

BU-827-m

Refining Binomial Confidence Intervals

GEORGE CASELLA*

A method for refining an equivariant binomial confidence procedure is presented which, when applied to an existing procedure, produces a new set of equivariant intervals that are uniformly superior. The family of procedures generated from this method constitute a complete class within the class of all equivariant procedures. In certain cases it is shown that this class is also minimal complete. Also, an optimality property, monotone minimaxity, is investigated and monotone minimax procedures are constructed.

KEY WORDS: Binomial parameter; Confidence procedures; Admissibility; Minimax.

*George Casella is Associate Professor, Biometrics Unit, Cornell University, Ithaca, NY 14853. This research was supported by National Science Foundation Grant No. MCS83-00875. This paper is BU-827-M in the Biometrics Unit Series, Cornell University.

1. INTRODUCTION

The problem of constructing confidence intervals for a binomial success probability has a long history, dating back to the work of Clopper and Pearson (1934). Although the problem seems to be a simple one, it is made complicated by the fact that we are working with a discrete distribution, and standard decision-theoretic techniques tend to break down under such circumstances.

A mathematical trick sometimes employed, to overcome the problems in dealing with a discrete distribution, is to add an independent uniform $(0,1)$ random variable to the binomial random variable, thereby creating a variable with a continuous distribution (see, for example, Stevens (1950), Lehmann (1959), or Blyth and Hutchinson (1960)). This trick should be treated with disdain: while it produces a problem that is tractable theoretically, it produces a solution that is useless in application. Here we will consider only nonrandomized confidence intervals.

The Clopper-Pearson confidence intervals, which are still in use today, are constructed by intersecting one-sided lower and upper intervals. The problem with this construction is that, in many cases, the true confidence coefficient (the infimum of the coverage probabilities) is strictly greater than the stated level. The Clopper-Pearson intervals are, therefore, too wide for the stated level of confidence, and leave room for improvement. In order to do so, however, it seems that one must consider alternative methods of construction.

Significant contributions to the construction of binomial confidence intervals were made by Sterne (1954), Crow (1956) and, most recently, Blyth and Still (1983). Sterne proposed a method for constructing shortest acceptance regions, which could then be inverted to produce confidence

regions. Crow refined Sterne's method to eliminate confidence regions that were not intervals, and also showed that this method produces a set of confidence intervals which minimizes the sum of the lengths. These intervals, referred to as Sterne-Crow intervals, have enjoyed fairly widespread use, and have been extensively tabled (Natrella, 1963).

As pointed out by Blyth and Still, however, the Sterne-Crow intervals contain many irregularities. (For example, in some cases, when the observed number of successes increases, the lower endpoint of the confidence interval remains unchanged.) Such problems result from the fact that, for many parameter values, there is more than one shortest acceptance region, and Crow used a somewhat arbitrary rule for selecting the one to use.

The technique of Blyth and Still was to generate all shortest acceptance regions for success probabilities that are multiples of .005. After eliminating those acceptance regions which would produce disconnected confidence regions, five rules were examined for deciding how to handle the nonunique shortest acceptance regions. The criterion decided upon was to select as the confidence limit the midpoint of possible values. The resulting procedure, which retains the property of minimizing the sum of the lengths, is also approximately unbiased with approximately equal probability tails.

In this paper binomial confidence intervals are constructed by a direct method, working with the intervals themselves rather than inverting acceptance regions. An algorithm is developed that, when applied to an existing confidence procedure, produces a new procedure that is a uniform improvement, in the sense of having uniformly shorter length for the same confidence coefficient. This refinement algorithm actually produces not

one procedure, but a family of procedures, where each member of the family has the property of minimizing the sum of the interval lengths. This family is a complete class of procedures when the risk is measured by expected length. In certain cases it is a minimal complete class.

The refinement construction is actually equivalent to a continuous version of the Blyth-Still construction, but has the advantage of being somewhat more natural and easier to work with. (The Blyth-Still intervals are therefore members of the complete class.) In the absence of prior information there is little reason to prefer one member of the complete class over another, so tables of the entire class are given for 95% and 99% confidence levels.

We also introduce and examine a particular optimality criterion, monotone minimaxity, which is a reasonable criterion if there is some belief that the success probability is 'in the middle.' Monotone minimax intervals are constructed and reported for 95% and 99% confidence levels.

Section 2 contains the necessary notation and a few preliminaries, and in Section 3 the refinement algorithm is defined and properties of a refined procedure are given. In Section 4 the family of refined intervals is constructed and examined, while Section 5 treats monotone minimax intervals. There is also an Appendix, which contains the proof of Theorem 4.1.

2. Preliminaries

Let X be a binomial random variable based on n trials with success probability θ , that is,

$$P(X = x|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} . \quad (2.1)$$

A confidence procedure C is a collection of $n+1$ intervals $[l_x, u_x]$, $x=0, \dots, n$. The coverage probability of a confidence procedure C is the probability that the random interval $[l_X, u_X]$ covers the true parameter value, and is given by

$$P(\theta \in C | \theta) = \sum_{x=0}^n I_{[l_x, u_x]}(\theta) \binom{n}{x} \theta^x (1-\theta)^{n-x} . \quad (2.2)$$

When using a confidence procedure one would like to assert a minimum coverage probability, i.e., a number $1-\alpha$ such that

$$\inf_{\theta} P(\theta \in C | \theta) \geq 1 - \alpha \quad (2.3)$$

If (2.3) holds for a procedure C , we usually call C a $1-\alpha$ confidence procedure. It may be the case that the inequality in (2.3) is strict so that the true confidence coefficient ($= \inf_{\theta} P(\theta \in C | \theta)$) is larger than the nominal level ($1-\alpha$), however the nominal level is the number usually asserted. Such procedures are called conservative and, in the binomial case, are clearly non-optimal since, for the specified $1-\alpha$, one can construct a dominating procedure simply by narrowing the intervals.

Throughout this paper we will only deal with intervals that satisfy the following two conditions:

1. Equivariance

Since the binomial distribution is invariant under the transformations $X \rightarrow n-X$ and $\theta \rightarrow 1-\theta$, we require that the confidence intervals be equivariant under these transformations, that is,

$$l_x = 1 - u_{n-x} \quad x=0, \dots, n . \quad (2.4)$$

Thus a confidence procedure $C = \{[l_x, u_x], x=0, \dots, n\}$ is uniquely determined by its lower endpoints. Although the adjective "equivariant" is sometimes omitted in what follows, it is to be implicitly assumed that all statements only apply to equivariant intervals.

2. Monotonicity of Endpoints

For fixed n , we require $l_{x+1} > l_x$ and $u_{x+1} > u_x$ for $x=0,1,\dots,n-1$. This is an intuitively appealing requirement, and is one of the four desiderata listed by Blyth and Still (1983). Also, although we offer no proof, it seems reasonable to conjecture that intervals without this property are inadmissible.

Confidence intervals are always defined as closed intervals, as they have been defined here. In theory, this is fine, and in practice, there is no real difference between considering the intervals open or closed. With the binomial distribution, however, technical difficulties can arise when locating the relative minima of coverage probabilities. Without going into details, we find it easiest to consider the intervals to be half open, of the form $(l,u]$, in our calculating formulas. This adjustment has no effect on the use of the intervals, where we take them to be closed.

3. Refined Intervals

3.1 The Refinement Algorithm

We now describe a simple algorithm which, when applied to any $1-\alpha$ binomial confidence procedure, will produce a new $1-\alpha$ confidence procedure with uniformly shorter length. The strategy is simply to move all the lower endpoints of the intervals as far to the right as possible.

Starting with a confidence procedure $C = \{(l_x, u_x], x=0, \dots, n\}$ a refined procedure, $C^* = \{(l_x^*, u_x^*), x=0, \dots, n\}$ will be constructed. In the implementation of the following algorithm, the equivariance constraint is to be maintained at all stages. Thus, if any l_k is increased, the corresponding $u_{n-k} = 1 - l_k$ is decreased by the same amount.

The Algorithm: For each $k=n, n-1, \dots, 1$ increase λ_k until one of the following occurs:

a) $P(\lambda_k \in C | \lambda_k) = 1 - \alpha$

b) $\lambda_k = u_j$ for some j .

If a) occurs first, set $\lambda_k^* = \lambda_k$, decrement k , and start again. If b) occurs first, check if $P(\lambda_k \in C | \lambda_k) > 1 - \alpha$. If this inequality holds continue increasing λ_k until a) or b) occurs again. If this inequality does not hold, set $\lambda_k^* = u_j$ and move to the next value of k .

In practice one could use a bisection method to solve a), and would also specify the precision desired in the equalities a) and b). It is easy to see that the algorithm stops after the n steps have been executed. The function $P(\lambda_k \in C | \lambda_k)$ must eventually start decreasing as λ_k increases, and since it is at least $1 - \alpha$ at the start of the step, either a) or b) must occur.

According to the strict definition of algorithm (Knuth, 1973, p. 4), the above description fails since it doesn't consist of a finite number of operations. (Knuth would call the above procedure a "computational method"). However, we use the word here, in a less formal sense, to mean "a procedure for solving a mathematical problem that involves repetition of an operation."

3.2 Properties of a Refined Confidence Interval

Starting with a confidence procedure $C = \{[\lambda_x, u_x], x=0, \dots, n\}$, let $C^* = \{[\lambda_x^*, u_x^*], x=0, \dots, n\}$ be a refinement of C . Since C^* is obtained by increasing lower endpoints, and since the procedures are invariant, it immediately follows that

$$u_x^* - \lambda_x^* \leq u_x - \lambda_x \quad x=0, 1, \dots, n \quad (3.1)$$

since

$$u_x^* - l_x^* = 1 - (l_{n-x}^* + l_x^*) \leq 1 - (l_{n-x} + l_x) = u_x - l_x \quad (3.2)$$

Thus, C^* obtains a uniform length reduction over C . C^* also has the property of minimizing the sum of the $n+1$ lengths over all $1-\alpha$ invariant intervals, a result which follows from the following interesting properties of C^* .

For each l_k^* of C^* , call l_k^* a coincidental endpoint (or just coincidental) if $l_k^* = u_m^*$ for some m . Let $A = \{x: l_x^* \text{ is coincidental}\}$. We then have the following lemmas.

Lemma 3.1 If $C^* = \{[l_x^*, u_x^*]\}$ is a refined $1 - \alpha$ confidence procedure,

$$\sum_{x=0}^n (u_x^* - l_x^*) = n+1 - 2 \sum_{x \in A} 1 - 2 \sum_{x \in A^c} l_x^* \quad (3.3)$$

That is, the sum of the $n+1$ lengths of C^* is independent of the placement of the coincidental points.

Proof: Since C^* is invariant $u_x^* = 1 - l_{n-x}^*$, hence

$$\sum_{x=0}^n (u_x^* - l_x^*) = \sum_{x=0}^n (1 - l_{n-x}^* - l_x^*) = n+1 - 2 \sum_{x=0}^n l_x^* \quad (3.4)$$

Now if l_k^* is coincidental, we have $l_k^* = u_m^*$ for some m , and, hence $l_k^* = 1 - l_{n-m}^*$, or $l_k^* + l_{n-m}^* = 1$. It also follows that l_{n-m}^* is also coincidental, since $l_{n-m}^* = 1 - u_m^* = 1 - l_k^* = u_{n-k}^*$. Substituting into (3.4) establishes (3.3). \parallel

Lemma 3.2: The non-coincidental endpoints of a $1-\alpha$ refined confidence procedure are unique.

Proof: Let C and C' be two refined $1-\alpha$ confidence procedures, and let λ'_k be any non-coincidental endpoint of C' . Also, let u'_m satisfy

$$u'_m = \min_i \{u'_i : u'_i > \lambda'_k\} \quad , \quad (3.5)$$

that is, u'_m is the smallest upper endpoint greater than λ'_k . We first show that, in the procedure C , u_m is the smallest upper endpoint greater than λ_k . If $u_m < \lambda_k$, then, according to the algorithm, λ_k can be increased beyond u_m and $1-\alpha$ confidence is retained, that is, for some $\delta > 0$,

$$P(\lambda_k + \delta \in C | \lambda_k + \delta) > 1-\alpha \quad \text{for} \quad \lambda_k = u_m \quad ,$$

or, in terms of the calculating formula,

$$\sum_{x=m+1}^{k-1} \binom{n}{x} (\lambda_k)^x (1-\lambda_k)^{n-x} > 1-\alpha \quad . \quad (3.6)$$

But if this were the case, λ'_k would be greater than u'_m , a contradiction. Hence $\lambda_k < u_m$.

Thus, both λ'_k and λ_k must satisfy

$$g(\lambda'_k) = g(\lambda_k) = 1 - \alpha \quad , \quad (3.7)$$

where $g(\theta) = \sum_{x=m}^{k-1} \binom{n}{x} \theta^x (1-\theta)^{n-x}$. If $m = 0$, $g(\theta)$ is strictly decreasing in θ , and it follows that we must have $\lambda'_k = \lambda_k$. If $m > 0$, the derivative of $g(\theta)$ can be written

$$\frac{d}{d\theta} g(\theta) = n\theta^{m-1}(1-\theta)^{n-m} \left[\binom{n-1}{m-1} - \binom{n-1}{k-1} \left(\frac{\theta}{1-\theta}\right)^{k-m} \right] \quad , \quad (3.8)$$

showing that $g(\theta)$ has a unique maximum, and the equation $g(\theta) = 1-\alpha$ can have at most two roots. However, the algorithm will always choose the rightmost root, so once again we have $\lambda'_k = \lambda_k$ and the lemma is established. \parallel

From Lemmas 3.1 and 3.2 we can conclude that, for a given α , all refined confidence procedures have the same value for the sum of the lengths of the intervals. This fact, together with the fact that the refinement algorithm uniformly decreases the interval lengths, gives us the following theorem.

Theorem 3.1: Any $1-\alpha$ refined confidence procedure has the property of minimizing the sum of the lengths of the individual intervals.

4. Families of Refined Procedures

For any given input, the refinement algorithm produces a single refined procedure. However, from this single procedure we can generate an entire family of equivalent procedures. Since the coincidental points may not be uniquely defined, any coincidental point that satisfies the probability constraint is allowable. By specifying the range of these allowable coincidental points we can define a family of refined intervals.

Suppose λ_k is a coincidental endpoint of a refined confidence procedure C , say $\lambda_k = u_m = r$. The coverage probabilities at λ_k and u_m are then

$$P(\lambda_k \in C | \lambda_k) = \sum_{x=m}^{k-1} \binom{n}{x} r^x (1-r)^{n-x}$$

and

(4.1)

$$P(u_m \in C | u_m) = \sum_{x=m+1}^k \binom{n}{x} r^x (1-r)^{n-x} ,$$

and each of these quantities is greater than $1-\alpha$. Note that even though $\lambda_k = u_m$, the two coverage probabilities in (4.1) are not equal because we are working with half-open intervals, as mentioned in Section 2. If we now define

$$m(r) = \min\{P(\lambda_k \in C | \lambda_k), P(u_m \in C | u_m)\} \quad , \quad (4.2)$$

then the allowable range of values for λ_k is $r_* \leq \lambda_k \leq r^*$, where

$$\begin{aligned} r^* &= \max\{r : m(r) \geq 1-\alpha\} \\ r_* &= \min\{r : m(r) \geq 1-\alpha\} \quad . \end{aligned} \quad (4.3)$$

This calculation can be done for each coincidental endpoint of C , the result being a family of refined intervals which we will denote by R_α .

Each procedure in R_α is a $1-\alpha$ refined confidence procedure, so it enjoys the minimum length property of the previous section. Moreover, the family R_α is a complete class of procedures; any procedure not in R_α can be dominated by a procedure in R_α . We now make this statement precise.

A natural measure of size for a binomial confidence procedure is its length. If we let $L(x) = u_x - \lambda_x$, $x=0,1,\dots,n$ where the collection $\{(\lambda_x, u_x)\}$ is some confidence procedure C , we can measure the risk of C with its expected length $E_\theta L(C)$, where

$$E_\theta L(C) = \sum_{x=0}^n L(x) \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad . \quad (4.4)$$

If, for fixed α , a confidence procedure C^0 is not in R_α , the arguments in the previous subsection show that C^0 must have at least one noncoincidental endpoint that is different from those in R_α . This means that the refinement algorithm will produce a nontrivial decrease in the length of at least one interval of C^0 , so it immediately follows that there exists $C^* \in R_\alpha$ with the property

$$\begin{aligned} E_\theta L(C^0) &\geq E_\theta L(C^*) \quad \text{for all } \theta \\ E_\theta L(C^0) &> E_\theta L(C^*) \quad \text{for some } \theta \quad , \end{aligned} \quad (4.5)$$

showing that C^0 is inadmissible. Thus, R_α contains all the $1-\alpha$ admissible confidence procedures, and hence is a complete class.

Table 1 gives the families R_α for $\alpha = .05, .01$ and $n = 6, 30$. (For $n \leq 5$, R_α has a unique member, with endpoints given by the intersection of two $1-\alpha$ level one-sided procedures, and agrees with the Blyth-Still intervals. In general, for each n there exists a sufficiently small α such that R_α is unique: one need only choose α to require the intervals so wide that $k_n < u_0$.)

Although R_α itself may contain inadmissible procedures, in general the members of R_α are not comparable; their risk functions will cross. The family of intervals provides great flexibility for an experimenter in allowing him to choose that procedure which is most suited to a particular experiment. For example, if it is believed that θ lies near 1, one would like to use a procedure which gives shorter intervals for larger X , and the coincidental points can be chosen to reflect this.

When examining the range of values of the coincidental points given in Table 1, it might at first appear that one can reverse the order of the endpoints. For example, for $n = 25$, $\alpha = .05$, $k_{14} = .354 \pm .024$, $k_{15} = .396 \pm .021$, so it seems that one can choose $k_{14} = .354 + .024 = .378$ and $k_{15} = .396 - .021 = .375$. However, according to our requirements we must always have $k_{14} < k_{15}$. It is possible to choose $k_{14} = .378$ or $k_{15} = .375$, as long as the other endpoint does not overlap.

When one obtains a complete class of procedures, a natural question is whether the class is also minimal, i.e., are all the rules in the class admissible? This is usually a difficult question to answer, and is not answered in general here. However, in certain situations, the class R_α is a minimal complete class.

Theorem 4.1: For fixed α , if the class R_α consists of procedures with only one pair of coincidental lower endpoints, then R_α is a minimal complete class.

Proof: Given in the Appendix.

If there is more than one pair of coincidental points, the question of whether R_α is still a minimal complete class is much more difficult to deal with. The risk functions become sufficiently complex so that determining whether or not two distinct ones must cross is quite involved. It seems likely, however, that risk functions of distinct procedures in R_α must either cross or be everywhere the same.

The family of refined intervals, for a fixed α , can also be generated by an acceptance region inversion scheme similar to that used by Blyth and Still (1983). It is sometimes the case that the shortest acceptance regions are not unique, which is equivalent to our occurrence of coincidental endpoints. The intervals of Blyth and Still are obtained by setting the coincidental points equal to the midpoint of their range, and hence are contained in Table 1.

5. Monotone Minimax Procedures

All of the procedures in R_α have the property of minimizing the sum of the lengths, so there is no basis for comparison on this measure. In terms of expected length, however, there is room for comparison, and a natural property to inquire about is minimaxity.

Definition 5.1: A $1-\alpha$ confidence procedure C is said to be *minimax* if, for any procedure C' satisfying $\inf_{\theta} P(\theta \in C' | \theta) \geq 1-\alpha$, $\max_{\theta} E_{\theta} L(C) \leq \max_{\theta} E_{\theta} L(C')$.

In dealing with the binomial distribution, some care must be taken when considering minimaxity. For a procedure $C = \{[l_x, u_x], x=0, \dots, n\}$, we have

$$E_{\theta} L(C) = \sum_{x=0}^n (u_x - l_x) \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad (5.1)$$

so the intervals corresponding to large values of $\binom{n}{x}$, i.e., those with x values near $n/2$, will receive a larger weight in the sum. Thus, one would expect a minimax procedure to have shorter intervals for x near $n/2$, and longer intervals for x near 0 or n .

This situation is contrary to our intuition since $\text{Var } X = n\theta(1-\theta)$ is largest for $\theta = \frac{1}{2}$ and smallest for $\theta = 0$ or 1, we would expect confidence intervals to be longer for x near $n/2$ (which is evidence that $\theta \approx \frac{1}{2}$) than for x near 0 or n . However, the minimax procedure will maintain $1-\alpha$ confidence by having wider outer intervals and narrower inner intervals. For example, if $n=4$, $1-\alpha=.5$, the (equivariant) minimax procedure has lower endpoint $l_0=0$, $l_1=.159$, $l_2=.335$, $l_3=.500$, $l_4=.665$ yielding interval lengths .335, .341, .330, .341, .335, respectively. Intuitively speaking, if the middle intervals are shortened, but the procedure is maintaining a $1-\alpha$ confidence coefficient, it is doing so at the expense of the outer intervals.

There are many situations, however, where it is reasonable to suspect that θ is a 'middle value,' and the question remains as to how to construct optimal expected length intervals for this situation, and not have the problem of using intervals that are 'too short.' A solution to this problem is to require the confidence procedure to retain a length ordering which mimics the variance structure.

Definition 5.2: A confidence procedure $C = \{[l_x, u_x], x=0, \dots, n\}$ is said to be *monotone* if $u_x - l_x$ is a nondecreasing function of x for $x \leq n/2$ and is a nonincreasing function of x for $x \geq n/2$.

Monotone procedures will not allow the middle intervals to become shorter than the outer ones, and from within the class of monotone procedures, we now seek a minimax procedure.

Determining the monotone minimax procedure is a relatively easy task, in direct contrast to the much more difficult problem of determining the minimax procedure. For a monotone confidence procedure C it can be shown that $E_\theta L(C)$ has a maximum at $\theta = \frac{1}{2}$. Thus, to minimize the maximum risk we only need be concerned with the risk at $\theta = \frac{1}{2}$, and the monotone minimax procedure is easily found as the solution to a standard linear programming problem.

Since the only variables in R_α are the coincidental endpoints, we only need deal with these points in determining the monotone minimax procedure. These endpoints were determined for $n = 6, 30$, and are given in Table 2. It should be noted that Lemma 3.2 can be used to show that both the minimax and monotone minimax procedures are unique.

For purposes of comparison, a graph of the expected lengths of the monotone minimax procedure, the Blyth-Still, Sterne-Crow, and Clopper-Pearson, is given in Figure 1. One can see the improvement in risk, for θ in the middle, that is realized by using the monotone minimax procedure. The Blyth-Still procedure should be preferred if there is no prior information at all about θ , while a Sterne-Crow-type procedure is preferable if θ is thought to be near 0 or 1. (By choosing all coincidental lower endpoints to be their minimum value, one obtains a procedure in R_α with risk performance similar to that of the Sterne-Crow procedure.)

For $n = 14$ and $\alpha = .05$, the minimax procedure was also determined. Although the procedure is not monotone, its risk function is and, to three places, identical to that of the monotone minimax procedure. In general, it is probably true that the minimax procedure will have a risk function close to that of the monotone minimax procedure but, in general, comparisons with the minimax procedure were not done because of the computational complexity.

Acknowledgement. The author thanks the reviewers for their valuable comments.

References

- Blyth, C. R., and Hutchinson, D. W. (1960), "Tables of Neyman - Shortest Confidence Intervals for the Binomial Parameters," *Biometrika*, 47, 381-391.
- Blyth, C. R., and Still, H. A. (1983), "Binomial Confidence Intervals," *Journal of the American Statistical Association*, 78, 108-116.
- Clopper, C. J., and Pearson, E. S. (1934), "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, 26, 404-413.
- Crow, E. L. (1956), "Confidence Intervals for a Proportion," *Biometrika*, 43, 423-435.
- Karlin, S. (1957), "Polya Type Distributions, II," *Annals of Mathematical Statistics*, 28, 281-309.
- Knuth, D. E. (1973), *The Art of Computer Programming*, 2nd Edition, Reading, MA.: Addison-Wesley.

Lehmann, E. L. (1959), *Testing Statistical Hypotheses*, New York: John Wiley & Sons.

Natrella, M. G. (1963), *Experimental Statistics*, National Bureau of Standards Handbook 91.

Sterne, T. E. (1954), "Some Remarks on Confidence or Fiducial Limits" *Biometrika*, 41, 275-278.

Stevens, W. L. (1950), "Fiducial Limits of the Parameter of a Discontinuous Distribution," *Biometrika*, 37, 117-129.

Appendix: Proof of Theorem 4.1

Fix α , and let $C = \{[\lambda_x, u_x], x=0, \dots, n\}$ be a member of R_α . We must show that there does not exist a $1-\alpha$ confidence procedure $C' = \{[\lambda'_x, u'_x], x=0, \dots, n\}$ with the property that $E_\theta L(C') \leq E_\theta L(C)$ for all θ , with strict inequality for some θ . Clearly, we only need consider $C' \in R_\alpha$, for if there exists $C' \notin R_\alpha$ which dominates C , C' is inadmissible and is dominated by some $C'' \in R_\alpha$, which also dominates C .

If $C' \in R_\alpha$, then C and C' have the same noncoincidental endpoints. Thus, any difference in risk need only be examined at the intervals with coincidental endpoints. We will only consider the case where C' differs from C at one coincidental lower endpoint.

Suppose λ_k is a coincidental endpoint of C and differs from $\lambda'_k \in C'$. Further, let $\lambda'_k = u_m$. C and C' differ at no more than four intervals, corresponding to $x=k, n-k, m$, and $n-m$. Using the fact that $\lambda'_k = u_m$, and the invariance restriction $1-\lambda'_k = u_{n-k}$, we can express these four intervals as

$$[\lambda_k, u_k], [\lambda_{n-k}, 1-\lambda_k], [\lambda_m, \lambda'_k], [1-\lambda'_k, u_{n-m}]$$

where we note that $u_k, \lambda_{n-k}, \lambda'_m$, and u_{n-m} are noncoincidental endpoints. A similar calculation for C' gives the intervals

$$[\lambda'_k, u_k], [\lambda_{n-k}, 1-\lambda'_k], [\lambda_m, \lambda'_k], [1-\lambda'_k, u_{n-m}] .$$

Since $\lambda'_k = u_m$, we must have $k > m$ (because $\lambda'_k = u_m < u_k$ implies $k > m$). Also, if k and m are conjugate, that is $m = n-k$, then $\lambda'_k = u_{n-k} = 1-\lambda'_k$, which implies $\lambda'_k = \frac{1}{2}$.

Note that when subtracting corresponding interval lengths the non-coincidental endpoints cancel each other. We then obtain for the difference in expected lengths

$$\begin{aligned} \Delta(\theta) &= E_{\theta} L(C) - E_{\theta} L(C') \\ &= (\lambda'_k - \lambda_k) \{ [P(X=k|\theta) + P(X=n-k|\theta)] - [P(X=m|\theta) + P(X=n-m|\theta)] \} . \quad (A1) \end{aligned}$$

We now show that for θ near 0 $\Delta(\theta)$ has a different sign from $\Delta(\frac{1}{2})$, which will show that C' cannot dominate C .

For convenience, define the function $r_t(\theta)$ by

$$r_t(\theta) = \theta^t (1-\theta)^{n-t} + \theta^{n-t} (1-\theta)^t , \quad (A2)$$

so we can write

$$\Delta(\theta) = (\lambda'_k - \lambda_k) \left[\binom{n}{k} r_k(\theta) - \binom{n}{m} r_m(\theta) \right] . \quad (A3)$$

At $\theta = \frac{1}{2}$, $r_k(\frac{1}{2}) = r_m(\frac{1}{2}) = \frac{1}{2}^{n-1}$, so

$$\Delta(\frac{1}{2}) = \frac{1}{2^{n-1}} \left[\binom{n}{k} - \binom{n}{m} \right] (\lambda'_k - \lambda_k) . \quad (A4)$$

As $\theta=0$, $r_t(0) = 0$ unless $t=0$ or n , in which case $r_0(0) = r_n(0) = 1$. Hence if $0 < k, m < n$, $\Delta(0) = 0$, that is C and C' have the same expected length at 0. This is the more difficult case to deal with, so we first consider the case of $\Delta(0) \neq 0$.

If $\Delta(0) \neq 0$, it must be the case that either i) $k=n$, $m>0$, or ii) $k<n$, $m=0$. For these cases we have

$$\begin{aligned}
 \text{i)} \quad \Delta(\frac{1}{2}) &= \frac{1}{2^{n-1}} \left[1 - \binom{n}{m} \right] (\lambda'_n - \lambda_n) \\
 \Delta(0) &= \lambda'_n - \lambda_n \\
 \text{ii)} \quad \Delta(\frac{1}{2}) &= \frac{1}{2^{n-1}} \left[\binom{n}{k} - 1 \right] (\lambda'_k - \lambda_k) \\
 \Delta(0) &= -(\lambda'_k - \lambda_k) \quad .
 \end{aligned} \tag{A5}$$

In either case $\Delta(\theta)$ has opposite signs at $\theta=0$ and $\theta=\frac{1}{2}$, so C' cannot dominate C .

Now we consider the case $0 < m < k < n$, which implies that $\Delta(0) = 0$. For this case it will be shown that if $\Delta(\frac{1}{2}) > 0$ then $\Delta(\theta)$ decreases as θ increases from 0, while if $\Delta(\frac{1}{2}) < 0$ then $\Delta(\theta)$ increases as θ increases from 0. This, once again, will establish that C' does not dominate C .

First recall the following application of Taylor's theorem: If a function $f(y)$ satisfies $f'(y_0) = f''(y_0) = \dots = f^{(j-1)}(y_0) = 0$, $f^{(j)}(y_0) \neq 0$, then for some $\epsilon > 0$,

$$f(y) = f(y_0) + f^{(j-1)}(\xi) \frac{(y-y_0)^{j-1}}{(j-1)!} \quad y_0 < y < y_0 + \epsilon \tag{A6}$$

where ξ is some number between y_0 and y . Since $f^{(j)}(y_0) = \lim_{y \rightarrow y_0} f^{(j-1)}(y)/(y-y_0)$ it follows that $f(y)$ increases from $f(y_0)$ if $f^{(j)}(y_0) > 0$, while if $f^{(j)}(y_0) < 0$, $f(y)$ decreases from $f(y_0)$.

To apply this result, we must determine the first nonzero derivative of $\Delta(\theta)$ at $\theta=0$. The function $r_t(\theta)$, defined in (A2), can be written as

$$r_t(\theta) = \sum_{i=0}^{n-t} \binom{n-t}{i} (-1)^i \theta^{t+i} + \sum_{i=0}^t \binom{t}{i} (-1)^i \theta^{n-t+i} \quad , \tag{A7}$$

a polynomial of degree n with minimum exponent equal to $\min(t, n-t)$. It then follows from (A3) that the first nonzero derivative of $\Delta(\theta)$ at $\theta=0$ will be $(d^{j^*}/d\theta^{j^*})\Delta(\theta)$, where $j^* = \min\{k, n-k, m, n-m\}$. We must consider two cases:

Case 1: $k \leq n - k$

Since $k > m$, it then follows that $n-m > n-k \geq k > m$, so $j^* = m$ and, moreover, $\binom{n}{k} > \binom{n}{m}$. The coefficient of θ^m in $r_m(\theta)$ is 1, $(d^m/d\theta^m)\theta^m = m!$ and, from (A3),

$$\left. \frac{d^m}{d\theta^m} \Delta(\theta) \right|_{\theta=0} = -m! \binom{n}{m} (\lambda'_k - \lambda_k) . \quad (A8)$$

Since $\binom{n}{k} > \binom{n}{m}$, we see from (A4) that $\Delta(\frac{1}{2})$ and (A8) have opposite signs, showing that C' does not dominate C .

Case 2: $k > n - k$

This case must be split further: we must consider $k+m < n$ and $k+m > n$ separately. If $k+m < n$, it can again be deduced that $j^* = m$ and $\binom{n}{k} > \binom{n}{m}$, so the proof proceeds as in Case 1. If $k+m > n$ and then $j^* = n-k$ and $\binom{n}{k} < \binom{n}{m}$. We then have

$$\left. \frac{d^{n-k}}{d\theta^{n-k}} \Delta(\theta) \right|_{\theta=0} = (n-k)! \binom{n}{k} (\lambda'_k - \lambda_k) , \quad (A9)$$

and comparison with $\Delta(\frac{1}{2})$ shows that (A9) and $\Delta(\frac{1}{2})$ have opposite signs.

Hence the theorem is established.

Table 1. Values of the Lower Endpoints of the Complete Class R_α for $\alpha = .05$ (Left Column) and $\alpha = .01$ (Right Column). The Upper Endpoints are Obtained from the Identity $u_x = 1 - l_{n-x}$.

In All Cases $l_0 = 0$ and $u_n = 1$.

x	n=6		n=7		n=8		n=9		n=10	
1	.009	.002	.007	.001	.006	.001	.006	.001	.005	.001
2	.063	.027	.053	.023	.046	.020	.041	.017	.037	.016
3	.153	.085	.129	.071	.111	.061	.098	.053	.087	.047
4	.271	.173	.225	.142	.193	.121	.169	.105	.150	.093
5	.406±.012	.295	.341	.237	.289	.198	.251	.171	.222	.150
6	$1-l_5$.465	.446	.357	.356±.044	.293	.314±.031	.250	.281±.022	.218
7			.623	.517	.500	.410	.442	.343	.381	.297
8					$1-l_8$.550	.557	.428±.027	.444±.049	.379±.009
9							$1-l_8$	$1-l_8$	$1-l_8$.488
10									$1-l_6$	$1-l_8$
x	n=11		n=12		n=13		n=14		n=15	
1	.005	.001	.004	.001	.004	.001	.004	.001	.003	.001
2	.033	.014	.030	.013	.028	.012	.026	.011	.024	.010
3	.079	.043	.072	.039	.066	.036	.061	.033	.057	.031
4	.135	.084	.123	.076	.113	.069	.104	.064	.097	.059
5	.200	.134	.181	.122	.166	.111	.153	.102	.142	.094
6	.255±.016	.194	.233±.012	.174	.224	.159	.206	.146	.191	.135
7	.333	.263	.294	.235	.261	.213	.235±.028	.195	.217±.025	.179
8	.400±.036	.339	.365±.026	.302	.336±.019	.273	.311±.014	.249	.294	.229
9	.500	.408	.455±.017	.348±.029	.413	.319±.020	.371	.294±.013	.333	.273±.009
10	$1-l_8$.500	$1-l_8$.451±.011	.480	.406	.423±.037	.363	.393±.029	.328
11	$1-l_8$.641	$1-l_8$	$1-l_{10}$.566	.477	.500	.416±.027	.465±.025	.385±.017
12			$1-l_8$	$1-l_9$	$1-l_8$.571	$1-l_{10}$.500	$1-l_{11}$.462±.009
13					.775	$1-l_9$	$1-l_8$	$1-l_{11}$	$1-l_{10}$	$1-l_{12}$
14							$1-l_7$	$1-l_2$.698	$1-l_{11}$
15									$1-l_7$	$1-l_9$

Table 1 (Cont.)

x	n=16		n=17		n=18		n=19		n=20	
1	.003	.001	.003	.001	.003	.001	.003	.001	.003	.001
2	.023	.010	.021	.009	.020	.008	.019	.008	.018	.008
3	.053	.029	.050	.027	.047	.025	.044	.024	.042	.023
4	.090	.055	.085	.052	.080	.049	.075	.046	.071	.044
5	.132	.088	.124	.082	.116	.077	.110	.073	.104	.069
6	.178	.125	.166	.117	.156	.109	.148	.103	.139	.097
7	.208±.024	.166	.189±.022	.155	.178±.021	.145	.168±.019	.137	.158±.018	.129
8	.272	.210	.253	.197	.236	.185	.222	.174	.209	.163
9	.303±.030	.258	.282±.028	.242	.264±.027	.227	.250±.024	.213	.237±.021	.200
10	.367±.024	.296	.345±.019	.266±.023	.325±.015	.249±.021	.312	.235±.019	.293	.222±.016
11	.434±.017	.359±.010	.406	.339	.375	.314	.345	.292	.320±.026	.274
12	.502	.421	.456	.381	.411±.035	.343±.026	.389±.029	.323±.021	.369±.025	.306±.017
13	1-l ₁₁	.474	.511	.433±.023	.470±.029	.407±.015	.445±.025	.384±.010	.422±.020	.363
14	1-l ₁₀	.549	.583	.500	1-l ₁₃	.469±.011	.500	.436	.474±.018	.399
15	1-l ₉	1-l ₁₁	1-l ₁₀	1-l ₁₃	1-l ₁₂	1-l ₁₄	1-l ₁₃	.485	1-l ₁₄	.444±.023
16	1-l ₇	.735	1-l ₉	.655	1-l ₁₀	1-l ₁₃	1-l ₁₂	.544	1-l ₁₃	.500
17			1-l ₇	1-l ₁₀	1-l ₉	1-l ₁₂	.684	1-l ₁₃	1-l ₁₂	1-l ₁₅
18					1-l ₇	1-l ₁₀	1-l ₉	1-l ₁₂	1-l ₁₁	.624
19							1-l ₇	1-l ₁₀	1-l ₉	1-l ₁₂
20									1-l ₇	1-l ₁₀

x	n=21		n=22		n=23		n=24		n=25	
1	.002	.000	.002	.000	.002	.000	.002	.000	.002	.000
2	.017	.007	.016	.007	.016	.007	.015	.006	.014	.006
3	.040	.022	.038	.020	.037	.020	.035	.019	.034	.018
4	.068	.041	.065	.039	.062	.038	.059	.036	.057	.034
5	.099	.065	.094	.062	.090	.059	.086	.057	.082	.054
6	.132	.092	.126	.088	.120	.084	.115	.080	.110	.077
7	.151±.018	.122	.144±.016	.116	.138±.015	.111	.132±.014	.106	.126±.013	.101
8	.197	.155	.187	.147	.178	.140	.169	.133	.161	.127
9	.226±.019	.189	.215±.017	.179	.206±.016	.171	.197±.014	.162	.189±.013	.155
10	.276	.211±.014	.260	.201±.012	.247	.192±.011	.234	.184±.009	.222	.177±.008
11	.303±.025	.258	.286±.025	.242	.273±.023	.229	.261±.021	.217	.250±.019	.205
12	.351±.021	.291±.014	.334±.018	.277±.011	.319±.016	.264±.009	.306±.014	.258	.296	.245
13	.401±.016	.340	.382	.317	.360	.298	.339	.280	.317	.259±.022
14	.449	.368±.024	.417	.350±.019	.390±.029	.334±.016	.369±.027	.319±.013	.354±.024	.306±.010
15	.494	.421±.017	.452±.032	.401±.012	.432±.028	.384	.413±.024	.362	.396±.021	.342
16	.545	.474±.014	.500	.450	.475±.024	.419	.457±.022	.387±.023	.438±.018	.369±.020
17	1-l ₁₃	1-l ₁₈	1-l ₁₅	.495	1-l ₁₈	.452±.025	.500	.432±.020	.479±.017	.413±.015
18	1-l ₁₂	1-l ₁₅	.611	.546	1-l ₁₅	.500	1-l ₁₈	.477±.017	1-l ₁₇	.457±.012
19	1-l ₁₁	1-l ₁₄	1-l ₁₂	1-l ₁₅	1-l ₁₄	1-l ₁₇	1-l ₁₅	1-l ₁₈	1-l ₁₈	.504
20	1-l ₁₀	1-l ₁₃	1-l ₁₁	1-l ₁₄	1-l ₁₂	.614	1-l ₁₃	1-l ₁₇	1-l ₁₅	1-l ₁₈
21	1-l ₉	1-l ₁₀	1-l ₁₀	1-l ₁₂	1-l ₁₁	1-l ₁₄	1-l ₁₃	1-l ₁₆	1-l ₁₄	1-l ₁₇
22			1-l ₇	1-l ₁₀	1-l ₉	1-l ₁₂	1-l ₁₁	1-l ₁₄	.696	1-l ₁₈
23					1-l ₇	1-l ₁₀	1-l ₉	.741	1-l ₁₁	1-l ₁₄
24							1-l ₇	1-l ₁₀	1-l ₉	1-l ₁₃
25									1-l ₇	1-l ₁₀

Table 1 (Cont.)

x	n=26		n=27		n=28		n=29		n=30	
1	.002	.000	.002	.000	.002	.000	.002	.00	.002	.000
2	.014	.006	.013	.006	.013	.005	.012	.005	.012	.005
3	.032	.017	.031	.017	.030	.016	.029	.015	.028	.015
4	.054	.033	.052	.032	.050	.031	.049	.030	.047	.028
5	.079	.052	.076	.050	.073	.048	.071	.046	.068	.045
6	.106	.073	.101	.071	.098	.068	.094	.065	.091	.063
7	.121±.012	.097	.117±.012	.093	.113±.011	.089	.109±.010	.086	.105±.010	.083
8	.154	.122	.148	.117	.142	.112	.136	.108	.131	.104
9	.182±.012	.149	.175±.011	.142	.169±.010	.137	.163±.009	.132	.157±.009	.127
10	.212	.170±.007	.202	.163±.006	.192	.163	.184	.156	.175	.151
11	.241±.018	.194	.232±.016	.184	.223±.015	.175	.215±.014	.170±.013	.208±.013	.163±.011
12	.282	.234	.269	.224	.257	.214	.246	.205	.236	.198
13	.304±.021	.251±.016	.291±.021	.240±.015	.279±.021	.231±.016	.268±.021	.222±.014	.259±.020	.215±.013
14	.340±.022	.294±.008	.327±.020	.284	.315±.018	.272	.304±.016	.260	.294±.014	.250
15	.381±.018	.322	.366±.015	.303±.018	.355	.290±.017	.339	.278±.017	.324	.268±.016
16	.421±.015	.354±.017	.402	.341±.014	.381	.328±.012	.362±.022	.316±.009	.344±.025	.308
17	.458	.397±.011	.430	.383	.407±.025	.364	.393±.025	.347	.379±.022	.329
18	.494	.437	.461±.029	.413	.444±.026	.386±.021	.429±.023	.370±.019	.414±.020	.357±.017
19	.535	.474	.500	.440±.023	.478±.021	.423±.019	.464±.021	.408±.015	.448±.018	.393±.012
20	1-l _{1e}	.514	1-l ₁₈	.479±.020	1-l ₁₉	.462±.017	.500	.445±.013	.482±.017	.429±.009
21	1-l ₁₅	.558	.585	1-l ₂₀	1-l ₁₈	.500	1-l _{1e}	.482±.012	1-l ₂₀	.462
22	1-l ₁₄	1-l ₁₇	1-l ₁₅	1-l ₁₉	1-l ₁₇	1-l ₂₀	1-l _{1e}	1-l ₂₁	1-l _{1e}	.494
23	1-l ₁₃	1-l ₁₆	1-l ₁₄	.616	.643	1-l ₁₉	1-l ₁₇	1-l ₂₀	1-l _{1e}	.531
24	1-l ₁₁	1-l ₁₄	1-l ₁₃	1-l ₁₆	1-l ₁₄	1-l ₁₈	1-l _{1e}	1-l ₁₉	1-l ₁₇	1-l ₂₀
25	1-l ₉	1-l ₁₃	1-l ₁₁	1-l ₁₅	1-l ₁₃	1-l ₁₈	1-l ₁₄	1-l ₁₈	1-l _{1e}	1-l ₁₉
26	1-l ₇	1-l ₁₀	1-l ₉	1-l ₁₃	1-l ₁₁	1-l ₁₅	1-l ₁₃	1-l _{1e}	1-l ₁₄	1-l ₁₈
27			1-l ₇	1-l ₁₀	1-l ₉	1-l ₁₃	1-l ₁₁	1-l ₁₅	1-l ₁₃	.690
28					1-l ₇	.837	1-l ₉	1-l ₁₃	1-l ₁₁	1-l ₁₅
29							1-l ₇	1-l ₁₁	1-l ₉	1-l ₁₃
30								1-l ₁₁	1-l ₇	1-l ₁₁

Table 2. Values of the Coincidental Lower Endpoints for the Monotone Minimax Procedure
for $\alpha=.05$ (left column) and $\alpha=.01$ (right column)

		n												
x	6	7	8	9	10	11	12	13	14	15	16	17		
5	.418	--	--	--	--	--	--	--	--	--	--	--	--	
6	--	--	.400	.345	.303	.268	.237	--	--	--	--	--	--	
7	--	--	--	--	--	--	--	--	.262	.230	.227	.212	--	
8	--	--	--	.455	.485	.388	.436	.378	.355	.318	--	--	--	
9	--	--	--	--	--	--	.472	.377	.339	.307	.282	.334	.310	
10	--	--	--	--	--	--	--	.462	--	.457	.422	.391	.361	.290
11	--	--	--	--	--	--	--	--	--	.443	.480	.402	.451	.369
12	--	--	--	--	--	--	--	--	--	--	.471	--	--	--
13	--	--	--	--	--	--	--	--	--	--	--	--	--	.456
14	--	--	--	--	--	--	--	--	--	--	--	--	--	--

		n																			
x	18	19	20	21	22	23	24	25	26	27	28	29	30								
7	.199	.188	.177	.169	.160	.148	.146	.139	.133	.129	.124	.112	.115								
8	--	--	--	--	--	--	--	--	--	--	--	--	--								
9	.286	.255	.258	.244	.222	.222	.211	.200	.190	.182	.174	.163	.166								
10	.337	.270	.253	.236	.225	.210	.203	.193	.185	.177	.169	--	--								
11	--	--	.344	.321	.292	.288	.282	.269	.259	.230	.238	.228	.184	.218	.175						
12	.441	.369	.418	.344	.394	.323	.369	.305	.344	.286	.319	.272	.310	--	--						
13	.482	.422	.462	.394	.442	.417	--	--	--	.296	.300	.260	.293	.255	.283	.241	.272	.235	.279	.228	
14	--	.480	--	.482	.392	.369	.412	.350	.392	.331	.369	.316	.362	.286	.342	.319	.320	.308	--	--	
15	--	--	--	.467	.438	.466	.413	.456	.421	.417	.389	.366	.315	.298	.298	.292	.292	.292	.279	--	
16	--	--	--	--	.488	--	.486	--	.463	.410	.450	.389	.436	.371	.355	.325	.385	.316	.358	--	
17	--	--	--	--	--	--	.475	--	.451	.488	.428	--	.408	--	--	.420	.411	--	.387	--	
18	--	--	--	--	--	--	--	--	.488	--	.469	--	--	--	--	.466	.407	.429	.389	.430	.374
19	--	--	--	--	--	--	--	--	--	--	--	--	--	.463	.484	.442	.473	.415	.448	.405	--
20	--	--	--	--	--	--	--	--	--	--	--	--	--	.489	--	.477	--	.451	.491	.438	--
21	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	.489	--	--	--

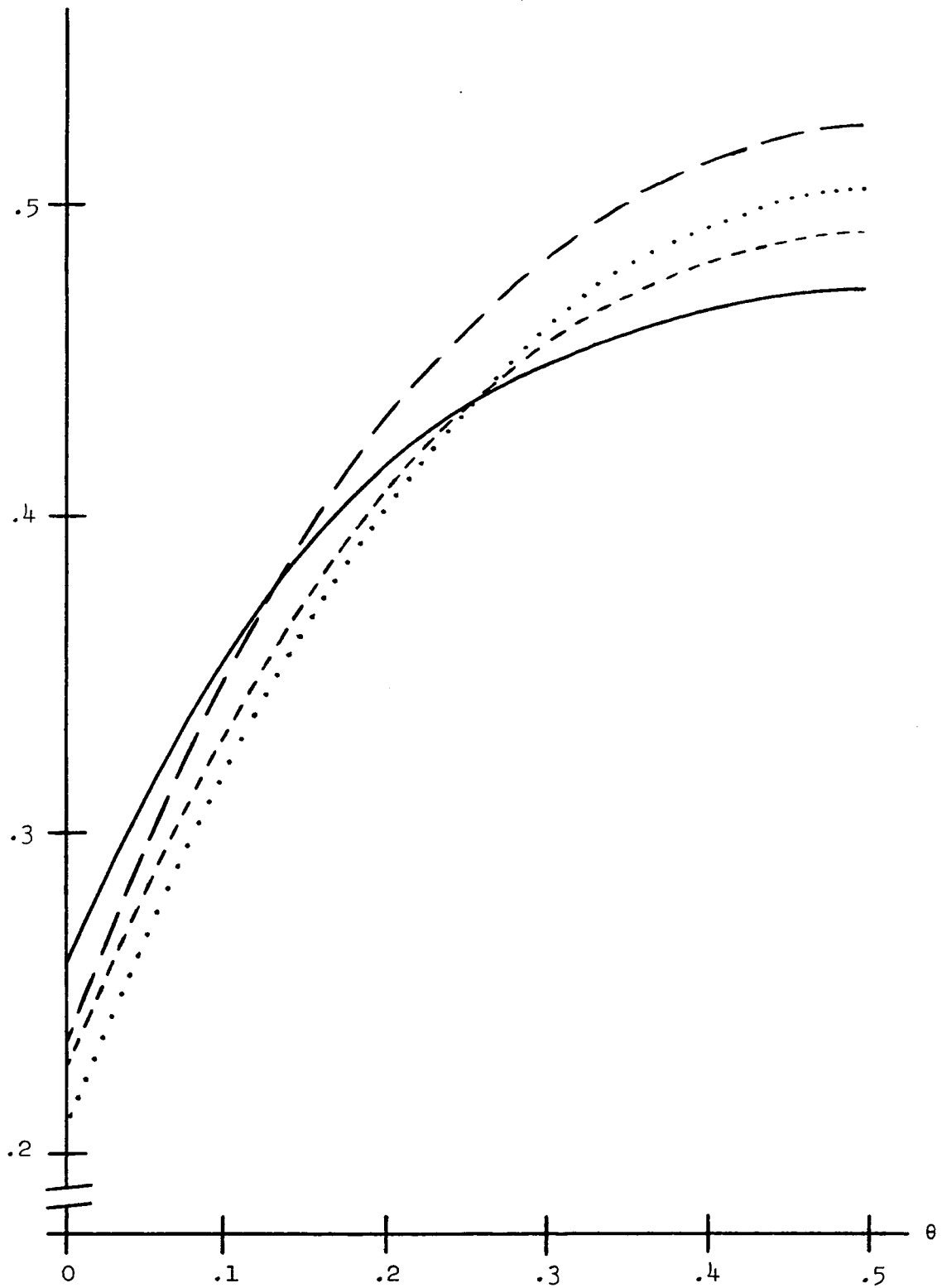


Figure 1. Expected Lengths, for $n=14$, $1-\alpha=.95$, of the Clopper-Pearson (long dashes), Sterne-Crow (dotted), Blyth-Still (short dashes), and Montone Minimax (solid line) Procedures.