28</td>
<td>19.050</td>
<td>39.763</td>
<td>19.051</td>
<td>40.353</td>
<td>19.051</td>
<td>40.514</td>
</tr>
<tr>
<td>29</td>
<td>19.050</td>
<td>40.944</td>
<td>19.814</td>
<td>41.824</td>
<td>20.345</td>
<td>42.625</td>
</tr>
<tr>
<td>30</td>
<td>20.335</td>
<td>41.753</td>
<td>20.848</td>
<td>42.840</td>
<td>20.973</td>
<td>43.322</td>
</tr>
<tr>
<td>31</td>
<td>21.364</td>
<td>43.453</td>
<td>21.364</td>
<td>43.856</td>
<td>21.364</td>
<td>43.934</td>
</tr>
<tr>
<td>32</td>
<td>21.364</td>
<td>44.259</td>
<td>22.322</td>
<td>45.328</td>
<td>22.816</td>
<td>46.112</td>
</tr>
<tr>
<td>33</td>
<td>22.945</td>
<td>45.260</td>
<td>23.352</td>
<td>46.343</td>
<td>23.433</td>
<td>46.805</td>
</tr>
<tr>
<td>34</td>
<td>23.762</td>
<td>47.025</td>
<td>23.762</td>
<td>47.357</td>
<td>23.762</td>
<td>47.411</td>
</tr>
<tr>
<td>35</td>
<td>23.762</td>
<td>47.889</td>
<td>24.825</td>
<td>48.829</td>
<td>25.330</td>
<td>49.635</td>
</tr>
<tr>
<td>36</td>
<td>25.400</td>
<td>48.744</td>
<td>25.852</td>
<td>49.843</td>
<td>25.990</td>
<td>50.332</td>
</tr>
<tr>
<td>38</td>
<td>26.306</td>
<td>51.293</td>
<td>27.324</td>
<td>52.329</td>
<td>27.885</td>
<td>53.192</td>
</tr>
<tr>
<td>39</td>
<td>27.735</td>
<td>52.150</td>
<td>28.349</td>
<td>53.341</td>
<td>28.524</td>
<td>53.901</td>
</tr>
<tr>
<td>40</td>
<td>28.966</td>
<td>53.719</td>
<td>28.966</td>
<td>54.354</td>
<td>28.966</td>
<td>54.539</td>
</tr>
<tr>
<td>41</td>
<td>28.966</td>
<td>54.990</td>
<td>29.821</td>
<td>55.827</td>
<td>30.479</td>
<td>56.782</td>
</tr>
<tr>
<td>42</td>
<td>30.017</td>
<td>55.511</td>
<td>30.843</td>
<td>56.839</td>
<td>31.413</td>
<td>57.507</td>
</tr>
<tr>
<td>43</td>
<td>31.675</td>
<td>56.966</td>
<td>31.675</td>
<td>57.850</td>
<td>31.675</td>
<td>58.174</td>
</tr>
<tr>
<td>44</td>
<td>31.675</td>
<td>58.178</td>
<td>32.365</td>
<td>59.323</td>
<td>33.103</td>
<td>60.399</td>
</tr>
<tr>
<td>45</td>
<td>32.277</td>
<td>58.838</td>
<td>33.336</td>
<td>60.335</td>
<td>33.798</td>
<td>61.412</td>
</tr>
<tr>
<td>46</td>
<td>34.048</td>
<td>60.240</td>
<td>34.356</td>
<td>61.346</td>
<td>34.403</td>
<td>61.940</td>
</tr>
<tr>
<td>47</td>
<td>34.665</td>
<td>61.900</td>
<td>34.665</td>
<td>62.367</td>
<td>34.665</td>
<td>62.456</td>
</tr>
<tr>
<td>48</td>
<td>34.665</td>
<td>62.814</td>
<td>35.828</td>
<td>63.830</td>
<td>36.477</td>
<td>64.801</td>
</tr>
<tr>
<td>49</td>
<td>36.030</td>
<td>63.488</td>
<td>36.846</td>
<td>64.841</td>
<td>37.139</td>
<td>65.528</td>
</tr>
<tr>
<td>50</td>
<td>37.667</td>
<td>64.947</td>
<td>37.667</td>
<td>65.851</td>
<td>37.667</td>
<td>66.200</td>
</tr>
</tbody>
</table>
Table 11: 95% confidence intervals for ST, B and LC-15 for \( x = 0, 1, 2, \ldots, 50 \). For ST if a confidence interval has gap (i.e. is a union of two intervals) then the upper and lower limits of the first interval are listed in the "UL" column and the upper and lower limits of the second interval are listed in the "UL" column.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( LL )</th>
<th>( UL )</th>
<th>( LL )</th>
<th>( UL )</th>
<th>( LL )</th>
<th>( UL )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.000</td>
<td>3.764</td>
<td>0.000</td>
<td>3.560</td>
<td>0.000</td>
<td>3.390</td>
</tr>
<tr>
<td>1</td>
<td>0.051</td>
<td>5.756</td>
<td>0.051</td>
<td>5.528</td>
<td>0.051</td>
<td>5.443</td>
</tr>
<tr>
<td>2</td>
<td>0.355</td>
<td>7.295</td>
<td>0.355</td>
<td>7.054</td>
<td>0.355</td>
<td>6.820</td>
</tr>
<tr>
<td>3</td>
<td>0.818</td>
<td>8.808</td>
<td>0.818</td>
<td>8.560</td>
<td>0.818</td>
<td>8.249</td>
</tr>
<tr>
<td>5</td>
<td>1.970</td>
<td>11.799</td>
<td>1.970</td>
<td>11.543</td>
<td>1.970</td>
<td>11.334</td>
</tr>
<tr>
<td>6</td>
<td>2.613</td>
<td>13.286</td>
<td>2.613</td>
<td>13.027</td>
<td>2.613</td>
<td>12.973</td>
</tr>
<tr>
<td>9</td>
<td>4.460</td>
<td>17.296</td>
<td>4.460</td>
<td>17.030</td>
<td>4.460</td>
<td>16.940</td>
</tr>
<tr>
<td>10</td>
<td>5.323</td>
<td>18.339</td>
<td>5.123</td>
<td>18.068</td>
<td>5.323</td>
<td>17.773</td>
</tr>
<tr>
<td>11</td>
<td>5.756</td>
<td>19.814</td>
<td>5.526</td>
<td>19.542</td>
<td>5.443</td>
<td>19.245</td>
</tr>
<tr>
<td>14</td>
<td>8.102</td>
<td>23.352</td>
<td>23.762</td>
<td>23.795</td>
<td>8.102</td>
<td>23.018</td>
</tr>
<tr>
<td>18</td>
<td>11.177</td>
<td>28.349</td>
<td>11.15</td>
<td>28.068</td>
<td>11.177</td>
<td>27.853</td>
</tr>
<tr>
<td>19</td>
<td>11.799</td>
<td>29.821</td>
<td>11.543</td>
<td>29.539</td>
<td>11.334</td>
<td>29.193</td>
</tr>
<tr>
<td>20</td>
<td>12.617</td>
<td>30.843</td>
<td>12.597</td>
<td>30.560</td>
<td>12.817</td>
<td>30.186</td>
</tr>
<tr>
<td>21</td>
<td>13.286</td>
<td>32.345</td>
<td>13.027</td>
<td>32.032</td>
<td>12.973</td>
<td>31.889</td>
</tr>
<tr>
<td>23</td>
<td>14.921</td>
<td>34.356</td>
<td>34.665</td>
<td>34.808</td>
<td>14.921</td>
<td>34.101</td>
</tr>
<tr>
<td>24</td>
<td>15.820</td>
<td>35.828</td>
<td>15.554</td>
<td>35.542</td>
<td>15.12</td>
<td>34.937</td>
</tr>
<tr>
<td>26</td>
<td>17.298</td>
<td>38.318</td>
<td>17.030</td>
<td>38.031</td>
<td>16.940</td>
<td>37.893</td>
</tr>
<tr>
<td>28</td>
<td>19.051</td>
<td>40.353</td>
<td>18.051</td>
<td>40.064</td>
<td>18.050</td>
<td>38.876</td>
</tr>
<tr>
<td>29</td>
<td>19.614</td>
<td>41.824</td>
<td>19.542</td>
<td>41.536</td>
<td>19.245</td>
<td>41.186</td>
</tr>
<tr>
<td>30</td>
<td>20.648</td>
<td>42.841</td>
<td>20.574</td>
<td>42.551</td>
<td>20.431</td>
<td>41.989</td>
</tr>
<tr>
<td>31</td>
<td>21.364</td>
<td>43.856</td>
<td>44.259</td>
<td>44.312</td>
<td>21.365</td>
<td>43.565</td>
</tr>
<tr>
<td>32</td>
<td>22.322</td>
<td>45.328</td>
<td>22.047</td>
<td>45.038</td>
<td>21.982</td>
<td>44.537</td>
</tr>
<tr>
<td>33</td>
<td>23.352</td>
<td>46.343</td>
<td>23.075</td>
<td>46.052</td>
<td>23.018</td>
<td>45.509</td>
</tr>
<tr>
<td>34</td>
<td>23.755</td>
<td>47.367</td>
<td>47.689</td>
<td>47.816</td>
<td>23.762</td>
<td>47.065</td>
</tr>
<tr>
<td>35</td>
<td>24.825</td>
<td>48.829</td>
<td>24.547</td>
<td>48.337</td>
<td>23.997</td>
<td>47.981</td>
</tr>
<tr>
<td>36</td>
<td>25.852</td>
<td>49.843</td>
<td>25.573</td>
<td>49.550</td>
<td>25.482</td>
<td>48.982</td>
</tr>
<tr>
<td>38</td>
<td>27.324</td>
<td>52.329</td>
<td>27.044</td>
<td>52.036</td>
<td>26.543</td>
<td>51.578</td>
</tr>
<tr>
<td>39</td>
<td>28.349</td>
<td>53.341</td>
<td>28.067</td>
<td>53.048</td>
<td>27.853</td>
<td>52.413</td>
</tr>
<tr>
<td>40</td>
<td>29.966</td>
<td>54.354</td>
<td>29.966</td>
<td>54.060</td>
<td>29.966</td>
<td>53.842</td>
</tr>
<tr>
<td>41</td>
<td>29.821</td>
<td>55.827</td>
<td>29.539</td>
<td>55.533</td>
<td>29.193</td>
<td>55.259</td>
</tr>
<tr>
<td>42</td>
<td>30.843</td>
<td>56.839</td>
<td>30.560</td>
<td>56.544</td>
<td>30.186</td>
<td>55.810</td>
</tr>
<tr>
<td>43</td>
<td>31.675</td>
<td>57.850</td>
<td>31.580</td>
<td>57.555</td>
<td>31.675</td>
<td>57.164</td>
</tr>
<tr>
<td>44</td>
<td>32.345</td>
<td>59.323</td>
<td>32.032</td>
<td>59.029</td>
<td>31.889</td>
<td>58.970</td>
</tr>
<tr>
<td>45</td>
<td>33.336</td>
<td>60.336</td>
<td>33.052</td>
<td>60.039</td>
<td>32.505</td>
<td>59.184</td>
</tr>
<tr>
<td>46</td>
<td>34.356</td>
<td>61.346</td>
<td>34.070</td>
<td>61.050</td>
<td>34.101</td>
<td>60.480</td>
</tr>
<tr>
<td>47</td>
<td>34.808</td>
<td>62.357</td>
<td>34.665</td>
<td>62.060</td>
<td>34.665</td>
<td>61.983</td>
</tr>
<tr>
<td>48</td>
<td>35.828</td>
<td>63.380</td>
<td>35.542</td>
<td>63.534</td>
<td>34.937</td>
<td>63.12</td>
</tr>
<tr>
<td>49</td>
<td>36.846</td>
<td>64.841</td>
<td>36.560</td>
<td>64.544</td>
<td>36.187</td>
<td>63.794</td>
</tr>
<tr>
<td>50</td>
<td>37.667</td>
<td>65.851</td>
<td>37.576</td>
<td>65.553</td>
<td>37.667</td>
<td>65.135</td>
</tr>
</tbody>
</table>
Appendix B: Proofs

Theorem 1: If a random variable $X$ has the Poisson distribution then $E(X) = \lambda$ and $Var(X) = \lambda$.

Proof

$$E(X) := \sum_{k=0}^{\infty} kP(X = k)$$

$$= \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!}$$

$$= e^{-\lambda} \lambda \sum_{k=0}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

$$= e^{-\lambda} \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$$

$$= e^{-\lambda} \lambda e^\lambda = \lambda$$

$$E(X(X - 1)) = \sum_{k=0}^{\infty} k(k - 1)P(X = k)$$

$$= \sum_{k=0}^{\infty} k(k - 1) \frac{e^{-\lambda} \lambda^k}{k!}$$

$$= e^{-\lambda} \lambda^2 \sum_{k=0}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}$$

$$= e^{-\lambda} \lambda^2 \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$$

$$= e^{-\lambda} \lambda^2 e^\lambda = \lambda^2$$

Therefore,

$$E(X^2) = E((X(X - 1) + X) = E(X(X - 1)) + E(X) = \lambda^2 + \lambda$$

So,
\( \text{Var}(X) = E(X^2) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \)

**Theorem 2:** If \( N_1, N_2, \ldots, N_j \) are independent Poisson random variables with parameters \( \lambda_1, \lambda_2, \ldots, \lambda_j \) respectively, then \( N = \sum_{j=1}^{j} N_j = N_1 + N_2 + \cdots + N_j \) is a Poisson random variable with parameter \( \lambda = \sum_{j=1}^{j} \lambda_j = \lambda_1 + \lambda_2 + \cdots + \lambda_j. \)

The following proof has been modeled after the proof found in Pitman(1993).

**Proof**

Suppose \( N_1 \) and \( N_2 \) are independent Poisson random variables with parameters \( \lambda_1 \) and \( \lambda_2 \) respectively. Then,

\[
P(N_1 + N_2 = k) = \sum_{x=0}^{k} P(N_1 = x)P(N_2 = k - x)
\]

\[
= \sum_{x=0}^{k} \left( \frac{e^{-\lambda_1} \lambda_1^x}{x!} \right) \left( \frac{e^{-\lambda_2} \lambda_2^{k-x}}{(k-x)!} \right)
\]

\[
= \sum_{x=0}^{k} e^{-(\lambda_1+\lambda_2)} \left( \frac{\lambda_1^x}{x!} \right) \left( \frac{\lambda_2^{k-x}}{(k-x)!} \right)
\]

\[
= \left( \frac{(\lambda_2 + \lambda_2)^k}{k!} \right) \left( \frac{k!}{(\lambda_2 + \lambda_2)^k} \right) \sum_{x=0}^{k} \left( \frac{\lambda_1^x \lambda_2^{k-x}}{x! (k-x)!} \right)
\]

\[
= e^{-(\lambda_1+\lambda_2)} \left( \frac{(\lambda_2 + \lambda_2)^k}{k!} \right) \sum_{x=0}^{k} \left( \frac{k!}{x! (k-x)!} \right) \left( \frac{\lambda_1^x \lambda_2^{k-x}}{(\lambda_2 + \lambda_2)^k} \right)
\]

\[
= e^{-(\lambda_1+\lambda_2)} \left( \frac{(\lambda_2 + \lambda_2)^k}{k!} \right) \sum_{x=0}^{k} \left( \frac{k!}{x! (k-x)!} \right) \left( \frac{\lambda_1^x \lambda_2^{k-x}}{(\lambda_2 + \lambda_2)^k} \right)
\]

\[
= e^{-(\lambda_1+\lambda_2)} \left( \frac{(\lambda_2 + \lambda_2)^k}{k!} \right) \sum_{x=0}^{k} \left( \frac{k!}{x! (k-x)!} \right) \left( \frac{\lambda_1^x \lambda_2^{k-x}}{(\lambda_2 + \lambda_2)^k} \right)
\]

\[
= e^{-(\lambda_1+\lambda_2)} \left( \frac{(\lambda_2 + \lambda_2)^k}{k!} \right) (1)
\]

\[
= p_{\lambda_1+\lambda_2}(k)
\]
because \( \sum_{x=0}^{k} \left( \frac{k!}{x!(k-x)!} \right) \left( \frac{\lambda_1}{\lambda_2 + \lambda_2} \right)^x \left( 1 - \frac{\lambda_1}{\lambda_2 + \lambda_2} \right)^{k-x} \) is the sum of the terms in a binomial \((k, p)\) with \( p = \frac{\lambda_1}{\lambda_2 + \lambda_2} \) and this sum is 1. Repeated application of this result for two terms will give the result for any finite number \( J \) terms.

**Theorem 3:** If a strict confidence procedure \( C \) has the Least Cardinality property then it has the Inability to be Shortened property.

**Proof**

Suppose a strict confidence procedure \( C \) has the Least Cardinality property. Now define a new confidence procedure \( C' \) that is identical to \( C \) except for the confidence interval for a fixed \( x_0 \), we increase the lower endpoint by \( \varepsilon \); that is, \( l'_x = l_x \) for all \( x \neq x_0 \) and \( u'_x = u_x \) for all \( x \), but \( l'_{x_0} = l_{x_0} + \varepsilon \). Then for all \( \lambda \in [l_{x_0}, l'_{x_0}) \), \( \text{cov}(\lambda \in C') = \text{cov}(\lambda \in C) - p_\lambda(x_0) \). So then the segments of coverage probability function of \( C' \) corresponding to the \( \lambda \)'s in \([l_{x_0}, l'_{x_0})\) will be of cardinality one less than that of \( C \). But, this implies all these \( \lambda \)'s will be in one fewer confidence interval. This means the coverage of \( C' \) must drop below the confidence level since by the Least Cardinality property every \( \lambda \) was already in the fewest possible CI's without dropping below the confidence level. Similarly, decreasing any upper endpoint also causes the coverage probability to fall below the confidence level. In either case a contradiction to the assumption that \( C \) is a strict confidence procedure occurs. Hence, \( C \) has the Inability to be Shortened property. \( \blacksquare \)

**Theorem 4:** Assume that \( x \in [l_x, u_x) \ \forall x \). Then if for some integer \( k \geq 0 \), the confidence procedure \( C \) has \( l_k > l_{k+1} \) (or \( u_k > u_{k+1} \)), then we can modify \( C \) so that \( l_k < l_{k+1} \) (or \( u_k < u_{k+1} \)) without affecting overall interval length but improving coverage.

**Proof**

Suppose that \( l_k > l_{k+1} \) and define a new confidence procedure \( C' \) that is exactly the same as \( C \) except \( l'_k = l_{k+1} \) and \( l'_{k+1} = l_k \). Note since we are assuming that \( x \in [l_x, u_x) \) we have that \( u_{k+1} \geq k + 1 > k \geq l_k = l'_{k+1} \).

Thus,
If \( \lambda \in (l_{k+1}, l_k) \) then \( \text{cov}_{C'}(\lambda) - \text{cov}_{C}(\lambda) = 0 \). If \( \lambda \in [l_{k+1}, l_k) \) then \( \frac{\lambda}{k+1} < 1 \) since \( \lambda \leq l_k \leq k \) because \( k \in [l_k, u_k) \) by assumption. Thus, \( \text{cov}_{C'}(\lambda) - \text{cov}_{C}(\lambda) > 0 \). Therefore, not only are the lower endpoints \( l'_k \) and \( l'_{k+1} \) in proper increasing order but, \( C' \) has better coverage than \( C \). Moreover, changing the values of \( l_k \) and \( l_{k+1} \) to the values \( l'_k \) and \( l'_{k+1} \) did not change the overall interval length of the confidence procedure since we decreased the interval for \( k + 1 \) by the same amount we increased the interval for \( k \).

The proof is similar if \( u_k > u_{k+1} \).

**Proposition 5:** The graph of \( P_{x,y}(\lambda) \) attains its maximum at \( \lambda = \lambda_{max}(x, y) \), where

\[
\lambda_{max}(x, y) = [(x)(x+1) \cdots (y)]^{1/(y-x+1)}.
\]

Furthermore, \( P_{x,y}(\lambda) \) is strictly increasing on \( \lambda \in (0, \lambda_{max}(x, y)) \) and is strictly decreasing on \( \lambda \in (\lambda_{max}(x, y), \infty) \).

**Proof**

For fixed values of \( x \) and \( y \) we would like to compute the maximum of \( P_{x,y}(\lambda) \). If \( x = 0 \) then \( P_{x,y}(\lambda) \) has a maximum value of 1 at

\[
\lambda = 0 = [(0)(0+1) \cdots (y)]^{1/(y-0+1)} = \lambda_{max}(x, y).
\]
So assume $x > 0$. Then, since for any $x > 0$ the function $P_{x,y}(\lambda) = 0$ for $\lambda = 0$, we can also assume $\lambda > 0$. Taking the derivative of

$$P_{x,y}(\lambda) = \sum_{i=x}^{y} p_{\lambda}(i),$$

we get,

$$\frac{d}{d\lambda} \sum_{i=x}^{y} p_{\lambda}(i) = \frac{d}{d\lambda} \sum_{i=x}^{y} \frac{e^{-\lambda} \lambda^i}{i!} = \sum_{i=x}^{y} \frac{d}{d\lambda} \frac{e^{-\lambda} \lambda^i}{i!} = \sum_{i=x}^{y} \left( \frac{e^{-\lambda} \lambda^i}{i!} + \frac{ie^{-\lambda} \lambda^{i-1}}{i!} \right) = \sum_{i=x}^{y} \left( \frac{e^{-\lambda} \lambda^i}{i!} + \frac{e^{-\lambda} \lambda^{i-1}}{(i-1)!} \right) = \sum_{i=x}^{y} \left( \frac{e^{-\lambda} \lambda^{i-1}}{(i-1)!} - \frac{e^{-\lambda} \lambda^i}{i!} \right) = \sum_{i=x}^{y} [p_{\lambda}(i-1) - p_{\lambda}(i)] = [p_{\lambda}(x-1) - p_{\lambda}(x)] + [p_{\lambda}(x) - p_{\lambda}(x+1)] + \cdots + [p_{\lambda}(y-2) - p_{\lambda}(y-1)] + [p_{\lambda}(y-1) - p_{\lambda}(y)] = p_{\lambda}(x-1) - p_{\lambda}(y).$$

Setting $\frac{d}{d\lambda} \sum_{i=x}^{y} p_{\lambda}(i) = 0$ and assuming $x > 0$ and $\lambda > 0$

$$\Rightarrow p_{\lambda}(x-1) - p_{\lambda}(y) = 0$$

$$\Rightarrow \frac{e^{-\lambda} \lambda^{x-1}}{(x-1)!} - \frac{e^{-\lambda} \lambda^y}{y!} = 0$$

$$\Rightarrow e^{-\lambda} \lambda^{x-1} \left( \frac{1}{(x-1)!} - \frac{\lambda^{y-x+1}}{y!} \right) = 0$$
\[ \Rightarrow \left( \frac{1}{(x-1)!} - \frac{\lambda^{y-x+1}}{y!} \right) = 0 \text{ since we assumed } \lambda > 0 \]
\[ \Rightarrow \frac{1}{(x-1)!} = \frac{\lambda^{y-x+1}}{y!} \]
\[ \Rightarrow \lambda = \left[ \frac{y!}{(x-1)!} \right]^{1/(y-x+1)} \]
\[ \Rightarrow \lambda = [(x)(x+1) \cdots (y)]^{1/(y-x+1)} \]

Now since \( P_{x,y}(\lambda) > 0 \) for \( \lambda > 0 \), \( P_{x,y}(\lambda) = 0 \) for \( \lambda = 0 \), and \( P_{x,y}(\lambda) = \sum_{i=x}^{y} \frac{e^{-\lambda} \lambda^i}{i!} \rightarrow 0 \) as \( \lambda \to \infty \), we have that \( \lambda = \lambda_{\text{max}}(x,y) \) is indeed the location of the maximum. Furthermore, \( P_{x,y}(\lambda) \) is strictly increasing on \( \lambda \in (0, \lambda_{\text{max}}(x,y)) \) and is strictly decreasing on \( \lambda \in (\lambda_{\text{max}}(x,y), \infty) \). \[ \blacksquare \]

**Proposition 6:** For \( \lambda \in (0, \infty) \), the graph of \( P_{x,y}(\lambda) \) intersects the graph of \( P_{x+1,y+1}(\lambda) \) at a single point, which is the maximum of the graph \( P_{x+1,y+1}(\lambda) \).

**Proof**

\[ P_{x,y}(\lambda) = P_{x+1,y+1}(\lambda) \]
\[ \Rightarrow \sum_{i=x}^{y} \frac{e^{-\lambda} \lambda^i}{i!} = \sum_{j=x+1}^{y+1} \frac{e^{-\lambda} \lambda^j}{j!} \]
\[ \Rightarrow \sum_{i=x}^{y} \frac{e^{-\lambda} \lambda^i}{i!} - \sum_{j=x+1}^{y+1} \frac{e^{-\lambda} \lambda^j}{j!} = 0 \]
\[ \Rightarrow p_\lambda(x) - p_\lambda(y+1) = 0 \]
\[ \Rightarrow \lambda = [(x+1)(x+2) \cdots (y+1)]^{1/(y+1)-(x+1)+1} \text{ by the proof of Proposition 5} \]
\[ \Rightarrow \lambda = \lambda_{\text{max}}(x+1, y+1). \blacksquare \]

**Proposition 7:**

a) \( P_{x,y}(\lambda) > P_{x+1,y+1}(\lambda) \) for all \( \lambda \in (0, \lambda_{\text{max}}(x+1, y+1)) \)
b) $P_{x,y}(\lambda) < P_{x+1,y+1}(\lambda)$ for all $\lambda \in (\lambda_{\text{max}}(x+1, y+1), \infty)$

c) $\lambda_{\text{max}}(x,y) < \lambda_{\text{max}}(x+1, y+1)$

d) $\max_{\lambda}\{P_{x,y}(\lambda)\} > \max_{\lambda}\{P_{x+1,y+1}(\lambda)\}$

Proof

$P_{x,y}(\lambda) \geq P_{x+1,y+1}(\lambda)$ according as $p_{\lambda}(x) - p_{\lambda}(y + 1) \geq 0$.

But by the proof of Proposition 5, this is precisely when $\frac{d}{d\lambda} \sum_{i=x+1}^{y+1} p_{\lambda}(i) \geq 0$; that is, when $P_{x+1,y+1}(\lambda)$ is increasing or respectively decreasing. So the statement of Proposition 5 implies

\[
\frac{d}{d\lambda} \sum_{i=x+1}^{y+1} p_{\lambda}(i) > 0 \text{ for all } \lambda \in (0, \lambda_{\text{max}}(x+1, y+1)) \text{ and } \frac{d}{d\lambda} \sum_{i=x+1}^{y+1} p_{\lambda}(i) < 0 \text{ for all } \lambda \in (\lambda_{\text{max}}(x+1, y+1), \infty).
\]

Therefore,

$P_{x,y}(\lambda) > P_{x+1,y+1}(\lambda)$ for all $\lambda \in (0, \lambda_{\text{max}}(x+1, y+1))$ and $P_{x,y}(\lambda) < P_{x+1,y+1}(\lambda)$ for all $\lambda \in (\lambda_{\text{max}}(x+1, y+1), \infty)$.

Now notice

$\lambda_{\text{max}}(x,y) \in (0, \lambda_{\text{max}}(x+1, y+1))$

since

\[
[(x)(x + 1) \ldots (y)]^{1/(y-x+1)} < [(x + 1)(x + 2) \ldots (y + 1)]^{1/(y-x+1)}.
\]

So we have $\lambda_{\text{max}}(x,y) < \lambda_{\text{max}}(x+1, y+1)$. Thus, $\max_{\lambda}\{P_{x,y}(\lambda)\} \geq \max_{\lambda}\{P_{x+1,y+1}(\lambda)\}$ follows from part a). \(\blacksquare\)

Proposition 8: Suppose $x, y, x, \bar{y}$ are nonnegative integers such that $\bar{x} \leq x$ and $y \leq \bar{y}$ where at least one of these inequalities is strict. Then $P_{x,y}(\lambda) < P_{\bar{x},\bar{y}}(\lambda)$ for all $\lambda \in (0, \infty)$. 

71
Case 1: $\bar{x} = x$ and $y < \bar{y}$
Case 2: $\bar{x} < x$ and $y = \bar{y}$
Case 3: $\bar{x} < x$ and $y < \bar{y}$

Case 1

$$P_{x,y}(\lambda) = \sum_{i=x}^{y} \frac{e^{-\lambda} \lambda^i}{i!}$$

$$= \sum_{i=\bar{x}}^{y} \frac{e^{-\lambda} \lambda^i}{i!}$$

$$< \sum_{i=\bar{x}}^{y} \frac{e^{-\lambda} \lambda^i}{i!} + \sum_{i=y+1}^{\bar{y}} \frac{e^{-\lambda} \lambda^i}{i!}$$

$$= \sum_{i=\bar{x}}^{\bar{y}} \frac{e^{-\lambda} \lambda^i}{i!} = P_{\bar{x},\bar{y}}(\lambda)$$

Cases 2 and 3 are proven similarly. ■

**Lemma 9**: If $C$ has the Least Cardinality property then as $\lambda$ increases the cardinality of the acceptance sets for $C$ never decrease.

**Proof**

Let $M(\lambda) = \min \{k: P_{j,j+k-1}(\lambda) \geq 1 - \alpha\}$ where $j$ is some positive integer; i.e., $M(\lambda)$ is the value of minimal cardinality at $\lambda$. Note that since $C$ has the Least Cardinality property it’s cpf must always use segments of the lowest cardinality which are still at or above the confidence level. During the construction of MS we defined a necklace of cardinality $k$ to be the function $N_k(\lambda) := \max_j \{P_{j,j+k-1}(\lambda): j = 0, 1, 2, \ldots\}$; i.e., $N_k(\lambda)$ is the upper envelope of all the acceptance curves of cardinality $k$. Since $P_{0,k-1}(0) = 1$, Proposition 5 implies $N_k(\lambda)$ is a strictly decreasing function and by Proposition 7 $N_k(\lambda) \leq N_{k+1}(\lambda) \forall \lambda \in (0, \infty)$. So given any $\lambda_0$, the following hold for each $k < M(\lambda_0)$.

\( (i) \quad N_k(\lambda_0) < 1 - \alpha \) by the definition of $M(\lambda_0)$;
(ii) \( N_k(\lambda) < 1 - \alpha \forall \lambda > \lambda_0 \) by (i) since \( N_k(\lambda) \) is decreasing;

(iii) from the definition of \( N_k(\lambda) \) it follows from (ii) that all acceptance curves of cardinality \( k \) are entirely below \( 1 - \alpha \forall \lambda > \lambda_0 \).

Hence, \( M(\lambda) \geq M(\lambda_0) \forall \lambda > \lambda_0 \).

**Theorem 10:** If \( C = \{(l_x, u_x), x = 0, 1, 2, \ldots\} \) has the Least Cardinality property then every \( u_x \) is coincidental with some \( l_y \) for \( y > x \).

**Proof**

Let \( \{a_i\}, \{b_i\} \) be the sequence of values of the acceptance sets \( P_{a_i,b_i}(\lambda) \) of \( C \) (Each \( P_{a_i,b_i}(\lambda) \) is a segment used for the cpf of \( C \) where the sequences are ordered by which segments are used for the cpf for \( C \) for increasing \( \lambda \). An upper endpoint occurs whenever \( a_{i+1} > a_i \), giving the upper endpoint for \( a_i \) and a lower endpoint occurs whenever \( b_{i+1} > b_i \), giving a lower endpoint for all integers \( b \) such that \( b_i + 1 \leq b \leq b_{i+1} \). But since \( C \) has the Least Cardinality Property by Lemma 9 whenever we have \( a_{i+1} > a_i \) we must have \( b_{i+1} > b_i \), otherwise the cardinality decreases. Thus, the upper endpoint \( u_{a_i} \) is coincidental with the lower endpoint \( l_{b_{i+1}} \) where \( b_i + 1 > a_i \).

**Theorem 11:** Let \( C' = \{(l'_x, u'_x), x = 0, 1, 2, \ldots\} \) be the confidence procedure produced by CG and let \( C = \{(l_x, u_x), x = 0, 1, 2, \ldots\} \) be any other confidence procedure which has the Least Cardinality property. Then

\[
\sum_{x=0}^{K} (u'_x - l'_x) \leq \sum_{x=0}^{K} (u_x - l_x) \forall k.
\]

**Proof**

First note that since CG chooses its endpoints to be the smallest of all possible choices for a procedure satisfying the Least Cardinality property we have \( u'_x \leq u_x \) for all \( x \). Also note \( l'_y \) is coincidental if and only if \( l_y \) is coincidental. Additionally by Theorem 10 for all \( x \leq K, u'_x \) and \( u_x \) are coincidental with lower endpoints \( l'_y \) and \( l_y \) respectively for \( y > x \).

Fix \( K \) and let \( H = \{y: y \leq K, l_y \text{ is coincidental}\} \). Therefore,

\[
\sum_{x=0}^{K} (u'_x - l'_x) - \sum_{x=0}^{K} (u_x - l_x)
\]
\[
= \sum_{x=0}^{K} (u'_x - u_x) + \sum_{x=0}^{K} (l_x - l'_x)
\]

\[
= \sum_{x=0}^{K} (u'_x - u_x) + \sum_{y \in H} (l_y - l'_y) + \sum_{y \in H} (l_y - l'_y)_{\text{=0}}
\]

\[
= \sum_{x=0}^{K} (u'_x - u_x) + \sum_{y \in H} (l_y - l'_y)
\]

\[
= \sum_{x \in f} (u'_x - u_x) \text{ where } f = \{x: u_x \text{ is coincidental with } l_y \text{ for } y > K\}.
\]

The last equality follows from the fact that \(u'_x - u_x + l_y - l'_y = 0\) whenever \(u'_x\) and \(u_x\) are coincidental with lower endpoints \(l'_y\) and \(l_y\) respectively. Thus, since all differences \(l_y - l'_y\) in the sum \(\sum_{y \in H} (l_y - l'_y)\) have a corresponding a difference \(u'_x - u_x\) in the sum \(\sum_{x=0}^{K} (u'_x - u_x)\) that cancels with it, the entire sum \(\sum_{y \in H} (l_y - l'_y)\) vanishes. After cancellation we are left with a sum of differences, \((u'_x - u_x)\) with values \(u'_x\) and \(u_x\) which have coincidental endpoints \(l'_y\) and \(l_y\) for \(y > K\). And since \(u'_x \leq u_x\ \forall x\) the remaining sum \(\sum_{x \in f} (u'_x - u_x) \leq 0\). ■