TOLERANCE LIMITS FOR THE EXPONENTIAL DISTRIBUTION

By

LEE J. BAIN

Bachelor of Science
Oklahoma State University
Stillwater, Oklahoma
1960

Submitted to the faculty of the Graduate School of
the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
MASTER OF SCIENCE
May, 1962

Thesis
1962
B162t
copy 2

OKLAHOMA
STATE UNIVERSITY
LIBRARY

NOV 6 1962

TOLERANCE LIMITS FOR THE EXPONENTIAL DISTRIBUTION

Thesis Approved:

David L. Weeks
Thesis Adviser

Carl E. Marshall

Robert MacVicar
Dean of the Graduate School

504256

PREFACE

In the past few years, parametric and non-parametric tolerance limits have been derived for many of the common density functions. This thesis is concerned with various aspects of tolerance limits associated with the exponential distribution. A problem associated with specification limits for the exponential distribution is also considerd.

The author is deeply indebted to Dr. David L. Weeks for suggesting the problem and for his guidance in the preparation of this thesis. An expression of gratitude is extended to Professor Carl E. Marshall for serving on the author's advisory committee, and to Miss Jane Duncan for typing this paper.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

A problem often encountered in industrial applications of statistics is that of determining if the output of some method of production is meeting certain specified standards. One situation is when these standards are fixed limits which must be met, and a sample of the output is taken to determine whether a prescribed proportion of the product is meeting these fixed standards. Fixed limits such as these are known as specification limits.

Another approach is to take a sample of the product and determine variable limits as functions of the sample observations which estimate standards that, according to some criterion, contain a certain proportion of the product. One common criterion is to choose functions such that the expected proportion of the population between the variable limits is greater than or equal to a prescribed amount. Another criterion is to choose limits which contain a certain proportion of the population with a specified probability. The criterion used determines what functions of the sample observations must be used as the limits. All variable limits such as these are known as tolerance limits.

This thesis develops the two types of tolerance limits mentioned above where the parent population is assumed to follow an exponential distribution given by

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & 0 \leq x < \infty \\ 0 & , \text{otherwise} \end{cases} .$$

If the unknown parameter θ were known, exact limits as a function of θ could be determined which would contain 100β per cent, say, of the population. Since θ is not known, it must be estimated and functions of the estimator used as tolerance limits.

There are several different useful estimators of θ commonly associated with the exponential distribution, and tolerance limits can be developed as functions of each of these estimators. For this reason, a summary of those methods of estimation commonly used and the distributional properties of the respective estimators are given first. Tolerance limits are then developed based on each of these estimators. The final chapter gives a solution for the problem associated with specification limits for the exponential distribution.

CHAPTER II

ESTIMATION PROCEDURES

Most problems associated with the exponential distribution are concerned with determining the length of life of some product. For example, a light bulb manufacturer may wish to estimate the mean life of his light bulbs. If the life of light bulbs is assumed to follow an exponential distribution, then it will be necessary to estimate the mean, θ, of the exponential distribution. A random sample of light bulbs would be chosen and placed on test, and the life of these light bulbs would be used to estimate θ. A special property of life testing problems of this type is that the sample observations will be ordered according to size as they become available. Chiefly due to this property, there are several different methods of estimating θ, each of which has certain advantages over the others. The most desirable method depends on the particular problem in question.

If a random sample of n items is chosen and placed on life test, and if the test is continued until all items have failed, then

$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the maximum likelihood estimate of θ, where x_i is the life of the i-th item on test. It is the "best" estimator of θ based on these n observations in the sense that it is an unbiased, minimum variance, efficient estimator of the unknown parameter θ (1).

The distribution of $\overline{x}$ can be derived by use of moment generating functions.

Let

$$M_x(t) = \int_0^\infty e^{tx} \frac{1}{\theta} e^{-x/\theta} dx$$

$$= \frac{1}{\theta} \int_0^\infty e^{-x(\frac{1}{\theta} - t)} dx$$

$$= \frac{1}{(1 - \theta t)} \int_0^\infty (\frac{1}{\theta} - t)\, e^{-x(\frac{1}{\theta} - t)}\, dx$$

$$= \frac{1}{(1 - \theta t)} \quad .$$

$$M_{\overline{x}}(t) = M_{\Sigma x_i}(t/n)$$

$$= (1 - \frac{t}{n}\theta)^{-n} \quad .$$

This is the moment generating function of a gamma density function with parameters $\beta = \frac{\theta}{n}$ and $\alpha = n - 1$, thus the density of $\overline{x}$ is

$$h(\overline{x};\, \theta) = \frac{1}{(n - 1)!} \left(\frac{n}{\theta}\right)^n \overline{x}^{\,n-1} e^{\frac{-n\overline{x}}{\theta}} , \quad 0 \leq \overline{x} < \infty ,$$

and $\frac{2n\overline{x}}{\theta}$ is distributed as a chi-square variable with $2n$ degrees of freedom. (2). Also, $E(\overline{x}) = \beta(\alpha + 1) = \theta$, and $Var(\overline{x}) = \beta^2(\alpha + 1) = \frac{\theta^2}{n}$.

Although $\overline{x}$ has optimum properties for the case mentioned above, it seems reasonable that some advantages may be gained by changing the sampling design, since $\overline{x}$ does not utilize the information given by the property that the i-th sample observation is the i-th order statistic. One method of utilizing this information is by using censored sampling. An example of a sample censored on the right is a life test experiment begun on n items and terminated after the r-th failure, where $r < n$. A sample may be similarly censored either on the left or doubly censored.

It has been shown for a sample censored on the right that

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

is the maximum likelihood, unbiased, minimum variance, and efficient estimator of θ, where r failures of n items placed on test are observed, and where $x_{i,n}$ denotes the time of the i-th failure for the n items on test. (3). The density function of $y = \hat{\theta}_{r,n}$ is given by

$$f_r(y) = \frac{1}{(r-1)!}\left(\frac{r}{\theta}\right)^r y^{r-1} e^{-\frac{ry}{\theta}}, \quad y > 0 \quad .$$

This is exactly the distribution of $\bar{x}$ for a sample of size r, where all r failures are observed. Thus, using $\hat{\theta}_{r,n}$ from a censored sampling procedure yields exactly the same precision and optimum properties for estimating θ as using $\bar{x}$ from a non-censored sample where only r items are placed on test.

The disadvantage of using a censored sample is that n-r more items have been placed on test for a time $x_{r,n}$, than if only r items are selected originally. The importance of this loss depends on the particular problem in question. The advantage of censored sampling is that much time may be saved, since the expected time for r out of n randomly selected items to fail is less than the expected time for r out of r randomly selected items to fail. The expected time of the r-th failure from a sample of n items is $E(x_{r,n})$. Thus, a measure of the time saved by censored sampling is $\alpha_{r,n} = \dfrac{E(x_{r,n})}{E(x_{r,r})}$.

The distribution of $x_{r,n}$, the r-th order statistic, is given by

$$f_{r,n}(x) = \frac{n!}{(r-1)!(n-r)!} [F_x(x)]^{r-1} f(x;\theta)[1 - F_x(x)]^{n-r} ,$$

where $F_x(x) = 1 - e^{-x/\theta}$ denotes the cumulative distribution function of the exponential density $f(x; \theta)$. (2). So,

$$f_{r,n}(x) = \frac{n!}{(r-1)!(n-r)!} \frac{1}{\theta} [1 - e^{-x/\theta}]^{r-1}[e^{-x/\theta}]^{n-r+1}.$$

Integration by parts yields

$$E(x_{r,n}) = \theta \sum_{i=1}^{r} a_i = \theta d \qquad \text{and}$$

$$\mathrm{Var}(x_{r,n}) = \theta^2 \sum_{i=1}^{r} a_i^2 ,$$

where $a_i = (n - i + 1)^{-1}$ and $d = \sum_{i=1}^{r} a_i$.

Thus

$$\alpha_{r,n} = \frac{\sum^{r} (n - i + 1)^{-1}}{\sum^{r} (r - i + 1)^{-1}}$$

is the ratio of the expected time for r out of n items to fail compared to the expected time for r out of r items to fail, and it is always less than 1 if $r < n$. Tables of values of $\alpha_{r,n}$ for small values of r and n have been computed. (4).

Sampling with replacement has also been considered for the exponential distribution. (4). That is, n randomly selected items are placed on life test and each time an item fails, it is immediately replaced by a new item selected at random. In this case the maximum likelihood estimator of θ is given by $\hat{\theta}'_{r,n} = \frac{n}{r} x'_{r,n}$, where the test is continued until r units have failed and $x'_{r,n}$ is the time of the r-th failure. $\hat{\theta}'_{r,n}$ has exactly the same distribution and optimum properties as $\hat{\theta}_{r,n}$. A measure of the proportion of the expected time saved by censored sampling compared to sampling without censoring when sampling with replacement is being used is

$$\hat{\alpha}'_{r,n} = \frac{E(x'_{r,n})}{E(x'_{r,r})} = \frac{r}{n} .$$

Clearly, the expected time for r failures will decrease when sampling with replacement compared to without replacement, since some of the replacement items can fail before r of the original n items have failed. Thus, sampling with replacement may be desirable in some cases.

The best estimator of θ based on the first r observations from a sample of size n, without replacement, was given before by

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

Since the last observation, $x_{r,n}$, is weighted by a factor of $n - r + 1$, it can be used alone to estimate θ and the first $r-1$ observations disregarded, if r is sufficiently less than n. Epstein (5) indicates that in general for $r \leq 2/3n$, the efficiency of using $x_{r,n}$ compared to $\hat{\theta}_{r,n}$ is greater than or equal to .90. For example, for $r = \frac{n}{2}$ with n even, the efficiency is greater than or equal to $2(\log 2)^2 = .9608$. Epstein has computed tables of efficiencies and unbiasing constants for $n = 1(1)\ 20\ (5)\ 30\ (10)\ 100$ and $r = 1\ (1)\ n$.

One can also consider which single ordered observation of the n sample observations would yield the most efficient estimator of θ compared to $\overline{x}$, for a given n. Harter (6) gives the best single order statistic $x_{k,n}$ to use to estimate θ for $n = 1\ (1)\ 100$. He also gives the unbiasing constant for the appropriate $x_{k,n}$ and the relative efficiency of that estimator compared to $\overline{x}$. The relative efficiency is given by

$$V_k = \frac{\left(\sum_{i}^{k} a_i \right)^2}{n \sum_{i}^{k} a_i^2} \ ,$$

where $a_i = (n - i + 1)^{-1}$. Hence, the proper $x_{k,n}$ was selected to minimize V_k by Harter. He also gives the best combination of pairs of order statistics to use to estimate θ for different values of n.

Thus, an experimenter can consider the different methods of sampling and estimation discussed above and determine the procedure most appropriate for his particular problem. These statistics and their properties can then be used to set tolerance limits and solve specification limit problems as shown in the following chapters.

CHAPTER III

γ TOLERANCE LIMITS

Problems associated with the exponential distribution will in general require one-sided tolerance limits. Hence, if $L(x_1, \ldots x_n)$ denotes a tolerance limit, let the interval (L, ∞), denoted by R_L, be the tolerance region associated with the limit L. Let

$$P_L = \int_L^{\infty} \frac{1}{\theta} e^{-x/\theta} dx$$

denote the content of the tolerance region R_L. That is P_L is the proportion of the population in the interval (L, ∞). R_L is a γ tolerance region for a proportion β if $\Pr[P_L \geq \beta] = \gamma$. In general, $\Pr[P_L \geq \beta]$ could depend on the unknown parameters of the parent distribution, but for the exponential distribution this probability is independent of θ for the particular tolerance regions to be considered.

Limits Based on $\bar{x}$

In order to construct a γ tolerance region, it is necessary to find a function $L(x_1, \ldots, x_n)$ such that $\Pr[P_L \geq \beta] = \gamma$ for specified β, γ, and n. Consider a function of $\bar{x}$ in the form $L_1 = -\bar{x}k \log \beta$. It is necessary to find k as a function of β, γ, and n, such that $\Pr[P_{L_1} \geq \beta] = \gamma$.

$$P_{L_1} = \int_{-k\bar{x} \log \beta}^{\infty} \frac{1}{\theta} e^{-x/\theta} dx$$

$$P_{L_1} = 1 - F_x(-k\bar{x} \log \beta)$$

$$P_{L_1} = e^{\frac{\bar{x} k \log \beta}{\theta}}$$

$$P_{L_1} = \beta^{\frac{k\bar{x}}{\theta}} .$$

$$\Pr[P_{L_1} \geq \beta] = \Pr[\beta^{\frac{k\bar{x}}{\theta}} \geq \beta]$$

$$= \Pr[\bar{x} \leq \frac{\theta}{k}]$$

$$= F_{\bar{x}}(\frac{\theta}{k})$$, where the cumulative distribution of a variable z will be denoted by $F_z(z)$.

Since $\bar{x}$ has a gamma density with parameters $\alpha = n - 1$ and $\beta = \frac{\theta}{n}$,

$$F_{\bar{x}}(\bar{x}) = 1 - \sum_{i=0}^{n-1} \frac{\left(\frac{n\bar{x}}{\theta}\right)^i e^{\frac{-n\bar{x}}{\theta}}}{i!} ,$$

and

$$F_{\bar{x}}(\frac{\theta}{k}) = 1 - \sum_{i=0}^{n-1} \frac{\left(\frac{n}{k}\right)^i e^{-n/k}}{i!} .$$

Thus, to obtain k, solve

$$F_{\bar{x}}(\frac{\theta}{k}) = 1 - \sum_{i=0}^{n-1} \frac{(\frac{n}{k})^i e^{-n/k}}{i!} = \gamma \text{ for } k.$$

A cumulative chi-square table can be used to determine k. In Pearson and Hartley's notation (7),

$$Q(\chi^2, \nu) = \sum_{j=0}^{\frac{1}{2}\nu - 1} \frac{e^{-\frac{1}{2}\chi^2} (\frac{1}{2}\chi^2)^j}{j!} ,$$

so

$$F_{\overline{x}}(\frac{\theta}{k}) = 1 - Q(\frac{2n}{k}, 2n) .$$

For example, if $n = 20$, $\gamma = .95$, then $Q = (\frac{40}{k}, 40) = .05$ and $\chi^2 = \frac{40}{k} = 55.758$, which gives $k = .717$. Hence for $n = 20$, $L_1 = -.717\,\overline{x} \log \beta$ yields

$$\Pr[P_{L_1} \geq \beta] = .95 .$$

If k is specified beforehand, $\Pr[P_{L_1} \geq \beta]$ can be evaluated for a given n by using the chi-square table to find $Q(\chi^2, \nu)$ for fixed χ^2 and ν. If both k and γ are specified, one can solve for the sample size required to satisfy $\Pr[P_L \geq \beta] = \gamma$, if such a sample size exists.

For a given limit $L_1 = -\overline{x} k \log \beta$ which satisfies $\Pr[P_{L_1} \geq \beta] = \gamma$, an experimenter may wish to know the probability that P_{L_1} is also less than or equal to $\beta + \varepsilon$, where $0 < \varepsilon \leq 1 - \beta$. This probability is given by

$$\Pr[\beta \leq P_{L_1} \leq \beta + \varepsilon] = \Pr[\beta \leq \beta^{\frac{k\overline{x}}{\theta}} \leq \beta + \varepsilon]$$

$$= \Pr[\frac{\theta \log(\beta + \varepsilon)}{k \log \beta} \leq \overline{x} \leq \frac{\theta}{k}]$$

$$= F_{\overline{x}}(\frac{\theta}{k}) - F_{\overline{x}}(\frac{c\theta}{k}) ,$$

where $$c = \frac{\log(\beta + \varepsilon)}{\log \beta} \quad .$$

Thus,

$$\Pr[\beta \leq P_{L_1} \leq \beta + \varepsilon) = Q[\frac{2cn}{k}, 2n] - Q[\frac{2n}{k}, 2n] \quad .$$

To illustrate, if $n = 20$, $\beta = .95$, $\varepsilon = .02$, and $k = .717$ as above, then $c = .5685$ and

$$\Pr[\beta \leq P_{L_1} \leq \beta + \varepsilon] = \Pr[.95 \leq P_{L_1} \leq .97]$$

$$= Q(31.72, 40) - Q(55.76, 40)$$

$$= .822 - .05 = .772.$$

An experimenter may wish to have a tolerance limit L_2 in a form which will allow him to fix the probability γ beforehand and then be assured of being able to solve for the sample size n which will satisfy the equation $\Pr[\beta \leq P_{L_2} \leq \beta + \varepsilon] = \gamma$, for arbitrarily fixed β and ε. For this purpose, consider the limit $L_2 = -\bar{x} \log(\beta + \frac{1}{2}\varepsilon)$. Since

$$\int_{-\theta \log(\beta + \varepsilon)}^{\infty} \frac{1}{\theta} e^{-x/\theta} dx = \beta + \varepsilon$$

for arbitrary ε, it follows that

$$\Pr[\beta \leq P_{L_2} \leq \beta + \varepsilon] = \Pr[-\theta \log(\beta + \varepsilon) \leq -\bar{x} \log(\beta + \frac{1}{2}\varepsilon) \leq -\theta \log \beta]$$

$$= \Pr[\frac{\theta \log(\beta + \varepsilon)}{\log(\beta + \frac{1}{2}\varepsilon)} \leq \bar{x} \leq \frac{\theta \log \beta}{\log(\beta + \frac{1}{2}\varepsilon)}] \quad .$$

Let $c_1 = \frac{\log(\beta + \varepsilon)}{\log(\beta + \frac{1}{2}\varepsilon)}$ and $c_2 = \frac{\log \beta}{\log(\beta + \frac{1}{2}\varepsilon)}$.

Then

$$\Pr[\beta \leq P_{L_2} \leq \beta + \varepsilon] = \Pr[c_1\theta \leq \bar{x} \leq c_2\theta]$$

$$= F_{\bar{x}}(c_2\theta) - F_{\bar{x}}(c_1\theta)$$

$$= Q(2c_1 n, 2n) - Q(2c_2 n, 2n) \quad .$$

Since $E(\bar{x}) = \theta$ and the distribution of $\bar{x}$ approaches a normal density as n increases, $\Pr[c_1\theta \leq \bar{x} \leq c_2\theta] = \gamma$ has a solution for n for any γ, since $c_1 < 1$ and $c_2 > 1$. More generally, a limit of the form $L_3 = -k\bar{x}\log(\beta + \frac{1}{2}\varepsilon)$ can be used, and a sample size n will exist which yields a specified probability γ, if $c_1 < k < c_2$. This follows since

$$\Pr[\beta \leq P_{L_3} \leq \beta + \varepsilon] = \Pr[\frac{c_1\theta}{k} \leq \bar{x} \leq \frac{c_2\theta}{k}] \quad ,$$

which increases as n increases when $c_1 < k < c_2$. Since the distribution of $\bar{x}$ becomes symmetric as n increases, one would expect a limit with k approximately midway between c_1 and c_2 to produce an equation, $\Pr[\frac{c_1\theta}{k} \leq \bar{x} \leq \frac{c_2\theta}{k}] = \gamma$, which would be satisfied by a minimum sample size n. For a fixed n, the k which maximizes $\Pr[\beta \leq P_{L_3} \leq \beta + \varepsilon]$ is the k which renders $1 - Q(\frac{2nc_2}{k}, 2n) = Q(\frac{2nc_1}{k}, 2n)$. This k is a function of n; therefore, in order to find the limit L_3 such that the minimum sample size n

will satisfy

$$\Pr[\beta \le P_{L_3} \le \beta + \varepsilon] = \gamma \quad ,$$

it is necessary to find k and n simultaneously by an iterative procedure. For example, set k = 1 and find n such that the content of the region associated with that limit is between β and $\beta + \varepsilon$ with probability γ. Then, find the k which maximizes the probability for that n. Now find the necessary n for this new k. Continue in this fashion until n does not change and this n is the minimum sample size.

As an example, let $\beta = .95$, $\varepsilon = .02$, $k = 1$, and $\gamma = .80$, then $c_1 = \dfrac{\log .97}{\log .96} = .746$, $c_2 = \dfrac{\log .95}{\log .96} = 1.256$, and $L_3 = -\bar{x} \log .96$.

$$\Pr[.95 \le P_{L_3} \le .97] = Q(\frac{2c_1 n}{k}, 2n) - Q(\frac{2c_2 n}{k}, 2n)$$

$$= Q[.746(2n), 2n] - Q[1.256(2n), 2n] \quad .$$

For

$$2n = 40, \ Q(29.84, 40) - Q(50.24, 40) = .88 - .12 \quad = .76$$

$$2n = 46, \ Q(34.32, 46) - Q(57.78, 46) = .89 - .11 \quad = .78$$

$$2n = 50, \ Q(37.30, 50) - Q(62.70, 50) = .908 - .101 \doteq .80 \quad .$$

Hence, n = 25 is the required sample size to yield $\Pr[.95 \le P_{L_3} \le .97] = .8$ for the tolerance limit $L_3 = -\bar{x} \log .96$. Since $Q(62.70, 50) = .101$ and $1 - Q(37.30, 50) = .092$ are almost equal, k = 1 seems to be sufficiently close to the k value which gives the minimum sample size n.

Since $\hat{\theta}_{r,n}$ and $\hat{\theta}'_{r,n}$ are distributed exactly the same as $\bar{x}$ with n replaced by r, γ tolerance regions based on $\hat{\theta}_{r,n}$ or $\hat{\theta}'_{r,n}$ can be obtained by simply substituting them for $\bar{x}$ and replacing n by r in all the formulae. Thus, γ tolerance regions based on single order statistics remain to be developed.

Limits Based on Single Order Statistics

Let $L_4 = -x_{r,n}\, k \log \beta$. Then a k is desired which satisfies $\Pr[P_{L_3} \geq \beta] = \gamma$ for specified β, γ, and n.

$$P_{L_4} = \int_{-x_{r,n}\, k \log \beta}^{\infty} \frac{1}{\theta} e^{\beta x/\theta}\, dx$$

$$= e^{\frac{k \log \beta\; x_{r,n}}{\theta}} = \beta^{\frac{k\, x_{r,n}}{\theta}} .$$

$$\Pr[P_{L_4} \geq \beta] = \Pr[\beta^{\frac{kx_{r,n}}{\theta}} \geq \beta]$$

$$= \Pr[x_{r,n} \leq \frac{\theta}{k}]$$

$$= F_{x_{r,n}}(\frac{\theta}{k}) .$$

Now,

$$F_{x_{r,n}}(x) = \int_0^x \frac{n!}{(r-1)!\,(n-r)!} \frac{1}{\theta}[1 - e^{-x/\theta}]^{r-1}[e^{-x/\theta}]^{n-r+1} dx .$$

Let $y = 1 - e^{-x/\theta}$, $dy = \frac{1}{\theta} e^{-x/\theta} dx$, then

$$F_{x_{r,n}}(x) = \int_0^{1-e^{-x/\theta}} \frac{n!}{(r-1)!\,(n-r)!}\, y^{r-1}(1-y)^{n-r)} dy$$

and

$$F_{x_{r,n}}(\frac{\theta}{k}) = \int_0^{1-e^{-1/k}} \frac{n!}{(r-1)!(n-r)!} y^{r-1}(1-y)^{n-r} dy \quad .$$

$g(y) = \frac{n!}{(r-1)!(n-r)!} y^{r-1}(1-y)^{n-r}$ is a Beta density with parameters $r-1$ and $n-r$, so following the usual notation,

$$F_{x_{r,n}}(\frac{\theta}{k}) = I_{1-e^{-\frac{1}{k}}}(r, n-r+1) \quad ,$$

where

$$I_x(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \int_0^x u^{a-1}(1-u)^{b-1} du \quad .$$

Thus, the equation $F_{x_{r,n}}(\frac{\theta}{k}) = \gamma$ can be solved for k by use of Incomplete Beta tables to obtain the desired k. (8).

For example, if an experimenter wishes to find a tolerance limit $L_4 = -x_{r,n} k \log \beta$, where $x_{r,n}$ is the median of a sample of size $n = 19$, say, such that $\Pr[P_{L_4} \geq \beta] = .90$, k is the solution of $I_{1-e^{-\frac{1}{k}}}(10, 10) = .90$. So, $I_{e^{-\frac{1}{k}}}(10, 10) = .10$ and the table gives $e^{-1/k} = .35793$, which gives $k = .974$ and $L_4 = -.974\, x_{r,n} \log \beta$. Now consider

$$\Pr[\beta \leq P_{L_4} \leq \beta + \varepsilon] = \Pr[-\theta \log(\beta+\varepsilon) \leq -k x_{r,n} \log \beta \leq -\theta \log \beta]$$

$$= \Pr[\frac{\theta \log(\beta+\varepsilon)}{k \log \beta} \leq x_{r,n} \leq \frac{\theta}{k}]$$

$$= F_{x_{r,n}}(\frac{\theta}{k}) - F_{x_{r,n}}(\frac{\theta \log(\beta+\varepsilon)}{k \log \beta})$$

$$= I_{c_1}(r,\ n-r+1) - I_{c_2}(r,\ n-r+1)\ ,$$

where $c_1 = 1 - e^{-1/k}$ and $c_2 = 1 - e^{-\frac{\theta \log(\beta+\varepsilon)}{k \log \beta}}$.

If, for example, $\beta = .95$, $\varepsilon = .02$, $n = 19$, and $k = .974$, as in the above example, then

$$c_1 = 1 - e^{-\frac{1}{.974}} = .642, \qquad c_2 = 1 - e^{-\frac{.5685}{.974}} = .443\ ,$$

and

$$\Pr[.95 \leq P_{L_4} \leq .97] = I_{.642}(10,\ 10) - I_{.443}(10,\ 10)$$

$$= .90 - .30$$

$$= .60\ .$$

In order to obtain a limit L based on $x_{r,n}$ with the property that $\Pr[\beta \leq P_L \leq \beta + \varepsilon]$ increases as n increases for arbitrarily specified ε, β, and γ, as was done for $\overline{x}$, consider the limit $L_5 = \dfrac{-kx_{r,n}\log(\beta + \frac{1}{2}\varepsilon)}{d}$, where $d = \sum_{i=1}^{r} \dfrac{1}{n-i+1}$ as given in Chapter I.

Then,

$$\Pr[\beta \leq P_{L_5} \leq \beta + \varepsilon] = \Pr[-\theta \log(\beta+\varepsilon) \leq \frac{-kx_{r,n}\log(\beta+\frac{1}{2}\varepsilon)}{d} \leq -\theta \log \beta]$$

$$= \Pr[\frac{\theta\, d \log(\beta+\varepsilon)}{k \log(\beta+\frac{1}{2}\varepsilon)} \leq x_{r,n} \leq \frac{\theta\, d \log \beta}{k \log(\beta+\frac{1}{2}\varepsilon)}]$$

$$= \Pr[\frac{\theta\, d\, c_1}{k} \leq x_{r,n} \leq \frac{\theta\, d\, c_2}{k}]\ ,$$

where $c_1 = \frac{\log(\beta + \varepsilon)}{\log(\beta + \frac{1}{2}\varepsilon)}$, and $c_2 = \frac{\log \beta}{\log(\beta + \frac{1}{2}\varepsilon)}$.

As n increases and r/n approaches some constant, the distribution of $x_{r,n}$ approaches a symmetric distribution about $E(x_{r,n}) = \theta d$. Hence, for $c_1 < k < c_2$, there will exist an n for arbitrary ε, β, and γ, such that $\Pr[\beta \leq P_{L_5} \leq \beta + \varepsilon] = \gamma$.

$$\Pr\left[\frac{\theta d c_1}{k} \leq x_{r,n} \leq \frac{\theta d c_2}{k}\right] = F_{x_{r,n}}\left(\frac{\theta d c_2}{k}\right) - F_{x_{r,n}}\left(\frac{\theta d c_1}{k}\right)$$

$$= I_{x_2}(r, n-r+1) - I_{x_1}(r, n-r+1) \quad ,$$

where $x_2 = 1 - e^{-\frac{c_2 d}{k}}$, and $x_1 = 1 - e^{-\frac{c_1 d}{k}}$.

Setting $I_{x_2}(r, n-r+1) - I_{x_1}(r, n-r+1) = \gamma$, it is desired to determine n and hence also r, where r is some prescribed function of n.

An iterative procedure is required to solve for n. In general, r, x_1, and x_2 must be recomputed for each n value chosen. These computations may be simplified for special cases. For instance, if the median for odd sample size is used, then $r = \frac{n+1}{2}$ and d is approximately log 2. $\lim_{n\to\infty} d = \log 2$ so a solution for n exists for the limit $L_5 = \frac{-kx_{r,n}\log(\beta + \frac{1}{2}\varepsilon)}{\log 2}$, where d is replaced by log 2 and $r = \frac{n+1}{2}$.

$$\Pr[\beta \leq P_{L_5} \leq \beta + \varepsilon] = I_{x_2}(r, r) - I_{x_1}(r, r) \quad ,$$

where $x_2 = 1 - e^{-\frac{c_2 \log 2}{k}}$, and $x_1 = 1 - e^{\frac{c_1 \log 2}{k}}$ are independent of n. For $\beta = .95$, $\varepsilon = .02$, $k = 1$, and $\gamma = .80$, then $c_1 = .7462$, $c_2 = 1.2565$, $x_1 = .5893$, $x_2 = .405$, and the r is desired which satisfies $I_{.5873}(r, r) - I_{.405}(r, r) = .80$.

For $r = 24$,

$$I_{.5893}(24, 24) - I_{.405}(24, 24) = .891 - .092 \doteq .80.$$

Thus, $r = 24$ and $n = 2r - 1 = 47$ compare to a sample size of $n = 25$ for a similar probability statement when $\overline{x}$ is used.

It should be noted that if $\varepsilon = 1 - \beta$, then

$$\Pr[\beta \leq P_L \leq \beta + \varepsilon] = \Pr[\beta \leq P_L \leq 1] = \Pr[P_L \geq \beta].$$

Hence, $\Pr[P_L \geq \beta]$ can be considered as a special case of $\Pr[\beta \leq P_L \leq \beta + \varepsilon]$.

CHAPTER IV

β-EXPECTATION TOLERANCE REGIONS

This chapter will be concerned with β-expectation tolerance regions for the exponential distribution. A tolerance region R_L will be called a biased β-expectation tolerance region if $E(P_L) \neq \beta$ for finite n and arbitrary β, but $\lim_{n\to\infty} E(P_L) = \beta$. If $E(P_L) = \beta$ for all n, then R_L will be an unbiased β-expectation tolerance region or simply a β-expectation tolerance region.

β-Expectation Limits Based on $\bar{x}$

Since

$$\int_{-\theta \log \beta}^{\infty} \frac{1}{\theta} e^{-x/\theta} dx = \beta,$$

it is reasonable to consider limits of the form $L = -\hat{\theta} \log \beta$, where $\hat{\theta}$ is an unbiased estimate of θ. Suppose $L_1 = -\bar{x} \log \beta$, then

$$P_{L_1} = \int_{-\bar{x}\log\beta}^{\infty} \frac{1}{\theta} e^{-x/\theta} dx = e^{\frac{\bar{x}\log\beta}{\theta}} = \beta^{\bar{x}/\theta} .$$

$$E(P_{L_1}) = E(e^{\frac{\bar{x}\log\beta}{\theta}}) = \int_0^{\infty} e^{\frac{\bar{x}\log\beta}{\theta}} \frac{1}{(n-1)!} (\frac{n}{\theta})^n \bar{x}^{n-1} e^{-\frac{n\bar{x}}{\theta}} d\bar{x}$$

$$= \int_0^{\infty} \frac{1}{(n-1)!} (\frac{n}{\theta})^n \bar{x}^{n-1} e^{-\frac{nx}{\theta} + \frac{x}{\theta}\log\beta} d\bar{x}$$

$$= \frac{\gamma^n}{(\frac{\theta}{n})^n} \int_o^\infty \frac{1}{(n-1)!\,\gamma^n} \;\bar{x}^{\,n-1} e^{-\bar{x}/\gamma} dx \quad ,$$

where $\gamma = \dfrac{\theta}{n - \log \beta}$. Since the integral of the density of $\bar{x}$ over its entire range is one, then

$$E(P_{L_1}) = \frac{\gamma^n}{(\frac{\theta}{n})^n} = (\frac{\theta}{n - \log \beta})^n (\frac{n}{\theta})^n = (\frac{n}{n - \log \beta})^n \quad .$$

$E(P_{L_1}) \neq \beta$, but $\lim_{n \to \infty} (\frac{n}{n - \log \beta})^n = \beta$, so R_{L_1} is a biased β-expectation tolerance region.

An unbiased β-expectation tolerance region can be obtained by using $L_2 = n(\beta^{-1/n} - 1)\bar{x}$, which gives $P_{L_2} = e^{\frac{-n(\beta^{-1/n} - 1)\bar{x}}{\theta}}$ and

$$E(P_{L_2}) = \int_o^\infty e^{\frac{-n(\beta^{-1/n} - 1)\bar{x}}{\theta}} \frac{1}{(n-1)!} (\frac{n}{\theta})^n \bar{x}^{\,n-1} e^{-\frac{n\bar{x}}{\theta}} d\bar{x}$$

$$= \int_o^\infty \frac{1}{(n-1)!} (\frac{n}{\theta})^n e^{-\frac{n(\beta^{-1/n})\bar{x}}{\theta}} \bar{x}^{\,n-1} d\bar{x}$$

$$= \frac{\left(\frac{\theta\beta^{1/n}}{n}\right)^n}{(\frac{\theta}{n})^n} \int_o^\infty \frac{1}{(n-1)!\,\delta^n} \bar{x}^{\,n-1} e^{-\bar{x}/\delta} d\bar{x} \quad ,$$

where $\delta = \dfrac{\theta\beta^{1/n}}{n}$.

Hence,

$$E(P_{L_2}) = (\frac{\theta\beta^{1/n}}{n})^n (\frac{\theta}{n})^{-n} = \beta \quad .$$

Similarly, $L = r(\beta^{-1/r} - 1)\,\hat{\theta}_{r,n}$ and $L = r(\beta^{-1/r} - 1)\,\hat{\theta}'_{r,n} = n(\beta^{-1/r} - 1)\,x'_{r,n}$ provide β-expectation tolerance regions based on $\hat{\theta}_{r,n}$ and $\hat{\theta}'_{r,n}$, since $\hat{\theta}_{r,n}$ and $\hat{\theta}'_{r,n}$ have the same distribution as $\bar{x}$ with n replaced by r.

β-Expectation Regions Based on Single Order Statistics

For a single order statistic $x_{r,n}$, let $L_3 = -ax_{r,n} \log \beta$, where $a = (\sum_{i=1}^{r} \frac{1}{n-i+1})^{-1}$. $E(ax_{r,n}) = \theta$, so $E(L_3) = -\theta \log \beta$ and this means $E(P_{L_3}) = \beta$ in the limit. For finite n,

$$P_{L_3} = \int_{-ax_{r,n}\log\beta}^{\infty} \frac{1}{\theta} e^{-x/\theta} dx$$

$$= e^{\frac{ax_{r,n}\log\beta}{\theta}}$$

$$= e^{-\frac{cx_{r,n}}{\theta}} , \qquad \text{where } c = -a \log \beta \quad .$$

$$E(P_{L_3}) = \int_0^{\infty} e^{-\frac{cx_{r,n}}{\theta}} \frac{n!}{(r-1)!(n-r)!} \frac{1}{\theta} [1 - e^{-\frac{x_{r,n}}{\theta}}]^{r-1} [e^{-\frac{x_{r,n}}{\theta}}]^{n-r+1} dx_{r,n}.$$

To evaluate this expectation, let $y = e^{-\frac{x_{r,n}}{\theta}}$, $dy = -\frac{1}{\theta} e^{-\frac{x_{r,n}}{\theta}}$.

$$E(P_{L_3}) = \int_0^1 y^c \frac{n!}{(r-1)!(n-r)!} y^{n-r} (1-y)^{r-1} dy = \frac{n!(n-r+c)}{(n-r)!(n+c)!} \quad .$$

$E(P_{L_3}) \neq \beta$, in general, therefore, $L_3 = -ax_{r,n} \log \beta = cx_{r,n}$ determines a biased β-expectation tolerance region. It is evident that a limit of the form $L_4 = kx_{r,n}$ will determine an unbiased β-expectation tolerance region for a given $x_{r,n}$ if, and only if, k satisfies the relation $\frac{n!(n-r+k)!}{(n-r)!(n+k)!} = \beta$. This equation may be written as $\frac{[n]_r}{[n+k]_r} = \beta$, where $[n]_r = n(n-1) \ldots (n-r+1)$. Then $[n]_r = [n+k]_r \beta$ is a polynomial of degree r in k. It has a unique real positive solution for k since all the coefficients are positive except the constant term which is equal to $(\beta-1)[n]_r$. This solution can be found readily by an iterative procedure for small n. For large n, k is approximately equal to $c = -a \log \beta$ as indicated above.

For example, if n = 5, r = 3, and $\beta = .95$, then

$$\frac{n!(n-r+k)!}{(n-r)!(n+k)!} = \frac{5!(2+k)!}{2!(5+k)!} = \frac{3 \cdot 4 \cdot 5}{(3+k)(4+k)(5+k)} . \text{ Solving } \frac{3 \cdot 4 \cdot 5}{(3+k)(4+k)(5+k)} = .95$$

yields $k = .0661$. Hence, for $L_4 = .0661\, x_{3,5}$, $E(P_{L_4}) = .95$. The approximate solution based on L_3 for this example is given by $c = -a \log \beta = -(1.2766)(-.051291) = .0655$.

Precision of the Tolerance Regions

The precision of a β-expectation tolerance region will be measured in terms of the variance of its content. The variances of the content for the tolerance regions based on L_2 and L_4 are derived below. Since these variances are difficult to compare except numerically, the variances of the associated limits are also considered, since they can be used for comparing the precision of different methods.

Consider first the unbiased β-expectation tolerance region based on $\bar{x}$ determined by $L_2 = n(\beta^{-1/n} - 1)\bar{x}$.

$$P_{L_2} = e^{-\frac{n(\beta^{-1/n} - 1)\bar{x}}{\theta}} \quad , \qquad E(P_{L_2}) = \beta \quad .$$

$$(P_{L_2})^2 = e^{-\frac{2n(\beta^{-1/n} - 1)\bar{x}}{\theta}} \quad ,$$

$$E(P_{L_2})^2 = \int_0^\infty e^{-\frac{2n(\beta^{-1/n} - 1)\bar{x}}{\theta}} \frac{1}{(n-1)!} (\frac{n}{\theta})^n e^{-\frac{n\bar{x}}{\theta}} \bar{x}^{n-1} d\bar{x}$$

$$= \int_0^\infty \frac{1}{(n-1)!} (\frac{n}{\theta})^n e^{-\frac{n\bar{x}}{\theta}[1 + 2(\beta^{-1/n} - 1)]} \bar{x}^{n-1} d\bar{x}$$

$$= (2\beta^{-1/n} - 1)^{-n} \int_0^\infty \frac{1}{(n-1)!} (\frac{n}{\theta}) (2\beta^{-1/n} - 1)^n \bar{x}^{n-1} e^{-\frac{n\bar{x}}{\theta}(2\beta^{-\frac{1}{n}})} d\bar{x}$$

$$= (2\beta^{-1/n} - 1)^{-n} \quad .$$

So,

$$\mathrm{Var}(P_{L_2}) = E(P_{L_2})^2 - (EP_{L_2})^2$$

$$= (2\beta^{-1/n} - 1)^{-n} - \beta^2 \quad .$$

Now consider $L_4 = kx_{r,n}$, where k satisfies $\frac{n!(n-r+k)!}{(n-r)!(n+k)!} = \beta$.

$$E(P_{L_4}) = E(e^{-\frac{kx_{r,n}}{\theta}}) = \frac{n!(n-r+k)!}{(n-r)!(n+k)!} = \beta .$$

Since k is a constant it follows that

$$E(P_{L_4})^2 = E(e^{-\frac{2kx_{r,n}}{\theta}}) = \frac{n!\,(n-r+2k)!}{(n-r)!\,(n+2k)!} .$$

So,

$$\text{Var}(P_{L_4}) = \frac{n!\,(n-r+2k)!}{(n-r)!\,(n+2k)!} - \beta^2 .$$

Similarly, the mean square errors for biased β-expectation tolerance limits can be determined.

Thus, for a given problem the variance of the content of the tolerance region to be used can be computed by the above formulae to give a measure of the precision which will be obtained. Comparisons of the variances of the contents of the regions based on each of the estimators can also be made numerically for a given problem to help determine the most desirable method for that situation.

The variances of the limits themselves can also be compared to indicate the relative efficiency of the different methods. For $L_2 = n(\beta^{-1/n} - 1)\,\overline{x}$, $\text{Var}(L_2) = n^2(\beta^{-1/n} - 1)^2 \text{ var } \overline{x}$. $\text{Var } \overline{x} = \frac{\theta^2}{n}$, so $\text{Var}(L_2) = n(\beta^{-1/n} - 1)^2\theta^2$. For $L_4 = kx_{r,n}$, $\text{Var}(L_4) = k^2\text{Var } x_{r,n}$. $\text{Var } x_{r,n} = \theta^2 \sum_{i=1}^{r} (\frac{1}{n-i+1})^2$, so $\text{Var}(L_4) = k^2\theta^2 \sum_{i=1}^{r} (\frac{1}{n-i+1})^2$.

Much simpler expressions can be obtained for comparisons for large sample sizes by using the concept of asymptotic variance. The asymptotic variance of an estimator is the variance of that estimator for the large sample or asymptotic distribution of that estimator.

If $\frac{r}{n}$ approaches a constant as $n \to \infty$, the limiting distribution of $x_{r,n}$ is normal with asymptotic mean $m(x_{r,n})$ defined by $\int_0^{m(x_{r,n})} f(x)\,dx = \frac{r}{n+1}$ and asymptotic variance, $AV(x_{r,n}) = \frac{r(n-r+1)}{n(n+1)^2} \frac{1}{f[m(x_{r,n})]^2}$. (9). Thus, for the exponential distribution,

$$\int_0^{m(x_{r,n})} \frac{1}{\theta} e^{-x/\theta}\,dx = 1 - e^{-\frac{m(x_{r,n})}{\theta}} = \frac{1}{n+1} ,$$

and

$$m(x_{r,n}) = -\theta \log \left(\frac{n-r+1}{n+1}\right) .$$

$$AV(x_{r,n}) = \frac{r(n-r+1)}{n(n+1)^2} \quad f\,[m(x_{r,n})]$$

$$= \frac{\theta^2 r(n-r+1)}{n(n+1)^2} \left(\frac{n-r+1}{n+1}\right)^{-2}$$

$$= \frac{r\theta^2}{n(n-r+1)} .$$

Now, $L_4 = kx_{r,n}$ approaches $-\theta \log \beta$ as n increases, and $x_{r,n}$ approaches $-\theta \log \left(\frac{n-r+1}{n+1}\right)$, so k must approach $\log \beta \left[\log\left(\frac{n-r+1}{n+1}\right)\right]^{-1}$ asymptotically as n increases. Thus,

$$AV(L_4) = (\log \beta)^2 \left[\log \left(\frac{n-r+1}{n+1}\right)\right]^{-2} AV(x_{r,n})$$

$$= (\log \beta)^2 \left[\log \left(\frac{n-r+1}{n+1}\right)\right]^{-2} \frac{r\theta^2}{n(n-r+1)} .$$

Now, consider $L_2 = n(\beta^{-1/n} - 1)\,\bar{x}$. $\lim_{n\to\infty} n(\beta^{-1/n} - 1) = -\log \beta$, so

$$AV(L_2) = (\log \beta)^2 \, AV(\bar{x})$$

$$= (\log \beta)^2 \, \frac{\theta^2}{n} \quad .$$

Thus, the ratio $\dfrac{AV(L_2)}{AV(L_4)} = \dfrac{n-r+1}{r} \left[\log \left(\dfrac{n-r+1}{n+1}\right)\right]^2$ is a measure of the relative efficiency of the limit L_4 compared to L_2.

For the limit $L = r(\beta^{-1/r} - 1)\ \hat{\theta}_{r,n}$ based on the first r observations, $AV(L) = (\log \beta)^2 \dfrac{\theta^2}{r}$. $AV(L) = \dfrac{n}{r}\ AV(L_2)$, so

$$\frac{AV(L)}{AV(L_4)} = \frac{n(r-r+1)}{r^2} \left[\log \left(\frac{n-r+1}{n+1}\right) \right]^2 \quad .$$

If $\frac{r}{n} \rightarrow p$, say, as $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} \frac{AV(L_2)}{AV(L_4)} = \frac{1-p}{p} \left[\log (1-p)\right]^2$$

and

$$\lim_{n \rightarrow \infty} \frac{AV(L)}{AV(L_4)} = \frac{1-p}{p^2} \left[\log (1-p)\right]^2 \quad .$$

For example, to compare limits based on the median with limits based on $\bar{x}$ and $\hat{\theta}_{r,n}$, let $p = \frac{r}{n} = \frac{1}{2}$. Then

$$\lim_{n \rightarrow \infty} \frac{AV(L_2)}{AV(L_4)} = (\log \tfrac{1}{2})^2 = .48$$

$$\lim_{n \rightarrow \infty} \frac{AV(L)}{AV(L_4)} = 2\,(.48) = .96 \quad .$$

Thus, the above considerations can be used to measure the precision of tolerance regions and to compare the relative efficiency of different methods for a particular problem.

CHAPTER V

SPECIFICATION LIMITS

Many times a manufacturer will be more concerned with whether his product is meeting certain fixed standards or specifications with a certain level of confidence, than with setting variable limits which may not be close to the required standards determined by other reasons.

Let $b > 0$ denote a one-sided specification limit for the exponential distribution. $P_b = \int_b^\infty \frac{1}{\theta} e^{-x/\theta} dx$ is the proportion of the population in the interval (b, ∞). Suppose it is desired that $100\,\beta$ per cent or more of the manufactured products have a life greater than or equal to b, i.e., that $\int_b^\infty \frac{1}{\theta} e^{-x/\theta} dx \geq \beta$. It cannot be determined whether a process is meeting this condition without complete testing. Samples of the product can be drawn, however, to determine if the process is satisfying this requirement with at least a certain confidence γ. If on the basis of a sample, there is less than γ confidence that $P_b \geq \beta$, then the manufacturing process must be altered since the products are not meeting specifications at the desired confidence level.

Large, sample means are associated with high confidence that $P_b \geq \beta$. Thus, if the calculated sample mean $\bar{x}_o$ is so small that the probability of getting an $\bar{x}_o$ that small or smaller is less than γ for every value of θ for which $P_b \geq \beta$, then the production will not be meeting specifications with the desired confidence γ.

Now,

$$P_b = \int_b^\infty \frac{1}{\theta} e^{-x/\theta} dx = e^{-b/\theta} . \qquad \text{So} \qquad P_b \geq \beta$$

when $\theta \geq -b/\log \beta$. For a given θ,

$$\Pr[\bar{x} \leq \bar{x}_o] = F_{\bar{x}}(\bar{x}_o) = 1 - \sum_{i=1}^{n-1} \frac{\left(\frac{n\bar{x}_o}{\theta}\right)^i e^{-\frac{n\bar{x}_o}{\theta}}}{i!} .$$

$F_{\bar{x}}(\bar{x})$ increases as $\bar{x}$ increases, so it must decrease as θ increases since $F_{\bar{x}}(\bar{x})$ can be written as a function of the ratio $\frac{\bar{x}}{\theta}$. The minimum θ which renders $P_b \geq \beta$ is $\theta = \frac{-b}{\log \beta}$, so if $\Pr[\bar{x} \leq \bar{x}_o] < \gamma$ for $\theta = \frac{-b}{\log \beta}$, then $\Pr[\bar{x} \leq \bar{x}_o]$ will be less than γ for all θ for which $P_b \geq \beta$.

Let us evaluate $\gamma_o = \Pr[\bar{x} \leq \bar{x}_o]$ when $\theta = \frac{-b}{\log \beta}$.

$$\gamma_o = \Pr[\bar{x} \leq \bar{x}_o \mid \theta = \frac{-b}{\log \beta}]$$

$$= 1 - \sum_{i=1}^{n-1} \frac{\left(\frac{n\bar{x}_o \log \beta}{-b}\right)^i e^{\frac{n\bar{x}_o \log \beta}{b}}}{i!}$$

$$= 1 - Q\left(\frac{-2n\bar{x}_o \log \beta}{b}, 2n\right)$$

$$= 1 - Q(\chi^2, \nu), \quad \text{where } \chi^2 = \frac{-2n\bar{x}_o \log \beta}{b}$$

and $\nu = 2n$. If $\gamma_o < \gamma$, then $\Pr[\bar{x} \leq \bar{x}_o]$ is less than γ for all θ for which $P_b \geq \beta$ and the manufacturing process is not meeting the requirement that 100β per cent of the products have life length greater than or equal to b with confidence γ.

CHAPTER VI

SUMMARY

There are four commonly used estimators of θ in the exponential distribution. The estimators are $\bar{x} = \sum_{i=1}^{n} x_i$, $\hat{\theta}_{r,n} = \frac{1}{r}\left[\sum_{i=1}^{r} (x_i) - (n-r)\, x_{r,n}\right]$ for censored sampling, $\hat{\theta}'_{r,n} = \frac{n}{r} x'_{r,n}$ for censored sampling with replacement, and $\left[\sum_{i=1}^{r} (n-r+1)^{-1}\right]^{-1} x_{r,n}$ for non-parametric estimation. Tolerance limits are derived based on each of these estimators.

A γ tolerance region for a proportion β based on $\bar{x}$ is given by $L_1 = -k\bar{x} \log \beta$ where k satisfies $1 - Q(\frac{2n}{k}, 2n) = \gamma$. Since the distribution of $\hat{\theta}_{r,n}$ and $\hat{\theta}'_{r,n}$ is the same as the distribution of $\bar{x}$ with n replaced by r, all results for $\bar{x}$ can be applied to $\hat{\theta}_{r,n}$ and $\hat{\theta}'_{r,n}$. The γ tolerance region based on $x_{r,n}$ is given by $L_2 = -kx_{r,n} \log \beta$ where k satisfies $I_{1-e^{-1/k}}(r, n-r+1) = \gamma$. In each case $\Pr[\beta \leq P_L \leq \beta + \varepsilon]$ can also be determined. Limits of the form $L_3 = -\bar{x} \log (\beta + \frac{1}{2}\varepsilon)$ and $L_4 = -\left[\sum_{i=1}^{r} (n-i+1)^{-1}\right]^{-1} x_{r,n} \log (\beta + \frac{1}{2}\varepsilon)$ can be used to solve for the sample size n which will satisfy $\Pr[\beta \leq P_{L_3} \leq \beta + \varepsilon] = \gamma$ and $\Pr[\beta \leq P_{L_4} \leq \beta + \varepsilon] = \gamma$ for arbitrarily fixed β, ε, and γ.

β-expectation tolerance regions based on $\bar{x}$ and $x_{r,n}$ are given by $L_5 = n(\beta^{-1/n} - 1)\bar{x}$ and $L_6 = k\bar{x}_{r,n}$, where k satisfies $\frac{[n]_r}{[n+k]_r} = \beta$. Variances of the contents associated with these regions and also the variances of the limits themselves are derived. A problem concerned with given specification limits is also considered.

BIBLIOGRAPHY

(1) Epstein, Benjamin. "Estimation from Life Test Data." Technometrics, II (November, 1960), 447-54.

(2) Mood, A. M. Introduction to the Theory of Statistics. New York: McGraw-Hill Book Company, Inc., 1950.

(3) Epstein, Benjamin, and Sobel, M. "Life Testing." Journal of the American Statistical Association, XLVII (1953), 485-502.

(4) Epstein, Benjamin. "Statistical Life Test Acceptance Procedures." Technometrics, II (November, 1960), 435-46.

(5) Epstein, Benjamin. "Efficiency of Estimators of the Mean of an Exponential Distribution Based Only on the r-th Smallest Observations in an Ordered Sample." Annals of Mathematical Statistics, XXIII (1952), 640.

(6) Harter, H. Leon. "Estimating the Parameters of Negative Exponential Populations from One or Two Order Statistics." Annals of Mathematical Statistics, XXXII (December, 1961), 1078-84.

(7) Pearson, E. S., and Hartley, H. O. Biometrika Tables for Statisticians. London: Cambridge University Press, 1956.

(8) Pearson, Karl. Tables of the Incomplete Beta-Function. London: Cambridge University Press, 1932.

(9) Hastings, C.; F. Mosteller; J. W. Tukey; and C. Winsor. "Low Moments for Small Samples: A Comparative Study of Order Statistics." Annals of Mathematical Statistics, XVIII (1947), 413-26.

VITA

Lee J. Bain

Candidate for the Degree of

Master of Science

Thesis: TOLERANCE LIMITS FOR THE EXPONENTIAL DISTRIBUTION

Major Field: Mathematical Statistics

Biographical:

Personal Data: Born near Newkirk, Oklahoma, January 11, 1939, the son of Ralph R. and Mae Bain.

Education: Attended grade school in Uncas, Oklahoma; attended high school in Ponca City and Newkirk, Oklahoma; graduated from Newkirk High School in 1956; received the Bachelor of Science degree from the Oklahoma State University, with a major in mathematics, in May, 1960; completed requirements for the Master of Science degree in May, 1962.