SEQUENTIAL TOLERANCE REGIONS[1]

By Sam C. Saunders

Boeing Scientific Research Laboratories

Summary. Consider a measurable space with a linear ordering on the space and the family of all probability measures which assign measure zero to each equivalence class induced by the ordering. For such a space and family of probability distributions sequential tolerance regions are defined. The procedure assigns for each finite sample a Borel set with boundaries determined by the order observations. The sampling terminates when the region remains unchanged for a certain number of observations. The coverage of the region thus sequentially determined is distribution free with respect to that family of distributions. Some relationships are derived between the distribution of the coverage and the generating function of the random sample size, which permit the determination of one in terms of the other. This paper includes as a special case the previous results of Jiřina on the distribution of coverage for his sequential procedure. Also, formulae are obtained for the expected sample sizes of the Jiřina procedure which were previously unknown. The results of Wilks for fixed sample tolerance limits are obtained as a limiting case and comparisons are made with sequential procedures in terms of coverage and expected sample size. For example it is shown that for one-sided tolerance limits no sequential procedure is as good as Wilks fixed sample procedure in the sense that if the expected sample sizes are the same the coverage of the Wilks procedure is stochastically greater than the coverage of the sequential procedure.

A discussion of past results. Let X be a random variable (r.v.) on an induced probability space $(\mathfrak{X}, \mathfrak{A}, P)$ and let $V = (X_1, X_2, \cdots, X_n)$ be a vector of n independent replications of X; denote the induced probability space on which V is a r.v. by $(\mathfrak{X}^*, \mathfrak{A}^*, P^*)$. If D is a function mapping $\mathfrak{X}^*$ into $\mathfrak{A}$, then the r.v. $D(V)$ has been called a distribution-free tolerance region whenever the distribution of the random coverage Q, defined $Q = P[D(V)]$, does not depend upon the measure P, under the condition that P belongs to some class of probability measures.

Such tolerance regions as outlined above were first introduced by S. S. Wilks [1] in the following special case: If $\mathfrak{X}$ is the real line, $\mathfrak{A}$ the Borel subsets of $\mathfrak{X}$ and $L(V)$, $U(V)$ are two statistics from $\mathfrak{X}^*$ into $\mathfrak{X}$ such that $U(V) \geqq L(V)$ almost surely (a.s.), then $U(V)$ and $L(V)$ are, respectively, upper and lower β-tolerance limits of probability level α for $\alpha, \beta \in (0, 1)$, if the coverage, letting $D(V)$ be the open random interval $(L(V), U(V))$, $Q = P[D(V)]$ is such that $P^*[Q > \beta] = 1 - \alpha$.

Received April 24, 1958; revised April 10, 1959.

[1] Sections of this paper are related to a thesis written at the University of Washington. That research was sponsored in part by the Office of Naval Research.

Under the condition that P assigns measure zero to all one point sets, Wilks has shown that if $U(V) = X_{(n+1-r)}$, $L(V) = X_{(s)}$ for s, r positive integers such that $n + 1 > r + s > 1$ where $X_{(j)}$ denotes the jth order statistic of V (ordering from the bottom) for $j = 1, \cdots, n$, then $U(V)$ and $L(V)$ are defined a.s. and

$$P^*[Q \leqq \beta] = I_\beta(n + 1 - r - s, r + s)$$

where

$$I_t(m + 1, n + 1) = \frac{(m + n + 1)!}{m!n!} \int_0^t z^m(1 - z)^n dz \quad \text{for} \quad t \, \varepsilon \, (0, 1)$$

denotes the incomplete beta function which is independent of P.

The work on tolerance limits or regions has been generalized to a great extent by A. Wald, H. Scheffé, J. W. Tukey, R. Wormleighton, D. A. S. Fraser, Irwin Guttman, and J. H. B. Kemperman ([2], [3], [4], [5], [6], [7], [8], [9], [10]). These investigations dealt with the construction of tolerance regions for multivariate r.v.'s defined on spaces with several distinct generalized orderings; however, all results were for a fixed sample size.

The first attempt at introducing sequential distribution-free tolerance limits was made by M. Jiřina [11] who proposed the following procedure for finding tolerance limits L and U under the same conditions which Wilks used. Let r, s, k be positive integers. During the first stage take $r + s$ observations and set $L^{(1)} = X_{(r)}$, $U^{(1)} = X_{(r+1)}$. During the jth stage, $j = 2, 3, \cdots$, continue sampling as long as

$$(*) \qquad L^{(j-1)} \leqq X_{t+i} \leqq U^{(j-1)}$$

and $i < k$ where t is the number of observations drawn during the preceding $j - 1$ stages. If $(*)$ holds for $i = k$, terminate the procedure and set $L = L^{(j-1)}$, $U = U^{(j-1)}$. If $X_{t+i} < L^{(j-1)}$ or $X_{t+i} > U^{(j-1)}$ and $i \leqq k$, set $L^{(j)} = X_{(r)}$ and $U^{(j)} = X_{(t+i+1-s)}$ and continue sampling for the $(j + 1)$th stage. He has shown that this procedure terminates a.s. and that if $Q = P[(L, U)]$ then

$$P_W[Q > \beta] = (1 - \beta)^{r+s} \exp \{(r + s) \sum_{i=1}^{k} \beta^i/i\}$$

where P_W is the probability measure in the probability space on which the r.v.'s L and U are defined.

1. Introduction. Suppose we have a sequence of independent observations of a r.v. such that any two of these observations may be compared and the worst (in some sense) determined without knowing the magnitude of either, e.g. they may be placed on a balance and the lighter one discovered without determining their weights. We continue to rank these observations until a particular predetermined arrangement of the ordered observations has occurred, e.g., the Jiřina case where the worst and the best have been unchanged for a number of rankings. The total number of observations ranked when such an event occurs is random. The proportion of the population which will be caught in the region so determined is random. We ask ourselves, apart from the mathematical difficulties one may encounter, what are the distributions of these r.v.'s and how may we optimize within such a class of sequential procedures?

To this end we merely formalize the relevant aspects of the ordered observations of a real r.v. with continuous distribution and the related concept of a distribution-free tolerance region and its coverage.

2. The basic sample space. Let P be a probability measure on a measurable space $(\mathfrak{X}, \mathfrak{A})$. Following the usual notational convention we shall let X be the r.v. on $\mathfrak{X}$, i.e., it is the identity function on $\mathfrak{X}$, and we shall use it to describe events as follows: If π is a proposition involving relations and/or functions for which $A = \{x \varepsilon \mathfrak{X}: \pi(x)\} \varepsilon \mathfrak{A}$ then we shall denote A by $[\pi(X)]$, loosely speaking we say $[\pi(X)]$ is the set of points in $\mathfrak{X}$ such that the relation π is true.

Let θ be a *balance* on $\mathfrak{X}$. By a balance we mean a triplet of binary relations on $\mathfrak{X}$, say $\theta = (\prec, \sim, \succ)$, where $\sim$ is an equivalence relation associated with the irreflexive relations $\prec, \succ$ such that for each $x, y \varepsilon \mathfrak{X}$ exactly one of $x \prec y$ or $x \succ y$ or $x \sim y$ must hold. The relations in the balance $\theta = (\prec, \sim, \succ)$ induce partial orderings on the set of subsets of $\mathfrak{X}$ and we write, e.g., for

$$S, T \subset \mathfrak{X}$$

that $S \prec T$ iff (read if and only if) $s \varepsilon S$, $t \varepsilon T$ imply $s \prec t$.

We say a set Z is *dense* in $\mathfrak{X}$ iff $x, y \varepsilon \mathfrak{X}$ and $x \prec y$ imply $x \prec z \prec y$ for some $z \varepsilon Z$.

(A) We assume there exists a countable set Z which is dense in $\mathfrak{X}$ with respect to θ.

It follows that if $S \prec T$ and $S \cup T = \mathfrak{X}$, then, whether or not S or T is empty,

$$S = \sup_{z \varepsilon ZS} [X \precsim z] \quad \text{or} \quad S = \inf_{z \varepsilon ZT} [X \prec z], \tag{2.1}$$

and we have also if $S' = S - \sup_{z \varepsilon ZS} [X \precsim z]$ is not empty, then

$$S' = [X \sim y] \quad \text{for some} \quad y \varepsilon T. \tag{2.2}$$

Now in order to assure the measurability of the sets under discussion let us define the class $\mathcal{S}$ of sets as follows:

$$\mathcal{S} = \{S \subset \mathfrak{X}: S \prec \mathfrak{X} - S\}.$$

(B) We assume the minimal σ-algebra of $\mathcal{S}$ is $\mathfrak{A}$.

As a point of comparison our assumptions (A) and (B) imply the assumptions (i) and (ii) of Kemperman [10] in his paper on generalized tolerance regions.

Assumptions (A) and (B) have been made stronger than Kemperman's so as to avoid such measurability considerations as arose in his paper. This is done by utilizing the concept of a Lusin space which originated with Blackwell [12]. The definition is as follows: a pair $(\Omega, \mathcal{B})$ is a Lusin space iff (a) $\mathcal{B}$ is separable, i.e., there is a sequence $\{B_n\}$ of elements of $\mathcal{B}$ such that $\mathcal{B}$ is the minimal σ-

algebra of $\{B_n\}$, and (b) the range of every real-valued $\mathcal{B}$-measurable function on Ω is an analytic set, i.e., a set which is the continuous image of the set of irrational numbers.

We have

THEOREM 2.1: *Under assumptions (A) and (B), $(\mathfrak{X}, \mathfrak{A})$ is a Lusin space and the atoms of $\mathfrak{A}$ are the sets of equivalence classes induced by θ on $\mathfrak{X}$.*

PROOF: Let $\mathfrak{X}^*$ denote the set of equivalence classes induced on $\mathfrak{X}$ by

$$\theta = (\prec, \sim, \succ)$$

and set $\{E_n\}_{n=1}^{\infty} = \{[X \precsim z]: z \,\varepsilon\, Z\}$; hence from (B) it follows by definition that $\mathfrak{A}$ is separable. That the atoms of $\mathfrak{A}$ are the points of $\mathfrak{X}^*$ follows from the definition of the atoms of a separable σ-algebra. To complete the proof we remark that in the natural topology $\mathfrak{I}$, with typical element $[y \prec X \prec x]$ we have $(\mathfrak{X}^*, \mathfrak{I})$ metrizable. This follows from the Urysohn Metrization Theorem (see, e.g., p. 125, Kelley, *General Topology*). Now a metric space is analytic if it is the continuous image of the set of irrational numbers (see Blackwell [12]). We define a function g on $\mathfrak{X}$ as follows:

$$g(x) = \sum_{n=1}^{\infty} e_n(x)/3^n$$

where e_n is the characteristic function of E_n. Now g is $1 - 1$ in the sense that $g(x) = g(y)$ implies $x \sim y$ and g is order reversing in the sense that $g(x) < g(y)$ implies $x \succ y$. Now g is clearly continuous. Express $r \,\varepsilon\, (0, 1)$ uniquely in its dyadic expansion $r = \sum_{n=1}^{\infty} a_n/2^n$. Now set $h(r) = \sum_1^{\infty} a_n/3^n$. Then h maps $(0, 1)$ onto $g[\mathfrak{X}]$ in a continuous manner, and so $\mathfrak{X}^*$ is the continuous image of the function $g^{-1}h$. But since the open unit interval is the continuous image of the irrational numbers the result is proved.

Now we remark that the atoms of a Lusin space need not be points and they are not in this case. We remark further that a Lusin space ensures a regularity which along with other advantages permits the identification of Borel and Baire functions and ensures the existence of conditional expectations.

The function defined on $\mathfrak{X}$ by $F(x) = P[X \precsim x]$ is the *distribution* of X and F maps $\mathfrak{X}$ into the unit interval.

THEOREM 2.2: *Now $U = F(X)$ is a* r.v. *with the uniform distribution on* $(0, 1)$ iff $P[X \sim x] = 0$ *for each* $x \,\varepsilon\, \mathfrak{X}$.

PROOF: If for some $x \,\varepsilon\, \mathfrak{X}$ we have $P[X \sim x] > 0$ then F is not onto $(0, 1)$ and U cannot be uniform. Let $u \,\varepsilon\, (0, 1)$ and set

$$S = [F(X) \leqq u], \qquad T = [F(X) > u].$$

Then making use of the properties (2.1) and (2.2) the proof follows.

A balance θ was said by Kemperman to be *continuous* with respect to the measure P iff $P[X \sim x] = 0$ for each $x \,\varepsilon\, \mathfrak{X}$.

Let $\mathcal{P}$ be the class of probability measures on $\mathfrak{A}$ which are continuous with respect to the balance θ and hereafter let P denote generically an element of $\mathcal{P}$.

We have a r.v. X defined on the probability space $(\mathfrak{X}, \mathfrak{A}, P)$ where $P \,\varepsilon\, \mathcal{P}$.

Let W denote the set of positive integers. We write $X_W = (X_1, \cdots, X_n, \cdots)$ for the r.v. on the probability space $(\mathfrak{X}_W, \mathfrak{A}_W, P_W)$ where $\mathfrak{X}_W$ is the countable cartesian product of $\mathfrak{X}$ with itself, $\mathfrak{A}_W$ is the σ-field of subsets of $\mathfrak{X}_W$ generated by all measurable cylinders in $\mathfrak{X}_W$ and P_W is the product measure on $\mathfrak{A}_W$ generated by P. From Blackwell's paper [12] we have the following:

THEOREM 2.3: *$(\mathfrak{X}_W, \mathfrak{A}_W, P_W)$ as defined above is a Lusin space.*

For $x_w \varepsilon \mathfrak{X}_W$ we label $x_{j,n}$ as the jth ordered element determined by θ from $x_1^n = (x_1, \cdots, x_n)$, where $x_{j,n} \precsim x_{j+1,n}$ for $j = 1, \cdots, n-1$. Thus a balance allows the determination of the r.v. $X_{j,n}$ which is the jth order observation of n with respect to θ from the random vector X_W^n.

Now extending our descriptive notation to elements of $\mathfrak{A}_W$ for a given $n \varepsilon W$ we set $K_{(1)}^n = [X_1 \prec X_2 \prec \cdots \prec X_n]$ and then let

$$K_{(i)}^n = [X_{i_1} \prec X_{i_2} \prec \cdots \prec X_{i_n}]$$

for each of the $n!$ permutations $(i) = (i_1, \cdots, i_n)$ of $(1, 2, \cdots, n)$

We will say that $B_n \varepsilon \mathfrak{A}_W$ is a *simplicial set* over $\{1, \cdots, n\}$ whenever there exists a set ψ which is a subset of $\{1, 2, \cdots, n!\}$ such that $B_n = \bigcup_{i \varepsilon \psi} K_{(i)}^n$ a.s. Now any simplicial set is a cylinder set and except for a set of probability zero is the union of simplexes and as such is a set which may be defined by arrangements of the ordered observations. If we let $c(\psi)$ denote the cardinality of the set ψ as defined above, then from the independence of $X_1, \cdots, X_n$ we have $P_W(B_n) = c(\psi)/n!$.

An event $A \varepsilon \mathfrak{A}_W$ is said to be of *structure* (d) on $T = \{t_1, \cdots, t_n\} \subset W$ iff there exists a measurable relation δ symmetric in its arguments and defined on the unit cube such that for any $P \varepsilon \mathcal{P}$, $A = [\delta(F(X_{t_1}), \cdots, F(X_{t_n}))]$ a.s. This nomenclature is adopted from Birnbaum and Rubin [13] because of the obvious similarity.

We have

THEOREM 2.4: *If B_n is some simplicial event and A_n is an event of structure* (d) *on $\{1, \cdots, n\}$, then the two events are independent.*

PROOF: It is sufficient to assume that for some ψ we have a.s.

$$B_n = \bigcup_{j \varepsilon \psi} K_{(j)}^n .$$

Now by the disjointedness of the $K_{(j)}^n$ and the nature of A_n,

$$\begin{aligned} P_W(A_n B_n) &= \sum_{j \varepsilon \psi} P_W[K_{(j)}^n \cap \delta(F(X_1), \cdots, F(X_n))] \\ &= \frac{1}{n!} \sum_{j \varepsilon \psi} n! P_W[\delta(F(X_{1,n}), \cdots, F(X_{n,n}))] \\ &= \frac{1}{n!} \sum_{j \varepsilon \psi} P_W[\delta(F(X_1), \cdots, F(X_n))] = \frac{c(\psi)}{n!} P_W(A_n) \end{aligned}$$

and hence we have independence.

THEOREM 2.5: *If B is a simplicial event on $\{1, \cdots, n\}$, B^* is a simplicial event on $\{n+1, \cdots, m\}$ and $C = AA^*$ where A is of structure* (d) *on $\{1, \cdots, n\}$*

and A^* is of structure (d) *on* $\{n + 1, \cdots, m\}$, *then the events B, B^* and C are independent.*

PROOF: This follows immediately from the preceding and from independence of the components of X_W.

3. Sequential sampling plans. Let $S = (S_1, \cdots, S_n, \cdots)$ be a sequence of disjunct measurable cylinders in $\mathfrak{X}_W$ such that each S_n is a simplicial set over $\{1, \cdots, n\}$. We call such a sequence a *sequential sampling plan* and the events S_n *stopping sets.* The stopping rule for our sequential sampling plan is: stop sampling after the nth observation iff $x_w \varepsilon S_n$. Because S_n is a cylinder set over $\{1, \cdots, n\}$ it is always known after n observations whether or not the event S_n has occurred. We wish to choose S so that for each $P \varepsilon \mathcal{P}$ we have $\sum P(S_n) = 1$, i.e., sampling terminates a.s. We define the r.v. N on $\mathfrak{X}_W$ into W by $N(x_w) = n$ iff $x_w \varepsilon S_n$. N will be called the *random sample size* and its distribution will depend upon our choice of S.

We now exhibit an obvious lemma for later reference which concerns the construction of a sequential sampling plan from a sequence of simplicial sets.

LEMMA 3.1: *If $B = (B_1, \cdots, B_n, \cdots)$ is a sequence of events and B_n is a simplicial set over $\{1, \cdots, n\}$ then $S_n = B_n \bigcap_{j=1}^{n-1} \overline{B_j}$ for $n \varepsilon W$ defines a sequence S of disjunct simplicial sets, and if we write*

$$p_n = P_W(S_n), \qquad q_n = P_W\Big(\bigcap_{j=1}^{n} \overline{B_j}\Big) \qquad \textit{for} \quad n \varepsilon W$$

it follows that

(i) $p_n = q_{n-1} - q_n$ *for $n \varepsilon W$ where $q_0 = 1$,*

(ii) $\sum p_n = 1$ iff $\lim_n q_n = 0$.

Let D be a function which maps $W \times \mathfrak{X}_W$ into $\mathfrak{A}$ which for fixed $n \varepsilon W$ is a function of x_w^n only. Let us write $D(x_w^n)$. The coverage of D is defined by

$$Q(x_w^n) = P(D(x_w^n)).$$

Hence for a given sequential sampling plan S which determines a random sample size N we have a *sequential tolerance region* $D(X_W^N)$ as a set valued r.v. on $(\mathfrak{X}_W, \mathfrak{A}_W, P_W)$ taking values in $\mathfrak{A}$ and the *random coverage* $Q(X_W^N)$ is a r.v. on the same probability space but assuming values in the unit interval.

We now begin a construction of S and D in terms of simplicial sets. Let $b = (b_1, \cdots, b_n, \cdots)$ be a non-decreasing sequence of positive integers such that

1° $n \leqq b_n$ for every $n \varepsilon W$ and if $b_n = n$ for some $n \varepsilon W$ then $b_{n+j} = n + j$ for all $j \varepsilon W$,

2° $\lim_n b_n/n = 1$.

We call b_n a *stopping number* and b is the sequence of sample sizes at which inspection takes place to ascertain if a stopping event has occurred.

For $x_w \varepsilon \mathfrak{X}_W$ we define a subset $A_j(x_w^n) = [x_{j-1,n} < X < x_{j,n}]$ for $j = 1, \cdots, n + 1$ with the obvious definition of A_0, A_{n+1}. Due to the continuity of

θ with respect to $P \varepsilon \mathcal{P}$ it follows for each $n \varepsilon W$ that $A_j(X_W^n)$ is a *statistically equivalent block*. This nomenclature follows from:

LEMMA 3.2: *For fixed $n \varepsilon W$ write $U_j = F(X_{j,n})$ for $j = 1, \cdots, n$. Then the random coverages for each $j = 1, \cdots, n+1$ defined by*

$$C_j(X_W^n) = P_W[A_j(X_W^n)] = U_j - U_{j-1}$$

(where we set $U_0 = 0$, $U_{n+1} = 1$) have the following properties:

(i) $\sum_1^{n+1} C_j(X_W^n) = 1,\ 0 \leqq C_j(X_W^n) \leqq 1$ a.s. *for* $j = 1, \cdots, n+1$,

(ii) *the distribution of the $C_j(X_W^n)$'s is completely symmetrical,*

(iii) $Q(X_W^n) = \sum_{j=1}^k C_{i_j}(X_W^n)$, *where $(C_{i_1}, \cdots, C_{i_{n+1}})$ is any arrangement of $(C_1, \cdots, C_{n+1})$, has the distribution $P_W[Q(X_W^n) \leqq q] = I_q(k, n-k+1)$* for $0 < q < 1$.

PROOF: From Theorem 2.2 we know that U_j's constitute a set of ordered observations from the r.v. with uniform distribution on the open unit interval and the properties (i), (ii) and (iii) are known consequences of this fact (see Tukey [4]).

Let $\lambda = (\lambda_1, \cdots, \lambda_n, \cdots)$ be a sequence of subsets of W such that there exists an $\eta \varepsilon W$ and λ_n is empty if $n < \eta$ and λ_n is a non-empty subset of $\{1, \cdots, n+1\}$ if $n \geqq \eta$, and further, such that $\bar{\lambda}_n = \{1, \cdots, n+1\} - \lambda_n$ has the property that for all $n \geqq \eta$, $c(\bar{\lambda}_n) = \eta$. We define D in terms of λ by

$$D(x_w^n) = \bigcup_{j \varepsilon \lambda_n} A_j(x_w^n) \qquad \text{for } n \geqq \eta,$$

and if additionally we have for each $n \varepsilon W$, $D(x_w^n) \subset D(x_w^{n+1})$, then λ will be called a *selection sequence* with *deletion number* η. Now λ_n tells us what union of statistically equivalent blocks forms our tolerance region after n observations and η is the number of statistically equivalent blocks which are deleted, this number remaining constant. The monotone restriction on D requires that the tolerance region not decrease with increasing sample size.

It is obvious that for each $n \varepsilon W$, $x_w \varepsilon \mathfrak{X}_W$

$$Q(x_w^n) = P(D(x_w^n)) = \sum_{j \varepsilon \lambda_n} P(A_j(x_w^n)) = \sum_{j \varepsilon \lambda_n} C_j(x_w^n)$$

and the r.v.'s so defined have the properties specified in Lemma 3.2.

We now describe a simplicial event in terms of the tolerance region and a type of stability, that is a simplicial event has occurred when after m observations the tolerance region is the same as it was after n observations ($m \geqq n$). More formally,

THEOREM 3.3: *Let a selection sequence λ with deletion number η be given and for fixed integers n, m such that $\eta \leqq n \leqq m$ set $B_{m,n} = [D(X_W^m) = D(X_W^n)]$. Then $B_{m,n}$ is a simplicial set over $\{1, \cdots, m\}$ and*

$$P_W(B_{m,n}) = \binom{n}{\eta} \Big/ \binom{m}{\eta}.$$

PROOF: Now $D(x_w^m) = D(x_w^n)$ iff $x_j \varepsilon D(x_w^n)$ for each $j = n + 1, \cdots, m$. From the independence of the X_i's we have

$$P_W\{\bigcap_{i=n+1}^{m} [X_i \varepsilon \bigcup_{j \varepsilon \lambda_n} A_j(x_w^n)] \mid X_W^n = x_w^n\} = (\sum_{j \varepsilon \lambda_n} C_j(x_w^n))^{m-n}.$$

By the properties of conditional expectation for Lusin spaces we have from Lemma 3.2, making the substitution $z = \sum_{j \varepsilon \lambda_n} C_j(x_w^n)$,

$$P_W(B_{m,n}) = \int_0^1 z^{m-n} dI_z(n - \eta + 1, \eta)$$

and integration yields the result. That $B_{m,n}$ is simplicial is obvious.

A couple (λ, b) we will call a *decision rule* and this nomenclature we make clear momentarily. For a given decision rule we can use the simplicial sets as defined in Theorem 3.3 to obtain a sequence of such sets by taking $m = b_n$ for each $n \geqq \eta$. Since we can without confusion omit the second subscript, let us do so and write B_{b_n} for the event as described. Let us define $B_{b_n} = \phi$ for $n = 1, \cdots, \eta - 1$. Now from the sequence $\{B_{b_n}\}_{n=1}^{\infty}$ we can use Lemma 3.1 to construct a sequential sampling plan $\{S_{b_n}\}$.

We have

THEOREM 3.4: *If from a given decision rule* (λ, b) *the sequential sampling plan* $\{S_{b_n}\}$ *is constructed in the above fashion then* $P_W(\sup S_{b_n}) = 1$.

PROOF: From (ii) of Lemma 3.1 it is sufficient to show that $q_{b_n} \to 0$ since S_{b_n} is simplicial. Now

$$q_{b_n} = P_W(\bigcap_{j=1}^{n} \bar{B}_{b_j}) \leqq P_W(\bar{B}_{b_n}) = 1 - P_W(B_{b_n}).$$

Since we know

$$\lim_{n\to\infty} \frac{n!n^{-h}}{(n - h)!} = 1 \quad \text{for any} \quad h \varepsilon W,$$

it follows from Theorem 3.3 and the definition of stopping numbers that

$$\lim_{n\to\infty} P_W(B_{b_n}) = \lim_{n\to\infty} \frac{(b_n - \eta)!n!}{b_n!(n - \eta)!} = 1.$$

We remark that we are assured of stopping sampling at the least $n \varepsilon W$ such that $b_n = n$ if such exists. Further: the proof of this theorem justifies the introduction of assumption 2° in the definition of $\{b_n\}$.

We examine more closely the structure of the sampling plan $\{S_{b_n}\}$ in the following primary:

THEOREM 3.5: *For a given decision rule* (λ, b) *let* σ *be the function defined by*

$$\sigma_n = \begin{cases} \max\{j \varepsilon W : b_j \leqq n - 1\} & \text{for } n \geqq b_1 + 1 \\ 0 & n = 1, \cdots, b_1. \end{cases}$$

In words: σ_n *is the largest subscript of the* $\{b_j\}$ *for which* $b_j < n$ *if* $n > b_1$ *and otherwise it is zero. Let* L_n *be the set defined* a.s. *by* $L_n = [X_n \varepsilon \bar{D}(X_W^{n-1})]$. *Then*

$$S_{b_n} = \begin{cases} B_{b_n} \cap \bigcap_{j=\eta}^{n-1} \bar{B}_{b_j} = B_{b_n} \cap L_n \cap \bigcap_{j=\eta}^{\sigma_n} \bar{B}_{b_j} & n \geqq \eta \\ \phi & n < \eta. \end{cases}$$

Proof: By construction in Lemma 3.1 we have only to show the equivalence mentioned for $n \geqq \eta$. Let n be fixed and denote the right-hand side by A. To prove that $A = S_{b_n}$ we will show that

$$L_k \subset \bigcap_{n=\sigma_k+1}^{k-1} \bar{B}_{b_n} \quad \text{and that} \quad S_{b_n} \subset A.$$

We have from the definition an equivalent expression

$$B_{b_n} = \bigcap_{i=n+1}^{b_n} [X_i \varepsilon D(X_W^n)]$$

but since $D(X_W^n) \subset D(X_W^{n+1})$ a.s. we have that $L_k \subset \bar{B}_{b_j}$ for all j such that $k > j > \sigma_k$, and hence $L_k \subset \bigcap_{j=\sigma_k+1}^{k-1} \bar{B}_{b_j}$. Now we see that $\bar{L}_n \cap B_{b_n} \subset B_{b_{n-1}}$ by noting that

$$[X_n \varepsilon D(X_W^{n-1})] \bigcap_{i=n+1}^{b_n} [X_i \varepsilon D(X_W^n)]$$

is contained in $\bigcap_{i=n}^{b_{n-1}} [X_i \varepsilon D(X_W^{n-1})]$, which proves the result.

We now have:

Corollary 3.6: *Using the notation introduced previously we have*

$$p_{b_n} = P(S_{b_n}) = q_{\sigma_n} \binom{n-1}{\eta-1} \Big/ \binom{b_n}{\eta} \quad \textit{for} \quad n \varepsilon W$$

with the understanding that $\binom{n}{\eta} = 0$ *if* $n < \eta$.

Proof: By the preceding theorem we have

$$p_{b_n} = P_W(B_{b_n} \cap L_n \cap \bigcap_{j=\eta}^{\sigma_n} \bar{B}_{b_j}).$$

Since $(L_n \cap \bigcap_{j=\eta}^{\sigma_n} \bar{B}_{b_j})$ is a simplicial event on $\{1, \cdots, n\}$ and B_{b_n} can be expressed as an event having structure (d) on $\{1, \cdots, n\}$, independence follows by Theorem 2.4. Apply this argument a second time along with Theorem 3.3 and simplify to obtain the result.

4. The distribution and generating function for a decision rule. Let $r = (\eta, \lambda, b)$ be a decision rule where λ is a selection sequence with deletion number η and b is a stopping sequence. This redundancy of notation has advantages as we shall see. Let R be the space of decision rules. Once $r \varepsilon R$ is chosen, the random sample size N, the tolerance region $D(X_W^N)$, and the coverage $Q(X_W^N)$

[all are r.v.'s defined on the probability space $(\mathfrak{X}_W, \mathfrak{A}_W, P_W)$ and all are functions of r, a fact disguised by our notation] are necessarily determined.

We now define three functions on $(0, 1) \times R$: the distribution G of the coverage, the generating function M of the sample size and a derived function Φ which is determined from M.

$$G(\beta, r) = P_W[Q(X_W^N) \leqq \beta],$$
$$M(\beta, r) = E(\beta^N),$$
$$\Phi(\beta, r) = M_\eta(\beta, r)/(\eta - 1)!$$

where $r = (\eta, \lambda, b)$ and subscripts of M denote derivatives with respect to β.

We now exhibit the main result concerning these functions:

THEOREM 4.2: *If $r = (\eta, \lambda, b)$ is fixed, then*

$$G(\beta, r) = \sum_{n=\eta}^{\infty} p_{b_n} I_\beta(b_n + 1 - \eta, \eta),$$

$$M(\beta, r) = \sum_{n=\eta}^{\infty} p_{b_n} \beta^{b_n},$$

and

$$\Phi(t, r) = \eta \sum_{n=\eta}^{\infty} \binom{n-1}{\eta-1} q_{\sigma_n} t^{b_n - \eta}$$

However, we have these relationships holding:

$$G(\beta, r) = \sum_{j=0}^{\eta-1} M_j(\beta, r) \frac{(1-\beta)^j}{j!} = \int_0^\beta (1-t)^{\eta-1} \Phi(t, r)\, dt,$$

$$M(\beta, r) = \int_0^\beta (\beta - t)^{\eta-1} \Phi(t, r)\, dt.$$

PROOF: Let N be the random sample size determined by r. Now by definition $N(x_w) = b_n$ iff $x_w \,\varepsilon\, S_{b_n}$ so the stated result for M is immediate. Now

$$G(\beta, r) = P_W[Q(X_W^N) \leqq \beta] = \sum_{n=1}^{\infty} P_W(S_{b_n}[Q(X_W^N) \leqq \beta])$$

and by utilizing theorems 2.4 and 2.5 we have

$$G(\beta, r) = \sum_{n=1}^{\infty} p_{b_n} P_W[Q(X_W^{b_n}) \leqq \beta \mid X_W \,\varepsilon\, S_{b_n}]$$

and by Lemma 3.2 the result for G follows. That Φ is as claimed follows from Corollary 3.6 and the definition and the remaining equations follow from repeated integration by parts.

5. A derivation of the Wilks and Jiřina tolerance region procedures. We shall call any family χ of decision rules a *tolerance region procedure* wherever one can make $G(\beta, r)$ for given $\beta \,\varepsilon\, (0, 1)$ arbitrarily small by proper choice of $r \,\varepsilon\, \chi$. This definition includes fixed sample procedures as well as sequential procedures.

We now use our results to obtain the known results.

Procedure 1: (Wilks [1]) Let $\chi = \{(\eta, \lambda, b) \varepsilon R: b_n = k$ for all $n \leqq k$, for some $k \varepsilon W\}$. Now any b has only one stopping number, say k, hence this is a fixed sample procedure where the k observations are drawn and the tolerance region determined by using λ.

If $r = (\eta, \lambda, b)$ and $b_n = k$ for all $\eta \leqq n \leqq k$, then we have $G(\beta, r) = I_\beta(k + 1 - \eta, \eta)$, $E(N) = k$. The proof is immediate since by definition we have $G(\beta, r) = I_\beta(k + 1 - \eta, \eta)\sum_{n=1}^{\infty} p_{b_n}$ and

$$M(\beta, r) = \beta^k, \qquad \Phi(t, r) = \eta \binom{k}{\eta} t^{k-\eta}.$$

It is clear that this is a procedure since the parameter k can be taken arbitrarily large for each fixed η.

Procedure 2: (Jiřina [11]) Let $\chi = \{(\eta, \lambda, b) \varepsilon R: b_n = k + n$ for each $n \varepsilon W$, for some $k \varepsilon W\}$. This is a fixed increment procedure in which sampling stops if the tolerance region obtained from the first n observations remains fixed during the observation of the next k.

We shall show that for any $r \varepsilon \chi$ for which $r = (\eta, \lambda, b)$ and b has increment $k \epsilon W$ such that $\eta < k + 1$ then $G(\beta, r) = 1 - (1 - \beta)^\eta \exp\{\eta \sum_{j=1}^{k} \beta^j/j\}$ and if $\eta \geqq 2$ then we have

$$E(N) = \eta(\eta - 1) \int_0^1 (1 - t)^{\eta-2} t^k \exp\left\{\eta \sum_{j=1}^{k} t^j/j\right\} dt$$

and for $\eta = 1$ we have

$$E(N) = \exp\{\textstyle\sum_{j=1}^{k} j^{-1}\}.$$

We now turn to the proof of the above results. In the light of the Theorem 4.1 it is sufficient that we determine the derived function Φ since from it we can determine both M and G.

Let us fix $r \varepsilon \chi$ as described above and omit its mention. By Theorem 4.2 we have

$$\Phi(t) = \eta \sum_{n=\eta}^{\infty} \binom{n-1}{\eta-1} q_{\sigma_n} t^{b_n - n}$$

where by definition

$$\sigma_n = \begin{cases} 0 & \text{for } n \leqq k + 1 \\ n - k - 1 & \text{for } n \geqq k + 2 \end{cases}$$

and $q_n = 1$ for $n = 0, 1, \cdots, \eta - 1$. Therefore, it follows that

$$\Phi(t) = t^k \eta \sum_{n=\eta}^{\infty} \binom{n-1}{\eta-1} q_{\sigma_n} t^{n-\eta}.$$

To determine Φ it is sufficient to determine γ where $\Phi(t) = t^k\eta\gamma(t)$. Therefore, using the definition of σ given above we have, since $\eta < k + 1$,

$$\gamma(t) = \sum_{n=\eta}^{k+1} \binom{n-1}{\eta-1} t^{n-\eta} + \sum_{n=k+2}^{\infty} \binom{n-1}{\eta-1} q_{n-k-1} t^{n-\eta}$$
$$= \sum_{n=\eta}^{k+\eta} \binom{n-1}{\eta-1} t^{n-\eta} + \sum_{n=\eta}^{\infty} \binom{n+k}{\eta-1} q_n t^{n+k+1-\eta}.$$

Let $\gamma = f + g$ where f and g are, respectively, the first and second terms in the expression above. Then, using the prime notation for derivatives,

$$\gamma'(t) = \sum_{n=\eta+1}^{k+\eta} \binom{n-1}{\eta-1} (n-\eta) t^{n-\eta-1}$$
$$+ \sum_{n=\eta}^{\infty} \binom{n+k}{\eta-1} (n+k+1-\eta) q_n t^{n+k-\eta}$$
$$= \eta \sum_{n=\eta+1}^{k+\eta} \binom{n-1}{\eta} t^{(n-1-\eta)} + \eta \sum_{n=\eta}^{\infty} \binom{n+k}{\eta} q_n t^{n+k-\eta}.$$

Using the recursion relation $q_n = q_{n-1} - p_n$ and the result of Corollary 3.6 we have upon simplification

$$\frac{g'(t)}{\eta} = \sum_{n=\eta}^{\infty} \binom{n+k}{\eta} q_{n-1} t^{n+k-\eta} - \sum_{n=\eta}^{\infty} \binom{n-1}{\eta-1} q_{\sigma_n} t^{n+k-\eta}$$
$$= \binom{n+k}{\eta} t^k + \sum_{n=\eta}^{\infty} \binom{n+1+k}{\eta} q_n t^{n+k+1-\eta}$$
$$- \sum_{n=\eta}^{k+n} \binom{n-1}{\eta-1} t^{n+k-\eta} - \sum_{n=\eta}^{\infty} \binom{n+k}{\eta-1} q_n t^{n+2k+1-\eta}.$$

Now by using the recursion relation for the binomial coefficient one has that

$$\sum_{n=\eta}^{\infty} \binom{n+1+k}{\eta} q_n t^{n+k+1-\eta} = g(t) + \frac{t}{\eta} g'(t)$$

hence

$$\frac{g'(t)}{\eta} = \binom{n+k}{\eta} t^k + \frac{t}{\eta} g'(t) + g(t) - t^k f(t) - tg(t)$$

and substitution shows that

$$\frac{\gamma'(t)}{\eta} = \sum_{n=\eta}^{k+n} \binom{n}{\eta} t^{n-\eta} + \frac{t}{\eta} g'(t) + g(t) - t^k f(t) - t^k g(t).$$

Again using the recursion relation for the binomial coefficient one sees that

$$\sum_{n=\eta}^{k+n} \binom{n}{\eta} t^{n-\eta} = f(t) + \frac{t}{\eta} f'(t),$$

hence we have that $\gamma'(t) = t\gamma'(t) + (1 - t^k)\eta\gamma(t)$. Regarding this last expression as a differential equation and integrating over $(0, t)$, we have, since $\gamma(0) = 1$,

$$\gamma(t) = \exp\{\eta \textstyle\sum_{j=1}^{k} t^j/j\}$$

and hence

$$\Phi(t) = \eta t^k \exp\{\eta \textstyle\sum_{j=1}^{k} t^j/j\}.$$

Integrating $\eta \int_0^\beta (1 - t)^{\eta-1}\gamma(t)\, dt$ by parts will show that

$$\eta \int_0^\beta (1 - t)^{\eta-1} t^k \gamma(t)\, dt = 1 - (1 - \beta)^\eta \gamma(\beta),$$

and hence we can now use theorem 4.2 to obtain

$$G(\beta, r) = 1 - (1 - \beta)^\eta \exp\{\textstyle\sum_{j=1}^{k} \beta^j/j\}$$

and

$$E(N) = \eta(\eta - 1) \int_0^1 (1 - t)^{\eta-2} t^k \gamma(t)\, dt \qquad \text{for } \eta \geqq 2,$$

which gives us one result.

As a consequence of theorem 4.2 $E(N) = G_1(1, r)$ for $\eta = 1$. We now must check only this case, and hence $G(\beta, r) = M(\beta, r) = 1 - \exp\{y(\beta, k)\}$ where

$$y(\beta, k) = -\ln(1 - \beta) - \sum_{j=1}^{k} \beta^j/j = \int_0^\beta t^k (1 - t)^{-1}\, dt.$$

Therefore

$$E(N) = \lim_{\beta \to 1} \frac{y'(\beta, k)}{\exp\{y(\beta, k)\}} = \lim_{\beta \to 1} \frac{\beta}{1 - \beta} \cdot \frac{1 - \beta}{\exp\left\{-\sum_{j=1}^{k} \beta^j/j\right\}}$$

which is the result claimed.

We notice here that for every $\alpha, \beta \,\varepsilon\, (0, 1)$ and every $\eta \,\varepsilon\, W$ we can choose k so that $G(\beta, r) \leqq \alpha$ by simply taking k large enough that

$$\sum_{j=1}^{k} \beta^j/j \geqq \frac{1}{k} \ln(1 - \alpha) - \ln(1 - \beta).$$

That this is always possible can be seen from the fact that

$$\lim_{k \to \infty} \textstyle\sum_{j=1}^{k} \beta^j/j = -\ln(1 - \beta).$$

6. Comparison of sequential procedures. In what we have considered so far we have determined the random sample size by a sequence of events which obtain or not from the sequence of observations. Since it follows from the above that $r \,\epsilon\, R$ determines a random sample size N_r and

$$G(\beta, r) = P_W[Q(X_w^{N_r}) \leqq \beta] = E_{N_r} P[Q(X_w^n) \leqq \beta \mid N_r = n].$$

It follows that for any independent r.v. N with the same distribution in W as N_r the *randomized decision rule* $r^* = (\eta, \lambda, N)$ would have the same distribution of coverage and generating function.

This extension of decision rules allows us to encompass a wider class for comparison.

If r, r' are decision rules we say they are *comparable* iff $E(N_r) = E(N_{r'})$. Now for comparable rules we see that r is better than r' at $\beta \varepsilon (0, 1)$ iff $G(\beta, r) < G(\beta, r')$. We modify our nomenclature in the natural way when the inequality holds *uniformly* for all β in some interval and a rule is said to be *best* when it is better than all the rules in some set of rules. (In the following sections primes affixed to the characters Φ, G and M do not refer to derivatives.)

THEOREM 6.1: *Let r, r' be decision rules with associated functions Φ and Φ', respectively, such that both have number η. If there exists $\beta_0 \varepsilon (0, 1)$ such that $\Phi' < \Phi$ on $(0, \beta_0)$ and $\Phi < \Phi'$ on $(\beta_0, 1)$,* i.e., *β_0 is the unique zero of $\Phi' - \Phi$ in $(0, 1)$, then and only then is $G' < G$ on $(0, 1)$.*

PROOF: By the results of Theorem 4.2 we have by letting $h = \Phi - \Phi'$ that $h > 0$ on $(0, \beta_0)$ and $-h > 0$ on $(\beta_0, 1)$ and hence for each $\beta \varepsilon (0, 1)$

$$G(\beta) - G'(\beta) = \int_0^\beta (1 - t)^{\eta-1}h(t)dt = -\int_\beta^1 (1 - t)^{\eta-1}h(t)\, dt,$$

with the last equation following, since $\int_0^1 (1 - t)^{\eta-1}h(t)\, dt = 0$. From the expression above necessary and sufficient conditions follow.

We also mention the immediate:

COROLLARY 6.1.1: *If Φ, Φ' are defined as above and are such that $\Phi' > \Phi$ on $(\beta_2, 1)$ and $\Phi' < \Phi$ on $(0, \beta_1)$ where $\beta_1 < \beta_2$,* i.e., *$\Phi' - \Phi$ has more than one zero on $(0, 1)$, then we know that $G' < G$* on *$(0, \beta_1)$ and on $(\beta_2, 1)$.*

We now have a criterion for comparison of error functions in the terms of the associated functions Φ and we have immediately:

COROLLARY 6.1.2: *If $\eta = 1$ and $M_1(1, r) < M_1(1, r')$ then there exists one $\beta_0 \varepsilon (0, 1)$ such that $G < G'$ on $(\beta_0, 1)$ and if $\eta = 2$ and $M_2(1, r) < M_2(1, r')$ then there exists a β_0 such that $G < G'$ on $(\beta_0, 1)$.*

We shall say of two procedures χ_1 and χ_2 that χ_1 *is better* than χ_2 at $\beta \varepsilon (0, 1)$ iff $r_1 \varepsilon \chi_1$, $r_2 \varepsilon \chi_2$ are comparable implies $G(\beta, r_1) < G(\beta, r_2)$. Further we shall say χ_1 is *uniformly better* than χ_2 on $(0, 1)$ iff comparability of $r_1 \varepsilon \chi_1$ and $r_2 \varepsilon \chi_2$ implies the inequality holds for all β in the unit interval.

Jiřina claims in [11] that for number $\eta = 1$, Wilks' procedure is uniformly better than his own on $(0, 1)$ but any number $\eta \geqq 2$ his procedure is better than Wilks' for β sufficiently close to unity. However, he assumes firstly, that the procedures are comparable by disregarding the difference between γ, the ASN for his decision rule, and $[\gamma]$ (the greatest integer less than γ), the ASN for the Wilks' decision rule, and secondly, makes no mention of the consideration that the neighborhood of 1 in which his decision rule is better might well depend upon γ.

We know that exp $\{\sum_{j=1}^{k} j^{-1}\}$, which is the expected sample size for $\eta = 1$ for the Jiřina procedure, is not integral for any $k \varepsilon W$ and hence the two procedures are not comparable for $\eta = 1$. From the complexity of the expression for the ASN for $\eta \geqq 2$ given in the procedure 2 it is not apparent that for any value of η one would find the procedures comparable.

However, lack of attention to these details does not vitiate Jiřina's theorem. In fact, by modifying the argument slightly so as to assure comparability and with slight adaptation, Jiřina's proof applies to any other procedure.

Let $\gamma > \eta$ and $\eta \varepsilon W$ be given as an ASN and deletion number, respectively. Now define N by letting $m = [\gamma]$ and setting

$$N = \begin{cases} m & \text{with probability } s \\ m+1 & \text{with probability } 1 - s \end{cases}$$

where $\gamma - m = 1 - s$. Such a rule we call a *randomized Wilks decision rule* $r = (\eta, \lambda, N)$. Let N' be any other random sample size which assumes value n with probability p_n such that $\sum_{n=\eta}^{\infty} np_n = \gamma$.

We quote in our terminology:

THEOREM 6.2 (*Jiřina*): *For $\eta = 1$ a randomized Wilks procedure is uniformly better on* $(0, 1)$ *than any other comparable procedure for every expected sample size.*

THEOREM 6.3 (*Jiřina*): *For $\eta \geqq 2$ and a randomized Wilks decision rule $r = (\eta, \lambda, N)$ and any other comparable decision rule $r' = (\eta, \lambda, N')$ there exists a unique $\beta_0 \varepsilon (0, 1)$ depending on (r, r') such that r is uniformly better on $(0, \beta_0)$ and r' uniformly better in $(\beta_0, 1)$.*

We shall not concern ourselves with the original proof of these theorems since it is lengthy and an alternate proof will be given later.

This last result leads us to:

THEOREM 6.4: *There does not exist a decision rule with number $\eta \geqq 2$ which is uniformly better on* $(0, 1)$ *than all comparable decision rules with number η.*

PROOF: Suppose that we have such a rule with associated function Φ' and ASN equal to γ. Then by Theorem 6.1 if Φ is the associated function of any other comparable rule and $h = \Phi - \Phi'$, then h must possess exactly one zero at, say, $\beta \varepsilon (0, 1)$, and $h > 0$ on $(0, \beta)$ and $-h > 0$ on $(\beta, 1)$. But by Theorem 4.2 we know

$$(*) \qquad \int_0^1 (\eta - 1)(1 - t)^{\eta-2} h(t)\, dt = \gamma - \gamma = 0,$$

and we also have

$$(**) \qquad \int_0^1 (1 - t)^{\eta-1} h(t)\, dt = 1 - 1 = 0.$$

Therefore, letting $g(t) = (1 - t)^{\eta-2} h(t)$ we have from $(*)$ $\int_0^\beta g = \int_\beta^1 -g$ and from $(*)$ and $(**)$ $\int_0^\beta tg(t)\, dt = \int_\beta^1 -tg(t)\, dt$. But

$$\int_0^\beta tg(t)\, dt < \beta \int_0^\beta g = \beta \int_\beta^1 -g < \int_\beta^1 -tg(t)\, dt,$$

which is a contradiction and proves the result claimed.

For his procedure Jiřina has constructed tables of the value of the parameter k needed to attain a value of the error function less than 0.1, .05, .01 for values of β equal to .8, .9, .95. Of course the question is, how does the point β_0 as defined in theorem 6.3 behave as we alter k?

This can be partially answered as follows.

THEOREM 6.5: *Let N be any random sample size and N' a random sample size with the same expectation which has positive probability at no more than two integers which are adjacent. Now define $r_n = (2, \lambda, N + n)$, $r_n' = (2, \lambda', N' + n)$ as translated rules where the number of both is $\eta = 2$. Then for any $\beta \varepsilon (0, 1)$ there exists $m \varepsilon W$ such that $n > m$ implies $G(\beta, r_n) > G(\beta, r_n')$.*

PROOF: Let $h_n(t) = G(t, r_n) - G(t, r_n')$; using theorem 4.2 we have

$$h_n(t) = t^n[G(t, r_0) - G(t, r_0') + \frac{1-t}{t} n(M(t, r_0) - M(t, r_0'))].$$

But it follows by theorem 6.2 that $M(t, r_0) > M(t, r_0')$ for all $t \varepsilon (0, 1)$ hence for any $\beta \varepsilon (0, 1)$ we can, by taking n sufficiently large, force h_n to be positive and hence r_n' is uniformly better than r_n in $(0, \beta)$.

Since one is usually concerned with values of β near 1, one might be led to think from theorem 6.3 that in practical tolerance estimation situations with $\eta = 2$ one could advantageously use a sequential procedure. Unfortunately, however, we are also interested in having α small which forces the ASN to be large. To help clarify this situation we examine $G(\beta, \cdot)$.

We know that for any $r \varepsilon R$ with number η, we have

$$G(\beta, r) = \sum_{n=\eta}^{\infty} p_n I_\beta(n + 1 - \eta, \eta),$$

where

$$I_\beta(n + 1 - \eta, \eta) = \sum_{k=0}^{\eta-1} \binom{n}{k} \beta^n \left(\frac{1-\beta}{\beta}\right)^k$$

by a well known identity, clearly $G(\beta, r)$ is only a linear combination of points on the graph of $I_\beta(\cdot, \eta)$. We now examine this function.

Set $f(x) = \beta^x \sum_{k=0}^{\eta-1} \gamma^k \binom{x}{k}$, where $\gamma = (1 - \beta)/\beta$. We wish to find for β fixed, the values of $x > 0$, where (1) f is convex and (2) f is concave. Clearly (1) iff $f''(x)\beta^{-x} > 0$, (2) iff $f''(x)\beta^{-x} < 0$. Now upon taking derivatives we have for all $\eta \varepsilon W$

$$f''(x) \cdot \beta^{-x} = (\ln \beta)^2 \sum_{k=0}^{\eta-1} \binom{x}{k} \gamma^k + 2 \ln \beta \sum_{k=0}^{\eta-1} \frac{\gamma^k}{k!} \sum_{i=1}^{k+1} S_k^i \, i x^{i-1} + \sum_{k=0}^{\eta-1} \frac{\gamma^k}{k!} \sum_{i=2}^{k+1} S_k^i \, i(i-1) x^{i-2}$$

where the S_k^i are Sterling numbers.

Evaluation in the following special cases yields: if

$$\eta = 1 \; f''(x)\cdot\beta^{-x} = (\ln\beta)^2$$

$$\eta = 2 \qquad = (\ln\beta)^2(1+\gamma\beta)^2$$

$$\eta = 3 \qquad = (\ln\beta)^2\left[1+\gamma x+\frac{\gamma^2}{2}(x^2-x)\right]+2\ln\beta\left[\gamma+\frac{\gamma^2}{2}(2x-1)\right]+\gamma^2.$$

We remark that for $\eta = 1$ f is a convex function for all $x > 0$ and hence that the Wilks' procedure is uniformly best on $(0, 1)$ which is of course in agreement with theorem 6.2. We also have:

THEOREM 6.6: *If $r = (2, \lambda, N)$ is such that $P_W[N = n] = 0$ for $n \leqq [x_\beta] + 1$ where $x_\beta = -2/\ln\beta - \beta/(1-\beta)$ and r' is the comparable rule with positive probability at no more than two adjacent integers and both have numbers $\eta = 2$, then r' is uniformly better than r on $(0, \beta)$.*

PROOF: Using the equation above for $\eta = 2$ we have

$$f''(x)\beta^{-x} > 0 \quad \text{iff} \quad x > \frac{-2}{\ln\beta} - \frac{\beta}{1-\beta} = x_\beta .$$

Hence r' is better than r at β and by the results of theorem 6.3 it must be uniformly better on $(0, \beta)$.

The theorem above also throws light on the results of theorem 6.3 as to why r' is not uniformly better on (0,1).

As an aid in computation we prove:

COROLLARY 6.6.1: *With f and x_β defined as above for $\eta = 2$ we have*

(1) $\quad f(x_\beta) = 2e^{-1}[1 + (\beta-1)^2/8 - (\beta-1)^3/8 + o(1-\beta)^4],$

(2) $\quad x_\beta = (1-\beta)^{-1} - (1-\beta)/6 - (1-\beta)^2/12 + o(1-\beta)^3.$

PROOF: (1) We write $f(x_\beta) = 2e^{-2}h(\beta)$ where $h = g^{-1}e^g$ and

$$g(\beta) = -\beta\ln[\beta/(1-\beta)].$$

Expand h in a power series in terms of g and simplify and one obtains the result.

(2) We write $h(1-\beta) = (1-\beta)x_\beta$ which has a power series about $\alpha = 1-\beta$. Expanding and simplifying yields the result.

A further result on the comparison of two decision rules in the case of the number $\eta = 2$ is as follows.

THEOREM 6.7: *Let $r = (2, \lambda, N)$ be given where $p_j, p_k > 0$ for some $j + 1 < k$, let us define $r' = (2, \lambda, N')$ by $P_W[N' = n] = p_n' \geqq 0$ with $p_j' = p_j - \epsilon$, $p_m' = p_m + \epsilon$, $p_l' = p_l + \epsilon$, $p_r' = p_r - \epsilon$ and $p_n' = p_n$ for n elsewhere, and we suppose that $j < m \leqq l < k$ and $m - j = k - l = s$, then r' is uniformly better than r on $(0, (j/k)^{s-1})$.*

PROOF: Let $\sum' = \sum_{n \varepsilon W'}$ where $W' = \{n \varepsilon W : n \neq j, m, l, k\}$. Let $jm^{-1} = \delta$, $lk^{-1} = \sigma$, $ml^{-1} = \gamma$, $k - m = l - j = vs$ where v is some rational number. From the above follows

$$0 < \gamma \leqq 1, \qquad 0 < \delta < \sigma < 1, \qquad 1 + \gamma\sigma\delta = \sigma(1+\gamma). \qquad (*)$$

Let us denote $M(\cdot, r')$ by M' and similarly for G', M, and G.

$$M'(\beta) = \sum p'_n\beta^n = \sum{}' p_n\beta^n + \beta^j(p_j - \epsilon) + (p_m + \epsilon)\beta^m + (p_l + \epsilon)\beta^l + (p_k - \epsilon)\beta^k$$

$$M'(\beta) = M(\beta) - \epsilon\beta^j + \epsilon\beta^m + \epsilon\beta^l - \epsilon\beta^k = M(\beta) - \epsilon\beta^j(1 - \beta^s)(1 - \beta^{sv})$$

$$E(N') = E(N) - \epsilon j + \epsilon m + \epsilon l - \epsilon k = E(N) + \epsilon(m - j) - \epsilon(k - l)$$

hence we have r, r' comparable.

$$\begin{aligned} G'(\beta) &= M'(\beta) + (1 - \beta)M'_1(\beta) \\ &= M(\beta) - \epsilon\beta^j(1 - \beta^s)(1 - \beta^{sv}) \\ &\qquad + (1 - \beta)[M_1(\beta) + m\epsilon\beta^{j-1}(-\delta + \beta^{m-j}) + \epsilon k\beta^{l-1}(\sigma - \beta^{k-l})]. \end{aligned}$$

Now upon simplification we obtain $G'(\beta) - G(\beta) = \epsilon\beta^{j-1}g(\beta)$ where we have $g(\beta) = -\beta(1 - \beta^s)(1 - \beta^{sv}) + (1 - \beta)[m(\beta^s - \delta) + \beta^{sv}k(\sigma - \beta^s)]$. We now must examine g; setting $t = \beta^s$, $h(\beta^s) = g(\beta)$ we have

$$h(t) = -t^{s^{-1}}(1 - t)(1 - t^v) + (1 - t^{s^{-1}})kf(t)$$

where $f(t) = (1/k)[m(t - \delta) + t^v k(\sigma - t)] = t^v(\sigma - t) + \gamma\sigma(t - \delta)$. Using $(*)$ above we see that $f(t) = (t^v - \gamma\sigma\delta)(\sigma - t) + (1 - \sigma)(t - \gamma\sigma\delta)$ from which we can see that $t < \gamma\sigma\delta$ implies $f(t) < 0$.

Set $t_0 = \gamma\sigma\delta$ therefore $h < 0$ on $[0, t_0]$ and we have $g < 0$ on $(0, \beta_0)$ where $\beta_0 = t_0^{s^{-1}}$ and hence $G' < G$ on the same interval.

COROLLARY 6.7.1: *If in the above theorem the number $\eta = 1$ then r' is uniformly better than r on $(0, 1)$.*

PROOF: The result follows immediately from the identity

$$M'(\beta) = M(\beta) - \beta^j(1 - \beta^s)(1 - \beta^{sv})$$

used in the preceding argument.

These last results have been carried out for the simpler cases for numbers $\eta = 1, 2$, and clearly for larger η the results are more tedious.

Acknowledgments: The author is much indebted to Prof. Z. W. Birnbaum, of the University of Washington, who first proposed this problem and directed the thesis research and then allowed this paper to be published under my name.

REFERENCES

[1] S. S. WILKS, "Determination of sample sizes for setting tolerance limits," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 91–96.

[2] A. WALD, "An extension of Wilks' method for setting tolerance limits," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 45–55.

[3] H. SCHEFFÉ AND J. W. TUKEY, "Non-parametric estimation: I. Validation of order statistics," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 187–192.

[4] J. W. TUKEY, "Nonparametric estimation: II. Statistically equivalent blocks and tolerance regions—the continuous case," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 529–539.

[5] J. W. TUKEY, "Nonparametric estimation: III. Statistically equivalent blocks and

tolerance regions—the discontinuous case," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 30–39.

[6] D. A. S. FRASER AND R. WORMLEIGHTON, "Nonparametric estimation: IV," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 294–298.

[7] D. A. S. FRASER, "Sequentially determined statistically equivalent blocks," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 372–381.

[8] D. A. S. FRASER, "Nonparametric tolerance regions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 44–55.

[9] D. A. S. FRASER AND IRWIN GUTMAN, "Tolerance regions," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 162–179.

[10] J. H. B. KEMPERMAN, "Generalized tolerance limits," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 180–186.

[11] MIROSLAV JIŘINA, "Sequential estimation of distribution-free tolerance limits," *Cehoslovack. Mat. Z.* 2(77)(1952), pp. 211–232; correction 3(78)(1953), p. 283.

[12] DAVID BLACKWELL, "On a class of probability spaces," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. II (1956), pp. 1–6.

[13] Z. W. BIRNBAUM AND H. RUBIN, "On Distribution-Free Statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 593–598.