BUEHLER CONFIDENCE BOUNDS

Margarita F. Guerrero, Ph.D.[1]
and
Herbert T. David, Ph.D.[2]

1. Introduction

A statistical problem that arises in various applications is the problem of obtaining lower and/or upper confidence bounds for real-valued functions of several unknown parameters from experimental data. This problem is encountered in engineering practice, for example, in the estimation of the reliability function of a system from the results of trials on the components of the system. The discussions of lower and upper confidence bounds are completely parallel, and it is therefore sufficient to develop this discussion in terms of upper confidence bounds.

Proponents of specific procedures for constructing confidence intervals or regions evaluate desirability of their results in terms of various optimality criteria which generally include minimizing "volume" or "expected volume" and minimizing either the probabilities of covering false values or the probabilities of failing to include the true value of the unknown parameter or parametric function.

Upper bound confidence procedures that yield bounds of small magnitude in a sense yield intervals of small volume. Buehler (1957) provides uniformly smallest upper bounds for the product of two

---

[1]Assistant Professor, Institute of Mathematical Sciences and Physics, University of the Philippines at Los Baños, College, Laguna.

[2]Professor, Department of Statistics, Iowa State University, Ames, Iowa.

binomial parameters and indicates the generalization of his procedure to an arbitrary discrete distribution. This optimality property is the basis of a number of approximation methods for bounding reliability functions of systems as reported in a survey by Harris and Soms (1981). In this paper, we lay down a theoretical framework for Buehler's optimum confidence bounds procedure.

In Section 2, the notion of ordered sample spaces is introduced and defined. The monotone confidence sets that are an implicit part of the Buehler methodology are characterized in Section 3 and Section 4 describes and discusses Buehler's confidence procedure and its extension to continuous sample spaces. A procedure for obtaining simultaneously Buehler-optimal bounds for a collection of parametric functions is also indicated here. Finally, Section 5 deals with the issue of existence of Buehler bounds and discusses possible simplifications in the construction of these bounds.

2. *Definitions and Notations*

The notion of "ordering" the sample space $\mathfrak{X}$ of a random vector $X = (X_1, X_2, \ldots, X_n)$ is essential to the type of confidence bounds procedure presented here. In this section, this concept and the terminology associated with it are defined.

Suppose each random variable X_k is discrete-valued and assumes a finite number of values. Then the sample space consists of a finite number of points, say N, and there are $N!$ ways of labeling the sample points $x = (x_1, x_2, \ldots, x_n)$ using the index set $I = (1, 2, \ldots, N)$. Each one of these $N!$ distinct labelings will be referred to as an ordering of the finite sample space $\mathfrak{X}$. Given a particular ordering, one can then refer to the first point, the second point, and in general, the *kth* point of the ordered sample space. The notation $x^{(k)}$ will be used to refer to the *kth* point in the ordering. Clearly, the number of possible orderings of a finite space increases with N; indeed, the notion of labeling individual sample points may be extended to sample spaces that are countably infinite, in which case, an uncountably infinite number of orderings is possible.

Another sense in which a sample space is said to be ordered is through a partitioning of the sample space into ordered sets. This can be achieved via a real-valued function, say S, that associates partition elements A with values of S in such a way that x_1 and x_2 belong to the same partition element A if $S(x_1) = S(x_2)$. In addition, a precedence relation among partition elements is induced by S, in that, if x_1 belongs to partition set A_i and x_2 belongs to partition set A_j, and if $S(x_1) < S(x_2)$, then A_i is said to precede A_j in the ordering of the partition sets A_i.

Harris and Soms (1980) refer to the functions S as ordering functions. Note that an ordering obtained via the labeling of individual points represents a maximal partition of $\mathfrak{X}$.

Ordered partitions are particularly appropriate in the countably infinite and the uncountable cases. In the latter case, the ordering of individual sample points that is possible with discrete sample spaces is not possible, and an ordering of the sample space will refer to an ordered partition entailing specification of an ordering function, or "statistic".

In either case, once an ordering of $\mathfrak{X}$ is specified, the information contained in a sample is conveyed by a scalar random variable. For discrete sample spaces, this random variable will be the random index $I(X)$ with value-space $(1, 2, \ldots, N)$, while for continuous sample spaces, it will be the real-valued statistics S.

Preliminary to the construction of confidence bounds that are optimal in the Buehler sense, we will present a procedure for obtaining optimal confidence regions for parametric vectors.

3. Monotone Confidence Regions

3.1 Discrete Distributions

Let $X = (X_1, X_2, \ldots, X_n)$ be a vector of random variables assuming finitely many values. Suppose the distribution function X belongs to a family of distributions parameterized by $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$ belonging to the parameter space Θ.

For a particular labeling of the sample points, the cumulative distribution function of the ordered points is given by:

$$F(x^{(k)}; \theta) = F(k; \theta)$$

$$= PROB\,(I(x) \leqslant k \mid \theta\,)$$

$$= \sum_{j \leqslant k} f(x^{(j)}; \theta\,)$$

where I is the random index generated under θ by the given labeling, with $I(x) = k$ iff x is the kth point in the ordered sample space and

$$f(\mathrm{x}; \theta\,) \;=\; PROB\,(X = x \mid \theta)$$

Consider now the following subsets of the parameter space:

$$\Omega\;(k) = (\theta \in \theta\;:\;F(k; \theta) > \alpha), \quad k = 1, 2, \ldots, N, \qquad (3.1.1)$$

where $0 < \alpha < 1$ is fixed. Since, for fixed θ $\mathrm{F}(k; \theta\,)$ is monotone nondecreasing in k with $F(N; \theta) = 1$, these subsets form a sequence of nested regions that are nondecreasing to Θ; i.e.,

$$\Omega(1) \subset \Omega\,(2) \subset \ldots \subset \Omega\;(k) \subset \ldots \subset \Omega\;(N) \;=\; \Theta.$$

In addition, if

$$\omega(1) = \Omega(1)$$

and

$$\omega(k) = \Omega(k) - \Omega(k-1), \quad k = 2, 3, \ldots, N,$$

then every θ belongs to exactly one of the disjoint sets $\omega(k)$. This follows from the observation that if $\theta\,'$ does not belong to any $\omega(k)$, then $F(k; \theta\,) \leqslant \alpha$ for all k. In particular, $F(N; \theta') < \alpha < 1$, so that $F(\cdot\,; \theta\,')$ cannot be a distribution function.

The above properties are useful in proving the next two theorems. Note that, given a labeling of $\mathfrak{X}$, $\Omega\,(k)$ is used to denote $\Omega\,(I(x)\,)$, with $I(x\,)$ the label assigned to x (i.e., $\Omega\,(x^{(k)}\,) = \Omega\,(k)\,)$.

Theorem 3.1.1: For any labeling of the sample space, the sets $(\Omega(x): x \in \mathfrak{X})$ constitute a family of $(1\text{-}\alpha)$ 100% confidence regions for θ.

Proof: Let the parameter point θ' belong to $\omega\ (n+1)$. The θ' is in $\Omega(n+1)$ but is not in $\Omega(n)$. The monotonicity of the regions $\Omega(k)$ further implies that θ' is in $\Omega(k)$ iff $k \geqslant n+1$. Hence,

$$\begin{aligned} PROB\,(\theta' \in \Omega(X) \mid \theta') &= PROB\,(\theta' \in \Omega(I(X)) \mid \theta') \\ &= PROB\,(I(X) \geqslant n+1 \mid \theta') \\ &= 1\text{-}F\,(n;\theta') \\ &\geqslant 1\text{-}\alpha \end{aligned}$$

where the inequality is implied by the assumption that θ does not belong to $\Omega(n)$ so that $F(n;\theta') \leqslant \alpha$. Q.E.D.

Theorem 3.1.2. Consider a fixed labeling of $\mathfrak{X}$ and let $(\Omega(x): x \in \mathfrak{X})$ be the monotone confidence regions corresponding to this labeling. Suppose $(D(x): x \in \mathfrak{X})$ is any other family of $(1\text{-}\alpha)$ 100% confidence regions for θ. Let $D(k)$ denote the D-region corresponding to the *kth* point in the given labeling. If $D(k\text{-}1) \subset D(k)$ for $k=2, 3, \ldots, N$, with $D(N) = \Theta$, then, $\Omega(x) \subset D(x)$ for all x.

Proof: Suppose for some k and hence some x, we can find a θ' such that θ' is in $\Omega(k)$, but not in $D(k)$. Since θ' is not in $D(k)$ and the regions $D(i)$ are monotone nondecreasing in i, $\theta' \in D(i)$ iff $i \geqslant k+1$. Hence,

$$\begin{aligned} PROB\,(\theta' \in D(X) \mid \theta') &= PROB\,(\theta' \in D\,(I(X)) \mid \theta') \\ &= PROB\,(I(\underset{\sim}{X}) > k+1 \mid \theta'\theta') \\ &= 1\text{-}F(k;\theta') \\ &< 1\text{-}\alpha \end{aligned}$$

where the strict inequality follows from the assumption that θ' is in $\Omega\ (k)$, which implies that $F(k;\theta') > \alpha$. This establishes that the sets $(D(x))$ cannot constitute a system of $(1\text{-}\alpha)$ 100% confidence regions for θ. Therefore it must be true that there are no vectors θ that are in $\Omega(k)$ but not in $D(k)$ and, hence, $\Omega(x) \subset D(x)$ for all x. Q.E.D.

These two theorems establish that, for a given ordering, the confidence regions $(\Omega(x))$ are uniformly smallest among all similarly ordered nested confidence regions that are nondecreasing to Θ.

The nature of the monotone confidence regions is determined by the nature of the cumulative distribution function F of the ordered sample points for fixed k; i.e., F as a function of θ. This is illustrated in the following example.

EXAMPLE 1: Let $X \sim$ binomial (2, p). Then $\Theta = [0, 1]$ and $\mathfrak{X} = (0, 1, 2)$. In this simple case, there are six possible ways of labeling the sample points. Table 2.2.1 lists these orderings and their corresponding 95% monotone confidence regions. Order (I) is the "natural" labeling of the sample space, for which the distribution function is simply the binomial distribution, and

$$\Omega(k) = (p\colon\ PROB\ (X \leqslant k \mid p) > \alpha) \qquad \text{for } k = 0, 1, 2.^{***}$$

Table 1. Optimal 95% monotone confidence regions for p

Order	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$\Omega(1)$	$\Omega(2)$
I	0	1	2	[0.0, .78]	[0.0, .975]
II	0	2	1	[0.0, .78]	[0.0, 1.0]
III	1	0	2	[.026, .974]	[0.0, .975]
IV	1	2	0	[.026, .974]	[.025, 1.0]
V	2	0	1	[.225, 1.0]	[.025, 1.0]
VI	2	1	0	[.225, 1.0]	[.025, 1.0]

We now proceed to extend the type of confidence region procedure developed here to include parameters of continuous distributions.

3.2. Continuous distributions

Let us now consider the case where X is a vector of "continuous" random variables (i.e., random variables with distributions absolutely continuous with respect to Lebesgue measure), parameterized by $\theta \in \Theta$. Suppose an ordered partition of the sample space is obtained via the ordering function (statistic) $S(X)$, with distribution function given by

$$F(s;\theta) = PROB\,(S(X) \leqslant s \mid \theta)$$

where s belongs to the range space, $S(\mathfrak{X})$, of the function S.

A system of monotone confidence regions for θ based on S may be obtained by a construction similar to that corresponding to Equation 3.1.1. as follows: for a sample point x with $S(x) = s$, a subset of the parameter space is determined by:

$$\Omega(S(x)) = \Omega(s) = (\theta \in \Theta : F(s;\theta) > \alpha).$$

Clearly, these sets are monotone increasing; i.e., $\Omega(s) \subset \Omega(s')$ whenever $s < s'$. To establish that these regions are indeed $(1\text{-}\alpha)$ 100% confidence regions for θ , we not that:

$$\begin{aligned} PROB\,(\theta \in \Omega(S(X)) \mid \theta) &= PROB\,(F(S(X);\theta) > \alpha \mid \theta) \\ &= PROB\,(S(X) > s(\alpha,\theta) \mid \theta) \\ &\geqslant 1\text{-}\alpha \end{aligned}$$

where $s(\alpha, \theta)$ is any $\alpha\%$ point of $S(X)$ under θ, or, more precisely, is the supremum of all s such that $F(s;\theta) \leqslant \alpha$.

We now present the continuous analogue of Theorem 3.1.2., which can be proved in a similar fashion.

Theorem 3.2.1: Suppose $(D(s))$ is any other family of $(1\text{-}\alpha)$ 100% confidence regions for θ based on the ordering funtion S. If $D(s) \subset D(s')$ whenever $s < s'$, then $\Omega(s) \subset D(s)$ for all $s \in S(\mathfrak{X})$.

We end this subsection with an example.

EXAMPLE 2: Let X_k $(k=1, 2, \ldots, n)$ be i.i.d. $N(\mu, \sigma^2)$. Consider two ordering functions: $S_1(x) = \bar{x}$ and $S_2(x) = t = \sqrt{n}\,\bar{x}/\hat{\sigma}$ where $\bar{x} = (1/n)\Sigma x_i$ and $\hat{\sigma}^2 = (1/(n\text{-}1))\Sigma\,(x_i\text{-}\bar{x})^2$. Then, (with z designating the normal distribution and $t_{v,\,\delta}$ designating the non-central t-distribution with v degrees of freedom and noncentrality parameter δ);

$$\Omega(\bar{x}) = ((\mu, \sigma^2):\ PROB\,(\bar{X} \leqslant \bar{x} \mid (\mu, \sigma^2)) > \alpha)$$
$$= ((\mu, \sigma^2):\ \mu \leqslant \bar{x} + z_{(1-\alpha)}\sigma/\sqrt{n})$$

and

$$\Omega(t) = ((\mu, \sigma^2):\ PROB\,(T \leqslant t \mid (\mu, \sigma^2)) > \alpha)$$
$$= ((\mu, \sigma^2):\ t_{n-1},\ \sqrt{n}\mu/\sigma) \geqslant t).^{***}$$

4. Optimal Confidence Bounds

4.1. Buehler bounds for discrete distributions

In this section, we assume that the vector of observations X is defined over a finite sample space; hence, an ordering of the sample space refers to a labeling of the sample points.

For a fixed labeling of the sample space, let us define a function of the sample points as follows:

$$b(x^{(k)}) = b(k) = \sup\,(H(\theta):\ F(k;\theta) > \alpha) \qquad (4.1.1)$$
$$= \sup_{\theta \epsilon \Omega(k)} H(\theta)$$

where $\Omega(k)$ is the monotone region defined in Equation 3.1.1 above. In the following theorems, certain properties of the function $b(\cdot)$ given by Buehler (1957) are established. Note that the function b assigns a value to the sample point x via its position in the ordering. In what follows, we will denote $b(x)$ by $b(k)$ iff $x = x^{(k)}$.

Theorem 4.1.1: $b(k) \leqslant b(k+1)$ for $k=1, 2. . . ,N$-1.

Proof: Consider the following sets of values of the function H:

$$A(k) = (H(\theta):\ \theta \epsilon \Omega\,(k)),\ k=1, 2 . . ., N.$$

Now, note that $\Omega(k) \subset \Omega(k+1)$ implies that $A(k) \subset A(k+1)$ and that $b(k)$ is the supremum of $A(k)$. The conclusion follows from the definition of supremum. Q.E.D.

The next theorem establishes that the function b generates a system of $(1\text{-}\alpha)$ 100% upper confidence bounds for $H(\theta)$.

Theorem 4.1.2: For any labeling of the sample points, $PROB\ (H(\theta) \leqslant b(X) \mid \theta) > 1\text{-}\alpha$ for all θ.

Proof: Let θ' belong to $\omega\ (n+1)$, where ω is as defined in subsection 3.1. For such θ',

$$\begin{aligned} PROB\ (H(\theta') \leqslant b(X) \mid \theta') &= PROB\ (H(\theta') \leqslant b(I(X)) \mid \theta') \\ &\geqslant PROB\ (\sup_{\theta\ \epsilon \Omega\ (n+1)} H(\theta') \leqslant b(I(X)) \mid \theta') \\ &= PROB\ (b(n+1) \leqslant b(I(X)) \mid \theta') \end{aligned}$$

By the monotone nondecreasing property of b established in Theorem 4.1.1. we may write:

$$\begin{aligned} PROB\ (b(n+1) \leqslant b(I(X)) \mid \theta') &= PROB\ (I(X) > n+1 \mid \theta') \\ &= 1\text{-}F(n;\theta') \\ &\geqslant 1\text{-}\alpha. \end{aligned}$$

The inequality follows from $\theta' \notin \Omega(n)$, since $\theta' \notin \Omega(n)$ implies that $F(n;\theta') \leqslant \alpha$. Q.E.D.

We now state and prove an optimal property of the system of upper confidence bounds generated by the function b.

Theorem 4.1.3: Consider a fixed labeling of the sample points. Suppose $d(X)$ generates any other system of $(1\text{-}\alpha)$ 100% upper confidence bounds for $H(\theta)$. Denote $d(x^{(k)})$ by $d(k)$. If $d(k) \leqslant d(k+1)$ for k=1, 2 . . ., N-1, then $b(x) \leqslant d(x)$ for all $x\ \epsilon\ \mathfrak{X}$.

Proof: Suppose that, for some k and hence some x, $d(k) < b(k)$. Then, since $b(k)$ is the supremum of H on $\Omega(k)$, it follows, that for some θ' in $\Omega(k)$, $d(k) < H(\theta')$. Hence, by the monotone increasing property of the function d, $d(i) \geqslant H(\theta')$ iff $i > k+1$, or:

$$\begin{aligned} PROB\ (H(\theta') \leqslant d(X) \mid \theta') &= PROB\ (H(\theta') \geqslant d(I(X)) \mid \theta') \\ &\leqslant PROB\ (I(X) > k+1 \mid \theta') \\ &= 1\text{-}F\ (k;\theta') \\ &< 1\text{-}\alpha \end{aligned}$$

with the strong inequality implied by the assumption that θ' belongs to $\Omega(k)$. Therefore, $d(x)$ cannot be a (1-α) 100% upper confidence procedure for $H(\theta)$, which is a contradiction. So it must be true that $d(x) \geqslant H(\theta)$ for all $\theta \in \Omega(x)$; hence, $b(x) \leqslant d(x)$ for all x. Q.E.D.

The bounds provided by the function $b(\cdot)$ will referred to as *upper Buehler bounds*. Theorem 4.1.3 establishes that among all similarly-ordered upper confidence bounds for $H(\theta)$, the upper Buehler bounds constitute a family of uniformly smallest ones.

Clearly, the same methodology extends to obtaining a system of uniformly largest lower confidence bounds for $H(\theta)$ via the function:

$$a(x^{(k)}) = a(k) = \inf(\, H(\theta) \colon F(k;\theta) > \alpha\,) = \inf_{\theta \in \Omega(k)} H(\theta).$$

These bounds will be referred to as *lower Buehler bounds*. The following theorem provides a characterization of these lower bounds that is analogous to that established for the upper bounds in Theorems 4.1.1–4.1.3.

Theorem 4.1.4: For only labeling of the sample space, the function $a(\cdot)$ satisfies the following properties:

(i) $a(k) \geqslant a(k+1)$, $k=1, 2, \ldots, N$.

(ii) $PROB(a(X) \leqslant H(\theta) \mid \theta) \geqslant 1\text{-}\alpha$ for all $\theta \in \Theta$.

(iii) If $d(X)$ is any other system of (1-α) 100% lower confidence bounds for $H(\theta)$ for which $d(k) \geqslant d(k+1)$ (where $d(x^{(k)}) = d(k)$), then $d(x) \leqslant a(x)$ for all x.

The proof of this theorem may be argued along the same lines as the proofs of Theorems 4.1.1 through 4.1.3 and will not be provided here. We provide the following simple example which illustrates the basic concepts presented so for in this section.

EXAMPLE 3: Let $X \sim$ binomial (2, p). Then $\Theta = [0, 1]$ and $\mathfrak{X} = (0, 1, 2)$. Table 1 exhibits the 95% monotone confidence regions for the six possible orderings of $\mathfrak{X}$. Consider the two parametric functions: $H_1(p) = p$ and $H_2(p) = (1\text{-}p)$. In Table 2 the 95% upper Buehler bounds for these parametric functions under each of the six possible labelings of $\mathfrak{X}$ are presented. Note that some orderings may be deemed to provide more reasonable upper bounds than others, which indicates that in general, some consideration must be given to the way one labels the sample points.

Table 2. 95% upper Buehler bounds for *p* and 1-p

Order	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$b_1(1)$	$b_1(2)$	$b_2(1)$	$b_2(2)$
I	0	1	2	0.780	0.975	1.000	1.000
II	0	2	1	0.780	1.000	1.000	1.000
III	1	0	2	0.974	0.975	0.974	1.000
IV	1	2	0	0.974	1.000	0.974	0.975
V	2	0	1	1.000	1.000	.780	.975
VI	2	1	0	1.000	1.000	.975	1.000

4.2 Buehler bounds of continuous distributions

Let X be a vector of continuous random variables and assume that the continuous sample space is ordered according to the ordering function or statistic $S(X)$. Then, the continuous analogue of Equation 4.1.1 is given by:

$$b(S(x)) = b(s) = \sup(H(\theta): F(s;\theta) > \alpha) = \sup_{\theta \epsilon \Omega(s)} H(\theta)$$

where $S(x) = s$, $F(s;\theta)$ is the distribution function of S, and $\Omega(s)$ is the monotone $(1\text{-}\alpha)$ 100% confidence region for θ based on S. Cleary, the function b provides the upper Buehler bound procedure for a function of the parameters of a continuous distribution. The following theorem states the continuous analogues of the properties ascribed to upper Buehler bounds in the discrete case. Since the assertions that we make in this theorem can be established by arguments similar to those employed in the discrete case, a formal proof will not be provided.

Theorem 4.2.1: Suppose an ordered partition of the sample space of a continuous random vector X is provided by the statistic S. Then,

(a) $b(s) \leqslant b(s')$ whenever $s < s'$.

(b) $PROB\ (H(\theta) \leqslant b(S(X)) \mid \theta) \geqslant 1\text{-}\alpha$ for all $\theta \in \Theta$.

(c) Suppose $\{d(s): s \in S(\mathfrak{X})\}$ is any other family of $(1\text{-}\alpha)$ 100% upper confidence bounds for $H(\theta)$ based on S. If $d(s) \leqslant d(s')$ whenever $s < s'$, then $b(s) \leqslant d(s)$ for all $s \in S(\mathfrak{X})$.

Let us now consider the following example:

EXAMPLE 4: (Sample Inspection by Variables) Suppose a sample of size n is drawn from a lot of some manufactured product for the purpose of deciding whether the lot is of acceptable quality. Assume that the quality of an item is characterized by a variable Y and an item is considered satisfactory if Y exceeds a given constant, say U. The probability of a defective or of an unsatisfactory item is then given by: $p = \Phi\ (\frac{U\text{-}\mu}{\sigma})$. Suppose we assume that the measurements $Y_1, Y_2, \ldots, Y_n$ constitute a sample from $N(\mu, \sigma^2)$ and an upper Buehler bound for p is to be constructed based on this sample. Since

$$p = \Phi\ (\frac{U\text{-}\mu}{\sigma})$$

where $\Phi(x) = \int_{-\infty}^{x} (1/\sqrt{2\pi})\exp(-\frac{1}{2}t^2)dt$, this would be equivalent to bounding the parametric function $H(\mu, \sigma^2) = (U\text{-}\mu)/\sigma$. Consider the ordering functions $S_1(y) = \bar{y}$ and $S_2(y) = \sqrt{n}(U\text{-}\bar{y})/\hat{\sigma} = t$. With the monotone confidence regions $(\Omega(\bar{y})$ and $(\Omega(t))$ as given in example 2, the upper Buehler bounds for H corresponding to S_1 and S_2 are, respectively:

$$b_1(\bar{y}) = \alpha \quad , \qquad \forall\, \bar{y}$$

and

$$b_2(t) = g(t)/\sqrt{n}$$

where

$$g(t) = \sup\ (\ \delta : PROB\ (\ t_{n\text{-}1;\,\delta} \leqslant t\ ) > \alpha\ )$$

or $g(t) = \delta$ iff t is the (1-α) % point of the $t_{n\text{-}1},\ _\delta$ distribution. Note again that some orderings provide more reasonable bounds than others.

4.3. Optimal Simultaneous Confidence Bounds

If a particular parametric function H is of interest, then we may construct optimal (1-α) 100% upper confidence bounds for H by computing upper Buehler bounds:

$$b(x^{(k)}) \equiv b(k) \equiv \sup\ (\ H(\theta) : \theta \epsilon \Omega(x^{(k)})\ ) \qquad (4.3.1)$$

Indeed, such bounds may be computed for any desired number of parametric functions by computing indiviudal supremums over a common region of optimization and will clearly be simultaneous (1-α) 100% confidence bounds.

Such simultaneous bounds are optimal not only in the sense of being based on the optimal regions $\Omega(\cdot)$ but also in the sense of being uniformly smallest-possible among all similarly ordered bounds. We may therefore conclude that optimal Buehler bounds possess this simultaneity property: they may be computed for as many different parametric functions as we might wish, and will be not only optimal in the Buehler sense, but also simultaneously (1-α) 100% bounds. It may be noted that this assertion rests only on the fact that the monotone regions are (1-α) 100% confidence regions, and not on their "inclusion" optimality as regions. The claim to optimality is based, rather, on our noticing that bounds for individual *H's* computed on the basis of Equations 2.4.1 or 2.4.2 happen to have the form of Buehler's bounds that claim "magnitude" optimality as scalars.

4.4. *The Question of Existence*

In the computation of Buehler bounds for a function $H(\theta)$ of the parameters of the distribution function of a discrete random vector X, we seek the supremum or infimum of H over the region $\Omega(x) = \Omega(x^{(k)}) \doteq (\theta \epsilon \Theta: F(k;\theta) > \alpha)$ where 1-α is the specified confidence level. Clearly, if this set is empty, the Buehler bound for the sample point x is undefined. Since the regions (Ω) are order-dependent and α-dependent, this indicates that Buehler bounds may not exist for certain orderings of the sample space of X and for certain α values.

Let $f(x;\theta)$ denote the likelihood function of x. If $f(x^{(1)};\theta')$ is greater than α for some θ' in Θ, then obviously, $F(x^{(i)};\theta')$ must be greater than α for all $i > 1$. This establishes the following theorem:

Theorem 4.4.1: Let $(x^{(1)}, x^{(2)}, \ldots, x^{(N)})$ denote a particular ordering of the sample space of a discrete random vector X. Then for this ordering, $(1\text{-}\alpha)$ 100% upper and lower Buehler bounds for a parametric function H are defined for all x iff $f(x^{(1)};\theta) > \alpha$ for some θ.

This theorem implies that orderings for which $\Omega(1)$ is empty need not be considered. In addition, it implies that if $f(x;\theta) \leqslant \alpha$ for all $x \epsilon \mathcal{X}$ and $\theta \epsilon \Theta$, then $(1\text{-}\alpha)$ 100% upper and lower Buehler bounds cannot be constructed for all x.

5. *Some Simplifications in the Construction of Buehler Bounds*

In this section, the following definitions and notations will be useful: for a specified ordering of the sample space let

$$h(k) = \sup_{\theta \epsilon \Omega(k)} H(\theta), \qquad \Omega(k) = (\theta \epsilon \Theta: F(k;\theta) > \alpha)$$

$$h'(k) = \sup_{\theta \epsilon \Omega'(k)} H(\theta), \qquad \Omega'(k) = (\theta \epsilon \Theta: F(k;\Theta) = \alpha)$$

$$h^*(k) = \sup_{\theta \epsilon \Omega^*(k)} H(\theta), \qquad \Omega^*(k) = (\theta \epsilon \Theta: F(k:\theta) \geqslant \alpha).$$

Note that $h(k)$ is simply the Buehler upper bound $b(k)$ for $H(\theta)$. The following properties follow directly from the definitions given above.

Lemma 5. 1 $h^*(k) = \max (h(k), h'(k))$.

Proof: By definition, $\Omega^*(k) = \Omega'(k) \cup \Omega(k)$ and the conclusion immediately follows from this relationship. Q.E.D.

Lemma 5.2 ($\Omega^*(k)$: $k=1, 2, \ldots, N$) is a family of $(1-\alpha)$ 100% confidence regions for θ.

Proof: This follows from the observation that for every k, $\Omega^*(k) \supset \Omega(k)$. Q.E.D.

Lemma 5.2 implies that if $h(k)=h'(k)$, then upper Buehler bounds for $H(\theta)$ can be equivalently computed as $h(k)$, $h'(k)$, or $h^*(k)$.

5.1. *Simplifications in the scalar case*

Let us initially consider the case where both X and θ are scalar quantities and let $F(k;\theta)$ with $\theta \in \Theta$ denote the distribution function of X under a specified labeling of the sample space, under which Buehler bounds are defined for all $x \in \mathfrak{X}$. For a fixed k, let

$$\theta(k) = \sup (\theta : F(k;\theta) > \alpha)$$

and

$$\theta'(k) = \sup (\theta : F(k;\theta) = \alpha).$$

The following theorem is easily established:

Theorem 5.1.1: Let $F(k;\theta)$ denote the distribution function of discrete random variable X under a specified ordering of the sample space. Suppose k is fixed.

(a) If F is decreasing in θ and $F(k;\theta) = \alpha$, for some θ, then $\theta(k) \leqslant \theta'(k)$.

(b) If F is decreasing in θ and $F(k;\theta) = \alpha$ for some θ and if H is a bounded, nondecreasing function on $\Omega^*(k)$, then $h(k) \leqslant h'(k)$.

(c) If F is decreasing in θ and $F(k;\theta) = \alpha$ for some θ and if H is bounded, nondecreasing and continuous on $\Omega^*(k)$, then $h(k) = h'(k)$.

Part (c) of the above theorem provides sufficient conditions under which the upper Buehler bounds may be computed over the region $\theta'(k)$ in place of the region $\theta(k)$.

5.2. The nonscalar case

Consider the general situation where $X = (X_1, \ldots, X_k)$ with X_i assuming a finite set of values and $\theta = (\theta_1, \ldots, \theta_p)$ with θ_j defined on an interval. Suppose for a given ordering of sample points, the cumulative distribution function at $x^{(k)}$ satisfies the following properties:

PROPERTY 1. $F(k;\theta)$ is decreasing in θ_j for some j
PROPERTY 2. $F(k;\theta) = \alpha$ for some θ.

The following theorem indicates when the upper Buehler bound is attained at a boundary point of the region $\Omega(k)$.

Theorem 5.2.1: Consider, for fixed k, a cumulative likelihood function $F(k;\theta)$ which satisfies properties 1 and 2.

(a) Suppose H is bounded on $\Omega^*(k)$ and $H(\theta)$ is nondecreasing in θ_j whenever $F(k;\theta)$ is decreasing in θ_j. Then $h(k) = h'(k)$.

(b) Suppose H is bounded on $\Omega^*(k)$, H and $F(k;\theta)$ are continuous in a θ_j for which H is nondecreasing in θ_j whenever $F(k;\theta)$ is decreasing in θ_j. Then $h(k) = h'(k)$.

In lieu of a formal proof of the theorem, let us consider the following important example:

EXAMPLE 5. Let X_k (k=1, 2, 3) be independent binomial (n_k, p_k) random variables. Then $f(x;p) = f(x_1, x_2, x_3; p)$ is given by:

$$\prod_{i=1}^{3} \binom{n_i}{x_i} p_i\, x_i\, q_i\, n_i - x_i \qquad , q_i = 1 - p_i,$$

and the parameter space $\Theta = ((p_1, p_2, p_3): 0 \leqslant p_k \leqslant 1, k=1, 2, 3)$

is a closed and bounded set. The $N = \prod_{k=1}^{3} (n_k + 1)$ sample points can be

strictly ordered, and correspondingly distinctly labeled by ($x^{(1)}$, $x^{(2)}, \ldots, x^{(N)}$), in $N!$ different ways, and, for any one of these labelings, the cumulative distribution function of the ordered points $x^{(i)}$ can be expressed as:

$$F(k;p) = \sum_{j \leqslant k} \prod_{i=1}^{3} \binom{n_i}{x_i^{(j)}} p_i^{x_i^{(j)}} q_i^{n_i - x_i^{(j)}}$$

For a fixed valued of k (k=1, 2, . . ., N), the function F is clearly a polynomial in p_1, p_2, and p_3 and hence is continuous on the parameter space.

Suppose the sample points are distinctly labeled in such a way that the following two conditions are met:

Condition 1. $x^{(1)} = (0, 0, 0)$ and $x^{(N)} = (n_1, n_2, n_3)$.

Condition 2. (Monotonicity Property) If $x_1 = (x_{11}, x_{12}, x_{13})$ and $x_2 = (x_{21}, x_{22}, x_{23})$ are any two points for which $x_{1i} \leqslant x_{2i}$ for all i, then x_1 precedes x_2 in the ordered sample space.

In the sense of Harris and Soms (1980), such an ordering is a "monotone ordering of a finest-possible partition" of the sample space. Obviously, these two conditions do not determine a unique distinct labeling for $\mathfrak{X}$.

We now consider some further properties of the function $F(k;p)$.

Proposition 1: Suppose F is the cumulative distribution function of a distinct labeling, $(x^{(1)}, x^{(2)}, \ldots, x^{(N)})$, of the N possible sample points, that satisfies conditions 1 and 2. Then, for all $k \neq N$, $\partial F(k;p)/\partial p_j < 0$ for all j.

Proof: Without loss of generality, we will show that the conclusion holds for j=3. For fixed k, let us define the following quantities:

$$x_1^{\circ} = \max_{j \leqslant k} x_1^{(j)}$$

$$x_2^{r} = \max_{j \leqslant k} (x_2^{(j)}: (r, x_2^{(j)}, x_3^{(j)})) \qquad (5.2.1)$$

$$x_3^{r,s} = \max_{j \leqslant k} (x_3^{(j)}: (r, s, x_3^{(j)}))$$

and the sets

$$A(r, s) = (x: X_1 = r, X_2 = s, X_3 \leqslant x_3^{r,s})$$

where r=0, 1, . . ., x_1° and s=0, 1, . . ., x_2^{r}. Note that the sets $A(r, s)$ are disjoint and every sample point $x^{(j)}$ for which $j \leqslant k$ is included in the union $\underset{r}{\cup}\underset{s}{\cup} A(r, s)$. Therefore, for any p and $k \neq N$, we may write:

$$\begin{aligned} F(k;p) &= PROB\,(X \in \underset{r}{\cup}\underset{s}{\cup} A(r,s) \mid p) \\ &= \sum_r \sum_s PROB(X_1 = r, X_2 = s, X_3 \leq x_3^{r,s} \mid p) \qquad (5.2.2) \\ &= \sum_r \sum_s PROB(X_1 = r \mid p_1) \times PROB\,(X_2 = s \mid p_2) \\ &\quad \times PROB(X_3 \leq x_3^{r,s} \mid p_3) \end{aligned}$$

Let $g(p_3) = PROB(X_3 \leq x_3^{r,s} | p_3)$. Then, $\partial g(p_3)/\partial p_3 < 0$ for all $x_3^{r,s}$. It follows from (5.2.2) that the partial derivative of F with respect to p_3 is negative. Q.E.D.

This proposition implies that for fixed values of k, p_1 *and* p_2, $F(k; (p_1, p_2, p_3))$ is a decreasing function of p_3.

Proposition 2: Suppose F is the cumulative distribution function of a distinct labeling $(x^{(1)}, x^{(2)}, \ldots, x^{(N)})$ of the N possible sample points that satisfies conditions 1 and 2. The for any fixed value of $k \neq N$, $F(k; p)$ maps the parameter space onto [0, 1].

Proof: Utilizing the quantities defined in expressions (5.2.1) and 5.2.2, the function $F(k; p)$, for fixed k, can be written as:

$F(k; p) = \sum_r \sum_s \sum_t PROB\,(X_1 = r|p_1) \mathrm{x} PROB\,(X_2 = s|p_2) \mathrm{x} PROB\,(X_3 = t|p_3)$ where $t = 0, 1, \ldots, x_3^{r,s}$. Since X_k is binomially distributed,

$$PROB(X_k = x \mid 0) = \begin{cases} 0 & \text{if } x \neq 0. \\ 1 & \text{if } x = 0. \end{cases} \qquad (5.2.3)$$

Hence, for $p = (0, 0, p_3)$, with $p_3 \in [0, 1]$ the product of probabilities in Equation 5.2.3 is nonzero if r=s=0, in which case conditions 1 and 2 imply:

$$F(k; (0, 0, p_3)) = PROB\,(X_3 \leq x_3^{0,0} \mid p_3)\,. \qquad (5.2.4)$$

Now $F(k; p)$ is a polynomial in p_3 and, hence, is continuous in p_3, Hence, by Equation 5.2.4 and the monotone property of a binomial distribution function F assumes all values on [0, 1]. Q.E.D.

This proposition establishes that the set

$$\Omega'(k) = (p: F(k;p) = \alpha)$$

is nonempty for all $k \neq N$ and for all $0 < \alpha < 1$.

Proposition 3: Suppose H is nondecreasing in each p_j. Then, for $k \neq N$, $h(k) \leqslant h'(k)$.

Proof: Without loss of generality, let us assume that H is nondecreasing in p_3. From Proposition 2, we know that, for fixed values of k, p_1, and p_2, $F(k; (p_1, p_2, p_3))$ is strictly decreasing in p_3. This property and the continuity property of $F(k; p)$ in p_3 imply that for every $p' = (p_1', p_2', p_3')$ in $\Omega(k)$, there exists a δ-neighborhood of p' which includes points in $\Omega'(k)$. In particular, if $F(k; (p_1', p_2', p_3')) > \alpha$, then $F(k; (p_1', p_2', p_3' + \delta)) = \alpha$ for some $\delta > 0$. Since H is nondecreasing in p_3, it follows that $H(p_1', p_2', p_3' + \delta) > H(p_1', p_2', p_3')$ and hence that $h(k) \leqslant h'(k)$. Q.E.D.

Proposition 4: Suppose H is bounded and continuous on the parameter space and is nondecreasing in each p_k. Then, $h(k) = h'(k)$ for $k=N$.

Proof: By the continuity property of $F(k; p)$ on Θ, $\Omega^*(k)$ is closed and bounded for all $k \neq N$; hence, the continuity of $H(p)$ implies that $h^*(k)$ is achieved on $\Omega^*(k)$. Let $h^*(k) = H(p^*)$ and suppose that $h(k) < h'(k)$. By Lemma 5.1, this implies that $h'(k) = H(p^*)$ and $p^* \epsilon\ \Omega'(k)$. Hence for some $\eta > 0$,

$$h'(k) - h(k) = H(p^*) - h(k) = \eta$$

Since H is continuous, there exists a neighborhood of p^* such that for all p in this neighborhood, $H(p^*) - \eta/2 \leqslant H(p) \leqslant H(p^*)$. In particular, since $F(k; p)$ is also continuous, there exists p° in this neighborhood such that $p^\circ \epsilon\ \Omega(k)$. But,

$$h(k) = H(p^*) - \eta < H(p^*) - \eta/2 \leqslant H(p^\circ)$$

implies that $h(k)$ cannot be the supremum of H on $\Omega(k)$. Hence, $h(k) = h'(k)$. Q.E.D.

Note that the relationships established in the above propositions hold only for $k \neq N$. To see why this is so, observe that $F(N; p) = 1$

for all p. Hence, for $\alpha \epsilon (0, 1)$, $\Omega'(N)$ is empty. In this case, $h^*(N) = h(N) = H(1, 1, 1)$ for nondecreasing H. Also note that if the function H is nonincreasing, then for the given labeling of sample points, we have the trivial result that $h^*(k) = h(k) = H(0, 0, 0)$ for all k. Clearly, the ordering specified is undesirable for nonincreasing H.

6. *Summary and Conclusions*

Sample space-order-dependent confidence bounds due to Buehler (1957) are shown to be uniformly shortest within the class of similarly-ordered bounds. The monotone confidence regions intrinsic to the Buehler methodology also exhibit an analogous optimality property for regions. Furthermore, they provide for the construction of simultaneous confidence bounds for an arbitrary collection of parametric functions which are also optimal in the Buehler sense.

We may observe that some orderings of the sample space produce more reasonable bounds than others and it is of interest to characterize "optimal" orderings for classes of parametric functions.

References

Buehler, R. J. (1957). Confidence intervals for the product of two binomial parameters. *Journal of the American Statistical Association.* 52:482-493.

Harris, B. and Soms, A. P. (1980). Bounds for optimal confidence limits for series systems. Statistics Department Technical Report No. 611, University of Wisconsin.

______________. (1981). Recent advances in statistical methods for system reliability using Bernoulli sampling of components. Statistics Department Technical Report No. 620, University of Wisconsin.