Binomial Confidence Intervals

Derek Webb

September 21, 1995

APPROVAL

of a writing project submitted by

DEREK WEBB

This writing project has been read by the writing project director and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Statistics Faculty.

Date

Robert J. Boik
Writing Project Director

Introduction

This article considers coinfidence intervals of the parameter θ from the binomial distribution

$$P(X = x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}I_{\{0,1,\ldots,n\}}(x). \tag{1}$$

The generation of confidence intervals by a particular method results in a collection of $n+1$ intervals of the form $[\ell_x, u_x]$, $x = 0, 1, ..., n$. The coverage probability of a method can be calculated as the probability that a closed random interval $[\ell_X, u_X]$ covers the true parameter θ. This is computed by

$$P(\ell_X \leq \theta \leq u_X) = \sum_{x=0}^{n}\binom{n}{x}\theta^x(1-\theta)^{n-x}I_{[\ell_x,u_x]}(\theta). \tag{2}$$

For a given set of confidence intervals, it is useful to look at the minimum coverage probability, namely $\inf_\theta P(\ell_X \leq \theta \leq u_X)$. The actual construction of a set of $1-\alpha$ confidence intervals for θ (where $0 < \alpha < 1$) may result in $\inf_\theta P(\ell_X \leq \theta \leq u_X) < 1-\alpha$. The optimal method would result in $\inf_\theta P(\ell_X \leq \theta \leq u_X) = 1-\alpha$.

This paper will focus on particular methods of generating confidence intervals that satisfy Blyth and Still's (1983) three restrictions.

(1) *Exact confidence intervals.* This requirement states that $\inf_\theta P(\ell_X \leq \theta \leq u_X) \geq 1-\alpha$.

(2) *Equivariance.* The family of binomial distributions is invariant under the transformation $g(X) = n - X$ where $\bar{g}(\theta) = 1-\theta$. The confidence intervals must be equivariant under $g(X)$ which requires $\ell_x = 1 - u_{n-x}$, $x = 0, 1, ..., n$. This requirement produces a

symmetry about $\theta = .5$ when one plots $P(\ell_X \leq \theta \leq u_X)$. Also, one need only know the lower endpoints to determine all $n+1$ confidence intervals.

(3) *Monotone in X*. This requires interval ends to be strictly increasing in X such that $\ell_{x+1} > \ell_x$ and $u_{x+1} > u_x$.

Clopper-Pearson Intervals

One of the oldest and most widely employed sets of confidence intervals are the Clopper-Pearson intervals (Clopper and Pearson, 1934). These can be generated by the statistical method (Casella and Berger, 1990). The statistical method states that for a discrete random variable X with CDF $F_X(x|\theta) = P(X \leq x|\theta)$, ℓ_X and u_X can be defined in the following manner. If $F_X(x|\theta)$ is a decreasing function of θ and $0 < \alpha < 1$ then, define ℓ_X and u_X by

$$P(X \geq x|\ell_X) = \frac{\alpha}{2}, \quad P(X \leq x|u_X) = \frac{\alpha}{2}. \tag{3}$$

The binomial CDF is a decreasing function of θ as is shown by the following:

$$\begin{aligned}
F_X(x|\theta) &= \sum_{k=0}^{x} \binom{n}{k} \theta^k (1-\theta)^{n-k} \\
&= (1-\theta)^n + \binom{n}{1}\theta(1-\theta)^{n-1} + \ldots + \binom{n}{x}\theta^x(1-\theta)^{n-x} \\
\Rightarrow \frac{\partial}{\partial\theta} F_X(x|\theta) &= -n(1-\theta)^{n-1} + n(1-\theta)^{n-1} - n\theta(n-1)(1-\theta)^{n-2} + \ldots \\
&\quad - \binom{n}{x-1}\theta^{x-1}(n-x+1)(1-\theta)^{n-x} + x\binom{n}{x}\theta^{x-1}(1-\theta)^{n-x} \\
&\quad - \binom{n}{x}\theta^x(n-x)(1-\theta)^{n-x-1} \\
&= -n\theta(n-1)(1-\theta)^{n-2} + \ldots - \binom{n}{x-1}\theta^{x-1}(n-x+1)(1-\theta)^{n-x}
\end{aligned}$$

$$+\binom{n}{x-1}\theta^{x-1}(n-x+1)(1-\theta)^{n-x} - \binom{n}{x}\theta^{x}(n-x)(1-\theta)^{n-x-1}$$

$$= -\binom{n}{x}\theta^{x}(n-x)(1-\theta)^{n-x-1}.$$

Thus $\frac{\partial}{\partial\theta}\mathrm{Fx}(\mathrm{x}|\theta) < 0.$

Reexpressing (3) one has

$$P(X \leq x-1|\ell_X) = 1-\frac{\alpha}{2} \text{ and } P(X \leq x|u_X) = \frac{\alpha}{2}.$$

Solving for $[\ell_x, u_x], x = 1, 2, ..., n$ can be done by the use of the binomial distributions relation to the beta distribution. Consider the random variable $Y \sim \mathrm{Bin}(n, \theta)$, then $P(Y \leq y) = P(B < 1-\theta)$ where $B \sim \mathrm{Beta}(n-y, y+1)$. This relationship leads to

$$P(B_L < 1-\ell_x) = 1-\frac{\alpha}{2} \text{ and } P(B_U < 1-u_x) = \frac{\alpha}{2}$$

where

$$B_L \sim \mathrm{Beta}(n-x+1, x) \text{ and } B_U \sim \mathrm{Beta}(n-x, x+1).$$

Now, the calculation of the confidence intervals can be done by noting that

$$1-\ell_x = \mathrm{Beta}^{-1}(1-\frac{\alpha}{2}, n-x+1, x)$$

$$1-u_x = \mathrm{Beta}^{-1}(\frac{\alpha}{2}, n-x, x+1)$$

where $q = \text{Beta}^{-1}(p, \alpha, \beta)$ means that

$$p = \int_0^q \frac{u^{\alpha-1}(1-u)^{\beta-1}}{B(\alpha,\beta)} du \quad \text{where} \quad B(\alpha,\beta) = \frac{\Gamma(\alpha,\beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

It is also possible to obtain the confidence intervals by solving the following polynomials

$$\ell_x = \left\{ \ell_x : \sum_{t=0}^{x-1} \binom{n}{t} \ell_x^t (1-\ell_x)^{n-t} = 1 - \frac{\alpha}{2} \right\}$$

$$u_x = \left\{ u_x : \sum_{t=0}^{x} \binom{n}{t} u_x^t (1-u_x)^{n-t} = \frac{\alpha}{2} \right\}.$$

These two polynomials are a direct result of (3). The polynomial solution method was used by Vollset (1993).

The computation of the Clopper-Pearson confidence intervals for $\alpha = .05$ and $n = 10$ was carried out in Matlab using the Beta^{-1} function. They are as follows,

x	0	1	2	3	4	5	6	7	8	9	10
ℓ_x	0	.0025	.0252	.0667	.1216	.1871	.2624	.3475	.4439	.5550	.6915

and $u_x = 1 - \ell_{n-x}$. The confidence intervals clearly satisfy restrictions (2) and (3). It is conveniant to graph $P(\ell_X \le \theta \le u_X) \;\; \forall \theta \in (0,1)$ to demonstrate that these are exact confidence intervals and satisfy (1). Figure 1 was constructed by having θ range from .001 to .999 in increments of .001 and then solving (2) for each value of θ. Although this method is exact, it is not optimal. That is, $\inf_\theta P(\ell_X \le \theta \le u_X) = .9610$ at $\theta = .3475$ and $\theta = .6525$. This implies that the confidence intervals may be shortened somewhat before one reaches the $1 - \alpha$ confidence level.

Due to the discrete nature of the family of binomial distributions, there are a number

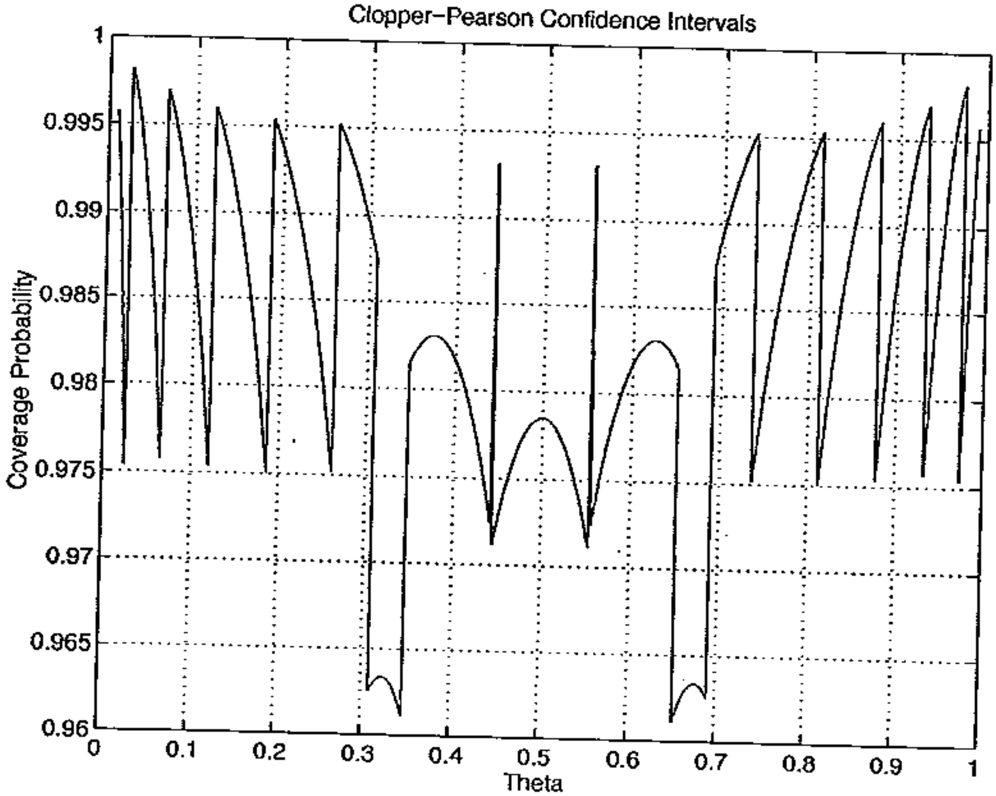

Figure 1: Graph of the coverage probabilities for Clopper-Pearson confidence intervals for $n = 10$.

of discontinuities in the function $P(\ell_X \leq \theta \leq u_X)$. In general, there are exactly $2(n+1)$ discontinuities. These occur where $\theta = \ell_x$ or $\theta = u_x$ for some $x \in \{0, 1, ..., n\}$. When the value of θ changes such that it moves into or out of a particular confidence interval a discontinuity results.

Crow Intervals

Crow (1956) approched the problem of computing confidence intervals by first computing acceptance regions and then inverting them. For a given hypothesis $\theta = \theta_o$ there is an acceptance region $[\ell_{\theta_o}, u_{\theta_o}]$ such that one accepts the hypothesis if, for an observed x, $\ell_{\theta_o} \leq$

$x \leq u_{\theta_o}$. After a set of acceptance regions has been calculated for a finite number of $\theta_o \in (0,1)$ they can be inverted to yield confidence intervals. This can be summarized as follows

$$\{\ell_x \leq \theta_o \leq u_x\} \Leftrightarrow \{\ell_{\theta_o} \leq X \leq u_{\theta_o}\}.$$

An example from Blyth and Still (1983) illustrates Crows method very well. Consider the case where $n = 8$ and $1 - \alpha = .95$. For $\theta_o = .3145$ the shortest acceptance region possible is $0 \leq X \leq 5$ where $P(0 \leq X \leq 5|\theta_o) \geq .95$. But, for $\theta_o = .3155$ through $\theta_o = .3995$ there are two shortest acceptance regions possible, $0 \leq X \leq 5$ and $1 \leq X \leq 6$. For $\theta_o = .4005$ there is again only one shortest acceptance region $1 \leq X \leq 6$. Crow was faced with choosing a $\theta_o \in [.3155, .3995]$ at which to stop using the acceptance region $0 \leq X \leq 5$ and start using $1 \leq X \leq 6$. When there was a choice to be made, he always chose the acceptance region farthest to the right, $1 \leq X \leq 6$ for this example. That is, he chose $1 \leq X \leq 6$ as the acceptance region for all $\theta_o \in [.3155, .3995]$.

Crow's confidence intervals were constructed using the criterion $P(\ell_{\theta_o} \leq X \leq u_{\theta_o}) \geq 1 - \alpha$ which ensures compliance with restriction (1). He also constructed all acceptance regions for $\theta_o \in (0, .5]$ then by symmetry he constructed the rest for $\theta_o \in (.5, 1)$. The acceptance regions were then inverted to obtain the confidence intervals. The construction method leads to confidence intervals that satisfy $\ell_x = 1 - u_{n-x}$ (restriction (2)). There are cases however, where Crow's confidence intervals do not satisfy restriction (3). That is, the interval endpoints are not strictly increasing in X. For example, using Crow's (1956) tables, the 95% confidence intervals corresponding to $n = 14$, $x = 6$, and $x = 7$ are $[.206, .688]$ and $[.206, .794]$, respectively. The endpoint ℓ_X is nonincreasing in X.

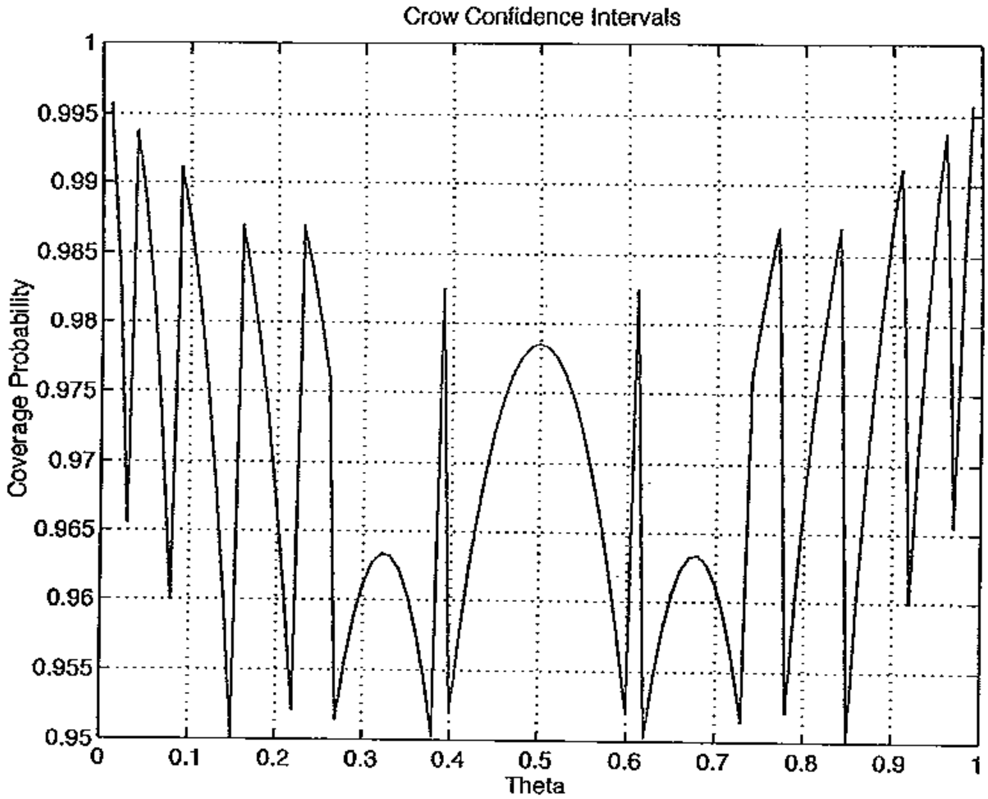

Figure 2: Graph of the coverage probabilities for Crow confidence intervals for $n = 10$.

Crow's confidence intervals may not always meet restriction (3), but they do have another very desirable quality. Crow (1956) proved that a set of $n+1$ confidence intervals generated by inverting minimum-length acceptance regions results in a set of minimum length intervals. That is the sum of the lengths of these confidence intervals is as small as posible. For certain θ_o more than one minium-length acceptance region is possible. Because of the different minimum-length acceptance regions one may choose, there is a family of confidence intervals with the property that the sum of their lengths is as small as possible.

The graph of coverage probabilities for Crow confidence intervals, where $n = 10$ and $\alpha =$ .05, is presented in figure 2 for comparison tothe Clopper-Pearson graph. The Crow confidence intervals actually attain the $1-\alpha$ confidence level for a countable number of θ.

Blyth-Still Intervals

As shown in the previous example; where $n = 8$ and $1 - \alpha = .95$ there were two shortest acceptance regions for $\theta_o \in [.3155, .3995]$, namely $0 \leq X \leq 5$ and $1 \leq X \leq 6$. Crow's method of choosing which acceptance region to use was to choose the one farthest to the right, that is choose $1 \leq X \leq 6$. This results in $u_0 = .3155$ and $\ell_6 = .3155$. Blyth and Still (1983) point out that this choice creates shorter confidence intervals for X near 0 or n and longer confidence intervals for X near $n/2$. Even though the sum of the lengths of Crow's confidence intervals is minimized, they may contain a certain confidence interval which is unnecessarily long. The reason being that confidence intervals near $X = n/2$ are already longer than for other choices of X.

Blyth and Still considered choosing minimum acceptance regions farthest to the left when there was a choice to be made at a particular θ_o. In the previous example, for $\theta_o \in [.3155, .3995]$ one would choose acceptance region $0 \leq X \leq 5$. This method lengthens confidence intervals for X near 0 or n and shortens them for X near $\frac{n}{2}$. This approach has the desirable property of minimizing the length of the longest confidence interval. One downfall to this method is that there are more numerous and extreme violations of restriction (3).

Blyth and Still (1983) decided to use an intermediate method for computing confidence intervals. The rule they followed for choosing acceptance regions for each $\theta_o \leq \frac{1}{2}$ is "take each confidence interval endpoint to be the midpoint of the interval of possibilities." Refering back to the example for $\theta_o \in [.3155, .3995]$ they would choose acceptance region $0 \leq X \leq 5$ for all $\theta_o \in [.3155, .3575)$ and choose acceptance region $1 \leq X \leq 6$ for all $\theta_o \in [.3575, .3995]$. This results in $u_o = .3575$ and $\ell_6 = .3575$. Blyth and Still's confidence intervals retain the

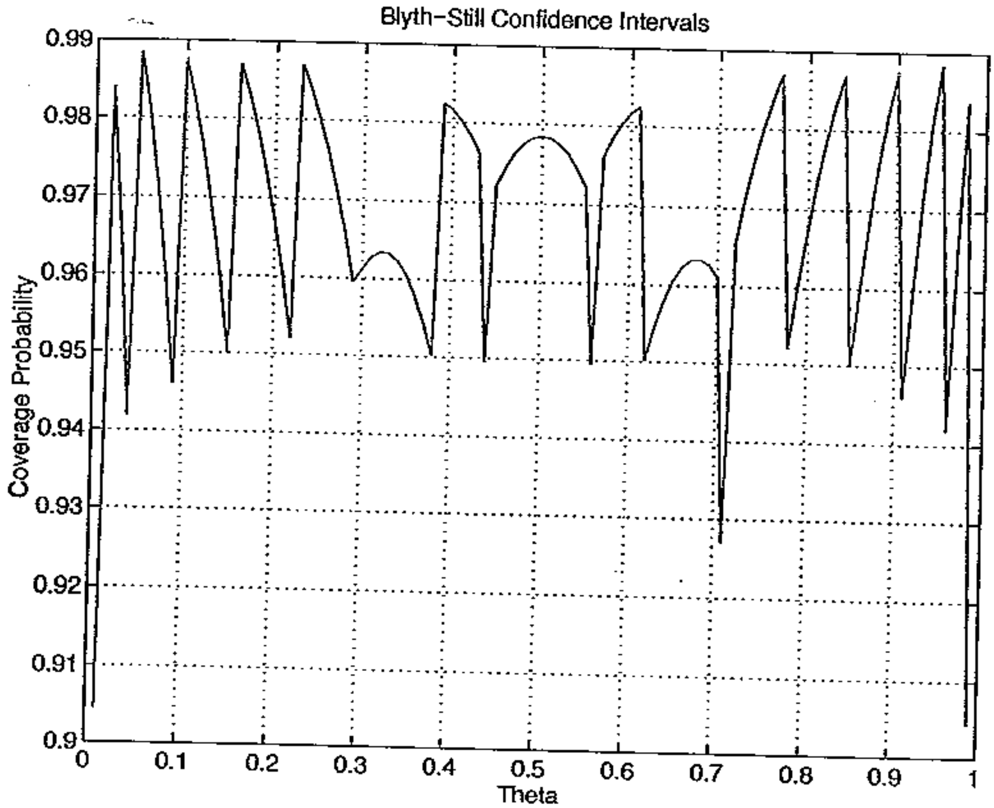

Figure 3: Graph of the coverage probabilities for Blyth-Still confidence intervals for $n = 10$.

property of minimizing the sum of the $n + 1$ possible lengths. All the confidence intervals they calculated for $1-\alpha = .95, .99$ and $n = 1, 2, ..., 30$ satisfy all three restrictions. Figure 3 is a graph of the coverage probabilities of the Blyth-Still published intervals for $n = 10$. Notice that the graph is not symmetric about $\theta = .5$ and it also dips below the 95% confidence level. These aberations are artifacts. Blyth and Still (1983) only tabulated their confidence intervals to two decimal places. Greater accuracy in the published tables would result in symmetric coverage that does not dip below 95%.

Casella Intervals

Casella (1986) presents an algorithm which, when applied, improves upon any set of equivariant confidence intervals that have minimum coverage probability greater than or equal to $1-\alpha$. Let $C = \{[\ell_x, u_x], x = 0,1,...,n\}$ be a set of equivariant confidence intervals. Let $C^* = \{[\ell_x^*, u_x^*], x = 0,1,...,n\}$ be the set of improved equivariant confidence intervals which result when the algorithm is applied to C. An improved set of confidence intervals satisfys $u_x^* - \ell_x^* \le u_x - \ell_x \ \forall x \in \{0,1,...,n\}$ All confidence intervals generated are to be closed, of the form $[\ell_X, u_X]$. But, in calculating coverage probabilities, Casella considers the confidence intervals to be half open, of the form $(\ell_X, u_X]$, where the coverage probability is given by

$$P(\ell_X < \theta \le u_X) = \sum_{x=0}^{n} \binom{n}{x} \theta^x (1-\theta)^{n-x} I_{(\ell_x, u_x]}(\theta).$$

The algorithm is as follows:

For each $k = n, n-1, ..., 1$ increase ℓ_k (simultaneously decrease $u_{n-k} = 1 - \ell_k$ by the same amount) until one of the following occurs:

(a) $P(\ell_X < \ell_k \le u_X) = 1 - \alpha$.

(b) $\ell_k = u_j$ for some j.

If (a) occurs first, set $\ell_k^* = \ell_k$, decrement k, and start again. If (b) occurs first, check if $P(\ell_X < \ell_k \le u_X) > 1 - \alpha$. If this inequality holds, continue increasing ℓ_k until (a) or (b) occurs again. If this inequality does not hold, set $\ell_k^* = u_j$ and move to the next value of k. Whenever a ℓ_k^* is defined, a u_{n-k}^* is defined.

The new set of confidence intervals C^* has uniformly shorter length than C. All of the confidence interval lengths, $u_X^* - \ell_X^*$, are at least as small as the origional lengths $u_X^* - \ell_X^*$, that is $u_x^* - \ell_x^* \le u_x^* - \ell_x^*$ for $x = 0,1,...,n$. This is easily shown using the fact that C^* is equivariant and the algorithm increases the lower endpoints:

$$u_x^* - \ell_x^* = 1 - \ell_{n-x}^* - \ell_x^* \leq 1 - \ell_{n-x} - \ell_x = u_x - \ell_x.$$

Casella also proved that the sum of the lengths of the $n+1$ confidence intervals in C^* for a given $1-\alpha$ confidence level is minimized.

Once C^* is produced by the algorithm, an entire family of equivariant confidence intervals can be generated. This family is complete (Casella and Berger,1990). That is, any C not in this family can be dominated by a C^* in this family. The family contains more than one menber because of the properties of a certain set of endpoints which Casella refers to as coincidental endpoints. Specifically, ℓ_k^* is a coincidental endpoint if $\ell_k^* = u_m^*$ for some m. Consider a particular C^* where ℓ_k^* is a coincidental endpoint. Denote the common value of ℓ_k^* and u_m^* by r. The coverage probabilities of ℓ_k^* and u_m^* are given by

$$P(\ell_X < \ell_k^* \leq u_X) = \sum_{x=m}^{k-1} \binom{n}{x} r^x (1-r)^{n-x}$$

$$P(\ell_X < u_m^* \leq u_X) = \sum_{x=m+1}^{k} \binom{n}{x} r^x (1-r)^{n-x}.$$

The two coverage probabilities are not equal because of the half open confidence intervals. Now define

$$m(r) = \min\{P(\ell_X < \ell_k^* \leq u_X), P(\ell_X < u_m^* \leq u_X)\}.$$

Then, ℓ_k^* can take on any value such that $r_* \leq \ell_k^* \leq r^*$ where

$$r_* = \min\{r : m(r) \geq 1-\alpha\} \text{ and } r^* = max\{r : m(r) \geq 1-\alpha\}.$$

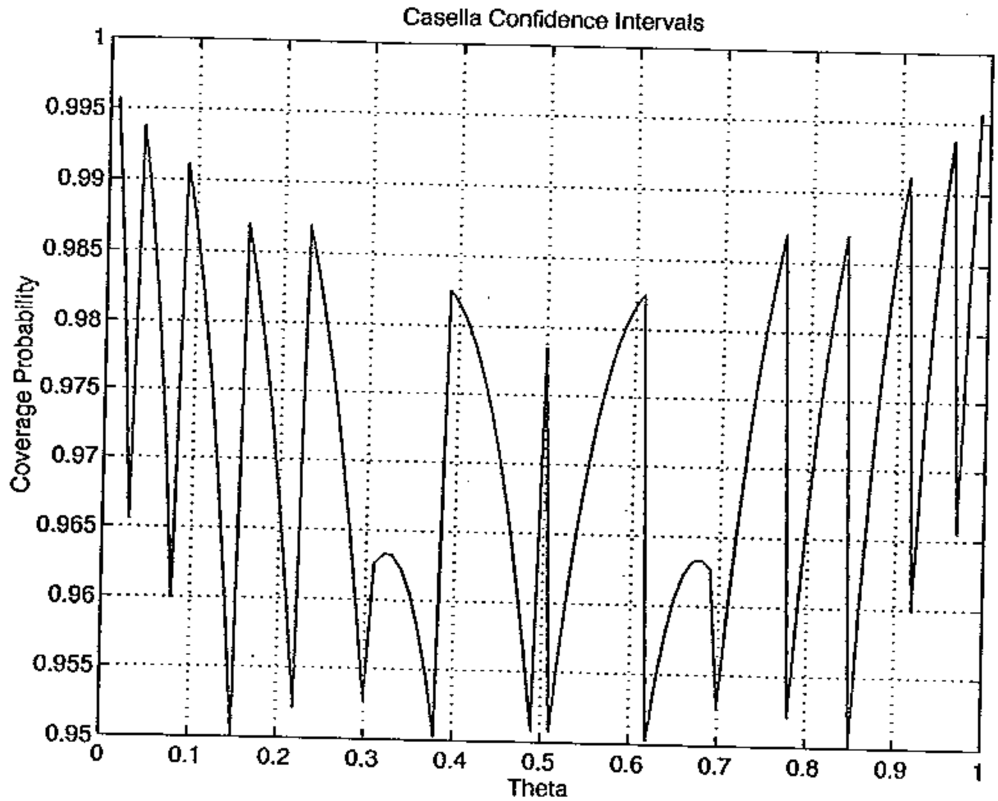

Figure 4: Graph of the coverage probabilities for Casella confidence intervals for $n = 10$.

The Blyth-Still intervals are members of the family of Casella intervals. They are obtained by setting the coincidental points equal to the midpoints of their range. The Crow intervals can be approximated by chosing all coincidental lower endpoints to have their minimum value. The Crow intervals agree with the family of Casella intervals on all noncoincidental endpoints. There are some Crow interval endpoints that do not fall within the range of the Casella coincidental endpoints. Even though some sets of Crow intervals fall within the Casella intervals, Crow intervals in general cannot be considered members of the family of Casella intervals because, as it was shown, they violate the monotonicity condition. The Crow intervals coverage probability equals $1 - \alpha$ for at least one θ and the sum of the lengths of the intervals is minimized and equal to the sum of the lengths of the Casella intervals.

To illustrate a complete family of equivariant confidence intervals using the above calculations, let $n = 10$ and $1 - \alpha = .95$. then, the following table gives the complete family of confidence intervals:

x	1	2	3	4	5	6	7	8	9	10
ℓ_x	.005	.037	.087	.150	.222	$.281 \pm .022$	.381	$.444 \pm .049$	$1 - \ell_8$	$1 - \ell_6$

The upper endpoints are calculated by $u_x = 1 - \ell_{n-x}$, and $\ell_0 = 0$. The coincidental endpoints occir at $x = 6, 8, 9, 10$. The table shows all the possible values these coincidental endpoints may take. Figure 4 illustrates the coverage probabilites of the above set of confidence intervals where the coincidental endpoints are chosen such that $\ell_6 = .303$ and $\ell_8 = .493$. If one has a set of confidence intervals that are not in Casella's complete family of confidence intervals, they can be improved upon by using the algorithm. The resultant set of intervals will then be in Casella's complete family. So, if one considers the loss function to be the sum of the confidence intervals, one can do no better than to use a set of confidence intervals tabled by Casella (1986).

Comparisons

All of the procedures discussed satisfy restrictions one through three, some actually attain the $1 - \alpha$ confidence level for particular θ. The casella intervals were shown to be superior in that the sum of the lengths of the $n + 1$ confidence intervals is as small as or smaller than any other method. Are the improvements over the initial set of confidence intervals given by Clopper and Pearson nontrivial? Is it worthwile to use updated tables to find confidence intervals? The following calculations help answer these questions.

One comparison that has been discussed is the length of confidence intervals. For a set of confidence intervals $[\ell_x, u-x], x = 0, 1, ..., n$ define the length $L(X)$ to be $L(x) = u_x - \ell_x, x = 0, 1, ..., n$. One can take the expected value of $L(X)$ and use it to compare different sets of confidence intervals. The expected value of $L(X)$ is defined to be

$$E[L(X)|\theta] = \sum_{x=0}^{n} L(x) \binom{n}{x} \theta^x (1-\theta)^{n-x}.$$

Figure 5 is a graph of the expected lengths of the Clopper-Pearson, Crow, Blyth-Still, and two particular sets of the Casella confidence intervals for $n = 10$ and $1 - \alpha = .95$. The set of Casella Max. intervals was chosen by setting the coincidental endpoints equal to their maximum value and the set of Casella Min. intervals was chosen by setting the coincidental endpoints equal to their minimum value. In figure 5, for $\theta = .5$, the Casella Max intervals are represented by the lower solid line and the Casella Min. intervals by the upper solid line. It is interesting to notice that the graphs of the Casella intervals completely bracket the graphs of the Blyth-Still and Crow intervals. Depending upon prior knowledge, one would be inclined to choose a different set of confidence intervals for different circumstances. If one knew θ to be near .5 then one should choose the Casella Max. intervals. If θ is thought to lie near 0 or 1 then one should use the Casella Min. intervals. The Blyth-Still intervals would be used if there was no prior knowledge about θ. It is interesting to note that the sum of the lengths of the Clopper-Pearson intervals, for $n = 10$ and $1 - \alpha = .95$, is 5.59 and the sum of the lengths for the other three, though minimal, is only 6.39% shorter. This is not a significant improvement.

To compare the methods on coverage probabilities, one can calculate the the area under the function $P(\ell_X \leq \theta \leq u_X)$ for $\theta \in (0, 1)$. A method that has $(1-\alpha)100\%$ coverage

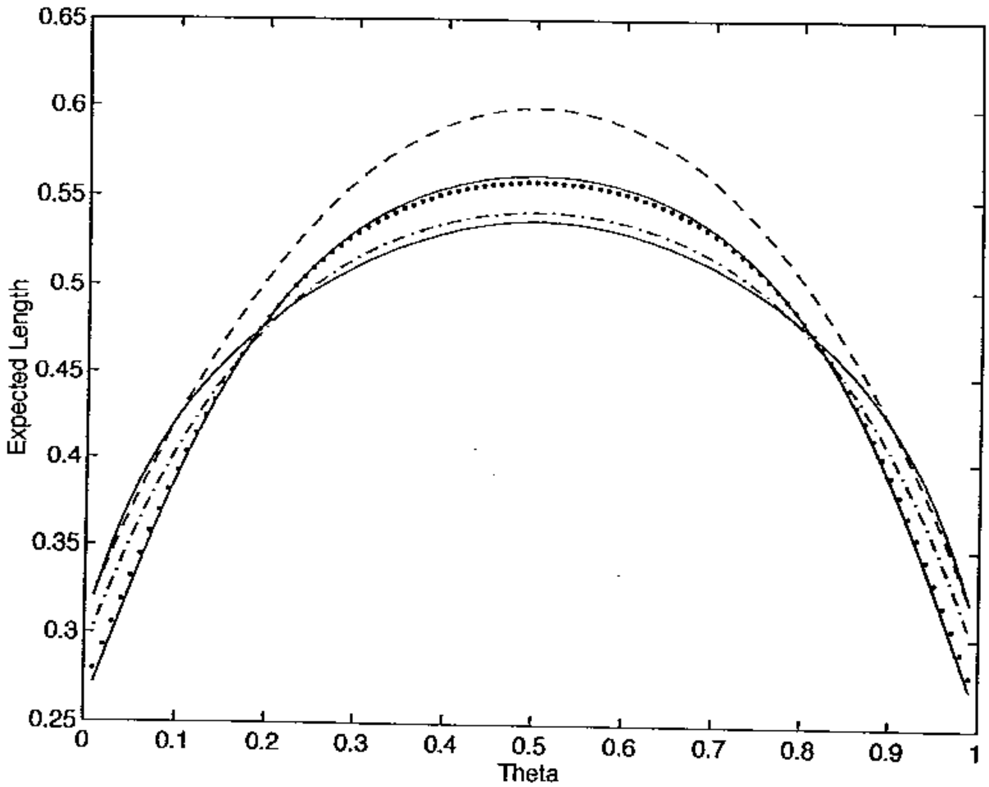

Figure 5: Graph of the expected lengths of the confidence intervals of Casella Max. and Min. (solid lines), Blyth-Still (dash dot line), Crow (dot line), and Clopper-Pearson (dash line), for $n = 10$ and $1 - \alpha = .95$.

probability for all θ has area $1 - \alpha$. For the above methods, which satisfy $\inf_\theta P(\ell_X \leq \theta \leq u_X) \geq 1 - \alpha$, one can subtract $1 - \alpha$ from the area calculations. For $\alpha = .05$, the areas are calculated as follows:

$$\begin{aligned}
A &= \int_0^1 \Big[\sum_{x=0}^{n} \binom{n}{x} \theta^x (1-\theta)^{n-x} I_{[\ell_x, u_x]}(\theta) - .95\Big] d\theta \\
&= \sum_{x=0}^{n} \binom{n}{x} B(x+1, n-x+1) \int_{\ell_x}^{u_x} \frac{\theta^x (1-\theta)^{n-x}}{B(x+1, n-x+1)} d\theta - .95 \\
&= \sum_{x=0}^{n} \binom{n}{x} \frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n-x+1+x+1)} \int_{\ell_x}^{u_x} f_X(x|\theta) d\theta - .95 \\
&= \sum_{x=0}^{n} \Big(\frac{n!}{x!(n-x)!}\Big)\Big(\frac{x!(n-x)!}{(n+1)!}\Big) \int_{\ell_x}^{u_x} f_X(x|\theta) d\theta - .95
\end{aligned}$$

$$= \frac{1}{n+1}\sum_{x=0}^{n}\int_{\ell_x}^{u_x} f_X(x|\theta)d\theta - .95,$$

where $f_X(x|\theta)$ is the pdf of Beta$(x+1, n-x+1)$.

Areas were calculated for the four methods for $n = 10$ and $1-\alpha = .95$. They are

Clopper	Crow	Blyth	Casella Max.	Casella Min.
.0338	.0214	.0221	.0220	.0208

where the coincidental endpoints were chosen at there maximum values of $\ell_6 = .303$ and $\ell_8 = .493$ for Casella Max. and chosen at their minimum values of $\ell_6 = .259$ and $\ell_8 = .381$ for Casella Min. The Casella Min. intervals have the smallest area. If one considered the loss function to be the area, then the Casella Min. intervals would be superior to the other methods for this example.

Example

The following example (Steel and Torrie, 1960) illustrates the different confidence interval procedures. A class in taxonomy near Itheca, NY studied *Dentaria*, a flowering plant. The students observed 26 plants and noted which flowered and which did not to estimate the proportion of flowering plants. They found that 6 flowered and 20 did not. The students were interested in generating exact 95% confidence intervals for the proportion of flowering plants. For $n = 26$ and $x = 6$, the confidence intervals for the four intervals described are

Clopper	Crow	Blyth	Casella Max.	Casella Min.
(.0897, .4365)	(.106, .421)	(.11, .42)	(.106, .436)	(.106, .406)

where the values were quoted from each authors published tables. Notice that the Clopper-Pearson interval is the longest, the Crow and Blyth-Still intervals are essentially the same length, and the Casella interval is either the shortest or the second longest. By choosing the upper coincidental endpoint to be as small as possible (Casella Min.), the Casella interval is shorter than the rest. But, by choosing the upper coincidental endpoint to be as large as possible (Casella Max.) one obtains a confidence interval longer than Crow's or Blyth's. So, it would be to the students advantage to choose the Casella Min. interval in this case.

Conclusion

The Clopper-Pearson intervals are still probabily the most widely used tabled set of exact confidence intervals. It was shown that the sum of lengths of the confidence intervals (for a particular n and α) for the other three methods discussed was shorter than for the Clopper-Pearson intervals. However, this difference in sum of lengths is not very substantial. If possible, one should choose confidence intervals from the family of Casella intervals. This not only guarentees that the interval selected is from the family of confidence intervals with the shortest sum of lengths (for a particular n and α) but, also affords the experimenter some flexibility due to the coincidental endpoints. As demonstrated in the preceeding example, if ones confidence interval contains a coincidental endpoint, it is possible to choose the endpoint that yields the shortest interval. This choice has to be made before the data are observed. For example, if one believed that θ lies near 1, then one could choose the coincidental endpoints to yield shorter intervals for larger X (Casella, 1986).

There is an interesting result in the graph of the expected lengths (figure 5). There are two points at approximately $\theta = .2$ and $\theta = .8$ where the expected lengths of the Crow, Blyth-Still and Casella method converge. Why do these two points exist, and do they exist for all n? These questions and the importance of these two points are the subjects of possible future research.

References

[1] BICKEL, P. J., and DOKSUM, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics.* Hoden-Day, Inc., Oakland, California.

[2] BLYTH, C. R., and STILL, H. A. (1983). Binomial confidence intervals. *Journal of the American Statistical Association*, 78, 108 – 116.

[3] CASELLA, G. (1986). Refining binomial confidence intervals. *The Canadian Journal of Statistics*, 14, 113 – 129.

[4] CASELLA, G. and BERGER, R. L. (1990). *Statistical Inference.* Wadsworth, Inc., Belmont, California.

[5] CROW, E. L. (1956). Confidence intervals for a proportion. *Biometrika*, 43, 423 – 435.

[6] DILLER, A. (1993). *Latex, Line By Line.* John Wiley & Sons Ltd, Westsussex , England.

[7] STEEL, R. G. D., and TORRIE, J. H. (1960). *Principles and Procedures of Statistics.* Mc-Graw-Hill, New York.

[8] VOLLSET, S. E. (1993). Confidence intervals for a binomial proportion. *Statistics In Medicine*, 12, 809 – 824.