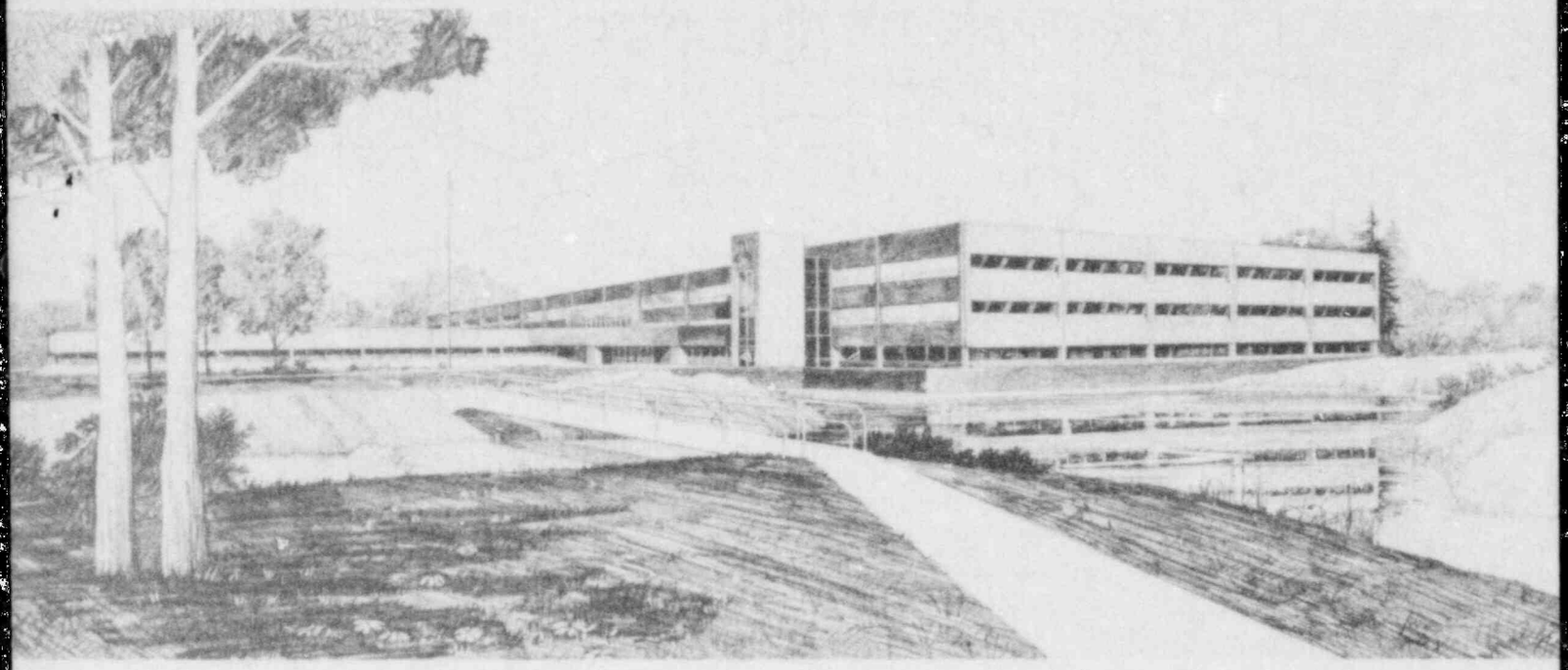
NUREG/CR-2714
EGG-2206



U.S. Department of Energy
Idaho Operations Office • Idaho National Engineering Laboratory

Analysis of Failure Rate Data with Compound Models

J. K. Shultis
N. D. Eckhoff
D. E. Johnson
G. A. Milliken
D. D. Drayer
D. R. Gallup
W. Buranapan

June 1982

Prepared for the
U.S. Nuclear Regulatory Commission
Under Contract No. DE-AC07-76IDO1570

EG&G Idaho

Available from

GPO Sales Program
Division of Technical Information and Document Control
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555

and

National Technical Information Service
Springfield, Virginia 22161

NUREG/CR-2714
EGG-2206
Distribution Category: RG, RD

ANALYSIS OF FAILURE RATE DATA WITH COMPOUND MODELS

J. K. Shultis
N. D. Eckhoff
D. E. Johnson
G. A. Milliken
D. D. Drayer
D. R. Gallup
W. Buranapan

Published June 1982

Department of Nuclear Engineering
Kansas State University
Manhattan, Kansas 66505

Prepared for EG&G Idaho, Inc.
Under Subcontract No. K-7717
and the U.S. Nuclear Regulatory Commission
Washington, D.C. 20555
Under DOE Contract No. DE-AC07-76IDO1570
FIN No. A6290

SUMMARY

One of the most useful statistical models to describe the failure of components with inherently low failure rates is the compound or Bayesian model. In this model the failure rate of a component is assumed to vary among all similar components according to some usually unknown prior distribution. The Poisson distribution, which gives the probability of having F component failures in operation time T for a given failure rate, is then averaged over all possible failure rates described by the prior distribution to obtain the compound or marginal distribution.

There are many procedures for determining the explicit form of the prior distribution that describes the failure rate variation among a given set of components. In this study, the prior distribution for a set of components was estimated from past failure data of the components in question by first assuming a functional form for the prior distribution (i.e., a prior family), and then estimating values of the prior parameters from observed failure data. Three methods were used to estimate values of the prior parameters: (1) matching the data moments to those of the prior distribution (PMMM), (2) matching the moments of the marginal distribution to those of the data (MMMM), and (3) the maximum likelihood method based on the marginal distribution (MMLM).

However, in the application of such a compound model to describe failure rate data, there are two major assumptions whose validity is often difficult to assess. The functional form assumed for the prior distribution is usually selected for mathematical convenience rather than for physical reasons. Consequently, the gamma distribution is almost universally used in a compound failure model because of the analytical simplification afforded by this conjugate function when combined with the Poisson failure distribution. While other prior families could, in principle, be used, the numerical evaluation of such non-conjugate compound models is exceptionally difficult. Although the conjugate gamma-Poisson model is mathematically convenient, it is unknown how the results of the failure analysis depend upon the a priori selection of the gamma family. To answer this question was one of the primary purposes of the present work.

The second major aspect of this work arises because of concern that the usual compound model in which the prior distribution is estimated from a given set of failure data is based on the assumption that all components in the data base can be described by the same prior distribution. This is not an obvious assumption. For example, one can well imagine that pumps from different manufacturers would have different prior distributions. Thus, in lumping together failure rate data from all pumps, one may actually be mixing two or more failure rate distributions. Hence the second major objective of this study was to investi-

gate methods for distinguishing failure data that came from different failure disributions.

In the first chapter of this report, the question of how close together two gamma prior distributions for the failure rate parameter can be before they cannot be distinguished from one another is considered. Failure rate data were generated from two gamma-Poisson marginal distributions with known gamma parameters. From each set of simulated failure data, the gamma parameters were estimated by each of the three estimation techniques (PMMM, MMMM, and MMLM), and likelihood ratio tests were performed to detemine whether the two data sets came from the same distribution. From these results on simulated data, power curves for the likelihood ratio test were empirically determined. These power curves give the fraction of time that data sets from different failure rate distributions were correctly identified as coming from different distributions versus a measure (one of the gamma parameters) of the actual difference in the distributions. It was found from these power curves that for small sample size (10) and for failure rate distributions typical of nuclear power plant components, that the PMMM and MMLM gave prior parameter estimates that were more readily distinguished as being different than the MMMM estimates. However, the marginal based estimation methods failed to give valid parameter estimates for an appreciable number of data sets, while the PMMM almost always yielded parameter estimates.

In the second chapter of this report, the study of outlier detection methods is summarized. Methods were compared for detecting outliers in three types of failure data: (1) data distributed according to a gamma distribution (i.e., failure rates), (2) data giving the number of failures F in operation time T for each component in the set, and (3) time-to-failure data which were distributed according to an exponential distribution. For data of type (1) three outlier detection methods were compared: Fisher's method, a normal conversion method, and an integration method. A novel way for semi-analytically obtaining power curves for each method was developed, and from the resulting power curves for each method, properties of the methods were determined. In all cases, the normal conversion method was the most powerful method for detecting upper outliers in data from a gamma distribution. For failure data of type (2), the cumulative marginal method and a binomial method were used to detect outliers. The power curves for these two methods were determined empirically from simulated failure data from known distributions. For data with an average failure rate smaller than 0.00005 per hour and all component test times equal, the binomial method is superior, while for data with greater average failure rates, the marginal cumulative method is more powerful in correctly identifying outliers. The detection of outliers in time-to-failure data were examined using a Fisher's test and a test originated by Dixon. Again power curves for the two outlier tests were generated from simulated time-to-failure data. In every case, Fisher's method was found to be superior to Dixon's method.

In the third chapter, several studies of the importance of the prior family selection in compound failure models are summarized. Simulated failure data of the (F,T) type were generated from conjugate gamma-Pois-

son and non-conjugate compound models. In particular, the prior distribution was taken as belonging to (1) the gamma family, (2) the Weibull family, (3) the lognormal family, or (4) the logbeta family. In an earlier study, methods were developed for estimating the two parameters of these prior families from failure date using the PMMM, MMMM, or the MMLM. The simulated failure data were analyzed by these three estimation methods assuming both conjugate and non-conjugate prior distributions. Many comparisons are presented in this report which use both conjugate and non-conjugate models to fit data from both conjugate and non-conjugate distributions. The main conclusion derived from all these studies is that the simpler conjugate compound model can be used to describe failure data even when the data come from non-conjugate distributions. There is no need to expend the large computational effort needed to use the non-conjugate compound models. The effect of the choice of the prior family is almost negligible. Of far greater importance, is the method used to estimate the parameters of the chosen prior family from the failure data.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

THE DISTINCTION OF FAILURE DATA BY THE LIKELIHOOD RATIO TEST

1.1 INTRODUCTION

When a group of similar components is viewed in a classical sense, all of the components are assumed to be completely identical. Because they are assumed identical, all of the components have the same failure rate or probability of failure per unit time. If the same group of components is viewed in a Bayesian sense, the components no longer are assumed identical but may exhibit variations in the failure rate, although each component still has its own distinct failure rate. The distribution of failure rates among the components in a given population is often referred to as the prior distribution and is a key feature of a Bayesian model. This prior distribution may represent the actual distribution of failure rates among similar components, or simply an estimate of the analyst's uncertainty about the true value of the failure rate. In either case, this use of a prior distribution leads to a much more flexible description of component failures.

Many models have been used to describe the failure rate distribution in a class of components. One of the desirable characteristics of a model for a prior distribution is that it should be able to assume many different shapes for different values of its parameters. Such flexibility in shape is beneficial when modeling actual failure rate distributions. Moreover, the choice of a particular prior family is often made on the basis of the ease with which it is incorporated into the analytical formulation of the failure model.

An important use of the Bayesian or compound failure model to describe the failure of components is its ability to distinguish between groups of components. For two sets of components, say from two different manufacturers, one is interested in determining whether or not the components have failure rates which come from the same prior distribution. To investigate this question, sets of failure data were simulated from an assumed failure rate distribution with known parameters. The simulated data were used to estimate parameters of the failure rate prior distribution. Several methods, including matching moments and maximum likelihood techniques, were used to estimate the parameters of the prior distribution from the simulated failure data. The likelihood ratio test was then used to compare the results for different simulated data sets obtained using different known prior distributions. From the results of these likelihood ratio tests, power curves were generated. With these empirical power curves it is then possible to draw conclu-

sions about how dissimiliar the prior distributions must be in order to state confidently that the failure rate distributions are in fact different.

1.2 THE HOMOGENEOUS AND COMPOUND FAILURE MODEL

Two different statistical models were used to describe component failures. The first of these models, the homogeneous model, is the simplest of the two. In this model the failure rate λ is assumed to be some unknown constant, which is equal for all components in a particular population.

It is often assumed that a component which fails is immediately repaired and brought back into operation without changing its failure rate probability, λ. With this assumption, the probability of observing F failures in T component-hours of test time for a particular component is given by the Poisson distribution,

$$f(F|\lambda;T) = \frac{(\lambda T)^F}{\Gamma(F+1)} e^{-\lambda T} \qquad (1\text{-}1)$$

where $\Gamma(F+1)$ is the gamma function.

If all the components in a class are then assumed to have the same failure rate λ, the above result becomes the homogeneous failure model in which T now represents the total number of component-hours of test time accumulated by all the components in operation. Of major practical interest with this model is the estimation of the failure rate for members of the population. It can be shown that the maximum likelihood estimator is given by

$$\hat{\lambda} = \sum_{i=1}^{n} F_i \Big/ \sum_{i=1}^{n} T_i \qquad (1\text{-}2)$$

i.e., the total number of failures divided by the total test time. For components which are designed to have very low failure probabilities, the homogeneous model is often inadequate since there are usually zero observed failures and the maximum likelihood estimate of the failure parameter thus has the unrealistic value of zero.

The second and more complex failure model, known as the compound model, is often better for modeling the failure rates of components, especially when the failure probabilities are inherently low. In this model the failure rate is assumed to vary from component-to-component within the class although it remains constant in time for any given component. The failure rate is further assumed to be distributed among the components of the population according to some function g( ). This distribution is called the failure rate prior distribution since it is usually determined from previous information about the components in a

given class. With $g(\lambda)$ known, the probability of experiencing F failures in T component-hours of operation is given by the marginal distribution

$$h(F;T,g) = \int_0^\infty f(F|\lambda;T)\, g(\lambda)\, d\lambda, \quad F=0,1,... \qquad (1\text{-}3)$$

where $f(F|\lambda;T)$ is called the likelihood or conditional distribution. For the failure rate case, f is the Poisson distribution of Eq. (1-1).

1.2.1 The Gamma-Poisson Failure Rate Model

The testing for the distinction of differences in the prior distributions was restricted in this study to the above failure rate problem. Further, the prior distribution was assumed to belong to the gamma family. This family is the natural conjugate to the Poisson distribution and allows the integration of Eq. (1-3) to be performed analytically. Using this conjugate failure model, the marginal distribution becomes

$$h(F;T,\alpha,\tau) = \frac{\Gamma(F+\alpha)}{\Gamma(\alpha)\Gamma(F+1)} \frac{T^F\, \tau^\alpha}{(T+\tau)^{F+\alpha}} \qquad (1\text{-}4)$$

where α and τ are the parameters of the prior gamma distribution

$$g(\lambda;\alpha,\tau) = \lambda^{\alpha-1}\, \tau^\alpha\, e^{-\lambda\tau}/\Gamma(\alpha)\ , \quad \alpha,\tau>0,\ 0\leq\lambda<\infty\ . \qquad (1\text{-}5)$$

The prior parameters can be estimated from previously observed component failure data by several methods. Let T_i and F_i denote the component test time and observed number of failures, respectively, for the i-th component in a given population of N components. The methods used in this study for estimating the parameters of the prior gamma distribution are summarized below.

1) PMMM - Matching data moments to those of the prior distribution. The parameter estimators are given by [1]

$$\hat{\alpha} = \bar{\lambda}^2/S^2 \quad \text{and} \quad \hat{\beta} \equiv 1/\hat{\tau} = S^2/\bar{\lambda} \qquad (1\text{-}6)$$

where

$$\bar{\lambda} \equiv \frac{1}{n} \sum_{i=1}^{n} \hat{\lambda}_i = \frac{1}{n} \sum_{i=1}^{n} (F_i/T_i) \qquad (1\text{-}7)$$

and

$$S^2 \equiv \frac{1}{n-1}\sum_{i=1}^{n}(\lambda_i-\bar{\lambda})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{F_i}{T_i}-\bar{\lambda}\right)^2 . \qquad (1\text{-}8)$$

2) MMMM - Matching data moments to those of the marginal distribution. The parameter estimators are given by [2]

$$\hat{\alpha} = \bar{\lambda}^2/[S^2 - \frac{\bar{\lambda}}{n}\sum_{i=1}^{n} T_i^{-1}] \qquad (1\text{-}9)$$

and

$$\hat{\beta} = [S^2 - \frac{\bar{\lambda}}{n}\sum_{i=1}^{n} T_i^{-1}]/\bar{\lambda} . \qquad (1\text{-}10)$$

3) MMLM - Marginal maximum likelihood method. The parameter estimators are solutions to the set of equations [2]

$$\hat{\alpha} = \left(\frac{F}{\hat{\beta}} - \sum_{i=1}^{n}\frac{T_i F_i}{1+\hat{\beta}T_i}\right) \Big/ \sum_{i=1}^{n}\frac{T_i}{1+\hat{\beta}T_i} \qquad (1\text{-}11)$$

and

$$\sum_{i=1}^{n}\left(\sum_{j=0}^{F_i-1}\frac{1}{\hat{\alpha}+1} - \ln(1+\hat{\beta}T_i)\right) = 0 \qquad (1\text{-}12)$$

where these two equations must be solved numerically.

1.3 LIKELIHOOD DISCRIMINATION PROCEDURE

In this study a gamma-Poisson compound failure model was assumed with specified values of parameters α and β and component test time T. With the marginal distribution of Eq. (1.2), simulated failure data were generated as follows. A random number between 0 and 1 was taken as the value of the cumulative marginal distribution. A value of F was then determined that yielded this value for the cumulative distribution of Eq. (1.2). By using a series of random numbers, a series of component failures F_i were generated which were distributed according to the specified marginal distribution.

Different sets of simulated failure data were thus generated from a gamma-Poisson marginal distribution with T = 10,000 hours and with values of the shape parameter α ranging between 1.0 and 2.5. The scale parameter β was chosen as the variable parameter for testing the ability of the likelihood ratio test to discriminate between different β values.

A standard value of β was selected as 0.00005 and then various other values were used to generate simulated failure data from marginal distributions that differed by various amounts from the standard one. The simulated failure data were grouped into sets of size 10, and each group was then used to estimate the gamma parameters by the three methods summarized in Section 1.2.1. These parameter estimates for each group can then be used to calculate the probability of observing the given data of each group, i.e., the sample likelihood

$$L(\alpha,\beta) = \prod_{i=1}^{n} h(F_i;T_i,\alpha,\beta) \qquad (1\text{-}13)$$

The null hypothesis (testing only for variations in β) is that all the failure data came from a gamma-Poisson distribution with $\beta = 0.00005$. The alternative hypothesis is that the failure data came from a distribution with a different value of the scale parameter. In both cases the shape paramter has the same known value (the value used to generate the simulated data). To test the null hypothesis, the likelihood ratio test was used. This test is based on the test statistic

$$T = -2\ell n\{L(\hat{\alpha}_o,\hat{\beta}_o)/[L(\hat{\alpha}_1,\hat{\beta}_1)\ L(\hat{\alpha}_2,\hat{\beta}_2)]\} \qquad (1\text{-}14)$$

which, asymptotically, has a chi-square sampling distribution with the number of degrees of freedom selected in accordance with the number of parameters estimated [3]. The quantity $L(\hat{\alpha}_o,\hat{\beta}_o)$ is the value of the likelihood function under the null hypothesis where the estimators $\hat{\alpha}_o$ and $\hat{\beta}_o$ are determined from the combined data sets while the quantities $L(\hat{\alpha}_i,\hat{\beta}_i)$ i=1,2 are evaluated using the parameter estimators $\hat{\alpha}_i$ and $\hat{\beta}_i$ obtained from each set separately. The selection of the number of degrees of freedom is discussed below. Thus, the decision rule is to reject the null hypothesis for a particular data set if $T > \chi^2(df,\gamma)$, where df is the appropriate number of degrees of freedom and γ is the significance level.

1.3.1 Power of the Likelihood Ratio Test

The power of the above likelihood ratio test is defined to be the fraction of the time the null hypothesis is correctly rejected for all data sets (of size 10) for which prior parameter estimators were obtained by the estimation technique under scrutiny. In many applications to data sets of such small size, the marginal matching moments (MMMM) or the maximum likelihood method (MMLM) may yield no parameter estimates. If estimates cannot be found for a particular set of data, this set is not used in calculating the power of the likelihood ratio test.

For example, say that the marginal matching moments method was able to find prior parameter estimates for only 151 sets of simulated failure

data out of a total of 250 generated sets. For these 151 pairs of estimators, it was found that the null hypothesis (all data came from a distribution with β =0.00005) was rejected 31 times. Thus the power of the likelihood ratio test based on the MMMM equals 31/151 not 31/250.

The statistic of Eq. (1-14) was used to test the null hypothesis that $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. The data were simulated from distributions with equal alpha parameters and different beta parameters to test the sensitivity of the test statistic. Thus, the simulated data was used to examine the power of the test only in the beta direction and not in the alpha direction. To generate the power of the test, the value of the likelihood function with a common alpha and different beta estimates should be obtained. However, available parameter estimation codes precluded a common alpha estimator for both data sets; rather separate alpha as well as beta estimators were used. The simulation reported here is based on using two degrees of freedom which yields too large a critical point so that the null hypothesis was rejected less often than it should have been. It can be shown that this procedure generates a lower bound for the power of the actual likelihood ratio test (i.e., one in which a common alpha estimate is used). An upper bound for the power of the test could be obtained by using one degree of freedom, although upper bounds are not reported here.

Lower bounds of the power curves, as a function of the scale parameter used to generate the simulated failure data, were obtained in the above manner for several different values of the shape parameter after using the three different estimation techniques (PMMM, MMMM, and MMLM) to find values of the gamma parameters from the simulated data sets. Typical results are shown in Figs. 1.1 and 1.2 As an approximation, these lower bounds can be used as power curves for the test of the null hypothesis. To obtain each empirical point on these power curves, 250 simulated failure data sets were used, and the significance level of the likelihood ratio test was taken as 0.05.

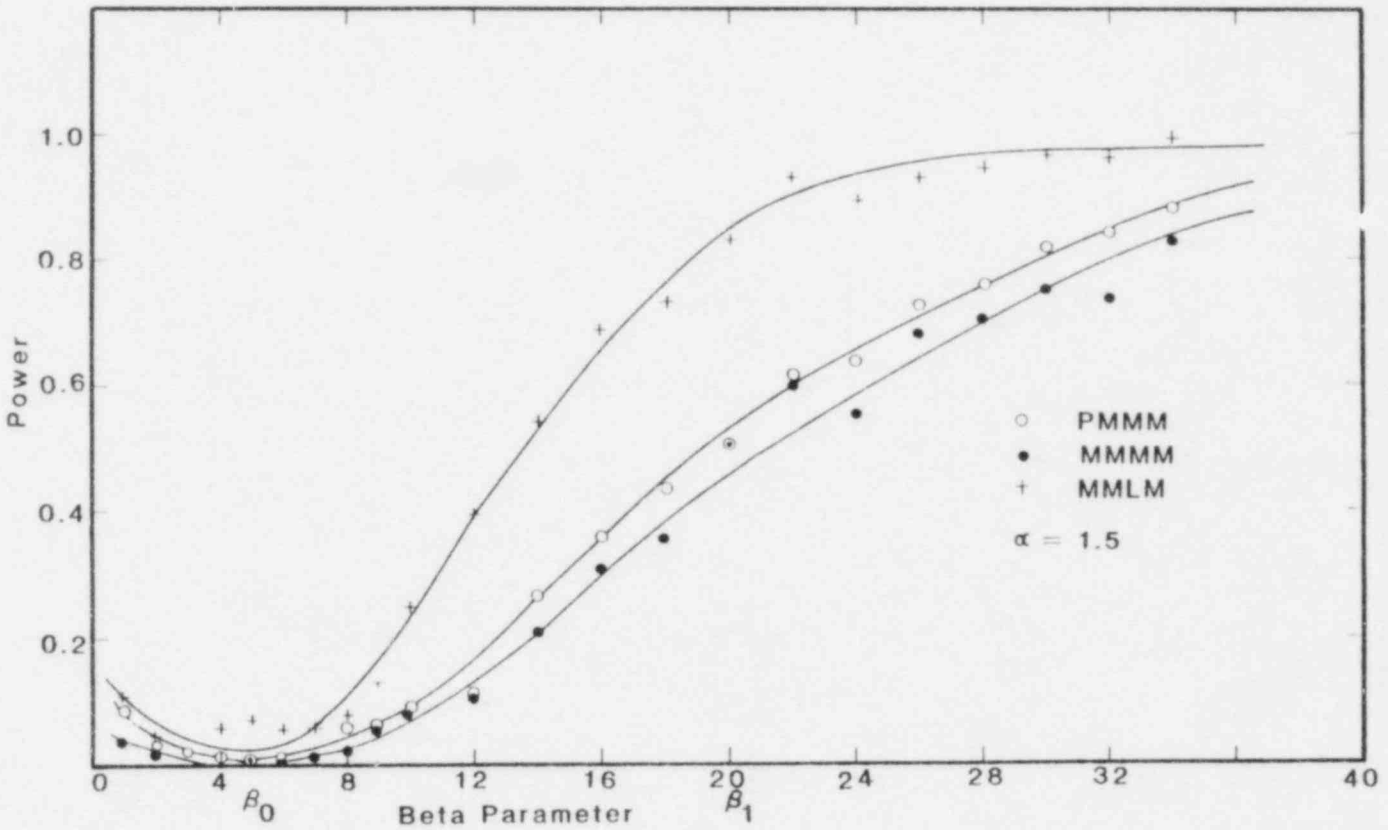


Figure 1.1: Empirical power curves with α = 1.5 for the three prior parameter estimation techniques

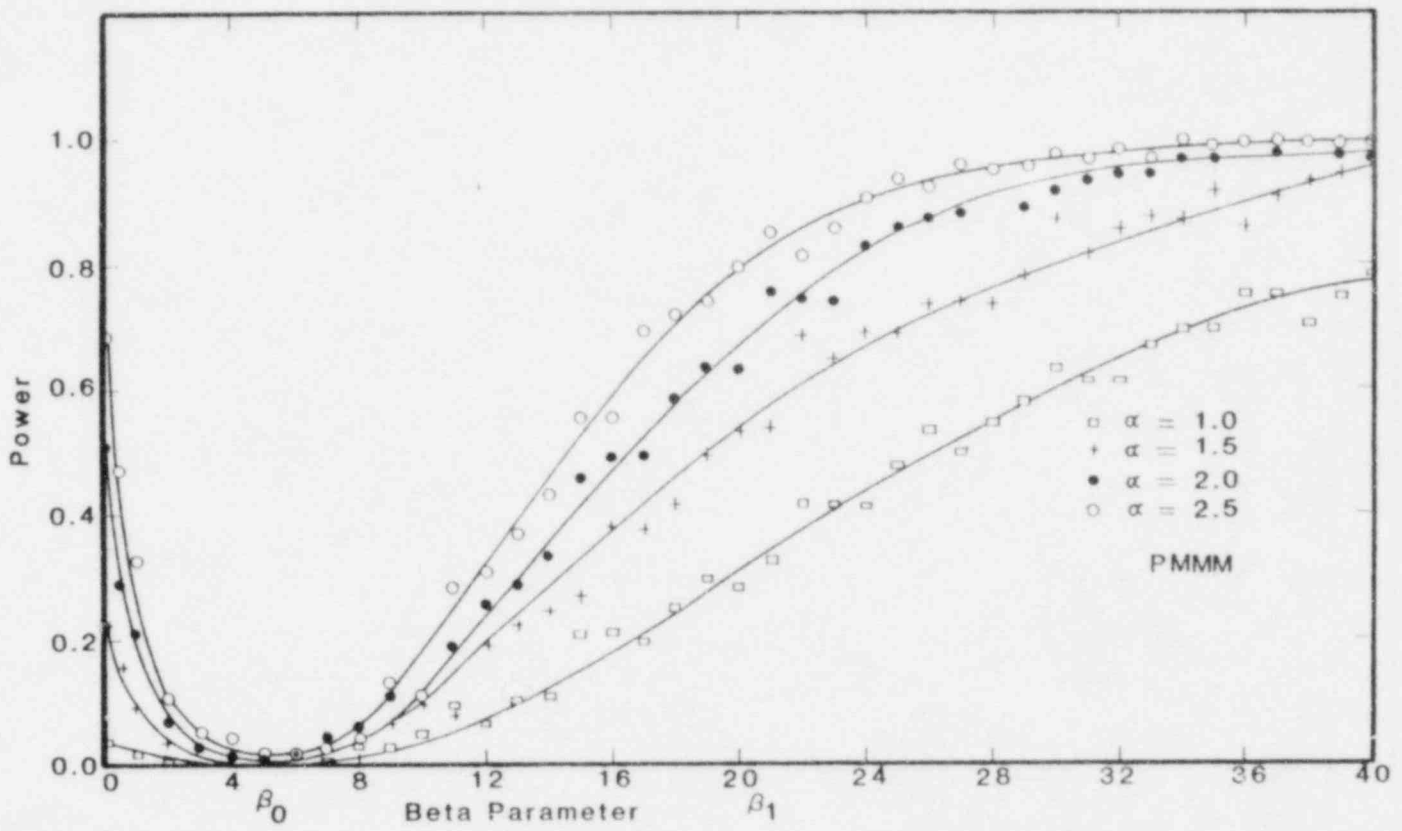


Figure 1.2: Empirical power curves with $\alpha = 2.0$ for the three prior parameter estimation techniques

1.4 DISCUSSION OF RESULTS

With the MMMM and MMLM, many sets of simulated failure data were excluded from the calculation of the power curves. For example with α=1.5 and β=0.00002, the MMLM yielded parameter estimates in only 41 out of the 250 data sets analyzed. With the same data sets, the MMMM yielded estimators in only 62 cases. By contrast, the simpler PMMM (prior matching moments method) gave estimators for 239 of the data sets. For the entire range of parameters used in this study (typical of pump failure rates in nuclear power plants), the two estimation techniques based on the marginal distribution tended to fail to yield estimators a significant fraction of the time. The MMMM failed when negative estimators were obtained (not allowed from the definition of the gamma distribution), while the MMLM failed when it was unable to find estimators within a specified region about the true values (taken as 100 times the true value). On the other hand, the PMMM failed only when all the observed failures in a data set were zero.

Despite the above difficulties, this work has shown that power curves can be generated from simulated failure data. The following observations about these power curves are of interest:

1. The power curves are not symmetrical about the standard value of the scale parameter, i.e., β=0.00005, since the allowed beta parameter space is zero to infinity. Perhaps a logarithmic scale for beta would yield more symmetric distributions.

2. All power curves exhibit a stochastic rather than a smooth behavior, as a result of the stochastic manner in which they were generated. Larger sample sizes would reduce these spurious stochastic fluctuations.

3. All power curves appear to approach unity asymptotically for large values of beta, indicating the null hypothesis is always rejected if the beta value used for the simulated data is sufficiently different from the standard value.

4. Each power curve has the shape of a negative pulse near the standard beta value. At β=0.00005 these curves indicate the power of the test is well below 0.05, although the exact minimum value cannot be determined exactly because of the stochastic nature of these empirical power curves.

The power curve negative pulse behavior near the standard beta value of 0.00005 can be characterized by its full width at half minimum (FWHM) or, since the power curves are assymetric about β_o, by the right or left half width at half minimum (RHWHM or LHWHM, respectively), i.e., the difference between the beta values at which the power is 0.5 and the value β_o. For beta values at which the power of the likelihood test equals 0.5 (endpoints of the FWHM), the null hypothesis is correctly rejected only 50 percent of the time. Ideally a power curve for a two sided test should dip to zero at the true parameter value (here beta = 0.00005) and should equal unity for all other parameter values. Thus

for the ideal case the FWHM equals zero. The smaller the value of FWHM, or the steeper the sides of the negative pulse of the power curve, the better is the test at distinguishing between marginal distributions.

In Table 1.1 the RHWHM for the power curves generated in this study are listed. From these results, it is seen that for a given α value the MMLM yields the smallest value for the RHWHM. The values obtained by the PMMM and MMMM are equal and are about 30 to 40 percent larger than the RHWHM from the MMLM. As the shape parameter of the gamma-Poisson distribution increases, it is seen that the RHWHM decreases, i.e. the likelihood ratio test becomes more discriminating.

TABLE 1.1

RHWHM for Generated Power Curves. Gamma shape parameter β_o=0.00005

Alpha	Est. Technique	RHWHM/β_o
1.5	PMMM	2.84
1.5	MMMM	3.25
1.5	MMLM	1.70
2.0	PMMM	2.32
2.0	MMMM	2.32
2.0	MMLM	1.45
1.0	PMMM	4.25
1.5	PMMM	2.80
2.0	PMMM	2.32
2.5	PMMM	1.95

Chapter 2

DETECTION OF OUTLIERS IN FAILURE DATA

2.1 INTRODUCTION

Frequently, failure rates are estimated from available failure data found in the literature. Unfortunately, the available failure data for nuclear power plant components are usually sparse, and there are often large variations in the data for similar components. Thus, the estimated values for failure rates have large uncertainties.

One method which can be used to reduce this uncertainty is to label data points that deviate greatly from the main body of the data as outliers. Once data points have been labeled as outliers, steps can be taken to reduce the effect which they have on the statistical interpretation of the remaining data. These steps can range from simply noting that outliers are present in order to reduce the confidence placed in the results, weighting the data in order to reduce the effect of the outliers, or rejecting the outliers before the data are interpreted.

The problem of outliers or discordant data has been recognized for decades. It is one of the most perplexing problems which occurs in the interpretation of data, because all elements of subjectivity must be removed and because the question of what to do with discordant data, once they are found, has no universally accepted answer. Ferguson [4] stated the problem as follows:

> "In a sample of moderate size taken from a certain population, it appears that one or two observations are surprisingly far away from the main group. The experimenter is tempted to throw away the apparently erroneous values, and not because he is certain that the values are spurious. On the contrary, he will undoubtedly admit that even if the population has a normal distribution there is a positive although extremely small probability that such values will occur in an experiment. It is rather because he feels that other explanations are more plausible, and that the loss in the accuracy of the experiment caused by throwing away good values is small compared to the loss caused by keeping even one bad value. The problem, then, is to introduce some degree of subjectivity into the rejection of outlying observations."

In this chapter, the specific problem of discordant data in failure rate data, with small failure rates, is addressed. The detection of outliers in three types of failure data were investigated, and the re-

sults are summarized in this chapter. Additional details on this phase of the investigation and many examples can be found in Ref. [5]. The following section deals with the idealized case in which the failure rates themselves are known and are assumed to be distributed according to a gamma distribution. Several methods of outlier detection for the gamma distribution are presented, some properties and comparisons among the methods are given.

Section 2.3 deals with the more practical case in which the data are in the form (F_i,T_i), where F_i is the number of component failures observed in test time T_i. For a particular component, the number of failures is assumed to be distributed according to the Poisson distribution, and the failure rates of similar components are assumed to be distributed according to a gamma distribution. Two tests for detecting outliers in data distributed according to the gamma-Poisson compound model are presented, and some properties and a comparison between the two methods are then given.

In Section 2.4, the problem of discordant values in time-to-failure data, i.e., data which are distributed according to the exponential distribution, is discussed. Two methods of detecting outliers in this type of data are presented, and some properties and a comparison between the two methods are given.

In the final section of this chapter, an overview of the problem of outliers in failure rate data is discussed, and the "best" methods for outlier detection are recommended.

2.2 OUTLIER DETECTION FOR FAILURE RATE DATA

In this section, methods for outlier detection in failure rate data, for which the failure rate gamma distribution is known, are discussed.

2.2.1 Fisher's Method for Outlier Detection

Fisher [6] developed a method for detecting a single outlier in a gamma distribution, $g(x;\alpha,\beta)$, with known parameter α. His method has since been expanded upon, notably by Fieller [7] who generalized the method to include the case of multiple outliers. Fisher's method of detecting outliers is a hypothesis test. The null hypothesis, H0, is that all data come from a single gamma distribution with a known fixed value of α. The alternative hypothesis, H1, is that a certain number of data points, $k<n$, come from a distribution(s) different from that associated with the main body of data.

The appliation of Fisher's method for detecting outliers consists of calculating a test statistic, TF, and comparing this statistic with the tabulated critical values, tF, of the statistic. If TF is greater that tF, then the null hypothesis is rejected, and the data points under scrutiny are declared to be discordant.

The test statistic for upper outliers is

$$TF = \sum_{i=n-k+1}^{n} x_i \Big/ \sum_{i=1}^{n} x_i, \qquad (2\text{-}1)$$

where $x_n \geq x_{n-1} \geq \ldots \geq x_1$ are the data and k is the number of outliers being tested for. The critical value of this test statistic is found from the equation [8]

$$P(TF \leq tF) \leq \binom{n}{k} P\left[F_{2k\alpha,2(n-k)\alpha} < \frac{(n-k)tF}{k(1-tF)}\right], \qquad (2\text{-}2)$$

where P(TF<tF) is the probability that the calculated value of Fisher's test statistic TF is less that its critical value tF, n is the number of data points, k is the number of outliers being tested for, and $F_{2k\alpha,2(n-k)\alpha}$ is the F-distribution with $2k\alpha$ and $2(n-k)\alpha$ degrees of freedom. There is no "right" or "wrong" way to select the value of P(TF<tF) which is used in determining tF. However, it is standard practice to use either 0.01 or 0.05. In doing so, the values of tF at either the 99% or the 95% confidence level, respectively, are obtained.

2.2.2 The Normal Conversion Method for Outlier Detection

The method of normal conversion consists of transforming a sample from a gamma distribution into a sample from a normal distribution, and then testing the normal sample for outliers. If the random variable X is distributed according to the gamma distribution $g(X;\alpha,\beta)$, $\alpha > 1$, it can be transformed into a random variable W which is approximately distributed according to the normal distribution

$$N\left[(\alpha\beta)^{1/3}\left(1 - \frac{1}{a\alpha}\right), \frac{1}{9}\left(\beta^2/\alpha\right)^{1/3}\right]$$

by taking the cube root of X [9]. Thus if $w_i = x_i^{1/3}$ are data, where the x_i are distributed according to a gamma distribution with the parameter α greater than unity, then w_i are approximately distributed according to a normal distribution.

Once the data have been transformed to normally distributed data, there are a number of methods available for detecting outliers. One of the more common methods, both because of its ease of application and its intuitive appropriateness, is the use of Grubbs-type statistics. These statistics, like the Fisher statistics, are based upon a hypothesis test. The null hypothesis H0 in this case is that all the transformed data come from a single normal distribution. The alternative hypothesis H1 is that a certain number of the data, k<n, come from a distribution(s) distinct from the normal distribution that describes the main body of data.

The Grubbs test statistic used to test the null hypothesis is calculated as follows [10]:

$$TN = \frac{1}{S}\left(\sum_{i=n-k+1}^{n} w_i - k\bar{w} \right) , \qquad (2\text{-}3)$$

where $w_1 \leq \ldots \leq w_n$ are ordered values of the normalized data, k is the number of suspected outliers, and S is the standard deviation of the transformed data. The critical value of the test statistic, tN, is found from the equation [8]

$$P(TN \geq tN) \leq \binom{n}{k} P\left\{ t_{n-2} > \left[\frac{n(n-2)\ tN^2}{k(n-k)(n-1) - ntN^2} \right]^{1/2} \right\} \qquad (2\text{-}4)$$

where P(TN>tN) is the probability that the Grubbs test statistic TN is greater than or equal to its critical value tN, t_{n-2} is Student's t-distribution with n-2 degrees of freedom, and tN is the critical value of TN for a given P(TN>tN). As before, the value of P(TN>tN) (selected to determine the critical value of TN) is usually chosen either as 0.01 or 0.05, so that either the 99% or 95% confidence level, respectively, is used.

2.2.3 The Integration Method For Outlier Detection

A third method for locating outliers in a gamma distribution involves the integration of the gamma distribution in question. As before, the null hypothesis H0 is that all data come from a single gamma distribution, while the alternative hypothesis H1 is that a few points, k<n, are from a different distribution(s) than the main body of the data.

The integration method is developed as follows. The probability that a data point from the gamma distribution $g(x;\alpha,\beta)$ is greater than a given value x_c is given by

$$p_i = \int_{x_c}^{\infty} g(x;\alpha,\beta)\, dx = 1 - G(x_c;\alpha,\beta) \quad , \qquad (2\text{-}5)$$

where G is the gamma cumulative distribution. The probability that any one of the n data points is greater than x_c (assuming independence among the data) is given by

$$P\left(\bigvee_{i=1}^{n} A_i \right) = 1 - P\left(\bigwedge_{i=1}^{n} A_i \right) = 1 - \prod_{i=1}^{n} (1-p_i) \quad , \qquad (2\text{-}6)$$

where A_i is the event $x_i > x_c$, A_i' is the complement of A_i, V is the union symbol, and is the intersection symbol. And if, according to the null hypothesis, all data come from the same distribution, then $p_1 = p_2 = \ldots = p_n \equiv p$, and thus, from Eq. (2-6)

$$P\left(\bigvee_{i=1}^{n} A_i\right) = 1-(1-p)^n \tag{2-7}$$

$$p = 1-\left[1-P\left(\bigvee_{i=1}^{n} p_i\right)\right]^{1/n} . \tag{2-8}$$

From Eq. (2-5)

$$\int_{x_c}^{\infty} g(x;\alpha,\beta)\, dx = p \tag{2-9}$$

which, solving for x_c, gives

$$x_c = G^{-1}(1-p;\alpha,\beta) \tag{2-10}$$

As in the previous methods, the 95% or 99% confidence levels are usually of interest, and so p is either taken as the value $1-.95^{1/n}$ or $1-.99^{1/n}$, accordingly.

It is not necessary to solve exactly Eq. (2-9) in order to detect outliers. Rather, the data points $x_n, x_{n-1} \ldots$, where $x_n \geq x_{n-1} \geq \ldots$, can be used as approximations of x_c. These values are substituted into Eq. (2-9), beginning with x_n, until a value x_{n-k+1} is reached such that

$$\int_{x_{n-k+1}}^{\infty} G(x;\alpha,\beta)\, dx < p \tag{2-11}$$

The upper k data points are then labeled as outliers.

2.2.4 Estimation of the Gamma Parameters

Until now, the values of the gamma parameters, α and β, have been assumed known. Often, however, values of these parameters will not be known *a priori*. When the Fisher's test statistic is applied to the failure data, the following algorithm can be used to determine the value of the alpha parameter to be used in finding the critical level of the test statistic. First $\hat{\alpha}$ is calculated from the non-suspicious failure data using the prior matching moments technique (PMMM) as follows [1]:

$$\bar{x}' = \frac{1}{n-k} \sum_{i=1}^{n-k} x_i \ , \qquad (2\text{-}12)$$

$$s'^2 = \frac{1}{n-k-1} \sum_{i=1}^{n-k} (x_i - \bar{x}')^2 , \qquad (2\text{-}13)$$

$$\hat{\alpha}' = \bar{x}'^2/s'^2 \ , \qquad (2\text{-}14)$$

where $x_1 \leq x_2 \leq \ldots \leq x_n$, and k are the suspected number of upper outliers. Then, the minimum outlier value, i.e., the minimum value at which a data point is labeled as outlier, is found using $\hat{\alpha}'$ to determine the critical value of the test statistic,

$$x_c' = \frac{tF'}{1-tF'} \sum_{i=1}^{n-k} x_i \ , \qquad (2\text{-}15)$$

where tF' is the critical value of Fisher's test statistic at $\hat{\alpha}'$. This minimum outlier value is then used to estimate $\hat{\alpha}$ as follows:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n-k} x_i + \frac{k}{n} x_c' \ , \qquad (2\text{-}16)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n-k} (x_i - \bar{x})^2 + \frac{k}{n-1} (x_c' - \bar{x})^2 \qquad (2\text{-}17)$$

$$\hat{\alpha} = \bar{x}^2/s^2 \ . \qquad (2\text{-}18)$$

The above value of $\hat{\alpha}$ is then used to find the critical value of the test statistic. Fisher's method of Section 2.2.1 is then used to determine whether or not the upper k data points are discordant.

When the integration method is being applied to locate outliers, the algorithm used to estimate values of the gamma parameters is very similar to that outlined above. First the data points that are <u>not</u> suspected as being discordant are used to estimate the gamma parameters as follows:

$$\bar{x}' = \frac{1}{n-k} \sum_{i=1}^{n-k} x_i \ , \qquad (2\text{-}19)$$

$$s'^2 = \frac{1}{n-k-1} \sum_{i=1}^{n-k} (x_i - \bar{x}')^2 \ , \qquad (2\text{-}20)$$

$$\hat{\alpha}' = \bar{x}'^2/s'^2 \ , \qquad (2\text{-}21)$$

$$\hat{\beta}' = \bar{x}'/s'^2 \ . \qquad (2\text{-}22)$$

These values of the gamma parameters, $\hat{\alpha}'$ and $\hat{\beta}'$, are then used to determine the minimum outlier value for the n-k data points from the equations

$$P = \int_0^{x'_c} g(x;\alpha',\beta')\, dx \equiv G(x'_c;\alpha',\beta'), \qquad (2\text{-}23)$$

which can be solved to give

$$x'_c = G^{-1}(P;\alpha',\beta'), \qquad (2\text{-}24)$$

where P is taken as either $0.95^{1/n}$ or $0.99^{1/n}$ depending on whether the 95% or 99% confidence level, respectively, is to be used for hypothesis testing.

Then, using the n-k non-suspicious data points and x', the gamma parameters, $\hat{\alpha}$ and $\hat{\beta}$, which are used to test the upper k data points for discordancy, are estimated as follows:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n-k} x_i + \frac{k}{n} x'_c \ , \qquad (2\text{-}25)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n-k} (x_i - \bar{x})^2 + \frac{k}{n-1} (x'_c - \bar{x})^2 \ , \qquad (2\text{-}26)$$

$$\hat{\alpha} = \bar{x}^2/s^2 \ , \qquad (2\text{-}27)$$

$$\hat{\beta} = \bar{x}/s^2 \ . \qquad (2\text{-}28)$$

Once the estimators $\hat{\alpha}$ and $\hat{\beta}$ have been computed, the integration method, described in Section 2.2.3, is used to determine if the upper k data points are, in fact, discordant.

2.2.5 Determination of the Number of Outliers

One method for determining the number of outliers in a set of data is to test repetitively the data, assuming for each test that one outlier may be present. That is, the data are tested first assuming the most extreme value may be discordant. If that value is not found to be discordant, then there are no outliers present, and the testing is complete. However, if the most extreme value is found to be an outlier, it is removed from the data, and the test is repeated assuming that the second most extreme value may be an outlier. If this value is found to belong to the distribution being tested, then the testing is complete, and there is only one outlier in the data. However, if this value is found to be discordant, it is removed from the data, and the third most extreme value is tested. This process is continued until a non-discordant value is found. If k+1 repetitions are necessary to find a non-discordant value, then there are k outliers in the data.

The repetition method for determining the number of outliers in a set of data seems, at first glance, to be quite reasonable. However, there is a major pitfall in its application arising from the phenomenon called masking. Suppose a set of data with two upper outliers is to be tested. And further suppose that the outlier detection technique to be applied is of the form D/S where D is a measure of separation of the most extreme value from the mean, and S is a measure of the spread of the data. The greater the value of the test statistic D/S, the more likely it is that the data point under scrutiny will be detected as an outlier. In this case, however, the second extreme value is also an outlier, and thus will cause the value of S to be larger and the value of D/S to be smaller than anticipated. As a result, the most extreme value may not be detected as being an outlier, and thus the two outliers would go undetected.

Another pitfall, known as swamping, arises in determining the number of outliers to be tested for when a non-discordant value is tested along with a discordant value. Suppose that a data set with only one outlier is to be tested for two outliers. And futher suppose that, as in the previous case, a test statistic of the form D/S is to be used. In this case, both D and S would be smaller than anticipated, and, depending on the discrepancy between the two tested values and the mean, both values would be labeled as either discordant or non-discordant.

The above considerations show that the simplistic testing of the data one-by-one for outliers is not the correct way to proceed. The correct way is as follows. The data are first examined to locate the largest gap between two consecutive data points, the data being ordered such that $x_1 \leq x_2 \leq \ldots \leq x_n$. If this gap occurs between the two points x_{n-k} and x_{n-k+1}, and if $x_{n-k+1} > \bar{x}$, the k data points x_{n-k+1}, x_{n-k+2}, ..., x_n

are considered to be potential upper outliers. These points are then tested to determine if they are, in fact, outliers or not. If they are not found to be discordant, then there are no outliers, and the testing is complete. If they are found to be discordant then they are removed from the data, and the test is repeated to determine whether or not there is another "cluster" of outliers present.

2.2.6 Power Curves and Their Generation

The properties and comparisons of the various tests which are described in the previous sections are determined by the use of 'power curves'. For our purposes, the y-axis of a power curve is the probability that a single outlier is detected, and the x-axis is a measure of the variation between the mean of the parent distribution and the mean of the distriution from which the outlier comes. In particular, the x variable is taken as $K = \mu_o/\mu_p$, where μ_p is the mean of the parent distribution and μ_o is the mean of the distribution generating the outlier. For the results presented in this chapter, the variances of the two distributions are set equal to each other. Note that $K = 1$ means that all data points come from the same distribution. Thus at $K = 1$, the probability of an outlier being detected will be less than or equal to 0.05 for the 95% confidence level.

Power curves have often been determined by generating sets of data from a parent distribution and then generating outliers from a second distribution (simulation of data). The simulated outliers were then tested for discordancy against the data sets from the parent distribution to obtain a point on the power curve. This procedure is then repeated for different outlier distributions until the entire power curve is obtained. Such an empirical approach for determining power curves results in apparent stochastic fluctuations in the curves because of the finite number of data sets used to constuct each point on the curves.

In this study, a modified approach for constructing power curves was taken. In this technique, it is necessary to generate simulated failure data sets from only the parent distribution. A semi-analytical treatment can then be used to obtain the probability that an outlier will be detected. This novel technique for generating power curves mitigates the stochastic nature of the empirical curves which is inherent in the conventional procedure. This new method is described below.

2.2.6.1 Power Curve Calculation for Tests of a Single Outlier

The modified approach for obtaining power curves for a hypothesis test for detecting a single upper outlier in a data set is developed as follows. The null hypothesis HO is that all the data $x_1 \leq x_2 \leq \ldots \leq x_n$ come from the same distribution, i.e., the suspected outlier x_n comes from the same distribution as the rest of the data. Suppose the hypothesis test is based on a statistic

$$T = f(x_1, x_2, \ldots, x_{n-1}, x_n) \qquad (2\text{-}29)$$

whose critical value $t(\alpha)$ is known at some α significance level. The null hypothesis is rejected if $T > t(\alpha)$. An equivalent test of the null hypothesis is to calculate a critical value x_c for x_n that yields $t(\alpha)$ as the test statistic T, i.e.,

$$t(\alpha) = f(x_1, x_2, \ldots, x_{n-1}, x_c) \qquad (2\text{-}30)$$

so that $T>t(\alpha)$ if and only if $x_n>x_c$. For a given data set consisting of n values, a random variable x_c can be calculated from the n-1 smallest values using the above equation, and the null hypothesis is rejected at the α-significance level if $x_n > x_c$.

The new technique for obtaining a power curve for the above hypothesis test is based on finding the density distribution of the random variable x_c, $f(x_c)$. For a specified sample size n and a specified parent distribution $g(x;\alpha,\beta)$ where α and β are the distribution parameters, the distribution of x_c can be determined empirically by generating multiple data sets from g and solving Eq. (2-30) for x_c from each data set.

Once $f_c(x_c)$ has been determined, the joint distribution of x_c and x_n is calculated as $f_c(x_c)g(x_n;\alpha_1,\beta_1)$ where $g(x_n;\alpha_1,\beta_1)$ is the distribution of x_n. From this joint distribution, the probability that the null hypothesis is rejected, i.e. $P\{x_n>x_c\}$, can be calculated as

$$P\{\text{H0 is rejected}\} = \iint_{\{x_n>x_c\}} f(x_c)\; g(x_n;\alpha_1,\beta_1)\; dx_c\; dx_n \qquad (2\text{-}31)$$

$$= \int_0^{\infty}\left(\int_0^{x_n} f_c(x_c)\; dx_c\right)\; g(x_n;\alpha_1,\beta_1)\; dx_n \qquad (2\text{-}32)$$

or, in terms of the cumulative distribution of x_c, $F_c(x_c)$,

$$P\{\text{H0 is rejected}\} = \int_0^{\infty} F_c(x_n)\; g(x_n;\alpha_1,\beta_1)\; dx_n\;. \qquad (2\text{-}33)$$

In practice, the cumulative distribution $F_c(x_c)$ will be determined empirically and hence will be in the form of an empirical distribution so that Eq. (2-33) will be replaced by its discrete analogue with the integral replaced by a sum of integrals over each class of the empirical distribution.

The advantage of this method for constructing a point on the power curve for a particular hypothesis test over the more conventional simulation procedure is considerable. In this new technique, only the dis-

tribution of critical x-values, f_c or F_c, must be obtained from multiple sets of simulated failure data $(x_1,\ldots,x_{n-1})$ generated from the parent distribution. In the more conventional procedure, additional simulated failure data must be obtained from $g(x_n;\alpha_1,\beta_1)$ and then each x_n tested against a large number of data sets from the parent distribution for many different values of x_n. The new technique, by using Eq. (2-33) or its discrete analogue, avoids the random sampling from $g(x_n;\alpha_1,\beta_1)$ as well as the repetitive testing of each x_n with a large number of data sets. The computational effort is thus greatly reduced while the stochastic errors in such power calculations are also reduced.

2.2.7 Properties and Comparison of Methods

There are two properties which all of the outlier tests have in common. First, as the value of the gamma parameter α increases, the power of the tests increases, i.e., saturation occurs at lower values of K. And second, except for the integration method, there is little variation in the power of the tests as the value of the parameter β changes [see Fig. (2.1)].

An interesting property of the normal conversion method is that as the value of the alpha parameter increases, the power of the normal conversion method seems to approach the power of Fisher's method. This type of behavior is expected, since the conversion of the data, which are distributed according to a gamma distribution, to approximately normally distributed data gives a better approximation as alpha increases [9]. A typical result is shown in Fig. (2.2).

The power of Fisher's method when the value of the alpha parameter is unknown is always less than the power of Fisher's method when alpha is known [see Fig. (2.3)]. Thus, when the alpha parameter is unknown, Fisher's method is conservative in that suspicious data are less likely to be labeled as outliers.

The integration method of detecting discordant data is not very powerful compared to the other methods as can be seen from Fig. (2.3). However, its power increases as the parameter alpha increases and the parameter tau decreases [see Fig. (2.4)].

The second most powerful method for all values of the gamma parameter alpha and tau which were tested is Fisher's method. However, it is not always possible to apply Fisher's method while staying within its theoretical limitations because the value of alpha is not always known.

The most powerful of the three outlier detection methods for all values of the gamma parameters is the normal conversion method. Not only is the normal conversion method the most powerful technique, but it can always be applied to failure data since it is not necessary to know the values of the gamma parameters. Also, this method is easy to use with a hand calculator. Thus, the normal conversion method is the "best" method for detecting a single outlier in failure data from a gamma disribu-

tion. This superiority was verified for gamma parameters with values $1 < \text{alpha} < 4$, and $100 < \text{tau} < 10$.

Although all of the power curves were generated at the 95% confidence level, the probability of outlier detection at K=1 is in every instance well below 0.01 [see Figs. (2.1) to (2.4)]. Thus, although the theoretical confidence level is 95%, type II errors, i.e., acceptance of the null hypothesis when it should be rejected, cause the actual confidence level to be well above 99%. Also, since the actual confidence levels[1] of the various tests are not equal, it may not be completely "fair" to compare their powers in the above manner.

---

[1] The probability of accepting H0 (all data are from the same distribution) is less than the critical level of significance, chosen a priori. For example, when a level of significance was chosen as 0.05 (a 5% chance of incorrectly rejecting H0), there was actually a much smaller chance than 5% of incorrectly rejecting H0.

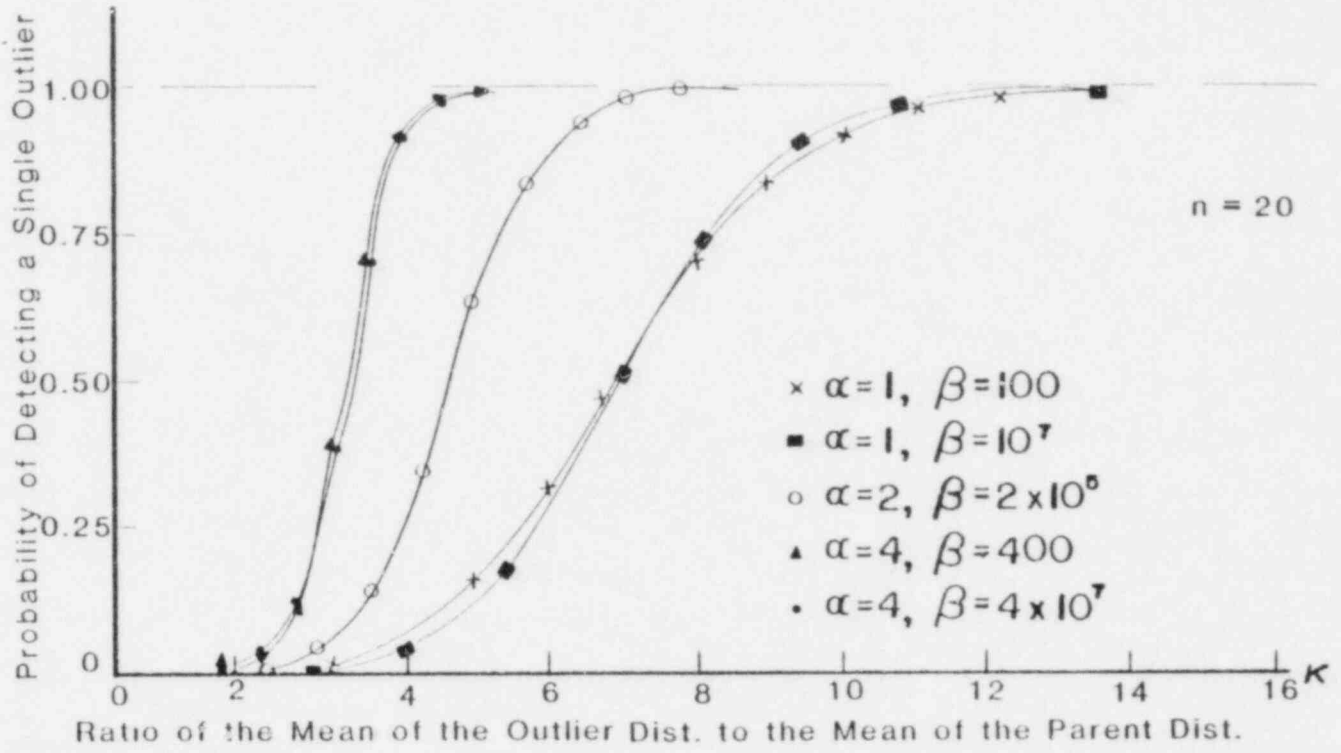


Figure 2.1: Variation in the power of Fisher's method for detecting a single outlier as the prior gamma parameters change.

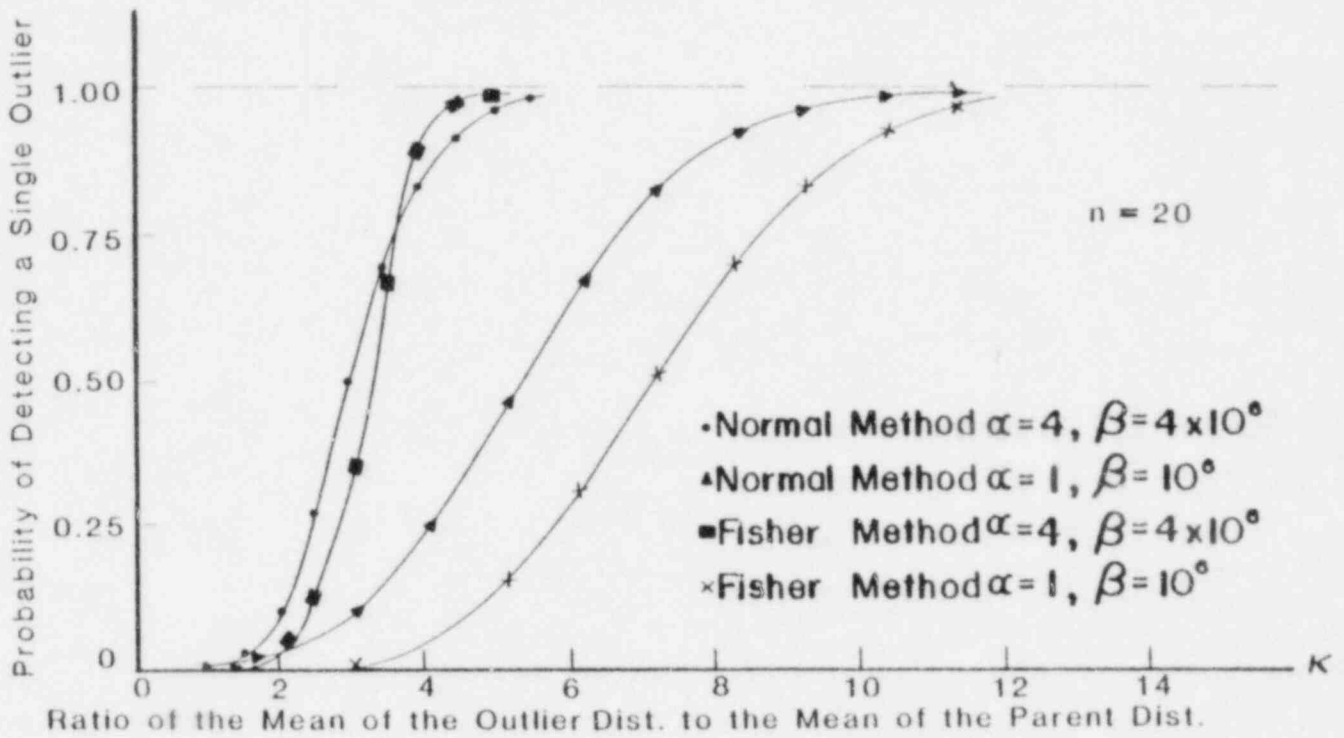


Figure 2.2: Relation between the powers of the normal conversion method and Fisher's method as the parameter alpha changes.

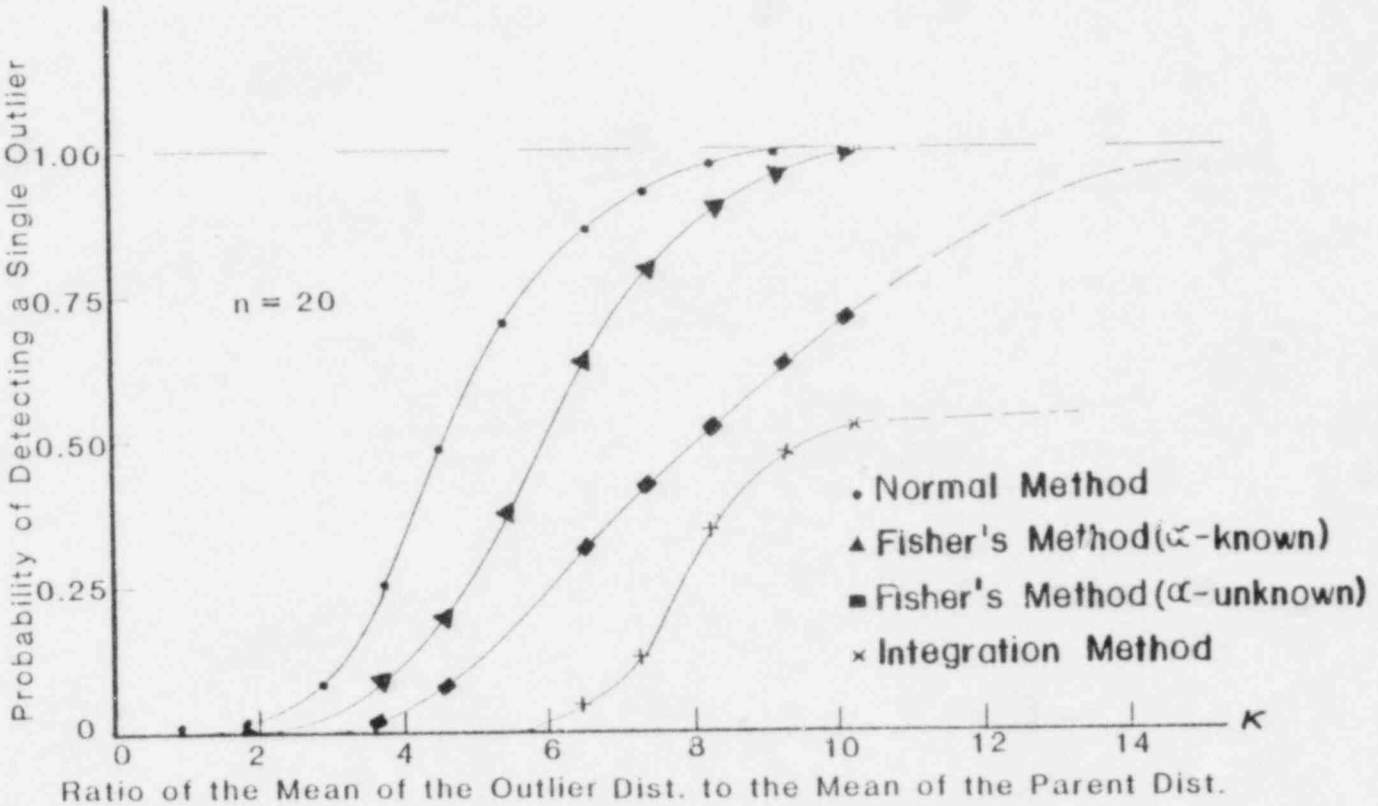


Figure 2.3: Power curves for the normal conversion method, the two Fisher's methods, and the integration method.

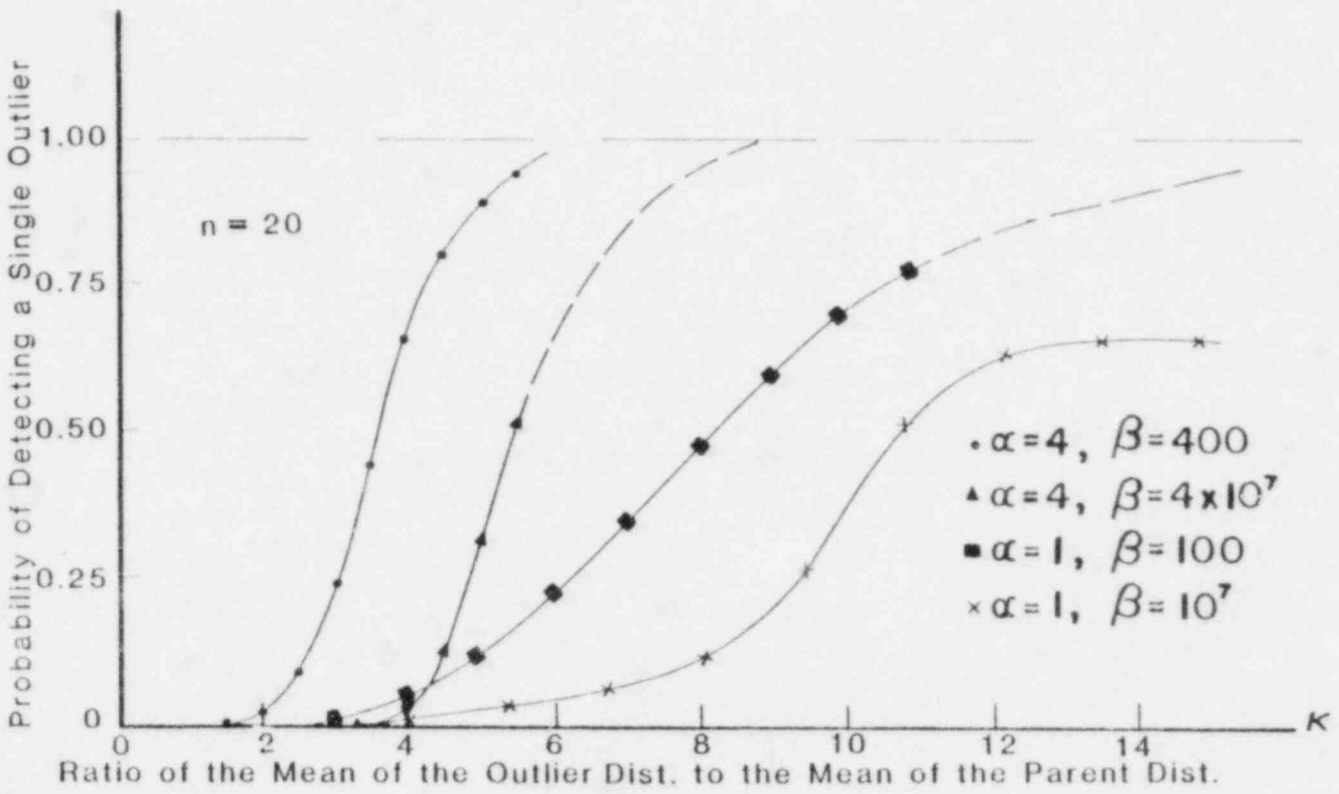


Figure 2.4: Variation in the power of the integration method as the prior gamma parameters change.

2.3 OUTLIER DETECTION IN COMPONENT FAILURE-TIME DATA

In the previous sections, the problem of detecting outliers in a gamma distribution was discussed in connection with failure rate data for which the failure rates themselves, i.e., the λ's, were observed. Normally, component failure data will consist of the number of failures F_i in a test time T_i for the i-th component.

2.3.1 Models for the (F,T) Data

2.3.1.1 The Homogeneous Model

In the classical description of a component, the failure rate of a component, λ , is regarded as an unknown constant which does not change even if the component fails and is subsequently repaired. Thus the number of failures F in operation time T is modeled by the Poisson distribution (sometimes called the likelihood distribution) of Eq. (1-1). A further asssumption is that similar components have the same failure rate λ. Thus, the failure rate can be estimated by pooling all the failure data for a given set of similar components. For example, the maximum likelihood estimator (MLE) of the failure rate based upon the homogeneous model is given by Eq. (1-2).

There are two major deficiencies with this model. First, the assumption that similar components have identical failure rates is usually not valid. Second, if a group of components has an extremely small failure rate, as is typical of components in nuclear power plants, and if the test time is not sufficiently long, it is possible that no failures would be observed. In this case, the estimated failure rate using the homogeneous model would be zero which is an unrealistic value.

2.3.1.2 The Compound Model

The compound model for the failure data generalizes the homogeneous model by allowing the failure rate to vary among the similar components. It is assumed that the failure rates are distributed according to a prior distribution, $g(\lambda)$. Thus the probability of obtaining F failures in a test time T from a component selected randomly from the component class described by g is given by Eq. (1-3) -- the marginal distribution.

For this phase of the study, it is assumed that the prior distribution is given by the gamma distribution of Eq. (1-5). This distribution is the natural conjugate to the Poisson likelihood distribution and allows the explicit evaluation of the marginal distribution (see Eq. (1-4)). In the next chapter the effect of using a non-conjugate prior family is addressed.

The use of a compound model requires values of the gamma distribution to be chosen so as to describe adequately the variation of failure rates

among the similar components. In Section 1.2.1, three methods for obtaining such parameter estimates from component failure data are summarized.

2.3.2 The Cumulative Marginal Method

For a set of failure data from n components which are assumed to be described by the same gamma prior distribution, one often encounters the problem of trying to decide if the number of failures for one component is abnormally high or low. One may reasonably decide that (F^*,T^*) is an outlier in a sample of size n if the probability of observing F^* failures in test time T^* for at least one component is less than some preselected value γ, i.e.,

$$P\left\{\bigvee_{i=1}^{n} F_i/T_i \geq F^*/T^*\right\} \leq \gamma \quad . \tag{2-34}$$

The value of γ should be selected as either 0.01 or 0.05 for the 99% or 95% confidence level, respectively.

If the data are independent (as is usually assumed), this equation may be rewritten as

$$1-P\left\{\bigwedge_{i=1}^{n} F_i/T_i < F^*/T^*\right\} \leq \gamma \ , \tag{2-35}$$

$$\prod_{i=1}^{n} P\{F_i/T_i < F^*/T^*\} \geq 1 - \gamma \tag{2-36}$$

$$\prod_{i=1}^{n} P\{F_i < F^*T_i/T^*\} \geq 1 - \gamma \ . \tag{2-37}$$

Expressed in terms of the marginal distribution, this last equation becomes

$$\prod_{i=1}^{n} \left| \sum_{F=0}^{[F^*T_i/T^*]'} h(F;T_i,\alpha,\beta) \right| \geq 1 - \gamma \ , \tag{2-38}$$

where $[F^*T_i/T^*]'$ is the largest integer strictly less than F^*T_i/T^*. In terms of the cumulative marginal distribution, Eq. (2-38) be expressed as

$$\prod_{i=1}^{n} H([F^*T_i/T^*]';T_i,\alpha,\beta) \geq 1 - \gamma \ , \qquad (2\text{-}39)$$

where the cumulative marginal distribution is

$$H(F;T,\alpha,\beta) = \sum_{i=0}^{F} h(i;T,\alpha,\beta) \ . \qquad (2\text{-}40)$$

To apply the above procedure to decide whether or not certain data points are outliers, values of the prior parameters must first be obtained. In this study, the gamma parameters alpha and tau were estimated from the entire set of failure data using one of the three estimation methods summarized in Section 1.2.1 (the PMMM, MMMM, or MMLM). A variation of this method is to exclude the data which are suspected of being outliers when estimating the gamma parameters. However, by making this change, the number of type I errors will increase, and the number of type II errors will decrease, i.e., H0 will be rejected more often when it should be accepted, and H0 will be accepted less often when it should be rejected.

This method can also be applied to failure/time data to determine whether or not the smallest number of observed failures in a data set are disproportionally small [2]. However, the testing for such lower outliers is usually not of interest for failure data from components in nuclear power plants.

2.3.3 Binomial Method

If all the test times for a set of components are equal, an alternative method can be used to test for upper outliers. Assume there are k suspected outliers in the data $F_1 \leq F_2 \leq \ldots \leq F_n$. The probability of having F_{n-k} or fewer failures is given by

$$p = \sum_{i=0}^{F_{n-k}} h(i;T,\alpha,\beta) = H(F_{n-k};T,\alpha,\beta) \quad , \tag{2-41}$$

where h and H are the density and cumulative marginal distributions, respectively. The probability of having n-k or more data points with F_{n-k} or fewer failures is thus given by

$$P_{n-k} = \sum_{i=n-k}^{n} \binom{n}{i} p^i (1-p)^{n-i} \quad . \tag{2-42}$$

Thus if $P_{n-k} \leq 0.05$ or 0.01 for the 95% or 99% confidence level, respectively, the k upper failure values are labeled as outliers.

The above binomial method can also be used in conjunction with the homogeneous model. In this case, the probability of having F_{n-k} or fewer failures is given by

$$p = \sum_{i=0}^{F_{n-k}} f(F_i|\lambda;T_i) \quad , \tag{2-43}$$

where f is the Poisson distribution given by Eq. (1-1). The probability obtained from this result is then substituted into Eq. (2-43) to determine whether or not discordant data are present.

2.3.4 Properties and Comparison of Tests

The properties and comparisons of the various tests are determined by the use of power curves. However, the determination of the power curves for both of the methods described in this section are complicated by the need to use discrete failure/time data. Because of the need to determine the power curves by a complete simulation (see Section 2.2.6), the power curves presented in this section have more uncertainty than those derived for tests on the failure rates.

The one property that all of the power curves have in common is that as the value of the prior gamma parameter tau increases, the power of the tests decreases, i.e., the power curve saturates at larger values of K (see Fig. (2.5)). The reason for this behavior is that for n = 20, T = 10,000 hr, and an average failure rate significantly less than 0.0001 per hour, data sets in which no failures are present appeared frequently in the simulated failure data.

Neither the cumulative marginal method nor the binomial method using the compound model, with the outlier included in the gamma parameter calculations, appear in any of the empirical power curves because, in no instance, did either of these methods detect an outlier. The reason for this lack of detection is that in data sets of size 20, an outlier will distort the estimated values of the parameters of the prior distribution to such an extent that it is not possible to detect the outlier.

The best method for detecting an outlier for $1 < \text{alpha} < 4$ and $10 < \text{tau} < 4{,}000{,}000$ (values for alpha and tau outside these ranges were not studied in this work) is the cumulative marginal method with the values of the prior parameters known (see Figs. (2.6) to (2.8)). However, in general, the values of the prior parameters are not known and thus, this method is not very practical.

The cumulative marginal method with the gamma parameters calculated by the marginal matching moments method (MMMM) is a very poor method, and should be used with caution for two reasons. First, for small values of the gamma parameter tau, the actual confidence level is less than the theoretical confidence level, e.g., at K=1 a good data point is labeled an outlier more often than it should be (see Fig. (2.6)). And secondly, for larger values of tau, the power of this method is lower than even the simple homogeneous model (see Figs. (2.7) and (2.8)).

For smaller failure rates, i.e., $\bar{\lambda} < 0.00005$ per hour, and all test times equal, the most powerful practical method of outlier detection is the binomial method used in conjuction with the homogeneous model (see Fig. (2.8)). If all the test times are not equal, the most powerful practical method is the cumulative marginal method with the gamma parameters estimated using the prior matching moments method (PMMM) without using the potential outlier.

For larger failure rates, i.e. $\bar{\lambda} > 0.00005$ per hour, the most powerful practical method is the cumulative marginal method with the gamma parameters calculated with the PMMM and without using the potential outlier.

Although the theoretical confidence level used in all calculations was 95%, except for the cumulative marginal method with alpha and tau calculated via the MMMM (without using the suspected outlier), the actual confidence level was always greater than 95%. This can be seen by noting that for K=1, the powers of the various methods are always below 0.05 (see Figs. (2.5) to (2.8)). However, since the actual confidence levels[2] of the various tests are not equal, it may not be completely fair to compare their powers in the above manner.

---

[2] The probability of accepting H0 (outlier and other data are from the same distribution) is less than the critical level of significance chosen a priori.

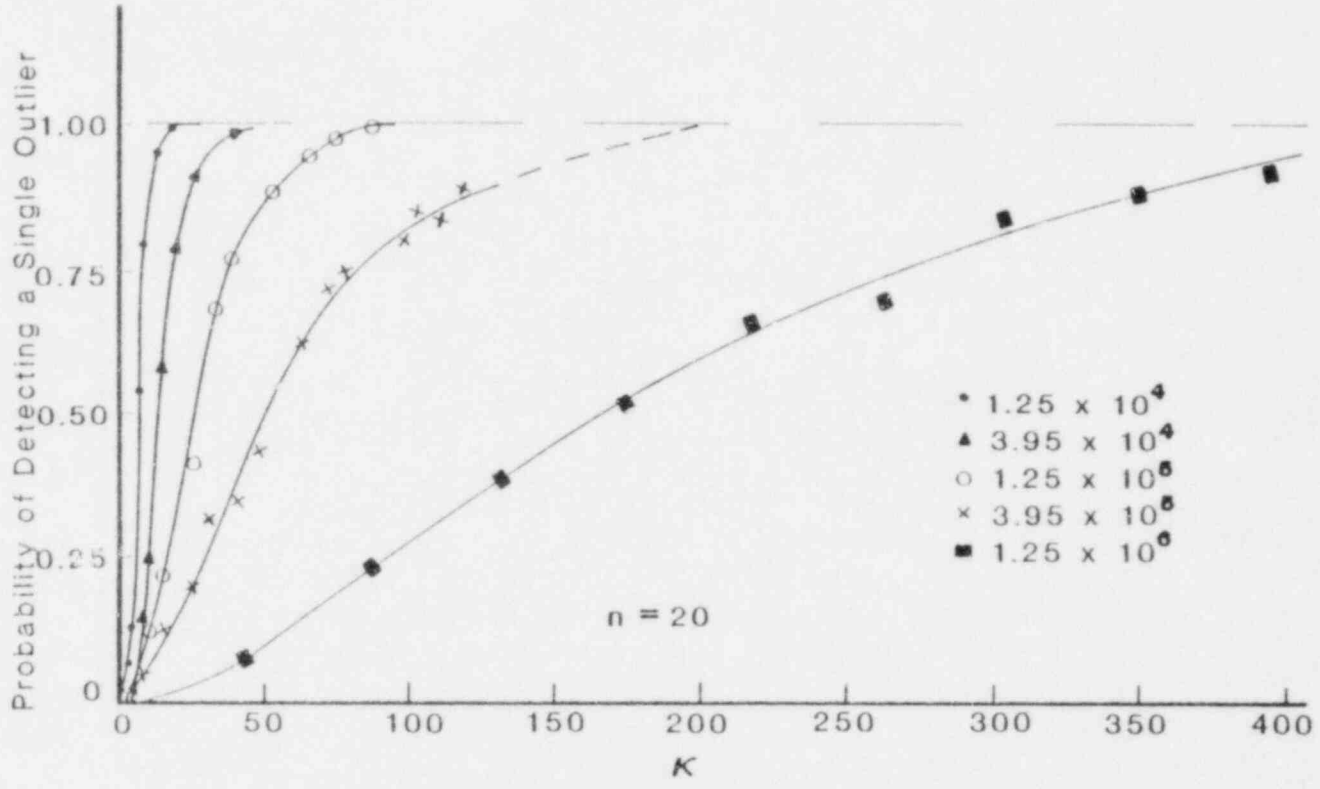


Figure 2.5: The power of the cumulative marginal method for single outlier detection for different tau values with alpha = 1.25

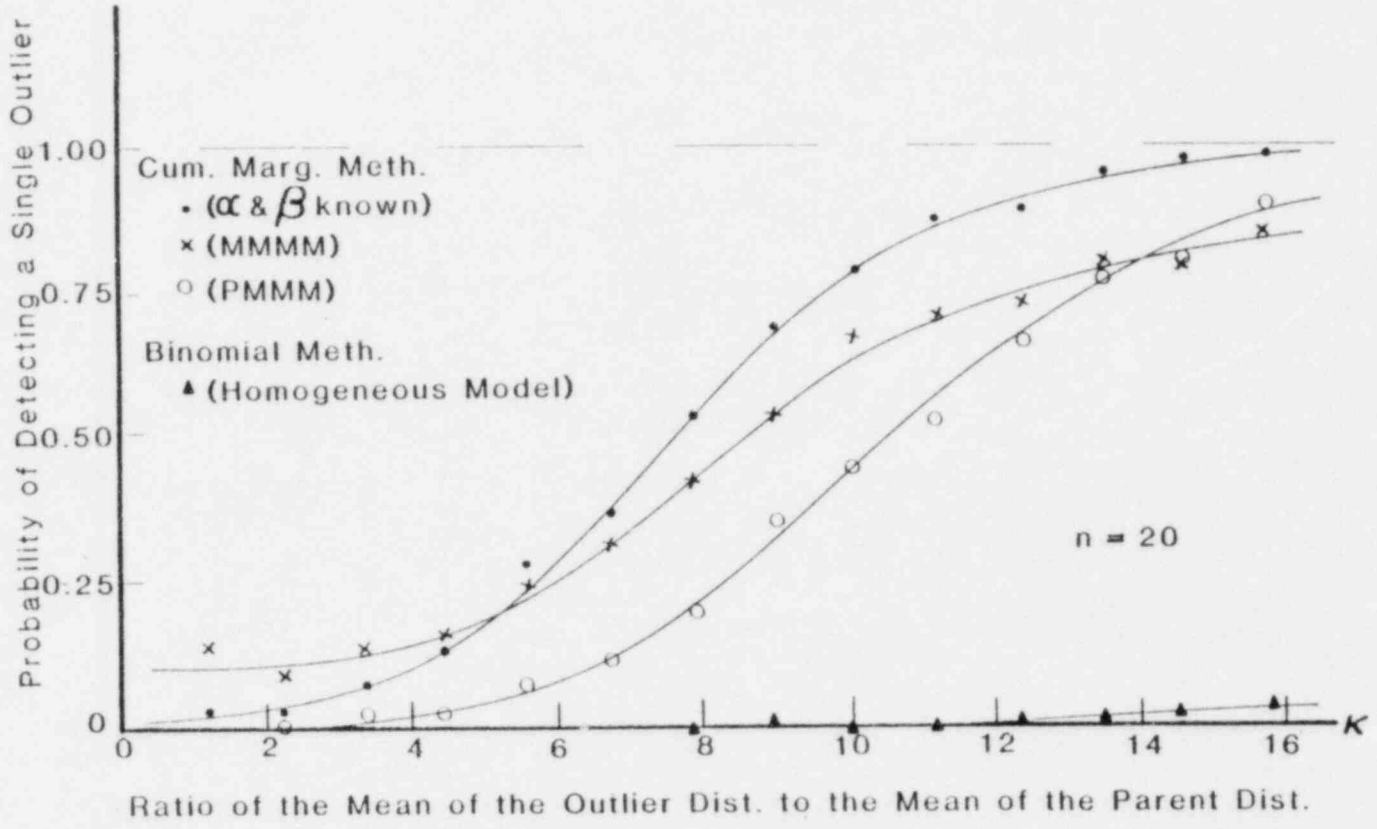


Figure 2.6: Power of single outlier detection tests with alpha = 1.25 and tau = 12,500

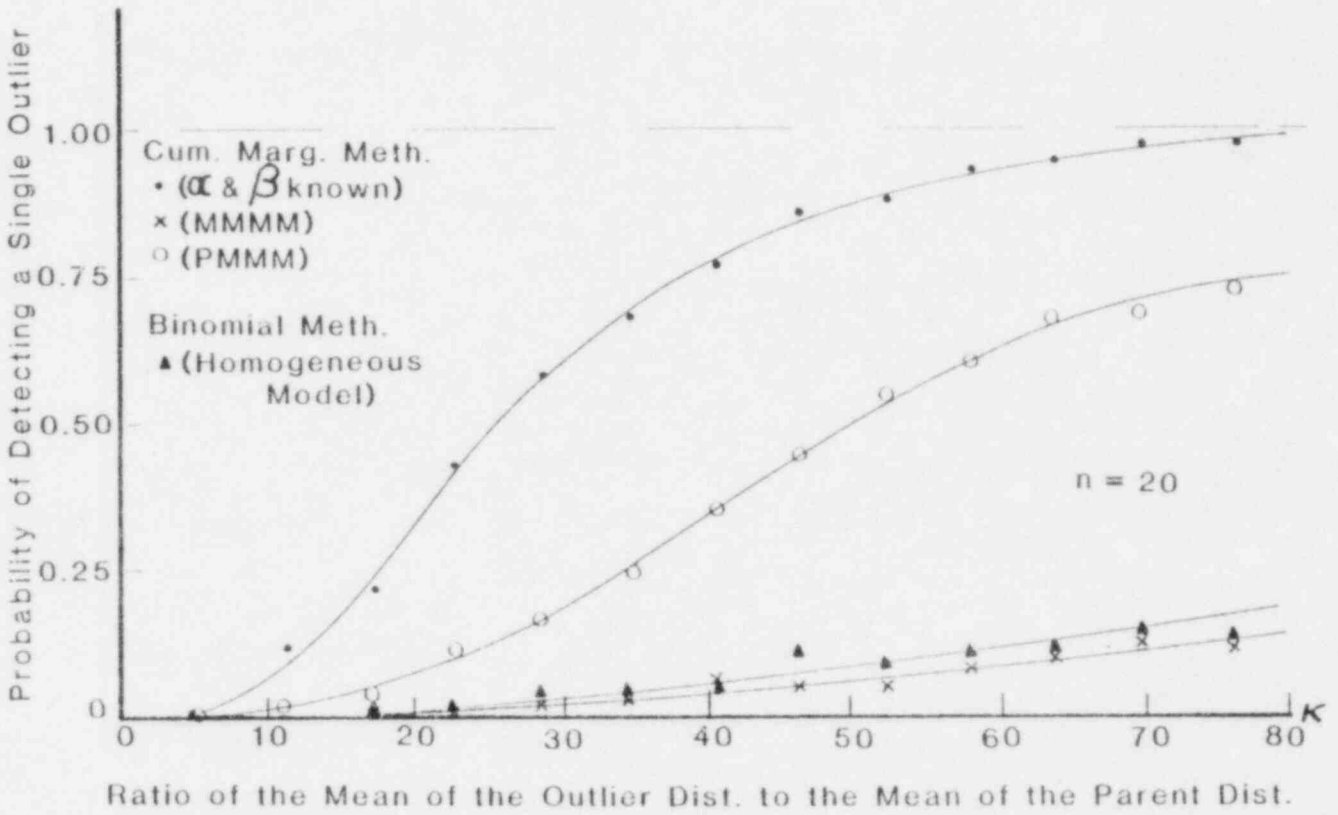


Figure 2.7: Power of single outlier detection tests with alpha = 1.25 and tau = 125,000.

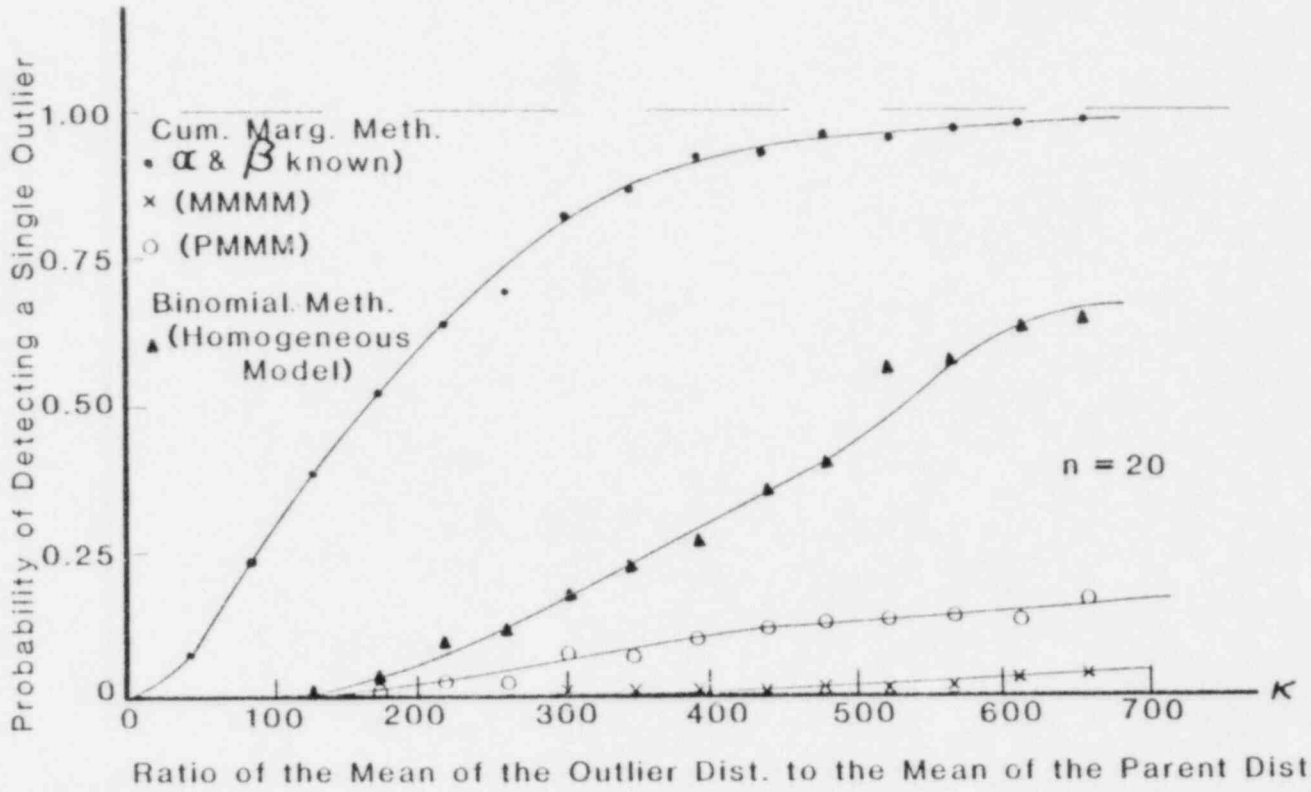


Figure 2.8: Power of single outlier detection tests with alpha = 1.25 and tau = 1,250,000.

2.4 OUTLIERS IN TIME-TO-FAILURE DATA

Often a basic premise of failure rate analyses is that the time-to-failure data will be distributed according to an exponential distribution. Since an exponenetial distribution is a special case of the gamma distribution, the methods for detecting outliers in an exponential distribution are similar to those used for the detection of outliers from a gamma distribution.

For outlier detection, there is one major difference between time-to-failure data and failure rate data (which are often assumed to be distributed according to a gamma distribution). Failure rate data for nuclear power plant components are usually assumed to contain no lower outliers because the inherent failure rates are so low and the test times are relatively short. That is, typical failure data reported in the literature consist of only a few observed failures F (often zero failures are reported) in the available test time T. Thus, it is impossible to detect lower outliers which might be present in such data, and so the assumption is made that no lower outliers are present. Time-to-failure data, on the other hand, may contain data in which a component failed either abnormally early or abnormally late, and thus lower, as well as upper, outliers are to be expected.

2.4.1 Fisher's Method of Outlier Detection

The foundation of Fisher's method for detecting outliers was discussed in Section 2.2.1. However, in that section it was applied to upper outliers only. For the case of lower outliers, the Fisher test statistic is defined by

$$TF = \sum_{i=1}^{k} x_i \Big/ \sum_{i=1}^{n} x_i \qquad (2\text{-}44)$$

where $x_1 \leq x_2 \leq \ldots \leq x_n$ are the observed time-to-failures, and k is the number of outliers being tested for.

The critical values of the test statistic are found from the following equation:

$$P\{TF \leq tF\} < \binom{n}{k} P\left\{F_{2k\alpha,2(n-k)\alpha} < \frac{(n-k)tF}{k(1-tF)}\right\} , \qquad (2\text{-}45)$$

where n is the number of data points, k is the suspected number of outliers, and $F_{2k\alpha,2(n-k)\alpha}$ is the F distribution with $2k\alpha$ and $2(n-k)\alpha$ degrees of freedom. As before, P(TF<tF) is given the value 0.01 or 0.05 in order to obtain either the 99% or 95% confidence level, respectively.

2.4.2 Dixon's Method of Outlier Detection

Dixon [11] developed a method for detecting outliers in normal samples, and Likes [12] extended this method to exponential distributions. Dixon's method, like Fisher's, is a hypothesis test. The null hypothesis, HO, is that all data come from the same exponenetial distribution. The alternative hypothesis, H1, is that a certain number of data points $k < n$, come from a distribution different from the distribution describing the main body of the data.

The application of Dixon's method is similar to that of Fisher's. A test statistic TD is evaluated from the data, and then the probability of it being as large as it is is determined. The Dixon test statistic is

$$TD = (x_s - x_r)/(x_q - x_p) \tag{2-46}$$

where $x_1 \leq x_2 \leq \ldots \leq x_n$ are the failure data, $1 \leq p \leq r < s \leq q \leq n$, and $q-p > s-r$. For the upper outlier case, the Dixon test statistic becomes

$$TD = (x_n - x_{n-k})/ (x_n - x_{a+1}) \tag{2-47}$$

where k is the number of upper outliers being tested for, and a is the suspected number of lower outliers. For the lower outlier test, Dixon's test statistic becomes

$$TD = (x_{k+1} - x_1)/(x_{n-b} - x_1) \tag{2-48}$$

where k is the number of lower outliers being tested for and b is the number of suspected upper outliers. The use of a and b will be discussed below.

The equation used to determine whether or not TD is too large is [8]

$$1 - F(tD) = \frac{(n-p)!}{(n-q)!}(1 - tD)$$

$$\times \left\{ \sum_{i=1}^{q-s} \sum_{k=1}^{s-r} \frac{(-1)^{i+k}(q-r-i)![(n-s+k)tD+(n-q+i)(1-tD)]^{-1}}{(i-1)!(k-1)!(q-s-i)!(s-r-i)!(q-p-i)!(n-s+k)} \right. \tag{2-49}$$

$$\left. + \sum_{j=1}^{r-p} \sum_{k=1}^{s-r} \frac{(-1)^{q-s+j+k}(s-r+j-1)![(n-s+k)tD+(n-r+j)(1-tD)]^{-1}}{(j-1)!(k-1)!(r-p-j)!(s-r-k)!(q-r+j-1)!(n-s+k)} \right\},$$

where F(tD) is the probability that TD is less than tD, P(TD<tD). The quantity TD is substituted for tD on the right hand side of Eq. (2-49), and the value obtained for F(TD) is the percentage confidence which can be used in labeling the suspicious data as outliers. Thus the critical value of F(TD) is taken as 0.95 if a 95% confidence level is desired. Note that in the above result, if q=s, the first double sum is zero, while if r=p the second double sum vanishes.

2.4.3 Determination of the Number of Outliers

It is much more difficult in the exponential case than in the gamma distribution case to determine the number of outliers because both upper and lower outliers are of concern. However, due to the simplicity of the outlier tests themselves, this difficulty is not a major obstacle.

When the Fisher test statistic is used, the number of data points being tested as outliers is determined in a different manner than that used in the gamma distribution case. The largest gap between consecutive data points relative to the upper gap is located. Then, depending upon whether this gap occurs above or below the mean value of the data, either upper or lower outliers, respectively, are tested for.

A difference between the gamma and exponential cases also exists in the criteria which are used to determine when the testing is complete. The first criterion, which must be met before testing is complete, is that both ends of the data spectrum must have negative tests for outliers. Even though there might be a larger relative gap between data in the upper region of the data than in the lower region, and no upper outlier is found, a lower outlier could still be present. This is because different distributions are used for the upper and lower regions to determine the critical values of TF. The second criterion which must be satisfed, therefore, is that lower outliers do not swamp upper outliers, and vice versa. Say, for example, that the upper region of a set of data is tested for outliers, and that some are found. Then the lower region is tested, and outliers are also found. In this case, the upper region must be retested, because outliers there might have been swamped by the lower outliers during the first test.

When the Dixon test statistic is used, the number of data points to be tested as being outliers is also determined by using the method of finding the largest relative gap in the data. However, due to its greater flexibility, there are some extra steps in the test for outliers. The greater flexibility of Dixon's method comes from the fact that when the testing is being applied to find the upper outliers, it can be taken into account that there might also be lower outliers present and vice versa. Thus, the effects of masking and swamping by outliers from the opposite end of the data spectrum can be reduced.

The method used to take into account that outliers might be present on the opposite end of the data spectrum is as follows. Suppose that suspicious data on the upper end of the spectrum are being tested. The

test is first applied assuming that no lower outliers are present, i.e., the value of a in Eq. (2-47) is given the value of zero. If these data are found to be discordant, they are labeled as outliers. However, if they are not discordant, the testing on them continues. First, using the gap method, the number of possible outliers is determined and becomes the value of a. This value of a is inserted into Eq. (2-47) and the test repeated. If the upper data are not found to be discordant this time, then they are not outliers, and the testing on them is complete, for the time being at least. If the upper data are found to be discordant when potential lower outliers are present, then the potential lower outliers must be tested for discordancy. This is done by using Eq. (2-48) and assuming that there are no upper outliers, i.e., b is given the value of zero. If the lower data are found to be discordant by this method, then both the upper and the lower data are outliers. If the lower data are not found to be discordant, then it is tested again assuming that the upper data might be outliers, i.e., b is given the value of the number of suspected upper outliers. If by this method, the lower data are found to be non-discordant, then neither the upper nor the lower data are discordant, and the testing is complete. If, on the other hand, the lower data are found to be non-discordant by this method, then it is not known whether the upper and the lower data are discordant or both are non-discordant. Thus, to be conservative, they must be assumed to be non-discordant and not labeled as outliers. The testing is then complete.

When Dixon's method is used, masking from the opposite end of the data is taken into account, and only one criterion need be met before the testing is complete. It is that both ends of the data must have undergone testing for outliers with the results being negative.

2.4.4 Properties and Comparison of the Outlier Tests

The power of the two tests for both upper and lower outliers show very little variation as the value of the failure rate varies. In fact, as the failure rate goes from 0.001 per hour to 10^{-7} per hour, the variation in the Fisher test is negligible (see Figs. (2.9) and (2.10)).

When upper outliers are being tested for, there is very little variation in the power of the tests as the number of data points in the data set varies. From Fig. (2.11) it is seen that as the number of data points varies from 10 to 30 there is insignificant variation in the power curves. However, when lower outliers are of interest, there is a distinct variation in the power of the tests as the number of data points changes. From Fig. (2.12) it is seen that as the size of the data set increases, the power of the tests decrease, i.e., saturation occurs at larger values of K.

From the studies performed during this study, it was found that Fisher's method of outlier detection is better than Dixon's method in every case. For detecting upper outliers, the power of Fisher's method is, in general, about 5% more powerful (see Fig. (2.13)). For detecting lower

outliers, Fisher's method is substantially more powerful than Dixon's method, although neither method is very powerful (see Fig. (2.14)).

As was the case for outlier detection in gamma distributions, the actual confidence level of detecting outliers was greater than 99% although the level of the tests used was at the 95% level. This is shown in Figs. (2.9) to (2.14) in which the power curves at K=1 or lnK=0 are below 0.01. However, since the actual confidence levels[3] of the various tests are not equal, it may not be completely fair to compare the power of the Fisher and Dixon tests in the above manner.

---

[3] The probability of accepting H0 (outlier and data are from the same distribution) is less than the critical level of significance chosen *a priori*. For example, a critical level of significance chosen as 0.05 (5% chance of incorrectly rejecting H0) led to power curves that indicated that there was a much smaller chance than 5% of incorrectly rejecting H0.

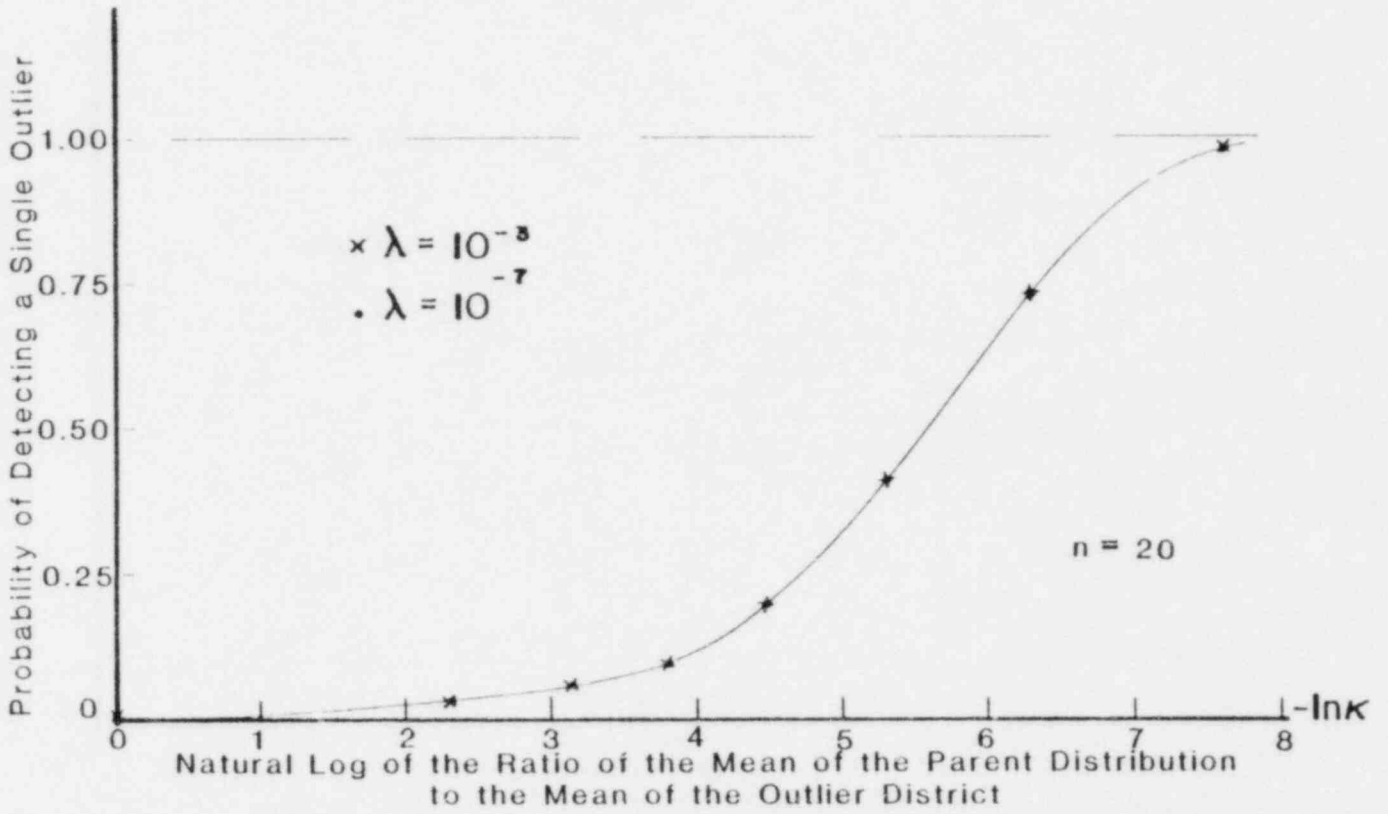


Figure 2.9: Lack of variation in the power of Fisher's method for a single lower outlier as the failure rate changes.

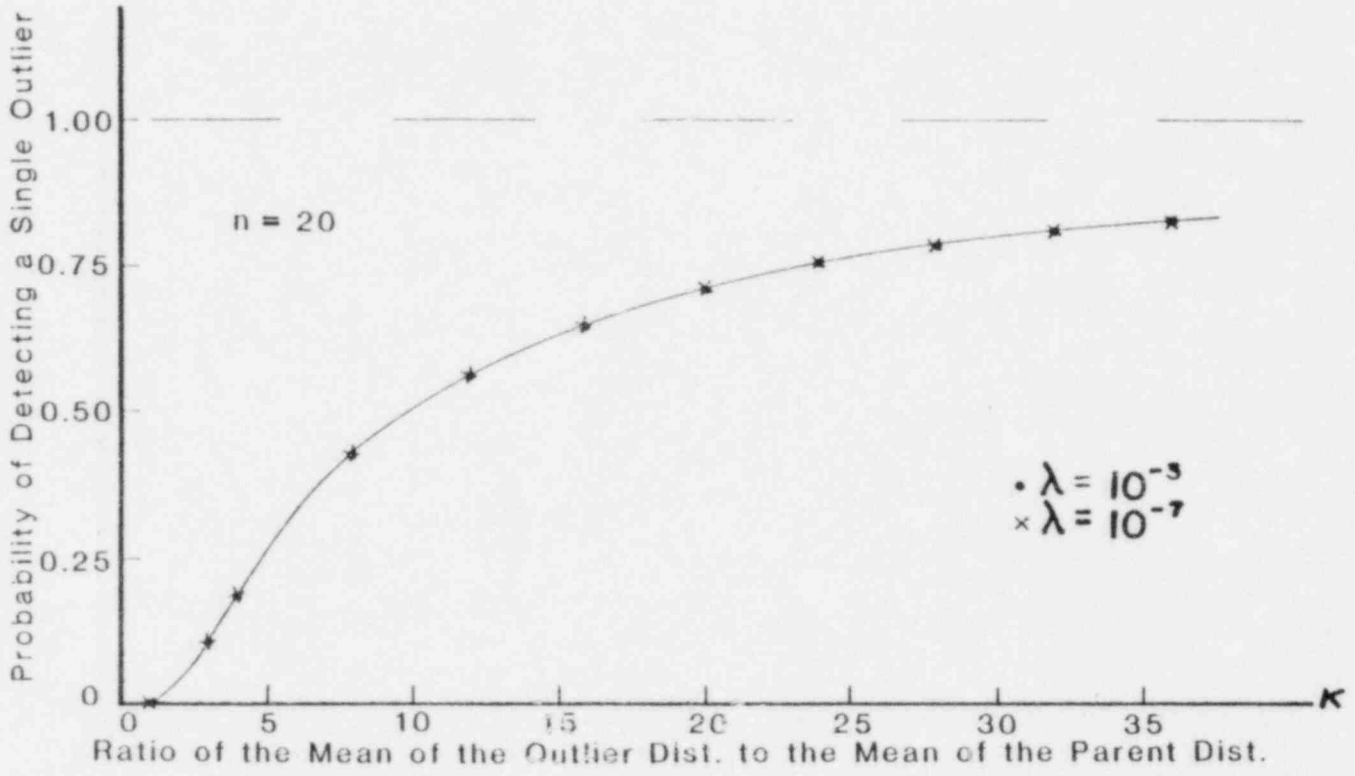


Figure 2.10: Lack of variation in the power of Fisher's method for detecting a single upper outlier as the failure rate changes.

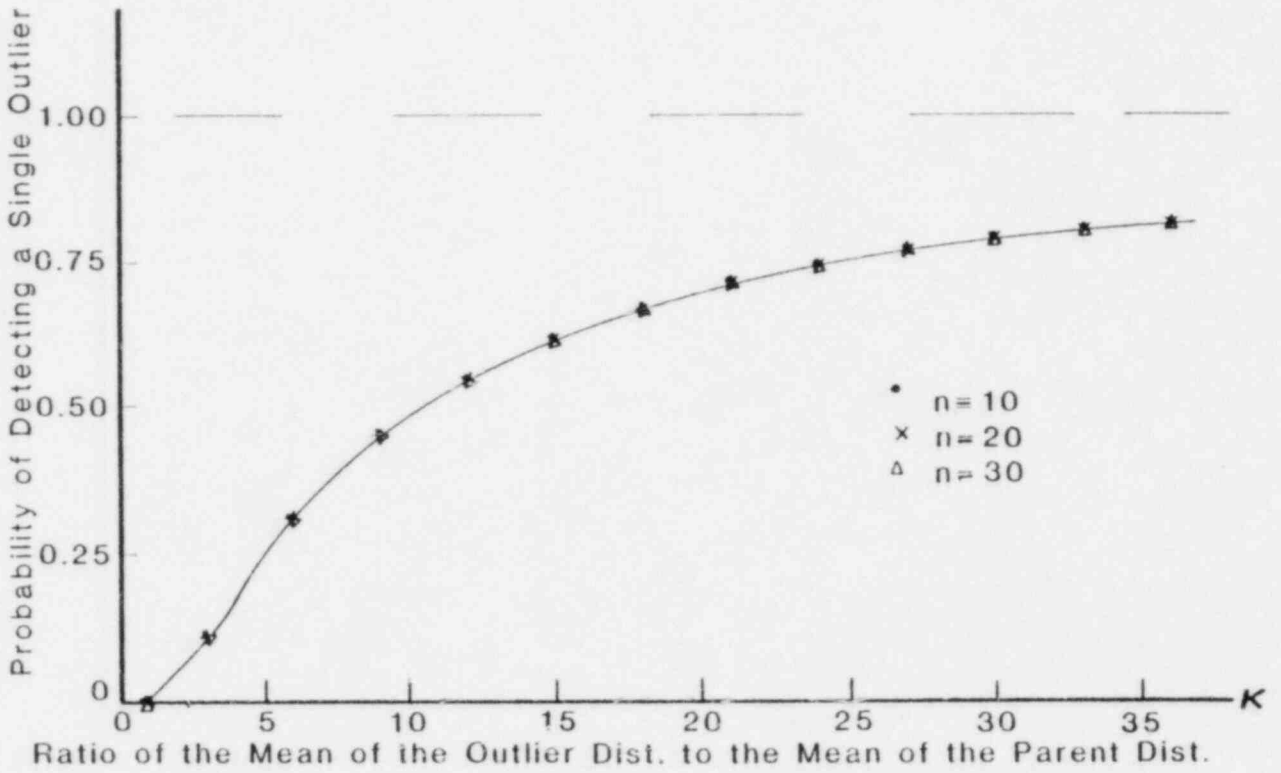


Figure 2.11: Lack of variation with Fisher's method for a single upper outlier as the sample size n varies.

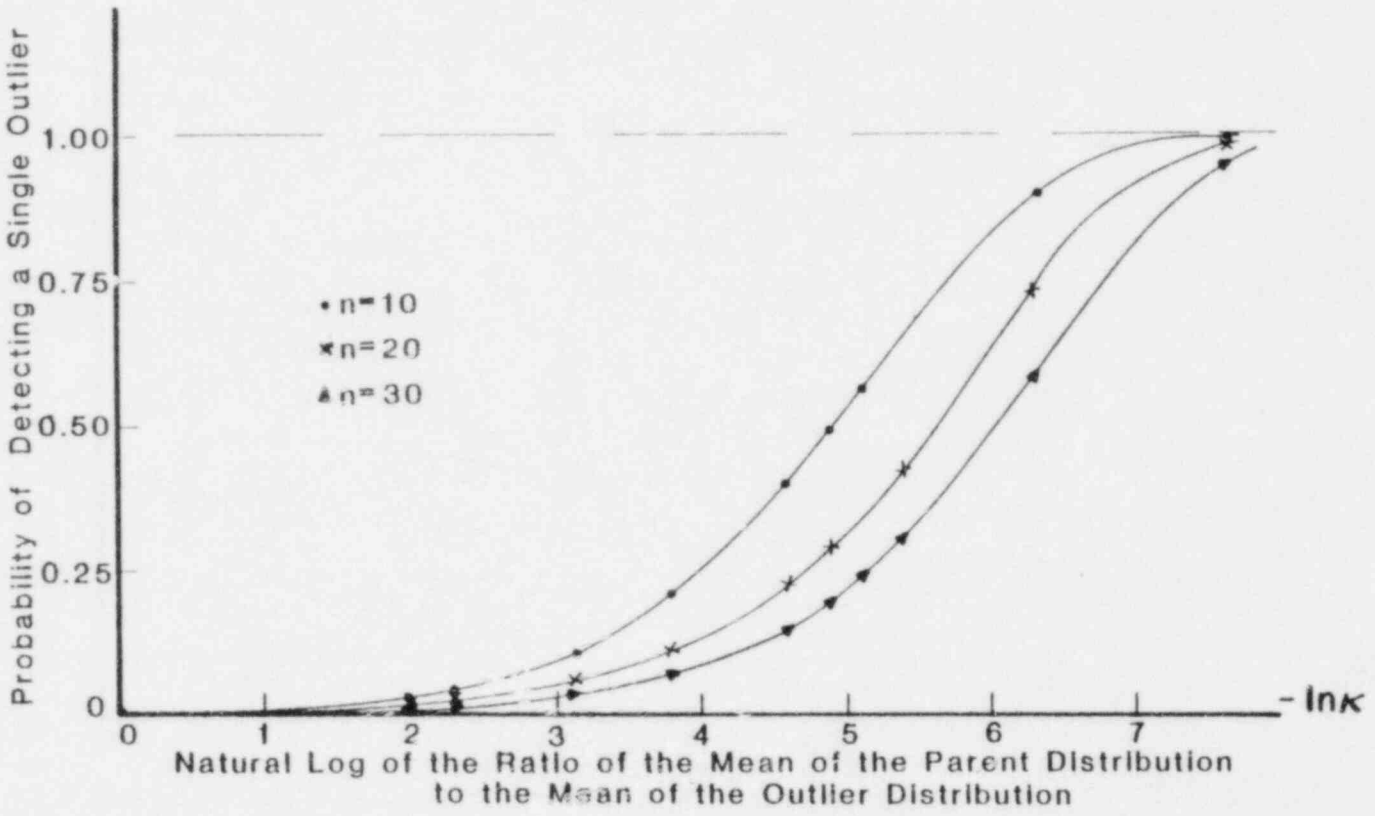


Figure 2.12: Variation with sample size of the power of Fisher's test for a single lower outlier (failure rate = 0.000001 per hr)

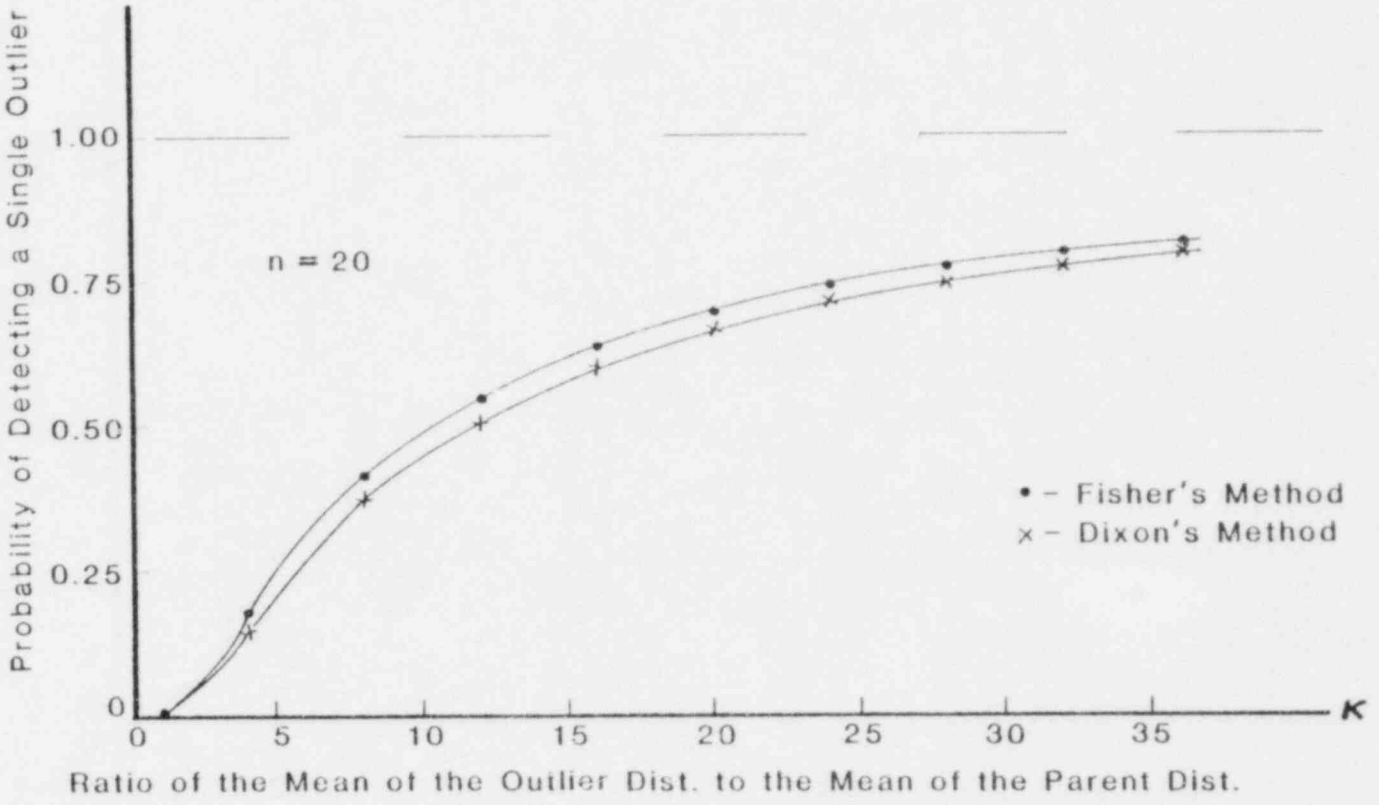


Figure 2.13: Comparison of Fisher's and Dixon's methods for detecting a single upper outlier (failure rate = 0.000001 per hr).

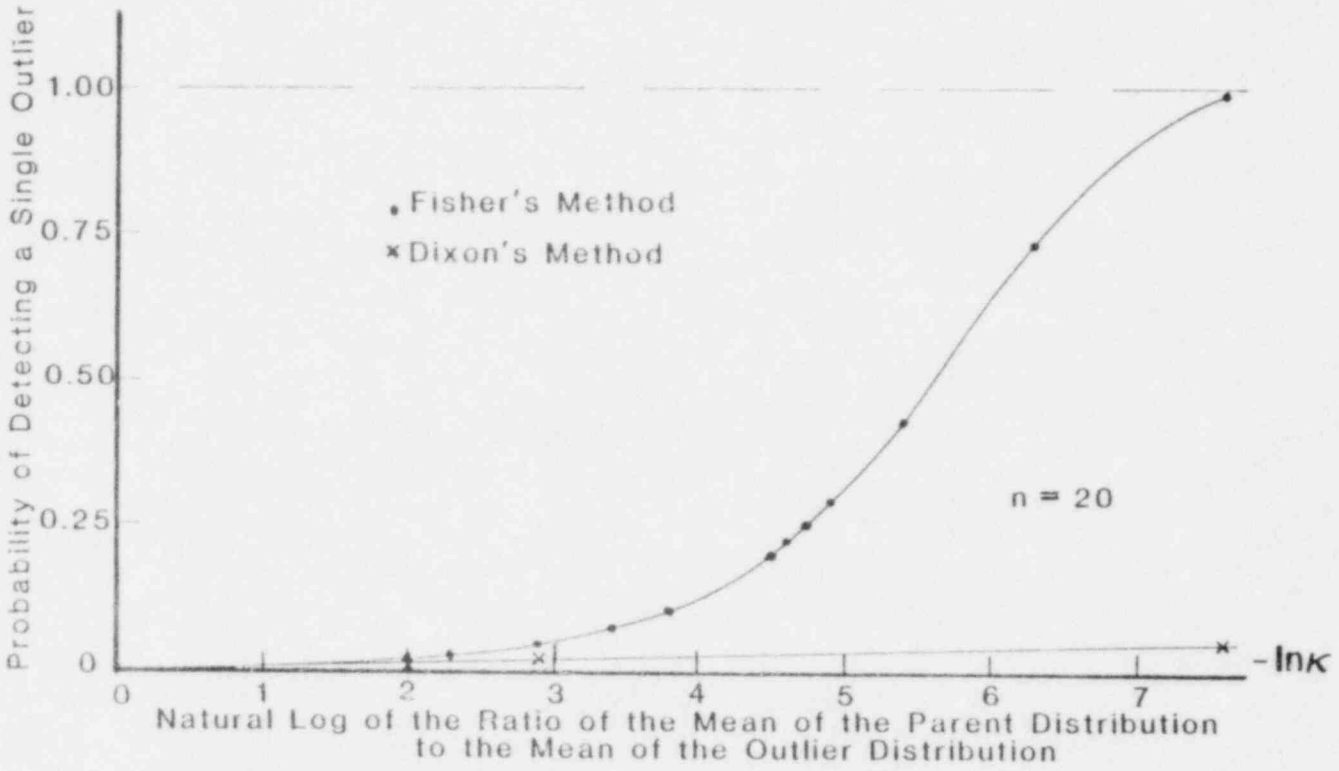


Figure 2.14: Comparison of Fisher's and Dixon's methods for a single lower outlier (failure rate = 0.000001 per hr).

2.5 SUMMARY OF CONCLUSIONS

For the detection of a single outlier in a gamma distribution, the least desirable method is the integration method for two reasons. First, the power of the integration method compared to the other available methods is low. Second, the computations which are necessary to apply the integration method are extremely cumbersome. Fisher'S method of outlier detection is much better than the integration method. However, when the value of the gamma alpha parameter is unknown (a situation often encountered in the analysis of failure data), the power of the Fisher method is substantially reduced. Surprisingly, the best method for detecting outliers in a gamma distribution is the conversion-to-normal method. In every case investigated, the normal conversion method was the most powerful method for detecting a single outlier. The normal conversion method is also attractive because of its simplicity, requiring only the use of a hand calculator and a table of critical values of the Grubbs test statistic.

For the detection of a single outlier in (F,T) data, the methods in which the gamma parameters are estimated using the suspected outlier fail completely. When there are only 20 data points per data set, an outlier will distort the gamma parameter estimates to such an extent that the outlier will not be detected. The cumulative marginal method, with the gamma parameters estimated by the MMMM is a poor method, and caution should be exercised in its use. This caution is necessary because for large failure rates, its actual confidence level is less than its theoretical confidence level, while for smaller failure rates, its power is extremely low. For small failure rates, i.e., $\bar{\lambda} < 0.00005$ per hour, with T = 10,000 hr, and n=20, the most powerful practical method of single outlier detection is the binomial method used in conjuction with the homogeneous failure model. For larger failure rates, i.e., $\bar{\lambda} > 0.00005$ per hour, the most powerful practical method is the cumulative marginal method with the gamma parameters estimated by the PMMM without using the suspected outlier. This method is also the best method for detecting outliers when $\bar{\lambda} < 0.00005$ per hour and all the test times are not equal.

For the detection of a single outlier in the exponential distribution, Fisher'S method is the best. It is approximately 5% more powerful than Dixon's method for detecting an upper outlier, and much more powerful for detecting a single lower outlier.

One property of all the recommended methods of outlier detection in the above cases is that when the tests are peformed for a theoretical confidence level of 95%, the actual confidence level is well above 99%. Thus the theoretical confidence level is seen to be conservative.

Chapter 3

PROPERTIES OF PRIOR ESTIMATION TECHNIQUES IN BAYESIAN ANALYSES

3.1 INTRODUCTION

To use the compound failure model in the analysis of failure rate data, the prior distribution describing the variation of the failure rates for a given set of components must be completely known, i.e., both the function family and its parameter values must be specified. Several procedures can be used to determine the prior distribution. One of the most widely used techniques is to assume the prior distribution belongs to a particular family of distributions and then to estimate appropriate values for the distribution parameters. In most applications of the compound model for analyzing failure rate data, a gamma distribution is chosen to represent the prior distribution, mainly because of the analytical simplicity of the subsequent analysis.

After the functional form of the prior distribution has been selected, values for the prior parameters must be determined. Although many techniques have been used for parameter estimation in Bayesian analyses, the methods investigated in this study are based on analyses of observed failure data for the components under consideration. In particular, the prior parameters are estimated by the following three methods [2]: (1) matching data moments to those of the prior distribution (PMMM), (2) matching data moments to those of the marginal distribution (MMMM), and (3) the maximum likelihood method applied to the marginal distribution (MMLM).

Once the prior distribution is determined, the compound or Bayesian failure model can be used to make predictions about failures of the components under investigation. However, the above procedure for determining the prior distribution for a given group of components involves two major assumptions which may affect the subsequent analyses. First, the effect of choosing the conjugate family to represent the prior distribution is generally unknown, and secondly the properties of the estimators for the prior parameters are generally unknown. The principal emphasis of this phase of the project was to investigate both of these questions for the failure rate problem.

Besides the gamma distribution, several other families of distributions are reasonable alternatives for representing the prior distribution. In this study, four families of distribution, namely, gamma, Weibull, lognormal and logbeta, were used as possible prior distributions. For each distribution family, the prior parameters are estimated from

given component failure data by three methods (PMMM, MMMM and MMLM). From this study, it was then possible to determine the effect of the prior family selection and the parameter estimation techniques on the compound failure model.

3.2 PROPERTIES OF THE PARAMETER ESTIMATORS

Because the gamma-Poisson failure model is widely used to describe the failure rate problem, the first part of this study was directed at determining the properties of various parameter estimation schemes. To investigate the properties of the parameter estimators, simulated failure data were generated from a marginal distribution whose functional form and parameter values are known. These simulated failure data were then grouped into samples of various sizes, and estimates of the prior parameters were obtained by each of the three estimation methods for each sample. With these estimates, the distributions of the estimators were obtained empirically. From these empirical distributions, the properties of the estimators for each of the three estimation methods were determined.

The generation of simulated failure data from a known marginal distribution and the subsequent data analysis can be performed in many ways by using different functions for the marginal and prior distributions. In this section, the results obtained for simulated failure data from a known conjugate gamma-Poisson model are described. Further, a gamma-Poisson model was then fitted to these simulated data by the three parameter estimation techniques (PMMM, MMMM and MMLM). Results for data from non-conjugate models and/or subsequent analysis by non-conjugate models are presented in later sections of this chapter.

The technique used to generate failure data which are distributed according to a given marginal distribution is described in Section 1.3 of this report. In the generation of simulated failure data from a known marginal distribution, the component test time T was always held constant at a specified value. The simulated data thus consisted of the number of failures of a component which has been operated for a time T and which belongs to a class described by the marginal distribution.

3.2.1 Gamma-Poisson Analysis of Gamma-Poisson Data: Set 1

Since this study was concerned primarily with components characterized by low failure rates, the gamma-Poisson marginal distribution used to generate the first set of simulated failure data (Set 1) was that of Eq. (1.3) with parameter values selected as alpha=1.2 and tau=100,000. With these parameter values, the mean, standard deviation and 95-th percentile of the assumed gamma prior distribution are 0.12(-4),[4] 0.109(-4)

---

[4] read 0.12(-4) as 0.12 times 10 to the -4 power, i.e., 0.000012.

and 0.337(-4) per hour, respectively. The component operation time T was taken as 50,000 hours (almost 6 years). The simulated failure data generated in this manner are characteristic of components whose failure rate distribution is a gamma distribution and which can be expected to fail , on the average, approximately 1 time in 9 years of operation.

With this gamma-Poisson marginal distribution, simulated failure data were generated by the method discussed in Section 1.3. A total of 10,000 values of component failures F_i were generated, and these failure data were then grouped into sets with four sample sizes, namely, 5, 10, 20 and 50 components. In all, there were 2,000, 1,000, 500 and 200 data sets of sample sizes 5, 10, 20 and 50, respectively. For these simulated failure data, estimates of the parameters alpha and tau of the gamma prior distribution were calculated by each of the three estimation methods summarized in Section 1.2.1 and discussed in detail in Ref. [1].

3.2.1.1 Ability of Estimation Methods to Yield Estimators

The three prior parameter estimation techniques, (MMMM, PMMM and MMLM) each failed for some of the failure data sets. The MMMM would occasionally yield unrealistic negative estimators for the prior paramters, while the MMLM would sometimes fail to find estimators, and at other times produce values that were more than 100 times the true values. The PMMM failed only when all failures in a sample were zero -- a not infrequent possibility for the particular gamma-Poisson model used to generate the data for this case. In fact, when the sample consisted of only zero failures, none of the methods yielded parameter estimates. In Table 3.1 a summary is presented which shows the frequency with which the three estimation methods yielded valid parameter estimators. Among the three estimation techniques, the PMMM always had the highest percentage success, ranging from 91% to 100%. For sample size greater than 20, this simple estimation method always worked. The two marginal-based methods had almost identical percentages of success ranging from 39% to 91% for samples of size 5 to 50. It is not completely clear why these marginal methods failed for certain samples (besides the case of all zero failures), although from Eq. (1-9) the MMMM fails whenever the term in square brackets is less than or equal to zero.

3.2.1.2 Distribution of Prior Parameter Estimators

The empirical frequency distributions of the estimators were constructed from the successful results for all three estimation methods. Typical results are shown in Fig. 3.1 and Fig. 3.2. The distribution of estimators for data set 1 is seen to be quite "spikey" although it is concentrated around the true parameter values. The reason for the spiked nature of these distributions is that for the particular gamma-Poisson distribution used to generate the failure data, only a few different samples (i.e., with different values of F) are likely and hence only a few different values for the prior parameter estimates are found.

TABLE 3.1

Ability of Estimators to Yield Valid Estimates: Data Set 1

Method	Sample Size	No. of Successful Solutions(a)	% of Success	Unreasonable Large Solutions(b)
PMMM	5	1814 (2000)	90.7	-
PMMM	10	991 (1000)	99.1	-
PMMM	20	500 (500)	100.0	-
PMMM	50	200 (200)	100.0	-
MMMM	5	786 (2000)	39.3	61
MMMM	10	597 (1000)	59.7	75
MMMM	20	396 (500)	79.2	-
MMMM	50	183 (200)	91.5	-
MMLM	5	777 (2000)	38.8	-
MMLM	10	547 (1000)	54.7	-
MMLM	20	363 (500)	72.6	-
MMLM	50	180 (200)	90.0	-

(a) Numbers in the parentheses are total number of samples.
(b) Estimates were greater than 100 times true values.

More interesting is the appearance of inordinately large estimates produced by the marginal-based methods. For the MMMM estimators, these large estimators arise from samples for which the term in square brackets in Eq. (1-9) is very small. Upon examination of the results for samples of size 5, these inordinately large estimates were found to be generated from only two types of data sets -- (2,2,1,0,0) and (3,1,1,1,0) -- but with the failures occurring in different orders. For samples of size 10, the marginal matching moments method yields 75 extrodinarily large estimators all of which came from the same data set (2,1,1,1,0,0,0,0,0,0). The MMMM alpha estimators for these three data sets are 3.87(15), 2.23(15) and 1.55(16) respectively, while the tau estimates are 1.93(20), 9.28(19) and 1.55(21), respectively. These values are obviously too large to be reasonable estimates of the prior parameters alpha and tau whose actual values are 1.2 and 100,000. However, these estimates of alpha and tau still give estimates of the mean of the gamma prior distribution as 0.20(-4), 0.24(-4) and 0.10(-4) each of which is very close to the actual value, 0.12(-4).

These estimates of alpha and tau are so large that, if included in the subsequent analyses, any values of statistics calculated from these

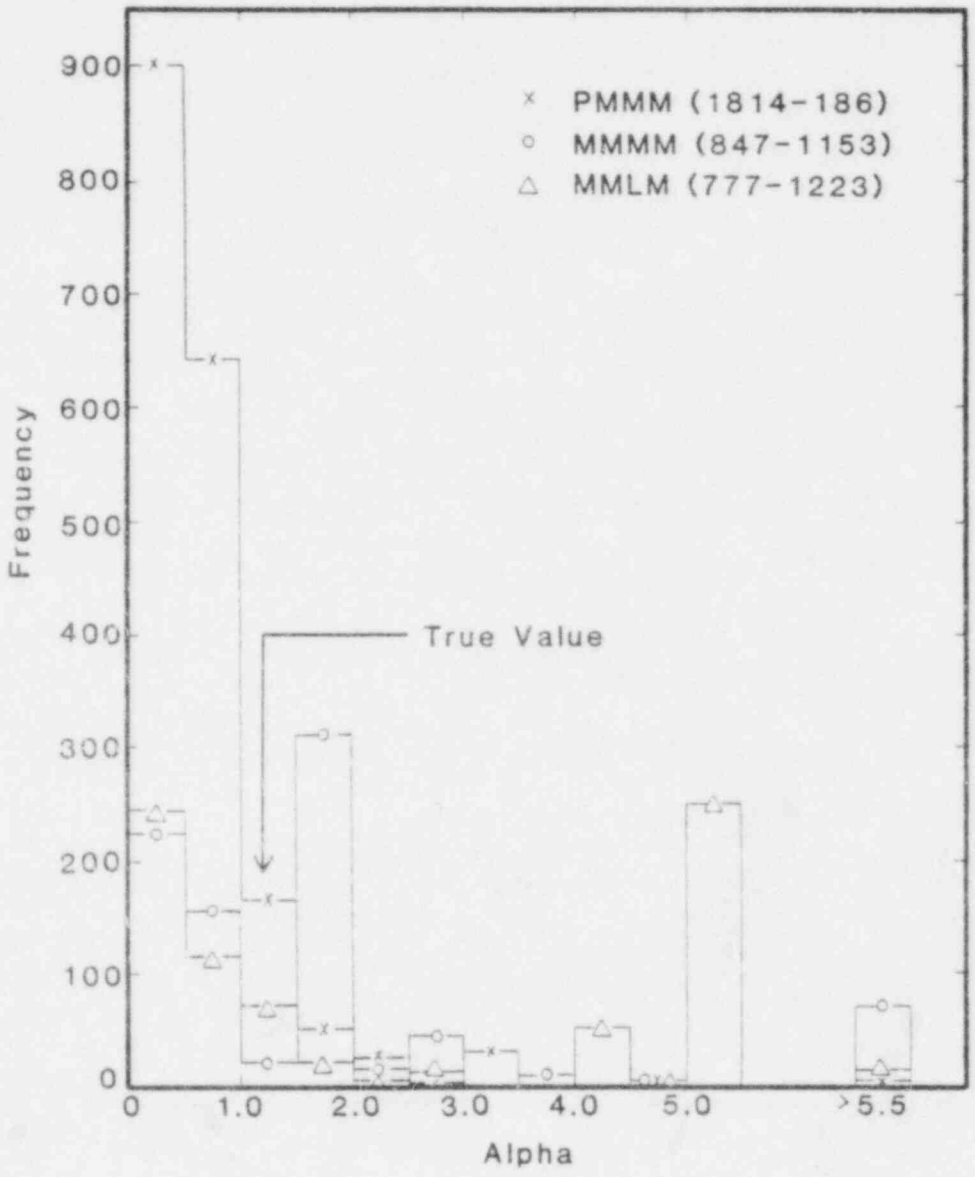


Figure 3.1: Distribution of the alpha estimator for sample of size 5 for data set 1. The true value of alpha is 1.2.

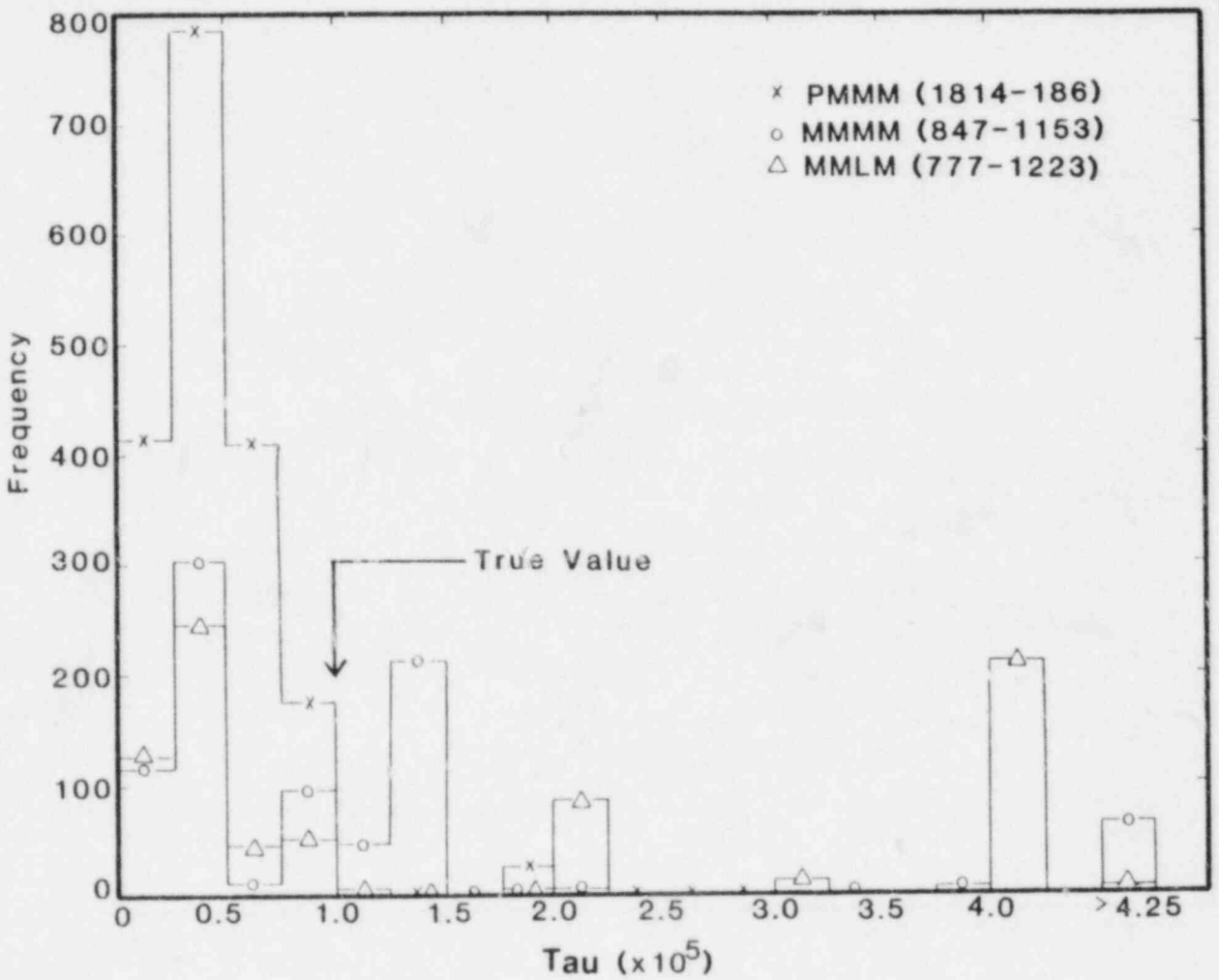


Figure 3.2: Distribution of the tau estimator for samples of size 5 for data set 1. The true value of tau is 100,000.

estimates would be unduly influenced by them. For example, the mean of the alpha estimates for samples of size 5 reduces from 9.0(14) to 1.4 when the 61 very large estimates are removed from the analysis. In this study, estimates which are 100 times greater than the true values are considered invalid solutions and are not included in the computation of any statistics except when indicated otherwise. Since 100 times the true parameter value was used as the boundary of the numerical solution space for the MMLM, all successful MMLM solutions are within the specified boundary and inordinately large parameter estimates were thus classified as "no solution".

The distributions of estimators for the PMMM also show a slight additional peak at the high parameter end of the distributions, but only for sample size 5. Most estimates center around the lower side of the true parameter values for all sample sizes. In other words, the mean of the distributions of the alpha and tau estimators are smaller than the true values.

3.2.1.3 Mean and Standard Deviation of the Estimators

The means of the parameter estimates are presented in Tables 3.2 and 3.3. The means of the PMMM parameter estimators are always smaller than the true values for this set of failure data. On the other hand, those from the two marginal-based estimation methods are almost always greater than the true values.

While the mean is an important parameter of a distribution because it indicates where the central part of a distribution is located, it alone usually does not give enough information to provide an adequate description of the distribution. Another measure is needed to indicate how spread out or dispersed the distribution is. One commonly used measure of dispersion is the standard deviation. The standard deviations of the distributions of alpha and tau estimators are also shown in Tables 3.2 and 3.3. For a given sample size, the minimum standard deviation was always obtained for the PMMM estimators.

3.2.1.4 Bias of Estimators

One criterion for a good estimator is how close the average value of the estimator comes to the parameter being estimated. It is desirable to have the mean of the estimator distribution equal the true parameter value. If this is the case, the estimator is called an unbiased estimator. When the estimator mean does not equal the true value, the estimator is said to contain a bias, and the estimator is said to be a biased estimator. The magnitude of bias is defined to be the difference between the true value and the mean value of the estimate.

The variation of the bias of alpha and tau estimates with sample size is shown in Fig. 3.3. All estimates of alpha and tau from the three

TABLE 3.2

Mean, standard deviation and square root of the mean square error of the alpha estimator for data set 1

Method	Sample Size	Mean	Standard Deviation	Sq. Root of MSE
PMMM	5	.646	.616	.828
PMMM	10	.509	.311	.758
PMMM	20	.458	.186	.765
PMMM	50	.429	.122	.781
MMMM	5	1.40	1.42	1.43
MMMM	10	2.19	3.57	3.70
MMMM	20	3.06	6.02	6.29
MMMM	50	2.24	3.36	3.51
MMLM	5	2.55	2.66	2.98
MMLM	10	1.62	2.14	2.18
MMLM	20	2.19	4.42	4.53
MMLM	50	2.16	3.89	4.00

estimation methods are found to be biased for all sample sizes used in this study. The PMMM estimates of alpha and tau for sample size 5 have the least bias of all estimates. However, all estimates from the PMMM have negative bias and the magnitude of the bias increases slightly as the sample size increases. In contrast, the marginal-based estimators almost always have positive bias. However, the bias shows no specific pattern of variation with sample size. This lack of a trend in the bias with sample size arises from the peculiarities of the data set used. For data set 1 the components have a very low failure rate, and for the test time chosen (T=50,000) most simulated failures were zero or unity. This limitation on the number of different failures results in only a few different parameter estimates being obtained. For such spikey estimator distributions (see Figures 3.1 and 3.2), no trends with sample size are apparent.

3.2.1.5 Mean Squared Error of Estimators

Between two unbiased estimators, the estimator with a smaller standard deviation (or smaller variance) is the better estimator. However for biased estimators, the mean squared error is normally used in such a comparison rather than the standard deviation . The magnitudes of both

TABLE 3.3

Mean, standard deviation and square root of the mean square error of the tau estimator for data set 1

Method	Sample Size	Mean	Standard Deviation	Sq. Root of MSE
PMMM	5	0.507(+5)	0.300(+5)	0.576(+5)
PMMM	10	0.443(+5)	0.202(+5)	0.592(+5)
PMMM	20	0.396(+5)	0.144(+5)	0.621(+5)
PMMM	50	0.360(+5)	0.090(+5)	0.646(+5)
MMMM	5	0.847(+5)	0.561(+5)	0.581(+5)
MMMM	10	1.48 (+5)	1.80 (+5)	1.87 (+5)
MMMM	20	2.53 (+5)	4.59 (+5)	4.84 (+5)
MMMM	50	1.93 (+5)	3.02 (+5)	3.15 (+5)
MMLM	5	1.68 (+5)	1.73 (+5)	1.86 (+5)
MMLM	10	1.28 (+5)	1.32 (+5)	1.35 (+5)
MMLM	20	1.82 (+5)	3.43 (+5)	3.52 (+5)
MMLM	50	1.87 (+5)	3.56 (+5)	3.66 (+5)

the bias and the variance of an estimate are included in the mean squared error (MSE) which is defined as

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\theta_i - \theta)^2 \qquad (3\text{-}1)$$

where θ is the parameter being estimated.

Very often the standard deviation is used in place of the variance because the standard deviation has the same units as the mean of the distribution. For the same reason, the square root of mean squared error is used instead of mean squared error itself.

Presented in Fig. 3.4 are the variation of square root of the mean square error of alpha and tau estimators as a function of the sample size. As with the bias, there appears to be a smooth variation of the MSE with sample size only for the PMMM estimators. The wild variation of the MSE for the marginal-based estimators is a peculiarity of the data set (see the discussion in Section 2.2.1.4 above). The MSE of the PMMM estimators changes very slightly as the sample size changes. Moreover, for all sample sizes, the PMMM estimators have the smallest MSE.

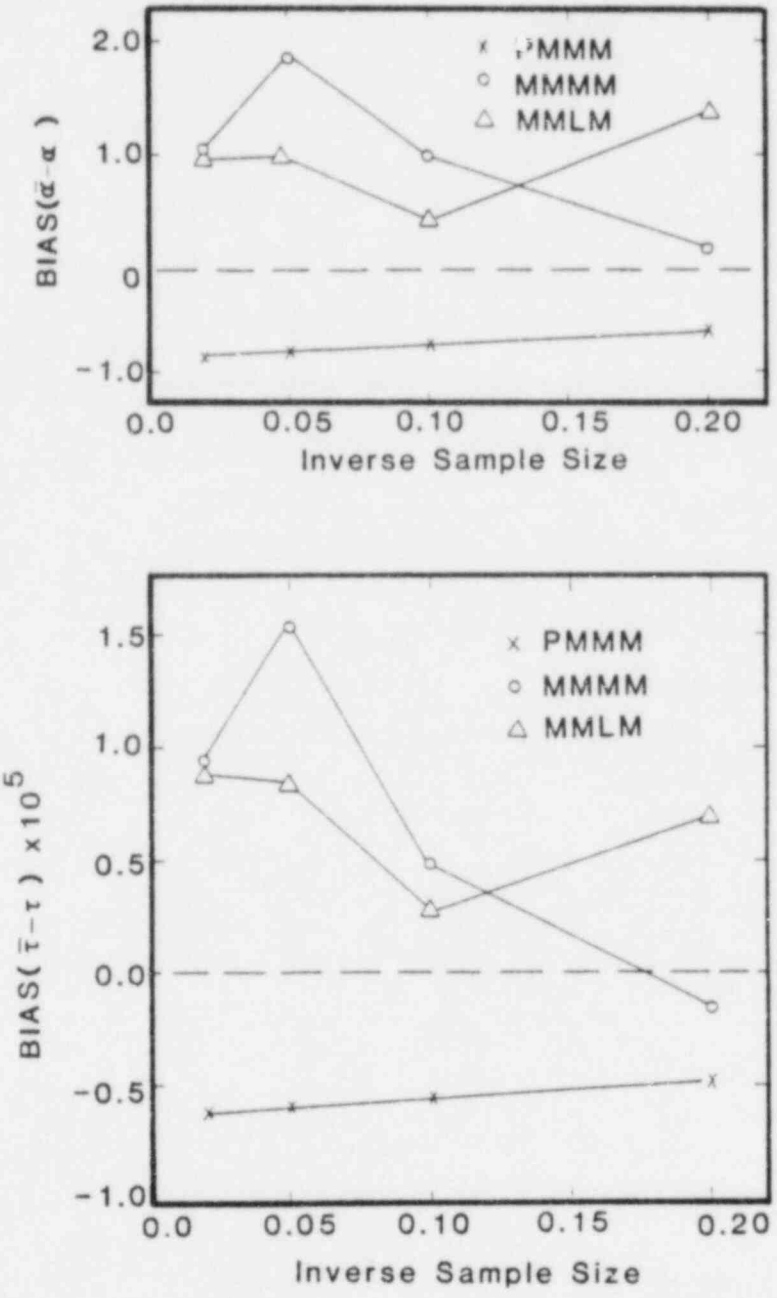
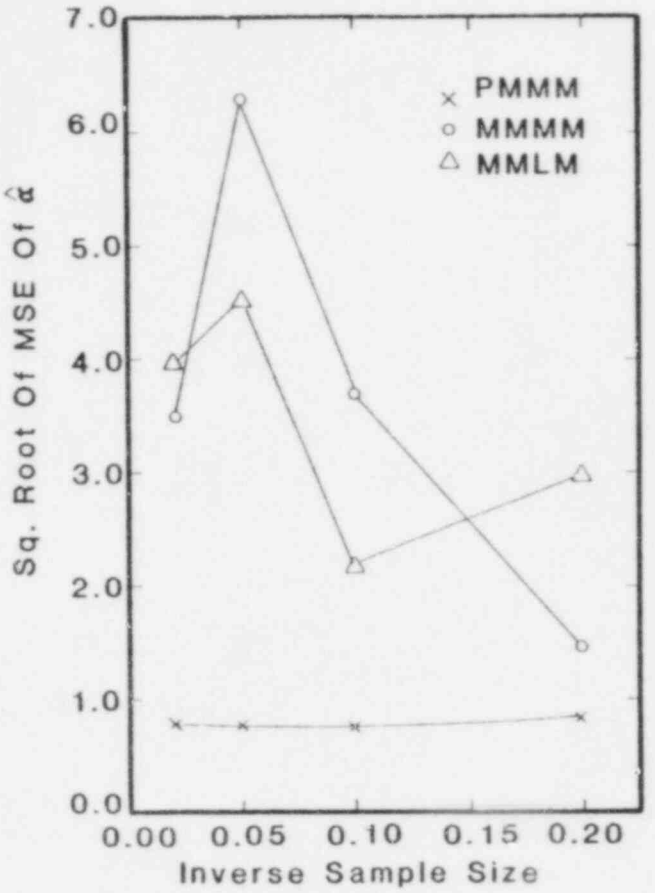


Figure 3.3: Bias of the alpha and tau estimators for data set 1



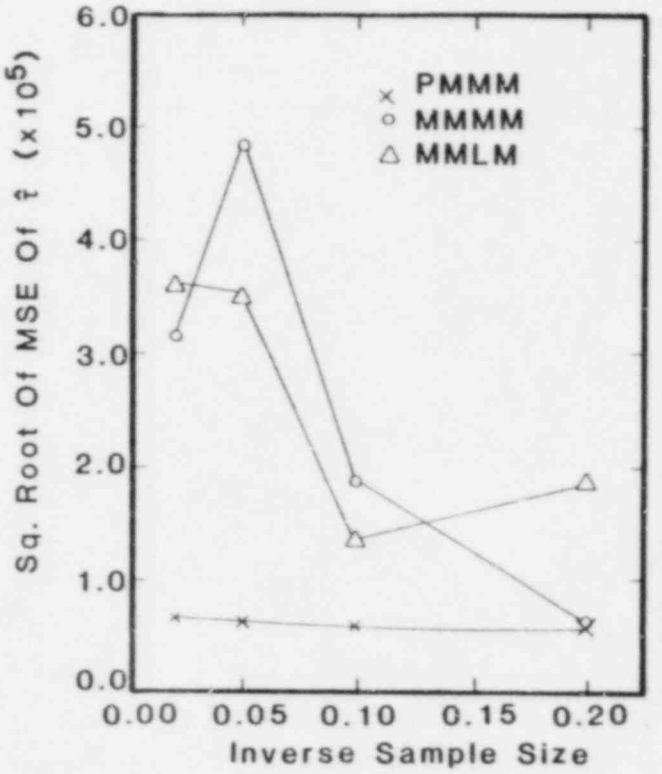


Figure 3.4: Square root of the MSE of the prior parameter estimators for data set 1

3.2.1.6 Median of Estimators

The median is another widely used measure of a distribution's central location and is less sensitive to the tail of the distribution than is the mean. The very large estimates obtained from the MMMM method have less effect on the value of the median than that of the mean. The variation of the median of the alpha and tau estimates with sample size are shown in Fig. 3.5. Medians of the PMMM estimators show a similar variation with sample size to that of the bias shown in Fig. 3.3. However, for all sample sizes, the medians of the marginal-based estimators come closer to the true parameter values than those from the PMMM.

3.2.1.7 Mean and Standard Deviation of the Estimated Gamma Prior Distribution

From each pair of alpha and tau estimates, the mean and standard deviation of the resulting prior gamma distribution were computed. If all three estimation methods yield valid parameter estimates, the means of the estimated gamma prior distribution should be the same. But in practice, not all data samples give valid parameter estimates by all three methods for the same failure data. Shown in Table 3.4 are the mean, standard deviation, and square root of the MSE for the mean of the gamma distribution obtained from the parameter estimators. The variation with sample size of the average of the means of the estimated gamma prior distributions are shown in Fig. 3.6. As would be expected, the average of the means approach the true value as the sample size increases for all three estimation methods. However, the PMMM estimates are the closest to the true value for all sample sizes.

Table 3.5 and Fig. 3.7 show the averages of the standard deviation of the estimated gamma distributions. The results from the MMMM and MMLM analyses approach the true value as the sample size increases, but the PMMM results diverge.

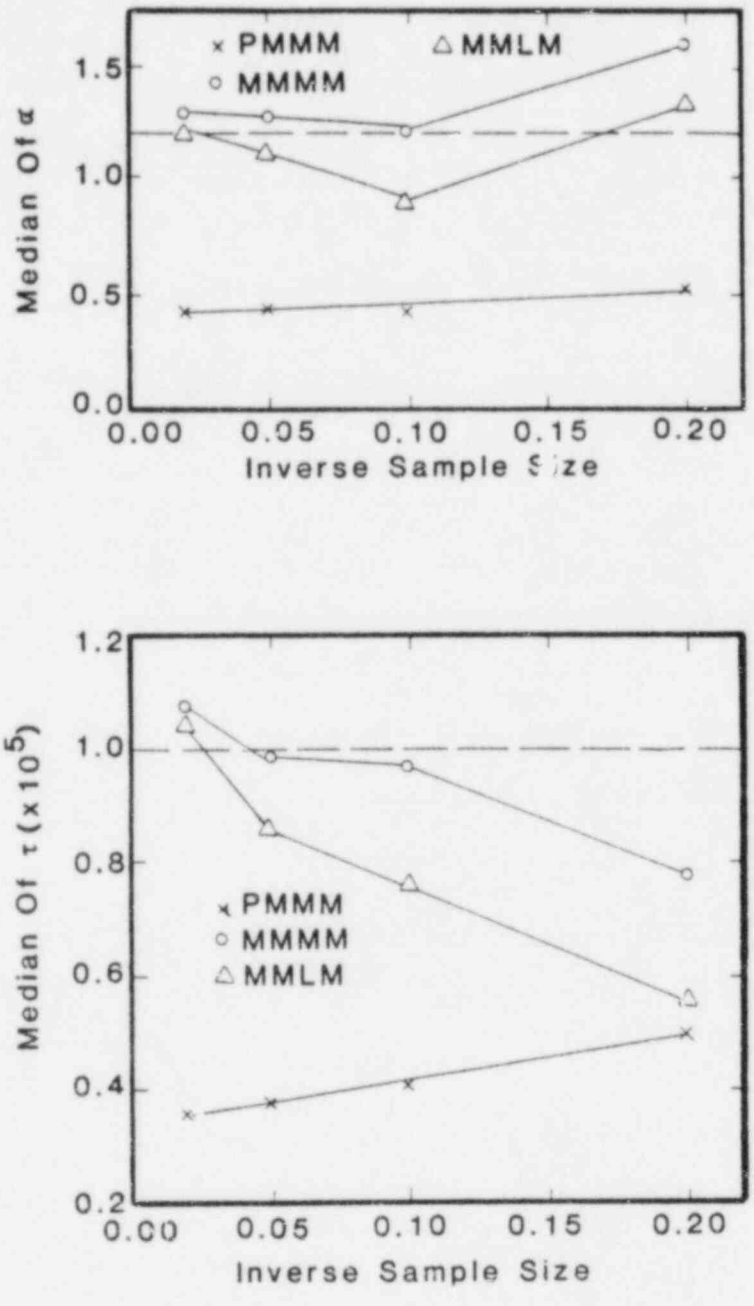


Figure 3.5: Medians of prior parameter estimators for data set 1

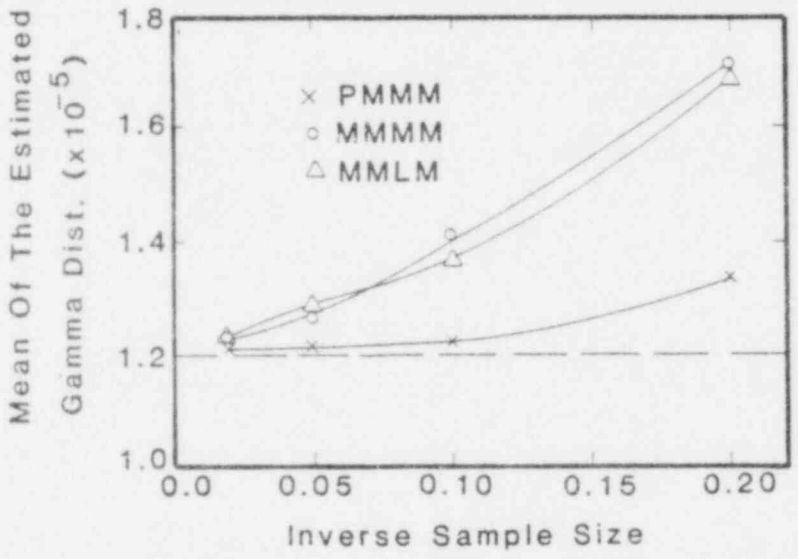


Figure 3.6: Mean of the estimated means of the prior distributions for data set 1.

TABLE 3.4

Mean, standard deviation and square root of the MSE of the estimated prior mean for data set 1

Method	Sample Size	Mean	Standard Deviation	Sq. Root of MSE
PMMM	5	1.34(-5)	0.818(-5)	0.829(-5)
PMMM	10	1.22(-5)	0.601(-5)	0.601(-5)
PMMM	20	1.21(-5)	0.423(-5)	0.422(-5)
PMMM	50	1.21(-5)	0.287(-5)	0.287(-5)
MMMM	5	1.71(-5)	0.813(-5)	0.960(-5)
MMMM	10	1.41(-5)	0.602(-5)	0.637(-5)
MMMM	20	1.27(-5)	0.422(-5)	0.428(-5)
MMMM	50	1.23(-5)	0.288(-5)	0.288(-5)
MMLM	5	1.69(-5)	0.785(-5)	0.925(-5)
MMLM	10	1.36(-5)	0.604(-5)	0.626(-5)
MMLM	20	1.28(-5)	0.431(-5)	0.438(-5)
MMLM	50	1.23(-5)	0.288(-5)	0.289(-5)

TABLE 3.5

Mean, standard deviation and square root of the MSE for estimated prior standard deviation for set 1

Method	Sample Size	Mean	Standard Deviation	Sq. Root of MSE
PMMM	5	1.77(-5)	0.915(-5)	1.14 (-5)
PMMM	10	1.77(-5)	0.734(-5)	0.998(-5)
PMMM	20	1.83(-5)	0.556(-5)	0.920(-5)
PMMM	50	1.87(-5)	0.373(-5)	0.860(-5)
MMMM	5	1.70(-5)	0.890(-5)	1.08 (-5)
MMMM	10	1.33(-5)	0.716(-5)	0.754(-5)
MMMM	20	1.14(-5)	0.594(-5)	0.594(-5)
MMMM	50	1.08(-5)	0.432(-5)	0.432(-5)
MMLM	5	1.74(-5)	1.26 (-5)	1.42 (-5)
MMLM	10	1.48(-5)	0.839(-5)	0.920(-5)
MMLM	20	1.25(-5)	0.616(-5)	0.635(-5)
MMLM	50	1.11(-5)	0.451(-5)	0.451(-5)

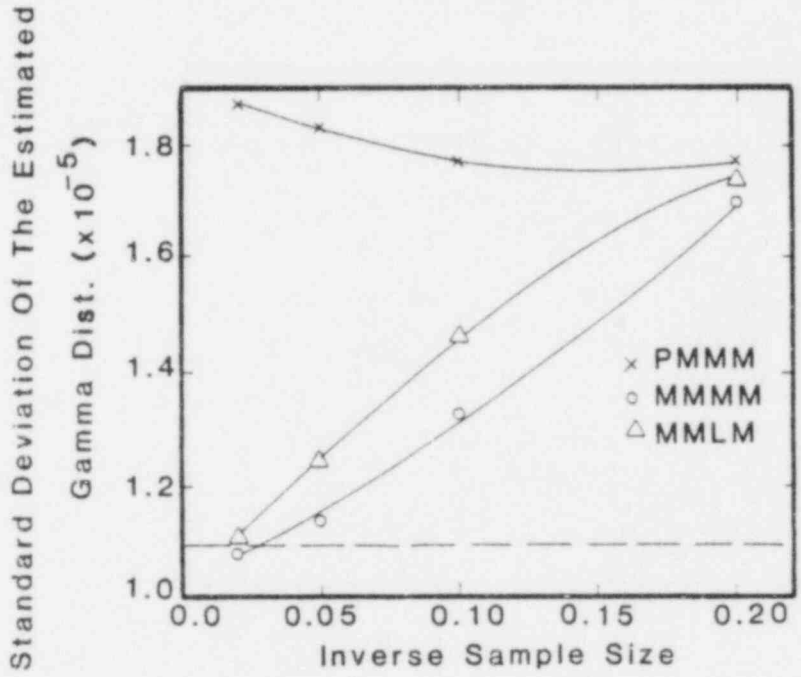


Figure 3.7: Mean of the estimated standard deviations of the prior distributions for data set 1.

3.2.1.8 Estimated Percentiles of the Gamma Prior Distribution

Of considerable interest in safety analysis is the failure rate distribution for high failure rates. One widely used measure of the high probability tail is the 95-th percentile, i.e., the failure rate above which there is only a 5 % chance that the true failure rate of a component lies. From each pair of alpha and tau estimates, the 95-th percentile of the estimated gamma prior distribution was computed.

In Fig. 3.8, the variation of the mean of these 95-th percentile estimates is shown as a function of sample size. The results of Fig. 3.8 show characteristics similar to the variation with sample size of the mean of the standard deviation of the estimated gamma distribution (see Fig. 3.7). The means of the MMMM and MMLM percentiles seem to be very similar, and both tend to approach the true value as the sample size increases. For sample size 50, the percentile estimates are almost exact. By contrast, the mean of the PMMM percentile estimator shows little variation with sample size -- a property earlier observed for the PMMM bias and MSE.

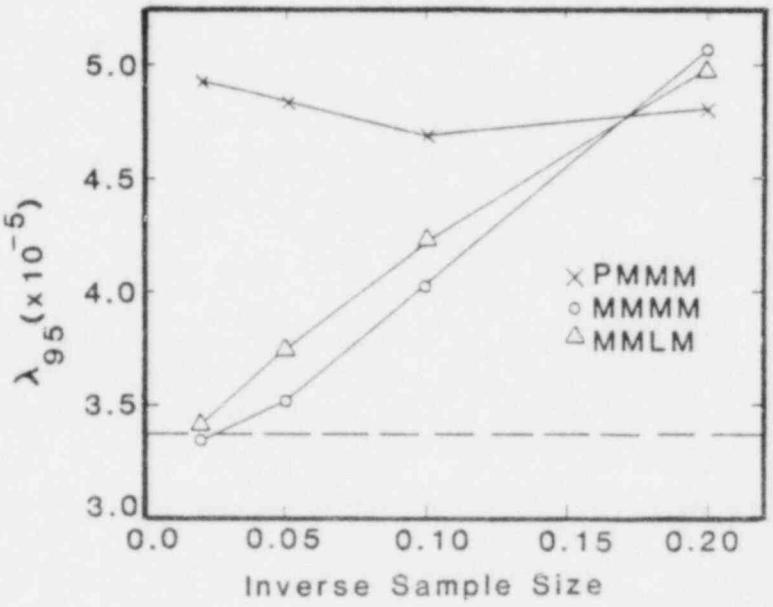


Figure 3.8: Mean of the 95-th percentiles of the estimated gamma distributions for data set 1

3.2.2 Gamma-Poisson Analysis of Gamma-Poisson Data: Set 2

The results shown in the previous section for the simulated date set 1 were obtained from simulated failure data based on a prior gamma distribution which has a mode. To determine whether or not the shape of the prior distribution has any effect on the parameter estimators, the same analysis was repeated with another set of simulated failure data. For the results presented in this section, failure data were simulated from a gamma-Poisson marginal distribution with alpha=0.3, tau=25000 and T=50000 hours. The prior gamma distribution of this data set has the same mean, but has twice as large a standard deviation as that of the previous data set. Specifically, the mean and the standard deviation of the prior gamma distribution are 0.12(-4) and 0.219(-4), respectively. The shape of this gamma prior distribution resembles a negative exponential. The failure data simulated in this section represents those which would be observed during approximately six years of operation from components whose failure rate distribution is a gamma distribution, and whose expected number of failure is the same as that for data set no. 1 (i.e., 1 failure in nine operating years).

Although this set of failure data were simulated based on the prior distribution whose shape is different from that of data set no. 1, the characteristics of this failure data set are very similar to those of the previous data set, namely, the majority of failures is still zero or unity. However, since the standard deviation of the prior distribution of this data set is twice as large as that of previous data set, more failures greater than one (e.g., as large as seven) were obtained.

It was found that (see Table 3.6) the percentage of successful parameter estimation by the PMMM was reduced compared to results for the first data set. This reduction is a direct effect of a higher probability of observing zero failures in data set no. 2 (0.719) compared to 0.615 in data set no. 1 so that more samples with all zero failures were obtained. On the other hand, the degree of success of the other two estimation methods improved significantly by about 20% to 10% from samples of size 5 to 50 as a result of the greater variation of the failure data in a sample.

As with the first data set, inordinately large estimates of the prior parameters from the MMMM still appeared in this data set, but now only 18 such estimates were observed for samples of size 5. Sixteen out of these 18 estimates were found to be generated from the same sample as that which caused the inordinately large estimates for data set no. 1 (namely, 2,2,1,0,0). The other three large estimates were obtained from (2,2,2,0,0) and (4,3,2,2,0). For samples of size 10, all 16 MMMM extraodinarily large estimates were computed from the set (2,1,1,1,0,0,0,0,0,0) which is also the only sample that yielded extraordinarily large prior parameter estimates in data set no. 1.

TABLE 3.6

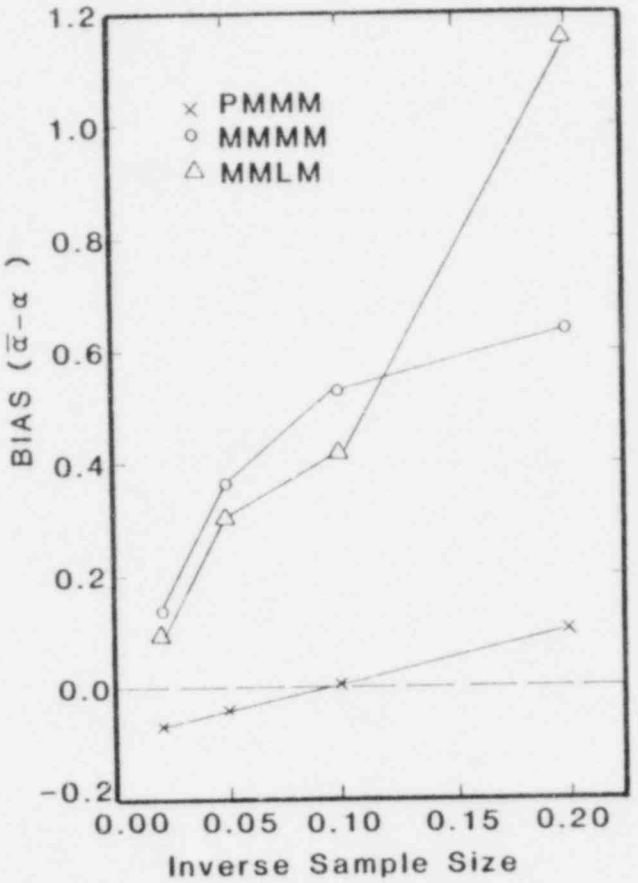
Percentage of successful estimates for data set 2

Method	Sample Size	No. of Successful Solutions(a)	% of Success	Unreasonably Large Solutions(b)
PMMM	5	1615 (2000)	80.7	-
PMMM	10	965 (1000)	96.5	-
PMMM	20	499 (500)	99.8	-
PMMM	50	200 (200)	100.0	-
MMMM	5	980 (2000)	49.0	18
MMMM	10	750 (1000)	75.0	16
MMMM	20	463 (500)	92.6	1
MMMM	50	199 (200)	99.5	-
MMLM	5	975 (2000)	48.7	-
MMLM	10	743 (1000)	74.3	-
MMLM	20	461 (500)	92.2	-
MMLM	50	199 (200)	99.5	-

(a) Numbers in the parentheses are total number of samples.
(b) Estimates were greater than 100 times true values.

The spikey nature of the distributions of prior parameter estimators previously obtained from the estimation methods based on the marginal distribution are also evident for this data set. Surprisingly, for samples of size 5, the spikes of both marginal-based estimation methods occurred at the same locations as those of data set no. 1 although with a smaller amplitude. Since the values of the prior parameters of this data set are smaller, the estimator distributions for all three estimation methods shifted slightly to the left towards the smaller side of the distribution.

The bias of alpha and tau estimators is shown in Fig. 3.9. The variations with sample size of the bias for the MMM and MML methods are again irregular, although both methods give positively biased estimates. Only the bias of the alpha and tau estimators for the PMM method changes from positive bias to negative bias in a smooth manner as the sample size increases. The unbiased estimates of both alpha and tau are observed from PMM method at sample size between 20 and 50.



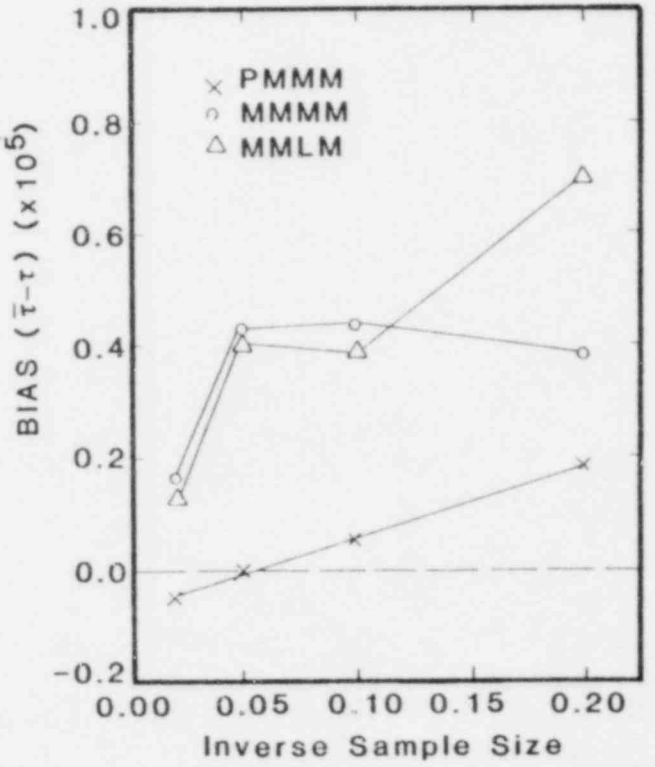


Figure 3.9: Bias of the alpha and tau estimators for data set 2

The same results, as were obtained with the first simulated data set, are obtained for the square root of MSE of the alpha and tau estimates as is shown in Fig. 3.10. The PMM method yields the smallest MSE of the alpha and tau estimates for all sample sizes. The MSE of alpha and tau estimates obtained from the two marginal-based estimation methods still changes irregularly as the sample size changes.

Figure 3.11 shows the variation with sample size of the means of the 95-th percentiles of the estimated gamma distribution. This figure presents only slightly different results from those for data set 1 (see Fig. 3.8), i.e., the variation of the percentile estimates with sample size has the same trend although the magnitude of the estimates are reduced somewhat. The estimates from the MMMM and MMLM are smaller than the true percentile value for sample sizes 20 and 50 but larger for the two smaller sample sizes.

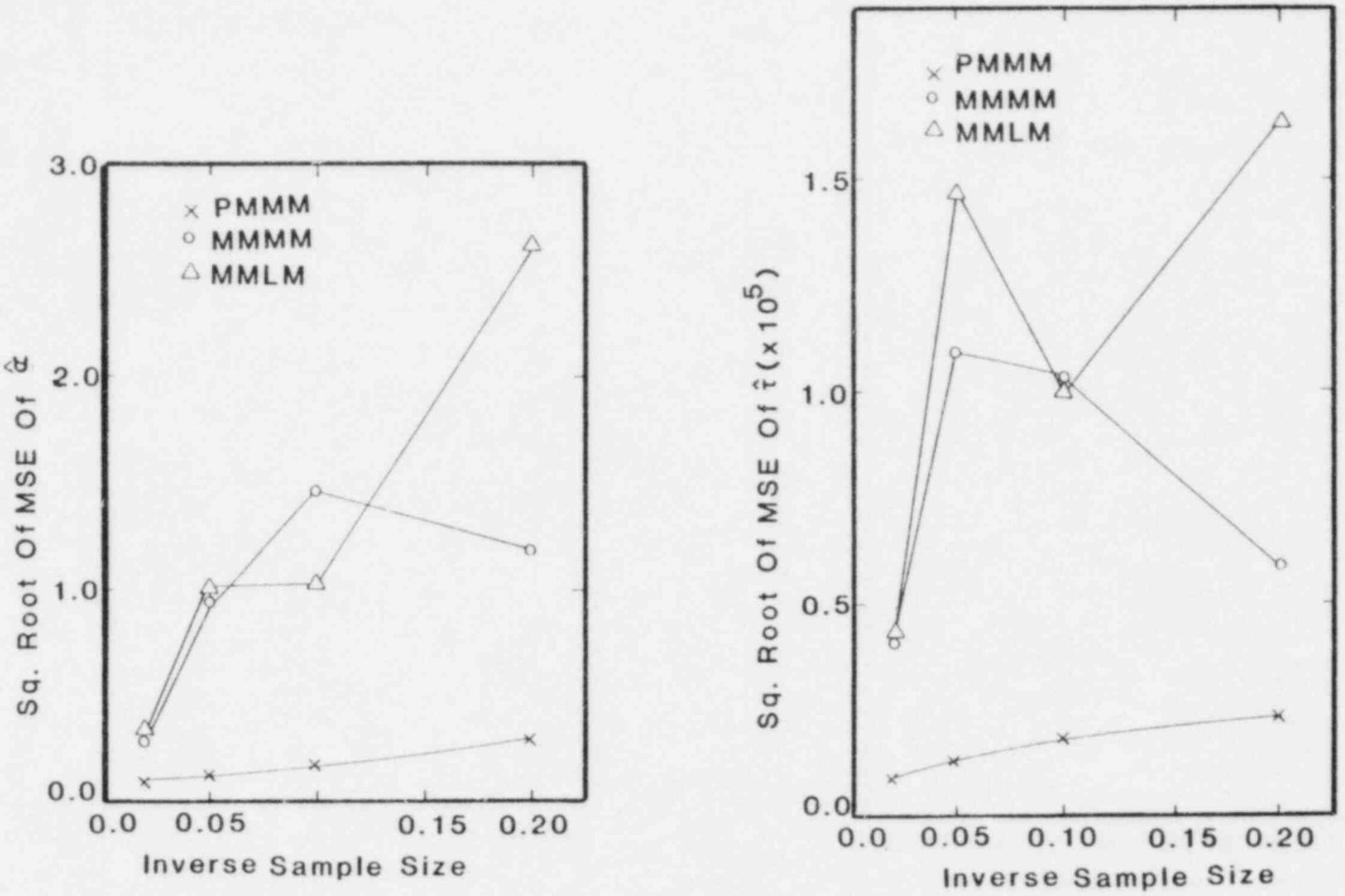


Figure 3.4: Square root of the MSE of the alpha and tau estimators for data set 2

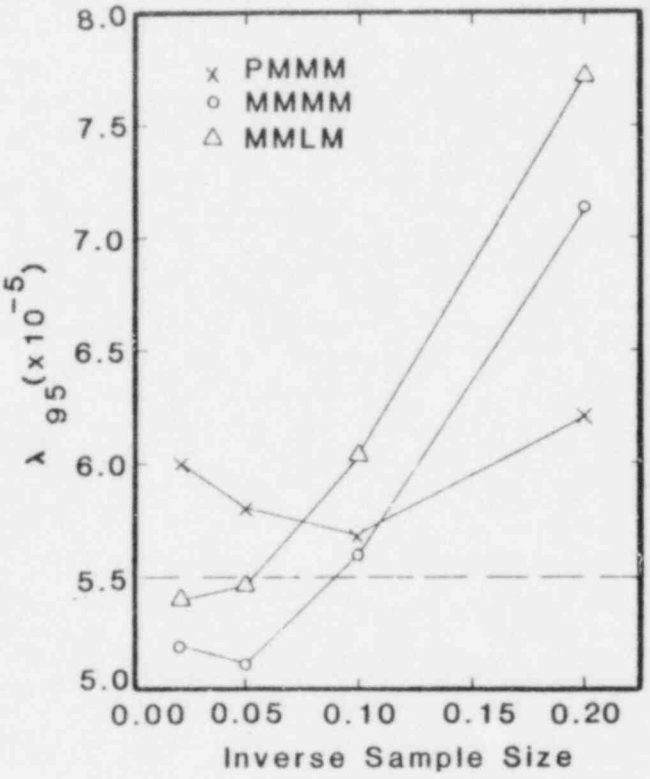


Figure 3.11: Mean of the 95-th percentiles of the estimated gamma distributions for data set 2.

3.2.3 Gamma-Poisson Analysis of Gamma-Poisson Data: Set 3

In an earlier study on the failure-on-demand problem [13], it was found that the prior matching moments method was better than the marginal matching moments and marginal maximum likelihood methods for data samples of size 5 to 50 in terms of the least biasedness, the least mean squared error of prior parameter estimates, and the most conservative estimates of the 95-th percentile of the estimated prior distributions.

In order to compare the results obtained in this study to those of the earlier failure-on-demand study, a third set of failure data was simulated from a gamma-Poisson marginal distribution with the same prior parameters as were used in the first analysis (Set 1) but with the observed time T increased to 1,000,000 hours so as to raise the mean number of observed failures from 0.6 to 12 . In this manner the majority of simulated failure data no longer consisted of zero or one failures, and a far greater variation in the failures in the samples was achieved. As a consequence, the distributions of prior parameter estimators obtained from these data were smooth with none of the spikes obtained with the previous two data sets.

The results obtained from this set of failure data (data set no. 3) match rather closely those obtained for the failure-on-demand case [13], even in terms of percentage of successful estimation of the prior parameters. The parameter estimators from the PMMM for all sample sizes and those from the MMMM and the MMLM for samples of size 20 and 50 have 100% success. Both marginal-based estimation methods yield 99.9% success for samples of size 10 and about 97% for samples of size 5 . The success of all three methods is dramatically improved by using data samples with more than just zero or one failure for each component.

The bias of the alpha and tau estimators as a function of sample size is shown in Fig. 3.12. All estimates from the three estimation methods have positive bias (except for the parameter estimates obtained from the PMMM for samples of size 50 which are only slightly underestimated). As the sample size increases, the bias for all estimation method decreases and approaches zero. Although the PMMM estimates change from positive bias to negative at about sample size 50, this method gives the least biased estimates for all sample sizes considered.

Presented in Fig. 3.13 are variations with sample size of the square root of the MSE of the alpha and tau estimators. The square root of the MSE of the parameter estimates for all methods decreases, as expected, with increasing sample size. The MMLM estimates have a larger MSE than those from the MMMM for samples of size 5 and 10, but have smaller values for samples of size 20 and 50. The PMM method also has the smallest MSE for all sample sizes, although for samples of size 50, all three methods give almost the same results.

The mean of the 95-th percentiles of the estimated prior distributons as a function of sample size are presented in Fig. 3.14. The MMMM and MMLM percentile estimates are underestimated for all sample sizes. Estimates from the PMMM are also underestimated for samples of size 5 and

10, but overestimated for samples of size 20 and 50 . For all sample sizes, except size 50, the PMMM yields percentile estimates closest to the true value. It should be noticed that even though at sample size 50, the MMLM estimate is closest to the true value, it is not a conservative estimate from the point of view of safety analyses for which failure rates should ideally be overestimated.

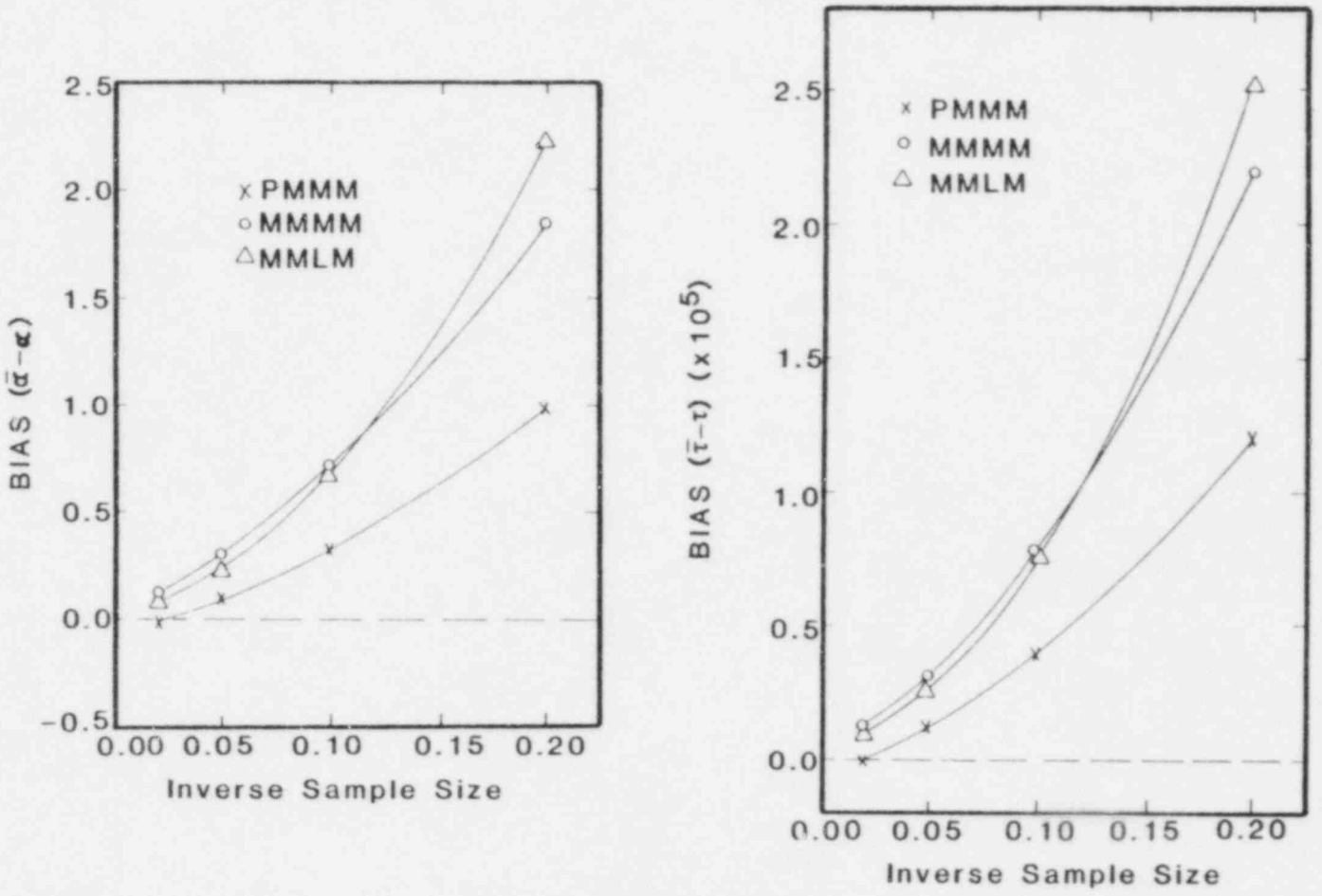


Figure 3.12: Bias of the alpha and tau estimators for data set 3.

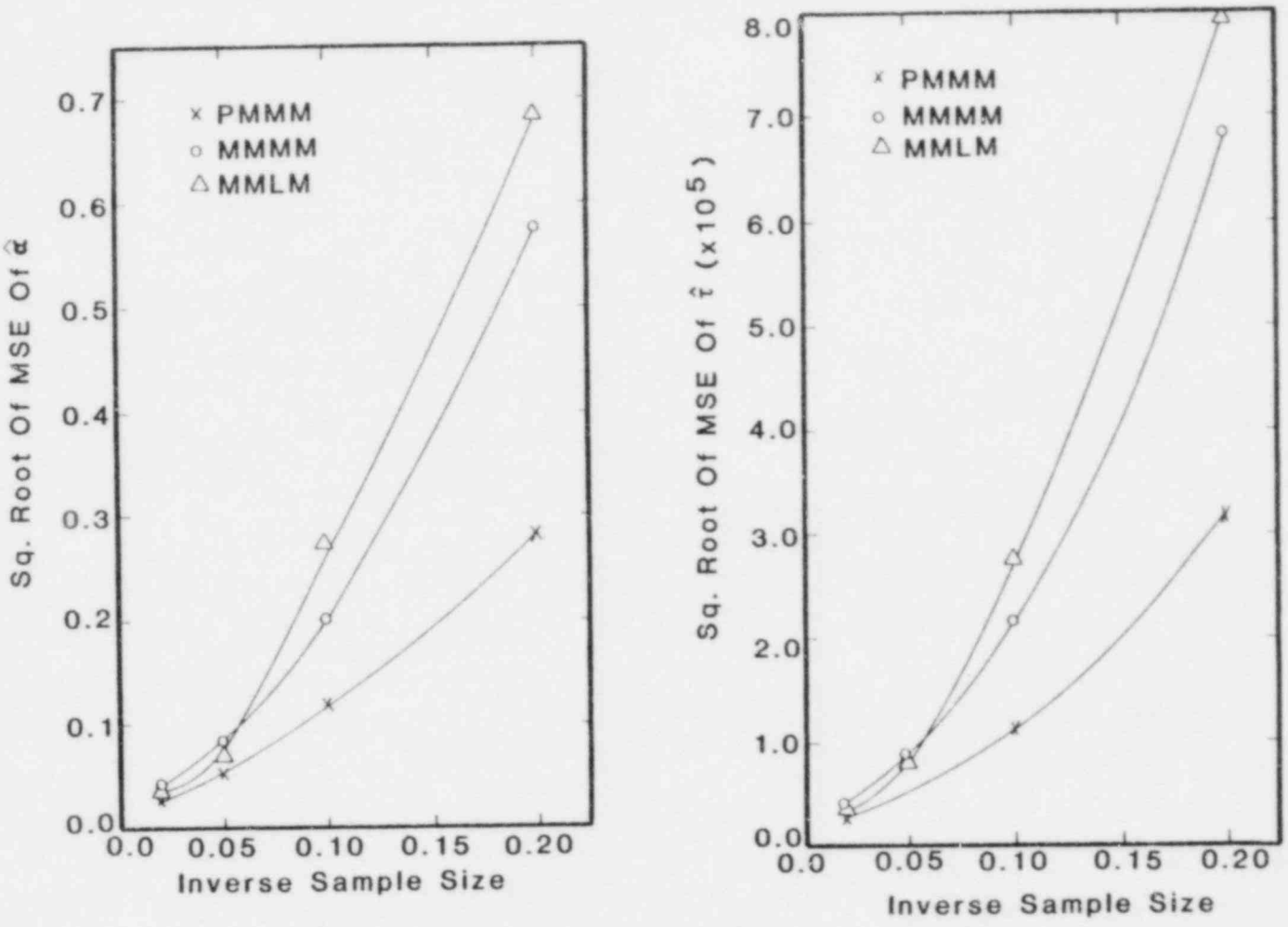


Figure 3.13: Square root of the MSE of the alpha and tau estimators for data set 3

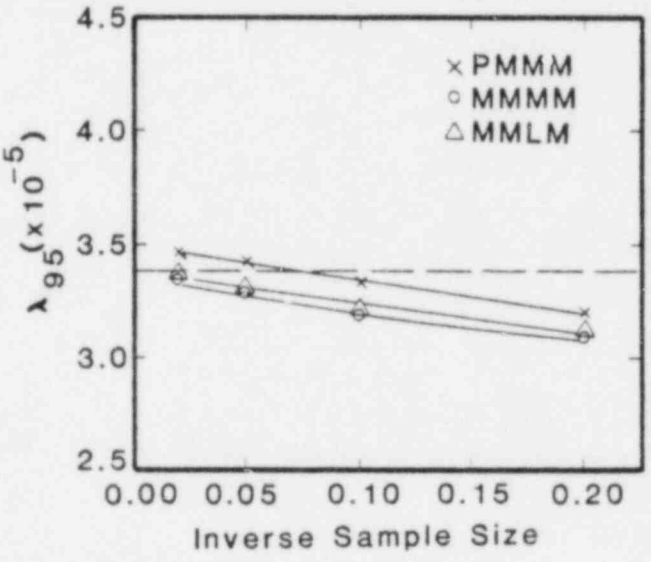


Figure 3.14: Mean of the 95-th percentiles of the estimated gamma prior distributions for data set 3.

3.3 CONJUGATE ANALYSIS OF NON-CONJUGATE FAILURE DATA

Many different distribution families can be used to describe the prior distribution of failure rates for a given set of components. However, using any distribution other than the conjugate gamma distribution presents major difficulties in obtaining numerical results from the compound failure model. Almost all computational steps in an analysis with a non-conjugate prior distribution require numerical methods which are time-consuming and often complicated to program [2]. Therefore, it is of interest to determine how well the conjugate gamma distribution can be used to approximate non-conjugate distributions in a compound failure model.

In this section, failure data were generated from Weibull-Poisson, lognormal-Poisson and logbeta-Poisson marginal distributions with known parameters. These failure data were then analyzed by the conjugate gamma-Poisson failure model and the results of the estimated gamma prior distribution were then compared to the actual non-conjugate prior distributions.

3.3.1 Non-Conjugate Compound Failure Models

A compound failure model is called non-conjugate if a distribution besides a gamma distribution is used to describe the prior distribution of the components' failure rates. Three families of distributions (the Weibull, lognormal and logbeta) were selected as non-conjugate prior distributions for this study. The Weibull and lognormal distributions have two shape parameters and one range (or shift) parameter, while the logbeta distribution has two shape parameters and two range parameters. In this study, the range parameters of these distributions were assumed to be known to simplify the parameter estimation process. Therefore only the two shape parameters of each distribution had to be estimated from the samples of simulated failure data.

3.3.1.1 Non-Conjugate Prior Families

Explicit expressions for the three non-conjugate prior distributions are presented in this section.

(a) Weibull Distribution:

$$g(\lambda;\alpha,\beta,\theta) = \frac{\alpha}{\beta}\left(\frac{\lambda-\theta}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{\lambda-\theta}{\alpha}\right)^{\beta}\right] , \quad \alpha,\beta > 0 , \quad \theta \geq 0 \qquad (3\text{-}2)$$

where α and β are the "scale" and "shape" parameters, respectively, and θ is a "shift" parameter (assumed known -- usually set to zero).

(b) Lognormal Distribution:

$$g(\lambda;\alpha,\beta,\theta) = [(\lambda-\theta)\sqrt{2\pi}\beta]^{-1} \exp[-(\ln(\lambda-\theta)-\alpha)^2/2\beta^2], \quad (3\text{-}3)$$

$$-\infty<\alpha<\infty \ , \ \beta>0, \ \theta>0$$

where α and β are the shape parameters, and θ is a "shift" parameter (assumed to be known -- usually set to zero).

(c) Logbeta Distribution:

$$g(\lambda;\alpha,\beta,a,b) = \frac{\Gamma(\alpha+\beta)}{\lambda\Gamma(\alpha)\Gamma(\beta)} (b-a)^{1-\alpha-\beta} (\ln\lambda-a)^{\alpha-1} (b-\ln\lambda)^{\beta-1} \quad (3\text{-}4)$$

$$\alpha,\beta > 0 \ , \ e^a < \lambda < e^b$$

where α and β are the "shape" parameters, and a and b are the "range" parameters which are assumed known in this study.

The non-conjugate prior distributions can be combined with the Poisson likelihood distribution to obtain the non-conjugate marginal distribution as shown by Eq. (1-3). However, numerical techniques, often quite elaborate, must be used to evaluate the integral in Eq. (1-3), and thus such compound non-conjugate failure models are rather cumbersome for routine analyses.

3.3.1.2 Parameter Estimation of Non-Conjugate Compound Failure Models

The prior parameter estimation techniques, employed in the conjugate compound failure model (i.e., PMMM, MMMM, and MMLM), can also be used for the non-conjugate models. The details of the parameter estimation methods for individual non-conjugate compound failure models can be found in Refs. [2] and [14].

3.3.2 Results of Using a Conjugate Model to Analyze Data from Non-Conjugate Models

A total of 500 failure data were simulated from each of the three non-conjugate distributions, and these data were then grouped into 25 sets of samples of size 20. Parameter values for the three non-conjugate prior distributions were chosen so that the prior means and standard deviations were the same as those of the gamma distribution used in Section 3.2.1 . Specifically, the parameters alpha and beta for the Weibull distribution are 0.12(-4) and 1.096; for the lognormal distribution they are -11.634 and 0.778 and for the logbeta distribution they are 281.6 and 140.4. The observed component test time was taken as 50,000 hours.

From the gamma parameters estimated by the PMMM, MMMM and MMLM from the simulated non-conjugate failure data, various characteristics of the resulting gamma distributions (i.e., those defined by the estimated prior parameters) were compared to the same characteristics of the known non-conjugate prior distributions. These comparisons are summarized in

Tables 3.7, 3.8 and 3.9 for failure data obtained from the known Weibull-Poisson, lognormal-Poisson and logbeta-Poisson marginal distributions, respectively.

The results obtained from using the gamma-Poisson model to analyze failure data simulated from non-conjugate failure models often matched reasonably well with the actual values. As will be seen in the next section, even when the correct models are used to analyze these same sets of simulated failure data, perfect agreement is still not obtained. Furthermore, in some cases, the approximate model (i.e., the gamma-Poisson model) even yielded parameter estimators closer to the actual values than did the correct models. Thus, for practical purposes, the gamma-Poisson model should be regarded as a very good model for analyzing failure data even though the data are believed to come from another failure distributon.

TABLE 3.7

Results of a gamma-Poisson analysis of data from a Weibull-Poisson distribution

Parameter & True Value	Method	Mean	Standard Deviation	Sq. Root of MSE	Bias
mean	PMMM	1.20(-5)	.397(-5)	.389(-5)	.004(-5)
1.20(-5)	MMMM	1.25(-5)	.395(-5)	.388(-5)	.053(-5)
	MMLM	1.24(-5)	.432(-5)	.421(-5)	.044(-5)
standard	PMMM	1.81(-5)	.496(-5)	.864(-5)	.714(-5)
deviation	MMMM	1.10(-5)	.568(-5)	.553(-5)	.004(-5)
1.10(-5)	MMLM	1.39(-5)	.568(-5)	.623(-5)	.029(-5)
5-th	PMMM	.009(-5)	.015(-5)	.075(-5)	-.074(-5)
percentile	MMMM	.235(-5)	.352(-5)	.375(-5)	.152(-5)
.083(-5)	MMLM	.096(-5)	.145(-5)	.141(-5)	.013(-5)
25-th	PMMM	.120(-5)	.106(-5)	.297(-5)	-.278(-5)
percentile	MMMM	.490(-5)	.379(-5)	.380(-5)	.092(-5)
.399(-5)	MMLM	.329(-5)	.295(-5)	.294(-5)	-.070(-5)
50-th	PMMM	.504(-5)	.287(-5)	.477(-5)	-.385(-5)
percentile	MMMM	.910(-5)	.390(-5)	.380(-5)	.021(-5)
.889(-5)	MMLM	.770(-5)	.415(-5)	.419(-5)	-.119(-5)
75-th	PMMM	1.53(-5)	.602(-5)	.608(-5)	-.147(-5)
percentile	MMMM	1.65(-5)	.555(-5)	.541(-5)	-.021(-5)
1.67(-5)	MMLM	1.65(-5)	.607(-5)	.588(-5)	-.020(-5)
95-th	PMMM	4.78(-5)	1.38(-5)	1.95(-5)	1.41(-5)
percentile	MMMM	3.44(-5)	1.39(-5)	1.35(-5)	.056(-5)
3.38(-5)	MMLM	4.00(-5)	1.46(-5)	1.54(-5)	.621(-5)

TABLE 3.8

Results of a gamma-Poisson analysis of data from a lognormal-Poisson distribution

Parameter & True Value	Method	Mean	Standard Deviation	Sq. Root of MSE	Bias
mean	PMMM	1.20(-5)	.418(-5)	.409(-5)	-.004(-5)
1.20(-5)	MMMM	1.25(-5)	.419(-5)	.411(-5)	.047(-5)
	MMLM	1.24(-5)	.459(-5)	.446(-5)	.038(-5)
standard	PMMM	1.79(-5)	.524(-5)	.863(-5)	.694(-5)
deviation	MMMM	1.07(-5)	.598(-5)	.582(-5)	-.023(-5)
1.10(-5)	MMLM	1.32(-5)	.604(-5)	.625(-5)	.221(-5)
5-th	PMMM	.010(-5)	.015(-5)	.221(-5)	-.221(-5)
percentile	MMMM	.247(-5)	.351(-5)	.342(-5)	.017(-5)
.230(-5)	MMLM	.121(-5)	.167(-5)	.195(-5)	-.109(-5)
25-th	PMMM	.123(-5)	.110(-5)	.403(-5)	-.389(-5)
percentile	MMMM	.505(-5)	.387(-5)	.377(-5)	-.006(-5)
.512(-5)	MMLM	.367(-5)	.324(-5)	.346(-5)	-.145(-5)
50-th	PMMM	.506(-5)	.299(-5)	.478(-5)	-.377(-5)
percentile	MMMM	.917(-5)	.406(-5)	.397(-5)	.034(-5)
.884(-5)	MMLM	.799(-5)	.442(-5)	.437(-5)	-.084(-5)
75-th	PMMM	1.52(-5)	.628(-5)	.616(-5)	.006(-5)
percentile	MMMM	1.64(-5)	.582(-5)	.580(-5)	.127(-5)
1.51(-5)	MMLM	1.64(-5)	.637(-5)	.630(-5)	.128(-5)
95-th	PMMM	4.74(-5)	1.46(-5)	2.08(-5)	1.51(-5)
percentile	MMMM	3.37(-5)	1.47(-5)	1.44(-5)	.148(-5)
3.22(-5)	MMLM	3.84(-5)	1.55(-5)	1.62(-5)	.622(-5)

TABLE 3.9

Results of a gamma-Poisson analysis of data from a logbeta-Poisson distribution

Parameter & True Value	Method	Mean	Standard Deviation	Sq. Root of MSE	Bias
mean	PMMM	1.20(-5)	.418(-5)	.409(-5)	-.004(-5)
1.20(-5)	MMMM	1.25(-5)	.419(-5)	.411(-5)	.047(-5)
	MMLM	1.24(-5)	.459(-5)	.446(-5)	.038(-5)
standard	PMMM	1.79(-5)	.524(-5)	.863(-5)	.694(-5)
deviation	MMMM	1.07(-5)	.598(-5)	.582(-5)	-.023(-5)
1.10(-5)	MMLM	1.32(-5)	.604(-5)	.625(-5)	.221(-5)
5-th	PMMM	.009(-5)	.015(-5)	.237(-5)	-.237(-5)
percentile	MMMM	.247(-5)	.351(-5)	.342(-5)	.001(-5)
.246(-5)	MMLM	.121(-5)	.167(-5)	.204(-5)	-.125(-5)
25-th	PMMM	.123(-5)	.110(-5)	.415(-5)	-.401(-5)
percentile	MMMM	.505(-5)	.387(-5)	.377(-5)	-.019(-5)
.524(-5)	MMLM	.367(-5)	.324(-5)	.351(-5)	-.158(-5)
50-th	PMMM	.506(-5)	.299(-5)	.480(-5)	-.380(-5)
percentile	MMMM	.917(-5)	.406(-5)	.396(-5)	.031(-5)
.886(-5)	MMLM	.799(-5)	.442(-5)	.437(-5)	-.087(-5)
75-th	PMMM	1.52(-5)	.628(-5)	.616(-5)	.020(-5)
percentile	MMMM	1.64(-5)	.582(-5)	.583(-5)	.141(-5)
1.50(-5)	MMLM	1.64(-5)	.637(-5)	.633(-5)	.142(-5)
95-th	PMMM	4.74(-5)	1.46(-5)	2.10(-5)	1.54(-5)
percentile	MMMM	3.37(-5)	1.47(-5)	1.44(-5)	.180(-5)
3.19(-5)	MMLM	3.84(-5)	1.55(-5)	1.63(-5)	.654(-5)

3.4 NON-CONJUGATE ANALYSIS OF DATA FROM NON-CONJUGATE MODELS

The results of the previous section showed that for failure data arising from non-conjugate marginal distributions, the gamma-Poisson conjugate model can be successfully used to analyze those data and to yield reasonably good estimates of the properties of the prior distribution.

To confirm these results, failure data generated from the three non-conjugate models were analyzed by the same marginal models (i.e., the Weibull-Poisson, lognormal-Poisson and logbeta-Poisson compound models) from which the data were generated.

Tables 3.10, 3.11 and 3.12 present the results of using the correct non-conjugate models to analyze the simulated failure data from the Weibull-Poisson, lognormal-Poisson and logbeta-Poisson marginal distributions, respectively. Only the two matching moments estimation methods (PMMM and MMMM) were used in this phase of the study. The marginal maximum likelihood estimation method applied to non-conjugate failure models requires an inordinate amount of computational effort for the analysis of a large number of data samples, and therefore this method was omitted in this phase of the study.

Of the four prior models investigated in this study, the logbeta-Poisson model presents the most difficulties in calculating parameter estimates even for the simple prior matching moments method. An elaborate numerical algorithm is still required to obtain estimates of alpha and beta from both the PMMM and MMMM parameter estimation methods [2]. On the other hand, estimates of alpha and tau can be obtained easily with a handheld calculator by PMMM and MMMM methods for the gamma-Poisson model.

The use of a Weibull-Poisson model to analyze failure data from a Weibull-Poisson marginal distribution (as is seen from Table 3.10) gave better estimators of alpha, but poorer estimators of beta from the MMMM than those from the PMMM. The means of the estimated prior distributions from two matching moments methods revealed only slight disagreement, which is basically due to the different number of successful estimates. The mean of the standard deviations of the estimated prior distributions obtained by the MMMM have less bias and smaller MSE than those from the PMMM. Only the 5-th percentile estimators of the PMMM gave better results than those of the MMMM. For the other four percentiles (25-th, 50-th, 75-th and 95-th), the MMMM yielded more desirable estimators than the PMMM, both in terms of bias and MSE. In all, the MMMM showed better results than the PMMM.

In comparison with results in Table 3.7, which were obtained by using a gamma-Poisson model to analyze failure data from a Weibull-Poisson marginal distribution, the estimators of the means and the standard deviation of the estimated prior distribution from two matching moments methods are the same for both failure models. The correctly assumed model gave better estimators of the 5-th and 95-th percentiles, but poorer estimators of the 75-th percentiles, with both estimation methods

TABLE 3.10

Results of a Weibull-Poisson analysis of data from a Weibull-Poisson distribution

Parameter & True Value	Method	Mean	Standard Deviation	Sq. Root of MSE	Bias
alpha	PMMM	.938(-5)	.399(-5)	.495(-5)	-.304(-5)
1.24(-5)	MMMM	1.25(-5)	.445(-5)	.433(-5)	.004(-5)
beta	PMMM	.683	.126	.432	-.413
1.10	MMMM	2.04	2.44	2.56	.944
mean	PMMM	1.20(-5)	.397(-5)	.389(-5)	.004(-5)
1.20(-5)	MMMM	1.25(-5)	.395(-5)	.388(-5)	.053(-5)
stan. dev..	PMMM	1.81(-5)	.496(-5)	.864(-5)	.714(-5)
1.10(-5)	MMMM	1.10(-5)	.568(-5)	.553(-5)	.005(-5)
5-th pctl.	PMMM	.017(-5)	.016(-5)	.068(-5)	-.066(-5)
.083(-5)	MMMM	.219(-5)	.329(-5)	.348(-5)	.136(-5)
25-th pctl.	PMMM	.168(-5)	.105(-5)	.253(-5)	-.231(-5)
.399(-5)	MMMM	.500(-5)	.371(-5)	.375(-5)	.101(-5)
50-th pctl.	PMMM	.561(-5)	.269(-5)	.421(-5)	-.328(-5)
.890(-5)	MMMM	.926(-5)	.393(-5)	.384(-5)	.037(-5)
75-th pctl.	PMMM	1.49(-5)	.569(-5)	.586(-5)	-.182(-5)
1.67(-5)	MMMM	1.65(-5)	.563(-5)	.548(-5)	-.022(-5)
95-th pctl.	PMMM	4.55(-5)	1.38(-5)	1.79(-5)	1.18(-5)
3.38(-5)	MMMM	3.40(-5)	1.39(-5)	1.36(-5)	.018(-5)

TABLE 3.11

Results of a lognormal-Poisson analysis of data from a lognormal-Poisson distribution

Parameter & True Value	Method	Mean	Standard Deviation	Sq. Root of MSE	Bias
alpha	PMMM	-12.03	.568	.683	-0.40
-11.63	MMMM	-11.67	.526	.514	-0.04
beta	PMMM	1.105	.148	.357	.326
.778	MMMM	0.734	.303	.299	-.044
mean	PMMM	1.20(-5)	.418(-5)	.409(-5)	-.004(-5)
1.20(-5)	MMMM	1.25(-5)	.429(-5)	.411(-5)	.047(-5)
stan. dev.	PMMM	1.79(-5)	.524(-5)	.863(-5)	.694(-5)
1.10(-5)	MMMM	1.07(-5)	.598(-5)	.582(-5)	-.023(-5)
5-th pctl.	PMMM	.085(-5)	.053(-5)	.053(-5)	-.162(-5)
.246(-5)	MMMM	.222(-5)	.212(-5)	.252(-5)	-.024(-5)
25-th pctl.	PMMM	.260(-5)	.138(-5)	.135(-5)	-.264(-5)
.524(-5)	MMMM	.449(-5)	.268(-5)	.326(-5)	-.075(-5)
50-th pctl.	PMMM	.573(-5)	.272(-5)	.266(-5)	-.314(-5)
.886(-5)	MMMM	.780(-5)	.334(-5)	.384(-5)	-.107(-5)
75-th pctl.	PMMM	1.27(-5)	.555(-5)	.545(-5)	-.226(-5)
1.50(-5)	MMMM	1.42(-5)	.541(-5)	.541(-5)	-.079(-5)
95-th pctl.	PMMM	4.06(-5)	1.68(-5)	1.66(-5)	.870(-5)
3.19(-5)	MMMM	3.59(-5)	1.73(-5)	1.81(-5)	.400(-5)

TABLE 3.12

Results of a logbeta-Poisson analysis of data from a logbeta-Poisson distribution

Parameter & True Value	Method	Mean	Standard Deviation	Sq.Root of MSE	Bias
alpha	PMMM	136.	39.3	150.	-145.
282	MMMM	136.(1)	259.(1)	274.(1)	108.(1)
beta	PMMM	70.9	17.0	71.5	-69.5
140	MMMM	653.	122.(1)	130.(1)	512.
mean	PMMM	1.20(-5)	.418(-5)	.409(-5)	-.004(-5)
1.20(-5)	MMMM	1.25(-5)	.420(-5)	.411(-5)	.047(-5)
stan. dev.	PMMM	1.79(-5)	.524(-5)	.863(-5)	.694(-5)
1.10(-5)	MMMM	1.07(-5)	.599(-5)	.583(-5)	-.023(-5)
5-th pctl.	PMMM	.100(-5)	.061(-5)	.143(-5)	-.130(-5)
.230(-5)	MMMM	.351(-5)	.319(-5)	.333(-5)	.120(-5)
25-th pctl.	PMMM	.303(-5)	.152(-5)	.257(-5)	-.209(-5)
.512(-5)	MMMM	.606(-5)	.342(-5)	.346(-5)	.095(-5)
50-th pctl.	PMMM	.649(-5)	.284(-5)	.364(-5)	-.234(-5)
.884(-5)	MMMM	.940(-5)	.379(-5)	.373(-5)	.056(-5)
75-th pctl.	PMMM	1.38(-5)	.531(-5)	.536(-5)	-.129(-5)
1.51(-5)	MMMM	1.51(-5)	.531(-5)	.517(-5)	.002(-5)
95-th pctl.	PMMM	4.04(-5)	1.31(-5)	1.52(-5)	.814(-5)
3.22(-5)	MMMM	3.16(-5)	1.33(-5)	1.30(-5)	-.065(-5)

than did the gamma-Poisson model. For the estimates of the 50-th percentile, the Weibull-Poisson model yielded less biased and smaller MSE estimators from the PMMM, but yielded the opposite results with the MMMM. Generally, both models showed comparable results from each estimation method considered in this section.

The lognormal-Poisson model also yielded results that favored the marginal matching moments estimation method. All estimates from the MMMM are less biased than those obtained from the PMMM, and almost all estimates from the MMMM have less MSE than those from the PMMM with the exception of the mean, and the 5-th and 25-th percentiles of the estimated prior distribution. A comparison of these lognormal-Poisson results to those obtained with a gamma-Poisson analysis (Table 3.8) shows the mean and standard deviation of the estimated prior distribution for both models are the same. However, the MMMM method gave better estimates of the 5-th, 25-th, 50-th and 95-th percentiles in terms of bias and MSE using a gamma-Poisson model than those from a lognormal-Poisson model which is the correct model. On the other hand the PMMM method did a better job with a lognormal-Poisson than with a gamma-Poisson model.

As is shown in Table 3.12, the PMMM gave better estimates for the parameters of the logbeta prior distribution than did the MMMM in terms having smaller bias and smaller of MSE. The mean of the PMMM estimated prior distribution has a slight advantage over that obtained from the MMMM, mainly because the MMMM failed to give successful estimates for six data sets. However, the MMMM showed better results for estimates of the prior standard deviation. The first three percentiles, i.e., the 5-th, 25-th and 50-th percentiles, favor the PMMM in terms of a smaller MSE, but estimates from the MMMM have smaller bias. For the 95-th percentile, an important characteristic in safety studies, estimates from the MMMM are better than those from the PMMM. A similar conclusion also holds for the 75-th percentile.

A comparison of the above results to those from the gamma-Poisson model in Table 3.9 shows that almost the same values are obtained for the mean and standard deviation of the estimated prior distribution. Estimates of the first three percentiles (i.e., 5-th, 25-th and 50-th percentiles) from the PMMM with a logbeta-Poisson model have smaller bias and smaller MSE, but estimates from the MMMM seem to give comparable values of MSE. For 75-th percentile, both the conjugate and non-conjugate models yielded values close to the true prior parameters for both parameter estimation methods. Finally, estimates from a logbeta-Poisson model have smaller bias and smaller MSE for both parameter estimation methods. However, both parameter estimation methods applied to the gamma-Poisson model yielded conservative estimates of the 95-th percentile while the least-biased estimate (which is obtained from the MMM method with a logbeta-Poisson model) is underestimated.

3.5 COMPARISON OF USING FOUR DIFFERENT PRIOR DISTRIBUTIONS WITH THE SAME FAILURE DATA

In another effort to show the effect of using different prior distributions in the failure rate analysis, two failure data sets of sample size 5 were selected from the simulated failure data used in Section `3.2.3` (data set 3). These data sets were then analyzed with the three non-conjugate failure models, i.e., Weibull-Poisson, lognormal-Poisson, and logbeta-Poisson. For each failure model, estimates of the prior parameters were calculated by all three estimation methods, and for each pair of prior parameter estimates, the estimated prior distributions were plotted. Five prior percentiles, i.e., 5-th, 25-th, 50-th, 75-th and 95-th percentiles, of each estimated prior distribution were also computed.

The two failure data sets chosen for this phase of the study were (2,3,10,11,28) and (4,8,10,12,13). The first data sample, called good data, was selected because it yielded prior gamma parameter estimates with small bias for all three parameter estimation methods. The other sample, called bad data, was used because it gave very poor prior parameter estimates. The parameter estimates for each prior distribution obtained with these data sets and each estimation method are given in Table 3.13.

Figures 3.15, 3.16 and 3.17 show the estimated prior distributions belonging to the four prior families with parameters estimated by the PMMM, MMMM and MMLM, respectively, for the good data sample. All three figures show the same characteristic, i.e., close similarity between the estimated gamma and Weibull distributions and between the lognormal and logbeta distributions. Even though the low failure rate tails of these distributions are somewhat different, the more important part, i.e., the high failure rate tails of the distributions, are very close to one another. This characteristic is confirmed by the results in Table 3.14 which gives the 5-th, 25-th, 50-th, 75-th and 95-th percentiles of the estimated prior distributions of Figures 3.14, 3.15 and 3.16.

Figures 3.18, 3.19, and 3.20 and Table 3.15 represent similar results obtained from the bad data sample. Close similarity between the lognormal and logbeta distributions are still evident for this bad data set. The estimated prior distributions from the gamma and Weibull distributions show some minor differences, except those obtained from the MMMM estimation method. However, the results in Table 3.15 reveals very close estimates of all five percentiles for the different estimated distributions. Thus it is concluded that any difficulty with estimating the prior distribution from the failure data is inherent in the data sample and not in the parameter estimation technique used.

TABLE 3.13

Parameter estimates for four prior models obtained by three estimation techniques for GOOD and BAD data

Data Set	Prior[a]	Method	Alpha	Beta[b]
"GOOD"	Gamma	PMMM	1.073	9.936(4)
		MMMM	1.191	1.103(5)
		MMLM	1.543	1.429(5)
	Weibull	PMMM	1.096(-3)	1.036
		PMMM	1.096(-3)	1.036
		MMLM	1.167(-3)	1.267
	Lognormal	PMMM	-11.76	0.811
		MMMM	-11.74	0.781
		MMLM	-11.78	0.860
	Logbeta	PMMM	257.7	130.7
		MMMM	262.0	133.5
		MMLM	242.7	123.1
"BAD"	Gamma	PMMM	6.903	7.348(5)
		MMMM	25.99	2.765(6)
		MMLM	82.94	8.823(6)
	Weibull	PMMM	1.055(-3)	2.849
		MMMM	1.014(-3)	5.921
		MMLM	9.996(-4)	7.786
	Lognormal	PMMM	-11.64	3.678(-1)
		MMMM	-11.59	1.943(-1)
		MMLM	-11.58	0.988(-1)
	Logbeta	PMMM	1.323(3)	6.595(2)
		MMMM	4.788(3)	2.371(3)
		MMLM	1.626(4)	8.038(3)

[a] The range or shift parameters are selected so that the support of the prior distribution is (0,).

[b] For the gamma distribution, these are the tau parameter.

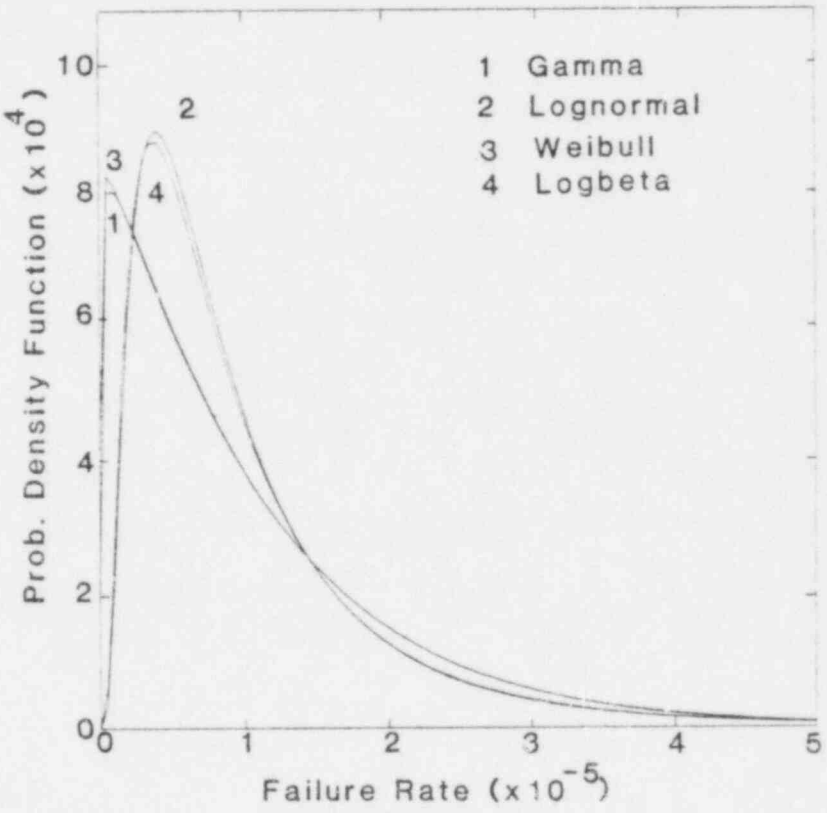


Figure 3.15: Four estimated prior distributions from the PMMM with GOOD DATA

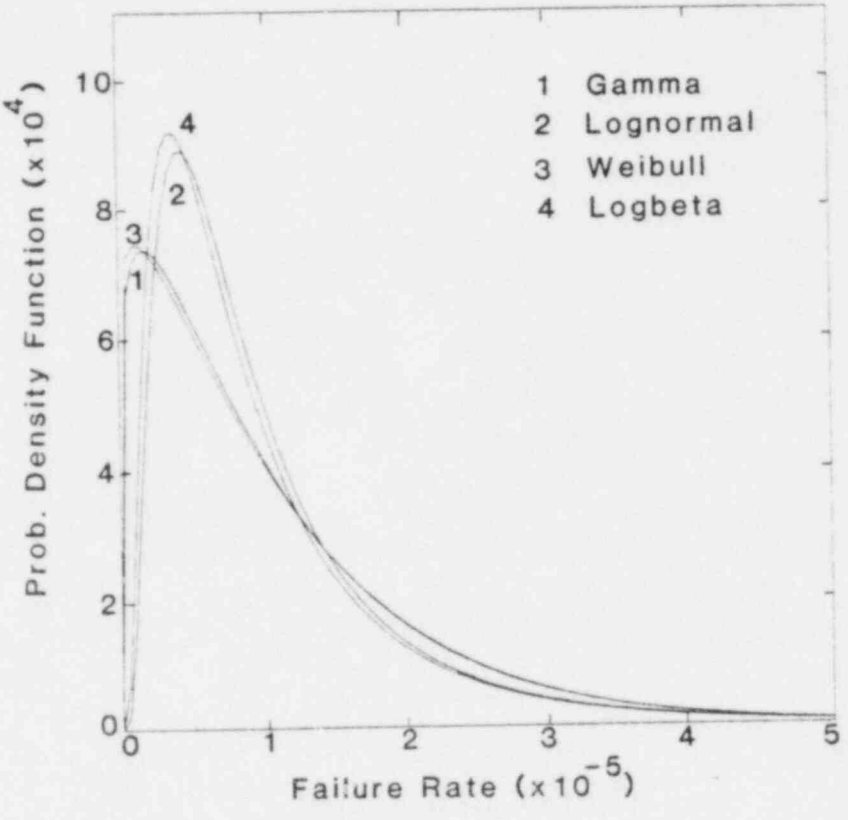


Figure 3.16: Four estimated prior distributions from the MMMM with GOOD DATA

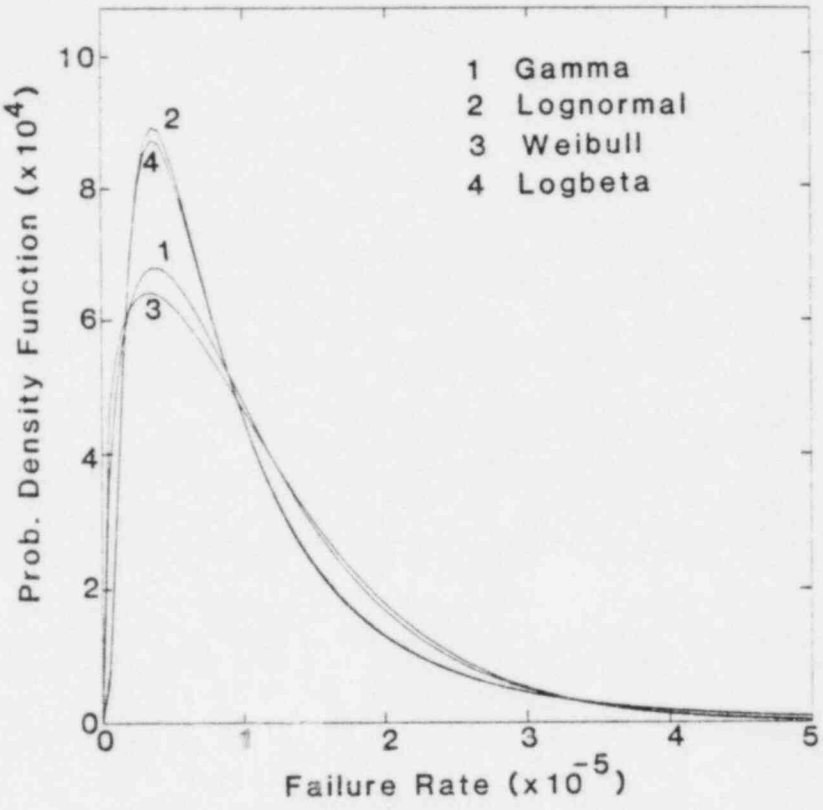


Figure 3.17: Four estimated prior distributions from the MMLM with GOOD DATA

TABLE 3.14

Percentiles of four prior distributions obtained from three estimation methods with GOOD DATA

Method	Prior Dist.	Percentiles				
		5-th	25-th	50-th	75-th	95-th
PMMM	gamma	.066(-5)	.333(-5)	.770(-5)	1.50(-5)	3.16(-5)
PMMM	Weibull	.062(-5)	.329(-5)	.769(-5)	1.50(-5)	3.16(-5)
PMMM	lognormal	.205(-5)	.449(-5)	.777(-5)	1.34(-5)	2.95(-5)
PMMM	logbeta	.190(-5)	.437(-5)	.773(-5)	1.36(-5)	2.99(-5)
MMMM	gamma	.082(-5)	.366(-5)	.797(-5)	1.49(-5)	3.04(-5)
MMMM	Weibull	.074(-5)	.357(-5)	.798(-5)	1.50(-5)	3.05(-5)
MMMM	lognormal	.220(-5)	.470(-5)	.796(-5)	1.35(-5)	2.87(-5)
MMMM	logbeta	.185(-5)	.423(-5)	.745(-5)	1.30(-5)	2.85(-5)
MMLM	gamma	.133(-5)	.445(-5)	.858(-5)	1.48(-5)	2.79(-5)
MMLM	Weibull	.112(-5)	.436(-5)	.874(-5)	1.51(-5)	2.78(-5)
MMLM	lognormal	.186(-5)	.429(-5)	.766(-5)	1.37(-5)	3.15(-5)
MMLM	logbeta	.181(-5)	.429(-5)	.773(-5)	1.38(-5)	3.11(-5)

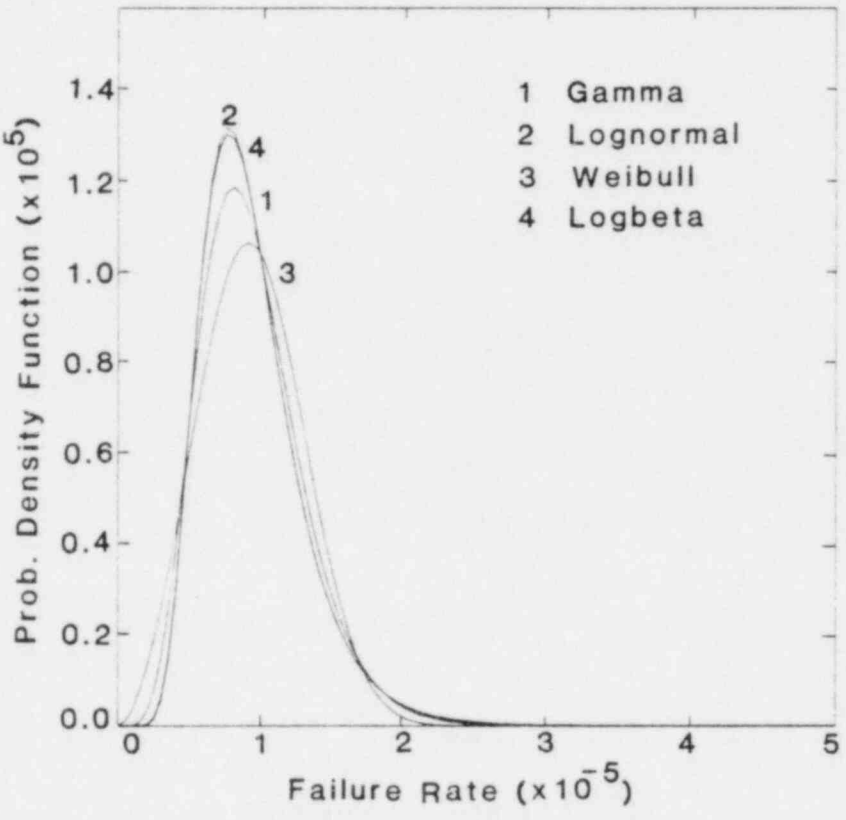


Figure 3.18: Four estimated prior distributions obtained by the PMMM with the BAD DATA

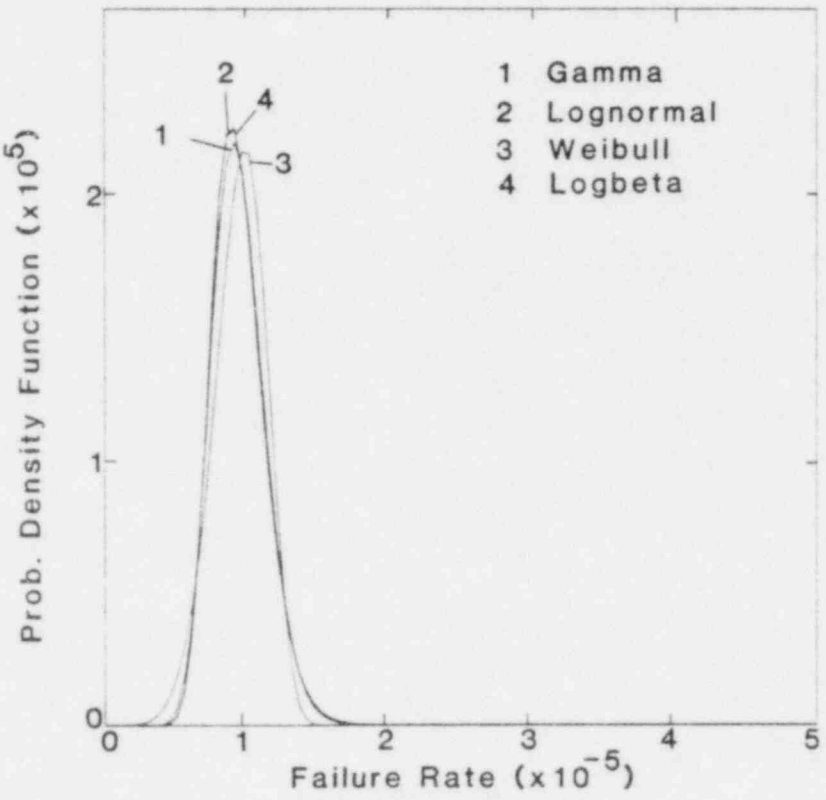


Figure 3.19: Four estimated prior distributions obtained by the MMMM with the BAD DATA

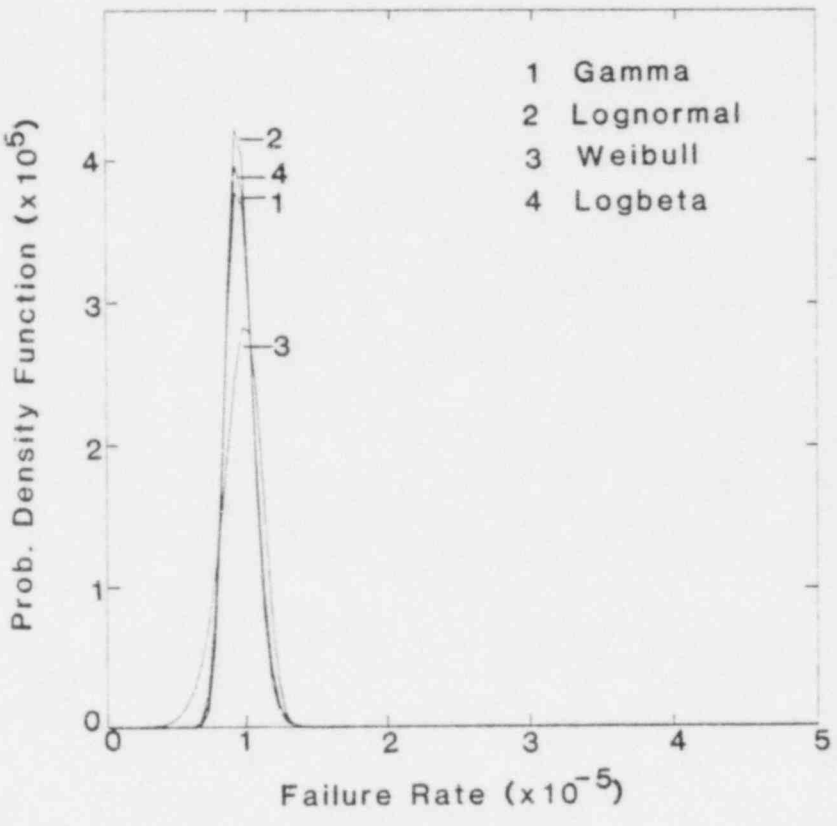


Figure 3.20: Four estimated prior distributions obtained by the MMLM with the BAD DATA

TABLE 3.15

Percentiles of four estimated prior distributions obtained with the three estimation methods and the BAD DATA

Method	Prior Dist.	Percentiles				
		5-th	25-th	50-th	75-th	95-th
PMMM	gamma	.438(-5)	.681(-5)	.895(-5)	1.15(-5)	1.59(-5)
PMMM	Weibull	.372(-5)	.681(-5)	.928(-5)	1.18(-5)	1.55(-5)
PMMM	lognormal	.480(-5)	.685(-5)	.878(-5)	1.13(-5)	1.61(-5)
PMMM	logbeta	.476(-5)	.684(-5)	.879(-5)	1.13(-5)	1.61(-5)
MMMM	gamma	.659(-5)	.810(-5)	.928(-5)	1.06(-5)	1.26(-5)
MMMM	Weibull	.614(-5)	.822(-5)	.953(-5)	1.07(-5)	1.22(-5)
MMMM	lognormal	.670(-5)	.809(-5)	.922(-5)	1.05(-5)	1.27(-5)
MMMM	logbeta	.669(-5)	.809(-5)	.923(-5)	1.05(-5)	1.27(-5)
MMLM	gamma	.777(-5)	.868(-5)	.936(-5)	1.01(-5)	1.12(-5)
MMLM	Weibull	.683(-5)	.852(-5)	.954(-5)	1.04(-5)	1.15(-5)
MMLM	lognormal	.795(-5)	.875(-5)	.935(-5)	1.00(-5)	1.10(-5)
MMLM	logbeta	.785(-5)	.870(-5)	.934(-5)	1.00(-5)	1.11(-5)

REFERENCES

1. J.K. Shultis, D.E. Johnson, G.A. Milliken, and N.D. Eckhoff, "GAMMA: A Code for the Analysis of Component Failure Rates with a Compound Poisson-Gamma Model," NUREG/CR-2373, U.S. Nuclear Regulatory Commision, Washington (1981).

2. J.K. Shultis, D.E. Johnson, G.A. Milliken, and N.D. Eckhoff, "Use of Non-Conjugate Prior Distributions in Compound Failure Models," NUREG/CR-2374, U.S. Nuclear Regulatory Commission, Washington (1981).

3. H. Cramer, Mathematical Methods of Statistics, Princeton Univ. Press, Princeton (1974).

4. T.S. Ferguson, "On the Rejection of Outliers," Procedings of the Fourth Berkely Symposium on Mathematical Statistics and Probability, Vol. 1, pp. 253-287 (1961).

5. D.R. Gallup, "Detection of Outliers in Failure Data," MS Thesis, Kansas State University, Manhattan, Kansas (1981).

6. R.A.Fisher, "Test of Significance in Harmonic Analysis," Proceedings of the Royal Society A, Vol. 124, pp. 54-59 (1929).

7. N.R.J. Fieller, "Some Problems Related to the Rejection of Outlying Observations," Ph.D. Thesis, University of Sheffield (1976).

8. V. Barnett and T. Lewis, Outliers in Statistical Data, John Wiley & Sons, New York (1978).

9. A.C. Kimber, "Tests for a Single Outlier in a Gamma Sample with Unknown Shape and Scale Parameters," Applied Statistics, Vol. 28, No. 3, pp. 243-250 (1979).

10. F.E. Grubbs, "Sample Criteria for Testing Outlying Observations," Annals of Mathematical Statistics, Vol. 21, pp. 27-58 (1950).

11. W.J. Dixon, "Analysis of Extrema Values," Annals of Mathematical Statistics, Vol. 21, pp. 488-506 (1950).

12. J. Likes, "Distribution of Dixon's Statistics in the Case of an Exponential Population," Metrika, Vol. 11, pp. 46-54 (1966).

13. J.K. Shultis, W. Buranapan, and N.D. Eckhoff, "Properties of Parameter Estimation Techniques for a Beta-Binomial Failure Model," NUREG/CR-2372, U.S. Nuclear Regulatory Commission, Washington (1981).

14. J.K. Shultis, N.D. Eckhoff, D.E. Johnson and G.A. Milliken, "SAFE-R and SAFE-D: Computer Codes for the Analysis of Failure Data," NUREG/CR-2375, U.S. Nuclear Regulatory Commision, Washington (1981).

NRC FORM 335 (7.77) U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

1. REPORT NUMBER (Assigned by DDC): NUREG/CR-2714

4. TITLE AND SUBTITLE (Add Volume No., if appropriate): Analysis of Failure Rate Data with Compound Models

2. (Leave blank)

3. RECIPIENT'S ACCESSION NO.

7. AUTHOR(S): J.K. Shultis, N.D. Eckhoff, D.E. Johnson, G.A. Milliken, D.D. Drayer, D.R. Gallup, and W. Buranapan

5. DATE REPORT COMPLETED: MONTH February, YEAR 1982

9. PERFORMING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code):
Kansas State University
Dept. of Nulcear Engineering
Manhattan, Kansas 66506

DATE REPORT ISSUED: MONTH, YEAR

6. (Leave blank)

8. (Leave blank)

12. SPONSORING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code)

10. PROJECT/TASK/WORK UNIT NO.

11. CONTRACT NO.

13. TYPE OF REPORT: Final - Technical

PERIOD COVERED (Inclusive dates)

15. SUPPLEMENTARY NOTES

14. (Leave blank)

16. ABSTRACT (200 words or less)

This report summarizes three studies into various applications of statistical models (both homogeneous and compound) for describing observed failure-rate data. In the first study, the likelihood ratio test is used to examine how different two prior gamma distributions must be for a difference to be detected. From simulated failure data, three prior parameter estimators were examined: (i) the prior matching moments method (PMMM), (ii) the marginal matching moments method (MMMM), and (iii) the marginal maximum likelihood method (MMLM).

In the second phase of this study, power curves for several tests for data outliers were generated from simulated failure data characteristic of components with inherently low failure probabilities. Three types of failure data were examined: (i) failure-rate data (λ_i), (ii) data consisting of the number of failures in a specified test time (F_i,T_i), and (iii) time-to-failure data (t_i).

Finally, an investigation into the use of non-conjugate compound-Poisson failure models is presented. For the three prior parameter estimators (PMMM, MMMM, and MMLM), four different prior families (gamma, Weibull, lognormal, and logbeta) were used to analyze simulated failure data. It is found that for data from components with low failure probabilities and of small sample size (<20) the compound model is rather insensitive to the choice of the prior family, but is strongly dependent on the method used to estimate the prior distribution from the failure data.

17. KEY WORDS AND DOCUMENT ANALYSIS

17a. DESCRIPTORS

Compound (Bayesian) Models
Outliers
Non-Conjugate Compound Models
Power curves

17b. IDENTIFIERS/OPEN-ENDED TERMS

18. AVAILABILITY STATEMENT

19. SECURITY CLASS (This report)

21. NO. OF PAGES

20. SECURITY CLASS (This page)

22. PRICE

NRC FORM 335 (7.77)

EG&G Idaho, Inc.
P.O. Box 1625
Idaho Falls, Idaho 83415