Estimation of Parameters of Mixed Exponentially Distributed Failure Time Distributions from Censored Life Test Data

by

William Mendenhall, 3rd

Institute of Statistics

Mimeo Series No. 171

May, 1957

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION

Scientists in many fields of study are interested in the distribution of length of life for animate or inanimate objects subjected to various environmental conditions. For instance, one might be interested in the distribution of length of life of a particular type of insect subjected to a toxic spray or in the distribution of length of life of television tubes under normal household operating conditions. The entomologist might utilize the distribution of length of life to predict the proportion of the total number of insects sprayed which would survive a specified length of time and the engineer might be interested in predicting the fraction of the total production of television tubes which would fail during the warranty period. Many other important questions concerning the life characteristics of a system can be answered by the utilization of knowledge concerning the distribution of length of life of that system, commonly called a failure distribution.

Two situations present themselves. In the first, the experimenter has knowledge, based on theoretical considerations or past experience, concerning the general form of the failure distribution but does not know the specific values of the parameters of the distribution. In the second case, no knowledge is available and the experimenter must determine not only specific parameter values but also the form of the failure distribution. The choice of the distributional form and the estimation of parameters are based upon the observation of the experimental results of what is called a life test.

No attempt will be made to discuss in detail the various methods of choosing a distribution except to say that some insight into the proper distributional form, when unknown, can be gained by examining the shape of a histogram of failures based on a random sample from the total population. A density function is then chosen which seems to fit most nearly the histogram. A second method which is gaining in popularity, is to examine the failure rate of the system, where the failure rate in any particular time interval is the number failing in that interval divided by the number entering the interval, and to choose its theoretical counterpart, the best fitting conditional failure density. The probability that a unit fails in the interval t to $t + dt$, given that the unit has survived to time t, is $z(t)dt$, where $z(t)$ is the conditional failure density.

Life tests are of different types. In general, a random sample of n units is placed on test and the length of time to failure for each unit is observed. If all of the n units are tested to failure, sampling is said to be without replacement and is neither truncated nor censored. In some cases, however, the experimenter prefers not to wait until all of the units fail. The test is terminated after a fixed number of units, say r, have failed or alternatively after a fixed length of time, T. If the number of units surviving the termination of the test is not known, the sample is said to have been drawn from a truncated distribution. If the number surviving the test is known, the sample is said to be censored.

If maximum utilization of test equipment is desired, a unit is replaced immediately upon failure so that, at any time during the test, n units will be on test. The total number of units tested will be a random variable greater than or equal to n. Thus samples can be drawn with or without replacement, censored or non-censored, or drawn from a truncated distribution. In addition, censored sampling can be subdivided into two types according to whether the test is terminated after a fixed number of units have failed or after a fixed length of time has elapsed. This, in brief, lists the most common methods of classifying life tests.

Attention has been given in the past few years to the classification of failures of a unit into two or more subclasses. A typical example is the classification by electrical engineers of electronic tube failures into (1) mechanical defects, (2) failures due to gas or air, and (3) failures due to deterioration of performance characteristics. Probability density functions which currently are used as models for failure distributions have ignored these sub-populations and their failure properties. An attempt is made in this dissertation to formulate a realistic model for the failure of a unit in which failure is viewed as belonging to one of two or more failure types and to consider the statistical problem of estimation of the parameters of this model under censored sampling with a fixed test termination time.

The need for a study of problems dealing with mixed failure distributions became apparent during the investigation of some censored life test data on cathode ray tubes. The failure rates for both of two separate lots of tubes were non-monotonic, rising and falling early in the life of the tubes and then settling to what might appear to be a constant failure rate. The problem of choosing an appropriate failure distribution function to model the observed failure characteristics was made more difficult by the fact that the conditional failure density is monotonic for all of the common theoretical failure distributions. However, it seemed intuitively possible that the high failure rate early in the life of the tubes could have been caused by the presence of a sub-population of tubes which were actually undetected defectives, and that, as soon as these tubes failed, the sample failure rate dropped to that of the nondefectives. A subsequent review of the literature tended to confirm the opinion that many failure populations can and should be considered as mixtures of failure sub-populations.

A mechanical or electronic system is usually composed of a large number of components. Most of the components will be very reliable with the result that system failures are caused by a relatively small number of components. When this situation occurs, the model for mixed failure populations can also be applied to problems involving system reliability. It would often be desirable to estimate the probability for each cause of failure and also the average length of life for units classified by cause of failure.

One method of modeling the failure characteristics of a system would be to assume that the life of the ith component of the system is given by the random variable t_i. Hence, if a system contains k components, each unit will have associated with it the vector $(t_1, t_2, \dots t_k)$ specifying the length of life of each component and the corresponding multivariate failure probability density function $f(t_1, t_2, \dots t_k)$. The probability that the system will fail due to the ith component is equal to the probability that $t_i < t_j$, $j = 1, 2, \dots k$ and $j \neq i$. The expected values of t_i, $E(t_i)$, could also be determined. The primary disadvantage of this model is fairly obvious. The mathematical and distributional problems are very complex unless $t_1, t_2, \dots t_k$ are assumed independent of one another, a very unrealistic assumption in most cases. The assumption of independence of component lives need not be made when the problem is viewed from the standpoint of mixed populations.

The mixed failure population and the corresponding model can be stated precisely as follows: a population is postulated which is composed of s sub-populations mixed in proportion $p_1 : p_2 : \dots : p_s$ where $0 \leq p_i \leq 1$, $i = 1, 2, \dots s$, and $\sum_{i=1}^{s} p_i = 1$, the sub-populations representing failure types. Each unit of the population conceptually contains a tag which indicates the population to which the unit belongs and hence defines the way in which that particular unit will fail. The information on the tag, i.e., the cause of failure, is obtained only after failure has occurred. Further, the failure times for the units within the ith sub-population are assumed to have a cumulative failure probability distribution defined by $F_i(t)$ and a probability density function defined by $f_i(t)$, $i = 1, 2, \ldots s$. Then the probability that a unit drawn at random from the population fails on or before time t is equal to $\sum_{i=1}^{s} p_i F_i(t)$ or,

$$F(t) = \sum_{i=1}^{s} p_i F_i(t)$$

Hence the density function for the population is

$$f(t) = \sum_{i=1}^{s} p_i f_i(t)$$

A sample of n units is drawn from the population and placed on test. The test is terminated at a fixed time, T, at which time r units have failed, r_i from sub-population (i), where $\sum_{i=1}^{s} r_i = r$, and $n - r$ units have not failed. The time of failure of the jth unit from sub-population (i), t_{ij}, $i = 1, 2, \ldots s$ and $j = 1, 2, \ldots r_i$, is observed. The $n - r$ units which have not failed yield no information as to the sub-population from which they were drawn. Note that r, r_i, and t_{ij}, $i = 1, 2, \ldots s$, $j = 1, 2, \ldots r_i$, are random variables.

This study is primarily devoted to the special case of two sub-populations with a probability density function as follows:

$$f(t) = p\frac{e^{-t/\alpha_1}}{\alpha_1} + \frac{(1-p)e^{-t/\alpha_2}}{\alpha_2} \quad \text{where } 0 \leq t < \infty$$

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

It was convenient to measure time in units of size T, the termination time. Hence, letting $x = \frac{t}{T}$, the probability density function becomes:

$$f(x) = p\frac{T}{\alpha_1} e^{-\frac{T}{\alpha_1}x} + (1-p)\frac{T}{\alpha_2} e^{-\frac{T}{\alpha_2}x} \quad \text{where } 0 \leq x < \infty$$

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

Methods were obtained for the estimation of $\frac{\alpha_1}{T}$, $\frac{\alpha_2}{T}$, and p where $\frac{\alpha_1}{T}$ and $\frac{\alpha_2}{T}$ are the average lives of units from sub-populations I and II respectively.

Chapter II

REVIEW OF LITERATURE

Literature related to this research is concentrated in three areas, namely, papers concerned with the life testing of mixed failure populations, those concerned with discussions of the conditional failure density, and papers concerned with estimation in the case of censored sampling and in the case of sampling from a truncated distribution.

The only literature specifically directed to estimation of parameters in the censored life testing of mixed populations was included in an example in Weibull's paper (1951). Weibull's primary interest was in presenting the derivation and use of his now famous Weibull failure distribution. However, he used a graphic method, involving trial and error, to separate two Weibull distributions.

A group of papers by Acheson and McElwee (1951), Steen (1952), Wilde (1952), Davies (1952), Epstein (1953), Herd (1953), and Madison (1955) contained the notion of the subdivision of failure into failure types and presented failure rate curves for some mixed failure populations.

Acheson and McElwee (1951) discussed the reliability of electronic tubes and presented failure data divided into three categories, (1) failures due to mechanical defects, (2) failures due to gas or air, (3) and failures due to the deterioration of performance characteristics. A graph was presented, showing rate of failure per hundred hours versus age for subminiature tubes, indicating a falling and then rising failure rate. The graph also indicated the fraction failing due to each cause as a function of time.

A similar paper by Wilde (1952), related to life testing of electronic tubes, subdivided failures by means of several methods and presented a rate of failure curve for subminiature tubes similar to that obtained by Acheson and McElwee (1951). The author stated finally that consideration of the rate of failure curve for both cathode ray tubes and miniature tubes leads one to say that the true density function is the sum of two density functions, an exponential density function representing early failures and a Gamma function representing late failures.

Epstein (1953), in discussing the validity of the exponential distribution as a failure model, mentioned the fact that some sets of exponential failure data were obtained after eliminating a large number of early failures, indicating the possibility of two kinds of failure.

Herd (1953) recognized the need for a model for mixed failure distributions. He used the same model as is considered in this dissertation,

$$f(t) = \sum_{i=1}^{k} \frac{p_i}{\alpha_i} e^{-t/\alpha_i} \quad \text{where} \quad \sum_{i=1}^{k} p_i = 1,$$

for k mixed exponentially distributed sub-populations. The conditional failure density,

$$z(t) = \frac{1}{\alpha_1} + \sum_{r=2}^{k} \frac{p_r\left(\frac{1}{\alpha_r} - \frac{1}{\alpha_i}\right)}{\sum_{i=1}^{k} p_i e^{-t\left(\frac{1}{\alpha_i} - \frac{1}{\alpha_r}\right)}}$$

was stated to be a monotonic decreasing function with bounds,

$$\frac{1}{\alpha_1} \leq z(t) \leq \sum_{i=1}^{k} \frac{p_i}{\alpha_i}$$

where, without loss of generality, it was assumed that $\alpha_i > \alpha_{i+1}$ for $i = 1, 2, \ldots, k - 1$.

Madison (1955), in evaluating the performance of electronic equipment, divided failures into two sub-populations which he designated as catastrophic failures and degradation failures. He suggested using a weighted linear combination of the sub-population density functions as the density function of the total population. This approach was used in the present research.

Davis (1952), in describing the analysis of some failure data, directed attention to the use of the failure rate and its theoretical counterpart, the conditional failure density, as a means of describing the failure characteristics of a product. Davis stated the mathematical relation between the cumulative failure distribution and the conditional failure density and graphically portrayed these functions for the normal distribution, the exponential distribution, and the human mortality distribution. Gunther (1956) derived the conditional failure density for most of the common failure distributions and obtained functions which were monotonic. It is interesting to note that the observed failure rate curves of the mixed populations of Acheson and McElwee (1951) and Wilde (1952) are non-monotonic.

Numerous papers have been written on the subject of estimation of parameters based on censored samples and sampling from truncated distributions. Hald (1949) considered the maximum likelihood estimation of the mean and standard deviation of a normal distribution for both the case of censored sampling and the case of sampling from a truncated distribution.

Cohen (1950) obtained maximum likelihood estimates for singly and doubly truncated normal distributions under censored sampling, where the censoring occurred at a fixed time. The solutions to the maximum likelihood equations were presented in a form amenable to the use of the standard tables of the normal distribution. Cohen (1951) obtained estimates of the parameters in truncated Pearson frequency distributions by substituting sample moments for the population moments in equations relating the moments and the parameters. Cohen states that these estimates are consistent and satisfactory as rough estimates or first approximations for iteration to obtain the maximum likelihood estimates.

Gupta (1952) considered the problem of estimation for a censored sample from a normal distribution where censoring occurs after a fixed number of units have failed. Tables were given to obtain the maximum likelihood estimates and their asymptotic variances and covariances. Due to the fact that the properties of the maximum likelihood estimates are unknown for small samples, the best linear estimates of the mean, α, and standard deviation, β, were obtained by utilizing the extended principles of least squares and minimizing

$$D = (X - BZ)'V^{-1}(X - BZ)$$

where the observations are the ordered variates $x_1 < x_2 < \cdots < x_r$, $r \leq n$, from a sample of size n.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_r \end{bmatrix}, \quad B = \begin{bmatrix} 1 & \alpha_1 \\ 1 & \alpha_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & \alpha_r \end{bmatrix}, \quad Z = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

The expected value of the ith order statistic from a normal population with mean equal to zero and variance equal to 1 was defined as α_i and the vector X was assumed to have a variance-covariance matrix $\beta^2 V$. The estimators are then

$$\hat{\alpha} = \sum_{i=1}^{r} b_i x_i$$

$$\hat{\beta} = \sum_{i=1}^{r} c_i x_i$$

The coefficients b_i and c_i were tabulated for samples up to size $n = 3$ to 10 and $r = 2$ to $n - 1$ and the variances and covariances given in each case. An alternative less efficient linear estimate was given for $n = 10$ by assuming that V is a unit matrix.

Sarhan and Greenberg (1956) extended the work of Gupta by obtaining the best linear unbiased estimates of the mean and the standard deviation β of a normal distribution when the sample is doubly censored. A doubly censored sample is one in which the smallest k_1 and the largest k_2 observations are missing. The necessary coefficients of the r observed ordered observations are given for samples of size $n \leq 10$.

Epstein and Sobel (1953) presented an estimator and a test based on the first r observations drawn from an exponential distribution under censored sampling. These results and more general results on life test estimation procedures for the exponential distribution were included in a paper by Epstein (1954) in which censored samples, both for a fixed termination time, T, and a fixed number of observed failures, r, with and without replacement were considered. Estimates in all four cases are a function of the total observed life. For samples of size n when r is fixed,

$$\text{Total life} = R(t) = \sum_{i=1}^{r} t_i + (n - r)t_r,$$

the maximum likelihood estimate of the average life α is

$$\hat{\alpha} = \frac{R(t)}{r}$$

The estimator, $\hat{\alpha}$, is unbiased with variance, $V(\hat{\alpha}) = \frac{\alpha^2}{r}$. The estimate $\hat{\alpha}$ is a minimum variance unbiased estimate of α and $\frac{2r\hat{\alpha}}{\alpha}$ is distributed as a Chi-Square with 2r degrees of freedom. In fact, $\hat{\alpha}_{r,n}$, based on r failures out of a sample of size n, has exactly the same distribution as $\hat{\alpha}_{r,r}$ where a sample of size r is tested to failure. In the replacement case,

$$R(t) = nt_r$$

$$\hat{\alpha} = \frac{nt_r}{r} .$$

The estimator, $\hat{\alpha}$, is unbiased with variance, $V(\hat{\alpha}) = \frac{\alpha^2}{r}$. When the sample is censored at a fixed time, T, and r is a random variable,

$$\text{Total life} = R(t) = \sum_{i=1}^{r} t_i + (n - r)T;$$

the maximum likelihood estimate of α is

$$\hat{\alpha} = \frac{R(t)}{r} .$$

In the replacement case,

$$R(t) = nT$$

and

$$\hat{\alpha} = \frac{nT}{r}$$

where r is the total number of items failing by time T. The estimate $\hat{\alpha}$, in both the replacement and non-replacement case, where censoring occurs at a fixed time T, possesses the properties of a maximum likelihood estimate but is not minimum variance unbiased. If r is large, the bias is negligible and $\frac{2R(t)}{\alpha}$ is approximately distributed as Chi-Square with 2r degrees of freedom. Epstein and Sobel (1954) investigated the properties of the maximum likelihood estimates of α when r is fixed. Various assumptions are made concerning the time of birth, A_i, for an ith subset of the sample from a distribution having the density function

$$f(x;\ \alpha,\ A) = \frac{1}{\alpha} e^{-\frac{(x-A)}{\alpha}} \qquad A \leq x < \infty$$
$$= 0 \qquad \text{otherwise.}$$

Deemer and Votaw (1955) obtained the maximum likelihood estimates for the parameter, c, where the assumed failure distribution is the exponential

$$f(x) = ce^{-cx} \qquad 0 \leq x < \infty$$
$$= 0 \qquad \text{otherwise.}$$

Estimation was considered for both the case of censored sampling with a fixed termination time and also the case of sampling from a truncated distribution. The values of the estimates in the truncated case were presented in tabular form as a function of $\bar{x}$ and the asymptotic variances compared for various values of cx_o, where x_o is the termination time, for both the truncated and censored cases.

The problem of estimation in the case of censored sampling was further generalized by Herd (1956) by considering multi-censored samples in which k_i units are assumed removed from the test at the time of the ith ordered failure and where the test is terminated when a fixed number of units, r, have failed. The likelihood for this general case, where F(x) and f(x) are the failure cumulative distribution function and density function, respectively, is

$$f(x_1, x_2, \ldots x_r; \theta_1, \theta_2, \ldots \theta_m) = \prod_{i=1}^{r} \left\{ n_i' \, f(x_i) \left[1 - F(x_i) \right]^{k_i} \right\}$$

where n = sample size, $n_i' = (n - \sum_{j=1}^{i-1} k_j - i + 1)$,

$k_0 = 0$, $x_0 = 0$, $n - r = \sum_{j=1}^{r} k_j$, and $\theta_1, \theta_2, \ldots, \theta_m$ are m unknown parameters of the distribution. Maximum likelihood estimates of the θ's were obtained for the exponential distribution, the normal distribution, and the Gamma distribution. Due to the fact that in many cases the maximum likelihood equations are difficult to solve, a second method of estimation was presented which is nonparametric and based on the use of quantiles. Note that in the case where $k_i = 0$, $i = 1, 2, \ldots, r - 1$, Herd's case reduces to the single censored case.

Chapter III

CHARACTERIZATION OF FAILURE

Cumulative Failure Distribution Function

The most common method of characterizing failure is by means of the cumulative failure distribution function, $F(t)$, and the associated probability density function $f(t)$. $F(t)$ is defined as the probability that a unit placed on test will fail on or before time t. If $F(t)$ is a continuous function, which it usually is in life testing, $dF(t) = f(t)dt$. Hypothetically, if the total population were placed on test, the fraction of the total population failing on or before time t would equal $F(t)$ and the fraction failing in the interval t to $t + dt$ would be $f(t)dt$. The density function has popular appeal in that it is proportional to the mathematical function which would pass through a histogram of failures.

Conditional Failure Density

Engineers frequently characterize failure by means of the failure rate per unit time. This represents the rate of failure over a particular time interval, given that the units have survived up to the time interval. It might be argued that the failure rate is actually the ideal function with which to characterize failure in that its value at a particular time is a measure of susceptibility to failure. Hence a unit with a rising failure rate is one in which the susceptibility to failure increases as a function of time. In the case where wearout is evident, this susceptibility would be a measure of fatigue.

The theoretical function which models the failure rate of a product is the conditional failure density, z(t), defined as the probability that a unit fails in the interval t to t + dt, given that the unit has survived to time t. The equation

$$z(t) = \frac{f(t)}{1 - F(t)}$$

establishes the relation between the conditional failure density and the cumulative failure distribution and shows that one is simply a transformation of the other. It is easy to derive the conditional failure densities corresponding to the common failure distributions as is shown by Gunther (1956).

Distribution	f(t)	z(t)
Exponential	$f(t) = \frac{1}{\alpha} e^{-t/\alpha}$	$z(t) = \frac{1}{\alpha}$
Weibull	$f(t) = \frac{m}{\alpha}(\frac{t}{\alpha})^{m-1} e^{-(t/\alpha)^m}$	$z(t) = \frac{m}{\alpha}(\frac{t}{\alpha})^{m-1}$
Gamma	$f(t) = \frac{e^{-t/\alpha}(t/\alpha)^{r-1}}{\alpha(r-1)!}$	$z(t) = \frac{\frac{1}{\alpha(r-1)!}(t/\alpha)^{r-1}}{1+t/\alpha+\ldots+\frac{1}{(r-1)!}(t/\alpha)^{r-1}}$
Normal	$f(t) = \frac{1}{\sqrt{2\pi}\,\beta} e^{\frac{-(t-\alpha)^2}{2\beta^2}}$	$z(t) = \frac{e^{\frac{-(t-\alpha)^2}{2\beta^2}}}{\sqrt{2\pi}\,\beta - \int_{-\infty}^{t} e^{\frac{-(t-\alpha)^2}{2\beta^2}} dt}$

It will be noted that z(t), the conditional failure density, is monotonic for the distributions shown above.

<u>Characterization of Failure for Mixed Populations</u>

The conditional failure density is probably the most useful means of describing the failure characteristics of a mixed population because a close look at its form gives insight into the peculiar behavior of failure rates involving mixtures.

Consider the case of s sub-populations for the population as defined in Chapter I. A new quantity is now introduced called the conditional mixture proportion,

$$p_i(t) = p_i \frac{G_i(t)}{\sum_{j=1}^{s} p_j G_j(t)} \qquad i = 1, 2, \ldots s$$

where $$G_i(t) = 1 - F_i(t)$$

and $$\sum_{i=1}^{s} p_i(t) = 1$$

If it were possible to place the entire population on test, then at time t, the s sub-populations would be mixed in proportion $p_1(t)$: $p_2(t)$: ....: $p_s(t)$ where $0 \leq p_i(t) \leq 1$ and $\sum_{i=1}^{s} p_i(t) = 1$. If at time t, a unit were chosen at random from the population on test, the probability that the unit would belong to sub-population (i) would be $p_i(t)$. Obviously,

$$p_i = p_i(0).$$

The probability that a unit is from population i and fails in the interval t to t + dt, given that it has survived to time t, is $p_i(t)z_i(t)$. These events are mutually exclusive for $i = 1, 2, \ldots s$.

Hence,

$$z(t) = \sum_{i=1}^{s} p_i(t)z_i(t).$$

For the special case under consideration in this paper, s = 2, define $p_1(t) = p(t)$ and $p_2(t) = 1 - p(t)$ and let the failures in sub-population (i) be distributed with a cumulative failure distribution,

$$F_i(t) = 1 - e^{-t/\alpha_i} \qquad i = 1, 2.$$

Then

$$z(t) = p(t)\left(\frac{1}{\alpha_1}\right) + \left[1 - p(t)\right]\left(\frac{1}{\alpha_2}\right)$$

or

$$z(t) = \frac{1}{\alpha_2} + p(t)\left[\frac{1}{\alpha_1} - \frac{1}{\alpha_2}\right].$$

Hence the conditional failure density for two exponentially distributed sub-populations has exactly the same form as p(t). The conditional mixture proportion for this case is:

$$p(t) = \frac{pe^{-t/\alpha_1}}{pe^{-t/\alpha_1} + qe^{-t/\alpha_2}}$$

$$= \frac{1}{1 + \frac{q}{p}e^{t/\alpha_1 - t/\alpha_2}}$$

where $q = 1 - p$ and $p = p_1(0)$.

Thus a mixture of two sub-populations, each distributed as an exponential with a constant conditional failure density, yields a conditional failure density which is monotonically decreasing and which asymptotically approaches the smaller of $\frac{1}{\alpha_1}$ or $\frac{1}{\alpha_2}$ as $t \to \infty$. If $\alpha_1 = \alpha_2 = \alpha$, then $z(t) = \frac{1}{\alpha}$, yielding the conditional failure density of the exponential. Figure 1 is a graph of the conditional failure density for a mixture of two exponentials,

$$z(t) = \frac{1}{\alpha_2} + \frac{\frac{1}{\alpha_1} - \frac{1}{\alpha_2}}{1 + \frac{q}{p} e^{t/\alpha_1 - t/\alpha_2}}$$

where $\alpha_1 = 100$, $\alpha_2 = 1000$, and $p_1 = p = 1/3$.

The conditional failure density, $z(t)$, for two exponential distributions can be shown to be monotonically decreasing function of t if the first derivative of $z(t)$, with respect to t, is negative for all values of t. It is easily seen that

$$\frac{dz(t)}{dt} = -\left[\frac{1}{\alpha_1} - \frac{1}{\alpha_2}\right]^2 p(t)^2 e^{t(1/\alpha_1 - 1/\alpha_2)}$$

is always negative.

A non-monotonic conditional failure density can be obtained by considering a mixture of two sub-populations where sub-population (1) is distributed as a Weibull function,

$$F_1(t) = 1 - e^{-(t/\alpha_1)^{m_1}}$$

and sub-population (2) is exponentially distributed. The conditional failure density is

$$z(t) = \frac{1}{1 + \frac{q}{p} e^{(t/\alpha_1)^{m_1} - t/\alpha_2}} \left[\frac{m_1}{\alpha_1^{m}} t^{m_1 - 1} - \frac{1}{\alpha_2} \right] + \frac{1}{\alpha_2}$$

The form of z(t) is not obvious. In the special case where $m_1 = 2$, $\alpha_1 = 100$, $\alpha_2 = 200$, and $p = .2$, the conditional failure density is

$$z(t) = \frac{1}{1 + 4e^{(\frac{t}{100})^2 - \frac{t}{200}}} \left[\frac{2t}{(100)^2} - \frac{1}{200} \right] + \frac{1}{200} \;.$$

A graph of this function, shown in Figure 2, indicates that z(t) is a non-monotonic function, rising at first and then asymptotically approaching $z(t) = .005$ as $t \longrightarrow \infty$.

Figure 1

Conditional Failure Density for a Mixture of Two Exponentials where $\alpha_1 = 100$, $\alpha_2 = 1000$, and $p = 1/3$

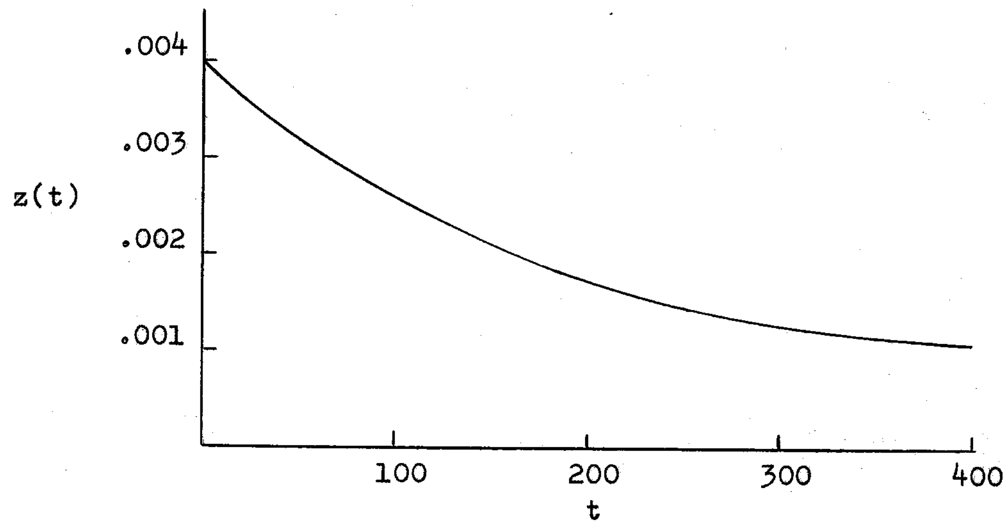
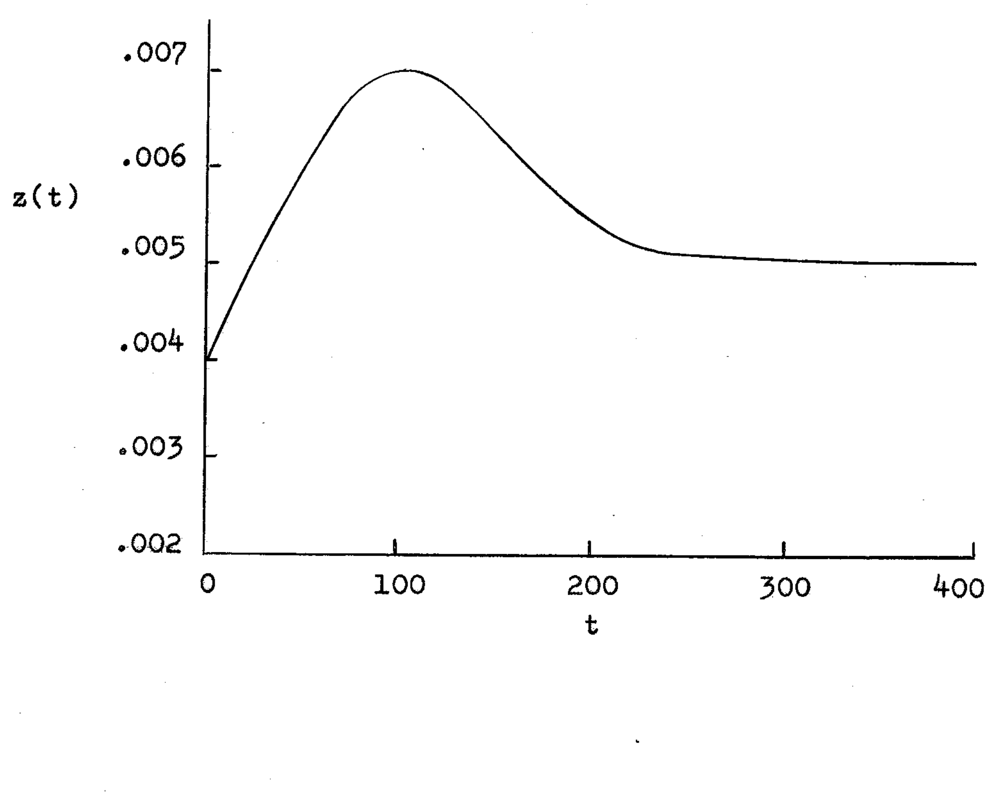


Figure 2

Conditional Failure Density for a Mixture of an Exponential and a Weibull Distribution

Sub-population (1): Weibull Distribution $\alpha_1 = 100$, $m_1 = 2$

Sub-population (2): Exponential Distribution $\alpha_2 = 200$, $p = .2$



Chapter IV

ESTIMATION IN THE CASE OF TWO MIXED EXPONENTIALLY DISTRIBUTED FAILURE POPULATIONS

Estimates Obtained by the Method of Maximum Likelihood

Consider sampling as stated in Chapter I. That is, from n units placed on test, r units are observed to fail before time T, r_i belonging to population (i), and the ordered failure times, t_{ij}, are recorded, $j = 1, 2, \ldots, r_i$, $i = 1, 2$. The density function for the ith sub-population is defined as $f_i(t) = \frac{1}{\alpha_i} e^{-t/\alpha_i}$ and that for the total population is therefore

$$f(t) = \frac{p}{\alpha_1} e^{-t/\alpha_1} + \frac{(1-p)}{\alpha_2} e^{-t/\alpha_2}; \quad 0 \leq t < \infty$$

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

where $p = p_1(0)$.

Let $x = \frac{t}{T}$ and $\beta_i = \frac{\alpha_i}{T}$. Then all measurements will be in units of size T and

$$f(x) = \frac{p}{\beta_1} e^{-x/\beta_1} + \frac{(1-p)}{\beta_2} e^{-x/\beta_2} \qquad 0 \leq x < \infty$$

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

$$F(x) = \int_0^x f(x)dx$$

$$G(x) = 1 - F(x).$$

The likelihood,

$$L = L(x_{11}, x_{12}, \dots x_{1r_1}, x_{21}, x_{22}, \dots, x_{2r_2}, r_1, r_2)$$

$$L = \frac{n!}{(n-r)!} \left[G(1)\right]^{n-r} p^{r_1} q^{r_2} \prod_{i=1}^{r_1} f_1(x_i) \prod_{i=1}^{r_2} f_2(x_i)$$

where $r_1 + r_2 = r$, $q = 1 - p$, and $G(1) = pe^{-1/\beta_1} + qe^{-1/\beta_2}$.

$$\ln L = \ln \frac{n!}{(n-r)!} + (n - r) \ln G(1) + r_1 \ln (p)$$

$$+ r_2 \ln (q) - r_1 \ln \beta_1 - \frac{r_1 \bar{x}_1}{\beta_1} - r_2 \ln \beta_2 - \frac{r_2 \bar{x}_2}{\beta_2} .$$

Taking the first partial derivatives of ln L,

$$\frac{\partial \ln L}{\partial \beta_1} = \frac{k(n-r)}{\beta_1^2} - \frac{r_1}{\beta_1} + \frac{r_1 \bar{x}_1}{\beta_1^2}$$

$$\frac{\partial \ln L}{\partial \beta_2} = \frac{(1-k)(n-r)}{\beta_2^2} - \frac{r_2}{\beta_2} + \frac{r_2 \bar{x}_2}{\beta_2^2}$$

$$\frac{\partial \ln L}{\partial p} = \frac{k(n-r)+r_1}{p} - \frac{(1-k)(n-r)+r_2}{q}$$

where

$$k = \frac{pe^{-1/\beta_1}}{pe^{-1/\beta_1} + qe^{-1/\beta_2}} = p(1)$$

$$= \frac{1}{1 + \frac{q}{p} e^{(1/\beta_1 - 1/\beta_2)}} .$$

Setting the partial derivatives equal to zero, we obtain

$$\beta_1 = \bar{x}_1 + \frac{k(n-r)}{r_1}$$

$$\beta_2 = \bar{x}_2 + \frac{(1-k)(n-r)}{r_2}$$

$$p = \frac{r_1}{n} + \frac{k(n-r)}{n} .$$

The similarity between these results and those for the single exponential distribution is fairly obvious. The estimate of average life for the single sub-population is $\frac{1}{r_i}$ times the total observed life. However, in the mixed case, the life for the n - r surviving units must be divided between the two sub-populations since it is impossible to tell from which sub-population a unit was drawn until after failure has occurred. The best estimate of the fraction of those surviving which belong to population (1), at time T, is the conditional mixture proportion, p(t = T), defined in Chapter III. It will be observed that p(t = T) = k. Hence the estimated total life for sub-population (1),

$$\sum_{j=1}^{r_1} x_{ij} + k(n-r),$$

is equal to the sum of the observed failure times for sub-population (1) plus the expected fraction of the observed life for those surviving. Similarly, the estimated total life for sub-population (2) is

$$\sum_{j=1}^{r_2} x_{2j} + (1-k)(n-r).$$

The estimates of β_1, β_2, and p must be obtained by iterative methods. Substituting into the equation relating k, β_1, β_2, and p produces

$$k = \frac{1}{1 + \frac{\hat{q}}{\hat{p}} e^{(1/\hat{\beta}_1 - 1/\hat{\beta}_2)}}$$

or

$$k = g(k).$$

Since k is bounded, $0 \leq k \leq 1$, it is relatively easy to obtain k by considering $g(k) - k$ versus k and obtaining the solution where $g(k) - k = 0$. The function $g(k) - k$ will be positive or zero when $k = 0$.

A good first approximation to k can be obtained by using the maximum likelihood estimate obtained by Deemer and Votaw (1955) for the case of samples drawn from a truncated exponential distribution. Their table is not very useful in this case since since they gave values of $\frac{1}{\beta_i}$ as a function of $\bar{x}_i$, it being easier to calculate the tables directly rather than to convert the values of $\frac{1}{\beta_i}$ which were given. The maximum likelihood estimate of β_i, where the sample is assumed to be truncated at time T, is the solution of

$$(\beta_i - \bar{x}_i)(e^{1/\beta_i} - 1) = 1.$$

The solutions, $\hat{\beta}_i$, can be obtained graphically from Figure 3 where $\hat{\beta}_i$ is given as a function of $\bar{x}_i$. Choose the smaller $\bar{x}$ and identify this as sub-population (1). Obtain the corresponding $\hat{\beta}_{10}$ from Figure 3. Then, substituting into

$$\hat{\beta}_{10} = \bar{x}_1 + k_0 \frac{(n-r)}{r_1} ,$$

solve for k_0. Using k_0, calculate $\hat{\beta}_{20}$ and $\hat{p}_0$ from

$$\hat{\beta}_{20} = \bar{x}_2 + (1-k_0)\frac{(n-r)}{r_2}$$

and

$$\hat{p}_0 = \frac{r_1}{n} + k_0 \frac{(n-r)}{n} .$$

The first approximations, $\hat{\beta}_{10}$, $\hat{\beta}_{20}$, and $\hat{p}_0$ are then used to determine $g(k_0)$ directly from

$$g(k_0) = \frac{1}{1 + \frac{\hat{q}_0}{\hat{p}_0} e^{(1/\hat{\beta}_{10} - 1/\hat{\beta}_{20})}} .$$

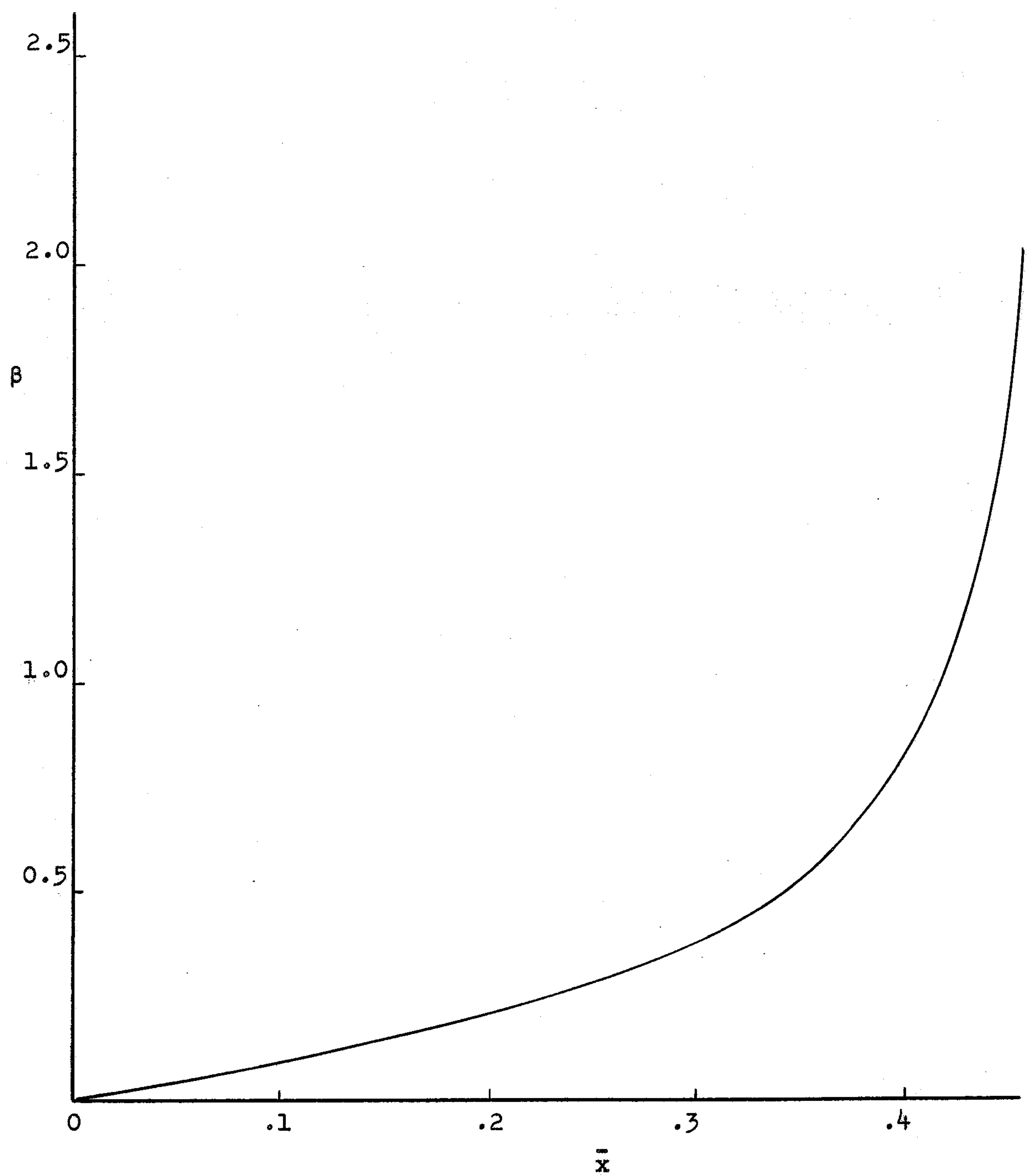
Calculate $g(k_0) - k_0$. Since $g(0)$ is greater than or equal to zero, the value of k which satisfies

$$g(k) - k = 0$$

must be $k < k_0$ or $k > k_0$ depending upon whether $g(k_0) - k_0$ is negative or positive.

Figure 3

Maximum Likelihood Estimate of β as a Function of $\bar{x}$
Based on a Sample from a Truncated Exponential Distribution
Measurements Expressed in Units of Truncation Time T



The first iteration involves the choice of k_1. Let

$$v = \frac{q}{p} e^{(1/\beta_1 - 1/\beta_2)}$$

and

$$D = g(k) - k$$

$$= \frac{1}{1 + v} - k.$$

Then

$$\frac{dD}{dk} = \frac{dg(v)}{dv} \frac{dv}{dk} - 1$$

and

$$dk = \frac{dD}{\frac{dg(v)}{dv} \frac{dv}{dk} - 1}$$

$$= - \frac{dD}{1 + \frac{1}{(1 + v)^2} \frac{dv}{dk}}$$

$$= - \frac{dD}{1 + g(k)^2 \frac{dv}{dk}} .$$

The quantity, $\frac{dv}{dk}$, can be obtained by taking the derivative of ln(v) with respect to k.

$$\ln(v) = \ln(q) - \ln(p) + \frac{1}{\beta_1} - \frac{1}{\beta_2}$$

$$\frac{dv}{dk} = v \left[- \frac{1}{q} \frac{dp}{dk} - \frac{1}{p} \frac{dp}{dk} - \frac{1}{\beta_1^2} \frac{d_{\beta_1}}{dk} + \frac{1}{\beta_2^2} \frac{d_{\beta_2}}{dk} \right]$$

where

$$v = \frac{q}{p} e^{(1/\beta_1 - 1/\beta_2)}$$

$$\frac{dp}{dk} = \frac{(n - r)}{n}$$

$$\frac{d\beta_1}{dk} = \frac{(n - r)}{r_1}$$

$$\frac{d\beta_2}{dk} = - \frac{(n - r)}{r_2}$$

have been previously calculated. It would seem reasonable to choose the value of dk_0 which would correspond to $dD_0 = - \left[g(k_0) - k_0 \right]$. Therefore,

$$dk_0 \approx \frac{g(k_0) - k_0}{1 + g(k_0)^2 \frac{dv_0}{dk_0}}$$

and $$k_1 = k_0 + dk_0.$$

The corresponding values of $\hat{\beta}_{11}$, $\hat{\beta}_{21}$, $\hat{p}_1$, $g(k_1)$, and $g(k_1)-k_1$ can now be computed. Linear interpolation usually will be satisfactory for obtaining k_2. The iterative process can be repeated until k is obtained to the desired degree of accuracy. An upper limit to the iterative error can be obtained since the solution will lie in an interval between two values of k, say k_i and k_j, such that $g(k_i)-k_i$ and $g(k_j)-k_j$ differ in sign.

The method of solving the likelihood equations, described above, is utilized in the following example.

An Example:

The data recorded in Tables 2 and 3, supplied by Dr. G. R. Herd of Aeronautical Radio, Inc., are times to failure taken from twelve ARC-1 VHF communication transmitter receivers of a single commercial airline. The data were collected over the calendar period from November, 1952 to January, 1954. Units which failed were removed from the aircraft for maintenance. However, in some cases the failures were unconfirmed, exhibiting satisfactory operation upon arrival at the maintenance center. Practical considerations make it desirable to estimate the fraction of unconfirmed failures in the population. Hence the sample of failures may be subdivided into confirmed failures, shown in Table 2, and unconfirmed failures shown in Table 3. The sample was censored at $T = 630$ hours as it was a general policy of the airline to remove units which had operated for 630 hours. Histograms plotted for both confirmed failures and unconfirmed failures suggest that both sub-populations of failures are exponentially distributed.

Considering unconfirmed failures as sub-population (1) and confirmed failures as sub-population (2), the data from Tables 2 and 3 yield the following:

$$n = 369$$

$$r_1 = 107$$

$$r_2 = 218$$

$$r = r_1 + r_2 = 325$$

$$n - r = 369 - 325 = 44$$

$$\bar{x}_1 = \frac{\bar{t}_1}{T} = .3034862$$

$$\bar{x}_2 = \frac{\bar{t}_2}{T} = .3644677.$$

The estimating equations are then

$$\hat{\beta}_1 = .3035 + .4112k$$

$$\hat{\beta}_2 = .5663 - .2018k$$

$$\hat{p} = .2900 + .1192k.$$

The process of obtaining the iterative solution is simplified by using a table similar to Table 1 shown below.

Table 1

Record of Iterations

i	k_i	$\hat{\beta}_{1i}$	$\hat{\beta}_{2i}$	$\hat{p}_i$	v	$g(k_i)$	$g(k_i)-k_i$
0	.186	.380	.529	.312	4.622	.1779	-.0081
1	.166	.3718	.5328	.3098	5.024	.1660	.0000
2	.167	.3721	.5326	.3099	5.002	.1666	-.0004
3	.165	.3713	.5330	.3097	5.046	.1654	.0004

The first step is to enter Figure 3 with $\bar{x}_1$ and obtain the first estimate of β_0, $\hat{\beta}_{10} = .380$. The corresponding value of k, $k_0 = .186$, can be obtained from the first estimating equation and then, utilizing k_0, $\hat{\beta}_{20}$ and $\hat{p}_0$ can be easily obtained. These values are shown in row i = 0 of Table 1.

The next step is to compute

$$g(k_0) = \frac{1}{1 + \frac{\hat{q}_0}{\hat{p}_0} e^{1/\hat{\beta}_{10} - 1/\hat{\beta}_{20}}} = .1779$$

and $g(k_0) - k_0 = -.0081$. This can easily be calculated on a desk calculator with the aid of tables of e^x. The value of k which corresponds to the solution of the maximum likelihood equations will occur when $g(k) - k = 0$. Since $g(k) - k$ is positive or zero when $k = 0$ and negative when $k = .186$, the solution for k must be $0 < k < .186$. Hence the value of k for i = 1 must be less than .186. The change in k, dk_0, can now be computed from

$$\frac{dv_0}{dk_0} = v_0 \left[- \frac{(n-r)}{nq_0} - \frac{(n-r)}{np_0} - \frac{(n-r)}{r_1\beta_{10}^2} - \frac{(n-r)}{r_2\beta_{20}^2} \right]$$

$$= (4.622) \left[- \frac{.1192}{.688} - \frac{.1192}{.312} - \frac{.4112}{6.92} - \frac{.2018}{3.57} \right]$$

$$= - 19.04$$

and

$$dk_0 \approx \frac{g(k_0) - k_0}{1 + g(k_0)^2 \frac{dv_0}{dk_0}}$$

$$\approx \frac{(-.0081)}{1 + (.1779)^2(-19.04)}$$

$$\approx -.02.$$

Hence

$$k_1 = k_0 + dk_0$$

$$= .186 - .02$$

$$= .166$$

$$\hat{\beta}_{11} = .3718$$

$$\hat{\beta}_{21} = .5328$$

$$\hat{p}_1 = .3098.$$

For all practical purposes, these estimates are the maximum likelihood estimates of the parameters since $g(k_1) - k_1 = .0000$. A bound on the iteration error can be obtained by calculating $g(k) - k$ for $k_2 = .167$ and $k_3 = .165$. Since $g(k_2) - k_2 = -.0004$ is negative and $g(k_3) - k_3 = .0004$ is positive and the solution for k is taken as .166, then clearly the absolute value of the iterative error for k is less than .001.

The estimate of the fraction of unconfirmed failures is $\hat{p} = .3098$ and their average life is estimated to be

$$\hat{\alpha}_1 = \hat{\beta}_1 T = (.3718)(630)$$

$$= 234.2 \text{ hours.}$$

The estimate of the average life of the confirmed failures is

$$\hat{\alpha}_2 = \hat{\beta}_2 T = (.5328)(630)$$
$$= 335.7.$$

It should be noted that a fairly accurate solution for the estimation equations was obtained in only two steps. All calculations, including those for the two boundary values, were made on a desk calculator in less than half an hour. An iterative scheme for an automatic computer can be programmed to deliver a much more accurate solution in a minute or less time.

It was previously mentioned that the parameter of primary interest is p. The estimates of the average life of units from the two sub-populations, $\hat{\alpha}_1$, and $\hat{\alpha}_2$, may be useful in the anticipating of maintenance requirements. In any case, this example represents an unusual and interesting application of the methods of estimation for mixed, exponentially distributed failure populations.

Table 2

Confirmed Failures

Hours to Failure for ARC-1 VHF Radio Transmitter Receivers*

16	224	16	80	128	168	144	176	176	568
392	576	128	56	112	160	384	600	40	416
408	384	256	246	184	440	64	104	168	408
304	16	72	8	88	160	48	168	80	512
208	194	136	224	32	504	40	120	320	48
256	216	168	184	144	224	488	304	40	160
488	120	208	32	112	288	336	256	40	296
60	208	440	104	528	384	264	360	80	96
360	232	40	112	120	32	56	280	104	168
56	72	64	40	480	152	48	56	328	192
168	168	114	280	128	416	392	160	144	208
96	536	400	80	40	112	160	104	224	336
616	224	40	32	192	126	392	288	248	120
328	464	448	616	168	112	448	296	328	56
80	72	56	608	144	408	16	560	144	612
80	16	424	264	256	528	56	256	112	544
552	72	184	240	128	40	600	96	24	184
272	152	328	480	96	296	592	400	8	280
72	168	40	152	488	480	40	576	392	552
112	288	168	352	160	272	320	80	296	248
184	264	96	224	592	176	256	344	360	184
152	208	160	176	72	584	144	176		

Table 3

Unconfirmed Failures

Hours to Failure for ARC-1 VHF Radio Transmitter Receivers*

368	136	512	136	472	96	144	112	104	104
344	246	72	80	312	24	128	304	16	320
560	168	120	616	24	176	16	24	32	232
32	112	56	184	40	256	160	456	48	24
200	72	168	288	112	80	584	368	272	208
144	208	114	480	114	392	120	48	104	272
64	112	96	64	360	136	168	176	256	112
104	272	320	8	440	224	280	8	56	216
120	256	104	104	8	304	240	88	248	472
304	88	200	392	168	72	40	88	176	216
152	184	400	424	88	152	184			

*Data supplied through the courtesy of Dr. G. R. Herd, Aeronautical Radio, Incorporated.

Chapter V

PROPERTIES OF THE ESTIMATES

The properties of the estimates of β_1, β_2, and p generated by the solution of the maximum likelihood equations, have been investigated by empirical sampling methods and also by obtaining the asymptotic variance-covariance matrix of the estimates.

Empirical Sampling Methods

The IBM 650 digital computer was utilized to generate empirical samples. Each sample of size n required the generation of n random numbers from a uniform distribution, the random numbers being randomly allocated in the ratio of p: 1-p to the two sub-populations. The number of units falling in the ith population was denoted by n_i, i = 1, 2, where $n_1 + n_2 = n$, $E(n_1) = np$ and $E(n_2) = n(1-p)$. The random numbers, y_{ij}, i = 1, 2, j = 1, 2, ... n_i, were then transformed to x_{ij} by the transformation

$$y_{ij} = F_i(x_{ij}) = 1 - e^{-\frac{x_{ij}}{\beta_i}} .$$

The x_{ij} are then exponentially distributed with average life, expressed in units of size T, equal to β_i for the ith sub-population. If $x_{ij} \leq 1$, the failure occurred prior to time T, the termination of the test. Hence each x_{ij} was placed in one of three mutually exclusive categories, namely a failure prior to time T belonging to sub-population (1), a failure prior to time T belonging to sub-population (2), and those which survived the test.

The survivors were a mixture from the two sub-populations, the number belonging to each sub-population assumed unknown. The number of units falling in each category, r_1, r_2, and (n-r) respectively, and the sums

$$\sum_{j=1}^{r_i} x_{ij},$$

when i = 1, 2 and $x_{ij} \leq 1$, were recorded.

The random numbers were generated in accordance with a method described by Moshman (1954). The jth random number generated is obtained from

$$y_j = Cy_{j-1} \pmod{10^{15}}$$

where $$C = 232630513987207$$

and $$y_0 = 1.$$

Mod 10^{15} means that only the last fifteen digits of the product, Cy_{j-1}, are retained. This method generates a repeating sequence of fifteen digit uniformly distributed random numbers with a period of $5(10^{12})$. The random allocation of the individual observations to the two sub-populations was based on the second and third high order digits of the random number. The sampling was conducted so that no part of the sequence of random numbers was used more than once, in order to avoid correlation of the estimates obtained from the samples.

Runs, comprising fifty samples, were made for each of twenty-five parameter points where a parameter point is specified by the parameters: n, β_1, β_2/β_1, and p. Estimates of the mean and variance of the estimates for each set of n, β_1, β_2, and p were then made on the basis of the fifty estimates obtained from each run. The parameter points were chosen with the intention of covering the regions in which the estimates were expected to be rather poor as well as regions where they were expected to be good. In practice a knowledge of the approximate values of p, β_1 and β_2 will enable the experimenter to choose the sample size n and the test termination time T large enough to avoid regions in which the estimation is known to be very poor.

The estimates for each sample were obtained by an iterative solution of the maximum likelihood equations. The method of iteration was essentially as described in Chapter IV, the solution for k being located in a region determined by a change in sign in $g(k) - k$. However, inasmuch as the computing was automatic, it was easier to choose successive values of k in a systematic manner. Since $g(0) \geq 0$, the computer was programmed to start at $k = .10$ and proceed towards $k = 1$ in intervals of .10 until $g(k) - k$ was observed to be negative in value. The solution is then known to fall in a specific interval of width equal to .10. The computer would then choose values of k starting at the beginning of this interval and proceed in steps of .01 until a change in sign in $g(k) - k$ was observed. The solution is then known to fall in a specific interval of width equal to .01. This procedure was repeated five times until the solution was isolated in an interval of width equal to .00001. The midpoint of this interval was chosen as the solution, producing an iteration error no greater than .000005 in absolute value.

The sample means and sample standard deviations, based on sampling experiments composed of N = 50 trials at each parameter point, are presented in Table 4 where

$$s_{\hat{\beta}_i} = \sqrt{\sum_{i=1}^{N} \frac{(\hat{\beta}_i - \bar{\hat{\beta}}_i)^2}{N - 1}}$$

and

$$s_{\hat{p}} = \sqrt{\sum_{i=1}^{N} \frac{(\hat{p} - \bar{\hat{p}})^2}{N - 1}} .$$

The estimated standard deviations of $\hat{\beta}_i$ and $\hat{p}$ are

$$\frac{s_{\hat{\beta}_i}}{\sqrt{N}} \quad \text{and} \quad \frac{s_{\hat{p}}}{\sqrt{N}}$$

and can be obtained approximately by dividing the values of $s_{\hat{\beta}_i}$ and $s_{\hat{p}}$ by seven. Table 4 also contains a listing of the parameter point identification numbers, the parameters, and the expected values of r_1 and r_2, $E(r_1)$ and $E(r_2)$ respectively, where

$$E(r_1) = npF_1(1)$$

$$= np(1 - e^{-1/\beta_1})$$

and

$$E(r_2) = nqF_2(1)$$

$$= nq(1 - e^{-1/\beta_2}).$$

Table 4

Estimated Means and Standard Deviations of the Estimates Based on 50 Samples at Each Parameter Point

Parameter Point	n	p	β_1	β_2/β_1	$E(r_1)$	$E(r_2)$	$\bar{\hat{p}}$	$s_{\hat{p}}$	$\bar{\hat{\beta}}_1$	$s_{\hat{\beta}_1}$	$\bar{\hat{\beta}}_2$	$s_{\hat{\beta}_2}$
1	100	.05	.6	3	4.1	40.6	.119	.1242	2.760	4.095	1.600	.414
2	100	.10	.4	2	9.2	64.2	.116	.0535	.679	.735	.800	.115
3	100	.10	.4	4	9.2	41.9	.120	.0829	.630	.829	1.578	.298
4	100	.10	.8	2	7.1	41.9	.147	.1166	1.589	1.708	1.550	.373
5	100	.10	.8	4	7.1	24.2	.195	.2080	2.380	3.611	2.993	1.097
6	100	.20	.2	3	19.9	64.9	.198	.0454	.219	.064	.584	.075
7	100	.20	.6	1	16.2	64.9	.208	.0592	.678	.373	.572	.098
8	100	.20	.6	3	16.2	34.2	.243	.1160	.734	.599	1.806	.444
9	100	.20	.6	5	16.2	22.6	.248	.1627	.899	.888	2.930	.912
10	100	.20	1.0	3	12.6	22.6	.280	.1623	1.646	1.329	2.655	1.006
11	100	.30	.4	2	27.5	49.9	.302	.0529	.439	.181	.807	.128
12	100	.30	.4	4	27.5	32.6	.297	.0483	.407	.129	1.704	.319
13	100	.30	.8	2	21.4	32.6	.329	.1227	.981	.618	1.453	.429
14	100	.30	.8	4	21.4	18.8	.394	.1729	1.219	.823	2.852	1.168
15	100	.35	.6	3	28.4	27.8	.382	.1071	.695	.406	1.727	.505
16	200	.10	.4	2	18.4	128.4	.105	.0245	.430	.204	.799	.057
17	200	.10	.8	4	14.2	48.4	.183	.1752	2.369	3.399	2.884	.840
18	200	.20	.6	3	32.4	68.4	.210	.0566	.660	.328	1.750	.235
19	200	.30	.4	2	55.0	99.8	.299	.0383	.386	.085	.803	.093
20	200	.30	.8	4	42.4	37.6	.369	.1468	1.110	.718	2.860	.893
21	50	.10	.4	2	4.6	32.1	.113	.0716	.650	.975	.788	.179
22	50	.10	.8	4	3.6	12.1	.224	.2291	2.677	5.500	3.008	1.279
23	50	.20	.6	3	8.1	17.1	.231	.1292	.957	1.115	1.654	.559
24	50	.30	.4	2	13.8	25.0	.311	.0668	.435	.274	.757	.212
25	50	.30	.8	4	10.7	9.4	.386	.1968	1.278	1.098	2.907	1.537

Appendix Table 9 presents the individual estimates obtained from the sampling experiment at parameter point 14 when n = 100, $\beta_1 = .8$, $\beta_2 = 3.2$, and $p = .3$. This parameter point was chosen to illustrate the behavior of the estimates in a region of the parameter space where estimation is poor, the estimates being badly biased and having large variances. An analysis of the estimates indicates that the excessive bias and large variance of the estimates are caused primarily by the estimates based upon samples 8, 15, 24, 26, 29, 33, 36, 41, and 48. These estimates have in common the fact that they are based upon values of the observed random variables where $\hat{\beta}_1 > \hat{\beta}_2$ even though $\beta_1 < \beta_2$. When $\hat{\beta}_1 > \hat{\beta}_2$, and $\beta_1 < \beta_2$ we will say that a crossover has occurred.

In many practical situations the experimenter knows that $\beta_1 < \beta_2$. If a crossover occurred, it would seem reasonable to choose $\hat{\beta}_1 = \hat{\beta}_2$ as an estimate. The maximum likelihood estimate of $\beta_1 = \beta_2 = \beta$ in this case is

$$\hat{\beta} = \frac{r_1\bar{x}_1 + r_2\bar{x}_2 + (n - r)}{r}$$

and

$$p = \frac{r_1}{r} .$$

Hence the rule of estimation will be to choose as estimates the solution of the original likelihood equations unless a crossover has occurred. If $\hat{\beta}_1 > \hat{\beta}_2$, assume that $\beta_1 = \beta_2 = \beta$ and obtain the adjusted estimates of β and p. This procedure will be called the adjusted estimation procedure to distinguish it from the procedure involving the solution of the original likelihood equations.

The latter will be called the ML procedure of estimation.

A comparison of the adjusted estimation procedure and the ML estimation procedure is presented in Tables 10 and 11. Table 10 lists the means of estimates of the parameters, based on N = 50 samples, for the two estimation procedures and Table 11 presents the corresponding estimated variances of the estimates. The symbol A is used to denote the number of crossovers and hence the number of adjusted estimates per group of N = 50 samples. The subscript A is used to identify the estimated means and variances for the adjusted procedure. A glance at these tables indicates that the estimates obtained by the adjusted procedure are less biased and have much smaller variances than those obtained by the ML procedure.

A clue to the behavior of the variance of the estimates as a function of n, p, β_1, and β_2 is afforded by examining the asymptotic variance of $\hat{\alpha}$ for a single exponentially distributed failure population.

$$V(\hat{\alpha}) = \frac{\alpha^2}{E(r)}$$

The parameter α is the average life and r is the number of items failing before time T. Since

$$\begin{aligned} E(r) &= nF(T) \\ &= n(1 - e^{-T/\alpha}), \end{aligned}$$

the asymptotic variance of $\hat{\alpha}$ for a single exponential distribution, with fixed parameter α, is

$$V(\hat{\alpha}) = \frac{\alpha^2}{nF(T)}$$

$$= \frac{\alpha^2}{n(1 - e^{-T/\alpha})} .$$

Obviously, for a fixed α, the asymptotic variance of $\hat{\alpha}$ decreases as n and T increase. Since $\beta = \frac{\alpha}{T}$, it follows that

$$V(\hat{\beta}) = \frac{1}{T^2} V(\hat{\alpha})$$

$$= \frac{\beta^2}{E(r)} .$$

It would seem likely that the variance of $\hat{\beta}_i$, in the case of mixed exponentially distributed sub-populations, is a complicated function of n, β_1, β_2, and p. However, it would appear that the relation,

$$V(\hat{\beta}_i) \approx \frac{C\beta_i^2}{E(r_i)} ,$$

where C is a constant of proportionality, holds reasonably well when n and T are large enough to provide a low frequency of cross-overs. This assumption is checked by comparing the approximate ratio of the variances of $\hat{\beta}_i$ at two different parameter points, say u and v,

$$\frac{V(\hat{\beta}_i)_u}{V(\hat{\beta}_i)_v} \approx \frac{\beta_{iu}^2}{\beta_{iv}^2} \frac{E(r_i)_v}{E(r_i)_u} = R_{u,v}$$

with the empirical ratio of the corresponding estimated variances, $s^2_{\hat{\beta}_{1u}}/s^2_{\hat{\beta}_{1v}}$. The ratio, $R_{u,v}$, is compared with the empirical ratio of the variances of $\hat{\beta}_1$ for the three parameter points, 6, 12, and 19, which contained no crossovers. The results are shown below in Table 5.

Table 5

A Comparison of $R_{u,v} \approx V(\hat{\beta}_1)_u/V(\hat{\beta}_1)_v$ with

the Empirical Ratio, $s^2_{\hat{\beta}_{1u}}/s^2_{\hat{\beta}_{1v}}$

u	v	$R_{u,v}$	$s^2_{\hat{\beta}_{1u}}/s^2_{\hat{\beta}_{1v}}$
6	12	.35	.24
6	19	.69	.57
19	12	.50	.41

The agreement between $R_{u,v}$ and the empirical ratio, of the estimated variances is good considering that the estimated variances are based on N = 50 samples.

A rough method of extrapolation from the empirical variances of the maximum likelihood estimates, given in Table 11, to variances at other parameter points would be to use the relation

$$V(\hat{\beta}_i)_u = R_{u,v}V(\hat{\beta}_i)_v$$

where it is assumed that $V(\hat{\beta}_i)_v$ is known and that

$$R_{u,v} = \frac{\beta_{iu}^2}{\beta_{iv}^2} \frac{E(r_i)_v}{E(r_i)_u} .$$

The quantities, $E(r_i)_u$ and $E(r_i)_v$ can easily be calculated since

$$E(r_1) = np(1 - e^{-1/\beta_1})$$

$$E(r_2) = nq(1 - e^{-1/\beta_2}) .$$

The results of Table 4 indicate that the estimates were fairly good at parameter point 6 where $n = 100$, $\beta_1 = .2$, $\beta_2 = .6$, and $p = .2$. Hence an additional 159 samples were drawn at this point, to make a total of 209, in order to shed some light on the approximate distributions of the estimates. These results are presented in the form of histograms in Figure 4. It will be noted that all three distributions are unimodal and skewed to the right although the skewness is moderate in the distribution of p. The estimated averages and variances of the estimates for the original $N = 50$ samples and for the total $N = 209$ samples are presented below in Table 6.

Figure 4

Histograms of Estimates at Parameter Point 6. N = 209

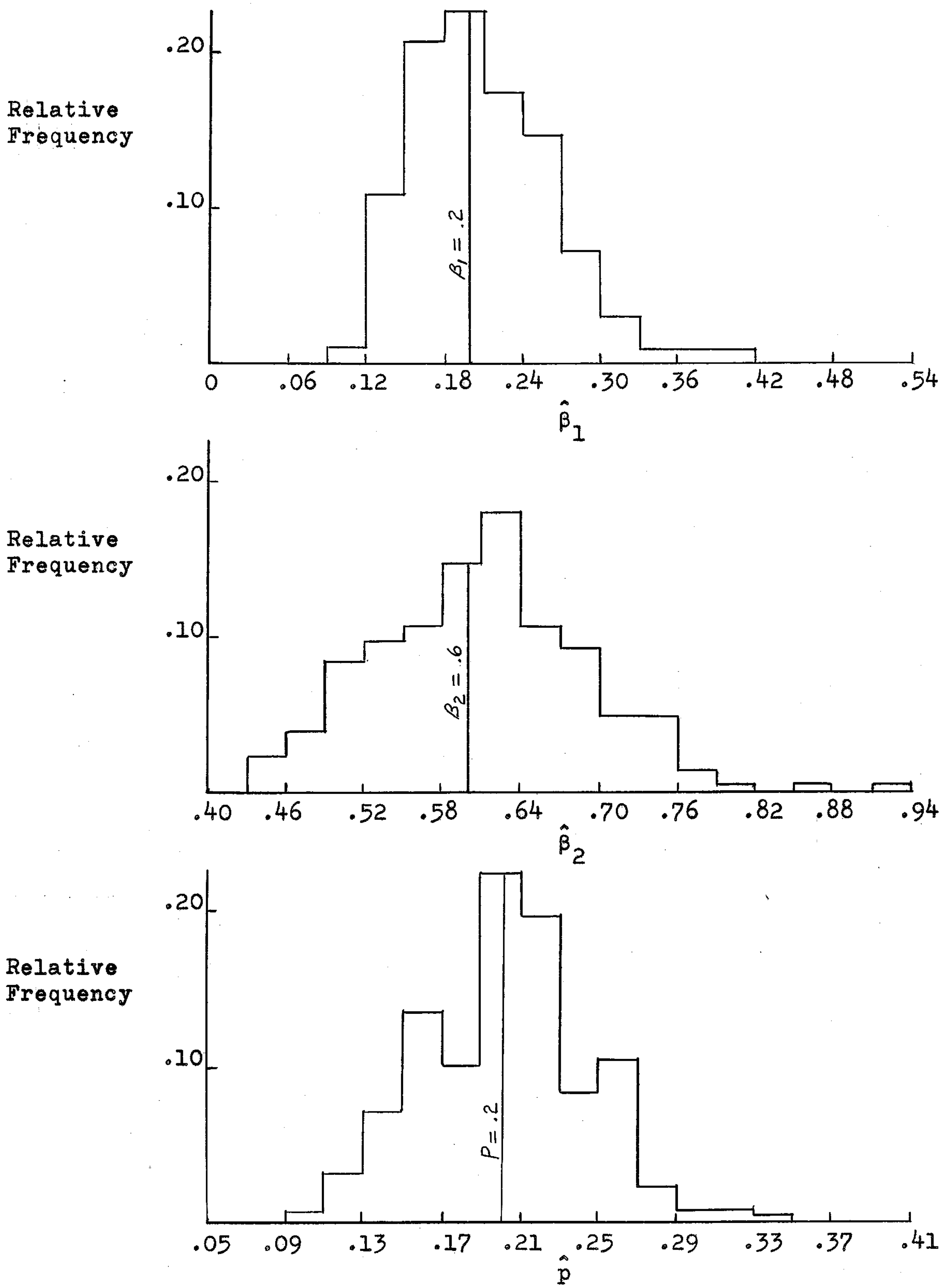


Table 6

Estimated Averages and Variances at Parameter

Point 6 for N = 50 and N = 209 Samples

$\beta_1 = .2 \quad \beta_2 = .6 \quad p = .2$

N	θ	$\hat{\theta}$	$s^2_{\hat{\theta}}$
50	β_1	.2193	.0040
	β_2	.5837	.0056
	p	.1978	.0021
209	β_1	.2096	.0030
	β_2	.6084	.0066
	p	.2008	.0019

Asymptotic Variances

The asymptotic variances and covariances of the estimates can be obtained by inverting the symmetric information matrix

$$I = \begin{bmatrix} -E\left\{\frac{\partial^2 \ln L}{\partial \beta_1^2}\right\} & -E\left\{\frac{\partial^2 \ln L}{\partial \beta_1 \partial \beta_2}\right\} & -E\left\{\frac{\partial^2 \ln L}{\partial \beta_1 \partial p}\right\} \\ & -E\left\{\frac{\partial^2 \ln L}{\partial \beta_2^2}\right\} & -E\left\{\frac{\partial^2 \ln L}{\partial \beta_2 \partial p}\right\} \\ & & -E\left\{\frac{\partial^2 \ln L}{\partial p^2}\right\} \end{bmatrix}$$

where the symbol E indicates the expected value of the quantity following.

The second partial derivatives of ln L are as follows:

$$\frac{\partial^2 \ln L}{\partial \beta_1^2} = \frac{(n - r)}{\beta_1^4}\left[k(1 - k) - 2k\beta_1\right] + \frac{r_1}{\beta_1^2} - \frac{2r_1\bar{x}_1}{\beta_1^3}$$

$$\frac{\partial^2 \ln L}{\partial \beta_2^2} = \frac{(n - r)}{\beta_2^4}\left[k(1 - k) - 2(1 - k)\beta_2\right] + \frac{r_2}{\beta_2^2} - \frac{2r_2\bar{x}_2}{\beta_2^3}$$

$$\frac{\partial^2 \ln L}{\partial p^2} = - \frac{(n - r)(k - p)^2}{p^2q^2} - \frac{r_1}{p^2} - \frac{r_2}{q^2}$$

$$\frac{\partial^2 \ln L}{\partial \beta_1\, \partial \beta_2} = - \frac{(n - r)k(1 - k)}{\beta_1^2\beta_2^2}$$

$$\frac{\partial^2 \ln L}{\partial \beta_1\, \partial p} = \frac{(n - r)k(1 - k)}{\beta_1^2 pq}$$

$$\frac{\partial^2 \ln L}{\partial \beta_2\, \partial p} = - \frac{(n - r)k(1 - k)}{\beta_2^2 pq} .$$

Utilizing the following expected values,

$$E(r_1) = npF_1(1)$$

$$E(r_2) = nqF_2(1)$$

$$E(n-r) = nG(1)$$

$$E(r_i\bar{x}_i) = E\left\{\sum_{j=1}^{r_i} x_{ij}\right\}$$

$$= \sum_{v=0}^{n} \text{Prob}\left\{r_i = v\right\} \quad E\left\{\sum_{j=1}^{v} x_{ij}\right\}$$

$$= \sum_{v=0}^{n} \text{Prob}\left\{r_i = v\right\} \; (v)E(x_i)$$

$$= \; E(r_i)E(x_i),$$

where $$E(x_i) = \beta_i - \frac{G_i(1)}{F_i(1)},$$

and taking the expected value of the second partial derivatives, the information matrix can be shown to be:

$$I = \begin{bmatrix} \frac{np}{\beta_1^2}\left[F_1(1) - \frac{(1-k)G_1(1)}{\beta_1^2}\right] & \frac{nG(1)k(1-k)}{\beta_1^2\beta_2^2} & -\frac{nG(1)k(1-k)}{\beta_1^2 pq} \\ \frac{nG(1)k(1-k)}{\beta_1^2\beta_2^2} & \frac{nq}{\beta_2^2}\left[F_2(1) - \frac{kG_2(1)}{\beta_2^2}\right] & \frac{nG(1)k(1-k)}{\beta_2^2 pq} \\ -\frac{nG(1)k(1-k)}{\beta_1^2 pq} & \frac{nG(1)k(1-k)}{\beta_2^2 pq} & \frac{n}{pq^2}\left[q-(1-k)G_1(1)\right] \end{bmatrix}$$

The asymptotic variances and covariances are obtained from the inverse of this matrix.

In the limit, as p approaches one, the information matrix can be shown to approach

$$\begin{bmatrix} \frac{nF_1(1)}{\beta_1^2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \infty \end{bmatrix} .$$

The variance covariance matrix is the inverse of the information matrix. This implies that $\sigma_{\hat{p}}^2 \rightarrow 0$, which is correct since all estimates of p would be unity. It also can be shown that, in the case of a single population with sampling censored at a fixed time T,

$$-E\left\{\frac{\partial^2 \ln L}{\partial \beta^2}\right\} = \frac{nF(1)}{\beta^2}$$

$$= \frac{E(r)}{\beta^2} .$$

Thus the information matrix agrees with the information for a single population in the case when p = 1.

It is well known that the maximum likelihood estimates are consistent and efficient as $n \rightarrow \infty$. The effect of increasing the test termination time, T, is not so obvious. It will be more helpful to examine the information matrix of $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{p}$ in order to determine the properties of the estimates as a function of T. Since

$$\frac{\partial^2 \ln L}{\partial \alpha_i^2} = \frac{1}{T^2} \frac{\partial^2 \ln L}{\partial \beta_i^2}$$

and

$$\frac{\partial^2 \ln L}{\partial \alpha_i \partial p} = \frac{1}{T} \frac{\partial^2 \ln L}{\partial \beta_i \partial p} \quad ,$$

the symmetric information matrix for $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{p}$ is

$$\begin{bmatrix} -\frac{1}{T^2}E\left\{\frac{\partial^2 \ln L}{\partial \beta_1^2}\right\} & -\frac{1}{T^2}E\left\{\frac{\partial^2 \ln L}{\partial \beta_1 \partial \beta_2}\right\} & -\frac{1}{T}E\left\{\frac{\partial^2 \ln L}{\partial \beta_1 \partial p}\right\} \\ & -\frac{1}{T^2}E\left\{\frac{\partial^2 \ln L}{\partial \beta_2^2}\right\} & -\frac{E}{T}\left\{\frac{\partial^2 \ln L}{\partial \beta_2 \partial p}\right\} \\ & & -E\left\{\frac{\partial^2 \ln L}{\partial p^2}\right\} \end{bmatrix} .$$

As T becomes large relative to α, the off-diagonal terms of the information matrix approach zero. Therefore, when T is large, the information matrix of $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{p}$ is approximately

$$\begin{bmatrix} \frac{npF_1(t = T)}{\alpha_1^2} & 0 & 0 \\ 0 & \frac{nqF_2(t = T)}{\alpha_2^2} & 0 \\ 0 & 0 & \frac{n}{pq} \end{bmatrix} .$$

Since the variance-covariance matrix of $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{p}$ is the inverse of the information matrix, it follows that

$$V(\hat{\alpha}_i) \approx \frac{\alpha_i^2}{np_iF_i(t = T)} = \frac{\alpha_i^2}{E(r_i)} \quad , \qquad i = 1, 2,$$

where $p_1 = p$ and $p_2 = q$,

and $$V(\hat{p}) \approx \frac{pq}{n} .$$

The variance of $\hat{\alpha}_i$ will decrease as $E(r_i)$ increases. Since $\beta_i = \frac{\alpha_i}{T}$ and α_i is assumed fixed, β_i decreases as T increases. Because the $E(r_i)$ is proportional to $F_i(1) = 1 - e^{-1/\beta_i}$, a change in the variances of $\hat{\alpha}_i$ for a fixed change in β_i will be greatest when β_i is large. It would seem that estimation of the parameters is best when $\beta_i = \frac{\alpha_i}{T}$ is less than .5 and preferably less than .4.

The asymptotic variances were computed for parameter points 3, 6, 14, 19, and 22, these points selected in order to obtain various combinations of $E(r_1)$ and $E(r_2)$. The results are shown in Table 7 where $\sigma^2_{\hat{p}}$, $\sigma^2_{\hat{\beta}_1}$, $\sigma^2_{\hat{\beta}_2}$, and $\sigma_{\hat{\beta}_1\hat{p}}$, $\sigma_{\hat{\beta}_2\hat{p}}$, $\sigma_{\hat{\beta}_1\hat{\beta}_2}$ represent the asymptotic variances and covariances of the estimates.

Table 8 presents the asymptotic variances and the corresponding estimated variances obtained from the sampling experiment. The asymptotic variances and the estimated variances based on samples of N = 50 estimates are seen to agree very well for sample points 6 and 19. It would appear that the actual variances and the asymptotic variances of the estimates differ but little for the given parameters when the sample size n and the test termination time T are large.

Table 7

Asymptotic Variances and Covariances of

the Estimates

Parameter Point	3	6	14	19	22
n	100	100	100	200	50
p	.1	.2	.3	.3	.1
β_1	.4	.2	.8	.4	.8
β_2	1.6	.6	3.2	.8	3.2
$E(r_1)$	9.2	19.9	21.4	55.0	3.6
$E(r_2)$	41.9	64.9	18.8	99.8	12.1
$\sigma^2_{\hat{p}}$	.0011	.0016	.0120	.0014	.0085
$\sigma^2_{\hat{\beta}_1}$	.0445	.0024	.2469	.0073	1.4857
$\sigma^2_{\hat{\beta}_2}$	.0625	.0056	.8268	.0077	.9619
$\sigma^2_{\hat{\beta}_1\hat{p}}$	.0025	.0001	.0465	.0012	.0932
$\sigma_{\hat{\beta}_2\hat{p}}$	.0005	.0000	.0529	.0007	.0275
$\sigma_{\hat{\beta}_1\hat{\beta}_2}$	-.0060	-.0001	-.2473	-.0024	-.3859

Table 8

A Comparison of the Asymptotic Variances

and the Estimated Variances

Parameter Point	3	6	14	19	22
$\sigma^2_{\hat{p}}$	.0011	.0016	.0120	.0014	.0085
$s^2_{\hat{p}}$	.0069	.0021	.0299	.0015	.0524
$s^2_{\hat{p}_A}$	.0018	.0021	.0169	.0015	.0126
$\sigma^2_{\hat{\beta}_1}$	.0445	.0024	.2469	.0073	1.4857
$s^2_{\hat{\beta}_1}$	.6876	.0040	.6771	.0072	30.2476
$s^2_{\hat{\beta}_{1A}}$	.1287	.0040	.3495	.0072	1.4379
$\sigma^2_{\hat{\beta}_2}$	.0625	.0056	.8268	.0077	.9619
$s^2_{\hat{\beta}_2}$	.0885	.0056	1.3653	.0087	1.6364
$s^2_{\hat{\beta}_{2A}}$	.0619	.0056	.9191	.0087	.8537

Conclusions

The maximum likelihood estimation procedure is good when n is large and T is large relative to α_1 and α_2. The asymptotic variances and covariances are inversely proportional to the sample size n since the information matrix, I, equals

$$I = nA$$

where A is a matrix whose elements are not a function of n. The estimation efficiency decreases rapidly as the ratio, $\beta_i = \alpha_i/T$ increases beyond $\beta_i = .5$. In addition, the relative magnitude of the covariances to the variances of the estimates is controlled entirely by the test termination time, T. The covariances approach zero as T approaches infinity. In general, it is desirable to choose T so that $\beta_i = \alpha_i/T$ is less than .5 and preferably less than .4.

The experimenter will usually know the relative magnitude of α_1 and α_2 and hence can reduce the bias and variance of the estimates by the use of a procedure to adjust for crossovers. A crossover occurs when $\hat{\beta}_i > \hat{\beta}_j$, given that $\beta_i < \beta_j$. In this case it is reasonable to assume that $\beta_1 = \beta_2$. The adjusted estimation procedure uses the maximum likelihood estimate,

$$\hat{\beta}_1 = \hat{\beta}_2 = \frac{\sum_{j=1}^{r_1} x_{1j} + \sum_{j=1}^{r_2} x_{2j} + (n - r)}{r},$$

when a crossover occurs. Otherwise the estimates are those obtained from the original maximum likelihood procedure.

It is assumed, as a rough approximation, that the variance of $\hat{\beta}_i$ varies as β_i^2 and inversely as $E(r_i)$. This rule would seem to hold reasonably well except in regions of the parameter space where the frequency of crossovers is excessive. The experimenter should choose T and n in combination so as to obtain the desired variance of the estimates. The variances of the estimates for parameter points other than those used in the empirical sampling experiment can be obtained by extrapolation.

For example, suppose that the experimenter anticipates that the values of α_1, α_2, and p are approximately 40 hours, 80 hours, and .25 respectively. As has been mentioned, it is desirable to choose T so that β_i is less than .5. Assume that the experimenter has chosen T = 100 and desires to determine the sample size n for which the variance of $\hat{\alpha}_1$ is approximately 100. Therefore,

$$\beta_1 = \alpha_1/T \approx .4$$

$$\beta_2 = \alpha_2/T \approx .8$$

$$p \approx .25$$

$$V(\hat{\beta}_1) = \frac{V(\hat{\alpha}_1)}{T^2} = \frac{100}{10000} = .01.$$

Extrapolate from a parameter point which has approximately the same value of p as that assumed and which has a low frequency of crossovers. In this case, choose parameter point 6. From Tables 4 and 11,

$$E(r_1)_6 = 19.9$$

$$V(\hat{\beta}_1)_6 = .004.$$

Use the relation,

$$\frac{V(\hat{\beta}_1)}{V(\hat{\beta}_1)_6} = \frac{\beta_1^2}{\beta_{1,6}^2} \frac{E(r_1)_6}{E(r_1)}$$

or

$$\frac{.010}{.004} = \frac{(.4)^2}{(.2)^2} \frac{19.9}{E(r_1)} ;$$

hence $$E(r_1) = 31.8.$$

The estimated sample size n can now be obtained from

$$E(r_1) = np(1 - e^{-1/\beta_1})$$

or $$n = \frac{31.8}{(.25)(1 - e^{-1/.4})}$$

$$= 139.$$

Chapter VI

RESULTS FOR A MORE GENERAL MODEL

Consider a mixture of s failure sub-populations, mixed in proportion $p_1: \ p_2: \ \ldots: p_s$ where $0 \le p_i \le 1$ and $\sum_{i=1}^{s} p_i = 1$. Assume that the sub-populations are distributed according to a distribution function $F_i(t: \alpha_1, \alpha_2 \ldots \alpha_m)$ which is independent of p_j, $j = 1, 2, \ldots s$. Then the cumulative distribution function is

$$F(t) = \sum_{i=1}^{s} p_i F_i(t).$$

A random sample of n units is tested to time $t = T$. The number of units, r_i, belonging to the ith sub-population and failing before time T, is recorded along with the actual failure times, t_{ij}, $i = 1, 2, \ldots s$; $j = 1, 2, \ldots r_i$. Let the $\sum_{i=1}^{s} r_i = r$. Then $(n - r)$ units, which cannot be identified as to sub-population, survive the test.

Assume that all measurements are in units of size T and let $x = \frac{t}{T}$ and $\beta_i = \frac{\alpha_i}{T}$. The conditional mixture proportion, defined in Chapter IV, is

$$p_i(x) = \frac{p_i G_i(x)}{G(x)}$$

where

$$G_i(x) = 1 - F_i(x),$$

$$G(x) = 1 - F(x) \ ,$$

$$0 \leq p_i(x) \leq 1 \ ,$$

$$\sum_{i=1}^{s} p_i(x) = 1 \ ,$$

$$p_i = p_i(0) \ ,$$

and

$$p_i(1) = k_i, \ i = 1, 2, \ldots s.$$

The likelihood L is

$$L = \frac{n!}{(n - r)!} G(1)^{n-r} \prod_{j=1}^{s} p_i^{r_i} \prod_{j=1}^{r_1} f_1(x_j) \prod_{j=1}^{r_2} f_2(x_j) \ \ldots \prod_{j=1}^{r_s} f_s(x_j).$$

It follows that the first partial derivative of lnL with respect to p_i is

$$\frac{\partial \ln L}{\partial p_i} = (n - r) \frac{d\{\ln G(1)\}}{dp_i} + \frac{r_i}{p_i} - \frac{r_s}{p_s}$$

$$= (n - r)\left[\frac{k_i}{p_i} - \frac{k_s}{p_s}\right] + \frac{r_i}{p_i} - \frac{r_s}{p_s} \ .$$

Setting the $\frac{\partial \ln L}{\partial p_i}$ equal to zero and simplifying,

$$p_i = \frac{r_i}{n} + \frac{(n - r)ki}{n} \ , \quad i = 1, 2, \ldots s.$$

The likelihood equations for p_i are thus linear functions of the corresponding conditional mixture proportions, $k_i = p_i(1)$, regardless of the distributional form of $F_i(x)$.

It would be desirable to consider a distribution function, $F_i(x)$, more general in form than the exponential in order to provide a model which will fit a larger set of failure populations. The frequent occurrence of exponentially distributed failure populations makes it desirable to choose a family of distributions for which the exponential would be a special case. The family of distributions represented by the Weibull function,

$$F_i(x) = 1 - e^{-(x/\beta_i)^{m_i}}$$

and

$$f_i(x) = \frac{m_i}{\beta_i}\left(\frac{x}{\beta_i}\right)^{m_i-1} e^{-(x/\beta_i)^{m_i}}$$

is one possibility since $F_i(x)$ is the exponential distribution when $m_i = 1$. Assuming various values for the shape parameter, m_i, provides a wide range of distributional shapes from which to choose.

Let

$$F(x) = \sum_{i=1}^{s} p_i F_i(x)$$

where

$$F_i(x) = 1 - e^{-(x/\beta_i)^{m_i}} \qquad i = 1, 2, \ldots s$$

and the values of m_i are assumed known. Taking the first partial derivative of lnL with respect to β_i yields

$$\frac{\partial \ln L}{\partial \beta_i} = \frac{m_i(n-r)k_i}{\beta_i^{(m_i+1)}} - \frac{m_i r_i}{\beta_i} + \frac{m_i \sum_{j=1}^{r_i} x_{ij}^{m_i}}{\beta_i^{(m_i+1)}} .$$

Setting the $\frac{\partial \ln L}{\partial \beta_i}$ equal to zero and simplifying,

$$\hat{\beta}_i = \sqrt[m_i]{\frac{\sum_{j=1}^{r_i} x_{ij}^{m_i} + (n-r)k_i}{r_i}} .$$

When $s = 2$ and $m_1 = m_2 = 1$, the estimating equations reduce to those obtained for the case of two exponentials.

It should not be too difficult to set up an iterative method for the solution of the maximum likelihood equations, regardless of the shape parameters, m_i, $i = 1, 2, \ldots s$, as long as the number of sub-populations, s, is small. In general, there will be $(2s - 1)$ equations to solve for the same number of parameters. The procedure used for $s = 2$ was to reduce the three equations, by substitution, to a single equation,

$$k = g(k),$$

and then determine the value of k such that $g(k) - k = 0$. In the general case, it is possible to reduce the $(2s - 1)$ equations to $s - 1$ equations of the form

$$g_i(K) - k_i = 0 \qquad i = 1, 2, \ldots, s - 1$$

where

$$k_i = p_i(T)$$

is the conditional mixture proportion for sub-population (i) at time $t = T$ and $K = (k_1, k_2, \ldots, k_{s-1})$. The iteration method would then involve the selection of a vector K, the solution of the (s - 1) simultaneous equations. If a unique maximum likelihood solution exists, the solution will be a point in a restricted region within a unit cube in the (s - 1) dimensional space of K since

$$0 \leq k_i \leq 1$$

and

$$\sum_{i=1}^{s} k_i = 1.$$

A digital computer would have little difficulty in locating the solution by trial and error when s is small. More efficient procedures for solving the equations by iterative methods are given by Scarborough (1930). Once K is determined, the estimates of β_i and p_i can be obtained from the original maximum likelihood equations.

Chapter VII

SUMMARY AND CONCLUSIONS

Summary

This dissertation has been concerned with failure populations which can be subdivided into sub-populations representing different types of failure. The s sub-populations were assumed to be mixed in unknown proportion $p_1 : p_2 : \ldots : p_s$, where $0 \leq p_i \leq 1$, $i = 1, 2, \ldots s$ and $\sum_{i=1}^{s} p_i = 1$. The sub-population cumulative failure distributions, $F_i(t)$, were assumed known except for the parameters which were to be estimated for censored sampling with a fixed test termination time, T.

Conditional mixture proportions, $p_i(t)$, $i = 1, 2, \ldots s$, $0 \leq p_i(t) \leq 1$, and $\sum_{i=1}^{s} p_i(t) = 1$, were defined as the proportions in which the sub-populations were mixed at time t.

Estimation problems were considered for the special case where s = 2 and the failures for each sub-population were exponentially distributed with average lives, β_1 and β_2, expressed in units of size T. The maximum likelihood estimates of the population parameters were shown to be linear functions of the conditional mixture proportion, $k = p_1(t = T)$, where T is the test termination time. A simple iterative computational procedure was introduced for the solution of the estimating equations. An adjusted estimation procedure was proposed to improve the estimation when the relative magnitude of the sub-population parameters is known.

The means and variances of the estimates were obtained empirically and presented in tabular form for various combinations of the parameters, β_1 = .2 to 1.0, β_2 = .6 to 3.2, p = .05 to .35, and sample sizes, n = 50, 100, 200. The asymptotic variances of the maximum likelihood estimates were computed for certain parameter points in order to enable a comparison with empirical results.

The mixed failure population model was generalized to include any number of sub-populations of failures distributed according to Weibull distributions having different, but known, shape parameters. The maximum likelihood equations for estimating the parameters were shown to be simple functions of the conditional mixture proportions, $k_i = p_i(T)$, i = 1, 2, ... s.

Conclusions

The maximum likelihood estimation procedure appears to give satisfactory results when the sample size n is large and T is large relative to α_1 and α_2. However, the maximum likelihood estimators are badly biased and have large variances when n and T are small. It would seem desirable to investigate the use of other estimators which would have better properties for small samples. One possibility is the use of the adjusted estimation procedure, discussed in this dissertation, when the relative magnitude of α_1 and α_2 is known. In most practical situations, the experimenter will possess this information. The bias and the variance of the adjusted estimates is much less than for the maximum likelihood estimates when the sample size is small.

The variances and covariances of the estimates vary inversely as the sample size, n, and the ratio of the covariances to the variances diminishes as the test termination time, T, increases. The efficiency of estimating $\alpha_i = \beta_i T$, the average life of the units from sub-population (i), decreases rapidly when T is chosen so that $\beta_i > .5$. In general, T should be chosen so that β_i is as small as possible and preferably less than .4.

For practical purposes, the variance of $\hat{\beta}_i = \frac{\hat{\alpha}_i}{T}$ varies approximately as β_i^2 and inversely as $E(r_i)$,

$$V(\hat{\beta}_i) = \frac{\beta_i^2}{E(r_i)} ,$$

where r_i is the number of units from sub-population (i) which fail before time T. Using this relation, a choice of the test termination time T, and approximate knowledge of α_1, α_2, and p, it is possible to estimate the sample size required to obtain a specified variance of $\hat{\alpha}_i$ by extrapolation from the empirically obtained variances.

The primary benefit to be derived from this investigation, is in the use of the mixed failure population approach to describe the failure characteristics of a product. The failure model, utilizing Weibull distributions, should be general enough to fit most mixed failure populations, including those possessing non-monotonic conditional failure density functions.

Suggested Extensions

Problems dealing with mixed failure populations offer a fertile and useful area for research. Extensions which are obvious are as follows:

1. Examine the properties of the maximum likelihood estimates for a mixture of Weibull distributions.
2. Consider censored sampling where the total number failing, r, is fixed and the test termination time, t_r, is random.
3. Consider sampling with replacement in order to achieve more efficient use of testing equipment.
4. Develop small sample estimators of the parameters for mixed failure distributions.
5. Develop procedures for testing hypotheses.

BIBLIOGRAPHY

Acheson, Marcus A. and McElwee, Eleanor M. 1951. Concerning the reliability of electron tubes. The Sylvania Technologist 4.

Cohen, A. C., Jr. 1950. Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. Ann. Math. Stat. 21:557-569.

Cohen, A. C., Jr. 1951. Estimation of parameters in truncated Pearson frequency distributions. Ann. Math. Stat. 22:256-265.

Davies, J. A. 1952. Life test acceptance sampling methods. Unpublished paper presented at a summer statistics conference, University of North Carolina.

Davis, D. J. 1952. An analysis of some failure data. Jour. Amer. Stat. Assoc. 47:113-150.

Deemer, Walter L., Jr. and Votaw, David F., Jr. 1955. Estimation of parameters of truncated or censored exponential distributions. Ann. Math. Stat. 26:498-504.

Epstein, B. 1953. Statistical problems in life testing. Proceedings of The Seventh Annual Convention, American Society of Quality Control.

Epstein, B. 1954. Life test estimation procedures. Unpublished, Technical Report No. 2, Department of Mathematics, Wayne State University.

Epstein, B. and Sobel, M. 1953. Life testing. Jour. Amer. Stat. Assoc. 48:486-502.

Epstein, B. and Sobel, M. 1954. Some theorems relevant to life testing from an exponential distribution. Ann. Math. Stat. 25:373-381.

Feller, William. 1950. An introduction to probability theory and its applications. John Wiley and Sons, Inc., New York.

Gunther, Paul. 1956. Techniques for statistical analysis of life test data. Unpublished report, General Electric Co., Schenectady, New York.

Gupta, A. K. 1952. Estimation of the mean and the standard deviation of a normal population from a censored sample. Biometrika 39:260-273.

Hald, A. 1949. Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. Skandinavisk Aktuarietidskrift 32:119-134.

Herd, G. R. 1953. Heterogeneous distributions. Unpublished technical note.

Herd, G. R. 1956. Estimation of the parameters of a population from a multi-censored sample. Ph.D. dissertation, Iowa State College.

Kao, John H. K. 1955. The Weibull distribution in life testing of electron tubes. Unpublished paper presented at annual meeting of American Statistical Association, New York.

Leiblein, J. and Zelen, M. 1956. Statistical investigation of the fatigue life of deep-groove ball bearings. Journal of Research of the National Bureau of Standards 57:273-316.

Madison, Ralph L. 1955. Applications of statistical methods in evaluating performance of electronic equipment. Proceedings of the Ninth Annual Convention, American Society for Quality Control.

Moshman, J. 1954. The generation of pseudo-random numbers on a decimal calculator, Journal Association for Computing Machinery 1:88-91.

Sarhan, A. E. 1955. Estimation of the parameters of a skewed distribution by linear systematic statistics. Jour. Amer. Stat. Assoc. 50:196-208.

Sarhan, A. E. and Greenberg, B. G. 1956. Estimation of location and scale parameters by order statistics from singly and doubly censored samples. Ann. Math. Stat. 27:427-451.

Sarhan, A. E. and Greenberg, B. G. 1957. Tables for best linear estimates by order statistics of the parameters of single exponential distributions from singly and doubly censored samples. Jour. Amer. Stat. Assoc. 52:58-87.

Scarborough, J. B. 1930. Numerical mathematical analysis. The Johns Hopkins Press, Baltimore.

Steen, J. R. 1952. Life testing of electronic tubes. Unpublished paper presented at a summer statistics conference, University of North Carolina.

Walsh, J. E. 1956. Estimating population mean, variance, and percentage points from truncated data. Skandinavisk Aktuarietidskrift 39:47-58.

Weibull, Waloddi. 1951. A statistical distribution function of wide applicability. Journal of Applied Mechanics 18:293-297.

Wilde, R. D. 1952. Nature of rate of failure curves. Unpublished paper presented at a summer statistics conference, University of North Carolina.

Table 9

Results of Sampling Experiment

at Parameter Point 14

$n = 100 \quad \beta_1 = .8 \quad \beta_2 = 3.2 \quad p = .30$

Sample Number	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{p}$	$\bar{x}_1$	$\bar{x}_2$	r_1	r_2
1	.7278	3.5463	.2276	.3891	.4809	17	19
2	.5103	3.0486	.3022	.3481	.5595	26	20
3	.6749	3.6860	.2722	.3788	.4047	21	17
4	.3506	3.3251	.2124	.2889	.3869	20	20
5	.9517	3.1061	.3842	.4149	.4837	25	17
6	.7128	3.0015	.4527	.3811	.3532	34	15
7	1.1325	2.5771	.3374	.4456	.5652	20	22
8	2.8749	.4807	.7941	.4221	.3370	23	18
9	.4488	2.9101	.2355	.3274	.4350	21	22
10	.6056	1.9330	.2480	.3658	.4263	20	30
11	.7593	3.1771	.2449	.3988	.5813	18	21
12	.9962	3.5333	.3943	.4191	.4952	25	15
13	.3238	3.5647	.1886	.2762	.5076	18	20
14	.7149	3.6878	.3716	.3878	.4985	28	15
15	2.0371	1.5213	.5426	.4532	.4423	21	22
16	1.8296	1.9543	.5595	.4982	.5073	24	18
17	.9604	2.1570	.3853	.4190	.4845	25	23
18	.8117	2.8855	.4933	.4024	.5073	35	15
19	.7710	3.3036	.2620	.3922	.4192	19	19
20	.5016	3.4042	.3008	.3446	.5198	26	18
21	1.4403	1.8893	.3760	.4616	.4891	19	26
22	.6291	3.7225	.2764	.3725	.4663	22	17
23	.7247	2.4568	.2411	.3852	.4213	18	25
24	3.0817	.5095	.8021	.4358	.3453	22	17
25	2.0361	2.7129	.6437	.4615	.4734	25	11

Table 9 (cont.)

Results of Sampling Experiment

at Parameter Point 14

$n = 100 \quad \beta_1 = .8 \quad \beta_2 = 3.2 \quad p = .30$

Sample Number	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{p}$	$\bar{x}_1$	$\bar{x}_2$	r_1	r_2
26	3.2937	.4707	.8069	.4514	.3347	21	17
27	.8296	3.0737	.3565	.4036	.4985	25	18
28	1.5612	2.1624	.4993	.4808	.5270	24	19
29	2.1786	2.0019	.5688	.4700	.4654	21	17
30	1.1322	2.9344	.3703	.4489	.6203	22	19
31	1.0780	3.4981	.3145	.4228	.4656	19	17
32	.8448	5.4790	.2022	.4006	.3420	14	13
33	2.2999	1.8445	.6197	.4833	.4674	22	16
34	1.9478	2.4458	.3738	.4560	.4638	15	21
35	.4340	3.4800	.1445	.3224	.4062	13	21
36	3.3393	.9574	.7375	.4578	.4132	19	17
37	.4379	3.3181	.2338	.3246	.4871	21	20
38	1.6171	2.0801	.5125	.4815	.5145	24	19
39	.4900	3.7595	.2071	.3392	.3547	18	18
40	1.2742	2.8069	.4395	.4431	.5097	24	17
41	2.5327	1.4711	.5765	.4987	.4542	19	21
42	.9977	2.6273	.2690	.4153	.4491	17	23
43	.7986	4.7504	.3907	.4032	.6730	28	12
44	.3794	6.0220	.2908	.3024	.5746	27	11
45	.4929	3.8389	.2647	.3419	.5137	23	17
46	.3874	5.2203	.3137	.3056	.5013	29	12
47	.9178	3.4128	.2110	.4104	.4680	14	20
48	2.4094	1.5187	.6093	.5081	.4622	21	19
49	1.5155	2.0567	.5248	.4972	.5555	26	19
50	1.1484	3.2640	.3258	.4335	.5186	19	18

Table 10

Comparison of the Means of the Maximum Likelihood Estimates and the Adjusted Estimates Based on 50 Samples Per Parameter Point

A = Number of Adjustments Per 50 Samples

Parameter Point	n	p	β_1	β_2	$\bar{\hat{p}}$	$\bar{\hat{p}}_A$	$\bar{\hat{\beta}}_1$	$\bar{\hat{\beta}}_{1A}$	$\bar{\hat{\beta}}_2$	$\bar{\hat{\beta}}_{2A}$	A
1	100	.05	.6	1.8	.119	.065	2.76	1.17	1.60	1.74	16
2	100	.10	.4	.8	.116	.101	.68	.47	.80	.83	8
3	100	.10	.4	1.6	.120	.109	.63	.50	1.58	1.61	4
4	100	.10	.8	1.6	.147	.107	1.59	.94	1.55	1.66	17
5	100	.10	.8	3.2	.195	.132	2.38	1.31	2.99	3.29	13
6	100	.20	.2	.6	.198	.198	.22	.22	.58	.58	0
7	100	.20	.6	.6	.208		.68		.57		
8	100	.20	.6	1.8	.243	.228	.73	.65	1.81	1.86	6
9	100	.20	.6	3.0	.248	.226	.90	.77	2.93	3.05	4
10	100	.20	1.0	3.0	.280	.239	1.65	1.31	2.66	2.83	13
11	100	.30	.4	.8	.302	.299	.44	.43	.81	.81	2
12	100	.30	.4	1.6	.297	.297	.41	.41	1.70	1.70	0
13	100	.30	.8	1.6	.329	.300	.98	.82	1.45	1.55	14
14	100	.30	.8	3.2	.394	.370	1.22	1.10	2.85	3.00	9
15	100	.35	.6	1.8	.382	.368	.70	.64	1.73	1.78	5
16	200	.10	.4	.8	.105	.104	.43	.42	.80	.80	3
17	200	.10	.8	3.2	.183	.125	2.37	1.29	2.88	3.14	10
18	200	.20	.6	1.8	.210	.207	.66	.64	1.75	1.76	1
19	200	.30	.4	.8	.299	.299	.39	.39	.80	.80	0
20	200	.30	.8	3.2	.369	.350	1.11	1.02	2.86	2.97	8
21	50	.10	.4	.8	.113	.097	.65	.37	.79	.82	10
22	50	.10	.8	3.2	.224	.159	2.68	1.31	3.01	3.38	11
23	50	.20	.6	1.8	.231	.203	.96	.71	1.65	1.74	8
24	50	.30	.4	.8	.311	.303	.43	.39	.76	.77	5
25	50	.30	.8	3.2	.386	.352	1.28	1.09	2.91	3.10	11

Table 11

Comparison of the Estimated Variances of the Maximum Likelihood Estimates and the Adjusted Estimates Based on 50 Samples Per Parameter Point

A = Number of Adjustments Per 50 Samples

Parameter Point	n	p	β_1	β_2	$s^2_{\hat{p}}$	$s^2_{\hat{p}_A}$	$s^2_{\hat{\beta}_1}$	$s^2_{\hat{\beta}_{1A}}$	$s^2_{\hat{\beta}_2}$	$s^2_{\hat{\beta}_{2A}}$	A
1	100	.05	.6	1.8	.015	.003	16.771	4.803	.171	.111	16
2	100	.10	.4	.8	.003	.001	.540	.045	.013	.007	8
3	100	.10	.4	1.6	.007	.002	.688	.129	.089	.062	4
4	100	.10	.8	1.6	.014	.003	2.917	.315	.140	.083	17
5	100	.10	.8	3.2	.043	.008	13.044	1.212	1.204	.559	13
6	100	.20	.2	.6	.002	.002	.004	.004	.006	.006	0
7	100	.20	.6	.6	.004		.139		.010		
8	100	.20	.6	1.8	.013	.006	.358	.142	.197	.127	6
9	100	.20	.6	3.0	.026	.010	.788	.258	.832	.453	4
10	100	.20	1.0	3.0	.026	.010	1.765	.675	1.013	.650	13
11	100	.30	.4	.8	.003	.002	.033	.022	.016	.014	2
12	100	.30	.4	1.6	.002	.002	.017	.017	.102	.102	0
13	100	.30	.8	1.6	.015	.007	.383	.137	.184	.096	14
14	100	.30	.8	3.2	.030	.017	.677	.349	1.365	.919	9
15	100	.35	.6	1.8	.011	.006	.165	.071	.255	.165	5
16	200	.10	.4	.8	.001	.001	.042	.027	.003	.003	3
17	200	.10	.8	3.2	.031	.003	11.546	.869	.705	.216	10
18	200	.20	.6	1.8	.003	.002	.107	.062	.055	.044	1
19	200	.30	.4	.8	.001	.001	.007	.007	.009	.009	0
20	200	.30	.8	3.2	.022	.012	.516	.272	.797	.490	8
21	50	.10	.4	.8	.005	.002	.950	.063	.032	.022	10
22	50	.10	.8	3.2	.052	.013	30.248	1.438	1.636	.854	11
23	50	.20	.6	1.8	.017	.005	1.244	.218	.312	.209	8
24	50	.30	.4	.8	.004	.003	.075	.022	.045	.037	5
25	50	.30	.8	3.2	.039	.022	1.205	.562	2.362	1.729	11