UNIFORMLY MOST ACCURATE

TOLERANCE LIMITS

By

GERALD DAVID FAULKENBERRY

Bachelor of Science
Southeastern State College
Durant, Oklahoma
1959

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1962

Submitted to the Faculty of the Graduate School of
the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY
May, 1965

UNIFORMLY MOST ACCURATE

TOLERANCE LIMITS

Thesis Approved:

David L. Weeks
Thesis Adviser

J. Leroy Folks

Robert D. Morrison

Joe W. Whiteman

R B Deal

[signature]
Dean of the Graduate School

ACKNOWLEDGEMENTS

I am indebted to Dr. David L. Weeks for serving as chairman of my graduate committee, for suggesting this thesis problem and for his guidance in the preparation of this thesis.

I appreciate Professors J. Leroy Folks, Robert D. Morrison, Roy B. Deal, and Joe Whiteman serving on my advisory committee and I am grateful for the encouragement and help of Professor Carl E. Marshall, head of the Statistical Laboratory.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

The primary objective of this thesis is to consider a measure of accuracy for tolerance limits and investigate the problem of determining the sample size required to obtain a specified degree of accuracy.

When considering the problem of determining confidence limits on a parameter of a density function, say $f(x;\theta)$, work has been done on several reasonable measures of accuracy. For example, if a two sided confidence interval is desired, then a minimum width interval might be considered (for given probability level and given sample size). Or a minimum expected width interval might be considered if the width is a random variable. If a one-sided confidence limit is desired then the uniformly most accurate property specifies a type of optimum choice. A <u>uniformly most accurate</u> confidence limit is one which minimizes the probability of covering a false value of the parameter. That is, a uniformly most accurate lower confidence limit for θ is a function of the sample, say $\underline{\theta}(X)$, such that $Pr(\underline{\theta}(X) < \theta) = a$, and if $\underline{\theta}^*(X)$ is any other limit such that $Pr(\underline{\theta}^*(X) < \theta) = a$, then $\theta' < \theta$ implies that $Pr(\underline{\theta}(X) < \theta') < Pr(\underline{\theta}^*(X) < \theta')$. A uniformly most accurate upper confidence limit is defined analogously.

The accuracy of tolerance limits has not been investigated as thoroughly as has the accuracy of confidence limits. This has been partly due to the lack of a suitable criterion of accuracy in the case of tolance limits. In this thesis tolerance limits which satisfy the condition $Pr(\text{coverage} > p) = a$ are considered, so measures of accuracy for this type limit are studied. Precision criterion for p-expectation tolerance limits have been thoroughly investigated by Fraser and Guttman (2).

One type of criterion which has been investigated for the type of tolerance limits we are considering is <u>expected coverage</u>. That is, if we had two tolerance limits both of which satisfied the condition $Pr(\text{coverage} > p) = a$, then the one with smaller expected coverage would be considered better. However this is not a good criterion to use for the problem of determining a sample size so that a given precision will be obtained, since it does not take into consideration the variation of the limit.

Goodman and Madansky (3) have formulated a criterion for comparing tolerance limits as follows: Suppose we have a tolerance interval, call it I, such that $Pr(\text{coverage of I} > p) = a$. Then I is called the <u>most stable</u> tolerance interval if for any other tolerance interval, say II, such that $Pr(\text{coverage of II} > p) = a$, and for every p_1 and p_2 such that $p_1 > p > p_2$ we have

$$Pr(p_1 > \text{coverage of I} > p) > Pr(p_1 > \text{coverage of II} > p)$$

and

$$Pr(p > \text{coverage of I} > p_2) > Pr(p > \text{coverage of II} > p_2) .$$

This is a desireable and practical criterion for comparing tolerance intervals since it not only requires that the coverage be greater than p with fixed probability, but it also takes into consideration how concentrated the distribution of the coverages is about p. The authors of this article did not study any general properties of this criterion, and they did not attempt to formulate it in a convenient manner so that the problem of sample size determination to achieve given stability requirements could be considered.

Bain (1) considers the problem of determining the sample size such that for given p, e, and a, the tolerance limit satisfies the statement $Pr(p+e > coverage > p) = a$. The best tolerance limit to use (different tolerance limits can be obtained by using different statistics) for this case is not considered. In fact, it is difficult to compare two different limits in this manner because $Pr(coverage > p)$ generally changes for different limits derived to satisfy the above statement. It would seem more practical and more consistent with the standard tolerance interval problem to be able to fix $Pr(coverage > p)$ and then impose some other condition for the coverage to satisfy.

In the practical problem of determining a tolerance limit for a distribution it seems useful to have the statement $Pr(coverage > p) = a$ satisfied, then have a criterion of accuracy which is meaningful in terms of the problem being considered and which provides answers to the following questions.

(1) Does the tolerance limit improve as the sample size is increased?

(2) If the limit does improve with increased sample size, can a sample size be determined to achieve a desired accuracy?

(3) Given two tolerance limits which are different functions of the sample values, can we say that one is better than the other?

(4) Does there exist a "best" limit?

In this thesis a criterion is formulated which gives reasonable answers to these questions. Some general properties of this criterion are considered. In particular, requirements on the distribution which insure that a best tolerance limit exists are given, and a method for obtaining this best limit is given. The problem of determining the sample size to obtain a given accuracy requirement is worked out and illustrated for some particular distributions, and some comparisons between different tolerance limit forms are made.

A variation of the tolerance limit problem is also considered. A common situation is to have a tolerance specification given, and the problem is then to determine what fraction of the population meets this specification. Rather than just estimate the fraction of the population which meets this specification, it might be more realistic in a given problem to set a lower confidence limit on this fraction. This is seen to be the same problem as considered before except that the limit is given and p is to be determined for a given value of a. A measure of precision is defined for this problem and the sample size determination problem is worked out and illustrated for the exponential distribution.

CHAPTER II

AN ACCURACY CRITERION FOR TOLERANCE LIMITS

We are considering tolerance limits which satisfy the statement Pr(coverage $>$ p) = a, and we would like to have a measure of how close the coverage is likely to be to p. In terms of the distribution of the coverage, we will have something such as in Figure 1. These might

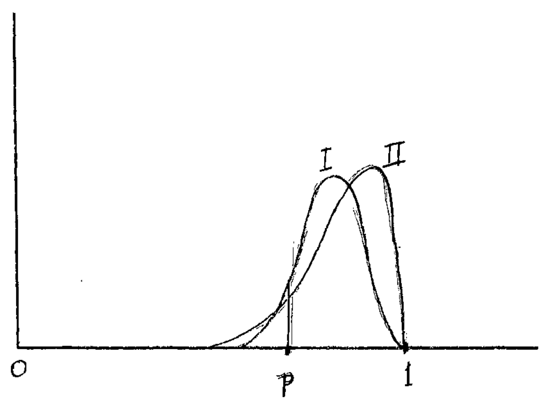

Figure 1

be the distributions of coverage for two different tolerance limits, that is, two different functions of the sample. The coverage in both cases will have probability a of being greater than p, however I will be considered better than II in that the distribution is concentrated more in the neighborhood of p.

Accuracy Criterion

We desire a means of measuring the concentration about p of the distribution of the coverage associated with a tolerance limit which will be simple to interpret and reasonably easy with which to work.

The criterion proposed here is a modification of the most stable property proposed by Goodman and Madansky (3). It corresponds to the same type of reasoning associated with the idea of a uniformly most accurate confidence limit. We propose to use Pr(coverage $>p'$) where $p' > p$ as a measure of accuracy for a tolerance limit which satisfies Pr(coverage $>p$) = a. Now this seems to be easy enough to interpret in terms of physical problems, and it gives the kind of measure of concentration about p for which we are looking. Its usefulness will depend on how well it lends itself to theoretical development and how well it performs for specific distributions.

We now see how this criterion can be used to answer the questions proposed at the end of Chapter I.

(1) Does the tolerance limit improve as the sample size is increased? In terms of the proposed criterion this is the same as asking if Pr(coverage $>p'$) is a decreasing function of the sample size.

(2) If the limit does improve with increased sample size, can a sample size be determined to achieve a desired precision? This is the same as choosing a $p' > p$ and an a'(usually small) and determining if there is a solution to the equation Pr(coverage $>p'$) = a'. (For intergral values of n this will have to be $\leq a'$)

(3) Given two tolerance limits which are different functions of the sample, can we say that one is better than the other? Here we will say that tolerance limit I is better than II if for all $p' > p$, we have

$$\Pr(\text{ coverage of I} > p') < \Pr(\text{ coverage of II} > p')$$

where

$$\Pr(\text{ coverage of I} > p) = \Pr(\text{ coverage of II} > p).$$

(4) Does there exist a best tolerance limit? This will be asking, "Is there a tolerance limit, say I, with Pr(coverage of I $>$p) = a such that for any other limit, II, with Pr(coverage of II$>$p) = a, we have for all $p' > p$

$$\Pr(\text{ coverage of I} > p') \leq \Pr(\text{ coverage of II} > p')?\text{"}.$$

If such a limit as I exists then we will call it the <u>uniformly</u> <u>most</u> <u>accurate</u> <u>tolerance</u> <u>limit</u> such that Pr(coverage$>$p) = a.

We see then that these questions can quite readily be formulated in terms of the proposed criterion. We now need to consider an example to which we can apply this criterion in order to study these questions and clarify the ideas.

Example: The Uniform Density

Let $X_1,\ldots,X_n$ be a random sample from a population with uniform density function

$$f(x;\theta) = 1/\theta \qquad \text{for } 0<x<\theta$$
$$= 0 \qquad \text{otherwise,}$$

and let $L(X_1,\ldots,X_n) = L$ denote a real valued function defined on the sample space. We then consider the problem of determining L such that

$$\Pr\left[\int_L^\theta (1/\theta)dx > p\right] = a,$$

that is, determine a lower tolerance limit on 100p% of the distribution with probability a. Now

$$\Pr\left[\int_L^\theta (1/\theta)dx > p\right] = \Pr(\ 1 - L/\theta > p\ )$$

$$= \Pr(\ L < \Theta(1 - p)\ ),$$

and $Z = \max_i(X_i)$ is the minimal sufficient statistic for Θ. Also $Y = Z/\Theta$ has density function

$$g(y) = ny^{n-1} \quad \text{for } 0 < y < 1$$
$$= 0 \quad \text{otherwise.}$$

Cumulative points on this density will be given by $\Pr\left[\ Y < (a)^{1/n}\right] = a$. Therefore we have

$$a = \Pr\left[\ Z/\Theta < (a)^{1/n}\right]$$
$$= \Pr\left[\ Z(1 - p)/(a)^{1/n} < \Theta(1 - p)\right],$$

and if we compare this with the statement above we see that a choice for L is

$$L = Z(1 - p)/(a)^{1/n}.$$

This choice of L then gives a lower a probability tolerance limit on 100p% of the density, and this statement holds for any sample size.

Now consider the statement $\Pr(\text{ coverage} > p'\ )$ for $p' > p$. We have

$$\Pr\left[\ \int_L^{\Theta} (1/\Theta)dx > p'\right] = \Pr\left[\ Z(1 - p)/(a)^{1/n} < \Theta(1 - p')\right]$$
$$= \Pr\left[\ Z/\Theta < \frac{1 - p'}{1 - p}\,(a)^{1/n}\right]$$
$$= \left[(1 - p')/(1 - p)\right]^n a.$$

But $p' > p$ implies $(1 - p')/(1 - p) < 1$, therefore $\Pr(\text{ coverage} > p'\ )$

is a strictly decreasing function of n. Hence considering question (1) we see that the tolerance limit does improve, in terms of the proposed criterion, as the sample size increases. In fact, each additional sample element decreases Pr(coverage $> p'$) by a factor of $(1 - p')/(1 - p)$.

If we wish to determine the sample size such that Pr(coverage $> p$) $= a$ and Pr(coverage $> p'$) $\leq a'$, then we will have to choose n such that

$$\left[(1 - p')/(1 - p)\right]^n a \leq a'$$

or

$$n \geq \log(a'/a)/\log\left[(1 - p')/(1 - p)\right].$$

For example if we want to have Pr(coverage $> .90$) $= .95$ and Pr(coverage $> .92$) $\leq .10$, then we must choose n such that

$$n \geq \log(.10/.95)/\log(.08/.10) = 10.08$$

or

$$n = 11.$$

This gives

$$L = Z(.10)/(.95)^{1/11} = (.1005)Z$$

for the tolerance limit, and the above probability statements are satisfied for this limit.

So as to have for illustration a different tolerance limit to compare with the one just derived, we consider the problem of obtaining a tolerance limit based on the minimum sample value. For notation let $Z_1 = \max_i(X_i)$ and $L_1 = Z_1(1 - p)/(a)^{1/n}$ and let $Z_2 = \min_i(X_i)$. Now if we let $W = Z_2/\theta$, then the density of W is given by

$$g(w) = n(1 - w)^{n-1} \qquad \text{for } 0 < w < 1,$$
$$= 0 \qquad \text{otherwise.}$$

Cumulative points on this density are given by $\Pr\left[W < 1 - (a)^{1/n}\right] = a$. So if we choose

$$L_2 = Z_2(1 - p)/\left[1 - (a)^{1/n}\right],$$

then we will have

$$\Pr\left[\int_{L_2}^{\theta}(1/\theta)dx > p\right] = \Pr(L_2/\theta < 1 - p)$$
$$= \Pr\left[Z_2/\theta < 1 - (a)^{1/n}\right]$$
$$= a.$$

Also

$$\Pr\left[\int_{L_2}^{\theta}(1/\theta)dx > p'\right] = \Pr\left(1 - Z_2(1 - p)/\theta\left[1 - (a)^{1/n}\right] > p'\right)$$
$$= \Pr\left(Z_2/\theta < \frac{1 - p'}{1 - p}\left[1 - (a)^{1/n}\right]\right)$$
$$= \left(1 - \frac{1 - p'}{1 - p}\left[1 - (a)^{1/n}\right]\right)^n$$
$$= \left[\frac{p' - p}{1 - p} + \frac{1 - p'}{1 - p}(a)^{1/n}\right]^n.$$

Now $p' > p$, so $(p' - p)/(1 - p) > 0$, and therefore

$$\left[\frac{p' - p}{1 - p} + \frac{1 - p'}{1 - p}(a)^{1/n}\right]^n > \left[\frac{1 - p'}{1 - p}(a)^{1/n}\right]^n$$
$$= \left[(1 - p')/(1 - p)\right]^n a$$

$$= \Pr\left[\int_{L_1}^{\Theta} (1/\Theta)dx > p' \right].$$

Therefore this shows that in terms of the proposed criterion L_1 is a better tolerance limit than L_2. Another consideration concerning the relative efficiency of the two limits which could be carried out here is the comparison of the sample sizes required to obtain a given accuracy requirement.

We now consider the problem of whether $L = Z(1 - p)/(a)^{1/n}$, where $Z = \max_i(X_i)$, is the best form which can be used for a lower tolerance limit for the uniform density. "Best" as we are considering it means the uniformly most accurate tolerance limit. In terms of the density we are considering this means

$$\Pr\left[\int_{L}^{\Theta} (1/\Theta)dx > p \right] = a$$

and for any other limit L^* such that

$$\Pr\left[\int_{L^*}^{\Theta} (1/\Theta)dx > p \right] = a,$$

we have

$$\Pr\left[\int_{L}^{\Theta} (1/\Theta)dx > p' \right] \leq \Pr\left[\int_{L^*}^{\Theta} (1/\Theta)dx > p' \right]$$

for all $p' > p$. In this case we can show that $L = Z(1 - p)/(a)^{1/n}$ is the best limit in the following manner. Consider testing the hypothesis $H_0: \Theta = \Theta_0$ with alternative $H_1: \Theta > \Theta_0$ at the $1 - a$ probability level.

The uniformly most powerful test of this hypothesis is to reject H_o if $Z > c$, where $Z = \max_i(X_i)$ and c is chosen such that when $\theta = \theta_o$, $\Pr(\ Z > c\ ) = 1 - a$. The density of Z is given by

$$g(z) = n(1/\theta)(z/\theta)^{n-1} \qquad \text{for } 0 < z < \theta$$
$$= 0 \qquad \text{otherwise.}$$

Then c must be chosen such that

$$1 - a = \int_c^{\theta_o} n(1/\theta_o)(z/\theta_o)^{n-1} dz$$

$$= 1 - (c/\theta_o)^n$$

or

$$c = \theta_o(a)^{1/n}.$$

Now

$$\Pr\left[\ Z > \theta_o(a)^{1/n}\right] = \Pr\left[\ Z(1-p)/(a)^{1/n} > \theta_o(1-p)\right]$$
$$= \Pr\left[\ L > \theta_o(1-p)\right]$$
$$= 1 - \Pr\left[\int_L^{\theta_o}(1/\theta_o)dx > p\right],$$

and if L^* is any other tolerance limit such that $\Pr(\ \text{coverage} > p\ ) = a$, then

$$\Pr\left[\int_{L^*}^{\theta_o}(1/\theta)dx > p\right] = 1 - \Pr\left[\ L^* > \theta_o(1-p)\right].$$

Let $b = \theta_o(1-p)$, and the power of the test can be written in the form $\Pr_\theta(\ L > b\ )$. Then from the uniformly most powerful property of the test we have

$$\Pr(\ L > b\ ) = \Pr(\ L^* > b\ ) \quad \text{for} \quad \theta = \theta_0$$

and

$$\Pr(\ L > b\ ) > \Pr(\ L^* > b\ ) \quad \text{for} \quad \theta > \theta_0 .$$

Now if we have $p' > p$ then $b = \theta(1 - p')$ implies that $\theta > \theta_0$, so

$$\Pr\left[\int_L^\theta (1/\theta)dx > p' \right] = \Pr\left[L < \theta(1 - p') \right]$$

$$= 1 - \Pr(\ L > b\ )$$

$$< 1 - \Pr(\ L^* > b\ )$$

$$= \Pr(\ L^* < b\ )$$

$$= \Pr\left[\int_{L^*}^\theta (1/\theta)dx > p' \right] .$$

Since this inequality holds for any L^* other than L, we have that L is the uniformly most accurate lower tolerance limit.

Conclusion

This example has served to illustrate the usefulness and feasibility of the criterion proposed in this chapter, and it has shown that, in some cases at least, there do exist uniformly most accurate tolerance limits. We now need to consider the general properties of this criterion and formulate a method for obtaining a uniformly most accurate tolerance limit when one exists.

CHAPTER III

GENERAL PROPERTIES OF THE ACCURACY CRITERION

This chapter deals with the problem of determining when a uniformly most accurate tolerance limit exists, and also considers a method for obtaining this limit when it does exist. We will need to determine sufficient conditions on the density function to insure that a uniformly most accurate limit exists, and then determine what function of the sample gives this uniformly most accurate limit. The approach taken here is to obtain a relation between tolerance intervals on the density and confidence intervals on the parameter, and attempt to use the optimum properties of confidence intervals to obtain optimum tolerance intervals. The primary interest will be in one-sided tolerance limits; however, a few general remarks regarding confidence sets and tolerance regions will be made first.

Relation Between Tolerance Regions and Confidence Sets

Let $f(x;\Theta)$ be a density function and denote

$$P_{\Theta}(A) = \int_A f(x;\Theta)dx \ .$$

Let $X_1,\ldots,X_n$ be a random sample from a population with density $f(x;\Theta)$, and let $C(X_1,\ldots,X_n) = C(X)$ denote an a probability confidence set for Θ, that is, $\Pr\left[\Theta \text{ in } C(X)\right] = a$. Let $S(X_1,\ldots,X_n) = S(X)$ denote a tolerance region and suppose that $P_\theta(S(X))$ has a distribution which is independent of the parameter θ. That is, $\Pr\left[P_\theta(S(X))>p\right]$ is independent of θ. Now the coverage, $P_\theta(S(X))$, associated with $S(X)$ varies with θ. That is, for given $S(X)$, $P_\theta(S(X))$ can be considered as a function of θ, but since we have a set, $C(X)$, which we expect θ to be in, we also expect to have $P_\theta(S(X)) = P_{\theta_0}(S(X))$ for some θ_0 in $C(X)$. Therefore we will expect the true coverage to be greater than $\inf_{\theta \text{ in } C(X)} P_\theta(S(X))$. So if we can choose the form of $S(X)$ so that $\inf_{\theta \text{ in } C(X)} P_\theta(S(X))$ takes on a preassigned value, say p, then we will expect the coverage to be greater than p. To state this more precisely, we have that if $C(X)$ is such that $\Pr(\theta \text{ in } C(X)) = a$, then

$$\Pr\left[P_\theta(S(X)) > \inf_{\theta \text{ in } C(X)} P_\theta(S(X))\right] \geq \Pr(\theta \text{ in } C(X)) = a.$$

Furthermore, if $S(X)$ can be determined such that $\inf_{\theta \text{ in } C(X)} P_\theta(S(X)) = p$, then we will have

$$\Pr\left[P_\theta(S(X))>p\right] = \Pr\left[P_\theta(S(X)) > \inf_{\theta \text{ in } C(X)} P_\theta(S(X))\right]$$

$$\geq \Pr(\theta \text{ in } C(X)) = a.$$

This then gives us a relation between the probability statement of a tolerance region for the density and the probability statement of a confidence set on the parameter. Also we note that in the special case where $P_\theta(S(X)) \leq p$ for θ not in $C(X)$, then we will have

$$\Pr\left[P_\theta(S(X))>p\right] = \Pr(\theta \text{ in } C(X)) = a.$$

Here we are able to express the tolerance interval probability statement equal to a confidence interval probability statement, and we see a possible correspondence between the proposed criterion of accuracy for tolerance regions and the probability of the confidence set covering the wrong value of the parameter, say θ'. Our idea of considering $\Pr\left[P_{\theta}(S(X)) > p' \right]$, for $p' > p$, as a measure of the desireability of tolerance regions which have a fixed value for $\Pr\left[P_{\theta}(S(X)) > p \right]$ corresponds to the idea that confidence sets will be more informative the less likely they are to cover false values of the parameter with a controlled probability of covering the true value.

Since we will be primarily concerned with one-sided tolerance limits, we will now formulate these ideas for this case, give an example, and arrive at some general results.

One-Sided Tolerance Limits

Let L be a real valued function defined on the sample space, and suppose the cumulative distribution function, $F(x;\theta)$, is a continuous and decreasing function of its real parameter θ (corresponding results hold for an increasing function). Then we have

$$\int_L^{\infty} f(x;\theta)dx = 1 - F(L;\theta),$$

and $1 - F(L;\theta)$ is an increasing function of θ. Let $\underline{\theta}$ be a lower a probability confidence limit for θ, that is $\Pr(\underline{\theta} < \theta) = a$, and consider $\min_{\underline{\theta} \leq \theta} \left[1 - F(L;\theta) \right]$. Since $1 - F(L;\theta)$ is an increasing function of θ, the minimum with respect to θ, where θ is restricted to $\underline{\theta} \leq \theta$, will occur at the smallest value θ can take on in this interval, namely $\theta = \underline{\theta}$.

Therefore we have

$$\min_{\underline{\theta}\leq\theta}\left[1 - F(L;\theta)\right] = 1 - F(L;\underline{\theta}).$$

Now suppose it is possible to solve for L in the equation

$$1 - F(L;\underline{\theta}) = p.$$

Then for L determined in this manner we will have, since $1 - F(L;\theta)$ is an increasing function of θ, that $1 - F(L;\theta) > p$ for $\underline{\theta} < \theta$ and $1 - F(L;\theta) \leq p$ for $\underline{\theta} \geq \theta$. Therefore $1 - F(L;\theta) > p$ if and only if $\underline{\theta} < \theta$, so

$$\Pr\left[\int_L^\infty f(x;\theta)dx > p\right] = \Pr(\underline{\theta} < \theta) = a.$$

Note that if $1 - F(L;\theta)$ were a decreasing function of the parameter then we would have the same situation except that we would want an upper a probability confidence limit on θ.

This not only offers a possible systematic technique for arriving at tolerance limits for some densities by using a confidence limit on the parameter, but it enables us to express the tolerance limit probability statement in terms of the confidence interval probability statement. We now consider an example to illustrate how this technique can be applied to derive a tolerance limit, and then see how the properties of the confidence limit used to obtain this tolerance limit can be used to study the properties of the tolerance limit.

Example: Suppose we are interested in determining a lower tolerance limit on a normal density with unknown mean, m, and known variance, v_o^2.

Then

$$\int_L^{\infty}(2\pi v_o^2)^{-\frac{1}{2}}\exp-\tfrac{1}{2}\left[\frac{x-m}{v_o}\right]^2 dx = 1 - F(L;m)$$

is an increasing function of m, so we will need a lower confidence limit on m. Let z_a denote the point such that if Z has the standard normal density then $\Pr(Z < z_a) = a$. Then $\bar{x} - z_a v_o(n)^{-\frac{1}{2}}$ is a lower a probability confidence limit for m and

$$\min_{\bar{x} - z_a v_o(n)^{-\frac{1}{2}} \leq m}\left[1 - F(L;m)\right] = 1 - F\left[L;\bar{x} - z_a v_o(n)^{-\frac{1}{2}}\right].$$

Now if we solve for L such that

$$1 - F\left[L;\bar{x} - z_a v_o(n)^{-\frac{1}{2}}\right] = p,$$

then we will have

$$\int_L^{\infty}(2\pi v_o^2)^{-\frac{1}{2}}\exp-\tfrac{1}{2}\left[\frac{x-\bar{x}+z_a v_o(n)^{-\frac{1}{2}}}{v_o}\right]^2 dx = p$$

or

$$L = \bar{x} - z_a v_o(n)^{-\frac{1}{2}} - z_p v_o.$$

This choice for L gives a tolerance limit such that

$$\Pr(\text{ coverage} > p\ ) = \Pr\left[\bar{X} - z_a v_o(n)^{-\frac{1}{2}} < m\right] = a.$$

Continuing with this example we can illustrate how the accuracy criterion for tolerance limits corresponds to the probability of the confidence interval on the parameter covering the wrong parameter value.

For the tolerance limit derived in this example consider $\Pr(\text{ coverage} > p'\ )$, where $p' > p$. We have

$$\Pr\left[\int_L^{\infty} (2\pi v_o^2)^{-\frac{1}{2}} \exp -\tfrac{1}{2}\left[\frac{x-m}{v_o}\right]^2 dx > p' \right] = \Pr(L < m - z_{p'} v_o\ )$$

$$= \Pr\left[\bar{X} - z_a v_o (n)^{-\frac{1}{2}} - z_p v_o < m - z_{p'} v_o \right]$$

$$= \Pr\left[\bar{X} - z_a v_o (n)^{-\frac{1}{2}} < m - (z_{p'} - z_p) v_o \right]$$

$$= \Pr\left[\bar{X} - z_a v_o (n)^{-\frac{1}{2}} < m' \right]$$

where $m' = m - (z_{p'} - z_p)v_o < m$. This then states that the probability that the coverage is greater than p' is equal to the probability that the confidence interval covers a particular false value of the parameter.

Now $\bar{X} - z_a v_o (n)^{-\frac{1}{2}}$ is the uniformly most accurate lower a probability confidence limit on m. That is, if $\underline{m}$ is such that $\Pr(\ \underline{m} < m\ ) = a$ and m' is any value such that $m' < m$, then $\Pr\left[\bar{X} - z_a v_o (n)^{-\frac{1}{2}} < m'\right] \leq \Pr(\ \underline{m} < m')$. Therefore if we can show that for any given tolerance limit for this distribution such that $\Pr(\text{ coverage} > p\ )$ we can find a corresponding confidence limit on the parameter such that $\Pr(\text{ coverage} > p\ ) = \Pr(\ \underline{m}^* < m\ )$, then we may be able to use the uniformly most accurate property of the confidence limit to show that the tolerance limit we have derived is the uniformly most accurate tolerance limit. That is, suppose L^* is a given lower tolerance limit for the density $N(m, v_o)$ such that

$$\Pr\left[\int_{L^*}^{\infty} (2\pi v_o^2)^{-\frac{1}{2}} \exp -\tfrac{1}{2}\left[\frac{x-m}{v_o}\right]^2 dx > p \right] = a.$$

Then

$$a = \Pr(\ L^* - m < -z_p v_o\ ) = \Pr(\ L^* + z_p v_o < m\ ),$$

and therefore $L^* + z_p v_o$ is a lower a probability confidence limit on

m. This shows that for a given tolerance limit on this distribution, a corresponding confidence limit on the parameter can be determined such that the tolerance limit probability statement can be expressed in terms of this confidence limit probability statement. Also we have

$$\Pr\left[\int_{L^*}^{\infty}(2\pi v_o^2)^{-\frac{1}{2}}\exp-\tfrac{1}{2}\left[\frac{x-m}{v_o}\right]^2 dx > p'\right] = \Pr(L^* - m < -z_{p'}v_o)$$

$$= \Pr(L^* + z_{p'}v_o < m)$$

$$= \Pr\left[L^* + z_p v_o < m - (z_{p'} - z_p)v_o\right]$$

$$= \Pr(L^* + z_p v_o < m')$$

where, as before, $m' = m - (z_{p'} - z_p)v_o < m$. Therefore if $L = \bar{X} - z_a v_o(n)^{-\frac{1}{2}} - z_p v_o$ and L^* is any other limit such that

$$\Pr\left[\int_{L^*}^{\infty}(2\pi v_o^2)^{-\frac{1}{2}}\exp-\tfrac{1}{2}\left[\frac{x-m}{v_o}\right]^2 dx > p\right] = a,$$

then by the uniformly most accurate property of the lower confidence limit $\bar{X} - z_a v_o(n)^{-\frac{1}{2}}$, we have, for $p' > p$ and $m' = m - (z_{p'} - z_p)v_o$,

$$\Pr\left[\int_{L}^{\infty}(2\pi v_o^2)^{-\frac{1}{2}}\exp-\tfrac{1}{2}\left[\frac{x-m}{v_o}\right]^2 dx > p'\right] = \Pr\left[\bar{X} - z_a v_o(n)^{-\frac{1}{2}} < m'\right]$$

$$\leq \Pr(L^* + z_p v_o < m')$$

$$= \Pr\left[\int_{L^*}^{\infty}(2\pi v_o^2)^{-\frac{1}{2}}\exp-\tfrac{1}{2}\left[\frac{x-m}{v_o}\right]^2 dx > p'\right].$$

This means that $L = \bar{X} - z_a v_o(n)^{-\frac{1}{2}} - z_p v_o$ is the uniformly most accurate lower tolerance limit for the normal density with known variance v_o^2.

We also note that for this problem

$$\Pr\left[\bar{X} - z_a v_0(n)^{-\frac{1}{2}} < m'\right] = \Pr\left(\frac{\bar{X} - m}{v_0(n)^{-\frac{1}{2}}} < \left[\frac{m' - m}{v_0}\right](n)^{\frac{1}{2}} + z_a\right)$$

$$= \Pr\left(Z < \left[\frac{m' - m}{v_0}\right](n)^{\frac{1}{2}} + z_a\right).$$

Now $m' < m$ so $\left[\frac{m' - m}{v_0}\right](n)^{\frac{1}{2}} + z_a$ is a decreasing function of n.

Therefore

$$\Pr(\text{ coverage} > p'\ ) = \Pr\left(Z < \left[\frac{m' - m}{v_0}\right](n)^{\frac{1}{2}} + z_a\right)$$

is a decreasing function of the sample size, so the problem of determining a sample size to obtain a desired accuracy can be considered. In fact for given $p, p', a,$ and a' we have

$$a' = \Pr\left[\int_L^{\infty}(2\pi v_0^2)^{-\frac{1}{2}}\exp-\tfrac{1}{2}\left[\frac{x - m}{v_0}\right]^2 dx > p'\right]$$

$$= \Pr\left(Z < \left[\frac{m' - m}{v_0}\right](n)^{\frac{1}{2}} + z_a\right)$$

$$= \Pr\left[Z < -(z_{p'} - z_p)(n)^{\frac{1}{2}} + z_a\right]$$

which implies that

$$z_a - (z_{p'} - z_p)(n)^{\frac{1}{2}} = z_{a'}$$

or

$$n = \left(\frac{z_a - z_{a'}}{z_{p'} - z_p}\right)^2.$$

<u>General Properties:</u> Following the pattern of this example, we now show that for suitable density functions if a tolerance limit is derived

by the method described by using the uniformly most accurate confidence limit on the parameter, then the tolerance limit obtained will be the uniformly most accurate tolerance limit. In order to do this we will need to first give some definitions and theorems necessary for the proof.

The discussion will be limited to single parameter families of densities. We first define a property which gives a sufficient requirement on the density to insure that a uniformly most accurate confidence limit exists. The real parameter family of densities $p(x;\theta)$ (x may be a vector) is said to have monotone likelihood ratio if there exists a real valued function $T(x)$ such that for any $\theta < \theta'$ the distributions P_θ and $P_{\theta'}$ are distinct, and the ratio $p(x;\theta')/p(x;\theta)$ is a nondecreasing function of $T(x)$. The important requirement here is that $p(x;\theta')/p(x;\theta)$ be monotone in some real valued function $T(x)$. Nondecreasing is specified to avoid considering cases.

For reference we will quote two theorems given by Lehmann in (4) which pertain to the problem we are considering and which will be used in a following proof.

Theorem 1. Let θ be a real parameter, and let the random variable X have probability density $p(x;\theta)$ with monotone likelihood ratio in $T(x)$.

(i) For testing $H:\theta \leq \theta_0$ against $K:\theta > \theta_0$, there exists a uniformly most powerful test, which is given by

$$(1) \qquad \phi(x) = \begin{cases} 1 \text{ when } T(x) > C \\ q \text{ when } T(x) = C \\ 0 \text{ when } T(x) < C \end{cases}$$

where C and q are determined by

(2) $\quad E_{\theta_0}\phi(X) = a.$

(ii) The power function

$$\beta(\theta) = E_\theta \phi(X)$$

of this test is strictly increasing for all points θ for which $\beta(\theta) < 1$.

(iii) For all θ', the test determined by (1) and (2) is UMP for testing $H':\theta \leq \theta'$ against $K':\theta > \theta'$ at level $a' = \beta(\theta')$.

This theorem is important at this point because the concepts of uniformly most powerful test, uniformly most accurate confidence limit, and uniformly most accurate tolerance limit as we have defined it are all closely related.

We now state the theorem which gives sufficient conditions for the existence of a uniformly most accurate confidence limit.

Theorem 2. Let the family of densities $p(x;\theta)$ have monotone likelihood ratio in $T(x)$ and suppose that the cumulative distribution function $F(t;\theta)$ of $T = T(X)$ is a continuous function of t for each fixed θ in the parameter space.

(i) There exists a uniformly most accurate confidence bound $\underline{\theta}$ for θ at each confidence level $1 - a$.

(ii) If x denotes the observed values of X and $t = T(x)$, and if the equation $F(t;\theta) = 1 - a$ has a solution $\theta = \hat{\theta}$ in the parameter space, then this solution is unique and $\underline{\theta} = \hat{\theta}$.

We now state and prove the theorem which we have been leading up to concerning the existence of a best tolerance limit.

Theorem 3. Let $f(x;\theta)$ be a real parameter family of densities on the real line with monotone likelihood ratio in the real variable x. Let $F(x;\theta)$ be the cumulative distribution function of X, and assume it is continuous in x for each θ. If there exists a uniformly most accurate lower confidence limit, $\underline{\theta}$, for θ based on a sample of size n, then there exists a uniformly most accurate lower tolerance limit L for $f(x;\theta)$ given by the equation

$$F(L;\underline{\theta}) = 1 - p.$$

(Monotone likelihood ratio in x will refer to nondecreasing, however a nonincreasing likelihood ratio will produce the same results with the lower confidence limit $\underline{\theta}$ replaced by the uniformly most accurate upper confidence limit $\overline{\theta}$ on θ.)

Proof: By theorem 1 there exists a uniformly most powerful test of the hypothesis $H:\theta \leq \theta_0$ against $K:\theta > \theta_0$ and the critical region is of the form $x > C$. Theorem 1 also says that the power function is an increasing function of θ, therefore we have $Pr_\theta(X > C) < Pr_{\theta'}(X > C)$ for $\theta < \theta'$, or $1 - F(x;\theta) < 1 - F(x;\theta')$ for $\theta < \theta'$. Therefore $1 - F(x;\theta)$ is an increasing function of θ for each x. Hence if $\underline{\theta}(x_1,\ldots,x_n)$ is the uniformly most accurate lower confidence limit on θ, then we have for any x,

$$\inf_{\underline{\theta} < \theta} \left[1 - F(x;\theta)\right] = 1 - F(x;\underline{\theta}).$$

Let L be such that $1 - F(L;\underline{\theta}) = p$. L will be a real valued function of $(x_1,\ldots,x_n)$. Then we have $1 - F(L;\theta) > p$ for $\underline{\theta} < \theta$ and $1 - F(L;\theta) \leq p$ for $\underline{\theta} \geq \theta$. Therefore

$$Pr\left[\int_L^\infty f(x;\theta)dx > p \right] = Pr\left[1 - F(L;\theta) > p\right]$$

$$= \Pr(\underline{\theta} < \theta) = a,$$

so L is a lower tolerance limit on the fraction p of the density $f(x;\theta)$ with probability a.

We now need to show that for any other tolerance limit L^* such that

$$\Pr\left[\int_{L^*}^{\infty} f(x;\theta)dx > p\right] = a$$

and for every $p' > p$ we have

$$\Pr\left[\int_{L}^{\infty} f(x;\theta)dx > p'\right] \leq \Pr\left[\int_{L^*}^{\infty} f(x;\theta)dx > p'\right].$$

We will do this by finding a confidence limit $\underline{\theta}^*$ on θ corresponding to L^* and then use the uniformly most accurate property of $\underline{\theta}$ to establish the inequality.

Let L^* be a function of $(x_1,\ldots,x_n)$ such that

$$\Pr\left[\int_{L^*}^{\infty} f(x;\theta)dx > p\right] = a.$$

Then let $\underline{\theta}^*$ be such that $1 - F(L^*;\underline{\theta}^*) = p$. Now $\underline{\theta}^*$ will be a function of L^* and consequently a function of $(x_1,\ldots,x_n)$, and since $1 - F(x;\theta)$ is an increasing function of θ, we have $1 - F(L^*;\theta) > p$ for $\underline{\theta}^* < \theta$ and $1 - F(L^*;\theta) \leq p$ for $\underline{\theta}^* \geq \theta$. Therefore

$$\Pr(\underline{\theta}^* < \theta) = \Pr\left[1 - F(L^*;\theta) > p\right]$$

$$= \Pr\left[\int_{L^*}^{\infty} f(x;\theta)dx > p\right] = a.$$

So $\underline{\theta}^*$ is a lower a probability confidence limit on θ.

Now consider $p' > p$. By the UMP property of the test in theorem 1 we have that $Pr_{\theta_o}(X > C(\theta_o)) = a$, $Pr_{\theta_1}(X > C(\theta_o)) > a$ and $Pr_{\theta_1}(X > C(\theta_1)) = a$ for $\theta_o < \theta_1$ implies $C(\theta_o) < C(\theta_1)$. So $C(\theta)$ is an increasing function of θ. Also, for any value of p' there exists a C such that $Pr_{\theta_o}(X > C(\theta_o)) = p'$, and since the power is an increasing function of θ, $Pr_{\theta_1}(X > C(\theta_o)) = p$ for some $\theta_1 < \theta_o$. Note that $Pr_{\theta_1}(X > C(\theta_o)) = p$ implies $1 - F(C(\theta_o);\theta_1) = p$, which implies that θ_1 is a function of θ_o. Let $g(\theta_o)$ be this function, that is $\theta_1 = g(\theta_o)$. Now $F(C(\theta_o);\theta_1) = 1 - p$ and $C(\theta)$ an increasing function of θ implies that $g(\theta)$ is an increasing function of θ. Now consider $Pr_\theta(X > C(\theta_o))$ and $Pr_{\theta'}(X > C(\theta_o))$ where $\theta' = g(\theta)$. Let $x_o = C(\theta_o)$. Then $\theta > \theta_o$ implies $g(\theta) > g(\theta_o)$, so that $1 - F(x_o;\theta) = Pr_\theta(X > x_o) > p'$ and $1 - F(x_o;\theta') = Pr_{\theta'}(X > x_o) > p$. Also $\theta < \theta_o$ implies that $1 - F(x_o;\theta) < p'$ and $1 - F(x_o;\theta') < p$. Therefore $1 - F(x_o;\theta) > p'$ if and only if $1 - F(x_o;\theta') > p$. Now $Pr_{\theta_o}(X > C(\theta_o)) = p'$ holds for any θ_o, and consequently x_o, therefore we have that $1 - F(x;\theta) > p'$ if and only if $1 - F(x;\theta') > p$, so

$$Pr\left[\int_L^\infty f(x;\theta)dx > p' \right] = Pr\left[1 - F(L;\theta) > p' \right]$$

$$= Pr\left[1 - F(L;\theta') > p \right]$$

$$= Pr(\underline{\theta} < \theta')$$

and

$$Pr\left[\int_{L^*}^\infty f(x;\theta)dx > p' \right] = Pr\left[1 - F(L^*;\theta) > p' \right]$$

$$= Pr\left[1 - F(L^*;\theta') > p \right]$$

$$= \Pr(\underline{\theta}^* < \theta').$$

Then by the uniformly most accurate property of $\underline{\theta}$ we have

$$\Pr(\underline{\theta} < \theta') \leq \Pr(\underline{\theta}^* < \theta')$$

so

$$\Pr\left[\int_L^\infty f(x;\theta)dx > p' \right] \leq \Pr\left[\int_{L^*}^\infty f(x;\theta)dx > p' \right].$$

Therefore L is the uniformly most accurate lower tolerance limit such that $\Pr(\text{coverage} > p) = a$.

Two Sided Limits

We have been limiting the discussion to the case of one sided tolerance limits. The technique used to derive the one sided limit may also be used to derive two sided limits, although there will generally not be a uniformly most accurate two sided tolerance interval just as there is generally not a uniformly most accurate two sided confidence interval. Also the equations involved may become considerably more complicated. However two tolerance intervals can still be compared using the proposed criterion of accuracy, and the problem of determining sample size to obtain a desired accuracy can be considered. An example will be given to illustrate these points.

Consider the problem of determining a tolerance interval for a normal density with unknown mean m and known variance v_o. Then

$$\int_{L_1}^{L_2} (2\pi v_o^2)^{-\frac{1}{2}} \exp -\tfrac{1}{2}\left[\frac{x - m}{v_o}\right]^2 dx = F(L_2;m) - F(L_1;m).$$

Now $(\bar{x} - z_{a^*}v_o(n)^{-\frac{1}{2}}, \bar{x} + z_{a^*}v_o(n)^{-\frac{1}{2}})$, where $a^* = \frac{1 + a}{2}$, is an a probability confidence interval for m. Denote this interval by $(\underline{m},\overline{m})$, and we have $\min\left[F(L_2;m) - F(L_1;m)\right]$ with respect to m where m is restricted to $(\underline{m},\overline{m})$ occurs at one of the end points of the interval. If the method of expressing the tolerance interval probability statement in terms of the confidence interval probability statement is to be used, we must have L_1 and L_2 such that $F(L_2;m) - F(L_1;m) > p$ for m in $(\underline{m},\overline{m})$ and $F(L_2;m) - F(L_1;m) \leq p$ for m not in $(\underline{m},\overline{m})$. Therefore L_1 and L_2 will have to be chosen such that the equations

$$F(L_2;\underline{m}) - F(L_1;\underline{m}) = p$$

and

$$F(L_2;\overline{m}) - F(L_1;\overline{m}) = p$$

are satisfied. The first equation gives

$$(L_1 - \underline{m})/v_o = -z_{p+d}$$

and

$$(L_2 - \underline{m})/v_o = z_{1-d}.$$

Then using the symmetry of the normal density, the second equation gives

$$(L_1 - \overline{m})/v_o = -z_{1-d}$$

and

$$(L_2 - \overline{m})/v_o = z_{p+d},$$

so we will have the solution if we can determine d. We have

$$L_1 = \underline{m} - z_{p+d}v_o \quad \text{and} \quad L_1 = \overline{m} - z_{1-d}v_o,$$

therefore

$$\overline{x} - z_{a*}v_o(n)^{-\frac{1}{2}} - z_{p+d}v_o = \overline{x} + z_{a*}v_o(n)^{-\frac{1}{2}} - z_{1-d}v_o$$

or

$$z_{1-d} - z_{p+d} = 2z_{a^*}(n)^{-\frac{1}{2}}.$$

Now this equation has a unique solution in d since $z_{1-d} - z_{p+d}$ increases from 0 to ∞ as d decreases from $(1 - p)/2$ to 0. For example if we want L_1 and L_2 such that

$$\Pr\left[\int_{L_1}^{L_2} (2\pi v_o^2)^{-\frac{1}{2}} \exp -\tfrac{1}{2}\left[\frac{x - m}{v_o}\right]^2 dx > .90 \right] = .95$$

then we have $p = .90$, $a^* = .975$, so d must satisfy the equation $z_{1-d} - z_{.90+d} = 2(1.96)(n)^{-\frac{1}{2}}$. If we choose an $n = 16$, then we find that $d = .01$, so $L_1 = \bar{x} - 1.83v_o$ and $L_2 = \bar{x} + 1.83v_o$.

More Than One Unknown Parameter

Tolerance limits are generally very difficult to determine when more than one unknown parameter is involved. In order to use the previous procedure we would first need to determine an appropriate confidence region in the parameter space, then we would have to determine the parameter values which minimize $1 - F(x;\theta)$ with respect to θ where the vector θ is restricted to the confidence region. Now if L is such that $\min\left[1 - F(L;\theta)\right] = p$ where θ is restricted to the confidence region, then $1 - F(L;\theta) > p$ for θ in the confidence region, but in general $1 - F(L;\theta)$ is not less than or equal to p for all θ outside the confidence region. Therefore we would have $\Pr(\ 1 - F(L;\theta) > p\ ) \geq \Pr(\ \theta$ is in the confidence region).

CHAPTER IV

APPLICATIONS

In this chapter some applications of the preceding material will be given. These applications will be chosen for there possible usefulness and as examples to follow when studying tolerance limits for a distribution. The problem of determining the sample size necessary to obtain a desired accuracy will be considered in each case.

The Exponential

The exponential distribution will be the main distribution considered in this chapter. Two different sampling plans will be considered because of their usefulness in life testing problems. The uniformly most accurate limit will be derived and the problem of determining the sample size to obtain a desired accuracy will be solved. Also a tolerance limit based on a single order statistic will be considered and the sample size problem solved. These limits will be compared.

Let

$$f(x;\theta) = \theta^{-1}\exp{-x/\theta} \quad \text{for} \quad x > 0,\ \theta > 0,$$

$$= 0 \qquad\qquad \text{otherwise.}$$

The first method of sampling considered, stated in terms of the life testing problem, is to place n items on test, waiting until r of the n have failed without replacing those that fail as they fail, and record the times to failure of the r items. Let $x_{i,n}$ be the i-th smallest sample value, and let

$$\hat{\theta} = r^{-1}\left[\sum_{i=1}^{r} x_{i,n} + (n - r)x_{r,n}\right].$$

If $r = n$ then this is just the sample mean. It is known that $2r\hat{\theta}/\theta$ has the chi-square distribution with $2r$ degrees of freedom. We use the notation $\Pr\left[\chi^2 < \chi^2_a(2r)\right] = a$. Therefore

$$a = \Pr\left[\,2r\hat{\theta}/\theta < \chi^2_a(2r)\right]$$

$$= \Pr\left[\,2r\hat{\theta}/\chi^2_a(2r) < \theta\right],$$

so $2r\hat{\theta}/\chi^2_a(2r)$ is a lower a probability confidence limit on θ. In fact, deriving the uniformly most powerful test of the hypothesis $H: \theta \leq \theta_0$ against $K: \theta > \theta_0$ shows that this is the uniformly most accurate lower confidence limit for θ. Now

$$\int_L^{\infty} \theta^{-1} \exp -x/\theta \, dx = \exp -L/\theta$$

is an increasing function of θ, therefore

$$\min_{\underline{\theta} \leq \theta} \exp -L/\theta = \exp -L/\underline{\theta}$$

where $\underline{\theta} = 2r\hat{\theta}/\chi^2_a(2r)$. Then $\exp -L/\underline{\theta} = p$ implies

$$L = -2r\hat{\theta}\log(p)/\chi^2_a(2r).$$

Therefore, for this choice of L we have

$$\Pr\left[\int_L^{\infty} \theta^{-1} \exp - x/\theta \, dx > p \right] = \Pr\left[2r\hat{\theta}/\chi_a^2(2r) < \theta \right] = a.$$

and according to the theory in the previous chapter this choice of L gives the uniformly most accurate lower tolerance limit for this type of sampling.

Now to get a measure of the accuracy of this limit, let p' be greater than p. Then we have

$$\Pr\left[\int_L^{\infty} \theta^{-1} \exp - x/\theta \, dx > p' \right] = \Pr(L < -\theta \ln(p'))$$

$$= \Pr\left[-2r\hat{\theta}\ln(p)/\chi_a^2(2r) < -\theta\ln(p') \right]$$

$$= \Pr\left[2r\hat{\theta}/\theta < \chi_a^2(2r)\frac{\ln(p')}{\ln(p)} \right].$$

But since $2r\hat{\theta}/\theta$ has a chi-square distribution with $2r$ degrees of freedom, this probability statement is equivalent to

$$\Pr\left[\chi^2(2r) < \chi_a^2(2r)\frac{\ln(p')}{\ln(p)} \right].$$

Now $\ln(p')/\ln(p) < 1$ so this probability is a decreasing function of r. Therefore if we want to specify the sample size such that we will obtain a tolerance limit with $\Pr(\text{ coverage} > p\ ) = a$ and $\Pr(\text{ coverage} > p'\ ) \leq a'$, where $p' > p$ and a' is small, then we must choose r such that

$$\Pr\left[\chi^2(2r) < \chi_a^2(2r)\frac{\ln(p')}{\ln(p)} \right] \leq a'$$

or

$$\chi_a^2(2r)\frac{\ln(p')}{\ln(p)} \leq \chi_{a'}^2(2r).$$

The chi-square tables can be used to determine r such that this inequality is satisfied. For example, if we wish to have a lower tolerance limit on the exponential distribution such that Pr(coverage $>.90$) = .90 and also have the sample size large enough so that the accuracy will be such that Pr(coverage $>.93$) $\leq .10$, then we will have $\ln(p')/\ln(p)$ = .6888, so r must be determined such that the inequality

$$(.6888)\chi^2_{.90}(2r) \leq \chi^2_{.10}(2r)$$

is satisfied. If we look in the .90 column and the .10 column of the chi-square table and increase the degrees of freedom until the inequality is first satisfied, we find the degrees of freedom to be 96. Therefore we must have $2r \geq 96$ or $r \geq 48$. If we choose $r = 48$ then our tolerance limit will be

$$L = \frac{-2(48)\hat{\theta}\ln(.90)}{114.1} = (.0886)\hat{\theta}.$$

With this limit based on a sample of size 48 we will have Pr(coverage $>.90$) = .90, and we will also know that Pr($.93 >$ coverage $> .90$) $\geq .80$.

We note that the accuracy is not a function of n except that we must have $r \leq n$. Therefore since the accuracy increases as r increases, the best accuracy is obtained by taking $r = n$. However, in life testing problems another thing to consider is the time involved. This is where n has an effect. If n is held constant and r increased, then the accuracy of the limit improves but time to obtain the r failures increases. If r remains fixed and n is increased, then the accuracy of the limit remains the same but the time required to obtain the r failures decreases.

The second method of sampling considered, again stated in terms of the life testing problem, is to place n items on test replacing those that fail as they fail with new items until a total of r have failed. For this type of sampling the uniformly most accurate lower a probability confidence limit on θ is $2nx_{r,n}/\chi^2_a(2r)$. This quantity has the chi-square distribution with $2r$ degrees of freedom, so the tolerance limit will be the same as the one just derived except $\hat{\theta}$ will be replaced by $nr^{-1}x_{r,n}$. That is,

$$L = -2nx_{r,n}\ln(p)/\chi^2_a(2r)$$

is the uniformly most accurate lower a probability tolerance limit on $100p\%$ of the exponential density for this type of sampling. The accuracy is again independent of n but of course the time required to obtain the r failures is a function of n. Note that if we put n items on test for each method of sampling, then the second sampling method uses more items $(n + r - 1)$ to achieve the same accuracy, but requires less time.

In some of the life testing type problems it is convenient and economical to use a single order statistic when sampling without replacement to determine a tolerance limit. We will now determine a tolerance limit for the exponential based on a single order statistic and study its accuracy properties. This limit will be compared with those previously obtained.

Let $x_{r,n}$ be the r-th smallest order statistic in a random sample of size n from the exponential density. Let $w = 1 - \exp{-x_{r,n}/\theta}$.

Then the density of w is independent of θ and is given by

$$g(w) = \frac{n!}{(r-1)!(n-r)!} w^{r-1}(1-w)^{n-r} \qquad 0<w<1.$$

This is the beta density with parameters r and $n - r + 1$. We denote by $I_y(b,d)$ the cumulative beta distribution with parameters b and d evaluated at y. Let y_a denote the point such that $I_{y_a}(b,d) = a$. If we use the same technique as in the previous chapter to obtain a tolerance limit on the exponential, we will need a lower confidence limit on θ based on $x_{r,n}$ since $1 - F(x;\theta)$ for the exponential is an increasing function of θ. We have that

$$a = \Pr(\, w<y_a \,) = \Pr(\, 1 - \exp{-x_{r,n}/\theta} <y_a \,)$$

$$= \Pr\left[-\frac{x_{r,n}}{\ln(1-y_a)} < \theta \right],$$

where y_a is such that $I_{y_a}(r,n-r+1) = a$. Therefore $-\frac{x_{r,n}}{\ln(1-y_a)}$ is a lower a probability confidence limit on θ based on the r-th smallest order statistic. Then

$$\min_{\underline{\theta}\leq\theta}\left[1 - F(L;\theta)\right] = 1 - F(L;\underline{\theta})$$

where $\underline{\theta} = -\frac{x_{r,n}}{\ln(1-y_a)}$, so if we solve for L such that

$$1 - F(L;\underline{\theta}) = p,$$

then we have

$$\exp\left[-L/ \frac{x_{r,n}}{\ln(1-y_a)}\right] = p$$

or

$$L = \frac{x_{r,n}\ln(p)}{\ln(1 - y_a)} .$$

This choice of L gives a tolerance limit based on the r-th smallest order statistic such that Pr(coverage >p) = a.

Now to consider the accuracy of this tolerance limit let $p' > p$. Then we have

$$\Pr\left[\int_L^{\infty} \theta^{-1}\exp{-x/\theta}\, dx > p' \right] = \Pr(-L/\theta > \ln(p'))$$

$$= \Pr\left[-\frac{x_{r,n}\ln(p)}{\theta\ln(1 - y_a)} > \ln(p') \right]$$

$$= \Pr\left(\exp{-x_{r,n}/\theta} > \exp\left[\frac{\ln(p')\ln(1 - y_a)}{\ln(p)}\right] \right)$$

$$= \Pr\left[w < 1 - (1 - y_a)^{\frac{\ln(p')}{\ln(p)}} \right].$$

Then if we wish to determine the sample size such that Pr(coverage >p') $\leq a'$, we must have

$$\Pr\left[w < 1 - (1 - y_a)^{\frac{\ln(p')}{\ln(p)}} \right] \leq a'$$

or

$$1 - (1 - y_a)^{\frac{\ln(p')}{\ln(p)}} \leq y_{a'}$$

or

$$\ln(1 - y_{a'}) \leq \frac{\ln(p')}{\ln(p)} \ln(1 - y_a).$$

Whether or not this inequality can be satisfied depends on both r and n. There are three possibilities for obtaining this inequality.

(1) Let n be fixed (and sufficiently large) and determine r such that the inequality is satisfied.

(2) Let r be fixed and determine n such that the inequality is satisfied.

(3) Let there be a functional relation between r and n such that when one is known the other is also known, then determine n such that the inequality is satisfied.

Tables of the Incomplete Beta-Function, by Karl Pearson can be used to perform the computations.

Table I on the following page is given so that a rough comparison of the accuracy of the tolerance limit based on Θ and the accuracy based on $x_{r,n}$ can be made. It can also be used to get an approximate idea of what sample size should be taken to achieve a desired degree of accuracy.

Table I

The lower tolerance limits for the exponential density are determined such that they satisfy the statement, Pr(coverage$>$.90) = .90. Entries in the table are the computed values of Pr(coverage$>$.93). The tolerance limits used are

$$L_1 = -2r\theta\ln(p)/\chi^2_a(2r) \quad \text{and} \quad L_2 = x_{r,n}\ln(p)/\ln(1 - y_a)$$

where

p = .90 and a = .90, and θ, $x_{r,n}$, and y_a are as defined previously.

r	based on L_1	based on L_2, n = 20	30	40	50
10	.52	.52	.51	.51	.51
20	.34		.37	.36	.33
30	.22			.26	.23
40	.13				.19
50	.09				

No interpolation techniques were used in computing those values based on L_2, so the values are only approximate. When r = n the incomplete beta tables are not adequate for computing even an approximate value.

The Weibull

As another application we consider another density which is used in life testing problems. This density is called the Weibull density and its function form is

$$f(x;b,c) = cb^{-c}x^{c-1}\exp-(x/b)^c \quad \text{for} \quad x,b,c>0.$$

This density was first proposed by Weibull in (6). The difficulty in using this density as a model is usually the problem of determining a good value for c. Methods of estimating c are discussed and referenced by Qureishi in (5). It is suggested that this model is often superior to the exponential even if a "rough" value has to be used for c. Therefore we will look at the accuracy of a tolerance limit for this density assuming that c is known and b is unknown. The cumulative distribution is

$$F(x;b) = \int_0^x cb^{-c}x^{c-1}\exp-(t/b)^c dt = 1 - \exp-(x/b)^c,$$

so

$$1 - F(x;b) = \exp-(x/b)^c$$

is an increasing function of b. Also we have that for

$$y = b^{-c}\sum_{i=1}^{n} x_i^c ,$$

the density of y is given by

$$g(y) = \frac{1}{\Gamma(n)} y^{n-1}\exp-y .$$

This is a gamma density and is independent of the parameter and can be used to obtain a confidence limit on b. We note that 2y is distributed as a chi-square with 2n degrees of freedom, and since chi-square tables are more accessible than gamma tables we will use the distribution of 2y to obtain a tolerance limit. We have

$$a = \Pr\left(2y < \chi_a^2(2n) \right)$$

$$= \Pr\left[2b^{-c}\sum_{i=1}^{n} x_i^c < \chi_a^2(2n) \right]$$

$$= \Pr\left(\left\{ \frac{2\sum x_i^c}{\chi_a^2(2n)} \right\}^{1/c} < b \right).$$

Therefore $\underline{b} = \left[2\sum_{i=1}^{n} x_i^c / \chi_a^2(2n) \right]^{1/c}$ is a lower a probability confidence limit on b. Now since $1 - F(x;b)$ is an increasing function of b, we have

$$\min_{\underline{b} \leq b} \left[1 - F(x;b) \right] = 1 - F(x;\underline{b}),$$

so if we solve for L such that $1 - F(L;\underline{b}) = p$, we have

$$\exp-(L/\underline{b})^c = p$$

implies

$$L = \underline{b}(-\ln p)^{1/c}$$

implies

$$L = \left[-2\ln(p)\sum_{i=1}^{n} x_i^c / \chi_a^2(2n) \right]^{1/c}.$$

This choice of L gives a lower tolerance limit on the Weibull where c is assumed known such that Pr(coverage >p) = a.

$\underline{b} = \left[2\sum_{i=1}^{n} x_i^c / \chi_a^2(2n) \right]^{1/c}$ is a uniformly most accurate lower confidence limit on b for known c, so L is the uniformly most accurate lower tolerance limit.

Now consider Pr(coverage $>$ p') for $p' > p$. We have

$$\Pr\left[\int_L^{\infty} cb^{-c}x^{c-1}\exp-(x/b)^c dx > p'\right] = \Pr\left[\exp-(L/b)^c > p'\right]$$

$$= \Pr\left[(L/b)^c < -\ln(p')\right]$$

$$= \Pr\left[-2\ln(p)\sum_{i=1}^{n} x_i^c \,/b^c \chi_a^2(2n) < -\ln(p')\right]$$

$$= \Pr\left[2b^{-c}\sum_{i=1}^{n} x_i^c < \frac{\ln(p')}{\ln(p)}\chi_a^2(2n)\right]$$

$$= \Pr\left[\chi^2(2n) < \frac{\ln(p')}{\ln(p)}\chi_a^2(2n)\right] .$$

This is exactly the same quantity as we obtained in the case of the exponential. We conclude then that the accuracy of this limit and that for the exponential based on $\bar{x}$ is the same. The computation can be carried out as described in the exponential case.

CHAPTER V

ACCURACY OF A CONFIDENCE LIMIT ON A POINT OF THE CUMULATIVE

Instead of the usual tolerance limit problem, that is determining a point such that a specified portion of the density is greater than this point, it is often of interest to determine what fraction of a density is greater that a given point. For example we may want to know what fraction of the tubes produced by a particular process will last more than 50 hours. Now instead of just estimating this fraction, it may be more desirable to be able to say with high confidence that at least a certain fraction will last more than 50 hours. This amounts to determining a lower confidence limit on the fraction of the density greater than 50.

Let $f(x;\theta)$ be the density under consideration and let $p(X)$ be a statistic. The problem is then to determine $p(X)$ such that for given x_o and a we have

$$\Pr\left[\int_{x_o}^{\infty} f(x;\theta)dx > p(X)\right] = a,$$

For example, consider the uniform density on the interval $(0,\theta)$. Then for given x_o and a we want to determine $p(X)$ such that

$$\Pr\left[\int_{x_o}^{\theta} \theta^{-1}dx > p(X)\right] = a.$$

Therefore we have

$$a = \Pr\left[1 - x_o\theta^{-1} > p(X)\right]$$

$$= \Pr\left[x_o/(1 - p(X)) < \theta\right].$$

Now in Chapter II we determined that $Z/(a)^{1/n}$, where $Z = \max_i(X_i)$, is a lower a probability confidence limit on θ. So if we choose $x_o/(1 - p(X)) = Z/(a)^{1/n}$, then the probability statement will be satisfied. This means that $p(X) = 1 - x_o(a)^{1/n}/Z$. Similarly we can show that $p^*(X) = 1 - x_o(1 - (a)^{1/n})/Y$, where $Y = \min_i(X_i)$, satisfies the above probability statement.

Now we would like to do for this problem something similar to what was done for the standard tolerance limit problem, that is

(1) formulate a systematic method for obtaining $p(X)$,

(2) formulate a criterion of accuracy which will be meaningful in terms of the physical problem,

(3) determine the sample size necessary to obtain a desired accuracy, and

(4) determine best $p(X)$ functions in terms of this criterion.

$p(X)$ can be obtained by using a confidence limit on the parameter in the same manner as the tolerance limit was obtained previously. Suppose $1 - F(x;\theta)$ is an increasing function of θ and let $\underline{\theta}(X)$ be a lower confidence limit on θ such that $\Pr(\underline{\theta} < \theta) = a$. Then

$$\min_{\underline{\theta} \leq \theta}\left[1 - F(x_o;\theta)\right] = 1 - F(x_o;\underline{\theta}).$$

Therefore let $p(X) = 1 - F(x_o;\underline{\theta})$ and since $1 - F(x_o;\theta) > 1 - F(x_o;\underline{\theta})$ if and only if $\underline{\theta} < \theta$, we have

$$\Pr\left[\int_{x_0}^{\infty} f(x;\theta)dx > p(X)\right] = \Pr(\underline{\theta} < \theta) = a.$$

Of course if $1 - F(x;\theta)$ is a decreasing function of θ then we need an upper confidence limit for θ, and the same procedure follows.

Now consider the distribution of $p(X)$. The range of $p(X)$ is between 0 and 1 and it has probability a of being less than $1 - F(x_0;\theta)$ for the true value of θ. Therefore the density might look like that in Figure (2).

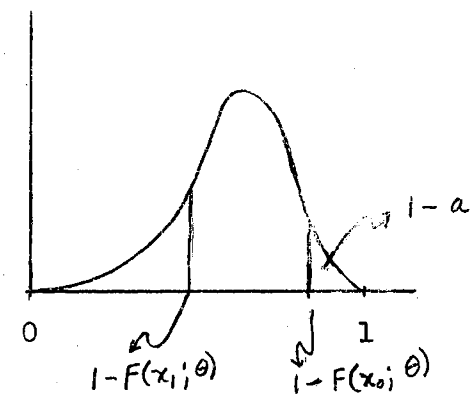

Figure (2)

We want $p(X)$ to be less than $1 - F(x_0;\theta)$ with probability a, but for accuracy we want it to be close to $1 - F(x_0;\theta)$. As a measure of this accuracy we might choose an x_1 such that $x_1 > x_0$ and determine what fraction of the density of $p(X)$ is between $1 - F(x_1;\theta)$ and $1 - F(x_0;\theta)$. For a $p(X)$ such that

$$\Pr\left[1 - F(x_0;\theta) > p(X)\right] = a,$$

and for $x_1 > x_0$, we can look at

$$\Pr\left[1 - F(x_1;\theta) > p(X)\right]$$

as the measure of accuracy. The smaller this quantity is the closer

$p(X)$ is likely to be to $1 - F(x_o;\theta)$. If we want to specify a value for this expression, then the sample size can be determined to achieve this.

Continuing with our example we have for $x_1 > x_o$

$$\Pr\left[\int_{x_1}^{\theta}\theta^{-1}dx > p(X)\right] = \Pr\left[1 - x_1\theta^{-1} > 1 - x_o(a)^{1/n}Z^{-1}\right]$$

$$= \Pr\left[x_1\theta^{-1} < x_o(a)^{1/n}Z^{-1}\right]$$

$$= \Pr\left[Z\theta^{-1} < x_o(a)^{1/n}x_1^{-1}\right]$$

$$= (x_o/x_1)^n a.$$

This is a decreasing function of n and for given a', an n can be determined such that $(x_o/x_1)^n a \leq a'$. We might also note that for $p^*(X)$ we have

$$\Pr\left[\int_{x_1}^{\theta}\theta^{-1}dx > p^*(X)\right] = \Pr\left[1 - x_1\theta^{-1} > 1 - x_o(1 - (a)^{1/n})Y^{-1}\right]$$

$$= \Pr\left[Y\theta^{-1} < (x_o/x_1)(1 - (a)^{1/n})\right]$$

$$= \left[1 - (x_o/x_1)(1 - (a)^{1/n})\right]^n$$

which is greater than $(x_o/x_1)^n a$. Therefore we would consider $p(X)$ better than $p^*(X)$.

In terms of the accuracy measure we have formulated, the best function $p(X)$ to use is the one which has the property of being the uniformly most accurate lower confidence limit on $1 - F(x_o;\theta)$. It seems reasonable that if we use a uniformly most accurate confidence limit on θ to derive $p(X)$ by the method given, then we will get the best function for $p(X)$. This is indeed the case in the special cases being considered here.

Theorem 4. Let $1 - F(x_0;\theta)$ be continuous in x and continuous and monotonic in θ (assume increasing for explicitness), and assume there exists a uniformly most accurate lower confidence limit $\underline{\theta}$ for θ, such that $\Pr(\underline{\theta} < \theta) = a$. If $p(X)$ is chosen such that

$$\min_{\underline{\theta} \leq \theta} \left[1 - F(x_0;\theta)\right] = p(X),$$

then

$$\Pr\left[1 - F(x_0;\theta) > p(X)\right] = \Pr(\underline{\theta} < \theta) = a,$$

and $p(X)$ is the uniformly most accurate lower a probability confidence limit for $1 - F(x_0;\theta)$.

Proof: Suppose we choose $x_1 > x_0$. Then $1 - F(x_1;\theta) < 1 - F(x_0;\theta)$, but since $1 - F(x;\theta)$ is monotone increasing in θ, there exists a θ' (some function of θ) such that $\theta' < \theta$ and $1 - F(x_1;\theta) = 1 - F(x_0;\theta')$. Therefore we have

$$\Pr\left[1 - F(x_1;\theta) > p(X)\right] = \Pr\left[1 - F(x_0;\theta') > p(X)\right]$$

$$= \Pr(\underline{\theta} < \theta').$$

But since $\underline{\theta}$ is the uniformly most accurate confidence limit for θ, $\Pr(\underline{\theta} < \theta')$ is a minimum. This means $\Pr\left[1 - F(x_1;\theta) > p(X)\right]$ is a minimum, so $p(X)$ is the uniformly most accurate lower a probability confidence limit on $1 - F(x_0;\theta)$.

Exponential

We will now use the preceding results in this chapter to solve the problem of setting a confidence limit on a point of the cumulative of the exponential which will satisfy specified accuracy criterion.

Let

$$f(x;\theta) = \theta^{-1}\exp -x/\theta \qquad \text{for} \quad x>0, \quad \theta>0$$

$$= 0 \qquad\qquad \text{otherwise.}$$

If $X_1,\ldots,X_n$ is a random sample from this density, then $2n\bar{x}/\theta$ has a chi-square distribution with $2n$ degrees of freedom, and $2n\bar{x}/\chi_a^2(2n)$ is a lower a probability confidence limit for θ. Also

$$\int_{x_o}^{\infty}\theta^{-1}\exp -x/\theta \, dx = 1 - F(x_o;\theta)$$

is an increasing function of θ, therefore using the method described previously, we have

$$\min_{\underline{\theta}\leq\theta} 1 - F(x_o;\theta) = 1 - F(x_o;\underline{\theta})$$

$$= \exp\left[-\frac{x_o\chi_a^2(2n)}{2n\bar{x}}\right].$$

Therefore, if we choose $p(X) = \exp\left[-x_o\chi_a^2(2n)/2n\bar{x}\right]$, $p(X)$ will be the uniformly most accurate lower a probability confidence limit for $1 - F(x_o;\theta)$ since $2n\bar{x}/\chi_a^2(2n)$ is the uniformly most accurate lower a probability confidence limit for θ. So we have the best $p(X)$ such that

$$\Pr\left[\int_{x_o}^{\infty}\theta^{-1}\exp -x/\theta \, dx > p(X)\right] = a.$$

Now for given x_1 such that $x_1 > x_o$ we have

$$\Pr\left[\int_{x_1}^{\infty}\theta^{-1}\exp -x/\theta \, dx > p(X)\right] = \Pr\left(\int_{x_1}^{\infty}\theta^{-1}\exp -x/\theta \, dx > \exp\left[-x_o\chi_a^2(2n)/2n\bar{x}\right]\right)$$

$$= \Pr\left(\exp -x_1/\theta > \exp\left[-x_o \chi^2_a(2n)/2n\bar{x}\right]\right)$$

$$= \Pr\left[2n\bar{x}/\theta < (x_o/x_1)\chi^2_a(2n) \right] .$$

But $2n\bar{x}/\theta$ is distributed as a chi-square variate with $2n$ degrees of freedom. Therefore if we want to determine n such that

$$\Pr\left[1 - F(x_1;\theta) > p(X) \right] \le a',$$

we will have to have

$$\Pr\left[2n\bar{x}/\theta < (x_o/x_1)\chi^2_a(2n) \right] \le a'$$

or

$$(x_o/x_1)\chi^2_a(2n) \le \chi^2_{a'}(2n).$$

Meaningful values of x_o and x_1 depend on the nature of the problem. For example, suppose a company produces a certain type of tube and a customer is interested in buying a large supply of these tubes. However the customer wants a guaranty as to what fraction of the tubes will last 40 hours or more. The company will need to know what fraction, say p, of the tubes produced will last 40 hours or more. p will likely be unknown, so an estimate must be used. The company will want this estimate to be less than p, that is they will want to be able to say with a high degree of confidence that at least a certain fraction of these tubes will last 40 hours or more. Therefore a $p(X)$ is needed such that $\Pr(p > p(X))$ is large, say .90. However, the company will not want to underrate its product by giving an estimate which may be considerably less than p, since it might loose the order as a result. The accuracy of $p(X)$ can be controlled then, by choosing another value, say 60, where p_1 is the fraction of the tubes produced by this process which last longer than 60 hours, and consider the probability that $p(X)$ fraction of the tubes actually last longer than 60 hours. That is, if $p(X)$ is less than p_1, the company is underrating its product by saying that at least $p(X)$ fraction of the tubes will last longer than 40 hours when in reality at least $p(X)$ fraction of the tubes will last longer than 60 hours. To control this kind of an error we can determine $p(X)$ such that $Pr(\ p > p(X)) = .90$, then determine what sample size should be taken so that we will have $Pr(\ p_1 > p(X)\ ) \leq .10$ (or some smaller value if more accuracy is desired). If we assume the the exponential distribution then we will have

$$p(X) = \exp\left[-40\chi^2_{.90}(2n)/2n\bar{x}\right]$$

and n must be determined such that the inequality

$$(40/60)\chi^2_{.90}(2n) \leq \chi^2_{.10}(2n)$$

is satisfied. Using the chi-square tables we see that the first degree of freedom row for which this is satisfied is 81. Therefore we need $2n \geq 81$, so we would choose a sample of size 41.

CHAPTER VI

SUMMARY

In this thesis a study is made of the accuracy of tolerance limits for densities which satisfy the statement $\Pr(\text{ coverage} > p\ ) = a$. A discussion is given on measures of accuracy which have been considered and the merits of these measures are discussed. It is pointed out that none of these measures is satisfactory for considering both the comparison of two different functional forms for a tolerance limit and the determination of a sample size to obtain a desired degree of accuracy.

For a tolerance limit for a density which satisfies the statement $\Pr(\text{ coverage} > p\ ) = a$, it is proposed that $\Pr(\text{ coverage} > p'\ )$, where $p' > p$, be used as a measure of the accuracy of the limit. The feasibility of this criterion for use in studying the above stated problems is discussed, and a limit is defined as uniformly most accurate if $\Pr(\text{ coverage} > p'\ )$ is a minimum for all $p' > p$. The uniform density is used as an example and the problem of determining a sample size necessary to obtain a desired degree of accuracy is solved. That is, a tolerance limit is determined such that $\Pr(\text{ coverage} > p\ ) = a$, then the sample size is determined so that for given p' and a' the statement $\Pr(\text{ coverage} > p'\ ) \leq a'$ will also be satisfied.

Let $f(x;\theta)$ be a density function and suppose the cumulative density function $F(x;\theta)$ of X is a decreasing function of θ for each x. Then if $\underline{\theta}$ is a lower a probability confidence limit on θ, we have

$$\min_{\underline{\theta}\leq\theta}\left[1 - F(x;\theta)\right] = 1 - F(x;\underline{\theta}).$$

Now if L is determined such that $1 - F(L;\underline{\theta}) = p$, then we have that $1 - F(L;\theta) > p$ if and only if $\underline{\theta} < \theta$. Therefore

$$\Pr\left[\int_L^{\infty} f(x;\theta)dx > p\right] = \Pr\left[1 - F(L;\theta) > p\right]$$

$$= \Pr(\underline{\theta} < \theta) = a.$$

This gives a useful relation between the probability statement of a tolerance limit on the density and the probability statement of a confidence limit on the parameter. Conditions on $f(x;\theta)$ are given so that if $\underline{\theta}$ is the uniformly most accurate confidence limit on θ and L is determined by the above relation, then L will be the uniformly most accurate tolerance limit on the fraction p of the density $f(x;\theta)$.

Some applications of the above techniques are given. For the exponential distribution, the uniformly most accurate lower tolerance limit is derived for each of two different sampling schemes. The problem of determining the accuracy of these limits is solved. Also a tolerance limit based on a single order statistic is derived and this limit is compared with the uniformly most accurate limit.

The problem of determining a lower confidence limit on the fraction of a density greater than a given value is briefly considered. It is shown that this confidence limit can be determined by using a confidence limit on the parameter similar to the technique given for the standard tolerance limit problem. A measure of precision is defined for this confidence limit and the problem of determining the sample size necessary to achieve a desired degree of accuracy is solved for the exponential distribution.

SELECTED BIBLIOGRAPHY

(1) Bain, L. "Tolerance Limits for the Exponential Distribution," unpublished M.S.thesis, Oklahoma State University, 1962.

(2) Fraser, D. A. S., and Guttman, I. "Tolerance Regions," Annals of Mathematical Statistics, Volume 27 (1956):162-179.

(3) Goodman, L. A., and Madansky, A. "Parametric-Free and Nonparametric Tolerance Limits; The Exponential Case," Technometrics, IV (February, 1962):75-95.

(4) Lehmann, E. L. Testing Statistical Hypotheses, John Wiley and Sons, New York, 1959.

(5) Qureishi, A. S. "The Discrimination Between Two Weibull Processes," Technometrics, VI (February, 1964):57-75.

(6) Weibull, W. "A Statistical Distribution of Wide Applicability," Journal of Applied Mechanics, Volume 18 (1951):293-297.

VITA

Gerald David Faulkenberry

Candidate for the Degree of

Doctor of Philosophy

Thesis: UNIFORMLY MOST ACCURATE TOLERANCE LIMITS

Major Field: Mathematics and Statistics

Biographical:

Personal Data: Born in Wapanucka, Oklahoma, May 28, 1937, the son of Elmer Lee and Susie Geraldine Faulkenberry.

Education: Attended grade school in Hartshorne, Oklahoma; attended high school in Hartshorne, Oklahoma and Cincinnati, Ohio; graduated from Western Hills High School, Cincinnati, Ohio in 1955; recieved the Bachelor of Science degree from Southeastern State College, Durant, Oklahoma, with a major in Mathematics, in May, 1959; recieved the Master of Science degree from Oklahoma State University, Stillwater, Oklahoma, with a major in Mathematics, in May, 1962; completed requirements for the Doctor of Philosophy degree in May, 1965.

Professional Experience: Graduate Assistant in the Department of Mathematics and Statistics, Oklahoma State University, Stillwater, Oklahoma, 1959-1965; worked as staff assistant in a National Science Foundation sponsored institute for high school teachers at Southeastern State College, Durant, Oklahoma, the summer of 1960; taught Mathematics the summer of 1961 at Southeastern State College, Durant, Oklahoma; worked as Operations Analyst at General Dynamics, Fort Worth, Texas, during the summer of 1963.