Statistical Research Report
Institute of Mathematics
University of Oslo

No 1
January 1969

DETERMINATION OF SAMPLE SIZE FOR DISTRIBUTION-FREE TOLERANCE LIMITS

by

Visiting professor
from University of Wyoming
William C. Guenther

DETERMINATION OF SAMPLE SIZE FOR DISTRIBUTION-FREE TOLERANCE LIMITS

WILLIAM C. GUENTHER
UNIVERSITY OF WYOMING

Many problems involving distribution-free tolerance limits can be solved quickly with a good binomial table. This paper considers several situations for which such a table yields under a given set of conditions the minimum sample size required.

1. INTRODUCTION

Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution having probability density function $f(x)$ and distribution function $F(x)$. Denote the order statistics by $Y_1, Y_2, \ldots, Y_n$ and let p and γ be given positive fractions less than 1. If

$$\Pr[F(Y_j) - F(Y_i) \geqq p] \geqq \gamma \tag{1.1}$$

then (y_i, y_j) is called a two sided tolerance interval with tolerance coefficient γ for a fraction p of the probability of the destribution of X. Similarly, if

$$\Pr[1 - F(Y_i) \geqq p] \geqq \gamma \tag{1.2}$$

or

$$\Pr[F(Y_j) \geqq p] \geqq \gamma \tag{1.3}$$

then (y_i, ∞) and $(-\infty, y_j)$ are respectively one sided tolerance intervals with tolerance coefficient γ for a fraction p of the probability of the distribution of X. The numbers y_i, y_j are called lower and upper tolerance limits respectively.

One type of problem which has received considerable attention in the literature requires as its solution for a given i or j or both the minimum n which satisfies (1.1), (1.2), or (1.3) depending upon which case is under consideration. Some papers which have dealt with this situation are those by Murphy [5], Scheffé and Tukey [7], Sommerville [8], and Wilkes [10]. In these papers are found specially constructed tables, an approximate formula for n, and solutions using the incomplete beta table of Pearson [6].

A second type of problem is obtained by imposing a second condition on the random interval $[(Y_i,Y_j),(Y_i,\infty)$, or $(-\infty,Y_j)]$. Suppose that for given γ, δ, p_0, p_1, where $p_1 > p_0$, we desire a lower tolerance limit y_i such that the inequalities

$$Pr[1 - F(Y_i) \geqq p_0] \geqq \gamma \qquad (1.4)$$

and

$$Pr[1 - F(Y_i) \geqq p_1] \leqq \delta \qquad (1.5)$$

are both satisfied for minimum n. Faulkenberry and Weeks [1] have considered the corresponding problem for three parametric cases (uniform, exponential, and normal). As motivation they have suggested a possible application in which a manufacturer desires a lower tolerance limit but does not want it to be unnecessarily small since this may make his product look inferior. Similarly for the two sided case we could seek tolerance limits y_i, y_j, such that the conditions

$$Pr[F(Y_j) - F(Y_i) \geqq p_0] \geqq \gamma \qquad (1.6)$$

and

$$Pr[F(Y_j) - F(Y_i) \geqq p_1] \leqq \delta \qquad (1.7)$$

are both satisfied for minimum n. To illustrate the use of this case a producer of a product with lower and upper specification limits may wish to assure with tolerance coefficient γ that at least a fraction p_0 of his product is between y_i and y_j but does not want these numbers unnecessarily far apart (and perhaps not in the range of defective material).

It is well known (and immediately obvious) that the left hand side of (1.2) can be written as

$$\Pr[1 - F(Y_i) \geqq p] = E(i;n,1-p) \qquad (1.8)$$

and the left hand side of (1.3) as

$$\Pr[F(Y_j) \geqq p] = E(n-j+1;n,1-p) \qquad (1.9)$$

where

$$E(r;n,p) = \sum_{w=r}^{n} \binom{n}{w} p^w (1-p)^{n-w}$$

The left hand side of (1.1) is slightly more difficult to handle but it is shown a number of places in the literature, including textbooks (i.e., Hogg and Craig [3, pp. 182-85]), that

$$\Pr[F(Y_j) - F(Y_i) \geqq p] = \Pr(V \geqq p)$$

where V has a beta distribution with parameters $j - i$ and $n - j + i + 1$. Using the relationship between the incomplete beta integral and the sum of binomial terms yields

$$\Pr[F(Y_j) - F(Y_i) \geqq p] = E(n-j+i+1;n,1-p) \qquad (1.10)$$

We will illustrate the usefulness of binomial tables (and Poisson tables for large n) in the solution of the type of problems described above.

2. FIRST TYPE OF PROBLEM; ONE CONDITION ON TOLERANCE INTERVAL

Using (1.10), (1.8), and (1.9) we can rewrite (1.1), (1.2), and (1.3) as

$$E(n-j+i+1;n,1-p) \geqq \gamma \quad (2.1)$$

$$E(i;n,1-p) \geqq \gamma \quad (2.2)$$

$$E(n-j+1;n,1-p) \geqq \gamma \quad (2.3)$$

respectively. As an example let $p = .90$, $\gamma = .95$ $i = 1$, $j = n - 1$. The above inequalities become

$$E(3;n,.10) \geqq .95$$

$$E(1;n,.10) \geqq .95$$

$$E(2;n,.10) \geqq .95$$

With the Ordnance Corps [9] table we find (by observation) that to satisfy the inequalities we must have $n \geqq 61$, $n \geqq 29$, $n \geqq 46$ respectively. Then in the two sided case the desired tolerance limits are y_1 and y_{60} with $n = 61$; in the lower tolerance limit case we use y_1 with $n = 29$; in the upper tolerance limit case we use y_{45} with $n = 46$.

3. SECOND TYPE OF PROBLEM; TWO CONDITIONS ON TOLERANCE INTERVAL

As a first example we consider a problem involving a lower tolerance limit. Let $p_0 = .85$, $\gamma = .90$, $p_1 = .96$, $\delta = .05$. Then inequalities (1.4) and (1.5) become

$$E(i;n,.15) \geqq .90 \ , \ E(i;n,.04) \leqq .05$$

The solution is obtained by trial starting with $i = 1$. From the binomial table we find with $i = 1$ that we must have $n \geqq 19$ and $n \leqq 1$ to satisfy the two inequalities so that no solution is possible. With $i = 2$ we find that we need $n \geqq 25$ and $n \leqq 9$ again an impossibility. Similarly, we get

with	$i = 3$	$n \geqq 34$ and	$n \leqq 21$
	$i = 4$	$n \geqq 43$ and	$n \leqq 34$
	$i = 5$	$n \geqq 52$ and	$n \leqq 50$
	$i = 6$	$n \geqq 60$ and	$n \leqq 66$

Hence, the minimum sample size to provide a solution is $n = 60$ and the lower tolerance limit is y_6. Under the assumption that X has a normal distribution Faulkenberry and Weeks [1] found that $n = 35$ for this problem.

As a second example suppose that p_0, p_1, γ, and δ are as in the first example but we desire a two sided tolerance interval. We get the same inequalities and the same solution except that i is replaced by $n - j + i + 1$. That is, $n = 60$, $60 - j + i + 1 = 6$ or $j - i = 55$. Possible choices are (y_5, y_{60}), (y_4, y_{59}), (y_3, y_{58}), (y_2, y_{57}), (y_1, y_{56}). We may prefer the "symmetric" interval (y_3, y_{58}).

As a third example suppose that we again use the same p_0, p_1, γ, δ but desire an upper tolerance limit. Again we get the same inequalities with the same solution except that i is replaced by $n - j + 1$. Thus $n = 60$, $60 - j + 1 = 6$, $j = 55$ and the upper tolerance limit is y_{55}, the 6th largest observation as contrasted with the 6th smallest in the lower tolerance limit case (as we would expect).

Finally, we consider a case for which the Poisson approximation is useful. Suppose that we desire a lower one sided tolerance limit when $p_0 = .95$, $p_1 = .98$, $\gamma = .90$, $\delta = .05$. Now the inequalities (1.4) and (1.5) become

$$E(i;n,.05) \geqq .90 \ , \ E(i;n,.02) \leqq .05$$

Since the solution requires that $n > 150$, the Ordnance Corps table is of no use. The Harvard [2] table contains n's up to 1000 but in steps of 20 in the range needed. Thus, it is convenient to replace the above inequalities by

$$E(i;.05n) \geqq .90 \ , \ E(i;.02n) \leqq .05$$

where

$$E(r;\mu) = \sum_{w=r}^{\infty} \frac{e^{-\mu}\mu^w}{w!}$$

As before increase i until a solution is found. With $i = 10$ and linear interpolation the Molina [4] table yields

$$E(10;14.21) = .90 \ , \ E(10;5.42) = .05$$

so that $.05n \geqq 14.21$, $n \geqq 284.2$ and $.02n \leqq 5.42$, $n \leqq 271$ but with $i = 11$ we find

$$E(11;15.4) = .90 \ , \ E(11;6.17) = .05$$

so that $.05n \geqq 15.4, n \geqq 308$ and $.02n \leqq 6.17$, $n \leqq 308.5$. Hence $n = 308$ and the tolerance limit is y_{11} .

REFERENCES

1. FAULKENBERRY, G.DAVID and WEEKS, DAVID L., 1968. Sample Size Determination for Tolerance Limits. Technometrics, 10, 343-348.

2. HARVARD UNIVERSITY COMPUTATION LABORATORY, 1955. Tables of the Cumulative Binomial Probability Distribution. Harvard University Press, Cambridge, Mass.

3. HOGG, ROBERT V. and CRAIG, ALLEN T., 1965. Introduction to Mathematical Statistics, (2nd. ed.) Macmillan, New York.

4. MOLINA, E.C., 1949. Poisson's Exponential Binomial Limit. D. Van Nostrand, Princeton, N.J.

5. MURPHY, R.B., 1948. Non-parametric Tolerance Limits. Annals of Mathematical Statistics, 19, 581-588.

6. PEARSON, KARL, 1968. Tables of the Incomplete Beta Function, (2nd. ed.). Cambridge University Press, Cambridge.

7. SCHEFFÉ, H. and TUKEY, J.W.,1944. A Formula for Sample Sizes for Population Tolerance Limits. Annals of Mathematical Statistics, 15,217.

8. SOMMERVILLE, PAUL N., 1958, Tables for Obtaining non-parametric Tolerance Limits. The Annals of Mathematical Statistics, 29, 599-600.

9. U.S. ARMY ORDNANCE CORPS, 1952. Tables of Cumulative Binomial Probabilities. Ordnance Corps Pamphlet ORD20-1. Office of Technical Services, Washington, D.C.

10. WILKS, S.S., 1941. Determination of Sample Sizes for Setting Tolerance Limits. Annals of Mathematical Statistics, 12, 91-96.