Title
Contributions to Interval Estimation for Parameters of Discrete Distributions

Author
Holladay, Bret Andrew

Publication Date
2019

Peer reviewed|Thesis/dissertation

University of California
Santa Barbara

Contributions to Interval Estimation for Parameters of Discrete Distributions

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Statistics and Applied Probability

by

Bret Andrew Holladay

Committee in charge:

Professor Sreenivasa Rao Jammalamadaka, Chair
Professor John Hsu
Professor Sang-Yun Oh

September 2019

The Dissertation of Bret Andrew Holladay is approved.

Professor John Hsu

Professor Sang-Yun Oh

Professor Sreenivasa Rao Jammalamadaka, Committee Chair

September 2019

Contributions to Interval Estimation for Parameters of Discrete Distributions

Dedicated to my best friend Scarlet

Acknowledgements

First and foremost I would like to express my deepest appreciation to my committee chair, Professor Sreenivasa Rao Jammalamadaka. Without your guidance and support this dissertation would not have been possible. You have a gift for teaching the most difficult of topics at a level that students can understand, and you are a big part of the reason I have decided to pursue a career in teaching at the university level.

I'd also like to thank Professor John Hsu and Professor Sang-Yun Oh for your support, and for serving as my committee members. In addition, I'm thankful to my family, friends, classmates, professors, and colleagues who helped me along my journey towards a doctoral degree in Statistics and Applied Probability. A special thanks to my classmates Yuanbo Wang and Nhan Huynh for always being there for me. I owe much of my success to you both.

I am also grateful to Heather Holladay, Matt Heitner, and Amber Gibson for all your help with my job applications. I can never thank you enough.

I especially want to thank my parents for their guidance and unconditional love. I would not be the man I am today without you.

And finally, Redilyn Holladay, thank you for your unwavering love and support. You stuck by my side through stress, tears, and hard times, and have sacrificed so much for me.

Curriculum Vitæ

Bret Andrew Holladay

Education

2019	Ph.D. in Statistics and Applied Probability, University of California, Santa Barbara.
2018	M.A. in Statistics, University of California, Santa Barbara.
2014	M.S. in Applied Mathematics, California State University, Northridge.
2011	B.S. in Applied Mathematics, California State University, Northridge.

Experience

2014-2019	Teaching Assistant, Department of Statistics and Applied Probability, University of California, Santa Barbara.
2016-2019	Teaching Associate, Department of Statistics and Applied Probability, University of California, Santa Barbara.
2018-2019	Lead Teaching Assistant, Department of Statistics and Applied Probability, University of California, Santa Barbara.
2011-2014	Teaching Associate, Department of Mathematics, California State University, Northridge.
2010-2014	Mathematics Tutor, Math Tutoring Center, California State University, Northridge.
2012 Summer	Consultant for California Department of Toxic Substances Control, Department of Mathematics, California State University, Northridge.
2011 Summer	Research Assistant, Department of Mathematics, California State University, Northridge.
2010 Summer	Preparing Undergraduates through Mentoring toward Ph.D.s Participant, Department of Mathematics, California State University, Northridge.
2008-2011	Subject Area Tutoring (SAT) Lab Tutor, Learning Resource Center, California State University, Northridge

Publications

2019	J.E. Gonzlez, J.F. Barquinero, B. Holladay, M. Giorgio, M. Higueras, Uncertainty calculation methods in dose assessment for dicentric chromosome assay (submitted for publication)
2017	Mark F. Schilling and Bret Holladay (2017), Length minimization for Poisson confidence procedures, Communications in Statistics - Theory and Methods, 46:2, 861-873

Awards

2013-2014	Outstanding Performance as Teaching Associate, California State University, Northridge.

Abstract

Contributions to Interval Estimation for Parameters of Discrete Distributions

by

Bret Andrew Holladay

We study interval estimation for parameters of discrete distributions, focusing on the binomial, Poisson, negative binomial, and hypergeometric distributions explicitly. We provide a broad treatment of the problem, covering both conventional and randomized confidence intervals, as well as Geyer and Meeden's concept of fuzzy confidence intervals. We take a graphical approach to the problem through the use of coverage probability functions and determine the optimal procedure under each of a wide variety of criteria, including multiple notions of length. Several new methods are proposed, including a method that produces length optimal fuzzy confidence intervals. Credible intervals and multi-parameter discrete distributions are also considered.

Contents

Chapter 1

Introduction

1.1 Introduction

The problem of finding optimal confidence intervals for one-parameter discrete distributions is a classical one with considerable history and literature behind it. However, after all this time, it appears that no consensus has been reached as to what may be considered "optimal" intervals. This is evidenced in many of the recent papers such as Blaker (2000) [16], Byrne & Kabaila (2005) [17], Geyer & Meeden (2005)[18], Schilling & Doi (2014)[23], and Schilling & Holladay (2017)[22]. Our goal here is to demonstrate that there are still potential improvements one can make, and look for what might be considered optimal procedures under various criteria. Clearly primary interest lies in finding intervals with the shortest possible length that maintain the nominal confidence level. In particular, we focus on what may be considered the preeminent discrete statistical models, namely the binomial, Poisson, hypergeometric, and negative binomial distributions explicitly. We also discuss confidence intervals for both the binomial sample size parameter n and the negative binomial parameter r that specifies the target number of successes. We treat the cases where the unknown parameter comes from a countable parameter space (e.g. n and r) separately.

We give a comprehensive graphical perspective to the problem, through the use of coverage probability functions, and argue in favor of particular class of procedures that correspond to acceptance regions of minimal size. We show that conventional methods for constructing confidence intervals have coverage probability consistently above the nominal level. This is partly an unavoidable characteristic of discrete data, which leads to conservative intervals. We mention how mean-coverage procedures and Bayesian approaches can often alleviate the over-coverage issue.

A third remedy to over-coverage is randomization that builds on randomized tests, as it allows for exact $(1-\alpha)$ coverage probability and consequently shorter intervals. However, some practitioners are critical of randomized procedures since two researchers using the same data, and confidence level may come to different intervals even with the same procedure due to the added randomness. Geyer and Meeden (2005) [18] introduced the notion of "fuzzy confidence intervals" which inherit the exact $(1-\alpha)$ coverage probability of randomized procedures. However, since fuzzy intervals do not require randomization, they avoid some of the issues of randomized confidence intervals. Geyer and Meeden's (2005) [18] fuzzy intervals are derived from the uniformly most accurate unbiased randomized intervals. We show by relaxing the unbiased condition shorter fuzzy intervals can be obtained. Lastly, we discuss applications of the work in the one-parameter case to the two-parameter situation. Credible intervals are also considered.

1.2 Background, Definitions, and Notation

Let X be a discrete real-valued random variable with integer-valued support $\mathcal{X}$. Suppose further that X has probability mass function (PMF) $p_X(x|\theta)$ with unknown parameter $\theta \in \Theta \subset \mathbb{R}$. Table 1.1 gives some basic models of interest, viz. the binomial, Poisson, negative binomial (NB). Here, $NB(r, \theta)$ is the distribution for the number of failures in successive Bernoulli trials before the rth success.

We initially deal with a single random variable X rather than a sample $X_1, ..., X_N$, both for convenience and the fact that this can be done without loss of generality, in models like the binomial, Poisson and the NB where X can represent the sufficient statistic $X = \sum_{i=1}^{N} X_i$ based on the sample.

Distribution	$p(x\|\theta)$	$\mathcal{X}$	Θ
Binomial(n, θ)	$\binom{n}{x}\theta^x(1-\theta)^{n-x}$	$\{0, 1, 2, ..., n\}$	$[0, 1]$
Poisson(θ)	$\frac{e^{-\theta}\theta^x}{x!}$	$\{0, 1, 2, ...\}$	$[0, \infty)$
NB(r, θ)	$\binom{x+r-1}{x}\theta^r(1-\theta)^x$	$\{0, 1, 2, ...\}$	$[0, 1]$

Table 1.1: Binomial, Poisson, and negative binomial (NB) distributions

In interval estimation, we are interested in constructing a family of random sets that contain the true parameter value θ with a guaranteed probability. The observed value of X influences the likelihood of any θ; for instance, in a binomial experiment a small number of successes gives a larger likelihood for the small values of the success parameter θ. Therefore, a different set of parameter values are indicated for each observed value $x \in \mathcal{X}$.

Definition 1.1 (Confidence Procedure) Let $0 < \alpha < 1$ be a fixed constant and let $C : \mathcal{X} \to \mathcal{P}(\Theta)$ be a map from the support of X to the power set of the parameter space Θ. The family of subsets $C(x)$ of Θ is called a *family of confidence sets* at confidence level $(1 - \alpha)$ if

$$\inf_{\theta \in \Theta} \mathbb{P}_\theta \{C(X) \ni \theta\} \geq 1 - \alpha. \tag{1.1}$$

In this case the function C defines a level $(1 - \alpha)$ *confidence procedure.*

C can be thought of as a function that assigns each x to a different subset $C(x)$ of Θ. $(1-\alpha)$ in (1.1) may sometimes serve only as an approximate lower bound. Moreover, it may sometimes be the case that the attained *confidence coefficient*, $\inf_{\theta\in\Theta}\mathbb{P}_\theta\left\{C(X)\ni\theta\right\}$ is strictly greater than the provided nominal level $(1-\alpha)$. For families of distributions with monotone likelihood ratio in a statistic $T(X)$ it is natural to restrict attention to confidence sets based on $T(X)$ that are intervals (Blaker (2000)[16]). When a confidence set $C(x)$ is an interval $[l(x), u(x)]$, it's called a *confidence interval* with lower and upper limits $l(x)$ and $u(x)$ respectively. The binomial(n, θ) and Poisson(θ) distributions, for instance, are two distributions for which confidence intervals make sense, and such confidence intervals will be of primary interest. Half-open confidence intervals of the form $[l, u)$ or $(l, u]$ will sometimes be used to avoid having sharp spikes (drops) in coverage probability that occur when the upper endpoint of one interval is tied with the lower endpoint of another.

Definition 1.2 (Acceptance Region) For any $\theta\in\Theta$ and confidence procedure C, define the *acceptance region* A_θ by,

$$A_\theta \coloneqq \left\{x\in\mathcal{X} : \theta\in C(x)\right\}. \tag{1.2}$$

The set of x's in $\mathcal{X}$ excluded from A_θ is called the *rejection region*.

A confidence procedure C defines both a family of confidence sets $\{C(x) : x\in\mathcal{X}\}$ and the corresponding family of acceptance regions $\{A_\theta : \theta\in\Theta\}$. The two families have an underlying confidence level that they inherit from C. Since the confidence level in Definition 1.1 serves only as a (approximate) lower bound for the probabilities $P_\theta\{C(X)\ni\theta\}$ we now consider the actual coverage probability of a procedure.

Definition 1.3 (Coverage Probability) Given a confidence procedure C, define the *coverage probability function* (CPF), $CP : \Theta \to [0,1]$ by,

$$CP(\theta) \coloneqq \mathbb{P}_\theta\{C(X) \ni \theta\}, \tag{1.3}$$

so that $CP(\theta)$ gives the probability that C covers the true parameter value θ.

In general the CPF for a discrete random variable is a discontinuous piecewise function that oscillates over the parameter space. Analyzing a confidence procedure through the structure of its CPF is crucial to a true understanding of its behavior and performance. Since $\theta \in C(x)$ if and only if $x \in A_\theta$, coverage may also written as $CP(\theta) = \mathbb{P}_\theta(X \in A_\theta)$. Thus, if the family of acceptance regions produced by C is used for testing the hypothesis $H_0 : \theta = \theta_0$ then the following relationship between coverage probability and type I error holds,

$$CP(\theta_0) = 1 - \mathbb{P}(\text{Type I error}). \tag{1.4}$$

It follows that,

$$\mathbb{P}_\theta\{X \notin A_\theta\} \leq \alpha \iff \mathbb{P}_\theta\left\{C(X) \ni \theta\right\} \geq 1 - \alpha. \tag{1.5}$$

Thus, $\{C(x)\}$ will be a family of $(1-\alpha)$ confidence sets if and only if for each $\theta_0 \in \Theta$, A_{θ_0} is an acceptance region of a level α test of $H_0 : \theta = \theta_0$.

Definition 1.4 (Acceptance curve) Let A_θ be the acceptance region for a given $\theta \in \Theta$. The function $\mathbb{P}_{\theta'}(X \in A_\theta)$ as a function of $\theta' \in \Theta$ is called the *acceptance curve* associated with θ.

The coverage probability function of a confidence procedure will be a piecewise function made up of a collection of acceptance curves. When X has a unimodal distribution (as is the case with the binomial and Poisson distributions for example) the only good acceptance regions will be comprised of contiguous values of x which surround the peak of the distribution. Thus, in such cases there will be an $a = \inf\{x \in A_\theta\}$ and $b = \sup\{x \in A_\theta\}$ such that $CP(\theta) = \mathbb{P}_\theta(a \leq X \leq b)$. Later we will see that the behavior of the sequences of $\{a\}$ and $\{b\}$ values obtained as θ increases play an important role in determining the performance of a procedure.

To help tie together the definitions given in this section, consider the family of 80% confidence intervals for the binomial$(4, \theta)$ distribution in Table 1.2. The associated CPF of this procedure is shown in Figure 1.1. Take for instance the confidence interval, $C(0) = [0.00, 0.44]$. For any $\theta \in C(0)$ it must follow that the observed value $x = 0$ is in A_θ. Thus, the CPF uses acceptance curves that involve $x = 0$ only for $\theta \in [0.00, 0.44]$. Notice the CPF's discontinuous oscillatory behavior which is due to the discrete nature of the random variable. Another characteristic of the CPF is that it is always greater than the nominal level 0.8. In terms of hypothesis testing this means the procedure will have type I error probabilities no bigger than 0.2, although the actual probabilities will be much smaller, since the CPF is far above 0.8 for most values of θ.

x	$C(x)$
0	$[0.00, 0.44]$
1	$[0.03, 0.68]$
2	$[0.14, .086]$
3	$[0.32, 0.97]$
4	$[0.56, 1.00]$

Table 1.2: 80% Confidence Procedure for binomial$(4, \theta)$ distribution.

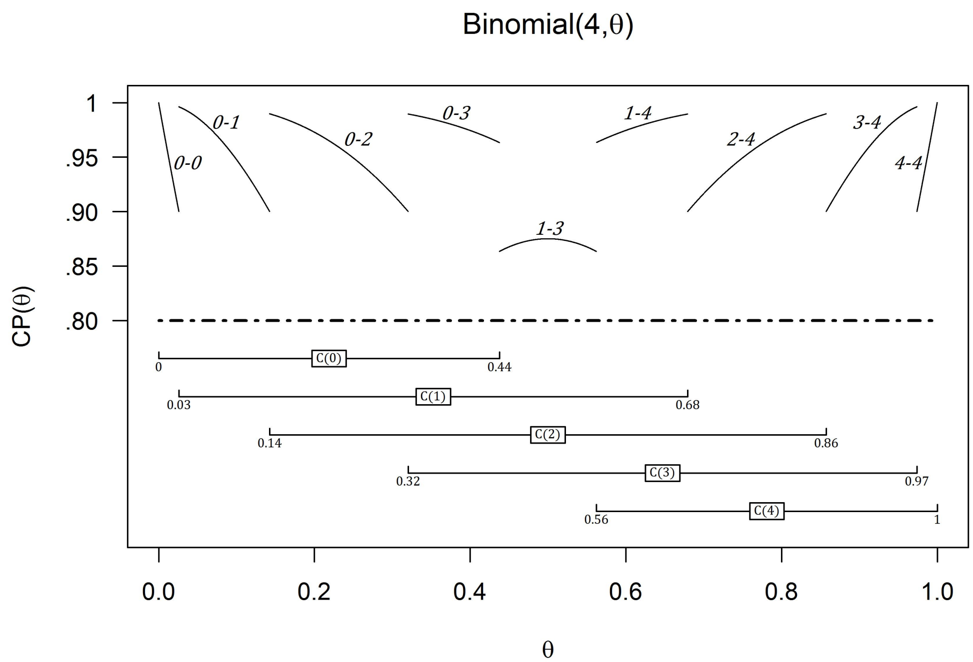

Figure 1.1: CPF of the 80% confidence procedure defined in Table 1.2. Acceptance curves $\mathbb{P}_\theta(a \leq X \leq b)$ are labeled $a - b$.

1.3 Performance of a Confidence Procedure

Suppose a practitioner hands you some observed data and tells you both the distribution from which it came and their preferred $(1-\alpha)$ confidence level. You in turn hand the practitioner back a confidence interval. What does the confidence level of this interval really mean? First note that one cannot talk about the confidence level of a single interval; a given confidence level must be associated with an entire confidence procedure. A $(1-\alpha)$ level of confidence implies that in repeated use of the confidence procedure, θ will be captured in the resulting confidence intervals at least $100(1-\alpha)\%$ of the time. Furthermore, since the user is attempting to estimate θ as accurately as possible we should supply them with the narrowest possible interval so that they can hone in on the true value of θ. Thus, although there are many properties to consider when choosing a confidence procedure, the two overriding conditions should be valid coverage and minimal length.

Property 1.1 (Valid Coverage) A $(1-\alpha)$ confidence procedure has *valid (or strict) coverage* if $CP(\theta) \geq (1-\alpha),\ \forall \theta \in \Theta$.

From among valid procedures, a procedure with intervals of the shortest possible length is desired. For distributions with finite support such as the binomial distribution one may compare two confidence procedures by their average interval length,

$$\frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} |u(x) - l(x)|. \tag{1.6}$$

However, for distributions with infinite support like the Poisson or negative binomial, confidence interval length is difficult to assess since such sums are infinite; moreover, the confidence intervals of one procedure may be shorter for some x's but longer for others.

Schilling & Holladay (2017)[22] suggest a different criterion for comparing the lengths of two procedures, as given below:

Schilling and Holladay Criterion: A strict confidence procedure $C'(x) = [l'(x), u'(x)]$ is *superior* to $C(x) = [l(x), u(x)]$ on length if there exist integer K_0 such that,

$$\sum_{|x|<K} |u'(x) - l'(x)| \leq \sum_{|x|<K} |u(x) - l(x)|,\ \forall K, \tag{1.7}$$

with

$$\sum_{|x|<K} |u'(x) - l'(x)| < \sum_{|x|<K} |u(x) - l(x)|,\ \forall K > K_0. \tag{1.8}$$

Thus C' is "superior" to C on length if it does no worse on length for initial intervals and eventually has smaller interval length sums.

It may also be useful to compare the expected length curves, $EL : \Theta \to \mathbb{R}^+$ defined by,

$$EL(\theta) \coloneqq \mathbb{E}_\theta[u(X) - l(X)]. \tag{1.9}$$

In addition to the length criteria mentioned above, there are also a of couple desirable length properties to consider. The first is what Byrne & Kabaila (2005) [17] call "inability to be shortened" property:

Property 1.2 (Inability to be Shortened) A strict $(1-\alpha)$ 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 θ.

Small acceptance regions might also be desirable:

Property 1.3 (Minimal Cardinality) A strict confidence procedure has the *minimal cardinality property* if its acceptance regions are of minimal size i.e. if decreasing the cardinality of any acceptance region A_θ, forces the CPF to drop below the confidence level.

Because we only consider acceptance regions composed of contiguous values of x, an acceptance region of minimal cardinality is said to be of *minimal span.* Note that since there may exists two different acceptance regions A'_θ and A_θ for θ of equal span, minimal cardinality procedures are not unique in general.

Schilling & Holladay (2017)[22] showed that all minimal cardinality procedures possess the inability to be shortened property.

Proposition 1.1 If a strict confidence procedure has the minimal cardinality property then it has the inability to be shortened property.

While valid coverage and short intervals should be the primary goals of any confidence procedure there are several other important attributes of a confidence procedure. For instance, other desirable criteria may include (i) intervals that are gapless, (ii) that are nested, or (iii) that have strictly increasing endpoints, which are discussed below.

Property 1.4 (Gapless) A confidence procedure is *gapless* if all confidence sets consists of a single continuous interval.

For instance, a gapless procedure will avoid confidence sets of the form $[a, b] \cup [c, d]$ with gap $[b, c]$, where $a < b < c < d$. Practitioners will clearly find it confusing to interpret a supposed confidence "interval" that is a union of two or more disjoint intervals since it has no true upper (or lower) bound. Thus, it may be desirable to avoid methods that produce intervals with gaps.

Property 1.5 (Strictly Increasing Endpoints) A confidence procedure $C(x) = [l(x), u(x)]$ has *strictly increasing interval endpoints* if $l(x) < l(x+1)$ and $u(x) < u(x+1)$ for all x.

For many distributions, including the binomial and Poisson, the point estimates $\hat{\theta}$ for θ increase with x and thus it makes sense to demand this property for the interval endpoints as well. In the binomial and Poisson cases in particular it can be shown that any strict procedure possessing endpoints that decrease can gain an increase in coverage, without loss in net interval length, by rearranging interval endpoints in proper order (Casella & Robert (1989)[13]). Hence, procedures with such misordered endpoints are inadmissible. In the NB case the maximum likelihood estimate (MLE) for θ decreases for increasing x and thus strictly decreasing endpoints might instead be desirable.

Property 1.6 (Nesting) A method for constructing confidence procedures is said to be *nested* if: $\alpha < \alpha'$ implies that intervals included in the corresponding $(1 - \alpha')$ procedure are completely contained in the $(1 - \alpha)$ procedure.

Nesting is desirable since it requires that any test that rejects $H_0 : \theta = \theta_0$ in favor of $H_a : \theta \neq \theta_0$ at level α will also reject at level $\alpha' > \alpha$ (here "rejection" occurs if the generated confidence interval does not contain θ_0). The last property considered is based on ideas of invariance, and is relevant specifically to the binomial distribution.

Property 1.7 (Equivariance for Binomial Procedures) A binomial confidence procedure $C(x) = [l(x), u(x)]$ is *equivariant* if $n-x$ generates the interval $[1-u(x), 1-l(x)]$.

A binomial confidence procedure will be equivariant if and only if it has symmetric CPF (Schilling & Doi (2014)[23]). Clearly if $[l(x), u(x)]$ is the confidence interval for θ then $[1-u(x), 1-l(x)]$ should yield a confidence interval for $1-\theta$. Thus, it is important for a binomial confidence procedure to be equivariant since switching the roles of 'success' and 'failure' switches the roles of x and $n-x$.

One might also require that a binomial procedure be monotone in n (see for instance Blyth and Still (1983, property 4)[12]. Or monotone in length; that is, length increases from $x = 0$ up to $x = n/2$ and consequently decreases there after due to the equivariance property (see Casella (1986)[6]).

1.4 Some Large Sample Methods

Numerous large sample methods have been proposed to produce some moderate-performing confidence intervals, but only a few well-known methods will be introduced here.

1.4.1 Wald's Method (W)

First the *Wald* (W) method is considered because it is one of the prevailing methods presented in elementary statistics textbooks. It is based on the large-sample theory for the MLE's $\hat{\theta}$ for θ which gives us,

$$\sqrt{I(\theta)}\left(\hat{\theta}-\theta\right) \stackrel{\rightarrow}{\sim} \mathrm{N}(0,1), \tag{1.10}$$

where $I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log p(X|\theta)\right)^2\right]$ is the Fisher information in X. This approximation is satisfactory, for instance, for sufficiently large n (and θ away from 0 and 1) in the binomial case and sufficiently large θ in the Poisson case.

The Fisher information will be a function of θ and is typically approximated by the observed information,

$$-\frac{\partial^2}{\partial\theta^2}\log p(x|\theta)\Big|_{\theta=\hat{\theta}}, \tag{1.11}$$

or by the expected information,

$$I(\theta)\Big|_{\theta=\hat{\theta}}. \tag{1.12}$$

Substituting the observed or expected information in (1.10) preserves the asymptotic normality because of Slutsky's theorem. The two approximations for $I(\theta)$ will often yield similar (or even equivalent; e.g. in the binomial, Poisson, and NB cases) results. Denoting either approximation of $I(\theta)$ by $\hat{I}(\theta)$ the approximate $(1-\alpha)$ W confidence set becomes,

$$\left\{\theta : \sqrt{\hat{I}(\theta)}\left|\hat{\theta}-\theta\right| \le z_{\alpha/2}\right\} = \left\{\theta : \hat{\theta} - z_{\alpha/2}\frac{1}{\sqrt{\hat{I}(\theta)}} \le \theta \le \hat{\theta} + z_{\alpha/2}\frac{1}{\sqrt{\hat{I}(\theta)}}\right\}, \tag{1.13}$$

where $z_{\alpha/2}$ is the $\left(1-\frac{\alpha}{2}\right)$th quantile of the Standard Normal distribution. This confidence set will be an interval with,

$$\begin{aligned} l(x) &= \hat{\theta} - z_{\alpha/2}\tfrac{1}{\sqrt{\hat{I}(\theta)}}, \\ u(x) &= \hat{\theta} + z_{\alpha/2}\tfrac{1}{\sqrt{\hat{I}(\theta)}}. \end{aligned} \tag{1.14}$$

Note that some textbooks present the Wald test statistic in the form

$$\frac{\hat{\theta}-\theta}{\sqrt{\mathbb{V}ar_\theta(\hat{\theta})}}, \tag{1.15}$$

since $\mathbb{V}ar_\theta(\hat{\theta}) = [I(\theta)]^{-1}$ for some common distributions including the binomial and Poisson distributions. Table 1.3 shows the resulting $(1-\alpha)$ Wald confidence procedures for the binomial, Poisson and NB cases.

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

Table 1.3: Wald Confidence Procedures.

Figure 1.2 shows the 95% CPF's for the Wald method for the binomial$(20,\theta)$ and Poisson(θ) distributions. The 95% W confidence procedures have coverage way below the nominal level. Although its performance in the binomial case improves for increased n and for θ away from the boundaries, this leaves the cases when n is small unaddressed. Different rules of thumb are often presented for using W in the binomial case, but they all reflect the fact that the Wald method has poor coverage when θ is near the boundaries, or when n is small. However, W performs poorly even in cases where such constraints are comfortably satisfied. See Brown, Cai & DisGupta (2001)[4] for a thorough analysis. In the Poisson case the CPF remains almost completely below the nominal level even for large θ, however, the amount at which it drops below decreases as θ increases.

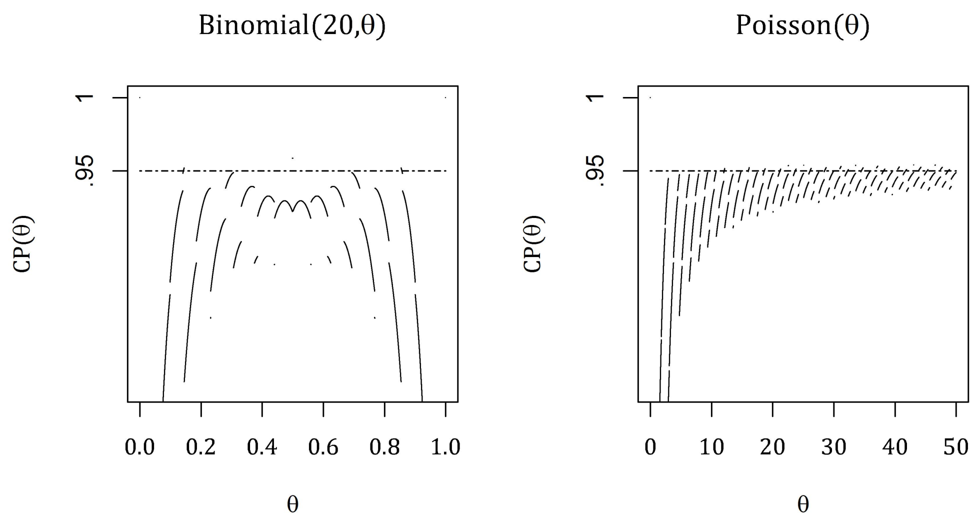

Figure 1.2: 95% Wald CPF's for the binomial$(20, \theta)$ and Poisson(θ) distributions.

1.4.2 Rao's Score Method (S)

Rao's confidence procedure (S) is based on the standardized "score" statistic,

$$Q(\theta|x) = \frac{\frac{\partial}{\partial\theta} \ln p(x|\theta)}{\sqrt{-\mathbb{E}_{\theta}\left[\frac{\partial^2}{\partial\theta^2} \ln p(x|\theta)\right]}}, \tag{1.16}$$

which has an asymptotic standard normal distribution (see for instance Casella & Berger (2002, p. 494)[15]). Note that under certain regularity conditions the denominator of Q will be equivalent to the square root of Fisher Information in X; in particular, this holds when $\{p(x|\theta)\}$ is an exponential family member. A $(1-\alpha)$ confidence set based on Q is given by,

$$\left\{\theta : |Q(\theta|x)| \leq z_{\alpha/2}\right\} \tag{1.17}$$

In certain special cases (e.g. binomial and Poisson) the Rao-Score confidence set in (1.17) reduces to,

$$\left\{\theta : \left|\sqrt{I(\theta)}\left(\hat{\theta} - \theta\right)\right| \leq z_{\alpha/2}\right\}, \tag{1.18}$$

which interestingly enough is the same as the Wald confidence set except that in this case $I(\theta)$ is not estimated. Table 1.4 shows the resulting $(1-\alpha)$ Rao-Score confidence procedures for the binomial, Poisson and NB distributions.

Distribution	Rao-Score Interval
Binomial(n,θ)	$\frac{1}{n+z^2_{\alpha/2}}\left[x+\frac{1}{2}z^2_{\alpha/2}\mp z_{\alpha/2}\sqrt{\frac{1}{n}x\,(n-x)+\frac{1}{4}z^2_{\alpha/2}}\right]$
Poisson(θ)	$x+\frac{1}{2}z^2_{\alpha/2}\mp z_{\alpha/2}\sqrt{x+\frac{1}{4}z^2_{\alpha/2}}$
NB(r,θ)	$\frac{1}{2r}\left[2r\hat{\theta}-z^2_{\alpha/2}\hat{\theta}^2\mp z_{\alpha/2}\hat{\theta}\sqrt{z^2_{\alpha/2}\hat{\theta}^2+4r(1-\hat{\theta})}\right]$

Table 1.4: Rao-Score Confidence Procedures.

Figure 1.3 shows the 95% CPF's for S for the binomial$(20,\theta)$ and Poisson(θ) distributions. For each distribution the coverage seems to almost average 0.95 over the parameter space.

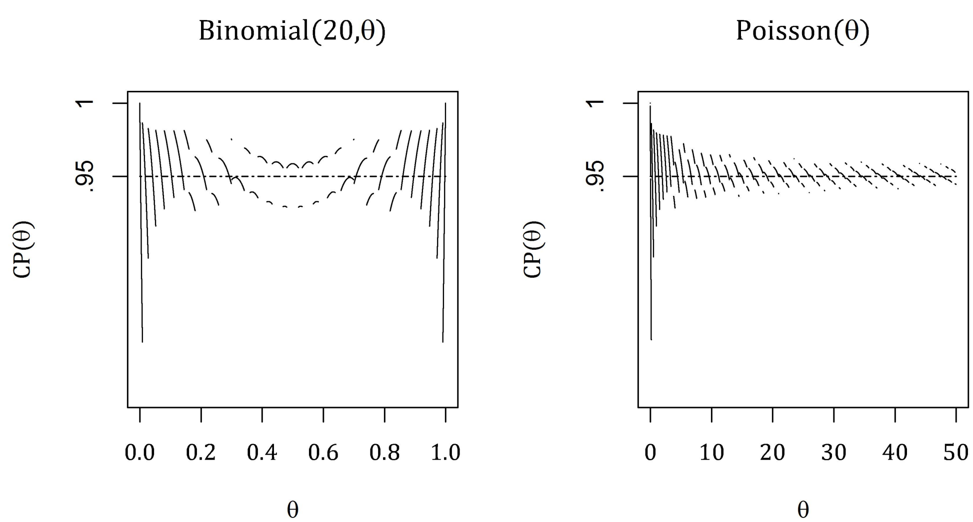

Figure 1.3: 95% Rao-Score CPF's for the binomial$(20,\theta)$ and Poisson(θ) distributions.

1.4.3 Wilks' Likelihood Ratio Method (LR)

The last large sample method considered is based on the large-sample theory of the likelihood ratio test. If $L(\theta)$ represents the likelihood function then the likelihood ratio test rejects $H_0 : \theta = \theta_0$ in favor of $H_a : \theta \neq \theta_0$ if $L(\theta_0)/L(\hat{\theta})$ is small or equivalently if $-2\ln\{L(\theta_0)/L(\hat{\theta})\} = 2[\ln L(\hat{\theta}) - \ln L(\theta_0)]$ is large. $\hat{\theta}$ here is again the MLE of θ. By Wilks' theorem

$$2[\ln L(\hat{\theta}) - \ln L(\theta)] \approx \chi^2_1. \quad (1.19)$$

Therefore an approximate confidence set for θ is given by,

$$\left\{\theta : 2[\ln L(\hat{\theta}) - \ln L(\theta)] \leq \chi^2_{1,1-\alpha}\right\} \quad (1.20)$$

where $\chi^2_{1,1-\alpha}$ is the $(1-\alpha)$th quantile of the chi-square distribution with 1 degree of freedom. Figure 1.4 shows the 95% CPF's for Wilks' Likelihood Ratio procedure (LR) for the binomial$(20, \theta)$ and Poisson(θ) distributions. The coverage of LR is similar to that of S.

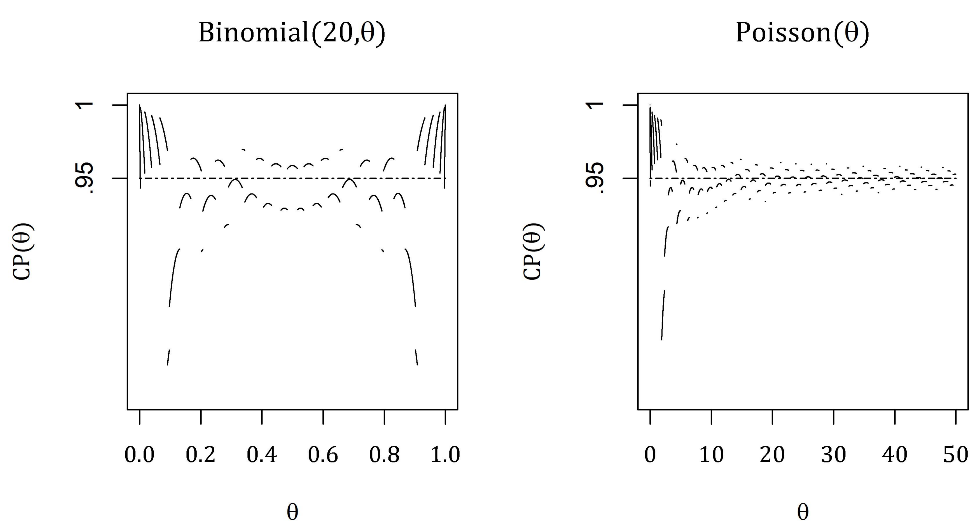

Figure 1.4: 95% LR CPF's for the binomial$(20, \theta)$ and Poisson(θ) distributions.

Many other approximate methods have been developed for large samples, some derived from other preexisting methods. Several are derived from the Wald and Score methods; for example, the so called adjusted Wald and continuity-corrected Score methods, just to name a few. See Agresti & Coull (1998)[1] or Byrne & Kabaila (2005) [17] for extensive discussion of such procedures in the binomial and Poisson cases respectively. Most of the proposed approximate methods aren't strict procedures. The ones that are strict tend to be too conservative and do not satisfy the inability to be shortened property. Thus, they are length inadmissible as they produce intervals that can be shortened without causing their CPF's to drop below the confidence level.

Chapter 2

Methods with Strict Coverage

Since the approximate (i.e. large sample) methods that are discussed in Section 1.4 which are commonly used, tend to perform poorly in most cases, this chapter introduces some strict methods more suitable for accurate estimation particularly when sample size is small. Because of the one-to-one correspondence of a confidence procedure with the acceptance region of a test, construction of an strict confidence procedure can be achieved through strategic choice of acceptance regions.

2.1 Clopper and Pearson's Method (CP)

One of first methods with valid coverage probability proposed for a discrete distribution is due to Clopper & Pearson (1934)[7] for the binomial problem. The analogs for the Poisson and NB cases were later developed by Garwood (1936) [11] and Casella & McCulloch (1984) [14] respectively. For a fixed θ, the idea of Clopper & Pearson's (CP) method is to fit as many x's into the rejection region tails as possible without allowing the probability of either tail to exceed $\alpha/2$. The x's that could not fit into either tail make up the acceptance region A_θ. The CP acceptance region selection process is illustrated diagrammatically in Figure 2.1 using the binomial support as an example.

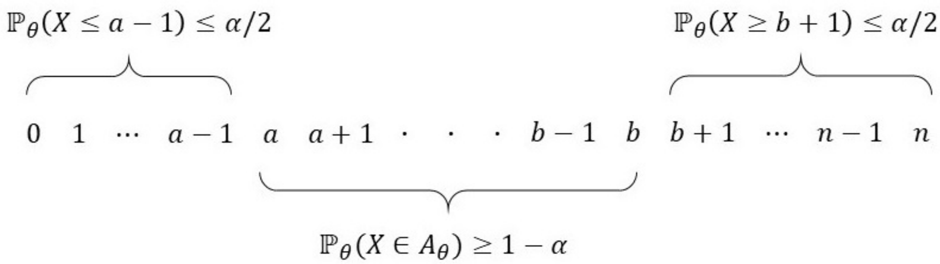

Figure 2.1: Illustratration of how CP's acceptance regions are determined.

In general it will not be possible to choose boundaries between the rejection tails and acceptance region such that the tail probabilities are each exactly equal to $\alpha/2$. However, it will be possible to choose an a and b so that,

$$\mathbb{P}_\theta(X \leq a-1) \leq \alpha/2 < \mathbb{P}_\theta(X \leq a),$$
$$\mathbb{P}_\theta(X \geq b+1) \leq \alpha/2 < \mathbb{P}_\theta(X \geq b).$$

Thus, the tail probabilities will each be as close to $\alpha/2$ as possible without exceeding it. Then the probability of the acceptance region will satisfy $P_\theta(a \leq X \leq b) \geq (1-\alpha)$, meeting the strict coverage requirement. For any given/fixed x, the boundaries or cutoff points for the values of θ that do not include x in their corresponding acceptance regions can be found by solving for θ in each of the equations $\mathbb{P}_\theta(X \geq x) = \alpha/2$ and $\mathbb{P}_\theta(X \leq x) = \alpha/2$. Such boundaries or cutoffs give the lower and upper bounds, $l(x)$ and $u(x)$ respectively. Many times $l(x)$ and $u(x)$ can be nicely expressed in terms of the quantiles of a continuous distribution. To see this observe the following identity

$$\mathbb{P}_\theta(X \geq x) = \sum_{k=x}^{\infty} \mathbb{P}_\theta(X = k) = \int_{\theta_{min}}^{\theta} \frac{\partial}{\partial t} \sum_{k=x}^{\infty} \mathbb{P}_t(X = k)dt, \tag{2.1}$$

where θ_{min} is the smallest possible value of θ. When the function

$$f(t) := \frac{\partial}{\partial t} \sum_{k=x}^{\infty} \mathbb{P}_t(X = k), \tag{2.2}$$

represents the density of a continuous random variable, then the integral in (2.1) corresponds to the cumulative distribution of such a continuous random variable. In the binomial, Poisson and NB cases the above quantity (2.2) has relation to the Beta, Chi-Square, and F distributions respectively (see for instance Casella & McCulloch (1984)[14] for details) and thus $l(x)$ and $u(x)$ can be nicely expressed in terms of the quantiles of these continuous distributions. For example, if X is binomial(n, θ) or Poisson(θ) then

$$\mathbb{P}_\theta(X \geq x) = \mathbb{P}\left\{B(x, n-x+1) \leq \theta\right\},$$

and

$$\mathbb{P}_\theta(X \geq x) = \mathbb{P}\left\{\chi^2_{2x} \leq 2\theta\right\}$$

respectively. The intervals for the $(1-\alpha)$ CP procedures for the Binomal, Poisson, and NB distributions are in Table 2.1.

Distribution	$l(x)$	$u(x)$
Binomial(n, θ)	$B_{\alpha/2}(x, n-x+1)$	$B_{1-\alpha/2}(x+1, n-x)$
Poisson(θ)	$\frac{1}{2}\chi^2_{2x,\alpha/2}$	$\frac{1}{2}\chi^2_{2(x+1),1-\alpha/2}$
NB(r, θ)	$\frac{1}{1+\frac{x+1}{r}F^{2(x+1)}_{2r,1-\alpha/2}}$	$\frac{\frac{r}{x}F^{2r}_{2x,1-\alpha/2}}{1+\frac{r}{x}F^{2r}_{2x,1-\alpha/2}}$

Table 2.1: CP Confidence Procedures

The plots of the 95% CPFs of CP for the binomial$(20, \theta)$ and Poisson(θ) distributions are shown in Figure 2.2. In general, the CPF for CP will be strict but quite conservative. Moreover, the CP intervals are often (e.g. in the binomial and Poisson cases) nested, gapless, and possess lower and upper limits that are monotonic in x. However, CP will not satisfy the minimal cardinality property in general. This results in the method producing very wide intervals. CP imposes the restriction that the rejection tail probabilities of the procedure be no more than $\alpha/2$; thus, improved intervals can be obtained by replacing this condition by the weaker condition that the total rejection region probability not exceed α. In what follows, are methods with asymmetric rejection region tail probabilities that potentially achieve better performance than CP.

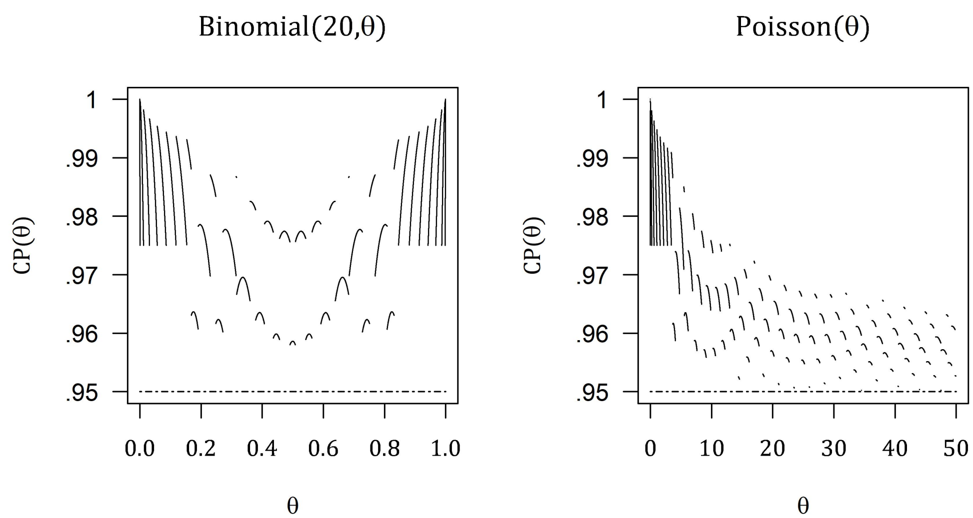

Figure 2.2: 95% CPF's of the Clopper & Pearson method for the binomial$(20, \theta)$ and Poisson(θ) distributions.

2.2 Blaker's Method (B)

One approach to constructing a confidence procedure with potentially asymmetric rejection region tail probabilities is due to Blaker (2000)[16]. Define the *min-tail* probability of an observed value x to be the minimum of $\mathbb{P}_\theta(X \leq x)$ and $\mathbb{P}_\theta(X \geq x)$. Then for a fixed θ, the Blaker method (B) includes x in A_θ if the probability of observing something with a min-tail probability as small as that of x exceeds α. In other words, given θ and x determine all values that have min-tail probabilities less than or equal to the min-tail probability of x (x is one of these values). If the sum of the probabilities for these values is greater than α then include x in A_θ. The 95% CPF's for Blaker's method for the binomial$(20,\theta)$ and Poisson(θ) distributions are shown in Figure 2.3.

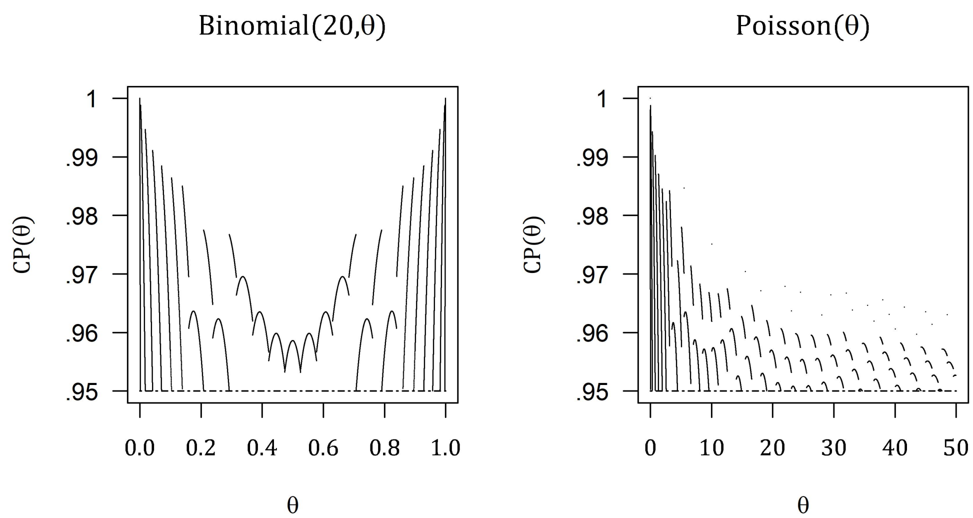

Figure 2.3: 95% Blaker CPF's for the binomial$(20,\theta)$ and Poisson(θ) distributions.

Blaker (2000) showed B is strict and produces nested intervals. Additionally, Blaker's method tends to perform quite well with regard to length and coverage. However, B does not possess the minimal cardinality property as its CPF consistently uses acceptance curves which are not of minimal span. Thus, the procedure is length inadmissible as it can be trivially shortened while maintaining valid coverage. We note however, that although the CPF for Blaker's method regularly uses acceptance regions that are not of minimal span, the regions where this occurs are typically relatively small; hence B will perform nearly as well as a minimal cardinality procedure on length.

2.3 Sterne's Method (ST) and Gaps

Another approach to constructing acceptance regions is due to Sterne (1954) [27]. The idea of Sterne's method (ST) is to choose the members of the acceptance region A_θ by entering the most probable x's one-by-one, until $\mathbb{P}_\theta(X \in A_\theta)$ exceeds $(1-\alpha)$. This construction gets coverage above the confidence level with the fewest number of x's possible in the acceptance regions and thus ST has the minimal cardinality property. However, since it may be possible to construct acceptance regions of the same cardinality as those of Sterne's method, but with smaller probability, ST will not be the only minimal cardinality procedure. It will however, be the minimal span procedure with maximal coverage. Figure 2.4 shows the 95% CPF's for Sterne's method for the binomial$(20, \theta)$ and Poisson(θ) distributions.

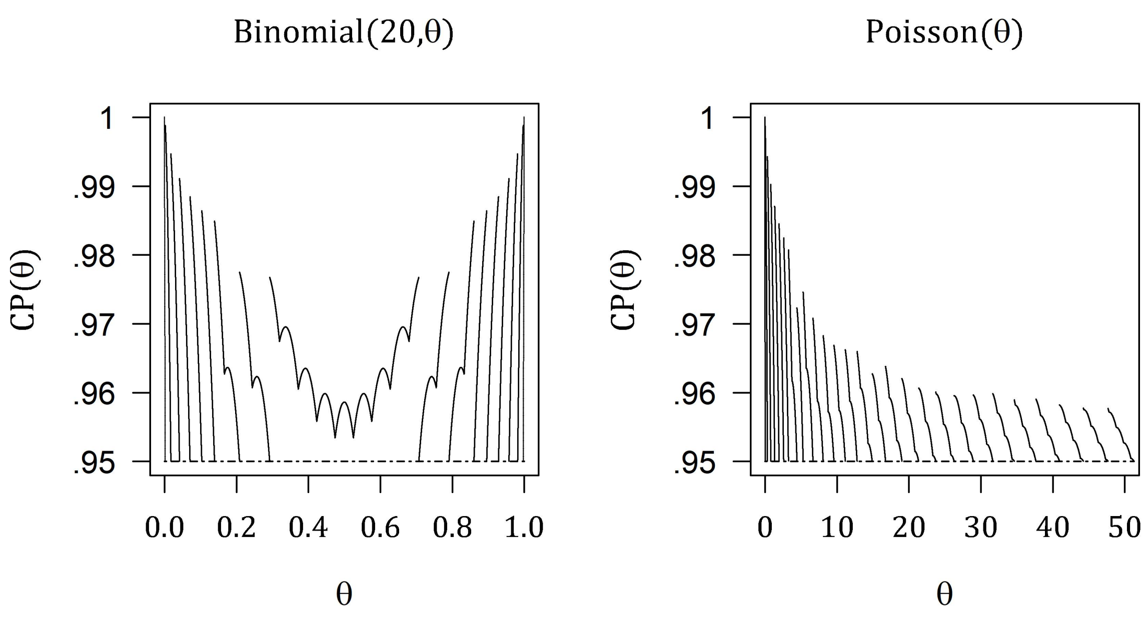

Figure 2.4: 95% Sterne CPF's for the binomial$(20, \theta)$ and Poisson(θ) distributions.

Although ST is strict and satisfies the minimal cardinality property, it unfortunately produces confidence intervals with gaps in the binomial, Poisson, and NB cases. In general gaps occur in the confidence interval for a particular x if the CPF of the procedure uses acceptance curves involving x on disjoint intervals of θ. Recall that since we restrict ourselves to acceptance regions of contiguous values of x, all acceptance curves will be of the form $\mathbb{P}_\theta(a \leq X \leq b)$ where, $a = \inf\{x \in A_\theta\}$ and $b = \sup\{x \in A_\theta\}$. Thus, the sequences of $\{a\}$ and $\{b\}$ values obtained as θ increases must be monotonic in order to obtain gapless intervals.

Figure 2.5 illustrates some specific instances where the CPF's for Sterne's method produces gaps for binomial and Poisson cases. For convenience denote a typical acceptance curve $\mathbb{P}_\theta(a \leq X \leq b)$ by $a - b$. In the binomial case Sterne's 90% procedure uses two acceptance curves $0-4$ and $0-5$, which both include $x = 0$, on disjoint intervals of θ. Thus, the values of θ between these two intervals, where the CPF uses $1-5$, is precisely the location of a gap for the confidence interval for $x = 0$.

One solution to the gap problem is to simply fill in the gap. This corresponds to the CPF transitioning directly from $0-4$ to $0-5$ and thus avoids using $1-5$ at all. However this will have effectively increased the overall interval length of the procedure by the length of the gap. A second option is to just remove the extra interval causing the gap as it is usually relatively small. However, in this case $x = 0$ will no longer be in acceptance regions of the θ's in the removed interval. Thus, the CPF will be using $1-5$, in place of $0-5$. But $1-5$ is of cardinality one less than $0-5$. Hence, this will cause the CPF to drop below the confidence level since Sterne's method has the minimal cardinality property.

Schilling & Doi (2014)[23] suggested a way to relocate the second interval of the confidence set that will preserve both Sterne's strict coverage probability and net interval length. The inset in Figure 2.5 (binomial case) shows how this may be done from perspective of the CPF. By using the dotted line which is a continuation of $1-6$ we can avoid using $0 - 5$. In this way the extra piece from the interval for $x = 0$ has been removed and merged with the interval for $x = 6$. This is good method for resolving the gap since the total length of the procedure remains unchanged and there is no violation in strictness. The corresponding intervals produced from the three approaches discussed above are provided in Table 2.2. Only the intervals for $x = 0$ and 6 are affected; all other intervals remain unchanged. What is ultimately gained from Schilling & Doi's (2014)[23] relocation approach is a strict confidence procedure which contains no gaps and satisfies the minimal cardinality property. And what is lost is a minuscule amount of coverage since for instance, in the above example, $0 - 5$ lies above $1 - 6$ in the small interval in which $0 - 5$ is swapped out for $1 - 6$. A similar problem occurs in the 95% Poisson case for the interval of $x = 23$. The inset of Figure 2.5 (Poisson case) shows how the problem may be remedied and Table 2.2 shows the resulting intervals

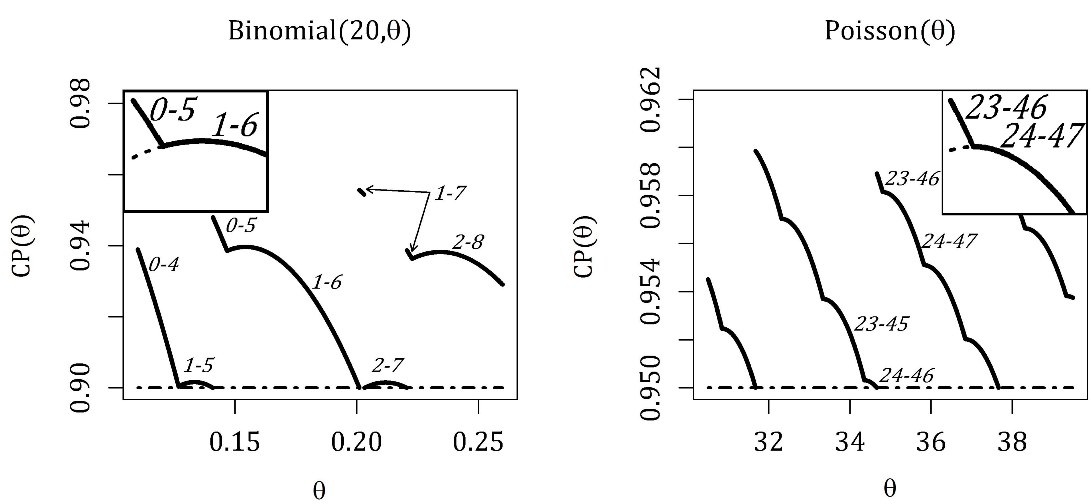

Figure 2.5: Illustration of an occurrence in Sterne's 90% binomial and 95% Poisson CPF's that produces gaps in the corresponding procedures. Curves $P_\theta(a \leq X \leq b)$ are labeled 'a-b.' Insets show how the problem may be corrected.

Solution	Binomial$(20,\theta)$		Poisson(θ)	
	$x=0$	$x=6$	$x=23$	$x=47$
Original	$[0,0.13)\cup[0.14,0.15)$	$[0.15,0.45)$	$[14.92,34.36)\cup[34.67,34.81)$	$[34.81,62.36)$
Fill	$[0,0.13,0.15)$	$[0.15,0.45)$	$[14.92,34.81)$	$[34.81,62.36)$
Remove	$[0,0.13)$	$[0.15,0.45)$	$[14.92,34.36)$	$[34.81,62.36)$
Relocate	$[0,0.13)$	$[0.14,0.45)$	$[14.92,34.36)$	$[34.67,62.36)$

Table 2.2: Gap Table

In order to prevent gaps in the confidence interval for a particular x, use of acceptance curves, that involve x, on disjoint intervals of θ need to be avoided. In order to achieve this, the sequences of $\{a\}$ and $\{b\}$ values of the acceptance curves $\{P_\theta(a \leq X \leq b)\}$ must be monotonic. Whenever, a procedure does produce gaps, analogous modifications to the ones discussed above can be made to resolve the gap. Call the procedure that modifies ST via Schilling & Doi's (2014)[23] approach, the Modified Sterne's Method (MST). MST in the Poisson case can summarized as follows:

(1) For each θ, choose the highest acceptance curve among those of minimal span, except:

(2) Whenever step (1) gives a curve $P_\theta(a \leq X \leq b)$ that results in a decrease in the $\{a\}$ sequence of the procedure, substitute for $P_\theta(a \leq X \leq b)$ the curve $P_\theta(a+1 \leq X \leq b+1)$.

In the Binomial case the procedure works a little differently due to the equivariance property:

(1) For $\theta \leq 0.5$ [$\theta > 0.5$], moving from left to right [right to left] choose the highest acceptance curve among those of minimal span, except:

(2) Whenever step (1) gives a curve $P_\theta(a \leq X \leq b)$ that results in a decrease [increase] in the $\{a\}$ [$\{b\}$] sequence of the procedure, substitute for $P_\theta(a \leq X \leq b)$ the curve $P_\theta(a+1 \leq X \leq b+1)$ [$P_\theta(a-1 \leq X \leq b-1)$].

Note that the actual number of occurrences of gaps, for a particular procedure, has been shown to be relatively small in binomial and Poisson cases (Schilling & Doi (2014) [23], Holladay (2014) [3]). Thus, MST will be nearly identical to ST except in those few locations where gaps occur.

2.4 Interplay Between CPF and Confidence Procedure

ST and B were defined in terms of their acceptances regions. Instead, however, they could have been described graphically through their choice of acceptance curves. In this way one can reverse engineer a high-performing procedure by first constructing a valid CPF possessing the properties that one desires. To help understand how the graphical properties of a CPF relate to the resulting confidence procedure's performance see Figure 2.6, which illustrates the interplay between the 95% MST Poisson CPF and the resulting 95% confidence intervals.

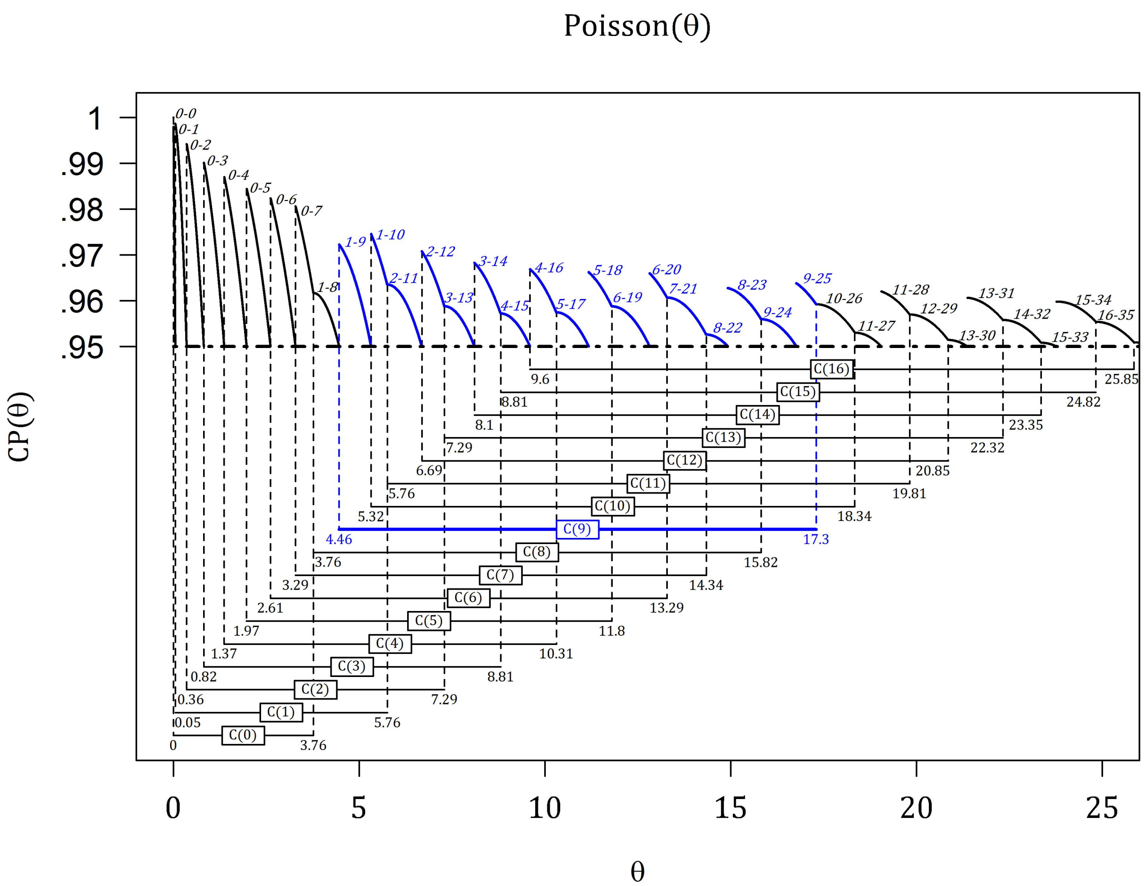

Figure 2.6: Interplay between the 95% MST Poisson CPF and resulting confidence intervals. Acceptance curves $P_\theta(a \le X \le b)$ are labeled as a-b.

Consider, for instance, the confidence interval for $x = 9$ in Figure 2.6. Highlighted in blue are all acceptance curves which involve $x = 9$. The first place the CPF uses such a curve is at $\theta = 4.46$ and the last place is at 17.3. Thus, the confidence interval for $x = 9$ is $[4.46, 17.3)$. This follows from the fact that $x \in A_\theta$ if and only if $\theta \in C(x)$. Repeating these steps for each x, one can always work backwards to obtain the entire family of confidence intervals produced from a given CPF.

2.5 Building a Minimal Cardinality Procedure

Suppose one would like to build a high-performing $(1-\alpha)$ confidence procedure from scratch through a strategic choice of acceptance curves. For each $\theta \in \Theta$ there is a minimum acceptance region size required for the coverage to exceed $(1-\alpha)$.

Definition 2.1 (Minimal Cardinality) Given a random variable X and a fixed confidence level $(1-\alpha) \in (0,1)$ define the minimal cardinality of θ as,

$$M(\theta) \coloneqq \min_{a,k}\{k : \mathbb{P}_\theta(a \leq X \leq a+k-1) \geq (1-\alpha)\},$$

where the minimum is taken over all $a \in \mathcal{X}, k \in \mathbb{Z}^+$.

Therefore, for each $\theta \in \Theta$ the CPF of a minimal span procedure will necessarily use acceptance curves with cardinality matching the value of $M(\theta)$. Figure 2.7 is a plot of $M(\theta)$ when $(1-\alpha) = 0.95$ for the binomial$(20,\theta)$ and Poisson(θ) distributions. In the binomial case $M(\theta)$ has an almost quadratic shape; whereas, in the Poisson case $M(\theta)$ is nondecreasing.

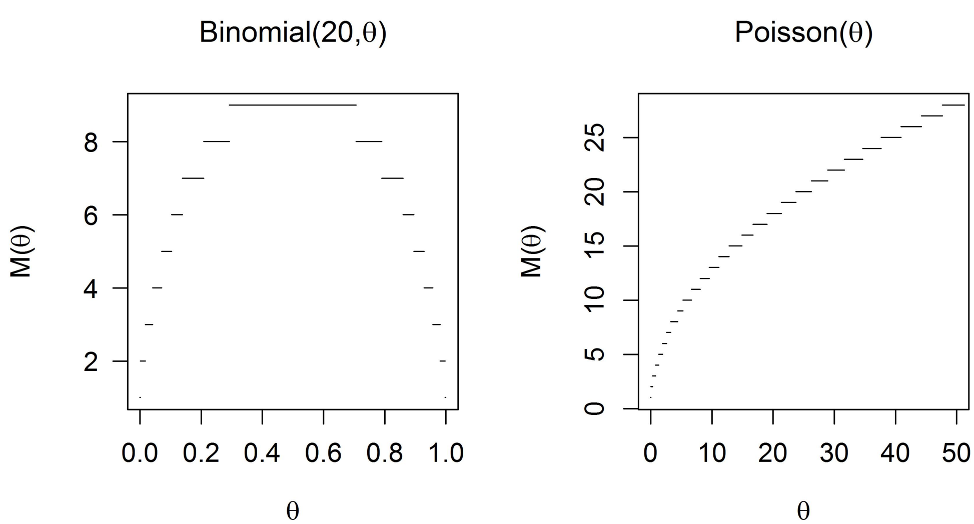

Figure 2.7: Minimal Cardinality of θ when $(1-\alpha) = 0.95$

Figure 2.8 shows the plots of all possible acceptance curves for a 95% gapless minimal cardinality procedure for the binomial$(20, \theta)$ and Poisson(θ) distributions. Building a piecewise function by choosing only from the eligible portions of the curves shown in the plots will automatically result in a 95% strict minimal cardinality procedure. Note that the curves of minimal span that would force either sequence of $\{a\}$ or $\{b\}$ values (in $\{P_\theta(a \le X \le b)\}$ as θ increases) to be nonmonotonic are omitted as use of these curves would cause gaps in the resulting procedures.

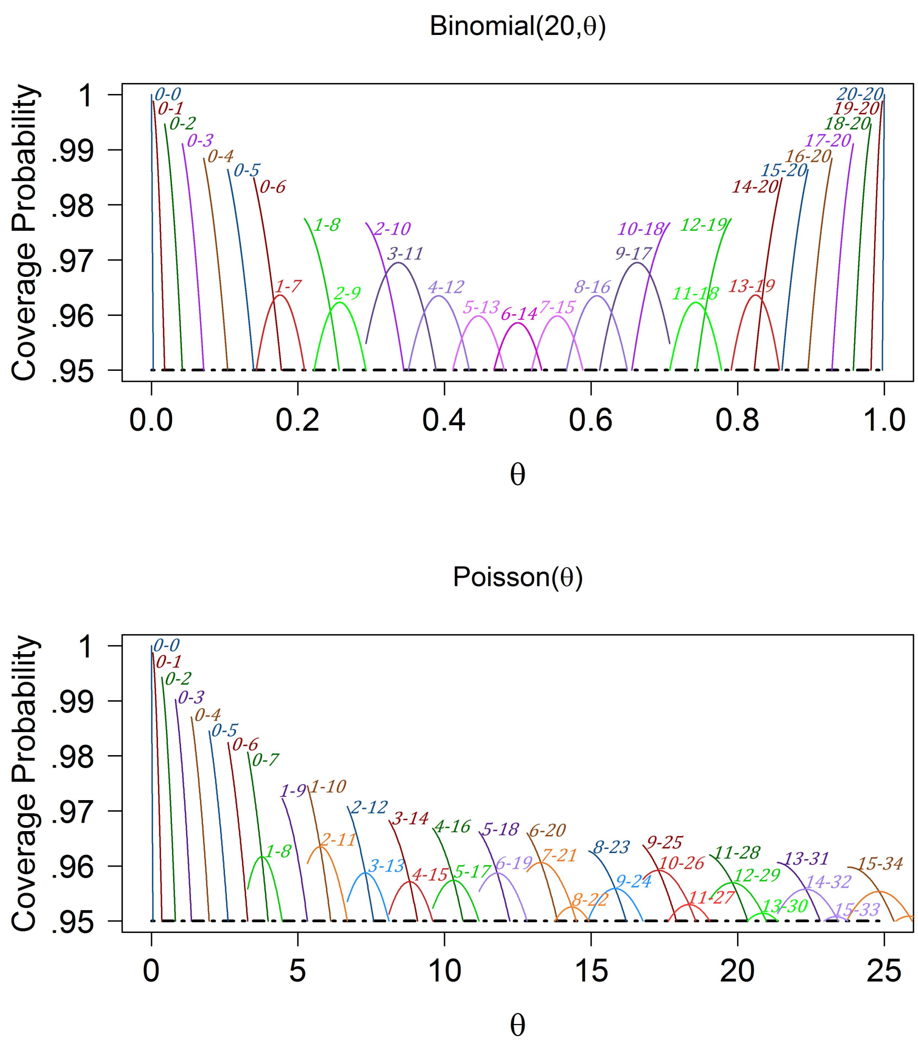

Figure 2.8: All 95% minimal cardinality acceptance curves except those that would cause gaps. We label acceptance curves $P_\theta(a \leq X \leq b)$ by a-b.

2.5.1 Acceptance Curve Behavior

As seen in Figure 2.8 for each θ there may be multiple acceptance curves of minimal cardinality to choose from. As a result, not only are we interested in the general behavior of acceptance curves, but we are interested in the relationship between acceptance curves of equal cardinality. In particular, we would like to investigate the relationship of pairs of acceptance curves of the form: $P_\theta(a \le X \le b)$ and $P_\theta(a+1 \le X \le b+1)$. Propositions 2.1 and 2.2 give several results on the general behavior of acceptance curves for the binomial and Poisson cases respectively. In addition, Figure 2.9 illustrates the general interplay between acceptance curve pairs: $P_\theta(a \le X \le b)$ and $P_\theta(a+1 \le X \le b+1)$.

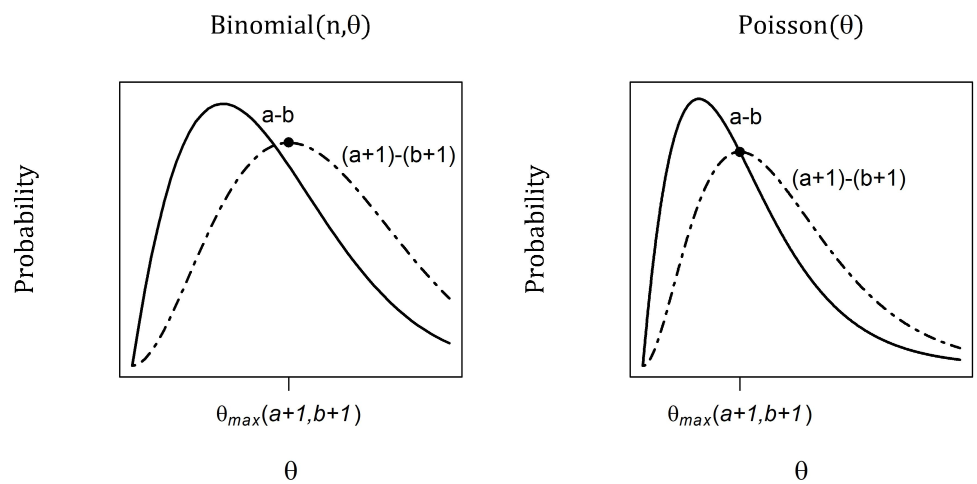

Figure 2.9: General interplay between acceptance curves $P_\theta(a \le X \le b)$ and $P_\theta(a+1 \le X \le b+1)$. We label acceptance curves $P_\theta(a \le X \le b)$ by *a*-*b*.

Proposition 2.1 Binomial Acceptance Curve Properties

(a) $P_\theta(a \le X \le b)$ attains its maximum at $\theta = \theta_{max}(a,b)$, where

$$\theta_{\max}(a,b) = \frac{\left[\frac{n-a+1}{n-b}\binom{n}{a-1}/\binom{n}{b}\right]^{1/(b-a+1)}}{\left[1+\frac{n-a+1}{n-b}\binom{n}{a-1}/\binom{n}{b}\right]^{1/(b-a+1)}}. \tag{2.3}$$

Furthermore $P_\theta(a \le X \le b)$ is strictly increasing on $(0, \theta_{\max}(a,b))$ and strictly decreasing on $(\theta_{\max}(a,b), 1)$

(b) $P_\theta(a \le X \le b)$ intersects $P_\theta(a+1 \le X \le b+1)$ at a single point, $\text{int}(a,b)$ where,

$$\text{int}(a,b) = \left[\frac{\binom{n}{a}/\binom{n}{b+1}}{1+\binom{n}{a}/\binom{n}{b+1}}\right]^{1/(b-a+1)}. \tag{2.4}$$

Furthermore, the location of this intersection occurs to the left (right) of the maximum of $P_\theta(a+1 \le X \le b+1)$ if $\frac{n-a}{n-b-1} > 1$ ($\frac{n-a}{n-b-1} < 1$)

(c) $P_\theta(a \leq X \leq b) \gtrless P_\theta(a+1 \leq X \leq b+1)$ according to $\theta \lessgtr \text{int}(a,b)$

(d) $\theta_{\max}(a,b) < \theta_{\max}(a+1,b+1)$

Proposition 2.2 Poisson Acceptance Curve Properties

(a) $P_\theta(a \leq X \leq b)$ attains its maximum at $\theta = \theta_{\max}(a,b)$, where

$$\theta_{\max}(a,b) = [(a)(a+1)\cdots(b)]^{1/(b-a+1)}. \tag{2.5}$$

Furthermore $P_\theta(a \leq X \leq b)$ is strictly increasing on $(0, \theta_{max}(a,b))$ and strictly decreasing on $(\theta_{max}(a,b), \infty)$

(b) $P_\theta(a \leq X \leq b)$ intersects $P_\theta(a+1 \leq X \leq b+1)$ at a single point which is $\theta_{\max}(a+1,b+1)$.

(c) $P_\theta(a \leq X \leq b) \gtrless P_\theta(a+1 \leq X \leq b+1)$ according to $\theta \lessgtr \theta_{\max}(a+1,b+1)$

(d) $\theta_{\max}(a,b) < \theta_{\max}(a+1,b+1)$

(e) $\max_\theta\{P_\theta(a \leq X \leq b)\} > \max_\theta\{P_\theta(a+1 \leq X \leq b+1)\}$

(f) Suppose a, b, a', b' are non-negative integers such that $a' \leq a$ and $b' \leq b$ where at least one of these inequalities is strict. Then, $P_\theta(a \leq X \leq b) < P_\theta(a' \leq X \leq b')$ for all $\theta \in (0, \infty)$.

The proofs of these propositions follow from elementary calculus and basic algebra. These propositions confirm what can already be seen in the Figure 2.9 about the general behavior and relationship of curves $P_\theta(a \leq X \leq b)$ and $P_\theta(a+1 \leq X \leq b+1)$. The take-away from these propositions is that both the location of the maximum of the acceptance curves and the location of the intersections between acceptance curve pairs $P_\theta(a \leq X \leq b)$ and $P_\theta(a+1 \leq X \leq b+1)$ have nice analytical form. Furthermore the most notable difference between the binomial and Poisson cases is that in the Poisson case the intersection between $P_\theta(a \leq X \leq b)$ and $P_\theta(a+1 \leq X \leq b+1)$ takes place at the maximum of $P_\theta(a+1 \leq X \leq b+1)$, but it occurs to left or the right of the maximum of $P_\theta(a+1 \leq X \leq b+1)$ in the binomial case, depending on the values of a and b under consideration. These propositions become especially useful for writing code for various minimal cardinality procedures and are often used to confirm that an idea which works for one value of n for the binomial(n, θ) distribution will work for all other cases.

2.5.2 Coincidental Endpoints

Recall our goal here is to reverse-engineer a high-performing confidence procedure by constructing a valid CPF. This can be achieved by selecting for each θ one of the eligible acceptance curves in Figure 2.8. As can be seen from Figure 2.8, for some values of θ there is a unique curve of minimal cardinality. For instance near 0 and 1 in the binomial case and near 0 for Poisson case certain acceptance curves must be used for the entire duration for which the cardinality of those curves is minimal. In others cases a curve must be used eventually since it is the only minimal cardinality curve available for a portion of θ's, but there is an entire range of choices to start using the curve.

As a result, it turns out that certain endpoints are fixed and any minimal cardinality procedure must agree on these endpoints, but other endpoints have an entire range of possible values. We call the latter coincidental endpoints. Coincidental endpoints occur whenever the CPF of a strict minimal cardinality procedure transitions (as θ increases) between two acceptance curves of equal cardinality. Shifting a coincidental endpoint within the range of where the two associated curves overlap (and remain above the confidence level) does not effect the net interval length of a confidence procedure. This is because a decrease in the length of one interval will be met with an increase of the same amount in another.

Coincidental endpoints are not that unusual; in fact, in general, the majority of confidence interval endpoints for minimal cardinality procedures are coincidental. The CPF's of any confidence procedure involve only two types of transitions (when reading the graph from left to right):

- Type I: A transition between acceptance curves of different cardinality
- Type II: A transition between two acceptance curves of equal cardinality.

For a minimal cardinality procedure all Type I transitions correspond to a fixed endpoint and all Type II transitions correspond to a coincidental endpoint. To see this consider the binomial$(4, \theta)$ case with $(1-\alpha) = .8$. The plot all possible minimal cardinality acceptance curves (that allow the sequences of a and b values obtained as θ increases to be monotonic) are shown in Figure 2.10.

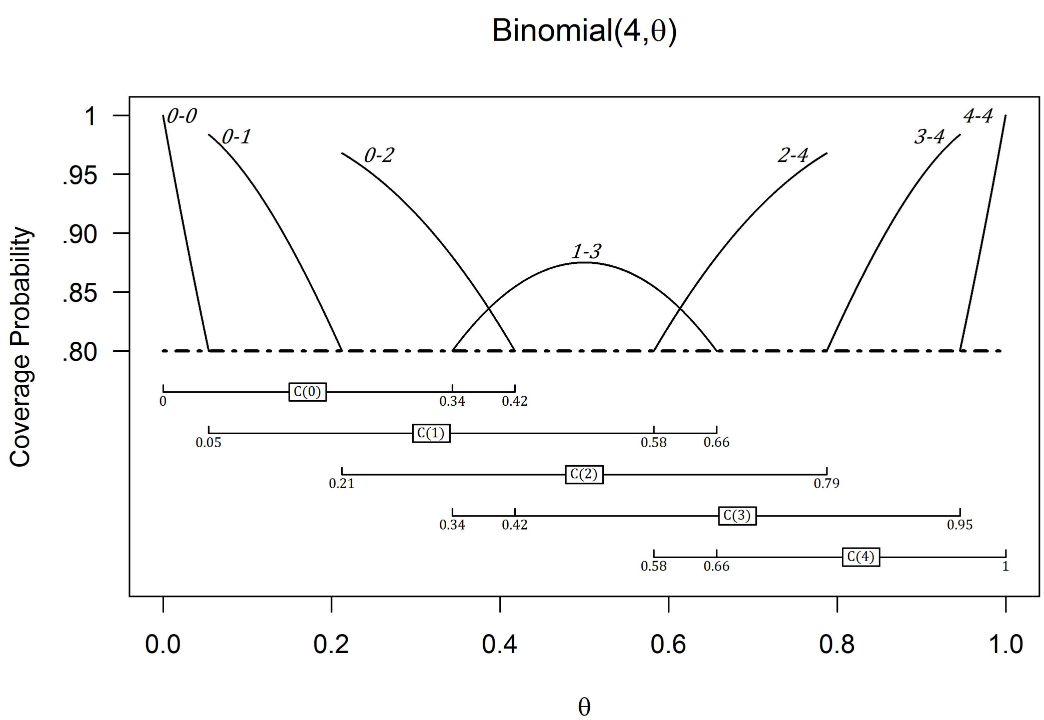

Figure 2.10: Coincidental Endpoint Illustration

Its easy to see that all Type I transitions must occur at the unique locations where the cardinality changes and thus all minimal cardinality procedures agree on the interval endpoints determined by these transitions. The only Type II transitions that can occur are from $P_{0,2}(\theta)$ to $P_{1,3}(\theta)$ and $P_{1,3}(\theta)$ to $P_{2,4}(\theta)$. The location of the two transitions can occur anywhere in [.34,.42] and [.58,.66] respectively. The location of these transitions determine the coincidental endpoints $u(0) = l(3)$ and $u(1) = l(4)$ respectively. This choice does not affect the overall length of the procedure since for example any increase in the upper limit of $x = 0$ will result in an exact increase in the lower limit for $x = 3$. Lastly, we note that we really only have one true choice for coincidental endpoint here because the equivariance requirement forces $u(0) = 1 - l(4)$ and $l(3) = 1 - u(1)$.

Note that when coincidental endpoints occur two confidence intervals $C(x)$ and $C(y)$ share the endpoint $u(x) = l(y)$. Thus, a decision must be made as to which interval ultimately should contain the particular value of θ which defines the coincidental endpoint. Otherwise A_θ will not be of minimal cardinality and a sharp spike in coverage would occur. We prefer the convention that the upper coincidental endpoint be left open while lower endpoint be closed as follows, $[l(x), u(x))$ and $[l(y), u(y))$ (except in the binomial case the roles are reversed when $x > n/2$ due to equivariance). Note that the choice to use half open intervals will not have any affect on the length properties of the procedures.

Since all Type I transitions correspond to fixed endpoints, all minimal cardinality procedures are completely distinguishable only through the location of Type II transitions; i.e. through their choice of coincidental endpoints. Thus, all known minimal cardinality procedures from the literature can be summarized through their coincidental endpoint choices. Note however, that the notions of minimal cardinality and coincidental endpoints were not exactly what all the authors of these procedures proposed.

MST chooses coincidental endpoints so that coverage is maximized. In light of Propositions 2.1 and 2.2 this means MST transitions between two curves of equal cardinality at the location of their intersection. Crow (1956) [9] and Crow & Gardner (1959) [10] developed a minimal cardinality procedure in the binomial case and its analog in Poisson case respectively. The motivation for their method, abbreviated CG, was to resolve the gaps of Sterne's method. CG chooses coincidental endpoints to be the smallest possible value, except in the binomial case, if $\theta > 0.5$, the transition is as late as possible in order to maintain equivariance. Byrne and Kabaila's (2005)[17], abbreviated BK, developed a procedure in the Poisson case. Their approach is the opposite extreme of CG: their procedure chooses coincidental endpoints to be the largest possible value, except in the binomial case if $\theta > 0.5$ the transition must be as early as possible.

Blyth and Still (1983)[6] point out that transitioning as early (late) as possible in the binomial case results in shorter (longer) intervals for x near 0 and n and longer (shorter) for x near $n/2$ in comparison to other minimal cardinality procedures. Instead, Blyth and Still (1983) [6] , abbreviated BS, propose that coincidental endpoints should be "about midway between the two extremes." They discuss several rules for which to choose coincidental endpoints (all resulting in a choice near the middle of possible values), but ultimately decide that coincidental endpoints should be chosen to be exactly in the center of the range of possible values. Casella (1986)[12] proposed making coincidental endpoints choices for the binomial problem that result in a minimax procedure with respect to expected length; that is, coincidental endpoints should be chosen so that $\max_\theta EL(\theta)$ is minimized. This rule may not work well for distributions like the Poisson where interval width becomes arbitrarily large for increasing x.

Note that a plot like the ones shown in Figure 2.8 is difficult to produce in the NB case as there is no acceptance curve above 0.95 with a reasonable (finite) cardinality when θ is near 0. As noted by Choi (2015) [2] if one shifts their attention to estimation of the the NB mean,

$$\mu = \mathbb{E}(X) = \frac{r(1-\theta)}{\theta},$$

then construction of a minimal cardinality procedure becomes much the same as for the

Poisson case. In this way we can construct a CPF by choosing acceptance curves from left to right, and only need to deal with acceptance curves of a reasonable cardinality, as in the Poisson case. Confidence intervals for the NB success probability θ can then be easily obtained from the relationship,

$$\theta = \frac{r}{\mu + r}.$$

2.5.3 Strictly Increasing Endpoints

Although CG is gapless in the Poisson and binomial cases, its choice of coincidental endpoints causes CG to run into another problem: it has endpoints of consecutive x's that are tied and thus does not have strictly increasing endpoints. To see this consider the 95% Poisson CPF for CG shown in Figure 2.11. Included in the plot are the corresponding confidence intervals for the procedure. As can be seen from Figure 2.11 CG regularly violates this property. For instance the lower endpoint for $x = 7$ and $x = 8$ are tied.

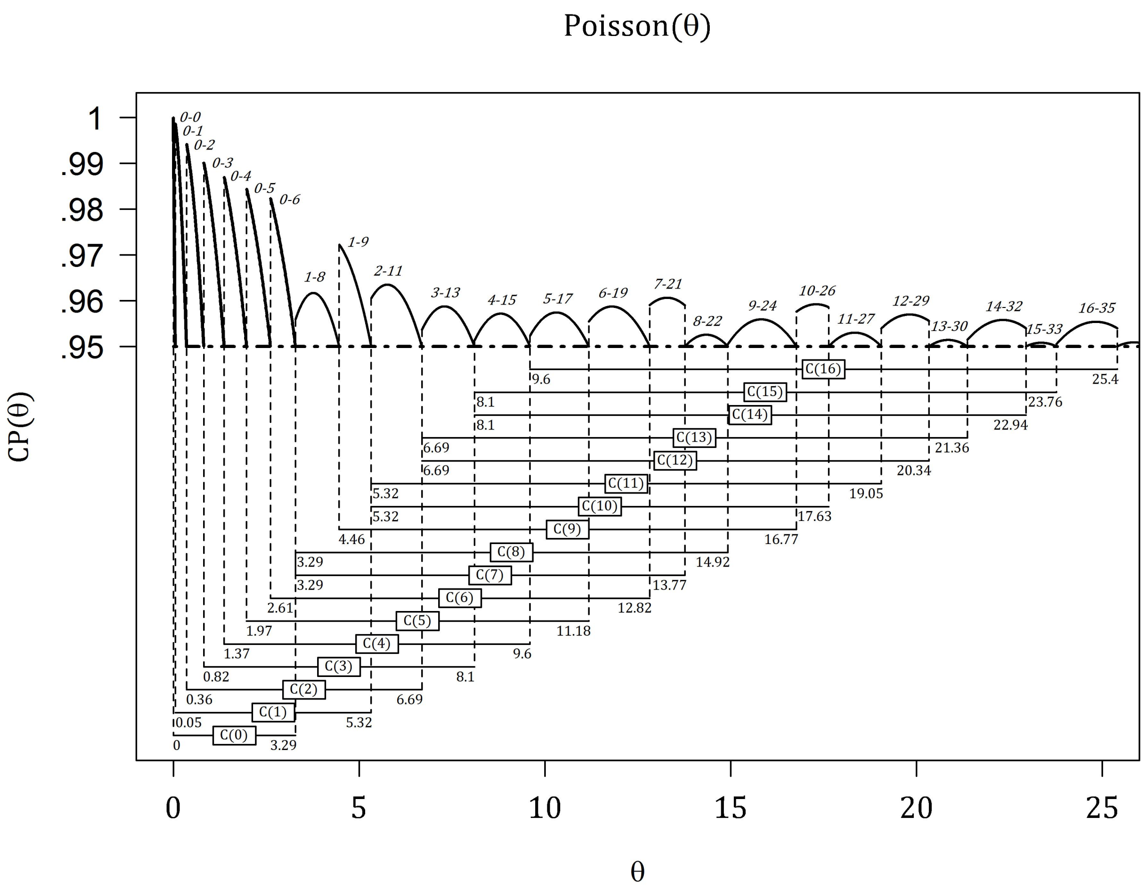

Figure 2.11: Interplay between Crow and Gardner's 95% Poisson CPF and resulting confidence intervals.

In general a minimal cardinality procedure will have tied endpoints if the CPF transitions from one acceptance curve to another that involves at least two new additional values of x. CG has tied endpoints for each of the binomial, Poisson and NB cases.

2.6 Length and Coverage Comparison for the Binomial

After a restriction is made to the collection of gapless minimal cardinality procedures attention can be shifted to comparing the interval width of different procedures. In the binomial case all minimal cardinality procedures have the same average interval width since different coincidental endpoint choices do not affect the net interval length of a procedure. Thus, another criterion such as coverage can be used as a tie breaker. Sterne's method is the unique minimal cardinality procedure with maximal coverage, but the procedure produces gaps in its confidence intervals. The next best option is MST since it is gapless and has near maximal coverage.

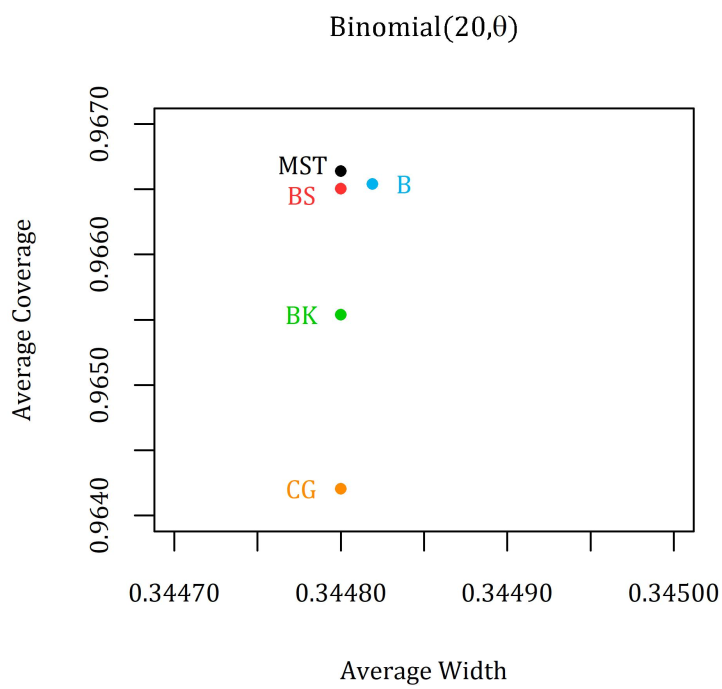

Figure 2.12: Plot average interval width v.s. mean coverage for the 95% binomial$(20, \theta)$ procedures: MST, BS, BK, CG, B

Figure 2.12 is the plot of average width versus average coverage for the binomial$(20, \theta)$ case for various minimal cardinality procedures. We also include B in the plot as it is competitive with the minimal cardinality procedures on length and coverage. The ideal procedure should located in the top left portion of the plot. The minimal cardinality procedures (MST, BS, BK, & CG) all have equal average interval width; whereas, B has a slightly larger interval width due to using acceptance curves which are not of minimal span. MST has larger coverage than both B and the other minimal cardinality procedures. A similar hierarchy occurs for all values of n. In particular, MST will remain the length/coverage optimal (gapless) procedure (Schilling & Doi (2014)[23])

2.7 Length and Coverage Comparison for the Poisson

In the Poisson case, Schilling and Holladay (2017)[22] show CG is the unique length optimal procedure according to their length criteria (see (1.7)-(1.8)). To see this consider the series of coincidental endpoints for the 95% Poisson case. The first coincidental endpoint is $u(0) = l(8)$ (see Figure 2.8). Since CG chooses its coincidental endpoints to be the smallest possible value it has the shortest interval for $x = 0$. But, it pays the price later by having the smallest value for the lower endpoint of $x = 8$. However, before it can pay the penalty it will have already made the upper endpoints for $x = 1, 2, ..., 7$ smaller than any other minimal cardinality procedure. So unlike the binomial case, in the Poisson case, CG never has to fully pay the price for having shorter intervals early on since there are infinitely many coincidental endpoints (Schilling and Holladay (2017)[22]). We sum up these facts with the following lemma and proposition:

Lemma 2.1 All upper endpoints of a Poisson minimal cardinality procedure are coincidental.

Proposition 2.3 CG attains the shortest interval width partial sums, $\sum_{x=0}^{K}(u(x)-l(x))$, of any Poisson minimal cardinality procedure.

For proofs of these results, see Schilling and Holladay (2017).[22]

Although CG is length optimal it unfortunately produces tied interval endpoints. Schilling & Holladay (2017) [22] propose a method for modifying CG in the Poisson case which has monotonic interval endpoints and near minimal length. Their proposal is as follows. Starting at $x = 0$ whenever

$$l(x+1) < ml(x) := l(x) + \min((.01)l(x), 0.1)$$

increase $l(x+1)$ to $ml(x)$. The resulting procedure will now have all consecutive endpoints separated by at least 1% if those endpoints are less than 10 or by at least 0.1 if the endpoints are 10 or greater. In this way small endpoints are changed less than the larger endpoints so that the relative change is small. Their rationale for these specific adjustments are as follows:

(i) Increases in endpoints need to be kept small in order to keep confidence interval lengths nearly as small as for CG, the length optimal confidence procedure;

(ii) Small endpoints (i.e., < 10) should be changed less than other endpoints so that the relative change in those endpoints is not large;

(iii) It is a fairly common practice to round confidence interval endpoints to three significant digits, except that for endpoints above 100 a single decimal digit is retained. With such a rounding protocol, all MCG endpoints will remain different after rounding;

(iv) A potential increase of much more than 0.1 leads to difficulties for large x because as x increases, progressively greater numbers of points are tied or nearly tied. [...] Modifying CG by spreading out the points in such a cluster by too large an amount would cause the modified points to run into the following points that are currently nearly evenly spaced out and not otherwise needing adjustment.

For more on MCG see Schilling & Holladay (2017)[22]. Similar adjustments may be made for other distributions. Adjustments should be large enough to comfortably satisfy the strictly increasing endpoints property, but small enough to keep coincidental endpoints within the restricted range of a minimal cardinality procedure. All in all, CG is the length optimal procedure in the Poisson case. However, since CG has tied endpoints, the next best option is arguably MCG as its endpoints are monotonic and has near minimal length.

In addition to considering the asymptotic behavior of confidence interval widths (via Schilling and Holladay's [22] criteria or another asymptotic criteria) we can compare procedures interval by interval. Figure 2.13 is plot of the interval width ranking for each strict procedure discussed with $(1-\alpha) = .95$. For instance for $x = 0$ the methods are ranked CG, MCG, B, BS, CP, MST and BK, in order of their interval length (shortest to widest). From this figure it is clear that CG has the shortest intervals of all 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. MCG is second best and is often tied with CG.

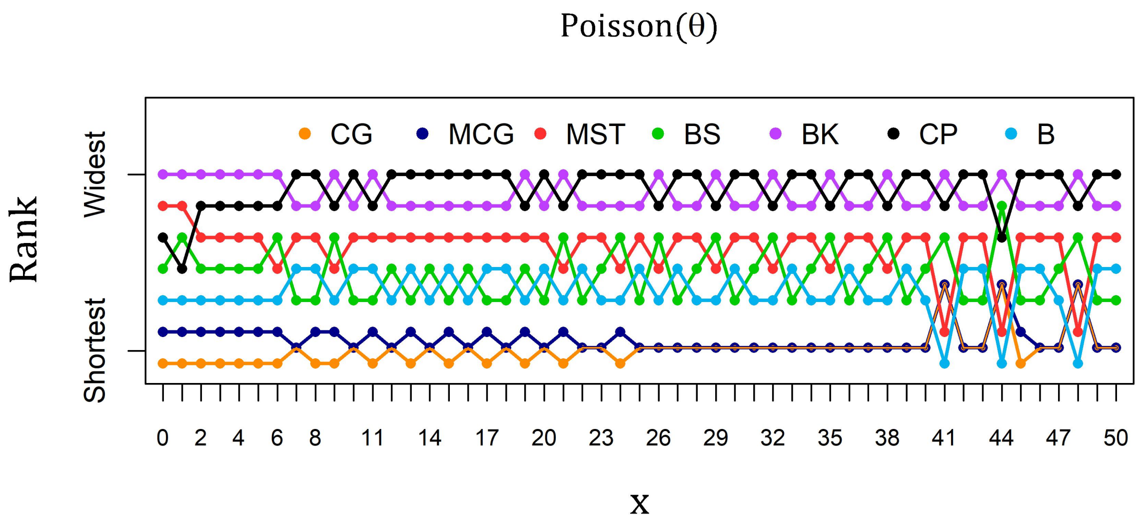

Figure 2.13: 95% Poisson confidence procedures ranked by confidence interval length for each x. Overlapping points indicate a tie has occurred for that x.

We can also compare the expected width,

$$\mathbb{E}_\theta\left[u(X) - l(X)\right] = \sum_{x \in \mathcal{X}} \left[u(x) - l(x)\right] \cdot P_\theta(X = x), \tag{2.6}$$

and the running average width,

$$\frac{1}{K}\sum_{x=0}^{K}(u(x) - l(x)) \tag{2.7}$$

(as a function of K). Figure 2.14 shows expected width and running average width for each strict procedure discussed relative to CP for $(1 - \alpha) = .95$; i.e., each given value (expected width or running average width) for each procedure is divided by the corresponding value for CP to produce the values shown in plots. CG clearly outperforms the other procedures as it has the smallest expected width and running average width of any procedure (Figure 2.14). MCG comes in for a close second place.

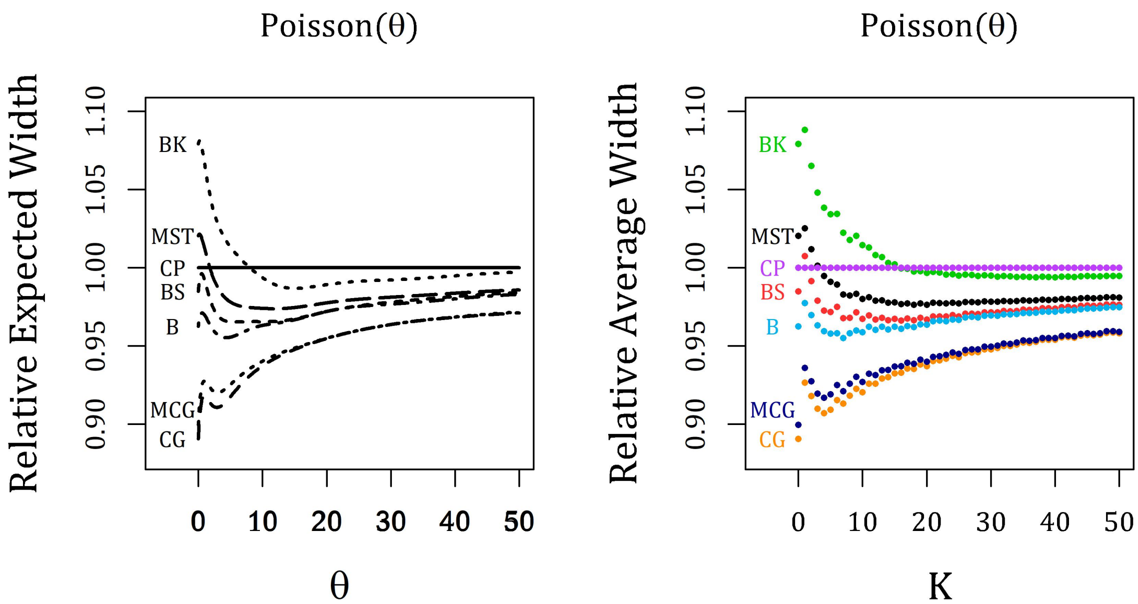

Figure 2.14: Plots of expected confidence interval widths (left) and running average confidence interval widths (right) for 95%Poisson procedures relative to that of CP.

In Figure 2.15 (left) we also provide plots of mean coverage on on $[0, \theta]$ (as a function of θ) for $(1-\alpha) = .95$. CP's wide intervals result in the largest coverage followed by B, MST and BS whom have very similar coverage. MST's coverage is almost indistinguishable from BS as their Type II transitions will often be very close to one another. Figure 2.15 (right) shows average confidence interval width for $x \in \{0, 1, ..., 50\}$ v.s. mean coverage on [0,50] for $(1-\alpha) = .95$. The ideal procedure would be located in the upper left corner. Since we often prioritize length over coverage, the procedures furtherest to the left are preferred.

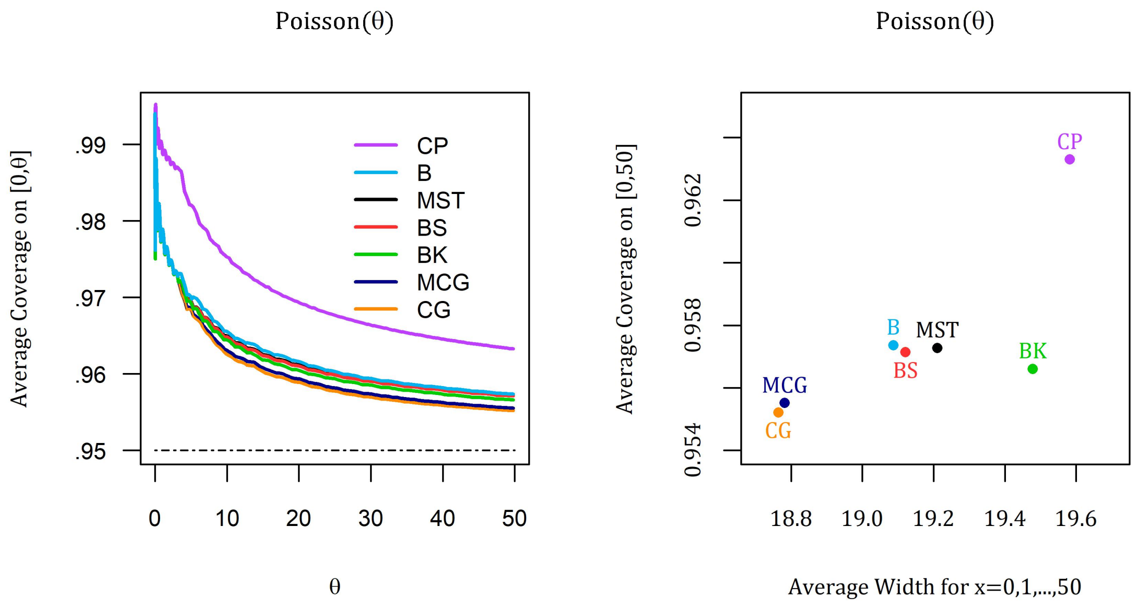

Figure 2.15: Plots of mean coverage on $[0, \theta]$ (left) and average confidence interval width for $x \in \{0, 1, ..., 50\}$ v.s. mean coverage on [0,50] (right) for 95% Poisson procedures.

2.8 Discussion: Conventional Intervals

In comparison to approximate methods, minimal cardinality procedures are not easy to compute and lack simple analytical expressions to compute by hand, but this should not be an issue with use of computers and free software such as R which can easily store tables of intervals or compute intervals quickly. Therefore, if an "optimal" confidence procedure is available for a particular distribution, it should be the preferred choice. If one decides to use approximate procedures, they should be aware of the shortfalls in the performance of such methods as these performance differences can be quite substantial in the discrete cases that we are discussing.

While coverage and interval length are the most important considerations in the selection of a procedure, if one procedure incurs a small sacrifice in length or coverage in order to satisfy another desirable property, it may be worth the cost. For instance in the binomial case if one uses MST rather the ST, one obtains a gapless length optimal procedure with near maximal coverage. For some it may be worth sacrificing a small amount of coverage in order to avoid a confidence interval with gaps. Or in the Poisson case if one uses MCG rather then CG, one can attain a procedure with strictly increasing endpoints that yields near length optimal interval widths: sacrificing optimal interval width to attain strictly increasing endpoints.

That is to say, in the Binomial case if a procedure with short intervals, high coverage, and no gaps is desired use MST. If the user doesn't mind gaps use ST as it has even higher coverage with intervals that are equally short. In the Poisson case if short intervals and increasing endpoints are desired use MCG. If the user doesn't mind having tied endpoints use CG as it produces even shorter intervals than MCG.

Unfortunately, no single procedure is globally superior to all the rest with respect to all of the desirable properties 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, one might prefer to use a minimal cardinality procedure, one with relatively short intervals, and one that possesses the desirable properties that the user requires. The choice among minimal cardinality procedures would depend on distribution and user preference.

Blaker (2000) proved that the family of minimal cardinality procedures will not have nested intervals in general. Blaker (2000) argued that any particular method may only satisfy at most two of the three properties: minimal cardinality, nestedness, and gapless. Schilling & Doi (2014) determined that MST rarely produces non-nested intervals for the binomial problem. Although not investigated we suspect this is probably true of the other methods and distributions. We leave it at the discretion of the practitioner to determine whether or not a sacrifice in interval length is necessary to achieve nested intervals, which is one selling point of Blaker's method.

2.9 Alternative Procedures

2.9.1 Mean Coverage Procedures

Even the best confidence procedures for discrete distributions have coverage often exceeding the confidence level – the result of this over-coverage being wider intervals. One remedy to the problem of over-coverage is the use of a *mean coverage* procedure i.e. a procedure whose average coverage equals the nominal confidence level. One can easily construct the mean-coverage version of a strict procedure by dropping the confidence coefficient (i.e. $\inf_{\theta\in\Theta} CP(\theta)$) of the procedure until an average coverage of $(1-\alpha)$ is reached. For a distribution with unbounded parameter space such as the Poisson, one may decide on a finite range for which to calculate the average coverage. The restriction can be made to include all reasonable values of the parameter. Minimal cardinality procedures are the perfect candidates for this process because one need not drop the confidence coefficient by much to achieve the nominal mean coverage. This is because minimal cardinality procedures tend to hover relatively close to the nominal confidence level and consistently hit $(1-\alpha)$ during each Type I transition.

Because of the over-coverage issue many statisticians have resorted to approximate methods. This is illustrated, for example, by the paper of Agresti & Coull[1], "Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions." We argue that a mean coverage procedure can be thought of as "approximate" in the sense that the coverage is approximately equal to $1-\alpha$ for each $\theta \in \Theta$. Moreover, since the mean coverage version of a procedure uses the distribution of X itself, it cannot only be made to produce shorter intervals than the other approximate procedures, but should be preferred based on that fact alone. A more thorough discussion of mean coverage procedures appears in Schilling & Doi(2014)[23] and Schilling & Holladay (2017)[22] for the binomial and Poisson cases respectively.

2.9.2 Bayes Credible Intervals

A second option to deal with over-coverage is to take a Bayesian approach to the problem. Let $\pi(\theta)$ denote the prior density of θ and let $\pi(\theta|x)$ be the posterior density of θ given $X = x$. If $\pi(\theta|x)$ is a continuous density then this discrete problem has effectively been transformed into a continuous one. For example in the binomial and Poisson cases the Beta and Gamma conjugate priors lead to continuous posterior densities (see Table 2.3) and thus allow for exact $(1-\alpha)$ credible probability.

Distribution	Conjugate Prior	Posterior	$\pi(\theta\|x) \propto$
binomial(n, θ)	Beta(α, β)	Beta$(\alpha + x, \beta + n - x)$	$\theta^{\alpha+x-1}(1-\theta)^{b+n-x-1}$
Poisson(θ)	Gamma(α, β)	Gamma$(\alpha + x, \beta + 1)$	$\theta^{\alpha+x-1}e^{-(1+\beta)\theta}$

Table 2.3: Conjugate priors for the Binomial(n, θ) and Poisson(θ) distributions.

The probability of an interval $[l(x), u(x)]$ is given by,

$$P(\theta \in [l(x), u(x))) = \int_{l(x)}^{u(x)} \pi(\theta|x)\, d\theta. \tag{2.8}$$

Thus, a length-optimal credible interval will be the shortest interval $[l(x), u(x)]$ satisfying,

$$\int_{l(x)}^{u(x)} \pi(\theta|x)\, d\theta = 1 - \alpha \tag{2.9}$$

Such an interval is obtained as

$$\{\theta : \pi(\theta|x) \geq k\}, \tag{2.10}$$

where k can be found by solving, if necessary iteratively, the equation

$$\int_{\{\theta:\pi(\theta|x)\geq k\}} \pi(\theta|x)\, d\theta = (1-\alpha). \tag{2.11}$$

A credible set of the form (2.10) is appropriately called the *highest posterior density* (HPD) region. It may be noted that when $\pi(\theta|x)$ is unimodal and symmetric, this simplifies to picking equal tails on either end of the posterior.

It is important to note that although such credible intervals achieve exact $(1-\alpha)$ probability, this does not translate to exact $(1-\alpha)$ coverage probability. The CPF of a procedure made up of credible intervals will be just as jagged and discontinuous as a typical conventional procedure. Credible coverage probability comes from the posterior distribution whereas classical coverage probability is taken only from the original PMF. 95% credible coverage means after combining our subjective prior knowledge of the parameter along with the given data, we are 95% confident of capture. On the other hand 95% classical coverage probability asserts that with repeated sampling θ will be captured 95% of the time.

An all-important question in Bayesian analysis is not only the choice of prior distribution but the choice of the hyperparameters (i.e. the parameters of the prior), when $\pi(\theta)$ is assumed to be of a known parametric form. In the absence of any prior information one can use a so-called vague prior, or the empirical Bayes method. Some common choices of vague priors are Jeffrey's prior (i.e. $\pi(\theta) \propto \sqrt{I(\theta)}$) which remains invariant under reparameterization and the uniform prior (i.e. $\pi(\theta) = 1$). Empirical Bayes methods, on the other hand, use the data itself to estimate the hyperparameters; hence, each observed value x results in a different prior. One common approach to estimate the hyperparameters is through ML-II estimation which estimates the hyperparameters by maximizing the marginal likelihood (see for instance Berger (1980, p. 99)[24])

Some practitioners, Bayesians and frequentists alike, question the use of such methods, when there is a lack of good prior information. However, when dealing with discrete data, a Bayesian approach can often transform a complicated discrete problem into a much simpler one involving a smooth density. Thus, in such cases, use of Bayesian machinery may seem more attractive, even if minimal prior information is available. A frequentist might then insist, however, that the resulting credible intervals should be subjected to the same evaluations as a conventional procedure. So how then do such credible intervals perform when evaluated under frequentist criteria?

Suppose, for example, θ is given Jeffrey's prior. In the binomial and Poisson cases this results in Beta$(x + 1/2, n - x + 1/2)$ and Gamma$(x + 1/2, 1)$ posterior densities respectively. Letting $C(x)$ be the resulting $(1 - \alpha)$ HPD region when x is observed, one can obtain a $(1 - \alpha)$ credible interval. The CPF's of the resulting confidence procedures are shown in Figure 2.16 for the binomial and Poisson cases.

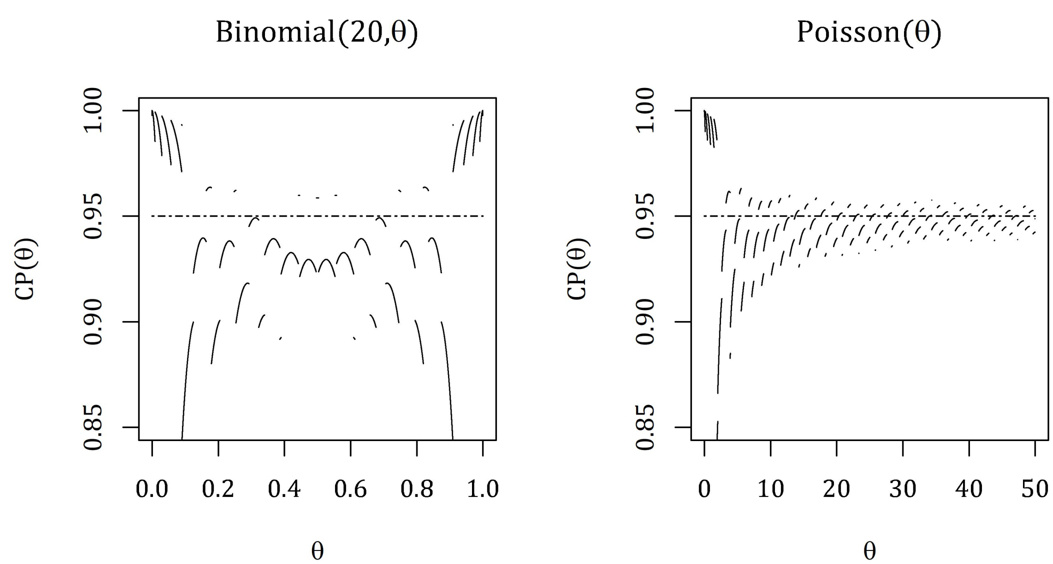

Figure 2.16: CPF of HPD regions when Jeffrey's prior is used.

Both the binomial and Poisson CPF's look similar to the corresponding CPF's produced by LR. Even with non-informative priors, modeling errors can often be underestimated, resulting in credible intervals that are too narrow. As a result, a Bayes approach will often yield coverage probabilities below the nominal level. To get an idea of the width of these credible intervals, Figure 2.17 compares the interval length of the 95% W and CP procedures with the 95% HPD regions produced by Jeffrey's prior. In both cases the length of the credible intervals are close to that of W; furthermore, both W and Jeffrey's HPD intervals are significantly shorter than the intervals produced by CP.

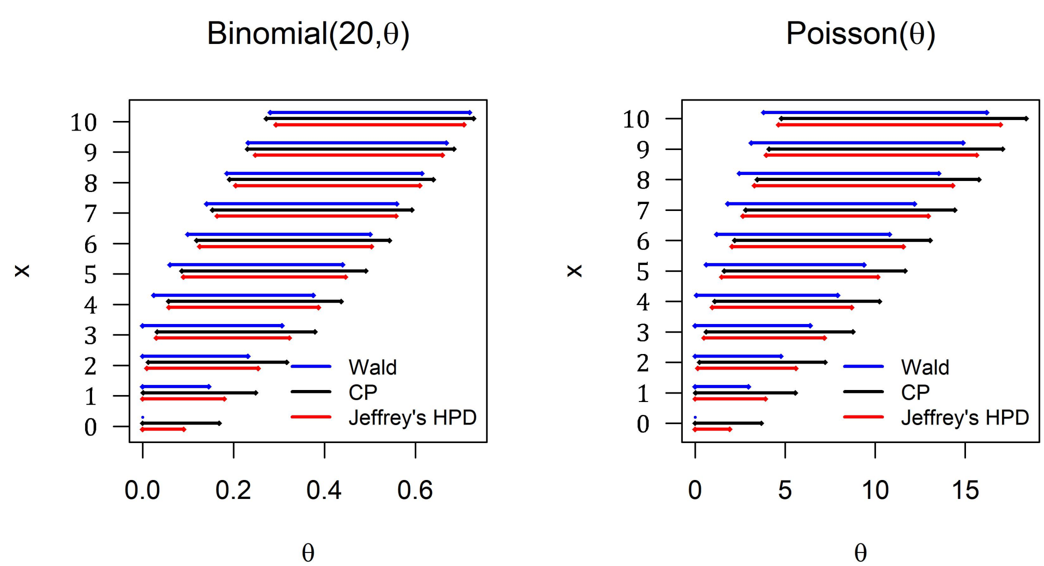

Figure 2.17: Comparison of interval length of the 95% Wald and CP procedures with the 95% HPD regions produced when Jeffrey's prior is used.

Chapter 3

Randomized and Fuzzy Confidence Procedures

Another approach to deal with over-coverage is through the use of randomization. We start our discussion of randomization with the following passage due to Stevens (1950)[21]:

> "It is the very basis of any theory of estimation, that the statistician shall be permitted to be wrong a certain proportion of times. Working within that permitted proportion, it is his job to find a pair of limits as narrow as he can possibly make them. If, however, when he presents us with his calculated limits, he says that his probability of being wrong is less than his permitted probability, we can only reply that his limits are unnecessarily wide and that he should narrow them until he is running the stipulated risk. Thus we reach the important, if at first sight paradoxical conclusion, that it is a statistician's duty to be wrong the stated proportion of times, and failure to reach this proportion is equivalent to using an inefficient in place of an efficient method of estimation."

The conventional procedures discussed thus far can be considered in Stevens' (1950)[21] own words "inefficient" methods of estimation. On the other hand, one of the attractive features of randomized procedures are that they allow for exact $(1-\alpha)$ coverage probability and consequently shorter intervals; i.e., they are an "efficient" method of estimation. The standard way to obtain a level $(1-\alpha)$ randomized confidence procedure is by inverting a family of level α randomized tests (see for instance Stevens (1950)[21] or Blyth & Hutchinson(1960)[5]).

However, many statisticians are critical of randomization and feel uneasy about letting a random number independent from the data, determine a decision. Randomization can result in a 90% confidence interval lying wholly inside of a 89% confidence interval, or in two researchers coming to opposite conclusions even with the same data, procedure and confidence level. Other objections to the use of randomized confidence intervals and some potential remedies to some of these objections are discussed in Kabaila (2013)[25]. A recent paper by Geyer & Meeden (2005)[18] introduces the notion of fuzzy confidence procedures, which can be thought of as an intermediate step between a conventional (i.e. non-random) and randomized procedure. Roughly speaking fuzzy intervals can be viewed as a graphical representation of randomized intervals before randomization has actually occurred. Since fuzzy intervals do not require randomization, they avoid some of the issues of randomized confidence intervals. We first discuss randomized tests and how they lead to randomized confidence intervals. We then show how the corresponding fuzzy intervals can be obtained. We propose several new fuzzy procedures and compare their performance with the fuzzy procedure proposed by Geyer & Meeden (2005)[18]. Unlike Geyer Meeden's fuzzy intervals that are derived from tests these new methods are derived from conventional confidence procedures.

3.1 Randomized Tests

Theorem 3.1 Consider testing the hypothesis $H_0 : \theta = \theta_0$ versus the 2-sided alternative $H_a : \theta \neq \theta_0$. If X has PMF of the form,

$$p_X(x|\theta) = h(x)c(\theta)e^{\theta T(x)}; \tag{3.1}$$

i.e., $\{p_X(x|\theta) : \theta \in \Theta\}$ is the so-called one-parameter exponential family, then there exists a Uniformly Most Powerful Unbiased (UMPU) randomized test of level α with the following test function (or critical function),

$$\phi(x, \theta_0) = \begin{cases} 1 & T(x) < C_1 \text{ or } T(x) > C_2 \\ \gamma_i & T(x) = C_i, \quad i = 1, 2 \\ 0 & C_1 < T(x) < C_2 \end{cases}, \tag{3.2}$$

where C_i's and γ_i's are determined by the level condition,

$$\mathbb{E}_{\theta_0}[\phi(X, \theta_0)] = \alpha \tag{3.3}$$

and the unbiased condition,

$$\mathbb{E}_{\theta_0}[T(X)\phi(X, \theta_0)] = \alpha \mathbb{E}_{\theta_0}[T(X)] \tag{3.4}$$

(see e.g. Lehmann & Romano (2005)[20], p. 111). The *critical function* ϕ by $x \mapsto \phi(x, \theta_0)$ gives the probability of rejecting H_0 at level α when X is x. The level condition ensures the probability of type I error is α and the *unbiased* condition ensures the probability of rejection is lowest when the null hypothesis is true. As with a confidence procedure there will always be some underlying level α associated with ϕ. Note that if the unbiased restriction is relaxed, Equation (3.2) along with (3.3) represent a larger class of level α randomized tests which we will study extensively. If γ_i's in (3.2) are both 0, this results in a non-random test. In this case, ϕ would only take the values 1 and 0, and can be expressed as,

$$\phi(x, \theta_0) = \begin{cases} 1 & x \notin A_{\theta_0} \\ 0 & x \in A_{\theta_0} \end{cases}, \tag{3.5}$$

where $A_{\theta_0} = \{x : C_1 \leq T(x) \leq C_2\}$. A confidence set for x based on ϕ (as in (3.5)) can then be constructed using the equivalence,

$$x \in A_\theta \iff \theta \in C(x).$$

An equivalent way to view the confidence set for x based on ϕ is through a series of hypothesis tests $H_0 : \theta = \theta_0$ versus $H_a : \theta \neq \theta_0$ carried out simultaneously for all $\theta_0 \in \Theta$. The confidence set for x consists of those θ_0's that would not reject H_0 if x was observed. Thus, the confidence set for x based on ϕ is,

$$C(x) = \{\theta : \phi(x, \theta) = 0\}.$$

Note the change in viewpoint from $\phi(x, \theta)$ as a function of x to a function θ for fixed x. If everything about ϕ is known then one can obtain both a complete family of level α tests and a corresponding family of level $(1 - \alpha)$ confidence sets.

The randomized test in (3.2) differs from the conventional test in (3.5) by changing what happens when $T(x)$ lands on the boundary of A_{θ_0}: if $T(x)$ lands on a boundary point C_i, flip a coin with probability of heads γ_i and reject only in the case of a heads-up toss. It is the added randomness to the decision at boundary points that allows for an exact level α test, which is generally not possible in the discrete case.

In practice the randomized test in (3.2) may be performed as follows. Simulate a random variable V uniformly from the interval $(0, 1)$ and Reject H_0 at level α if $V < \phi(x, \theta_0)$. Thus, do not reject if $C_1 < T(x) < C_2$, since $\mathbb{P}[V < \phi(x, \theta_0)] = \mathbb{P}(V < 0) = 0$, and reject if $T(x) < C_1$ or $T(x) > C_2$ since $\mathbb{P}[V < \phi(x, \theta_0)] = P[U < 1] = 1$. If $T(x) = C_i$, reject if $V < \gamma_i$ which has probability $\mathbb{P}[V < \phi(x, \theta_0)] = \mathbb{P}(V < \gamma_i) = \gamma_i$ of occurring. Both the binomial and Poisson distributions can be expressed in the exponential family form given in (3.1) with $T(x) = x$ and thus are two examples of distributions with UMPU test functions given by equation (3.2) with C_i's and γ_i's determined by (3.3)-(3.4).

3.2 Fuzzy and Randomized Procedures

Interest lies in constructing a $100(1-\alpha)\%$ confidence procedure based on a level α test of the form given in (3.2) satisfying (3.3) (UMPU or otherwise). In the special case where the family of tests are in fact UMPU (e.g., exponential family + (3.4)) the resulting confidence procedure is known to be uniformly most accurate unbiased (UMAU); that is, the procedure minimizes the probability of capturing false values of θ among all other $100(1-\alpha)\%$ unbiased confidence procedures. Note a confidence procedure is said to be *unbiased* if the probability of capturing false values of θ does not exceed the confidence level.

Analogous to the non-random case, to generate a randomized confidence set for x simultaneously carry out randomized tests for all $\theta_0 \in \Theta$ and collect those θ_0's whose tests did not reject H_0 when x was observed. The key to making this work is to use same realization of V for all the tests. Each distinct realization of V would result in a different interval. Let v be a particular realization of V then the $(1-\alpha)$ confidence set for x corresponding to v is,

$$C(x, v) = \{\theta : 1 - \phi(x, \theta) > v\},$$

where the procedure C has now been written as a function of both the observed x and the random component v.

Leaving V unrealized and considering

$$\psi(x, \theta) \coloneqq 1 - \phi(x, \theta) \tag{3.6}$$

as a function of θ results in an entity called a *fuzzy confidence interval*. Two statisticians can avoid the problem of getting different intervals from the same data by keeping them fuzzy since it is the different realizations of V that cause the discrepancy. However, if an immediate decision does need to be made one could carry out the simulation of V to obtain a realized random confidence interval. We may view the *membership function* ψ as two different functions:

- For fixed θ_0, $x \mapsto \psi(x, \theta_0)$ is called the *fuzzy decision* of level α. It provides the degree (probability) of acceptance of the hypothesis $H_0 : \theta = \theta_0$ when x is observed.
- For fixed x_0, $\theta \mapsto \psi(x_0, \theta)$ is called the *fuzzy confidence interval* with coverage $(1 - \alpha)$. It provides the degree (probability) of membership (inclusion) of each θ into the confidence interval of x_0.

We note that when the function $\theta \mapsto \psi(x, \theta)$ is viewed strictly as a fuzzy interval (without intention of carrying out a randomization) it represents the *degree of membership* of each θ into the fuzzy confidence interval for x. However, when there is an intention to randomize, the degree of membership is really the probability of inclusion of θ into the randomized confidence interval for x. Such a distinction is minor, but the change in language plays an important role in distinguishing fuzzy and randomized confidence intervals. These are really just two interpretations of the same function and we will often prefer the latter description as we suspect the reader is more familiar with randomization than fuzzy set theory.

Figure 3.1 shows three different 95% (UMAU) fuzzy confidence intervals (derived from the UMPU tests) in the binomial$(10, \theta)$ case. The plot shows the fuzzy intervals for $x = 0$ (dashed curve), $x = 3$ (solid curve), and $x = 9$ (dotted curve). These are the intervals proposed by Geyer & Meeden (2005) [18]. As done by Geyer & Meeden (2005) [18] we borrow some terminology from fuzzy set theory. The θ's that have positive probability of inclusion into the randomized confidence interval for each x make up the *support* of the fuzzy interval and those with probability 1 make up the *core*. The portion of θ's with positive inclusion probability less than one define the *fuzzy edges* of each fuzzy interval.

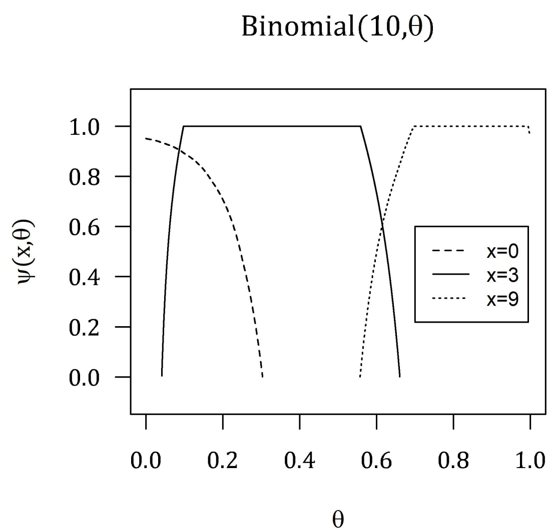

Figure 3.1: 95% UMAU fuzzy confidence intervals of the binomial$(10, \theta)$ distribution for observed data $x = 0$ (dashed curve), $x = 3$ (solid curve), and $x = 9$ (dotted curve). Only the nonzero portions of the intervals are plotted.

A fuzzy interval that takes only values 0 and 1 is called *crisp*. Hence, conventional confidence intervals are a special case of fuzzy confidence intervals. That is, conventional confidence intervals are really just crisp fuzzy confidence intervals and therefore any of our fuzzy confidence interval theory is just a generalization of current theory.

In practice researchers could display their fuzzy intervals as in Figure 3.2 which may be more visually appealing.

Figure 3.2: A fuzzy interval display suggestion.

Since randomized confidence procedures depend both on the random variable X and an auxiliary random component V (typically, $V \sim U(0, 1)$ and X, V are independent) a

$(1-\alpha)$ randomized confidence interval, in general, is given by,

$$C(x,v) = [l(x,v), u(x,v)],$$

so that each observation x and realization v of V generates a different confidence interval. Figure 3.3 illustrates how randomized intervals may be obtained from their corresponding fuzzy intervals. Suppose x is observed and v is drawn uniformly from $(0,1)$. The figure illustrates how v determines the portion of the fuzzy edges that are included in the randomized interval.

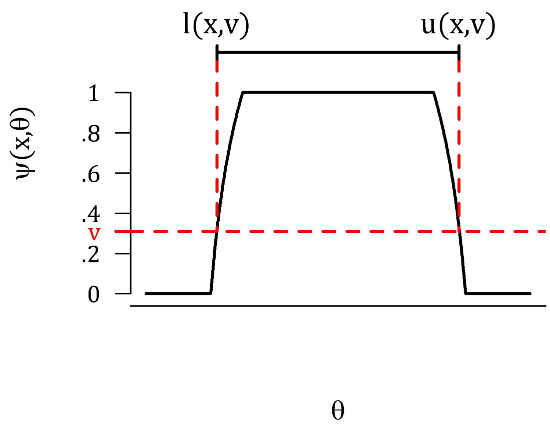

Figure 3.3: Illustration on how one may generate a randomized interval from a fuzzy one.

When constructing a randomized confidence procedure based on a given membership function ψ, like the one depicted in Figure 3.3, the confidence intervals limits are given by,

$$l(x,v) = \min\{\theta : \psi(x,\theta) = v\},$$

$$u(x,v) = \max\{\theta : \psi(x,\theta) = v\},$$

so that $C(x, v)$ comprises all values of θ for which $\psi(x, \theta) > v$.

The coverage probability of a confidence procedure based on a given membership function ψ can be expressed as,

$$CP(\theta) = \mathbb{E}[\psi(X, \theta)] = \sum_x \psi(x, \theta) \cdot \mathbb{P}_\theta(X = x).$$

Thus, if a critical function ϕ corresponds to an exact level α tests (i.e. $\alpha = \mathbb{E}[\phi(X, \theta)]$) then the corresponding randomized/fuzzy confidence procedure produced by ψ will have exact $(1 - \alpha)$ coverage probability for all $\theta \in \Theta$. In the case that ϕ/ψ generates a conventional test/procedure then $\mathbb{E}[\psi(X, \theta)] \geq (1 - \alpha)$.

The following theorem due to Pratt (1961)[19] gives an important relationship between false coverage probabilities and expected interval width:

Theorem 3.2 Let $C(x, v) = [l(x, v), u(x, v)]$ define a randomized confidence procedure. Then, for any value $\theta \in \Theta$,

$$\mathbb{E}_\theta \left[u(X, V) - l(X, V)\right] = \int_{\theta' \neq \theta} \mathbb{P}_\theta(\theta' \in C(X, U)) \tag{3.7}$$

Theorem 3.2 says that the expected length of a confidence procedure $C(X, V)$ is equal to an integral of false coverage probabilities. Since UMAU confidence procedures minimize the probability of false coverage it follows that:

Corollary 3.1 UMAU confidence procedures have the smallest expected interval width among unbiased procedures.

3.3 UMAU Fuzzy Confidence Procedures

In order to produce a graph of a $100(1-\alpha)\%$ fuzzy interval corresponding to a fixed x and α one needs to plot $\theta \mapsto \psi(x, \theta)$. Recall the membership function $\psi(x, \theta) := 1-\phi(x, \theta)$

and thus is determined by the values of the C_i's and γ_i's in (3.2). In the UMPU/UMAU case the C_i's and γ_i's can be found by solving the level and unbiased conditions (see (3.3) and (3.4)) for each $\theta \in \Theta$. The level condition can be simplified to,

$$(1-\alpha) = \mathbb{P}_\theta[C_1 < T(X) < C_2] + \sum_{i=1}^{2} (1-\gamma_i)\mathbb{P}_\theta[T(X) = C_i]. \tag{3.8}$$

And the unbiased condition can be written as,

$$(1-\alpha)\mathbb{E}_\theta[T(X)] = \mathbb{E}_\theta\left[T(X)\mathbb{1}_{\{C_1 < T(X) < C_2\}}\right] + \sum_{i=1}^{2} C_i(1-\gamma_i)\mathbb{P}_\theta[T(X) = C_i]. \tag{3.9}$$

In the special case where $C_1 = C_2$ is a solution then $\gamma_1 = \gamma_2$ also and the solution becomes,

$$C_1 = C_2 = \mathbb{E}_\theta[T(X)],$$

$$\gamma_1 = \gamma_2 = 1 - \frac{1-\alpha}{2p}.$$

Otherwise fix C_1 and C_2 such that $C_1 < C_2$. Then, solving the system for γ_1 and γ_2 yields,

$$\gamma_1 = 1 - \frac{(1-\alpha)(C_2 - \mathbb{E}_\theta[T(X)]) + \mathbb{E}_\theta\left[T(X)\mathbb{1}_{\{C_1<T(X)<C_2\}}\right] - C_2\mathbb{P}_\theta[C_1 < T(X) < C_2]}{(C_2 - C_1)\mathbb{P}_\theta[T(X) = C_1]} \tag{3.10}$$

$$\gamma_2 = 1 - \frac{(1-\alpha)(\mathbb{E}_\theta[T(X)] - C_1) - \mathbb{E}_\theta\left[T(X)\mathbb{1}_{\{C_1<T(X)<C_2\}}\right] + C_1\mathbb{P}_\theta[C_1 < T(X) < C_2]}{(C_2 - C_1)\mathbb{P}_\theta[T(X) = C_2]} \tag{3.11}$$

Since γ_1 and γ_2 can be completely expressed in terms of C_1 and C_2 determining $\phi(x, \theta)$ in the UMPU case reduces to a simple search of appropriate values of C_1 and C_2. Any choice of C_i's from the support of X such that $C_1 < C_2$ will automatically satisfy (3.8) and (3.9) as long as γ_i's are defined by (3.10)-(3.11). However, in order for the solution to be meaningful the boundary probabilities must satisfy, $0 \leq \gamma_i \leq 1,\ i = 1, 2$. As a result the search for a solution comes down to finding a pair of C_i's that produce γ_i's in $[0, 1]$.

Figure 3.4 shows the 95% UMAU fuzzy confidence procedures for the binomial and Poisson cases. For the Poisson case intervals up to $x = 10$ are provided. Observe that in the binomial case the intervals for $x = 0$ and 1 start at 0.95 and the interval for $x = 9$ and 10 end at 0.95. Geyer & Meeden (2005)[18] show that this behavior is expected (for UMPU fuzzy intervals) for the cases when $x = 0, 1, n-1$, and n. Suppose that the range of $T(X)$ is bounded below (as it is for the binomial and Poisson distributions). Then, Geyer & Meeden (2005) prove that as θ approaches $-\infty$ the UMPU critical function $\phi(x, \theta)$ converges to α for each x such that $T(x)$ is equal to either of its two smallest values (i.e. when $x = 0$ and 1 in the binomial and Poisson cases). It converges to 1 for all other values of x. An analogous convergence occurs as θ approaches ∞ when $T(x)$ bounded above (as is the case for the binomial, but not for Poisson).

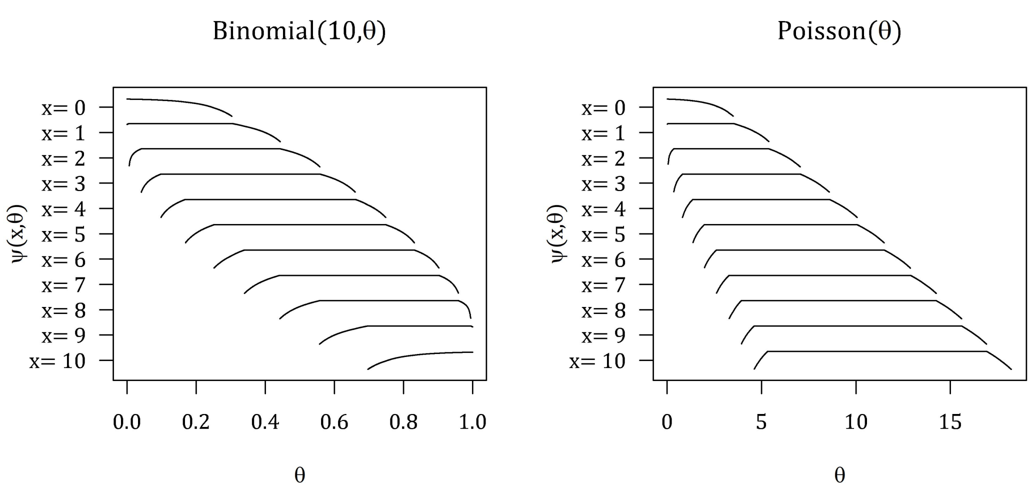

Figure 3.4: 95% fuzzy UMAU confidence procedures of the binomial$(10, \theta)$ and Poisson(θ) distributions. Only the nonzero portions of the intervals are plotted.

3.4 Alternative Fuzzy Procedures

It is known that confidence intervals derived from UMPU tests are UMAU and have the shortest expected length of any unbiased confidence procedure, but what happens if the unbiased condition is relaxed? Would they still be length optimal in this larger class of procedures? In this section we relax the unbiased condition and look for the optimal fuzzy procedures under various criteria. From a practical perspective, confidence intervals (fuzzy or otherwise) are supposed to contain the parameter with high probability. Thus in the case of fuzzy intervals, by minimizing the length of the support of the intervals, we do a better job of honing in the true value of the parameter. Moreover, when comparing fuzzy intervals, support length is arguably more important than the (expected) interval length after randomization because the randomization never actually occurs and therefore the interval endpoints are never realized. Therefore, fuzzy intervals with shortest support may be preferred.

To develop new fuzzy confidence procedures we capitalize on the following idea. Any conventional confidence interval can be thought of as fuzzy interval that takes only values 0 and 1 (i.e. a crisp interval). This of course results in a violation of the level condition; i.e, the procedure has coverage exceeding $1-\alpha$. By allowing the degree of membership of some θ's from the core of the interval to be less than 1, essentially turning parts of the crisp interval fuzzy, we can achieve $1-\alpha$ coverage exactly.

3.4.1 Equal-Tail Fuzzy Intervals

A description of the process to transform a conventional procedure into a fuzzy one is best done through an example. A natural procedure to start with is Clopper & Pearson's (1934)[7] method (CP) as it is often considered the "gold standard" of conventional confidence procedures. Each interval $C(x)$ from CP can viewed as a crisp interval by giving each $\theta \in \Theta$ an inclusion probability of 1 if it contained in $C(x)$ and an inclusion probability of 0 otherwise. This results in the membership function,

$$\psi(x,\theta) = \begin{cases} 1 & x \in A_\theta \\ 0 & \text{else} \end{cases} . \tag{3.12}$$

Each acceptance region $A_\theta = \{a, a+1, ..., b\}$ is a set of consecutive integers. Following the lead of the membership functions derived from the UMPU tests we assign exclusion probabilities $\gamma_1, \gamma_2 \in [0,1]$ only to the boundaries a and b of the acceptance region, producing a new membership function,

$$\psi(x,\theta) = \begin{cases} 1 & x \in \{a+1, a+2, ..., b-1\} \\ 1-\gamma_1 & x = a \\ 1-\gamma_2 & x = b \\ 0 & \text{else} \end{cases} . \tag{3.13}$$

For each $\theta \in \Theta$ the pair (γ_1, γ_2) can be any two probabilities that result in $\mathbb{E}_\theta\left[\psi(X,\theta)\right] = 1-\alpha$. Note that the values of a ,b, γ_1, and γ_2 in (3.13) are functions of θ. As a result, we may choose a different pair $(\gamma_1, \gamma_2) = (\gamma_1(\theta), \gamma_2(\theta))$ for each θ. Since CP was derived by bringing the rejection region tail probabilities $\mathbb{P}_\theta(X < a)$ and $\mathbb{P}_\theta(X > b)$ as close to $\alpha/2$ as possible without exceedance, a natural choice for exclusion probabilities would be values which bring the rejection region tail probabilities to exactly $\alpha/2$. In other words we should choose γ_1 and γ_2 so that,

$$\begin{aligned} \mathbb{P}_\theta(X < a) + \gamma_1 \mathbb{P}_\theta(X = a) = \alpha/2, \\ \mathbb{P}_\theta(X > b) + \gamma_2 \mathbb{P}_\theta(X = b) = \alpha/2. \end{aligned} \tag{3.14}$$

Hence,

$$\begin{aligned} \mathbb{E}_\theta\left[\psi(X,\theta)\right] &= 1 - \left[P_\theta(X < a) + \gamma_1 P_\theta(X = a) + P_\theta(X > b) + \gamma_2 P_\theta(X = b)\right] \\ &= 1 - \alpha. \end{aligned} \tag{3.15}$$

In the case that $A_\theta = \{a\}$ is a singleton (i.e. $a = b$), ψ reduces to

$$\psi(x,\theta) = \begin{cases} 1-\gamma & x = a \\ 0 & \text{else} \end{cases}. \tag{3.16}$$

Thus, γ must be chosen so that

$$(1-\gamma)\cdot \mathbb{P}(X = a) = 1-\alpha. \tag{3.17}$$

To see an example of this process in action take a look at Figure 3.5. In black are the 80% crisp CP intervals and corresponding coverage probability function. Blue dashed lines represent where the fuzzy version of CP differs from the crisp version after undergoing the above process. Consider the θ's corresponding to the region where the CPF uses the acceptance curve $0-0$. For each θ in this region $A_\theta = \{0\}$ is a singleton. Thus,

$$\psi(x,\theta) = \begin{cases} 1-\gamma & x = 0 \\ 0 & \text{else} \end{cases}. \tag{3.18}$$

where γ is the solution of $(1-\gamma)\cdot \mathbb{P}_\theta(X = 0) = 0.80$. Note that γ is different for each θ. Now consider for instance the θ's corresponding to the region where the CPF uses the acceptance curve $1-3$. In this case the membership function becomes,

$$\psi(x,\theta) = \begin{cases} 1 & x = 2 \\ 1-\gamma_1 & x = 1 \\ 1-\gamma_2 & x = 3 \\ 0 & \text{else} \end{cases}, \tag{3.19}$$

and the γ_i's should be chosen so that

$$\begin{aligned} \mathbb{P}_\theta(X < 1) + (1-\gamma_1)\mathbb{P}_\theta(X = 1) = 0.1, \\ \mathbb{P}_\theta(X > 3) + (1-\gamma_2)\mathbb{P}_\theta(X = 3) = 0.1. \end{aligned} \tag{3.20}$$

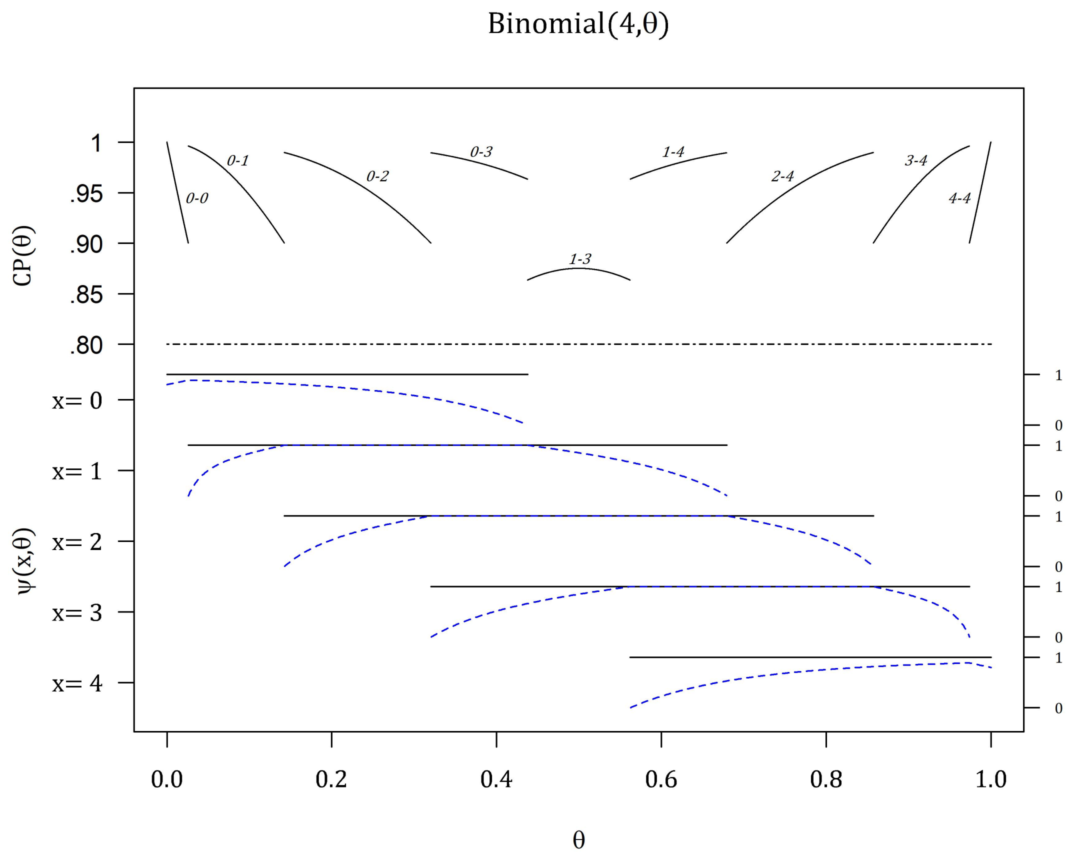

Figure 3.5: 80% crisp CP (black) and fuzzy CP (blue dashed) confidence procedures for the binomial$(4, \theta)$ distribution. Only the nonzero portions of the intervals are plotted. The blue dashed line shows where the two procedures differ.

Figure 3.6 shows the 95% CP fuzzy confidence procedures for the binomial$(10, \theta)$ and Poisson(θ) distributions. They do not appear to be much different from the fuzzy UMAU intervals in Figure 3.4 in terms of length and shape. And as we will see later they are especially similar in the Poisson case.

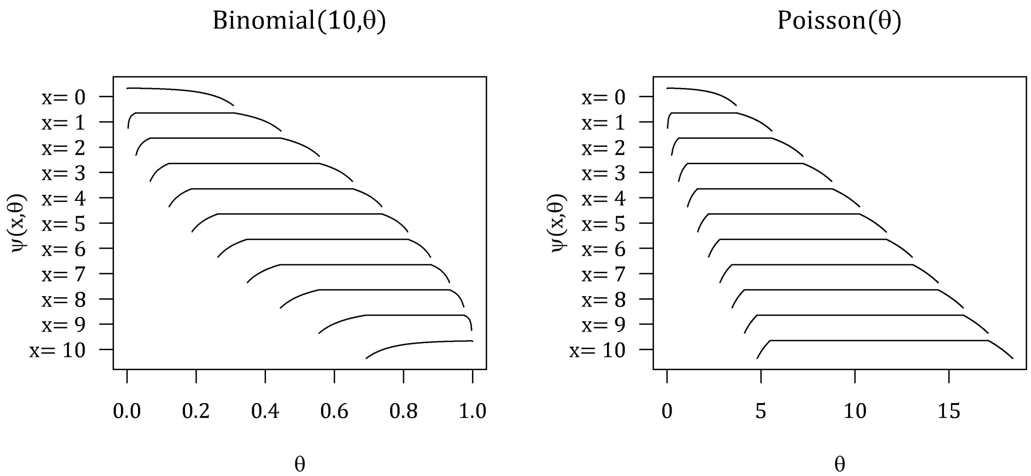

Figure 3.6: 95% fuzzy CP confidence procedures of the binomial$(10, \theta)$ and Poisson(θ) distributions. Only the nonzero portions of the intervals are plotted.

3.4.2 Minimal Cardinality Fuzzy Intervals

In this section we describe the process for transforming a conventional minimal cardinality procedure into a fuzzy procedure. The process discussed above of transforming a conventional procedure into a fuzzy one, by essentially bending the edges of the core of the crisp interval downwards, will result in a fuzzy procedure whose support length matches that of the original procedure. Thus, since minimal cardinality procedures produce short intervals, by executing a similar process, the fuzzy versions of these procedures will have equally short support lengths, which is our motivation for the process. As we did when transforming CP into a fuzzy procedure we think of each minimal cardinality conventional procedure as a crisp fuzzy procedure. We then assign exclusion probabilities $\gamma_1, \gamma_2 \in [0, 1]$ only to the boundaries a and b of the acceptance regions $A_\theta = \{a, a+1, ..., b\}$, producing a membership function of the form in (3.13).

As before for each $\theta \in \Theta$ the pair (γ_1, γ_2) can be any two probabilities satisfying $\mathbb{E}_\theta [\psi(X, \theta)] = 1-\alpha$. However unlike CP, minimal cardinality procedures have asymmetric rejection region tail probabilities, and thus there is no obvious criteria for choosing the values of (γ_1, γ_2). Note that because the acceptance sets are of minimal cardinality, only one of the γ_i's need be nonzero to achieve coverage $\mathbb{E}_\theta [\psi(X, \theta)] = 1 - \alpha$ for any given value of $\theta \in \Theta$. To further understand the problem at hand, take a look at Figure 3.7 which shows the 80% CPF for MST in the binomial$(4, \theta)$ case along with the fuzzy intervals that we will ultimately propose. Note because of the equivariance property we need only define our proposed fuzzy intervals for $\theta \in [0, 0.5]$.

First consider the θ's corresponding to where the CPF uses $0 - 0$. Since $A_\theta = \{0\}$ is a singleton the membership function reduces to (3.18) as it did for CP, where γ is again the solution of $(1 - \gamma) \cdot \mathbb{P}_\theta(X = 0) = 0.80$.

Moving to the right, consider the θ's associated with the curve $0 - 1$. In this case $A_\theta = \{0, 1\}$ and the membership function takes the form,

$$\psi(x, \theta) = \begin{cases} 1 - \gamma_1 & x = 0 \\ 1 - \gamma_2 & x = 1 \\ 0 & \text{else} \end{cases}, \tag{3.21}$$

Since the left fuzzy edge of $x = 0$ has already reached 1 it would be undesirable to allow it drop below 1 a second time as the interval would then have 2 cores, leading to interpretation issues. Hence for each θ is this region, we set $\gamma_1 = 0$, which in turn defines γ_2.

Moving over to the next acceptance curve $0 - 2$ we get a membership function of the form,

$$\psi(x, \theta) = \begin{cases} 1 & x = 1 \\ 1 - \gamma_1 & x = 0 \\ 1 - \gamma_2 & x = 2 \\ 0 & \text{else} \end{cases}, \tag{3.22}$$

There are infinitely many choices for (γ_1, γ_2) that satisfy, $\mathbb{E}_\theta\left[\psi(X,\theta)\right] = 0.80$. Setting γ_1 (γ_2) to 0 gives the maximum possible value for γ_2 (γ_1). The resulting fuzzy edges from these two different choices are shown in the Figure 3.7 in blue and red dashed lines respectively. Any choice of γ_i's between the corresponding blue and red curves will result in 80% coverage probability as long as after choosing either γ_1 or γ_2 the other is chosen so that $\mathbb{E}_\theta\left[\psi(X,\theta)\right] = 0.80$. Selecting γ_i's using a criteria such as the minimization of the area under the fuzzy interval or the minimization of the expected length (after randomization) will result in a fuzzy interval that is discontinuous on its support, which we find extremely undesirable. For simplicity and aesthetic appeal we propose a *straight-line* approach for γ_1 as shown in black which in turn defines γ_2. The "optimal" selection of the γ_i's is still an open problem. Note for now, our primary concern is support length and shape (for ease of interpretation), neither of which is harmed by the straight-line approach.

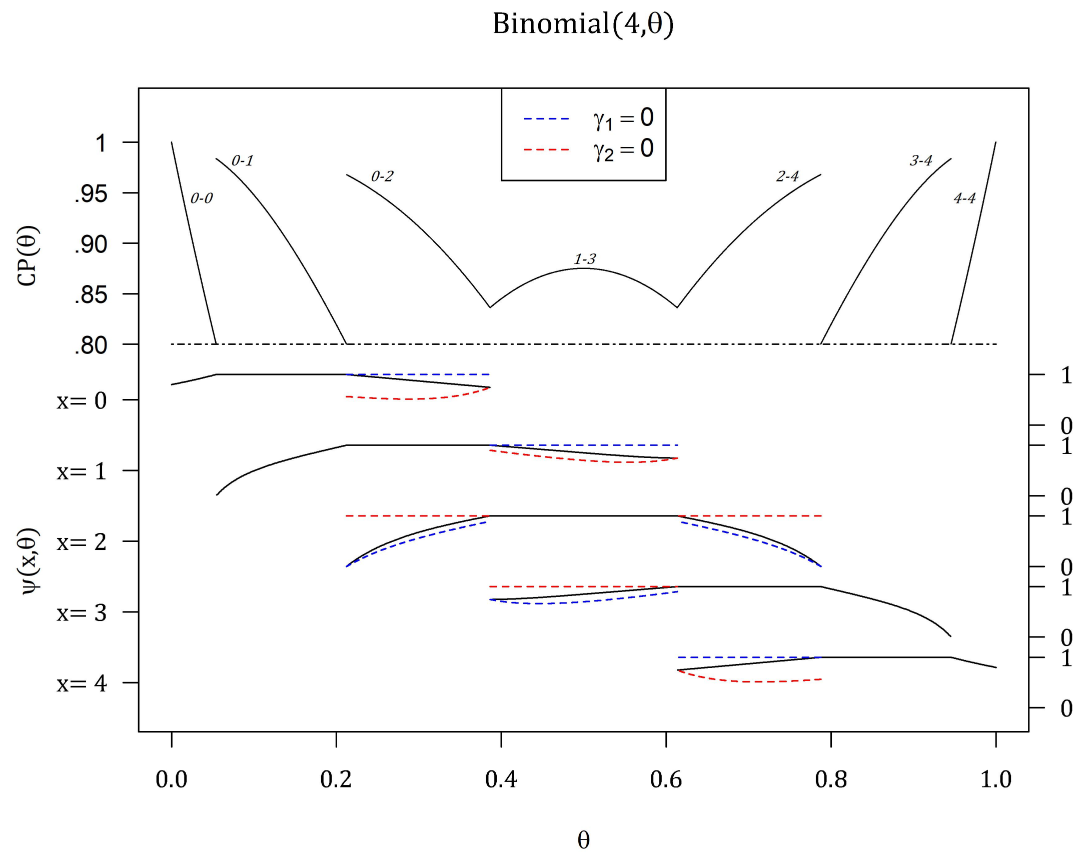

Figure 3.7: 80% MST CPF and fuzzy confidence procedure for the binomial$(4, \theta)$ distribution. The blue and red dashed lines show how the procedure would look if we were to set $\gamma_1 = 0$ or $\gamma_2 = 0$ respectively.

Lastly we consider the region where the CPF uses the acceptance curve $1-3$, which produces a membership function of the form,

$$\psi(x,\theta) = \begin{cases} 1 & x = 2 \\ 1-\gamma_1 & x = 1 \\ 1-\gamma_2 & x = 3 \\ 0 & \text{else} \end{cases}, \tag{3.23}$$

Since the fuzzy edge of both $x = 1$, and 3 cross over $\theta = 0.5$ we need to be wary of the equivariance property, as equivariance requires $\psi(x,\theta) = \psi(n-x, 1-\theta)$. Hence in order to avoid discontinuities of the fuzzy intervals at $\theta = 0.5$ we must choose values of (γ_1, γ_2) so that $\psi(1, 0.5) = \psi(3, 0.5)$. We may then use the straight-line approach for the right fuzzy edge of $x = 1$ by connecting its core to the corresponding value of $\psi(1, 0.5)$ determined in the previous step. This defines the right fuzzy edge of $x = 1$ for $\theta \leq 0.5$ which in turn defines the left fuzzy edge of $x = 3$ for $\theta \leq 0.5$. We have now finished defining the procedure $\psi(x, \theta)$ for all $\theta \leq 0.5$. For $\theta > 0.5$ use the equivariance requirement, $\psi(x, \theta) = \psi(n - x, 1 - \theta)$, to determine all remaining values of ψ.

The above process can extended to work for all other values of n and all confidence levels. In the Poisson case the process is similar, except we need not worry about the equivariance requirement. Moreover the above approach works for any of our minimal cardinality procedures. Figure 3.8 shows the 95% CG fuzzy confidence procedures for the binomial$(10, \theta)$ and Poisson(θ) distributions.

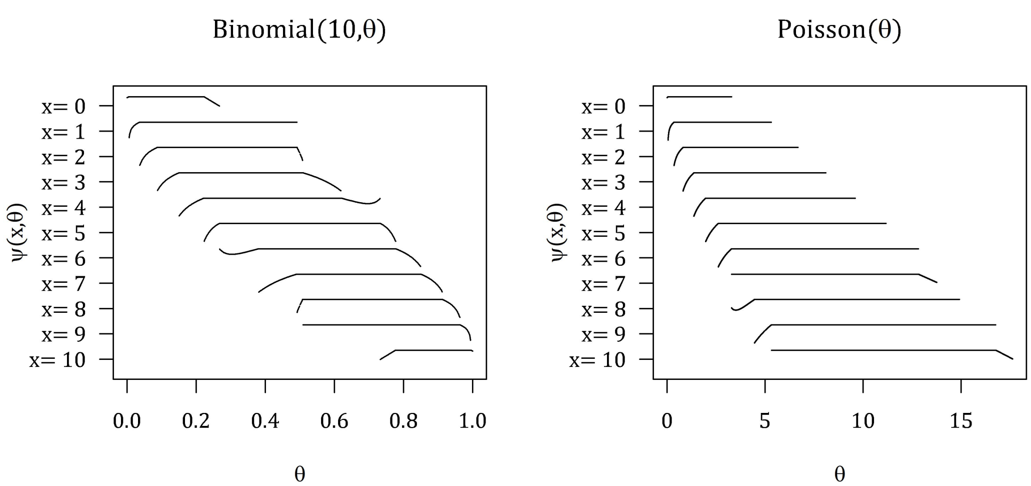

Figure 3.8: 95% fuzzy CG confidence procedures of the binomial$(10, \theta)$ and Poisson(θ) distributions. Only the nonzero portions of the intervals are plotted.

Notice the *curls* that occur in the fuzzy edges of some intervals. The curls are most prominent in the right edge of $x = 4$ (left edge of $x = 6$) in the Binomial case and the left edge of $x = 8$ in the Poisson case. Such curls occur when the following conditions are true. First in the Poisson case there would need to be a Type II transition between acceptance curves $P_\theta(a \le X \le b)$ and $P_\theta(a+1 \le X \le b+1)$ that has occurred to the left of the maximum of $P_\theta(a+1 \le X \le b+1)$. Then if either γ_i is set to 0 for all θ between the intersection of these two curves and the maximum of $P_\theta(a+1 \le X \le b+1)$, then the other γ_i will necessarily be increasing (as θ increases) to counteract the fact that the CPF is increasing here. Curls occur for the analogous reason in the Binomial case.

Figure 3.9 shows the 95% MST fuzzy confidence procedures for the binomial$(10, \theta)$ and Poisson(θ) distributions.

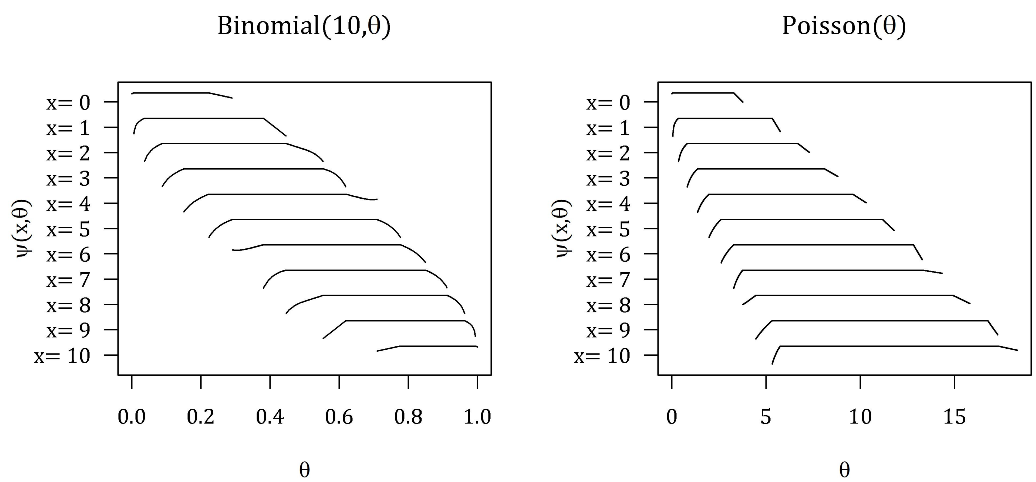

Figure 3.9: 95% fuzzy MST confidence procedures of the binomial$(10, \theta)$ and Poisson(θ) distributions. Only the nonzero portions of the intervals are plotted.

In the Poisson case, MST has Type II transitions between acceptance curves $P_\theta(a \le X \le b)$ and $P_\theta(a+1 \le X \le b+1)$ at the maximum of $P_\theta(a+1 \le X \le b+1)$. Hence the fuzzy version of MST will never have curls. By the same token any minimal cardinality procedure whose Type II transitions are to the right of MST (e.g. BK) will avoid curls (for the Poisson case). Note that in the Binomial case, curls cannot be avoided in general as the range of possible values for Type II transitions often do not include the maximum of the curve of interest. However, the length of the curl can be reduced by a incurring the latest (earliest) Type II transition when $\theta \leq 0.5$ ($\theta > 0.5$). Thus, BK produces the shortest curls of any of the minimal cardinality procedures in the Binomial case. Moreover, you'll notice the curl in the right edge of $x = 4$ (left edge of $x = 6$) is less noticeable for MST than for CG (in the Binomial$(10, \theta)$ case) due to having later (earlier) Type II transition when $\theta \leq 0.5$ ($\theta > 0.5$).

3.5 Support Comparison for Fuzzy Intervals

We now compare the support lengths of the fuzzy UMAU, fuzzy CP, and proposed fuzzy minimal cardinality intervals. Figure 3.10 compares the support lengths of the 95% MST, CP, and UMAU fuzzy procedures for the Binomial$(10, \theta)$ case. The plot on the left shows the supports for each procedure. The plot on the right shows the support lengths at each x for each procedure. Since all minimal cardinality procedures have the same interval length and hence the same support length we need only include one of these procedures (namely MST) in the plot. As can be seen MST performs best on length followed by CP. The UMAU supports are even wider than CP, which is notorious for producing exceptionally wide conventional (crisp) intervals.

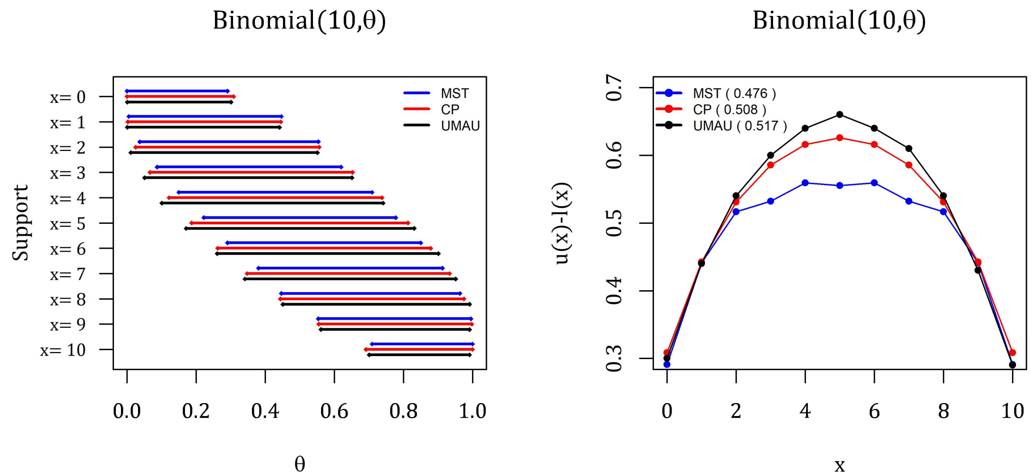

Figure 3.10: Comparison of the support lengths of the 95% MST, CP, and UMAU fuzzy procedure for the Binomial$(10, \theta)$ case. On the left is a plot of the supports for each procedure. On the right is a plot of the support lengths at each x for each procedure. The legend shows the average support length.

Figure 3.11 compares the support lengths of the 95% CG MST, CP, and UMAU fuzzy procedure for the Poisson(θ) case. The plot on the left shows the supports for each procedure. The plot on the right shows the support lengths at each x for each procedure.

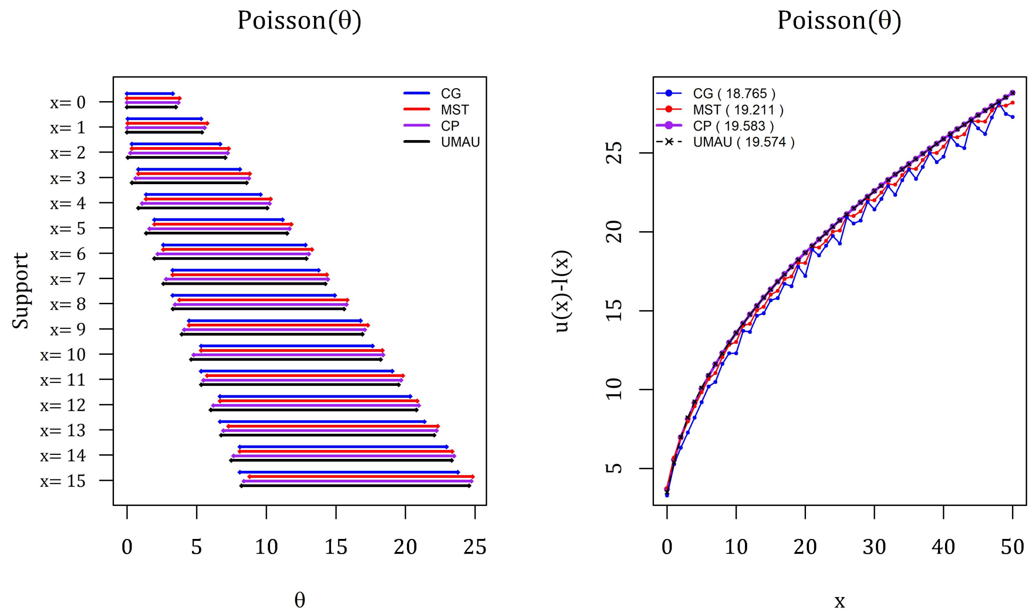

Figure 3.11: Comparison of the support lengths of the 95% CG, MST, CP, and UMAU fuzzy procedure for the Poisson(θ) case. On the left is a plot of the supports for each procedure. On the right is a plot of the support lengths at each x for each procedure. The legend shows the average support length for $x \in \{0, 1, ..., 50\}$.

As expected CG is best on length followed by MST. Except for the first few values of x, the CP and UMAU intervals have almost indistinguishable support lengths; moreover, the CP supports seems be shifted slightly right of the UMAU supports.

3.6 Discussion: Fuzzy Intervals

Using the words of Stevens (1950)[21], use of a conventional procedure is a choice to use an "inefficient" method of estimation over an "efficient" one. However, when the alternative is randomization it can be difficult to bring practitioners on board. First and foremost because two researchers using the same confidence level, data, and method may come to two different intervals. Geyer and Meeden (2005)[18] proposed a new method of interval estimation by graphing the membership function, that would otherwise, in combination with random number generation, produce randomized intervals. By plotting these membership functions for each x in their functional forms we avoid the need for random number generation and obtain a new type of interval estimator called a fuzzy confidence interval.

Geyer and Meeden (2005)[18] proposed fuzzy intervals that are derived from UMPU test and thus are UMAU. Therefore they are length optimal among unbiased procedures. However by relaxing the unbiased condition, shorter fuzzy intervals can be obtained. One way to produce short fuzzy intervals is by transforming minimal cardinality conventional procedures into fuzzy procedures. The process works by first considering each conventional interval of the procedure as a crisp fuzzy interval. Then we essentially bend the edges of the crisp interval downwards until an exact $(1-\alpha)$ coverage probability is obtained for all $\theta \in \Theta$. Consequently, we produce a fuzzy procedure whose support length matches that of the original procedure. Thus, in the Binomial case all minimal cardinality procedures produce fuzzy procedures that are equal on length; however, we saw BK will produce the shortest curls due to the location of its Type II transitions. In the Poisson case CG is shortest, but produces the longest curls of any minimal cardinality procedure due to having the earliest Type II transitions. MST then produces the shortest fuzzy procedure without curls in this case.

Agresti & Gottard (2005) (in published comments on Geyer and Meeden (2005)[18]) state: “There is no reason, however, that a statistical procedure needs to be unbiased to have good practical performance.” Unbiasedness is desirable but, so is shortness. Though we do not wish to take sides on the matter of unbiasedness, once the requirement is relaxed, we open ourselves up to a whole range of possibilities. Our goal here was to present all the options so that the reader can make their own informed decision on which method to use in practice. Additionally, since Geyer & Meeden’s (2005)[18] paper focused on the binomial case we hoped to shed some more light on the other distributions such as the Poisson distribution.

3.7 One-Sided Intervals: Conventional, Randomized, and Fuzzy

In this section we discuss one-sided confidence intervals, both conventional and randomized/fuzzy.

3.7.1 Conventional One-Sided Intervals

One-sided confidence intervals have the form $[l(X), \sup\Theta]$ or $[\inf\Theta, u(X)]$ where the lower and upper bounds $l(X)$ and $u(X)$ satisfy $\mathbb{P}_\theta\left[l(X) \geq \theta\right] \geq 1-\alpha$ and $\mathbb{P}_\theta\left[U(X) \leq \theta\right] \geq 1-\alpha$ respectively. Similarly to the way we obtained the upper and lower limits for CP, the lower and upper bounds $l(x)$ and $u(x)$ can be found by solving each of the equations $\mathbb{P}_\theta(X \geq x) = \alpha$ and $\mathbb{P}_\theta(X \leq x) = \alpha$ for θ. The resulting $(1-\alpha)$ lower and upper bounds are provided in Table 3.1 for the Binomal, Poisson, and NB distributions.

Distribution	$l(x)$	$u(x)$
Binomial(n,θ)	$B_\alpha(x, n-x+1)$	$B_{1-\alpha}(x+1, n-x)$
Poisson(θ)	$\frac{1}{2}\chi^2_{2x,\alpha}$	$\frac{1}{2}\chi^2_{2(x+1),1-\alpha}$
NB(r,θ)	$\frac{1}{1+\frac{x+1}{r}F^{2(x+1)}_{2r,1-\alpha}}$	$\frac{\frac{r}{x}F^{2r}_{2x,1-\alpha}}{1+\frac{r}{x}F^{2r}_{2x,1-\alpha}}$

Table 3.1: One-Sided Confidence Intervals

Because one-sided confidence intervals produce rejection regions made-up of $x's$ from only one of the tails of the distribution, minimal cardinality acceptance regions are unique in this case and the corresponding lower and upper bounds are none other than those already provided in Table 3.1.

3.7.2 Randomized and Fuzzy One-Sided Intervals

Recall that in general, one-sided tests lead to one-sided intervals. Likewise, uniformly most powerful (UMP) tests lead to uniformly most accurate (UMA) intervals. Hence, we start our discussion of randomized/fuzzy one-sided intervals with some standard theory on one-sided UMP tests.

Theorem 3.3 Let the random variable X have PMF of the form,

$$p_X(x|\theta) = h(x)c(\theta)e^{\theta T(x)}; \tag{3.24}$$

i.e., $\{p_X(x|\theta) : \theta \in \Theta\}$ is a one-parameter exponential family. Then there exists a UMP test for testing $H_0 : \theta \leq \theta_0$ against $H_a : \theta > \theta_0$, , which is given by

$$\phi(x, \theta_0) = \begin{cases} 1 & T(x) > C \\ \gamma & T(x) = C \\ 0 & T(x) < C \end{cases}, \tag{3.25}$$

where C and γ are determined by the level condition,

$$\mathbb{E}_{\theta_0}[\phi(X, \theta_0)] = \alpha \tag{3.26}$$

(Lehmann & Romano (2005)[20], p. 67). The analogous result holds for testing $H_0 : \theta \geq \theta_0$ against $H_a : \theta < \theta_0$. Since both the binomial and Poisson distributions are one-parameter exponential families with $T(x) = x$ they have UMP randomized tests as given by (3.25)-(3.26). Furthermore, when $T(x) = x$ the level condition becomes,

$$\alpha = \mathbb{E}_{\theta_0}[\phi(X, \theta_0)] = \gamma \mathbb{P}_{\theta_0}(X = C) + \mathbb{P}_{\theta_0}(X > C) \tag{3.27}$$

Thus C will be the unique value of x satisfying,

$$\mathbb{P}_{\theta_0}(X \geq x + 1) \leq \alpha/2 < \mathbb{P}_{\theta_0}(X \geq x). \tag{3.28}$$

After finding C, γ is determined by,

$$\gamma = \frac{\alpha - \mathbb{P}_{\theta_0}(X > C)}{\mathbb{P}_{\theta_0}(X = C)}. \tag{3.29}$$

The process for determining the corresponding values of C and γ is similar for testing $H_0 : \theta \geq \theta_0$ against $H_a : \theta < \theta_0$. The membership function $\psi(x, \theta) = 1 - \phi(x, \theta)$ as a function of θ determines the UMP one-sided fuzzy interval for each $x \in \mathcal{X}$. Note C is chosen in the same manner as the boundary of the acceptance regions for the conventional one-sided intervals discussed in the previous section. Thus, each of these fuzzy intervals will have support equal to the corresponding conventional one-sided interval. As a result, in the one-sided case, there is no need to transform our minimal cardinality procedures into fuzzy ones as they would produce fuzzy intervals identical to the corresponding UMP intervals. Figure 3.12 left (right) shows both the 95% conventional lower (upper) one-sided intervals of Table 3.1 in black solid lines, as well as, the corresponding UMP one-sided fuzzy intervals in blue dashed lines for the binomial$(10, \theta)$ case.

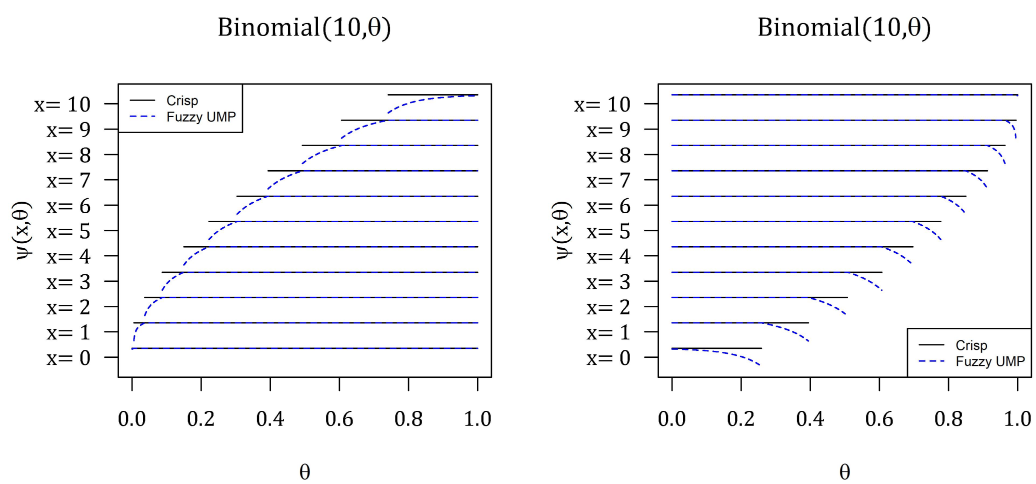

Figure 3.12: 95% one-sided fuzzy confidence procedures (blue dashed) derived from the UMP test for the binomial$(10, \theta)$ distribution. The left plot shows the lower one-sided intervals and the right plot shows the upper one-sided intervals. We also include the corresponding crisp one-sided confidence intervals of Table 3.1 (black solid).

As can be seen the support of the UMP fuzzy intervals match the support of the crisp conventional intervals. The shape of the fuzzy edges are not too different from those of the two-sided case. We also include similar plots for the 95% Poisson case in Figure 3.13. Only the intervals up to $x = 10$ are plotted. Note in Poisson case the support of lower one-sided intervals are of the form $[l(x), \infty)$. In this case intervals with larger values of $l(x)$ are considered shorter, despite the fact that all intervals are of infinite length.

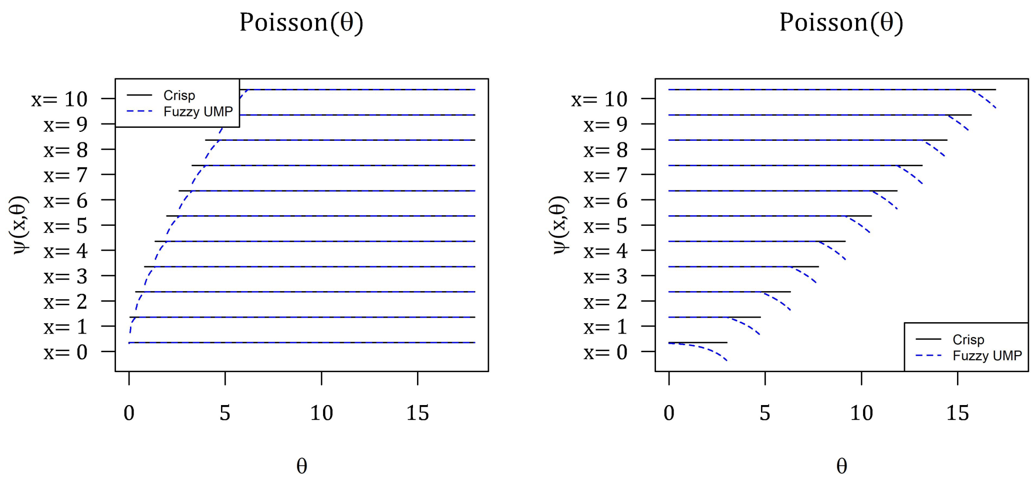

Figure 3.13: 95% one-sided fuzzy confidence procedures (blue dashed) derived from the UMP test for the Poisson(θ) distribution. The left plot shows the lower one-sided intervals and the right plot shows the upper one-sided intervals. We also include the corresponding crisp one-sided confidence intervals of Table 3.1 (black solid). Only intervals up to $x = 10$ are plotted.

Chapter 4

Confidence Intervals when the Parameter Space is Countable

In this chapter we consider estimation of parameters from a countable parameter space. In particular we consider the binomial with n unknown while p known, the negative binomial (NB) with r unknown, but p known, and the hypergeometric (HG) with each of its parameters singly unknown (see Table 4.1). Here the hypergeometric distribution gives the probability of drawing x units possessing a particular attribute when we randomly sample n units without replacement from a finite population of size N; M represents the total number of units in the population possessing the attribute of interest. As before we choose a negative binomial random variable to be the number of failures in successive Bernoulli trials before the rth success. Note that in this chapter when the probability of success in a Bernoulli trial is known we denote the probability by p rather than θ.

Distribution	$p(x\|\theta)$	$\mathcal{X}$	Parameter Restrictions
$HG(N, M, n)$	$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$\max\{0, n+M-N\} \leq x \leq \min\{M, n\}$	$\{N, M, n \in \mathbb{Z} : 0 \leq M, n \leq N\}$
Binomial(n, p)	$\binom{n}{x}p^x(1-p)^{n-x}$	$\{0, 1, 2, ..., n\}$	$p \in [0, 1]$, $\{n \in \mathbb{Z} : n \geq 0\}$
NB(r, p)	$\binom{x+r-1}{x}p^r(1-p)^x$	$\{0, 1, 2, ...\}$	$p \in [0, 1]$, $\{r \in \mathbb{Z} : r \geq 1\}$

Table 4.1: Hypergeometric (HG), binomial, and negative binomial (NB) distributions

Since we are now interested in estimating integer-valued parameters, we consider (countable) confidence sets over integers, rather than (uncountable) intervals over all real values. Thus, in the context of this section an interval $[l, u]$ denotes the set $\{l, l+1, \ldots, u-1, u\}$. Our goal here is to construct confidence intervals for each of the following 5 cases: HG with each of N, M, and n unknown, binomial with n unknown, and NB with r unknown. To start, we provide some examples where we might want to estimate the aforementioned parameters and also include the natural point estimates in each case:

- **HG with N unknown**: N is often of interest in cases where the capture-recapture method is used. For example, an ecologist might *capture*, mark and release M fish in a lake. Later the ecologist *recaptures* a random sample of size n and records the number of fish X from the sample who were previously marked. The ecologist then wishes to estimate the total number of fish N in the lake. The proportion of marked fish in the sample $\hat{p} = X/n$ can then be matched to the proportion of marked fish from the population $p = M/N$. Thus, it's reasonable to estimate that N will be near Mn/X.

- **HG with M unknown**: An example where M is of interest occurs when we wish to estimate the total number of units from a finite population possessing a particular characteristic. Since it may be costly or even impossible to sample the entire population, we take a sample of size $n \leq N$ (without replacement) and measure the number of units X from the sample possessing the characteristic. A natural point estimator for the proportion $p = M/N$ possessing the characteristic is then $\hat{p} = X/n$ and thus we might expect that M will be somewhere near $N\hat{p}$.

- **HG with n unknown**: Suppose there are N fish in a lake. Suppose further M fish are captured, marked and released back into the population. An unknown number n of the fish migrate downstream. A device is placed downstream to count the number of marked fish X participating in the migration. Assuming the fish migrating downstream constitute a random sample taken from the population we then wish to estimate the total number of fish n which migrated downstream. We can reasonably conclude n to be near XN/M.

- **Binomial with n unknown**: Suppose an unknown number n fish migrate downstream. Suppose it is known that the proportion of fish in the population which like a particular brand of bait, to be some known value p. A trap containing the bait is placed downstream to capture the fish. The number of fish X caught in the trap is then observed. Assuming each fish is trapped with probability p independent of all other fish, we wish to estimate n. We can reasonably estimate n to be near X/p.

- **NB with r unknown**: Woodpeckers search for bugs by making tiny holes in the tree bark. Suppose it is known that any particular hole will contain an insect with probability p. A woodpecker will continue searching for bugs until satiated. Once the woodpecker's search is over an ecologist counts the number of holes drilled X and wishes to estimate the total number of insects r that the woodpecker consumed; i.e., how many insects it takes to make the woodpecker full. Since we expect $p \approx r/(r + X)$, it's reasonable to conclude that r should be near $Xp/(1 - p)$.

Note that these five cases can actually be reduced down to four due to the following proposition on the duality of M and n for the hypergeometric distribution:

Proposition 4.1 $\text{HG}(N, M, n)$ and $\text{HG}(N, n, M)$ have the same probability distribution.

The result can easily be checked after expanding the factorials of the corresponding probability mass functions. Since $\text{HG}(N, M, n)$ and $\text{HG}(N, n, M)$ are equivalent distributions the confidence procedure for n in $\text{HG}(N_0, M_0, n)$ is equal to the confidence procedure for M in $\text{HG}(N_0, M, M_0)$. Hence once a procedure has been developed for M nothing needs to be done for the n case because we can simply exchange the value of n and M in any of our methods for M, to produce the appropriate confidence procedure for n.

As with the binomial distribution where we require equivariance in p, it is important for a HG confidence procedure to be equivariant in M, since switching the roles of 'success' and 'failure' switches the roles of x and $n-x$ (here success means drawing a unit possessing the attribute of interest). This follows from the fact that if X is $\text{HG}(N, M, n)$ then $n - X$ counts the number of failures in the sample and is $\text{HG}(N, N - M, n)$.

Property 4.1 (Equivariance in M for HG Procedures) A HG confidence procedure $C(x) = [l(x), u(x)]$ for M is *equivariant* if $n - x$ generates the confidence interval $[N - u(x), N - l(x)]$

Note that such a requirement forces the CPF of a confidence procedure for M to be symmetric around $N/2$ as will be seen later in our CPF plots.

4.1 Clopper and Pearson (CP)

The analogs of Clopper & Pearson's (1934)[7] method (CP) can found in all four cases. We include some example CPF plots in Figures 4.1 and 4.2 for the 90% HG($N, 15, 10$), HG($75, M, 20$), binomial($n, 0.35$), and NB($r, 0.84$) cases. Note that the acceptance curves here are not quite "curves" at all since they are only defined on the integers.

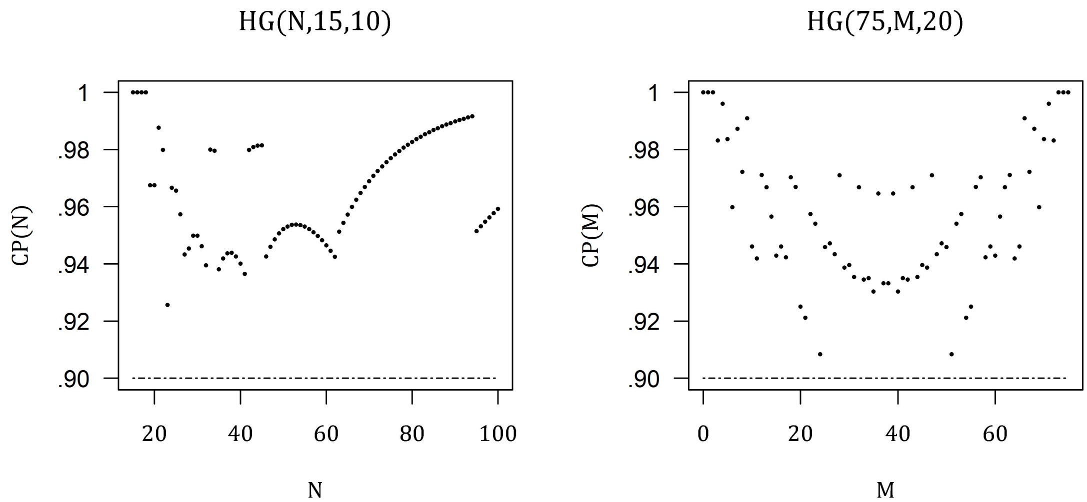

Figure 4.1: 90% CPFs of the Clopper & Pearson method for HG($N, 15, 10$) (left) and HG($75, M, 20$) (right) cases. For HG($N, 15, 10$) case only CPF values for N up to 100 are shown.

In the HG($N, 15, 10$) case the CPF behaves quite wildly. Additionally only values up to $N = 100$ are plotted, although the domain of N is $N \geq \max\{n, M\}$. Notice in the HG($75, M, 20$) case the CPF is symmetric due to equivariance; the domain of the CPF in this case is $0 \leq M \leq N$. In both the binomial($n, 0.35$) and NB($r, 0.84$) cases the CPF is plotted only up to 50 although the domains of the CPF's are not bounded above. Lastly observe that the CPF's in each of the four cases are quite conservative.

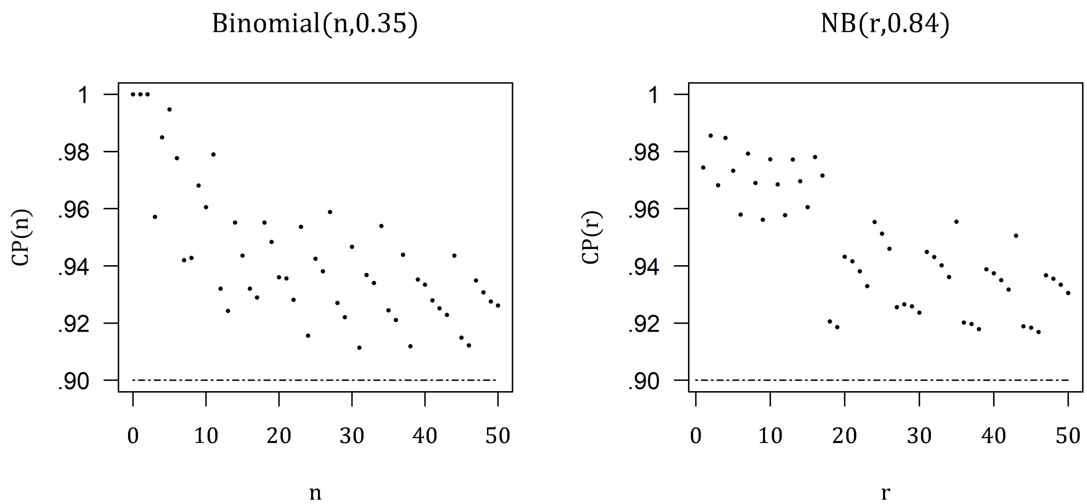

Figure 4.2: 90% CPFs of the Clopper & Pearson method for the binomial($n, 0.35$) (left) and NB($r, 0.84$) cases.

4.2 Blaker

We can also construct the analogs of Blaker (2000)'s method [16] (B). The 90% CPF's are shown in Figures 4.3 and 4.4 for the HG($N, 15, 10$), HG($75, M, 20$), binomial($n, 0.35$), and NB($r, 0.84$) cases.

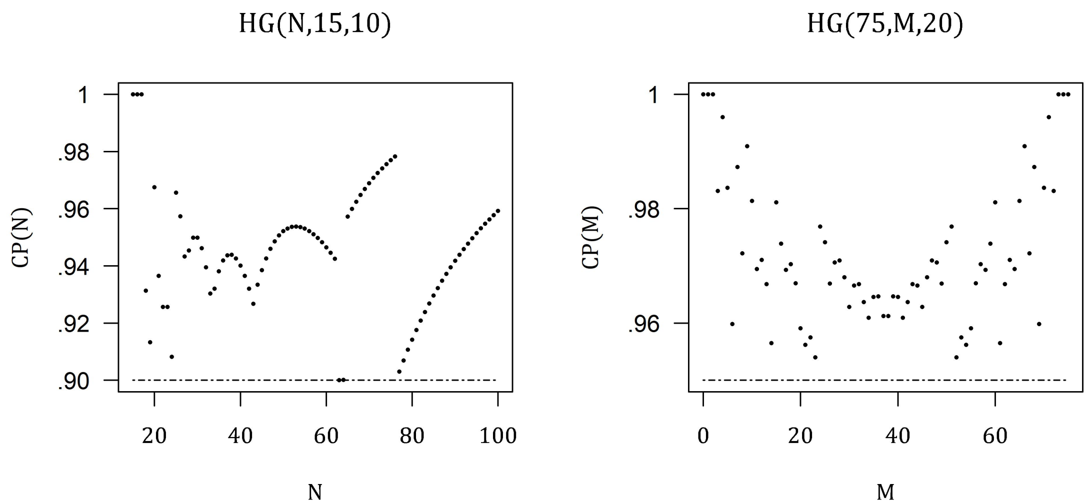

Figure 4.3: 90% CPFs of the Blaker's method for HG(N, 15, 10) (left) and HG(75, M, 20) (right) cases. For HG(N, 15, 10) case only CPF values for N up to 100 are shown.

The CPF's for B maintain a similar shape structure to the corresponding ones for CP; however, the CPF's for B are much less conservative, ultimately leading to shorter intervals in each case.

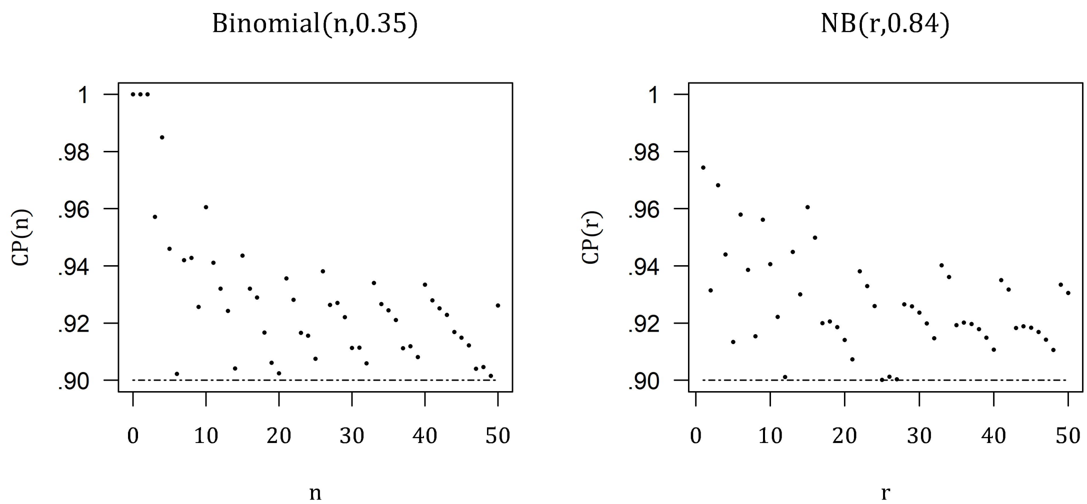

Figure 4.4: 90% CPFs of the Blaker's method for binomial($n, 0.35$) (left) and NB($r, 0.84$) (right) cases. For HG($N, 15, 10$) case only CPF values for N up to 100 are shown.

In the negative binomial case Blaker's method will sometimes produce gaps. To see this take a look at Figure 4.5, which shows the 90% CPF of B for the NB($r, 0.90$) case. The red crosses in the figure indicate the region that causes a gap. Notice the CPF transitions from $0-6$, to $1-6$, and back to $0-6$ again, causing a gap in the confidence interval for $x = 0$. This is because the probability of observing a value of X with a min-tail probability as small as that of $x = 0$ is greater than 0.10 when $r = 26$ and 28, but not when $r = 27$.

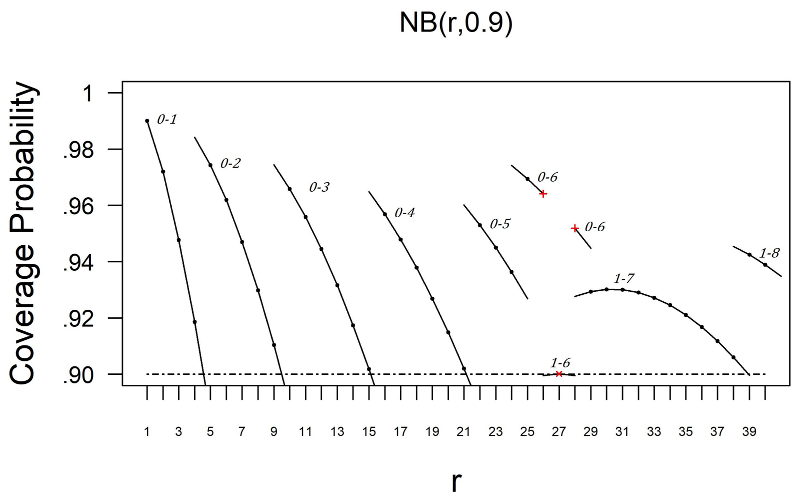

Figure 4.5: Illustration of an occurrence in Blaker's 90% CPF for the NB$(r, 0.90)$ that produces a gap in the corresponding procedure. Red crosses highlight the region that cause the gap.

4.3 Minimal Cardinality Procedures

In this section we examine the class of minimal cardinality procedures in each case and look for the optimal procedure under our various criteria. In particular we will discuss the analogs of Sterne, MST, CG, and BK and discuss which of the four methods is "best" for each of the four cases. We approach each of the four problems in a similar manner as before, by considering only confidence sets which correspond to acceptance regions of minimal size. By then plotting the eligible acceptance curves we can better understand which series of acceptance regions produce "optimal" confidence sets in each case, where optimal may be user-defined.

4.3.1 HG with N unknown

Figure 4.6 shows all 90% minimal cardinality acceptance curves in the $\text{HG}(N, 15, 10)$ case, except those that would cause gaps. We do however include the curves (in black dashed lines) that cause gaps in Sterne's method. For better visualization we plot the CPF on $[15, 100]$ in a separate plot due to the high concentration of curves in this region. We plot CPF for $100 \leq N \leq 1600$ is second plot. Additionally, to better understand the behavior of the acceptance curves, lines are placed between adjacent points $P_N(a \leq X \leq b)$, $P_{N+1}(a \leq X \leq b)$ to indicate that the two points lie on the same acceptance curve.

Unlike the Poisson(θ) and binomial(n, θ) (n known) cases small values of x here point to large parameter values. Thus the CPF starts (reading from left to right) with acceptance curves involving the largest values of x first. Furthermore the CPF appears at first glance to be a rather arbitrary configuration of curves. There isn't quite the same repetitive structure as seen in the CPF plots for the other cases. Note that different choices for M and n will produce similarly structured CPF's.

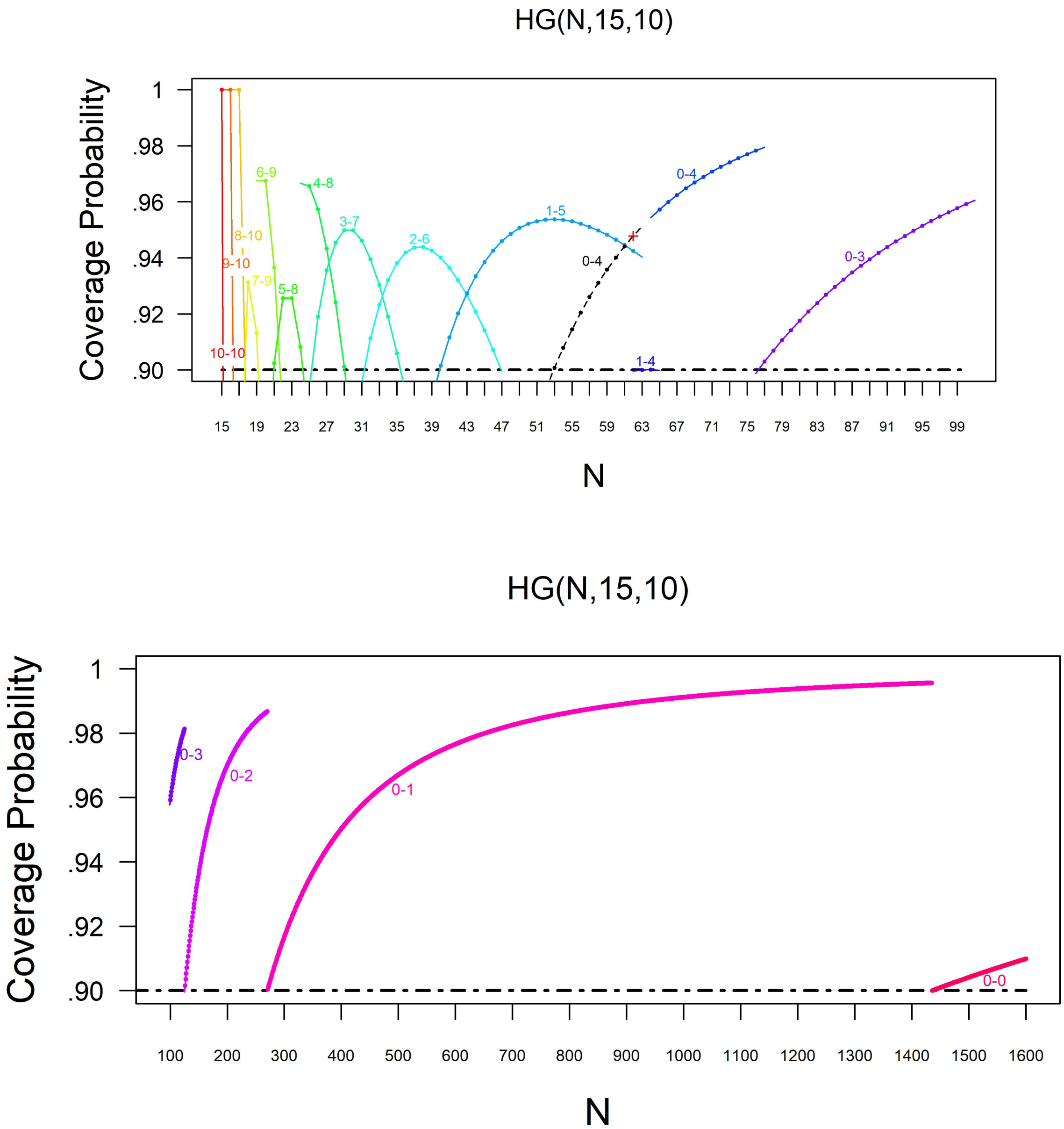

Figure 4.6: All 90% minimal cardinality acceptance curves in the $\mathrm{HG}(N, 15, 10)$ case, except those that would cause gaps. We do however include the curves (in black dashed lines) that would cause gaps in Sterne's method. We label acceptance curves $P_N(a \leq X \leq b)$ by a-b. Colors are used to better distinguish nearby curves. For better visualization we plot the CPF on $[15, 100]$ in a separate plot due to the high concentration of curves in this region. This figure only includes coverage values up to $N = 1600$. To better understand the behavior of the acceptance curves, lines are placed between adjacent points $P_N(a \leq X \leq b)$, $P_{N+1}(a \leq X \leq b)$ to indicate that the two points lie on the same acceptance curve.

For significantly large N the only reasonable value for X is 0 as drawing a unit possessing the attribute of interest in this case would be much like finding a needle in a haystack. For N is the thousands only the acceptance curves $0-0$ and $0-1$ make the cut. As N gets closer to M we would expect to observe a larger number of units possessing the attribute of interest; hence, the remaining values of x are more likely to be observed, causing a higher concentration of curves for $N \leq 100$.

We now describe how the analogs of ST, MST, CG, and BK choose their acceptance curves. For each N ST chooses from among all minimal span curves the one with highest coverage. For the remaining methods the selection of curves is reduced to those minimal span curves $\{P_N(a \leq X \leq b)\}$ that would keep the sequences of $\{a\}$ and $\{b\}$ values monotonic (when moving from right to left). These are the curves shown in Figure 4.6. Then, for each N, whenever there are multiple acceptance curves $\{P_N(a \leq X \leq b)\}$ available, CG chooses the one with largest value of a, BK chooses the one with smallest value of a, and MST chooses the one with highest coverage. The corresponding confidence intervals produced by these four approaches are shown in Table 4.2.

Note that ST and MST differ only when use of the highest minimal span curves cause the sequences of $\{a\}$ and $\{b\}$ values to be nonmonotonic. For the HG$(N, 15, 10)$ case this occurs when $N = 62$. Here ST uses the curve $0-4$ as it has higher coverage than $1-5$. This causes a gap in the interval for $x = 0$ and additionally causes the length of the interval for $x = 5$ to differ from MST by one unit as shown in red in Table 4.2.

	$C(x)$			
x	CG	MST	BK	ST
0	$[65, \infty)$	$[65, \infty)$	$[65, \infty)$	$\{62\} \cup [65, \infty)$
1	$[47, 1435]$	$[43, 1435]$	$[40, 1435]$	$[43, 1435]$
2	$[36, 270]$	$[34, 270]$	$[32, 270]$	$[34, 270]$
3	$[30, 125]$	$[28, 125]$	$[26, 125]$	$[28, 125]$
4	$[25, 76]$	$[25, 76]$	$[25, 76]$	$[25, 76]$
5	$[22, 62]$	$[22, 62]$	$[21, 62]$	$[22, 61]$
6	$[20, 46]$	$[20, 42]$	$[20, 39]$	$[20, 42]$
7	$[18, 35]$	$[18, 33]$	$[18, 31]$	$[18, 33]$
8	$[17, 29]$	$[17, 27]$	$[17, 25]$	$[17, 27]$
9	$[16, 21]$	$[16, 21]$	$[16, 20]$	$[16, 21]$
10	$[15, 17]$	$[15, 17]$	$[15, 17]$	$[15, 17]$

Table 4.2: 90% Confidence Procedures for HG$(N, 15, 10)$ distribution for CG, MST, BK and ST.

Because $A_N = \{0\}$ for significantly large N (specifically any $N > 1435$ in the HG$(N, 15, 10)$ case) the upper limit for $x = 0$ is ∞. As a result comparing interval width cannot be achieved through normal means such as comparing average length, despite the fact that there are only finitely many intervals to compare.

Definition 4.1 (Length Difference) Let $C(x) = [l(x), u(x)]$ and $C'(x) = [l'(x), u'(x)]$ be two different confidence procedures for N. Then we define the *length difference* D_L of $C(x)$ relative to $C'(x)$ to be,

$$D_L = \sum_x |C(x) - C'(x)| - |C'(x) - C(x)|, \tag{4.1}$$

so that D_L is a measure of the difference in total number of elements between the two procedures.

Note that here for any two sets A and B, $A - B$ denotes the set of elements from A that do not appear in B. Furthermore the operator $|\cdot|$ denotes the map which takes any set to its cardinality. $D < 0/D = 0/D > 0$ indicate that $C(x)$ has a lesser/equal/greater number of elements in it's intervals than $C'(x)$. Recall that different choices of minimal span acceptance curves do not affect the overall length of a procedure as these various choices would shorten (lengthen) one interval by the same amount it lengthens (shortens) another interval. Hence the value D produced from comparing any pair of these four minimal cardinality procedures will be zero.

We might also be interested in the difference in coverage of two procedures.

Definition 4.2 (Coverage Difference) Let $C(x) = [l(x), u(x)]$ and $C'(x) = [l'(x), u'(x)]$ be two different confidence procedures for N. Then we define the *coverage difference* D_C of $C(x)$ relative to $C'(x)$ by,

$$D_C = \sum_N CP_C(N) - CP_{C'}(N). \tag{4.2}$$

Recall that for the Poisson case comparing the coverage of two procedures over the entire parameter space is typically not possible because such comparisons need to be made on a infinite range. However, in this case (HG case with N unknown), although N unbounded, the CPF's of the minimal cardinality procedures agree for all large N and thus the comparison need only be made on a finite range. For instance in the HG$(N, 15, 10)$ case the CPF's agree for all $N \geq 63$ since only one choice of minimal span acceptance curve exists for all N in this range. Thus the sum would only need be considered for $15 \leq N \leq 62$. Alternatively we could compute the average coverage for $N \in \{15, 16, ..., 62\}$,

$$\sum_{N=15}^{62} CP(N)/48. \tag{4.3}$$

The resulting values are shown in Table 4.3 for CG, MST and BK. MST has the largest average coverage followed by CG.

Method	Avg. Cov.
CG	0.9425
MST	0.9464
BK	0.9424

Table 4.3: Average coverage for $N \in \{15, 16, ..., 62\}$, $\sum_{N=15}^{62} CP(N)/48$ for the 90% CG, MST, and BK procedures in the HG$(N, 15, 10)$ case.

4.3.2 HG with M unknown

Figure 4.7 is a plot of all 90% minimal cardinality acceptance curves for the HG$(75, M, 20)$ case, except those that would cause gaps. However, we do include the curves (in black dashed lines) that cause gaps in Sterne's method. The CPF behaves much like that of the binomial(n, θ) case (with n known and θ unknown). The CPF is symmetric around $N/2$ due to equivariance. Moreover, any two consecutive curves $P_M(a \leq X \leq b)$, $P_M(a+1 \leq X \leq b+1)$ of equal cardinality intersect to the left (right) of the maximum of $P_M(a+1 \leq X \leq b+1)$ ($P_M(a \leq X \leq b)$) when $M \leq N/2$ ($M \geq N/2$).

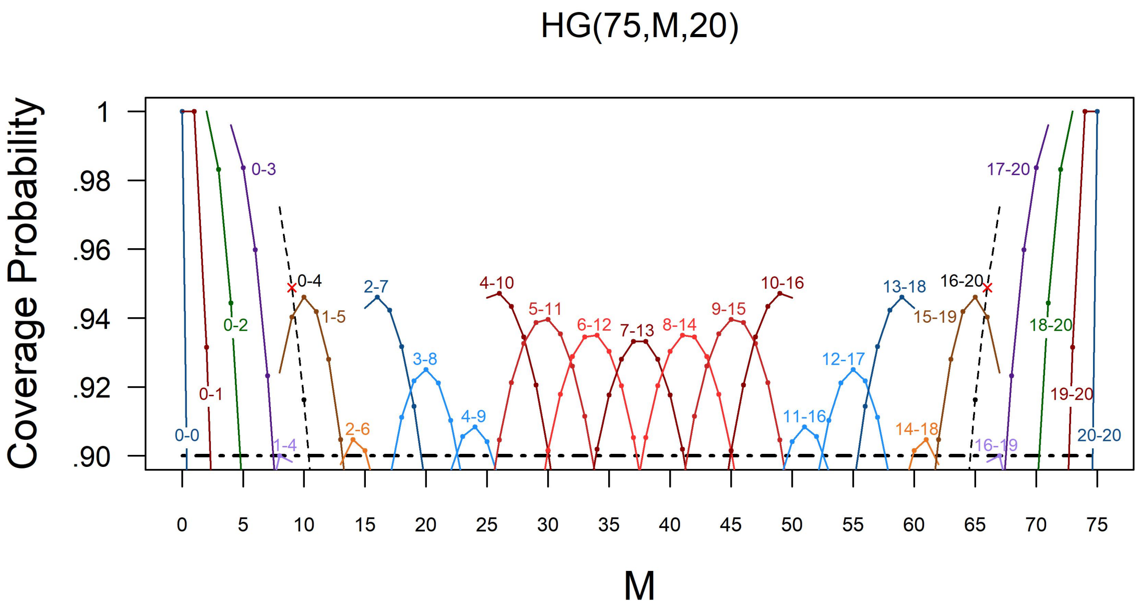

Figure 4.7: All 90% minimal cardinality acceptance curves in the HG$(75, M, 20)$ case, except those that would cause gaps. However, we do include the curves (in black dashed lines) that would cause gaps in Sterne's method. We label acceptance curves $P_M(a \le X \le b)$ by a-b. Colors are used to better distinguish nearby curves. Nearby curves that are of different shades of the same color indicate that the curves have equal cardinality. For better visualization lines are placed between adjacent points $P_M(a \le X \le b)$, $P_{M+1}(a \le X \le b)$ to indicate that the two points lie on the same acceptance curve.

The analogs of ST, MST, CG, and BK choose acceptance curves in the following way. For each M, ST chooses from among all minimal span curves the one with highest coverage. For the remaining methods the selection of curves is reduced to those minimal span curves $\{P_M(a \le X \le b)\}$ that would keep the sequences of $\{a\}$ and $\{b\}$ values monotonic (when moving from left to right). These are the curves shown in Figure 4.7. Then, for each M, whenever there are multiple curves $\{P_M(a \le X \le b)\}$ available, CG chooses the one with largest (smallest) value of a when $M \le N/2$ ($M \ge N/2$), BK chooses the one with smallest (largest) value of a when $M \le N/2$ ($M \ge N/2$), and MST chooses the one with highest coverage. The corresponding confidence intervals produced by these four approaches are shown in Table 4.4. The intervals for $11 \leq x \leq 20$ can be determined by equivariance.

	$C(x)$			
x	CG	MST	BK	ST
0	$[0,7]$	$[0,7]$	$[0,7]$	$[0,7] \cup \{9\}$
1	$[1,13]$	$[1,13]$	$[1,13]$	$[1,13]$
2	$[3,17]$	$[3,18]$	$[3,19]$	$[3,18]$
3	$[5,22]$	$[5,22]$	$[5,22]$	$[5,22]$
4	$[8,25]$	$[8,28]$	$[8,30]$	$[8,28]$
5	$[9,29]$	$[9,31]$	$[9,33]$	$[10,31]$
6	$[14,33]$	$[14,35]$	$[14,37]$	$[14,35]$
7	$[16,41]$	$[16,39]$	$[16,37]$	$[16,39]$
8	$[18,45]$	$[19,43]$	$[20,41]$	$[19,43]$
9	$[23,49]$	$[23,46]$	$[23,44]$	$[23,46]$
10	$[26,49]$	$[26,49]$	$[26,49]$	$[26,49]$

Table 4.4: 90% confidence procedures for the HG$(75, M, 20)$ distribution for CG, MST, BK, and ST. The missing intervals $C(x)$ for $11 \leq x \leq 20$ can be determined through equivariance.

When $M = 9$ ST uses the acceptance curve $0 - 4$ as it is the minimal span curve with maximum coverage. As a result ST produces a gap in the confidence interval for $x = 0$ and will additionally produce an interval for $x = 5$ that is of one unit shorter than that of MST, as shown in Table 4.4. Due to equivariance the analogous thing occurs at $M = 66$.

Since confidence procedures for M contain only finitely many intervals, each of finite length, we may compare the lengths of confidence procedures by their average interval length,

$$\sum_{x=0}^{n} (u(x) - l(x))/(n+1). \tag{4.4}$$

Since the different choices among minimal cardinality acceptance curves do not affect the overall length of the procedures all minimal cardinality procedures will have equal average interval width here. Therefore, we might instead be interested in comparing the average coverage of the procedures,

$$\sum_{M=0}^{N} CP(M)/(N+1). \tag{4.5}$$

Table 4.5 shows the average coverage of CG, MST, and BK for the HG$(75, M, 20)$ case. MST has the highest coverage followed by BK.

Method	Avg. Cov.
CG	0.9303
MST	0.9352
BK	0.9319

Table 4.5: Average coverage $\sum_{M=0}^{75} CP(M)/76$ for the 90% CG, MST, and BK procedures in the HG$(75, M, 20)$ case.

4.3.3 Binomial with n unknown

Figure 4.8 shows all 90% minimal cardinality acceptance curves in the binomial$(n, 0.35)$ case, except those that would cause gaps. However, we do include the curves (in black dashed lines) that would cause gaps in Sterne's method. The analogs of ST, MST, CG, and BK choose acceptance curves in the following way. For each n, ST chooses from among all minimal span curves the one with highest coverage. For the remaining methods the selection of curves is reduced to those minimal span curves $\{P_n(a \leq X \leq b)\}$ that would keep the sequences of $\{a\}$ and $\{b\}$ values monotonic (when moving from left to right). These are the curves shown in Figure 4.8. Then, for each n, whenever there are multiple curves $\{P_n(a \leq X \leq b)\}$ available, CG chooses the curve with largest value of a, BK chooses the one with smallest value of a, and MST chooses the one with highest coverage.

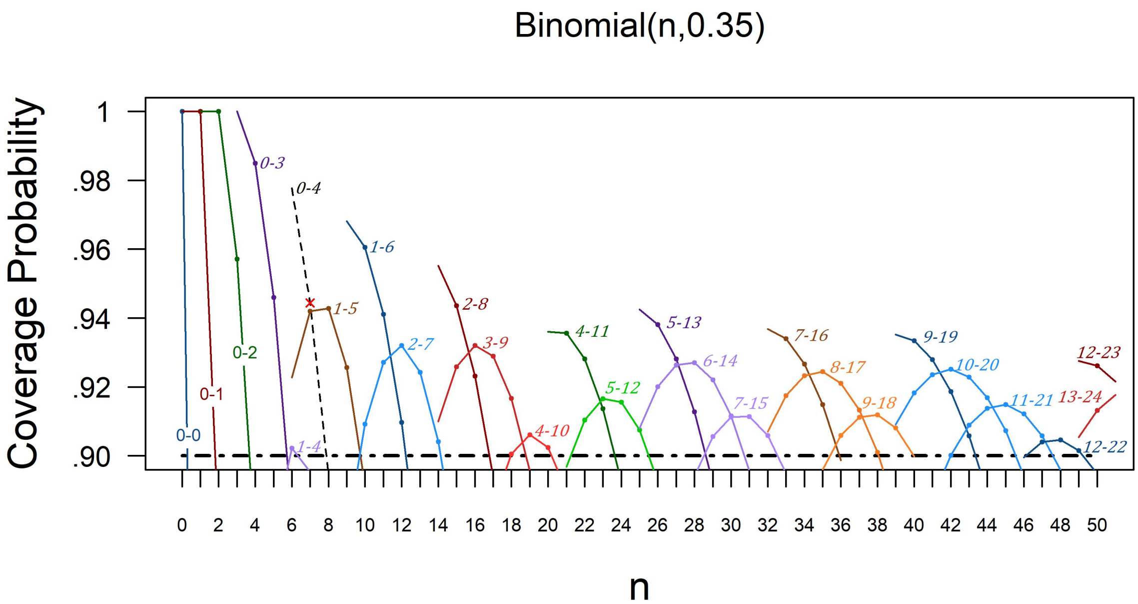

Figure 4.8: All 90% minimal cardinality acceptance curves in the binomial$(n, 0.35)$ case, except those that would cause gaps. However, we do include the curves (in black dashed lines) that would cause gaps in Sterne's method. We label acceptance curves $P_n(a \le X \le b)$ by a-b. Colors are used to better distinguish nearby curves. Nearby curves that are of different shades of the same color indicate that the curves have equal cardinality. For better visualization lines are placed between adjacent points $P_n(a \le X \le b)$, $P_{n+1}(a \le X \le b)$ to indicate that the two points lie on the same acceptance curve.

The corresponding confidence intervals produced by these four approaches are shown in Table 4.6 for $0 \le x \le 10$. We also include the corresponding intervals for CP and B. $n = 7$ is the only location that the CPF's for ST and MST disagree. By using the curve $0 - 4$ at $n = 7$, ST produces a gap in the confidence interval for $x = 0$ and additionally possesses a lower limit for $x = 5$ that is one unit larger than that of MST.

				$C(x)$		
x	CG	MST	BK	ST	B	CP
0	$[0,5]$	$[0,5]$	$[0,5]$	$[0,5]\cup\{7\}$	$[0,5]$	$[0,10]$
1	$[1,9]$	$[1,11]$	$[1,12]$	$[1,11]$	$[1,11]$	$[1,17]$
2	$[2,14]$	$[2,15]$	$[2,16]$	$[2,15]$	$[2,15]$	$[2,22]$
3	$[4,17]$	$[4,18]$	$[4,18]$	$[4,18]$	$[4,18]$	$[4,28]$
4	$[6,21]$	$[6,22]$	$[6,23]$	$[6,22]$	$[6,22]$	$[7,33]$
5	$[7,25]$	$[7,27]$	$[7,28]$	$[8,27]$	$[7,26]$	$[10,39]$
6	$[10,28]$	$[10,30]$	$[10,30]$	$[10,30]$	$[10,29]$	$[12,44]$
7	$[10,32]$	$[12,34]$	$[13,35]$	$[12,34]$	$[12,34]$	$[15,49]$
8	$[15,35]$	$[15,37]$	$[15,38]$	$[15,37]$	$[15,36]$	$[18,54]$
9	$[15,39]$	$[16,41]$	$[17,43]$	$[16,41]$	$[16,41]$	$[21,59]$
10	$[18,41]$	$[19,44]$	$[19,45]$	$[19,44]$	$[19,44]$	$[24,64]$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 4.6: 90% confidence procedures for the binomial$(n, 0.35)$ distribution for CG, MST, BK, ST, B, and CP. Only intervals up to $x = 10$ are included.

Because confidence procedures for n in the binomial case contain infinitely many intervals, comparing the interval width of two procedures is not quite as simple as computing the average interval width. Figure 4.9 is plot of the interval width rankings for CG, B, MST, BK, and CP with $p = 0.35$ and $(1 - \alpha) = .90$. For instance for $x = 5$ the methods are ranked CG, B, MST, BK, and CP in order of their interval length (shortest to widest). Overlapping points indicate a tie has occurred. CG tends to do best on length followed by B and MST. BK will often have wider intervals than even CP.

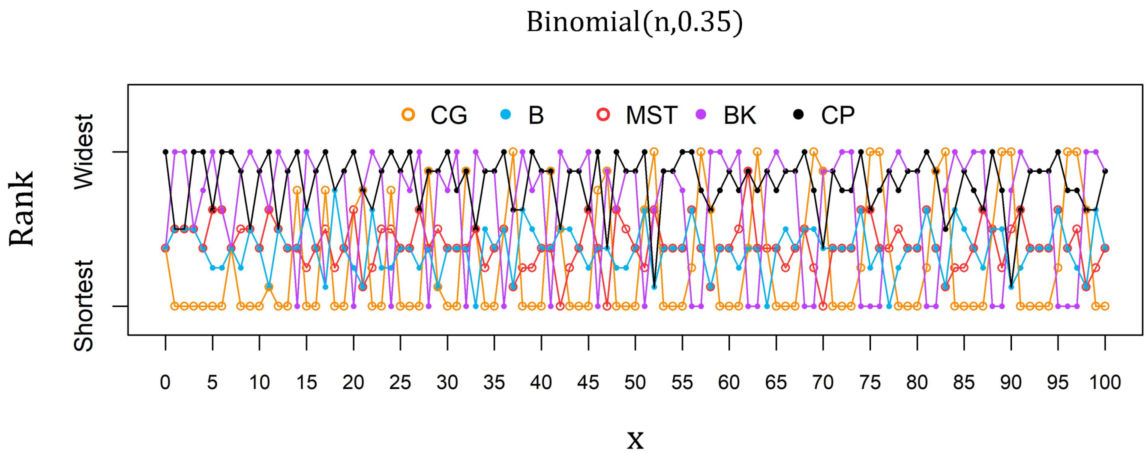

Figure 4.9: 90% binomial$(n, 0.35)$ confidence procedures ranked by confidence interval length for each x. Overlapping points indicate a tie has occurred for that x.

We might also be interested in the expected width,

$$\mathbb{E}_\theta\left[u(X) - l(X)\right] = \sum_{x=0}^{n} \left[u(x) - l(x)\right] \cdot P_n(X = x), \tag{4.6}$$

or the running average width,

$$\frac{1}{K} \sum_{x=0}^{K} (u(x) - l(x)) \tag{4.7}$$

(as a function of K). Figure 4.10 shows expected width and running average width for CG, MST, BK, B, and CP relative to MST for $p = 0.35$ and $(1 - \alpha) = .90$; i.e., each given value (expected width or running average width) for each procedure is divided by the corresponding value for MST to produce the values shown in the plots. CG clearly outperforms the other procedures as it has the smallest expected width and running average width of any procedure. B and MST are close for second place.

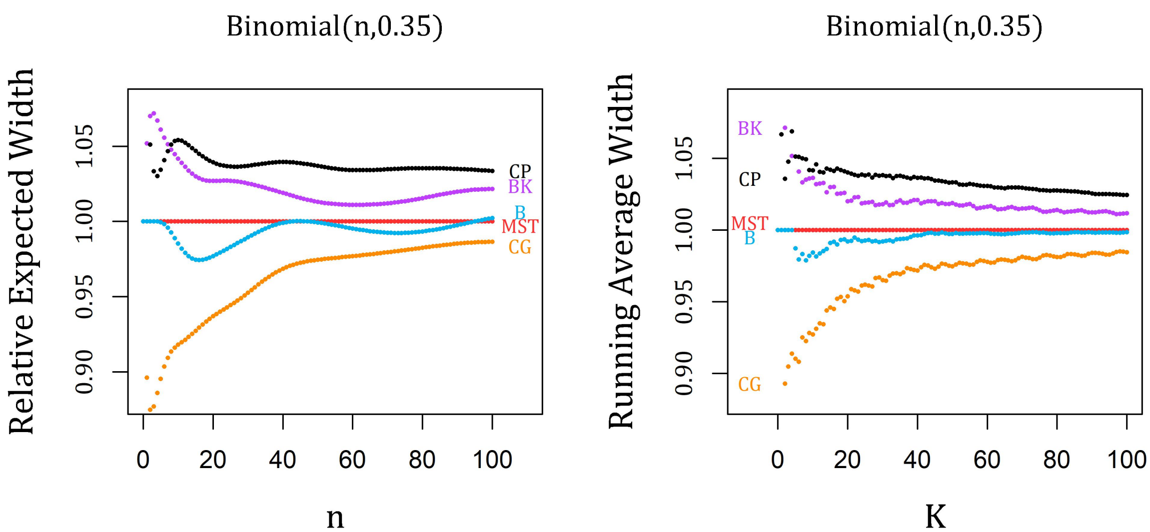

Figure 4.10: Plots expected width (left) and running average width (right) for CG, MST, BK, B, and CP relative to MST for the 90% binomial$(n, 0.35)$ case.

Figure 4.11 shows average confidence interval width for $x \in \{0, 1, ..., 100\}$ v.s. mean coverage on [0,100] for $(1 - \alpha) = .90$. The ideal procedure would be located in the upper left corner. Since we often prioritize length over coverage, the procedures furtherest to the left are preferred.

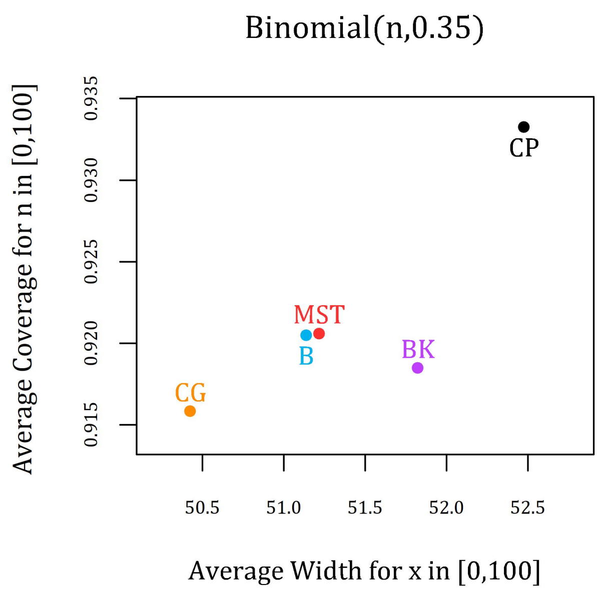

Figure 4.11: Plot of average confidence interval width for $x \in [0, 100]$ v.s. mean coverage on [0,100] for 90% binomial(n, 0.35) procedures.

Lastly, we may compare the interval length of two procedures under Schilling & Holladay's (2017)[22] asymptotic length criterion. Just like the Poisson case CG will be length optimal under this criteria due to choosing acceptance sets which prioritize the length of smaller x values.

4.3.4 NB with r unknown

Figure 4.12 shows all 90% minimal cardinality acceptance curves in the NB(r, 0.84) case except those that would cause gaps. However, we do include the curves (in black dashed lines) that would cause gaps in Sterne's method. The analogs of ST, MST, CG, and BK choose acceptance curves in a similar fashion to the binomial case with n unknown. For each r, ST chooses from among all minimal span curves the one with highest coverage. For the remaining methods the selection of curves is reduced to those minimal span curves $\{P_r(a \le X \le b)\}$ that would keep the sequences of $\{a\}$ and $\{b\}$ values monotonic (when moving from left to right). These are the curves shown in Figure 4.12. Then, for each r, whenever there are multiple curves $\{P_r(a \le X \le b)\}$ available, CG chooses the curve with largest value of a, BK chooses the one with smallest value of a, and MST chooses the one with highest coverage.

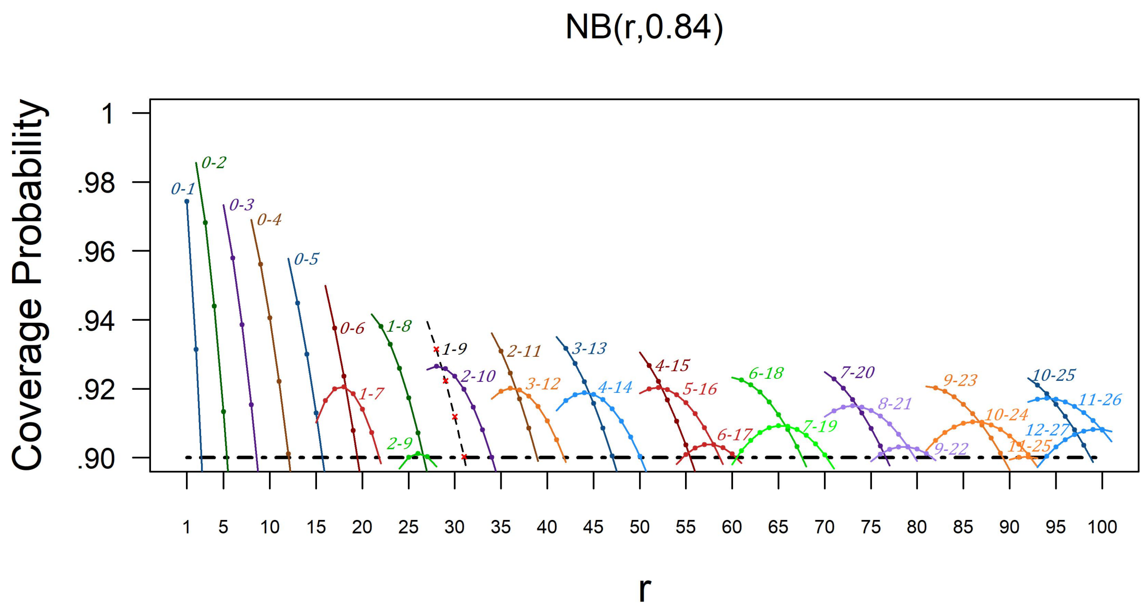

Figure 4.12: All 90% minimal cardinality acceptance curves in the NB$(r, 0.84)$ case except those that would cause gaps. However, we do include the curves (in black dashed lines) that would cause gaps in Sterne's method. We label acceptance curves $P_r(a \le X \le b)$ by a-b. Colors are used to better distinguish nearby curves. Nearby curves that are of different shades of the same color indicate that the curves have equal cardinality. For better visualization lines are placed between adjacent points $P_r(a \le X \le b)$, $P_{r+1}(a \le X \le b)$ to indicate that the two points lie on the same acceptance curve.

The corresponding confidence intervals produced by these four approaches are shown in Table 4.7 for $0 \le x \le 10$. We also include the corresponding intervals for CP and B. $n = 28$ is only location that the CPF's for ST and MST disagree. The intervals where ST differs from MST as a result of this disagreement are highlighted in red in Table 4.7.

	$C(x)$					
x	CG	MST	BK	ST	B	CP
0	$[1, 15]$	$[1, 18]$	$[1, 19]$	$[1, 18]$	$[1, 16]$	$[1, 17]$
1	$[1, 24]$	$[1, 26]$	$[1, 26]$	$[1, 26] \cup \{28\}$	$[1, 24]$	$[1, 26]$
2	$[1, 34]$	$[1, 36]$	$[1, 38]$	$[1, 36]$	$[3, 34]$	$[2, 35]$
3	$[6, 41]$	$[6, 44]$	$[6, 47]$	$[6, 44]$	$[6, 42]$	$[4, 43]$
4	$[9, 50]$	$[9, 52]$	$[9, 55]$	$[9, 52]$	$[9, 50]$	$[7, 50]$
5	$[13, 54]$	$[13, 58]$	$[13, 58]$	$[13, 58]$	$[13, 56]$	$[10, 57]$
6	$[16, 60]$	$[16, 65]$	$[16, 67]$	$[16, 65]$	$[15, 63]$	$[13, 64]$
7	$[16, 70]$	$[19, 73]$	$[20, 76]$	$[19, 73]$	$[17, 71]$	$[16, 71]$
8	$[22, 75]$	$[22, 79]$	$[22, 79]$	$[22, 79]$	$[22, 76]$	$[20, 78]$
9	$[25, 81]$	$[27, 86]$	$[27, 89]$	$[27, 86]$	$[25, 84]$	$[24, 85]$
10	$[28, 90]$	$[28, 94]$	$[28, 98]$	$[29, 94]$	$[28, 90]$	$[27, 92]$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 4.7: 90% confidence procedures for the NB$(r, 0.84)$ distribution for CG, MST, BK, ST, B, and CP. Only intervals up to $x = 10$ are included.

Confidence procedures for r in the NB case contain infinitely many intervals. Therefore we will compare the length of the different procedures in a similar fashion as we did for the binomial case (with n unknown). Figure 4.13 is a plot of the interval width rankings for CG, MST, BK, B, and CP with $p = .84$ and $(1 - \alpha) = .90$. For instance for $x = 1$ the methods are ranked CG, B, CP, MST, and BK in order of their interval length (shortest to widest). Overlapping points indicate a tie has occurred. CG tends to do best on length followed by B. BK tends to have wider intervals than even CP.

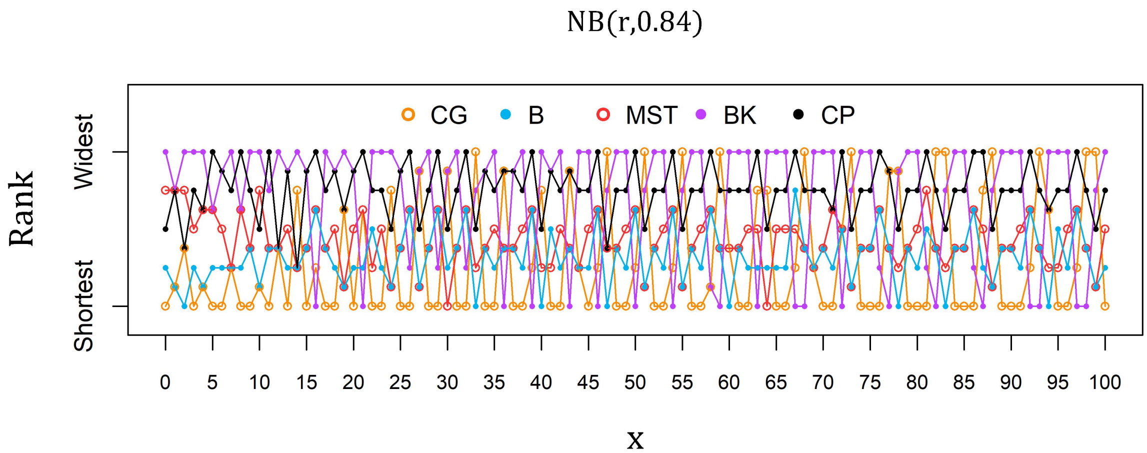

Figure 4.13: 90% $\mathrm{NB}(r, 0.84)$ confidence procedures ranked by confidence interval length for each x. Overlapping points indicate a tie has occurred for that x.

We might also be interested in the expected width,

$$\mathbb{E}_\theta\left[u(X) - l(X)\right] = \sum_{x=0}^{\infty}\left[u(x) - l(x)\right] \cdot P_n(X = x), \tag{4.8}$$

or the running average width. Note the expected width in this case is an infinite sum.

Figure 4.14 shows expected width and running average width for CG, MST, BK, B, and CP relative to MST for $p = 0.84$ and $(1 - \alpha) = .90$; i.e., each given value (expected width or running average width) for each procedure is divided by the corresponding value for MST to produce the values shown in the plots. Here, we approximate the infinite sum in (4.8) by,

$$\sum_{x=0}^{100}\left[u(x) - l(x)\right] \cdot P_n(X = x), \tag{4.9}$$

CG has the smallest expected width and running average width of any procedure. B is second best on length.

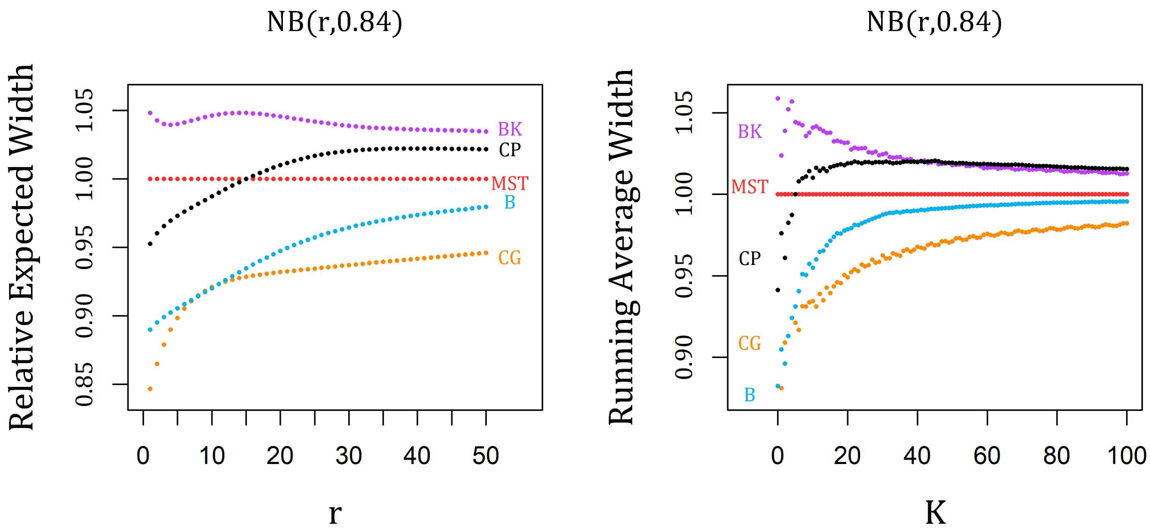

Figure 4.14: Plots expected width (left) and running average width (right) for CG, MST, BK, B, and CP relative to MST for the 90% NB$(r, 0.84)$ case.

Figure 4.15 shows average confidence interval width for $x \in \{0, 1, ..., 100\}$ v.s. mean coverage on [0,100] for $p = 0.84$ and $(1 - \alpha) = .90$. The ideal procedure would be located in the upper left corner. Since we often prioritize length over coverage, the procedures furtherest to the left are preferred.

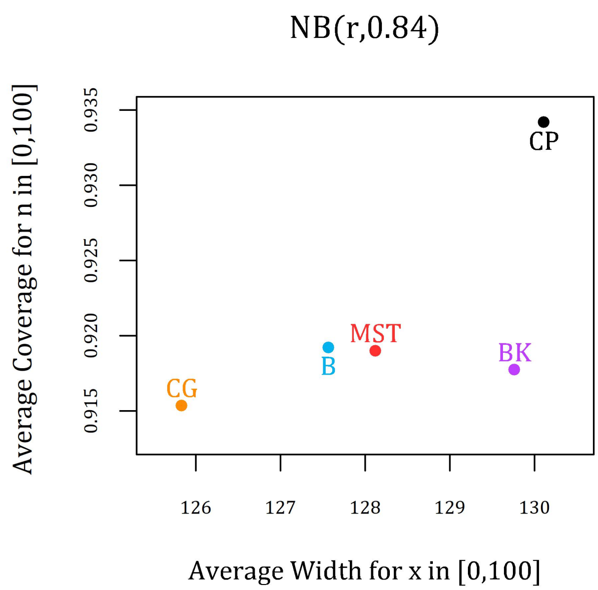

Figure 4.15: Plot of average confidence interval width for $x \in [0, 100]$ v.s. mean coverage on [0,100] for 90% NB(r, 0.84) procedures.

Because the confidence procedures for the NB case with r unknown have infinitely many intervals the length performance for the different methods will resemble that of the Poisson and binomial (n unknown) cases. CG will tend to produce the shortest interval lengths for each x, and therefore will have shortest expected length and running average length. Moreover it will be the length optimal procedure according to Schilling & Holladay's (2017)[22] asymptotic length criterion.

4.3.5 Coincidental Endpoints & Sterne's Coverage Ties

Coincidental endpoints work a little bit differently when working with integer-valued parameters as the transition between two curves $P_\theta(a \leq X \leq b)$ and $P_\theta(a{+}1 \leq X \leq b{+}1)$ of equal cardinality is not immediate, but rather takes place over the span of one unit of the parameter. The location of this transition determines the upper endpoint for $x = a$

and the lower endpoint $x = b + 1$. Specifically the two endpoints have the relationship $u(a) + 1 = l(b + 1)$, where $u(a)$ is the last location where the CPF uses the acceptance curve $P_\theta(a \leq X \leq b)$ and $l(b + 1)$ represents the location where the CPF first begins to use the curve $P_\theta(a + 1 \leq X \leq b + 1)$. The difference from before being that the coincidental endpoints are not equal, but rather differ by one unit.

ST/MST chooses from among those curves of minimal cardinality the one with highest coverage, but what happens when two curves $P_\theta(a \leq X \leq b)$ and $P_\theta(a+1 \leq X \leq b+1)$ are tied for highest coverage. Choosing one curve over the other makes either the confidence interval for $x = a$ or $b + 1$ shorter by one unit. Due to identifiability issues a choice needs to be made in such situations. An example where ties occurs quite frequently is in binomial$(n, 0.5)$ case. Figure 4.16 is a plot of all 90% minimal cardinality acceptance curves in the binomial$(n, 0.5)$ case. Ties occur here between every pair of curves $P_\theta(a \leq X \leq b)$ and $P_\theta(a + 1 \leq X \leq b + 1)$.

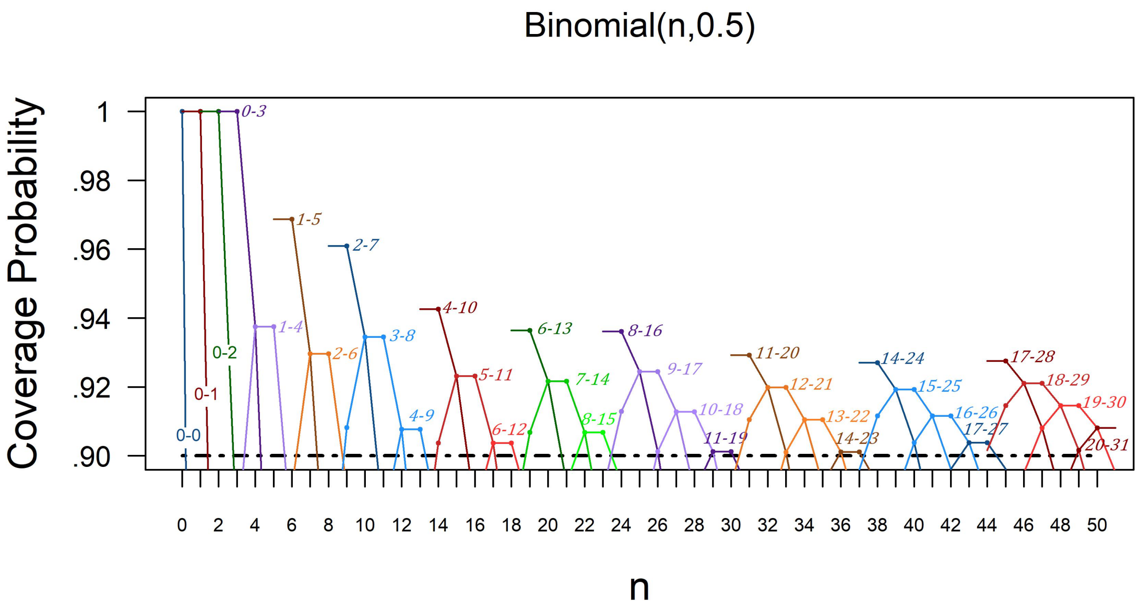

Figure 4.16: All 90% minimal cardinality acceptance curves in the binomial$(n, 0.5)$ case, except those that would cause gaps. We label acceptance curves $P_n(a \leq X \leq b)$ by a-b. Colors are used to better distinguish nearby curves. Nearby curves that are of different shades of the same color indicate that the curves have equal cardinality. For better visualization lines are placed between adjacent points $P_n(a \leq X \leq b)$, $P_{n+1}(a \leq X \leq b)$ to indicate that the two points lie on the same acceptance curve.

We prefer the convention that when multiple curves $\{P_\theta(a \leq X \leq b)\}$ of minimal cardinality are tied for highest coverage, ST/MST will use the one with the largest value of a. This choice not only resolves the identifiability issue, but will create a shorter procedure according to Schilling & Holladay's (2017)[22] asymptotic length criterion for cases when a procedure consist of infinitely many intervals. This is due to the same reason that CG is length optimal in such cases as it chooses shorter intervals for small x at the expense of larger values x.

4.4 Discussion: Countable Parameters

For estimation of integer-valued parameters we can employ similar methods/strategies to those that were used in the Poisson and binomial (n known) cases. Although there are a lot of similarities between the different cases one needs to consider each case separately and carefully as the CPF's can behave quite wildly like in the case of the HG with N unknown. Additionally we typically have to modify our length and coverage criteria to fit each case. However, there does seems to be a reoccurring theme in that when dealing with cases of finitely many intervals all minimal cardinality procedures are length optimal; hence, MST is typically preferred as it has the highest coverage of all gapless procedures. And when there are infinitely many intervals CG will be the shortest. However, CG will typically produce intervals with tied endpoints. This includes each of the five cases discussed above. Thus, the analog of MCG may be desired in such cases.

We stress that the comparisons among procedures in this chapter was by no means exhaustive as we made specific choices for the known parameters of each distribution. However, because the methods perform similarly in all cases investigated the results shown serve as a representative illustration of what to expect for the general case. Furthermore, over-coverage occurs in these fives cases just as they did in the Poisson and binomial (n known) cases. Thus, we might then be interested in more "efficient" randomized and fuzzy procedures. Even when UMPU tests do not exists fuzzy intervals can be obtained from conventional methods. In addition even non-exponential family members with MLR will have UMP/UMA test/procedures and thus the one-sided fuzzy versions of these procedures will inheirit the nice length properties associated with UMA procedures.

4.5 Fuzzy Credible Intervals

We saw in the case of the binomial(n, θ) (n known) and Poisson(θ) distributions that deriving credible intervals was rather uncomplicated as dealing with a continuous posterior distribution avoids the issues inherent to discrete distributions and thus exact $(1-\alpha)$ credible intervals are attainable. However, in the case of estimation of parameters of a countable parameter space, the posterior will be discrete and hence we face similar issues to those of conventional confidence procedures.

When the posterior is discrete the credible probability of an interval $[l(x), u(x)]$ is given by,

$$\mathbb{P}(\theta \in [l(x), u(x)]) = \sum_{\theta=l(x)}^{u(x)} \pi(\theta|x), \tag{4.10}$$

where $\pi(\theta|x)$ is the posterior PMF. Hence, achieving exact $(1-\alpha)$ credible probability will not be possible in general. The $(1-\alpha)$ HPD region for θ is given by,

$$H(x) = \{\theta : \pi(\theta|x) \geq k\}, \tag{4.11}$$

where the cutoff k is a function of α and in the discrete case can be chosen to be largest value such that,

$$\mathbb{P}(\theta \in H(x)) \geq 1-\alpha \tag{4.12}$$

Note this is analogous to Sterne's approach in that we enter the most probable values of the parameter into the credible set one by one until the first time the credible probability exceeds $(1-\alpha)$.

It might be possible for two values θ_1, θ_2 to have equal posterior probabilities satisfying $\pi(\theta_1|x) = \pi(\theta_2|x) = k$. When this happens the HPD region in (4.11) would then include both θ_1 and θ_2 even if only one is needed to achieve a $(1-\alpha)$ credible probability. As a result both $H(X)-\{\theta_1\}$, $H(X)-\{\theta_2\}$ will be shorter $(1-\alpha)$ credible sets. In such cases a convention should be chosen so that the shortest possible credible interval is obtained.

4.5.1 Credible Intervals for M in the HG

To further explore the issues that arise when constructing credible intervals from a discrete posterior, let X follow the HG(N, M, n) distribution with M unknown. A common prior for this situation is the Polya(α, β, N) [or beta-binomial(N, α, β)] distribution with probability mass function,

$$\pi(M|N, \alpha, \beta) = \frac{\binom{\alpha+M}{M}\binom{\beta+N-M}{\beta}}{\binom{\alpha+\beta+N+1}{\alpha+\beta+1}}, \quad M = 0, 1, ..., N, \tag{4.13}$$

where, α and β can be chosen so that prior best matches the practitioner's preconceived beliefs about M (Dyer & Pierce (1993)[8]). The Polya distribution is quite versatile due to its richness in PMF shapes. In addition to unimodal members the Polya prior encases the discrete uniform prior ($\alpha = 1$,$\beta = 1$), and both strictly increasing ($\beta = 0$) and decreasing ($\alpha = 0$) members (Dyer & Pierce (1993)[8]). Using the Polya prior in (4.13) we obtain the posterior,

$$\pi(M|x; N, n, \alpha, \beta) = \frac{\binom{\alpha+M}{\alpha+x}\binom{\beta+N-M}{\beta+n-x}}{\binom{\alpha+\beta+N+1}{\alpha+\beta+n+1}}, \quad M = x, ..., N - n + x, \tag{4.14}$$

(Peskun (2016)[26]). The Polya prior is a near conjugate prior, in that the posterior of $M - x$ follows a Polya distribution (Peskun (2016)[26]).

For illustration purposes suppose $N = 20$, $n = 7$, $\alpha = \beta = 1$, and $x = 2$ is observed. Figure 4.17 shows the corresponding PMF plot for $\pi(m|2)$. k_{90}, k_{95}, and k_{99} represent the corresponding 90%, 95%, and 99% cutoffs for the HPD region in (4.11). Table 4.8 shows the resulting HPD regions. For instance $H(2) = \{M : \pi(M|2) \geq k_{90}\} = [3, 10]$ is the 90% HPD region when $x = 2$ is observed.

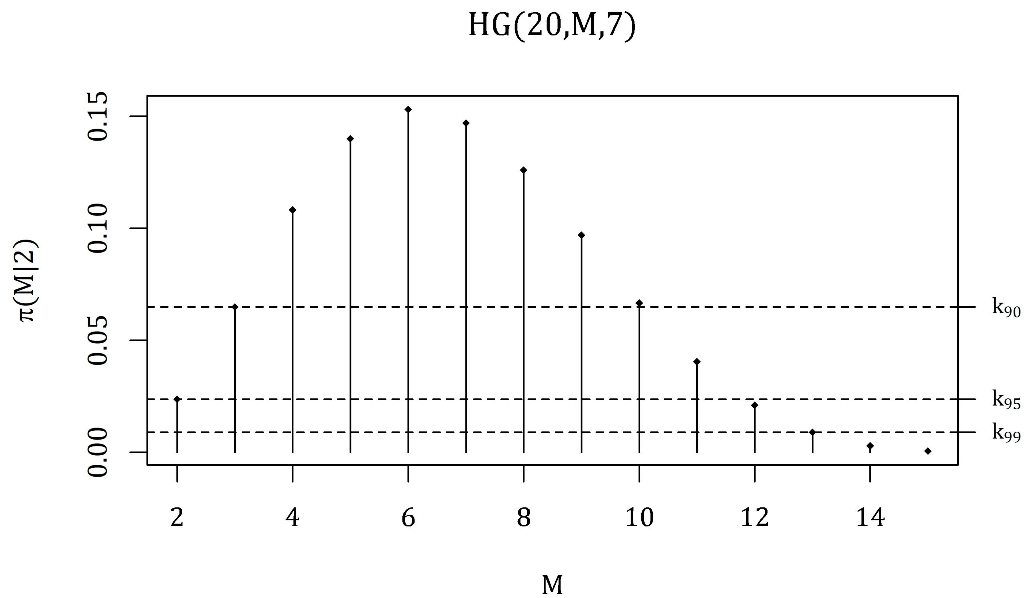

Figure 4.17: Plot of the posterior PMF in (4.14) when $N = 20$, $n = 7$, $\alpha = \beta = 1$, and $x = 2$ is observed. k_{90}, k_{95}, and k_{99} represent the corresponding 90%, 95%, and 99% cutoffs for the HPD region in (4.11).

The table displays both the target $(1 - \alpha)$ credible probability and the true credible probability $\mathbb{P}(\theta \in H(x))$ actually obtained. The two values will rarely ever be equal.

Credible Interval	Target Credible Probaility	True Credible Probaility
$[3, 10]$	0.90	0.9027
$[2, 11]$	0.95	0.9668
$[2, 13]$	0.99	0.9967

Table 4.8: 90%, 95%, and 99% HPD regions using the posterior in (4.14) with $N = 20$, $n = 7$, $\alpha = \beta = 1$, and $x = 2$.

4.5.2 Fuzzy Credible Intervals for M in the HG

Since in general exact $(1 - \alpha)$ credible probabilities cannot be obtained, in the spirit of fuzzy confidence intervals, we propose the following idea for fuzzy credible intervals.

Suppose a $(1-\alpha)$ credible interval $[l(x), u(x)]$ is obtained from a discrete posterior. Define the *fuzzy credible interval* by,

$$\psi(\theta|x) = \begin{cases} 1 & l(x) < \theta < u(x) \\ \gamma_1 & \theta = l(x) \\ \gamma_2 & \theta = u(x) \\ 0 & \text{else} \end{cases}, \tag{4.15}$$

where γ_1, γ_2 are any two probabilities such that,

$$1-\alpha = \gamma_1 \cdot \mathbb{P}(\theta = l(x)|x) + \mathbb{P}(l(x) < \theta < u(x)|x) + \gamma_2 \cdot \mathbb{P}(\theta = u(x)|x) \tag{4.16}$$

The function $\psi(\theta|x)$ provides the degree (probability) of membership (inclusion) of each θ into the credible confidence interval of x. Hence $l(x)$ and $u(x)$ would be included in the corresponding *randomized credible interval* with probability γ_1 and γ_2 respectively; where as, values of θ in $[l(x)+1, u(x)-1]$ are guaranteed to be included.

γ_1 and γ_2 can be chosen, for instance, so that their ratio γ_1/γ_2 is as close as possible to the ratio of the credible probabilities $\mathbb{P}(\theta = l(x)|x)/\mathbb{P}(\theta = u(x)|x)$; that is, so that,

$$\frac{\gamma_1}{\gamma_2} \approx \frac{\mathbb{P}(\theta = l(x)|x)}{\mathbb{P}(\theta = u(x)|x)}. \tag{4.17}$$

In this way a higher degree of membership is given to values of θ with large credible probabilities. When the two ratios can be made exactly equal it follows that,

$$\begin{aligned} \gamma_1 &= \frac{\gamma_2 \cdot \mathbb{P}(\theta = l(x)|x)}{\mathbb{P}(\theta = u(x)|x)}, \\ \gamma_2 &= \frac{(1-\alpha) - \mathbb{P}(l(x) < \theta < u(x)|x)}{[\mathbb{P}(\theta = l(x)|x)]^2/\mathbb{P}(\theta = u(x)|x) + \mathbb{P}(\theta = u(x)|x)}. \end{aligned} \tag{4.18}$$

If (4.18) would cause either γ_i to exceed 1 then set that γ_i to 1 which will in turn define the remaining γ_i. As a result, the two ratios γ_1/γ_2 and $\mathbb{P}(\theta = l(x)|x)/\mathbb{P}(\theta = u(x)|x)$ will be as close as possible. Note because these HPD regions are of minimal length $(\gamma_1, \gamma_2) = (1, 1)$ gives a credible probability of at least $(1-\alpha)$ and $(\gamma_1, \gamma_2) = (0, 0)$ gives a credible probability below $(1-\alpha)$; hence a solution will always exist. In particular, setting either $\gamma_i = 1$ will provide two possible solutions. Thus, our choice of γ_i's is well-defined.

Such a construction allows a Bayesian to obtain exact $(1-\alpha)$ credible probability even when the posterior is discrete. Figure 4.18 is a plot of the corresponding fuzzy credible intervals obtained from the intervals in Table 4.8.

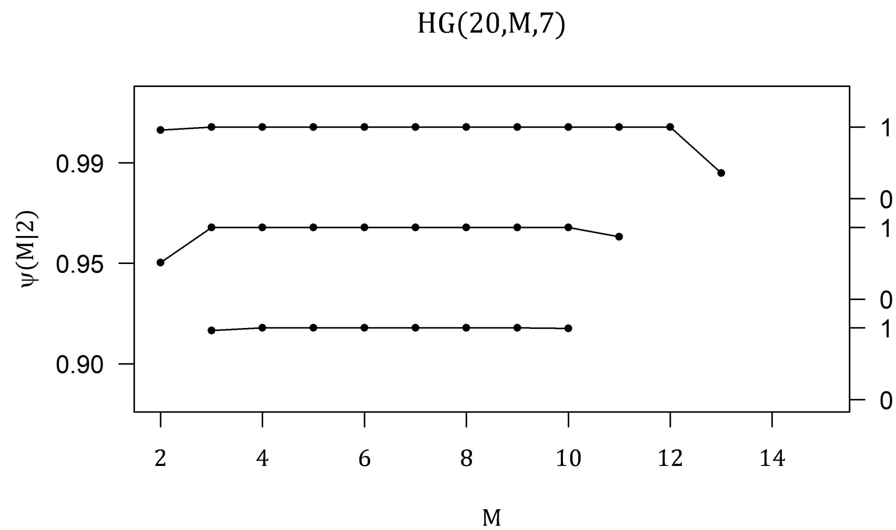

Figure 4.18: Plot of the 90%, 95%, and 99% proposed fuzzy credible intervals obtained from the intervals in Table 4.8.

Chapter 5

Two-Parameter Confidence Regions

In this chapter we discuss some applications to the two-parameter situation. Suppose vector $\mathbf{X} = (X_1, X_2, ..., X_N) \in \boldsymbol{\mathcal{X}}$ has a multivariate discrete distribution with joint probability mass function $p(\mathbf{x}|\boldsymbol{\theta})$. $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_r) \in \boldsymbol{\Theta}$ is a vector of unknown parameters of the distribution. Confidence regions for $\boldsymbol{\theta}$ based on observed points $\mathbf{x} = (x_1, ..., x_N)$ are desired. We focus primarily on the particular case when $N = r = 2$.

5.1 Bivariate Case

In the bivariate case with $r = 2$ unknown parameters $\mathbf{X} = (X, Y) \in \boldsymbol{\mathcal{X}}$ has bivariate discrete distribution with joint probability mass function $p_{X,Y}(x, y|\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \boldsymbol{\Theta}$ is unknown. Confidence regions for (θ_1, θ_2) based on observed points (x, y) are desired.

5.1.1 Independent Case

Let us first consider the case when $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ are two independent discrete random variables with probability mass functions, $p_X(x|\theta_1)$ and $p_Y(y|\theta_2)$. Suppose

$(X, Y) = (x, y)$ are observed and a $100(1-\alpha)\%$ confidence region for $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$ is desired. Here $\boldsymbol{\Theta} = \Theta_1 \times \Theta_2$ and $\boldsymbol{\mathcal{X}} = \mathcal{X} \times \mathcal{Y}$. If X and Y were continuous random variables, then independence makes for a somewhat uninteresting problem, since for any observed pair (x, y) one can take $C(x, y) = [l_X(x), u_X(x)] \times [l_Y(y), u_Y(y)]$ for two $(1-\alpha)^{\frac{1}{2}}$ confidence intervals for θ_1 and θ_2 respectively. Or if convenient, one may choose any two confidence levels that have product equal to $(1-\alpha)$. In the discrete case, a strict level $(1-\alpha)$ confidence procedure rarely has coverage exactly equal to $(1-\alpha)$. Instead it is only required that the minimum coverage exceeds $(1-\alpha)$. One can still obtain a $(1-\alpha)$ confidence region by crossing two marginal $(1-\alpha)^{\frac{1}{2}}$ confidence intervals. However, by considering the problem of estimating both θ_1 and θ_2 in combination it should be possible to drop the overall coverage probabilities even closer to the confidence level without going below it, which will potentially allow for smaller confidence regions.

For example, consider the case when X and Y are independent Poisson random variables with means θ_1 and θ_2. One way to construct acceptance regions for (θ_1, θ_2) is to extend Sterne's (1954) [27] method to the multi-parameter case; that is, enter the most probable points $\mathbf{x} = (x, y)$ into $A_{\boldsymbol{\theta}}$ one-by-one in decreasing order of probability, until the first time the probability of $A_{\boldsymbol{\theta}}$ exceeds $1-\alpha$. Figure 5.1 shows some examples of Sterne's 95% confidence regions for various observed pairs $\mathbf{x} = (x, y)$. The blue boxes are the resulting confidence regions from crossing two $\sqrt{95}\%$ one-parameter Sterne procedures. The red dots represent the MLE's for (θ_1, θ_2) in each case. The "guitar pick" shaped confidence regions certainly have a smaller squared area then the rectangles formed from crossing two one parameter $\sqrt{95}\%$ Sterne procedures.

Figures 5.2 and 5.3 show the corresponding CPF's for the two-parameter and crossed Sterne procedures respectively. The black planes in the figures represent the 0.95 confidence level. Colors are used to help in comparing the coverage in the two-parameter case with the respective values in the case of crossing 2 one-parameter procedures. From

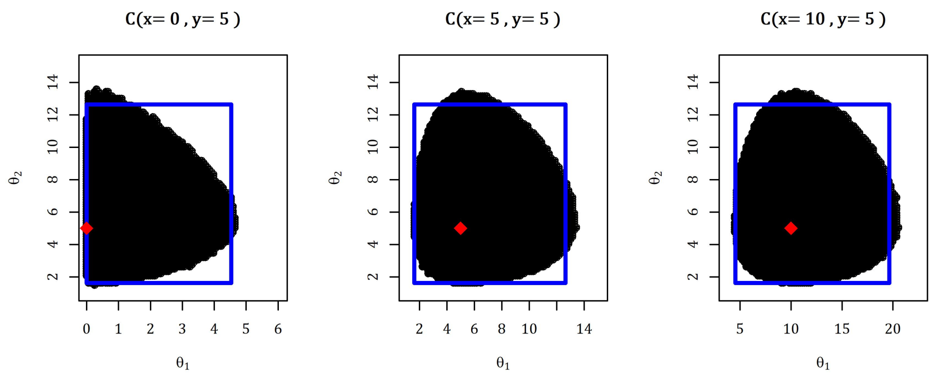

Figure 5.1: Comparison of the 95% two-parameter Sterne procedure (black regions) with the region produced by crossing two one-parameter $\sqrt{95}\%$ Sterne procedures (blue boxes) for the case of two independent Poisson random variables. Red points represent the MLE's of (θ_1, θ_2) in each case.

these two figures, it is apparent that the two-parameter Sterne procedure has significantly lower coverage than the coverage produced by crossing 2 one-parameter Sterne procedures, and therefore results in smaller confidence regions.

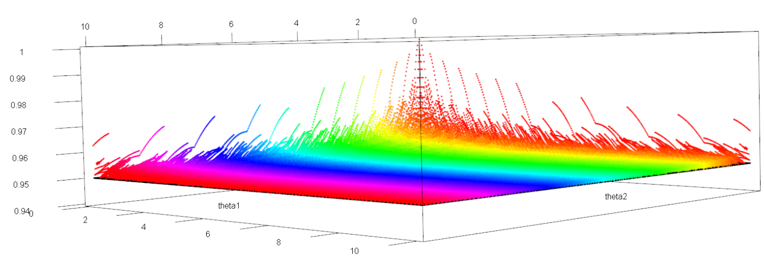

Figure 5.2: Two-parameter 95% Sterne CPF for the case of two independent Poisson random variables.

5.1.2 Compound Distributions

We can also apply the extension of Sterne's (1954) method to compound distributions where we deal with random sums of random variables such as:

$$Y = \sum_{i=1}^{N} Y_i$$

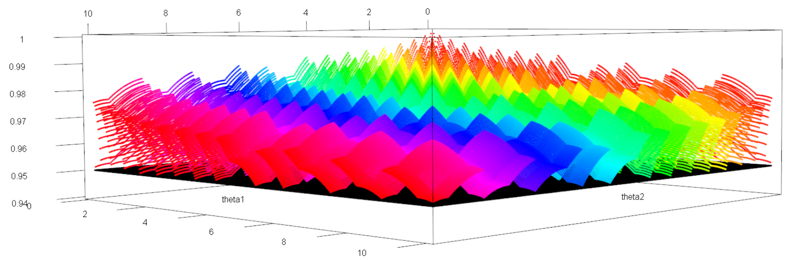

Figure 5.3: 95% crossed Sterne CPF for the case of two independent Poisson random variables.

where N and $\{Y_i\}$ are random variables. Given the identifiability problem we will assume that both the number of elements in the sum N and and the value of the sum Y are observed. Let $N \sim p_N(x|\theta_1)$ and $(Y|N) = \sum_{i=1}^{N} Y_i \sim p_{(Y|N)}(x|\theta_2)$ where, Y_i's are independent identically distributed random variables whose distribution depends on an unknown parameter θ_2. Given only the observed values of N and Y we'd like to construct a confidence region for (θ_1, θ_2). Such a distribution has applications in insurance, for example, where N represents the number of insurance claims in a given time period and Y_i's represent some attribute of claims such as the claim amount. Y_i's can even be binary random variables representing some true or false statement about each claim.

For example, let $N \sim \text{Poisson}(\theta_1)$ and $(Y|N) \sim \text{binomial}(N, \theta_2)$. Then, based on an observed pair $(N, Y) = (n, y)$ we would like to estimate (θ_1, θ_2) with confidence region $C(n, y)$. Here,

$$\begin{aligned} p_{N,Y}(n, y) &= p_{Y|N=n}(x|y) \times p_N(n) \\ &= \left[\binom{n}{y} \theta_2^y (1-\theta_2)^{n-y}\right] \left[\frac{e^{-\theta_1} {\theta_1}^n}{n!}\right] \end{aligned}$$

Figure 5.4 shows several 95% Sterne confidence regions for various points (n, y). Figure 5.5 shows some example of Sterne's 95% acceptance regions for various points (θ_1, θ_2).

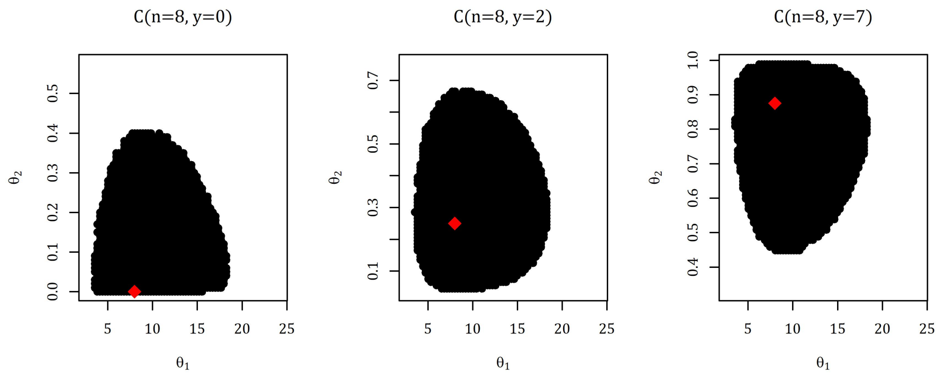

Figure 5.4: 95% Sterne confidence regions for the Poisson-binomial compound distribution.

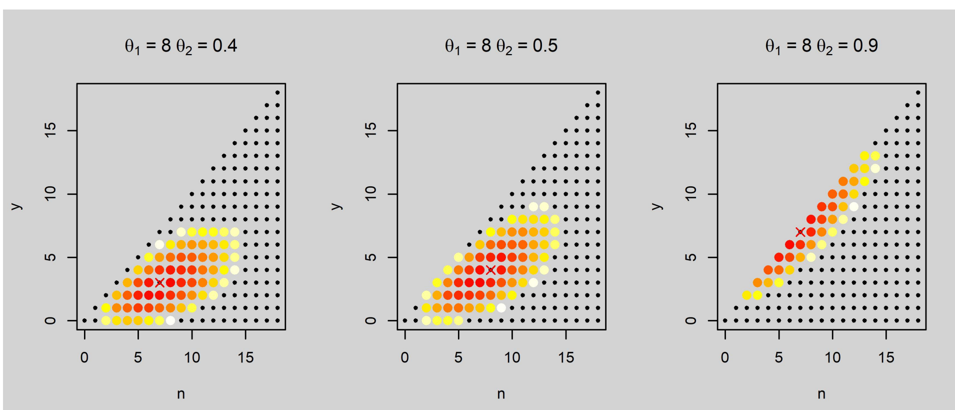

Figure 5.5: 95% Sterne acceptance regions for the Poisson-binomial compound distribution. Points in the acceptance regions are heat colored by probability.

Chapter 6

Future Work

Recall that during the construction of the fuzzy version of a conventional confidence procedure, there were regions for which there was infinitely many choices for the pair (γ_1, γ_2). The approach in this work is to use a straight-line due to the simplicity and aesthetic appeal of such an approach. Instead, one might define the fuzzy edges of the procedure by imposing yet another optimality criteria, which we wish to investigate.

The topic of estimation of the parameters of the hypergeometric distribution encompasses quite an intricate problem due to having 3 parameters, multiple parameterizations, and several extensions including versions of the noncentral hypergeometric and negative hypergeometric distributions. Moreover, often researchers are interested in estimation of the proportion $p = M/N$ of the population possessing an attribute of interest. This is a direct application of confidence intervals for M (when N, n known) as the confidence limits for M can be divided by N to obtain a confidence procedure for p. However, a related open problem is estimation of N and M when p and n are known. An interested party could also extend the ideas of this paper to deal with other parameterizations of the HG and with estimation of the parameters from the many versions of the noncentral hypergeometric and negative hypergeometric distributions. Other research ideas could involve having more than one unknown parameter or exploring Bayesian methods for parameters other than M.

On that note one might also consider a Bayesian approach for estimation of the other countable-parameters, such as n and r for the binomial and NB distributions respectively. It would also be interesting to extend the methods of this dissertation to other discrete distributions including some of the discrete circular distributions such as the wrapped Poisson or Mardia's (1972, p. 50) [28] biased roulette wheel distribution.

We barely scratched the surface on the topic of estimation in the multi-parameter case. Our hope is that much of the work in the one-parameter case can be extended to handle the wide variety of situations stemming from the multi-parameter case. In particular, the ideas of this dissertation on fuzzy confidence intervals can be extended to the two-parameter (or multi-parameter) situation. Brown et al., in their comments on Geyer and Meeden (2005) suggested the idea that fuzzy intervals can be extended to simultaneous regions for multiple binomial proportions. Let $X_i \sim \text{Binomal}(n_i, \theta_i)$ for $i = 1, 2$, where X_1 and X_2 are not necessarily independent. Then one may wish to construct a $1-\alpha$ level confidence region (or the randomized or fuzzy analogs) for (θ_1, θ_2). But, there is also an opportunity for extending this idea to general two-parameter (or multi-parameter) discrete distributions.

Our hope is that researchers will begin to use some of the methods proposed in this dissertation; however, this is more likely to happen if the intervals are easily computable and accessible. As a result, some future work on this topic would consist of publishing R code and making the intervals accessible online for the various methods and distributions discussed.

References

[1] Agresti, A. and Coull, B. A. Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician*, 52(2):119–126, 1998.

[2] Choi, S. A. Confidence Intervals for the Success Parameter of a Negative Binomial Random Variable. Master's thesis, California State University, Northridge, 2015.

[3] Holladay, B. Optimal Confidence Intervals for the Expectation of a Poisson Random Variable. Master's thesis, California State University, Northridge, 2014.

[4] Brown, L. D., Cai, T. T. and DasGupta, A. Interval Estimation for a Binomial Proportion. *The Canadian Journal of Statistics*, 16(2):101–133, 2001.

[5] Blyth, C. and Hutchinson, D. Table of Neyman-Shortest Unbiased Confidence Intervals for the Binomial Parameter. *Biometrika*, 47(3/4):381–391, 1960.

[6] Blyth, C. and Still, H. Binomial Confidence Intervals. *Journal of the American Statistical Association*, 78(381):108–116, 1983.

[7] Clopper, C.J and Pearson, E. S. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*, 26(4):404–413, 1934.

[8] Dyer, D. and Pierce, R. On the Choice of the Prior Distribution in Hypergeometric Sampling. *Communications in Statistics - Theory and Methods*, 22(8):2125–2146, 1993.

[9] Crow, E. Confidence Intervals for a Proportion. *Biometrika*, 43(3/4):423–435, 1956.

[10] Crow, E. and Gardner, R. Confidence Intervals for the Expectation of a Poisson Variable. *Biometrika*, 46(3/4):441–453, 1959.

[11] Garwood, F. Fiducial Limits for the Poisson Distribution. *Biometrika*, 28(3/4):437–442, 1936.

[12] Casella, G. Refining Binomial Confidence Intervals. *The Canadian Journal of Statistics*, 14(2):113–129, 1986.

[13] Casella, G. and Robert, C. Refining Poisson Confidence Intervals. *The Canadian Journal of Statistics*, 17(1):45–57, 1989.

[14] Casella, G. and McCulloch, C. E. Confidence Intervals for Discrete Distributions. *Biometrics Unit Technical Reports: Number BU-839-M*, 1984.

[15] Casella, G. and Berger, R. L. *Statistical Inference*. Duxbury, second edition, 2002.

[16] Blaker, H. Confidence Curves and Improved Exact Confidence Intervals for Discrete Distributions. *The Canadian Journal of Statistics*, 28(4):783–798, 2000.

[17] Byrne, J. and Kabaila, P. Comparison of Poisson Confidence Intervals. *Communications in Statistics - Theory and Methods*, 34(3):545–556, 2005.

[18] Geyer, C. J. and Meeden, G. D. Fuzzy and Randomized Confidence Intervals and P-Values. *Statistical Science*, 20(4):358–366, 2005.

[19] Pratt, J. Length of Confidence Intervals. *Journal of the American Statistical Association*, 56(295):549–567, 1961.

[20] Lehmann, E. L. and Romano, J. P. *Testing Statistical Hypotheses*. Springer, third edition, 2005.

[21] Stevens, W. L. Fiducial Limits of the Parameter of a Discontinuous Distribution. *Biometrika*, 37(1/2):117–129, 1950.

[22] Schilling, M. and Holladay, B. Length Minimization for Poisson Confidence Procedures. *Communications in Statistics - Theory and Methods*, 46(2):861–873, 2017.

[23] Schilling, M. and Doi, J. A Coverage Probability Approach to Finding an Optimal Binomial Confidence Procedure. *The American Statistician*, 68(3):133–145, 2014.

[24] Berger, J. O. *Statistical Decesion Theory and Bayesian Analysis*. Springer-Verlag, second edition, 1980.

[25] Kabaila, P. On Randomized Confidence Intervals for the Binomial Probability. `https://arxiv.org/abs/1302.6659v1` [math.ST], 2013. Accessed: 2017-05-06.

[26] Peskun, P. Some Relationships and Properties of the Hypergeometric Distribution. `https://arxiv.org/abs/1610.07554` [math.ST], 2016. Accessed: 2019-05-18.

[27] Sterne, T. Some Remarks on Confidence or Fiducial Limits. *Biometrika*, 41(1/2):275–278, 1954.

[28] Mardia, K. V. *Statistics of Directional Data*. Academic Press, 1972.