Summer 1989

Software Reliability Models

Syed Afzal Hossain
Old Dominion University

SOFTWARE RELIABILITY MODELS

by

Syed Afzal Hossain
B.S., 1977, University of Dhaka, Bangladesh
M. Sc. 1979, University of Dhaka, Bangladesh

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY
Statistics

OLD DOMINION UNIVERSITY

August 1989

Approved by:

Ram C. Dahiya (Director)

ABSTRACT

SOFTWARE RELIABILITY MODELS

Syed Afzal Hossain
Old Dominion University, 1989
Director: Dr. Ram Chandra Dahiya

The problem considered here is the building of Non-homogeneous Poisson Process (NHPP) model. Currently existing popular NHPP process models like Goel-Okumoto (G-O) and Yamada *et al* models suffer from the drawback that the probability density function of the inter-failure times is an improper density function. This is because the event no failure in $(0, \infty]$ is allowed in these models. In real life situations we cannot draw sample(s) from such a population and also none of the moments of inter-failure times exist. Therefore, these models are unsuitable for modelling real software error data. On the other hand if the density function of the inter-failure times is made proper by multiplying with a constant, then we cannot assume finite number of expected faults in the system which is the basic assumption in building the software reliability models.

Taking these factors into consideration, we have introduced an extra parameter, say c, in both the G-O and Yamada *et al* models in order to get a new model. We find that a specific value of this new parameter gives rise to a proper density for inter-failure times. The G-O and Yamada *et al* models

are special cases of these models corresponding to $c = 0$. This raises the question - "Can we do better than existing G-O and Yamada *et al* models when $0 < c < 1$?". The answer is 'yes'.

With this objective, the behavior of the software failure counting process $\{N(t), t > 0\}$ has been studied. Several measures, such as the number of failures by some prespecified time, the number of errors remaining in the system at a future time, distribution of remaining number of faults in the system and reliability during a mission have been proposed in this research. Maximum likelihood estimation method was used to estimate the parameters. Sufficient conditions for the existence of roots of the ML equations were derived. Some of the important statistical aspects of G-O and Yamada *et al* models, like conditions for the existence and uniqueness of the ML equations, were not worked out so far in the literature. We have derived these conditions and proved uniqueness of the roots for these models. Finally four different sets of actual failure time data were analyzed.

ACKNOWLEDGEMENTS

My sincere thanks to Professor Ram Chandra Dahiya for being my advisor in this research. I feel, he put the right man (myself) at the right time in the right field of research. Besides that, his patient advice and encouragement helped me to complete this work. I would also like to thank Professor John Swetits, Professor N. R. Chaganty and Professor Larry Lee for their valuable assistance.

I am also ever grateful to my alma mater Dhaka University, Bangladesh, where I got the foundation of my present academic career.

My heartfelt thanks to my grand-father and mother for their help, inspiration and prayer which raised me to this height of academic scholarship and achievement. Last but not the least, is my life-partner Suraiya's constant help and my son Imran's presence to this world acted as a catalyst in completing this thesis. I share my achievement with them.

to

my mother

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

1. INTRODUCTION

The increasing pace of change and complexity in computing technology has necessitated a greater emphasis on the development of cost-effective and reliable software. Reliability is probably the most important of the characteristics inherent in the concept "software quality". It concerns itself with how well the software, functions to meet the requirements of the customers. The importance of software has been further enhanced by the fact that the ratio of software to hardware cost continues to grow as technological advances keep reducing the hardware cost. In short, software reliability measurements can be used to guide managerial and engineering decisions on projects involving software. They can also guide customers and users of systems that have software components in purchasing and operating these systems. They can help focus software engineering research and its application by determining those methods that are most effective in enhancing reliability.

The software quality is dependent on the tools and techniques used during its development and operation. An important performance criterion is the nature and frequency of software failures. A failure is said to occur when a fault, a specific manifestation of an error, in the program is evoked by some input data resulting in the computer program not correctly computing the required function. It is intimately connected with the failure of the system and thus detecting and correcting the failures is the prime objective of software reliability study. There are in general two ways of characterizing failure occurrences in time:

1. inter-failure time, and
2. cumulative failures experienced up to a given time (grouped data).

Software reliability models first appeared in the literature almost one and half decades ago and according to a recent survey (cf. Dale (1982)) some 40 models have been developed. The first study of software reliability appears to have been conducted by Hudson (1967). He viewed software development as a birth and death process. The next major steps were taken by Jelinski and Moranda (JM) (1972) and Shooman (1972). Both these studies assumed a failure rate that was piecewise constant and proportional to the number of faults remaining. The failure rate changes at each fault correction by a constant amount. Shooman characterized the failure rate in terms of inherent fault density of the program, the proportion of unique instructions executed, a bulk constant and the faults corrected per instruction per unit time. The bulk constant represented the proportion of faults that cause failures.

Another early model was proposed by Schick and Wolverton (1973). The failure rate assumed was proportional to the product of the number of remaining faults and the time. They also proposed a modified version of it with the failure rate as a parabolic function instead of a linear function in time.

Shortly after some of the early work, Musa (1975) presented an execution time model for software reliability. He postulated that execution time, the actual processor time utilized in executing the program, was the best practical measure of the failure-inducing stress that was being placed on the program. Musa also had observed that when rates were taken with respect to execution time, the fault correction rate was generally proportional to the fault detection or failure rate.

A Bayesian approach to software reliability measurement was used by Littlewood and Verrall (1973). Most of the models assume that the failure rate is a function of the number of faults remaining in the system. But Littlewood and Verrall modeled it as a random variable. One of the parameters of the distribution of this random variable is assumed to vary with the number of failures encountered. It thus characterizes reliability change. They proposed various functional forms for the description of this variation. The values of the parameters of each functional form that produce the best fit for that form are determined. Then the functional forms are compared and the best fitting form is selected.

The differential fault model proposed by Littlewood (1981) may be viewed as a variant of the general Littlewood-Verrall model. It is similar in viewing the failure rate as a random variable and in using Bayesian inference.

Goel and Okumoto (1978) developed a modification of the JM model for the case of imperfect debugging. It is based on the conception of debugging as a Markov process, with the appropriate transition probabilities between states. Goel and Okumoto (1979) also described failure detection as a non-homogeneous Poisson process (NHPP) with an exponential decay rate function. The cumulative number of failures detected and the distribution of the number of remaining failures are both found to be Poisson. A modification of the NHPP model was investigated by Yamada, Ohba, and Osaki (1983), where the cumulative number of failures detected is described as an S-shaped curve.

Musa and Okumoto (1983) developed the logarithmic Poisson execution time model which combines simplicity with the high predictive validity. This model is based on a non-homogeneous Poisson process with an intensity function that decreases exponentially with the expected failures experienced.

In the next chapter we review some important basic models. In the third chapter we will investigate some inference problems of Goel-Okumoto (G-O), and Yamada *et al* models. In the first two sections of the third chapter we have proved the necessary and sufficient condition for the existence of solution of the Maximum Likelihood (ML) equations for the G-O model, both for inter-failure and grouped data. The rest of the sections deal with the Yamada *et al* model. We have the necessary and sufficient condition for the existence of the solution of the ML equations obtained from the Yamada *et al* model for inter-failure data and only the sufficient condition for that of grouped data. These conditions provide us with the prior knowledge of whether the solutions of the ML equations will or will not exist. In all these cases, except the last one, we have proved that the solutions, if exist, are unique.

The fourth chapter considers some new models which are modified G-O model and modified Yamada *et al* models. In the first few sections of this chapter we have shown the different performance measures of the modified G-O model. Also, we have derived the method of obtaining the density function for both the cases of inter-failure and grouped data. The next section deals with the estimation of parameters and therein we determine the sufficient condition for the existence of the roots of the ML equations. The rest of the sections are devoted on our suggested modified Yamada *et al* model. The last chapter contains some examples of real life applications of the new models and their comparison with other similar models.

2. REVIEW

In this chapter, we review the literature dealing with software reliability models for estimating the number of bugs in a system. First section reviews some important models of various nature and Section 2.2 deals with the Non-Homogeneous Poisson Process (NHPP) models.

2.1 SOME BASIC MODELS ON SOFTWARE RELIABILITY

(I) JELINSKI-MORANDA MODEL

One of the earliest and probably the most commonly used model for assessing software reliability is the model given by Jelinski and Moranda (1972). It is based on the following assumptions.

1. The times between successive failures of the program, $T_1, T_2, \ldots, T_n$ are independent and exponentially distributed random variables, where n is the number of failures observed.

3. Each time a failure occurs, the bug that caused the failure is removed immediately by the programmer.

4. Each bug contributes the same amount, ϕ, to the overall failure rate of the program.

It asserts that the software failure rate, or the hazard function, at any time is proportional to the current fault content of the program. In other words, the hazard function during t_k, the time between the $(k-1)$ and k failures, is given by

$$\lambda_k = \phi[N - (k-1)], \tag{2.1}$$

where ϕ is a proportionality constant.

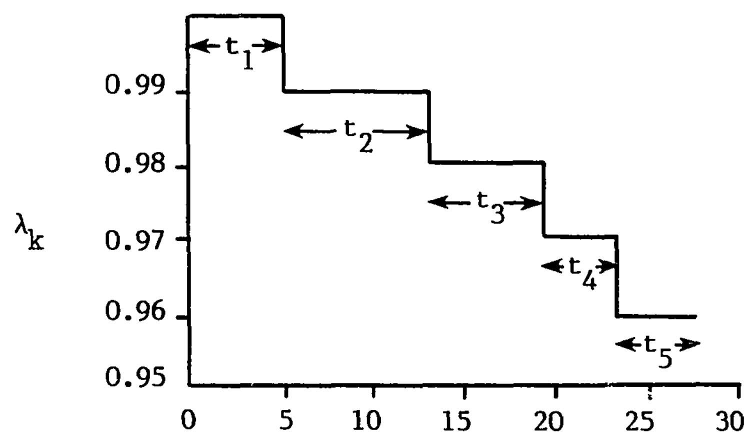

Fig 2.1 A typical plot of λ_k for the JM model ($N = 100, \phi = 0.01$).

Note that this failure rate function is constant between failures but decreases in steps of size ϕ following the removal of each fault. A typical plot of the hazard function, for $N = 100$ and $\phi = 0.01$ is shown in the Fig 1.

The likelihood function, given n observed failures and interfailure times $t_1, t_2, \ldots, t_n$, is

$$L(N, \phi) = \prod_{k=1}^{n} (N - k + 1)\phi e^{-(N-k+1)\phi t_k}, \tag{2.2}$$

where N represents the initial errors.

The ML equations obtained, for N and ϕ from the likelihood function does not have any closed form solution but can be solved by using numerical methods. Littlewood and Verrall (1981) proved that the necessary and sufficient condition for the existence of a finite solution for the ML equations is given by

$$\frac{\sum_{k=1}^{n}(k-1)t_k}{\sum_{k=1}^{n}(k-1)} > \frac{\sum_{k=1}^{n} t_k}{n}. \tag{2.3}$$

For the case of the failure process being observed until a fixed time t_0 and the number of observed failures is n, the likelihood function is given by:

$$L(N,\phi) = \phi^n e^{-\phi t_0 (x+N-n)} \prod_{k=1}^{n} (N-k+1), \tag{2.4}$$

$$\text{where,} \quad x = \frac{1}{t_0}\sum_{k=1}^{n} s_k, \quad s_k = \sum_{k=1}^{k} t_i \quad \text{and} \quad s_n < t_0$$

This likelihood function was studied by Joe and Reid (1985a, 1985b) and the following results were obtained:

The ML estimator, $\hat{N}$, of N is an increasing function of x. For each n, there is an increasing sequence of cut points:

$$m_{k,n} = [1 - (\frac{k-n}{k})^{\frac{1}{n}}]^{-1} - k + n, \qquad k > n, \tag{2.5}$$

$$\text{such that} \qquad \hat{N} = k, \qquad \text{if} \quad m_{k,n} \leq x < m_{k+1,n}, \tag{2.6}$$

$$\text{and} \qquad \hat{N} \to \infty, \quad \text{if} \quad x \geq \frac{n+1}{2}. \tag{2.7}$$

The ML estimate is unique unless $x = m_{k,n}$, for some k. The probability distribution of $\hat{N}$ is given below

$$P_r(\hat{N}=0) = e^{-\phi N},$$

$$P_r(\hat{N}=1) = N(1-e^{-\phi})e^{-\phi(N-1)},$$

$$P_r(\hat{N}=k) = \sum_{n=2}^{min(k,N)} \binom{N}{n}(1-e^{-\phi})^n e^{-\phi(N-n)} \times$$

$$[F_n(m_{k+1,n}) - F_n(m_{k,n})], \qquad k = 2,3,\ldots,$$

$$P_r(\hat{N}\to\infty) = \sum_{n=2}^{N}\binom{N}{n}(1-e^{-\phi})^n e^{-\phi(N-n)}\left[1 - F_n(\frac{n+1}{2})\right],$$

$$F_n(x) \equiv (1-e^{-\phi})^{-n}\sum_{j=0}^{n-1}\binom{n}{j}(-1)^j I(x>j)\times$$

$$\left[e^{-\phi j} - e^{-\phi x}\sum_{i=0}^{n-1}\frac{\phi^i(x-j)^i}{i!}\right], \tag{2.8}$$

$$\text{where,} \qquad I(u) = \begin{cases} 1, & \text{if } u \text{ is true,} \\ 0, & \text{otherwise.} \end{cases} \tag{2.9}$$

(II) LITTLEWOOD-VERRALL MODEL

Littlewood and Verrall (1973) took a different approach to the development of a model for times between failures. They argued that software reliability should not be specified in terms of the number of errors in the program. In their model, the times between failures are assumed to follow an exponential distribution but the parameter of this distribution is treated as a random variable with a gamma distribution, viz.

$$f(t_k|\lambda_k) = \lambda_k e^{-\lambda_k t_k}, \tag{2.10}$$

$$g(\lambda_k|\alpha, \psi(k)) = \frac{[\psi(k)]^k [\lambda_k]^{\alpha-1} e^{-\psi(k)\lambda_k}}{\Gamma(\alpha)}, \tag{2.11}$$

where the parameter $\psi(k)$ describes the ability of the programmer to determine the functional form that will give the best fit.

The unconditional distribution of T_k is

$$\begin{aligned} f(t_k) &= \int_0^\infty f(t_k|\lambda_k) g(\lambda_k|\alpha, \psi(k))\, d\lambda_k \\ &= \alpha \left[\frac{\psi(k)}{t_k + \psi(k)} \right]^\alpha \frac{1}{t_k + \psi(k)}, \end{aligned} \tag{2.12}$$

which is a Pareto distribution. The mean time to failure between the

$(k-1)$th and kth failures, and the failure rate are given by $\frac{\psi(k)}{\alpha}$ and $\frac{1}{t_k + \psi(k)}$, respectively.

In practice, $\psi(k)$ is assumed to have come from some parametric family, for example

$$\psi(k) = \beta_1 + \beta_2 k \tag{2.13}$$

and then the problem reduces to one of inference concerning the three parameters α, β_1 and β_2, which completely specify the model. This can be solved using maximum likelihood methods. If there is no apriori reason for assuming the linear growth function (2.13), then some means of choosing the best among several parametric families is required. Assuming that enough data is available to solve these inference problems, the model can be used in various ways. Most importantly, it is possible to obtain the pdf's of future inter-failure time. From these, the usual reliability measures such as failure rate, mean time to failure, reliability function and percentiles of time-to-failure distribution can be obtained.

(III) SCHICK-WOLVERTON (S-W) MODEL

This model is based on the same assumptions as the JM model except that the hazard function is assumed to be proportional to the current fault content of the program as well as the time elapsed since the last failure and is given by

$$\lambda(t_k) = \phi(N - k + 1)t_k, \tag{2.13}$$

where, N =total number of initial errors,

ϕ =constant of proportionality which keeps the area under the probability curve equal to unity,

t_k =the kth time debugging interval; i.e., the time between the kth and (k-1)th errors discovered.

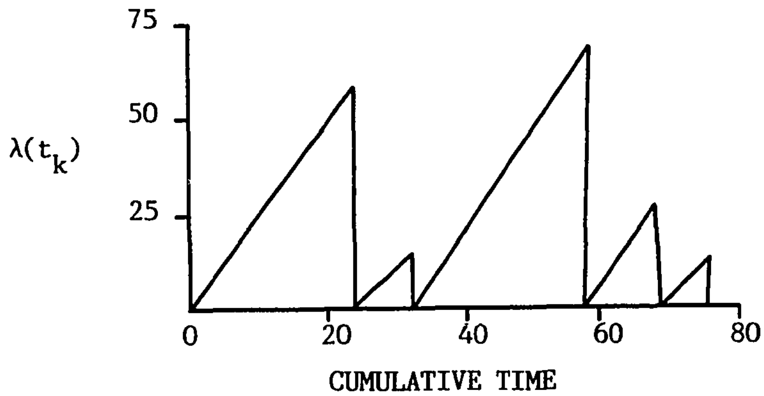

Fig 2.2 A typical plot of $\lambda(t_k)$ for the SW model ($N = 150, \phi = 0.02$)

We note that the above hazard rate is linear with time within each failure interval, returns to zero at the occurence of a failure, and increases linearly again but at a reduced slope, the decrease in slope being proportional to ϕ. A typical behavior of $\lambda(t_k)$ for $N = 150$ and $\phi = 0.02$ is shown in Fig.2.

The reliability function derived for this case is the well-known Rayleigh distribution given by

$$R(t_k) = e^{-\phi[N-(k-1)]\frac{t_k^2}{2}}. \tag{2.14}$$

The equations to estimate the total time, T required to find all remaining errors and standard deviation, σ for this estimate are

$$\hat{T} = \frac{1}{\phi}\left[\sum_{k=1}^{n}\frac{1}{k}\right], \tag{2.15}$$

and

$$\hat{\sigma} = \frac{1}{\phi}\left[\sum_{k=1}^{n}\frac{1}{k^2}\right]^{\frac{1}{2}}. \tag{2.16}$$

A modification of the above model was proposed by Schick and Wolverton (1978), whereby the failure function is assumed to be parabolic in test time and is given by

$$\lambda(t_k) = \phi[N-(k-1)](-at_k^2 + bt_k + c), \tag{2.17}$$

where a, b and c are constants and the other quantities are as defined above. This function consists of two factors. The first is basically the hazard function of the JM model and the second factor indicates that the likelihood of a failure occurring increases rapidly as the test time accumulates within a testing interval. At failure times $t_k = 0$, the failure rate function is proportional to that of the JM model.

(IV) SHOOMAN EXPONENTIAL MODEL

Shooman (1972) characterized the hazard rate in terms of the inherent fault density of the program ω_1, the proportion of unique instructions processed β_1, a bulk constant β_2, and the faults corrected per instruction per unit time β_3. The bulk constant represents the proportion of faults that cause failures. For this model the hazard function is of the form

$$\lambda(t) = \left[\omega_1 - \int_0^t \beta_3(x)\, dx\right] \beta_1 \beta_2. \tag{2.18}$$

This model is similar to the Jelinski-Moranda model.

2.2 NHPP MODELS ON SOFTWARE RELIABILITY

The Poisson process simply refers to the probability distribution of the value of the process at each point in time. The term non-homogeneous indicates that the characteristics of the probability distributions that make up the random process vary with time. In this section we discuss a non-homogeneous Poisson process as a stochastic model for the software failure phenomenon. Similar models have also been used to describe hardware reliability growth. The following definitions characterizing a software reliability growth aspect in software testing should be introduced before we discuss NHPP models.

1. The mean value function $m(t)$ is an increasing error detection rate (IEDR) function if the error detection rate per error, $d(t)$, is non- decreasing in t, $t \geq 0$.

2. The mean value function $m(t)$ is decreasing error detection rate (DEDR) function if $d(t)$ is non-increasing in t, $t \geq 0$.

3. The mean value function $m(t)$ is a constant error detection rate (CEDR) function if $d(t)$ is constant in t, $t \geq 0$.

(I) GOEL-OKUMOTO MODEL

Goel and Okumoto (1979) proposed a NHPP model which describes a software failure detection phenomenon. The mean value function showing an exponential growth curve is given by

$$m(t) = a[1 - e^{-bt}], \qquad a > 0 \quad \text{and} \quad b > 0, \tag{2.19}$$

where a represents the expected number of errors to be eventually detected and b represents the error detection rate per error at an arbitrary testing time. It is clear that $m(t)$ is a CEDR function since the error detection rate per error, say $d(t)$ is given by

$$d(t) = b, \qquad t \geq 0.$$

The intensity function also known as the error detection rate, $\lambda(t)$ is given by

$$\lambda(t) = abe^{-bt}, \tag{2.21}$$

which is clearly a decreasing function of testing time t. For $t \geq 0$ the total number of errors occuring in $(0, t)$, say $N(t)$, has a Poisson distribution with expected value $m(t)$, i.e.

$$P\{N(t) = y\} = \frac{[m(t)]^y}{y!}e^{-m(t)}, \qquad y = 0, 1, 2, \ldots; \tag{2.22}$$

Thus, the stochastic behavior of software failure phenomena is described through the $N(t)$ process.

Let S_k denote the time to k failures. Then the joint density function of $S_1, S_2, \ldots, S_n$ is given by

$$f_{S_1,\ldots,S_n}(s_1,\ldots,s_n) = e^{-a(1-e^{-bs_n})} \prod_{k=1}^{n} abe^{-bs_k}, \quad a > 0 \quad \text{and} \quad b > 0. \tag{2.23}$$

The ML equations are given by

$$\frac{n}{a} = 1 - e^{-bs_n}, \tag{2.24}$$

$$\frac{n}{b} = \sum_{k=1}^{n} s_k + as_n e^{-bs_n}, \tag{2.25}$$

which can be solved numerically.

For the case of grouped data, let $y_1, y_2, \ldots, y_n$ denote the cumulative number of failures detected by times $t_1, t_2, \ldots t_n$ respectively. Then the joint density function of the observed values is

$$P_r[N(t_1) = y_1, \ldots, N(t_n) = y_n] = e^{-m(s_n)} \prod_{k=1}^{n} \frac{[a(e^{-bt_{k-1}} - e^{-bt_k})]^{y_k - y_{k-1}}}{(y_k - y_{k-1})!}. \tag{2.26}$$

The ML equations are given by

$$\frac{1}{a}\sum_{k=1}^{n}(y_k - y_{k-1}) - (1 - e^{-bt_n}) = 0, \qquad (2.27)$$

$$-as_n e^{-bs_n} - \sum_{k=1}^{n}(y_k - y_{k-1})\left[\frac{t_{k-1}e^{-bt_{k-1}} - t_k e^{-bt_k}}{e^{-bt_{k-1}} - e^{-bt_k}}\right] = 0, \qquad (2.28)$$

which can be solved by using numerical methods.

This model has parameters with a physical interpretation and can yield quantitative measures for software performance assessment. Also of interest is the applicability of the model over a broad class of projects and the estimability of parameters when the available data is in the form of times between errors or as number of errors in given time intervals.

(II) YAMADA AND OSAKI MODEL

Yamada and Osaki (1985) proposed a non-homogeneous error detection rate model on the assumption that there exist two types of errors: Type 1 errors which are easy to detect and Type 2 errors which are difficult to detect. This NHPP model, called the modified exponential software reliability growth model (SRGM) , has a mean value function of

$$m(t) = a \sum_{k=1}^{2} p_k (1 - e^{-b_k t}),$$

$$\text{where,} \quad 0 < b_2 < b_1 < 1,$$

$$\text{and,} \quad \sum_{k=1}^{2} p_k = 1, \quad 0 < p_k < 1, \qquad (k = 1, 2).$$

Here

b_k is the error detection rate per Type k error (k=1,2), and

p_k is the content proportion of Type k errors, i.e., $p_k a$ is the expected initial error content of Type k errors (k=1,2).

The error detection rate per error at testing time t is given by

$$d(t) = \sum_{k=1}^{2} \left[\frac{p_k e^{-b_k t}}{p_1 e^{-b_1 t} + p_2 e^{-b_2 t}} \right] b_k.$$

It can be shown that $m(t)$ is a DEDR function.

(III) MUSA-OKUMOTO LOGARITHMIC POISSON EXECUTION TIME MODEL

Musa and Okumoto (1984), introduced a model where the observed number of failures by some time t is assumed to be a NHPP, similar to the Goel-Okumoto model, but with a different mean value function given by

$$m(t) = \frac{1}{\theta}\log(\lambda_0\theta t + 1), \tag{2.31}$$

where, λ_0 and θ represent the initial failure intensity and the rate of reduction in the normalized failure intensity per failure, respectively. This model is also closely related to Moranda's geometric de-eutrophnication model and can be viewed as a continuous version of this model. It has an intensity function that decreases exponentially with expected failures experienced and is given by

$$\lambda_t = \lambda_0 e^{-\theta m(t)}. \tag{2.32}$$

The conditional reliability function and the mean time to failure function are given by

$$R(t_k|t_{k-1}) = \left[\frac{\lambda_0\theta t_{k-1}+1}{\lambda_0\theta(t_{k-1}+t_k)+1}\right]^{\frac{1}{\theta}} \tag{2.33}$$

and

$$MTTF(t_{k-1}) = \frac{\theta}{1-\theta}(\lambda_0\theta t_{k-1}+1)^{1-\frac{1}{\theta}}, \tag{2.34}$$

respectively.

This model has a calendar time component and the total expected number of failures is infinite. It is very likely that the number of inherent faults in a program is finite. The model should be able to accommodate simultaneously an infinite number of failures and a finite number of faults and this was done by assuming a time-varying fault reduction factor of a specific form. The problem of parameter prediction for the logarithmic Poisson model will probably be more difficult than that for the basic model. As a result, the basic model is likely to be superior for initial, approximate determination of behavior. It should be pointed out that the parameter θ may be related to the efficiencies of a testing method and a repair activity; a larger value of θ implies a higher efficiency since the failure intensity is reduced at a faster rate.

(IV) YAMADA *et al* MODEL

In a software error removal phenomenon it should be assumed that a testing process consists of not only a software failure detection process, but also a software error isolation process. Yamada, Ohba and Osaki (1983) offered the ***delayed S-shaped Software Reliability Growth Modeling*** for such an error detection process in which the observed growth curve of the cumulative number of detected errors is S-shaped. This NHPP model has a mean value function given by

$$m(t) = a[1 - (1 + bt)e^{-bt}], \qquad b > 0, \tag{2.35}$$

which is a S-shaped growth curve. The parameter b represents the failure detection rate (and the error isolation rate). It can be shown that $m(t)$ is an IEDR function since the error detection rate per error, given by

$$d(t) = \frac{b^2 t}{1 + bt}$$

is monotonically increasing in testing time t.

The intensity function or the error detection rate, $\lambda(t)$, is given by

$$\lambda(t) = ab^2te^{-bt}. \tag{2.37}$$

Clearly, $\lambda(0) = 0$ and $\lambda(\infty) = 0$.

For $t \geq 0, N(t)$ has a Poisson distribution with expected value $m(t)$, i.e.

$$P\{N(t) = y\} = \frac{[m(t)]^y}{y!} e^{-m(t)}, \qquad y = 0, 1, \ldots \tag{2.38}$$

Thus, the stochastic behavior of software failure phenomena is described through the $N(t)$ process. Equations (2.35) and (2.38) constitute the basic model.

The joint density function of $S_1, \ldots, S_n$ is given by

$$f_{S_1,\ldots,S_n}(s_1, \ldots, s_n) = a^n b^{2n} e^{-a[1-(1+bs_n)e^{-bs_n}]} \prod_{k=1}^{n} s_k e^{-bs_k}, \quad \text{where} \quad a, b > 0. \tag{2.39}$$

The ML equations are obtained to be

$$\frac{n}{a} = 1 - (1 + bs_n)e^{-bs_n}, \tag{2.40}$$

$$\frac{2n}{b} = \sum_{k=1}^{n} s_k + abs_n^2 e^{-bs_n}, \tag{2.41}$$

which can be solved numerically.

In the case of grouped data, the joint density function is given by

$$P_r[N(t_1) = y_1, \ldots, N(t_n) = y_n] = e^{-a[1-(1+bt_n)e^{-bt_n}]} \times$$

$$\prod_{k=1}^{n} \frac{[a\{(1+bt_{k-1})e^{-bt_{k-1}} - (1+bt_k)e^{-bt_k})\}]^{y_k - y_{k-1}}}{(y_k - y_{k-1})!}, \quad (2.42)$$

where $\qquad a, b > 0.$

The ML equations for this case are given by

$$y_n = a[1 - (1+bt_n)e^{-bt_n}] \qquad \text{and} \qquad (2.43)$$

$$abt_n^2 e^{-bt_n} = -\sum_{k=1}^{n}(y_k - y_{k-1}) \frac{b(t_k^2 e^{-bt_k} - t_{k-1}^2 e^{-bt_{k-1}})}{(1+bt_{k-1})e^{-bt_{k-1}} - (1+bt_k)e^{-bt_k}} (2.44)$$

which can also be solved by using numerical methods.

(V) OHBA MODEL

The Ohba model describes a softrware failure detection phenomenon with a mutual dependence of detected errors. In the error detection process, the more failures we detect, the more undetected failures become detectable. This NHPP model has a mean value function of

$$m(t) = a\frac{1 - e^{-bt}}{1 + ce^{-bt}}, \tag{2.45}$$

where, $a > 0, \quad b > 0, \quad$ and $\quad c > 0,$

which shows an S-shaped growth curve. The parameters b and c represent the failure detection rate and the inflection factor, respectively. It can be shown that $m(t)$ is an IEDR function since

$$d(t) = \frac{b}{1 + ce^{-bt}} \tag{2.46}$$

is monotonically increasing in testing time t.

In fact, in Japan, some computer manufacturers and software houses use the logistic and Gompertz growth curve models. Those curves were originally developed to predict demand trend, economic growth, or future population. The expected cumulative number of errors detected up to testing time t for the logistic growth curve model is given by

$$m(t) = \frac{k}{(1 + me^{-pt})}, \qquad \text{where} \qquad m > 0, p > 0, k > 0, \tag{2.47}$$

and for the Gompertz growth curve model it is given by

$$m(t) = ka^{(b^t)}, \qquad \text{where} \qquad 0 < a < 1, 0 < b < 1, k > 0. \qquad (2.48)$$

The parameters are to be estimated by regression analysis. The parameter k in both models is the expected initial error content of a software system.

3. GOEL-OKUMOTO and YAMADA *et al* MODELS

3.1 G-O MODEL DEVELOPMENT

Let $\{N(t), t \geq 0\}$ be a counting process representing the cumulative number of failures up to time t. It is reasonable to assume that for any finite collection of times $t_1 < t_2 < ... < t_n$, the n random variables $N(t_1), N(t_2) - N(t_1), ..., N(t_n) - N(t_{n-1})$ are independent, implying that the counting process $\{N(t), t \geq 0\}$ has independent increments.

Model Assumptions.

1. The number of errors in a software system at the start of the debugging process is a random variable.
2. Each time failure occcurs, the bug which caused the failure is immediately removed by the programmer.
3. The time between failures k-1 and k depends on the time to k-1 failures.

Let $m_1(t)$ represent the expected number of software failures upto time t. Then, $m_1(t)$ is a bounded and non-decreasing function of t with the following boundary conditions:

$$m_1(t) = \begin{cases} 0, & \text{for } t = 0, \\ a, & \text{for } t \to \infty, \end{cases} \tag{3.1}$$

where a is the expected number of software errors to be eventually detected. It is assumed that the error detection rate per error, say b, is constant. i.e.,

$$\frac{1}{a - m_1(t)} \frac{dm_1(t)}{dt} = b, \quad b > 0. \tag{3.2}$$

Solving differential equation (3.2) under the boundary condition (3.1) the solution for $m_1(t)$ is given by :

$$m_1(t) = a(1 - e^{-bt}). \tag{3.3}$$

The intensity function or the error detection rate, $\lambda_1(t)$, is obtained by differentiating equation (3.3) with respect to t, which yields

$$\lambda_1(t) = abe^{-bt}.$$

For $t \geq 0$, $N(t)$ has a Poisson distribution with expected value $m_1(t)$, i.e.

$$P\{N(t) = y\} = \frac{[m_1(t)]^y}{y!} e^{-m_1(t)}, \qquad y = 0, 1, 2, \ldots \tag{3.4}$$

Thus, the stochastic behavior of software failure phenomena is described through the $N(t)$ process. Equations (3.3) and (3.4) constitute the basic models.

3.2 ESTIMATION OF PARAMETERS OF G-O MODEL

Case of Interfailure Data.

Let X_k be the time between failures (k-1) and k, and S_k the time to k failures. Then the joint density function of $S_1, S_2, ..., S_n$ is given by

$$f_{S_1,...,S_n}(s_1,...,s_n) = e^{-m(s_n)} \prod_{k=1}^{n} abe^{-bs_k}, \qquad a > 0, \quad b > 0.$$

Given the failure times $s = (s_1, s_2, ..., s_n)$, the log likelihood function is given by

$$L_1(a, b|s) = n \log(a) + n \log(b) - a(1 - e^{-bs_n}) - b\sum_{k=1}^{n} s_k. \tag{3.5}$$

Differentiating equation (3.5) with respect to a and b separately and equating to zero, the maximum likelihood (ML) equations are found to be:

$$\frac{n}{a} = 1 - e^{-bs_n}, \tag{3.6}$$

$$\frac{n}{b} = \sum_{k=1}^{n} s_k + as_n e^{-bs_n}. \tag{3.7}$$

So far no results are available regarding the solution of these equations. We plan to find a necessary and sufficient condition for the existence of finite positive roots of these equations.

The following lemma is used to prove Theorem 1 and Lemma 2, concerning the solution of equations (3.6) and (3.7).

LEMMA 1. *The function*

$$g(b) = \frac{n}{b} - \sum_{k=1}^{n} s_k - \frac{ns_n}{e^{bs_n} - 1} \tag{3.8}$$

is a decreasing function of b.

PROOF: Differentiating $g(b)$ with respect to b and simplifying further it can be shown that

$$g'(b) = \frac{-ne^{2bs_n} + 2ne^{bs_n} - n + nb^2 s_n^2 e^{bs_n}}{b^2(e^{bs_n} - 1)^2}. \tag{3.9}$$

The numerator of $g'(b)$ is $w(b) = ne^{bs_n}(-e^{bs_n} + 2 - e^{-bs_n} + b^2 s_n^2)$

Let $x = bs_n$, then

$$w(b) = 2ne^x(1 + \frac{x^2}{2} - \frac{e^x + e^{-x}}{2})$$
$$= 2ne^x(-\frac{x^4}{4!} - \frac{x^6}{6!} - \quad \dots \quad) < 0 \tag{3.10}$$

Therefore, $g'(b) \leq 0$ and as such $g(b)$ is decreasing in b. □

THEOREM 1. *The necessary and sufficient condition for the equations (3.6) and (3.7) to have finite positive roots is*

$$s_n > \frac{2}{n}\sum_{k=1}^{n} s_k. \tag{3.11}$$

PROOF: From equation (3.6) we have $a = n(1 - e^{-bs_n})^{-1}$. Now a is finite and positive if $b > 0$. Substituting this value of a into the equation (3.7) we obtain the following equation involving b alone,

$$\frac{n}{b} - \sum_{k=1}^{n} s_k - \frac{ns_n}{e^{bs_n} - 1} = 0. \tag{3.12}$$

Now it suffices to show that (3.11) is the necessary and sufficient condition for the existence of a positive root of equation (3.12). Note that the left hand side of (3.12) is the function $g(b)$ of Lemma 1. Now, $\lim_{b\to 0} g(b)$ and $\lim_{b\to\infty} g(b)$ are given by

$$\begin{aligned}
\lim_{b\to 0} g(b) &= \lim_{b\to 0}\left[\frac{n}{b} - \sum_{k=1}^{n} s_k - \frac{ns_n}{e^{bs_n}-1}\right] \\
&= -\sum_{k=1}^{n} s_k + n\lim_{b\to 0}\left[\frac{e^{bs_n}-bs_n-1}{b[e^{bs_n}-1]}\right] \\
&= -\sum_{k=1}^{n} s_k + n\lim_{b\to 0}\left[\frac{s_n e^{bs_n}-s_n}{e^{bs_n}+bs_n e^{bs_n}-1}\right] \\
&= -\sum_{k=1}^{n} s_k + n\lim_{b\to 0}\left[\frac{s_n^2 e^{bs_n}}{2s_n e^{bs_n}+bs_n^2 e^{bs_n}}\right] \\
&= -\sum_{k=1}^{n} s_k + \frac{ns_n}{2} \qquad (3.13)
\end{aligned}$$

and

$$\lim_{b\to\infty} g(b) = -\sum_{k=1}^{n} s_k < 0. \qquad (3.14\ )$$

On using the fact that $g(b)$ is a decreasing function of b, as proved in Lemma 1, it follows that equation (3.12) has a positive root if and only if (3.11) is true. □

Now we prove the uniqueness of the MLE's in the following Lemma.

LEMMA 2. *There exist a unique positive constants $\hat{a}$ and $\hat{b}$ satisfying the ML equations (3.6) and (3.7) if and only if the condition (3.11) is satisfied.*

PROOF: Since $g(b)$ is decreasing for all $b > 0$, therefore, by Lemma 1, $g(b) = 0$ has unique root. By substituting this root into equation (3.6) we obtain the unique solution for a. $\square$

Case of Grouped Data.

Let $y_1, y_2, ..., y_n$ denote the cumulative number of failures detected by times $t_1, t_2, ...t_n$, respectively. Then the joint density function of the observed data is

$$P[N(t_1) = y_1, ..., N(t_n) = y_n] = \prod_{k=1}^{n} P_r[N(t_k) - N(t_{k-1}) = y_k - y_{k-1}]$$

$$= e^{-m(t_n)} \prod_{k=1}^{n} \frac{[m(t_k) - m(t_{k-1})]^{y_k - y_{k-1}}}{(y_k - y_{k-1})!}$$

$$= e^{-a(1-e^{-bt_n})} \prod_{k=1}^{n} \frac{[a(e^{-bt_{k-1}} - e^{-bt_k})]^{y_k - y_{k-1}}}{(y_k - y_{k-1})!}.$$

where $t_0 = 0$ and $y_0 = 0$.

The likelihood function for the parameters is simply the joint probability density of $y_1, y_2, ..., y_n$, given by

$$L_2(a, b|\tilde{y}, \tilde{t}) = e^{-a(1-e^{-bt_n})} \prod_{k=1}^{n} \frac{[a(e^{-bt_{k-1}} - e^{-bt_k})]^{y_k - y_{k-1}}}{(y_k - y_{k-1})!}. \qquad (3.15)$$

Taking the natural logarithm of equation (3.15) yields:

$$\log L(a,b|y,t) = y_n \log a + \sum_{k=1}^{n} (y_k - y_{k-1}) \log(e^{-bt_{k-1}} - e^{-bt_k}) - \sum_{k=1}^{n} (y_k - y_{k-1})! - a(1 - e^{-bt_n}). \tag{3.16}$$

The ML equations are given by:

$$\frac{1}{a}\sum_{k=1}^{n}(y_k - y_{k-1}) - (1 - e^{-bt_n}) = 0, \tag{3.17}$$

$$-\sum_{k=1}^{n} (y_k - y_{k-1}) \left[\frac{(t_{k-1}e^{-bt_{k-1}} - t_k e^{-bt_k})}{e^{-bt_{k-1}} - e^{-bt_k}}\right] - at_n e^{-bt_n} = 0. \tag{3.18}$$

The following lemma is used in proving Theorem 2 concerning the solution of equations (3.17) and (3.18).

LEMMA 3. *Let* $g_1(b) = c^2(e^{bd} + e^{-bd} - 2) + d^2(2 - e^{bc} - e^{-bc})$, *where* $b > 0$ *and* c *and* d *are constants satisfying* $0 < c < d$. *Then* $g_1(b)$ *is positive for all* $b > 0$.

PROOF: Differentiating $g_1(b)$ with respect to b yields:

$$\begin{aligned} g_1'(b) &= dc^2(e^{bd} - e^{-bd}) + cd^2(e^{-bc} - e^{bc}) \\ &= dc^2\left[(e^{bd} - e^{-bd}) + \frac{d}{c}(-e^{bc} + e^{-bc})\right]. \end{aligned}$$

Let $r = \frac{d}{c} > 1$ and $bc = a > 0$. Note that $\sinh(0) = 0$. Then,

$$\begin{aligned}
g_1'(b) &= dc^2 \left[(e^{ar} - e^{-ar}) - r(e^a - e^{-a})\right] \\
&= 2dc^2 \left[\sinh(ar) - r\sinh(a)\right] \\
&= 2rdc^2 \left[\frac{\sinh(ar)}{r} - \sinh(a)\right] \\
&= 2ardc^2 \left[\frac{\sinh(ar) - \sinh(0)}{ar} - \frac{\sinh(a) - \sinh(0)}{a}\right]. \qquad (3.19)
\end{aligned}$$

Let $g_2(y) = \sinh(y)$. Then in the equation (3.19), the first expression is the slope of the secant joining $(o, g_2(0))$ and $(ar, g_2(ar))$ and the second expression is the slope of the secant joining $(0, g_2(0))$ and $(a, g_2(a))$. Note that $0 < a < ar$. Again from the properties of sinh we know that it is a convex function. Therefore,

$$\frac{\sinh(ar) - \sinh(0)}{ar} - \frac{\sinh(a) - \sinh(0)}{a} > 0.$$

Hence, $g_1'(b) > 0$. Note again that $g_1(0) = 0$. Therefore, $g_1(b) > 0$ for all $b > 0$. $\square$

We now derive a necessary and sufficient condition for the MLE's to be finite and positive in the following theorem.

THEOREM 2. *The necessary and sufficient condition for the existence of finite positive root of the equations (3.17) and (3.18) is*

$$t_n y_n \geq \sum_{k=1}^{n} (y_k - y_{k-1})(t_k + t_{k-1}). \tag{3.20}$$

PROOF: From equation (3.17) we obtain $a = y_n(1 - e^{-bt_n})^{-1}$. Clearly, a is finite and positive if $0 < b < \infty$. Substituting this value of a into equation (3.18) we obtain the following equation in b:

$$0 = -\sum_{k=1}^{n} (y_k - y_{k-1}) \left[\frac{(t_{k-1}e^{-bt_{k-1}} - t_k e^{-bt_k})}{(e^{-bt_{k-1}} - e^{-bt_k})} \right] - \frac{t_n \sum_{k=1}^{n} (y_k - y_{k-1}) e^{-bt_n}}{1 - e^{-bt_n}}$$

$$= -\sum_{k=1}^{n} (y_k - y_{k-1}) \left[\frac{t_{k-1}e^{b(t_k - t_{k-1})} - t_k}{e^{b(t_k - t_{k-1})} - 1} + \frac{t_n}{e^{bt_n} - 1} \right]. \tag{3.21}$$

Therefore, we have to show that (3.20) is the necessary and sufficient condition for the existence of a positive root for equation (3.21).

Let $h_1(b)$ denote the right hand side function of the equation (3.21). i.e.,

$$h_1(b) = -\sum_{k=1}^{n} (y_k - y_{k-1}) \left[\frac{t_{k-1}e^{b(t_k - t_{k-1})} - t_k}{e^{b(t_k - t_{k-1})} - 1} + \frac{t_n}{e^{bt_n} - 1} \right]$$

$$= -\sum_{k=1}^{n} (y_k - y_{k-1}) \left[\frac{t_{k-1}e^{b(t_k - t_{k-1})} - t_k}{e^{b(t_k - t_{k-1})} - 1} \right] - \frac{t_n y_n}{e^{bt_n} - 1}. \tag{3.22}$$

To prove the sufficient condition let us first determine $\lim_{b\to 0} h_1(b)$ and $\lim_{b\to\infty} h_1(b)$.

$$\lim_{b\to 0} h_1(b) = -\lim_{b\to 0}\sum_{k=1}^{n}(y_k - y_{k-1})\left[\frac{t_{k-1}e^{b(t_k-t_{k-1})} - t_k}{e^{b(t_k-t_{k-1})}-1} + \frac{t_n}{e^{bt_n}-1}\right]$$

$$= -\sum_{k=1}^{n}(y_k - y_{k-1})t_{k-1} + \lim_{b\to 0}\sum_{k=1}^{n}(y_k - y_{k-1})\times$$

$$\left[\frac{(t_k - t_{k-1})}{(e^{b(t_k-t_{k-1})}-1)} - \frac{t_n}{(e^{bt_n}-1)}\right]$$

$$= -\sum_{k=1}^{n}(y_k - y_{k-1})t_{k-1} + \lim_{b\to 0}\sum_{k=1}^{n}(y_k - y_{k-1})\times$$

$$\left[\frac{(t_k - t_{k-1})e^{bt_n} - t_n e^{b(t_k-t_{k-1})} - t_k + t_{k-1} + t_n}{e^{b(t_n+t_k-t_{k-1})} - e^{b(t_k-t_{k-1})} - e^{bt_n} + 1}\right]$$

$$= -\sum_{k=1}^{n}(y_k - y_{k-1})t_{k-1} + \lim_{b\to 0}\sum_{k=1}^{n}(y_k - y_{k-1})\times$$

$$\left[\frac{t_n(t_k - t_{k-1})e^{bt_n} - t_n(t_k - t_{k-1})e^{b(t_k-t_{k-1})}}{(t_n + t_k - t_{k-1})e^{b(t_n+t_k-t_{k-1})} - (t_k - t_{k-1})e^{b(t_k-t_{k-1})} - t_n e^{bt_n}}\right]$$

$$= -\sum_{k=1}^{n}(y_k - y_{k-1})t_{k-1} + \lim_{b\to 0}\sum_{k=1}^{n}(y_k - y_{k-1})\times$$

$$\left[\frac{t_n^2(t_k - t_{k-1})e^{bt_n} - t_n(t_k - t_{k-1})^2 e^{b(t_k-t_{k-1})}}{(t_n + t_k - t_{k-1})^2 e^{b(t_n+t_k-t_{k-1})} - (t_k - t_{k-1})^2 e^{b(t_k-t_{k-1})} - t_n^2 e^{bt_n}}\right]$$

$$= -\sum_{k=1}^{n}(y_k - y_{k-1})t_{k-1} + \sum_{k=1}^{n}(y_k - y_{k-1})\times$$

$$\left[\frac{t_n^2(t_k - t_{k-1}) - t_n(t_k - t_{k-1})^2}{(t_n + t_k - t_{k-1})^2 - (t_k - t_{k-1})^2 - t_n^2}\right]$$

$$= -\sum_{k=1}^{n}(y_k - y_{k-1})t_{k-1} + \frac{1}{2}\sum_{k=1}^{n}(y_k - y_{k-1})(t_n - t_k + t_{k-1})$$

$$= \frac{1}{2}\left[t_n y_n - \sum_{k=1}^{n}(y_k - y_{k-1})(t_k + t_{k-1})\right]. \tag{3.23}$$

Again, it can easily be shown from step two of equation (3.23) that

$$\lim_{b\to\infty} h_1(b) = -\sum_{k=1}^{n}(y_k - y_{k-1})t_{k-1}, \tag{3.24}$$

which is a negative quantity. Therefore, the sufficient condition for the existence of positive root of the equation $h_1(b) = 0$ is:

$$\frac{1}{2}\left[t_n y_n - \sum_{k=1}^{n}(y_k - y_{k-1})(t_k + t_{k-1})\right] > 0$$

i.e.,

$$t_n y_n > \sum_{k=1}^{n}(y_k - y_{k-1})(t_k + t_{k-1}). \tag{3.25}$$

To prove that the condition (3.20) is necessary, let us prove that the function $h_1(b)$ is decreasing in b. Taking derivative of $h_1(b)$ with respect to b we obtain:

$$h_1'(b) = -\sum_{k=1}^{n}(y_k - y_{k-1})\left[\frac{(t_k - t_{k-1})^2 e^{b(t_k - t_{k-1})}}{[e^{b(t_k - t_{k-1})} - 1]^2} - \frac{t_n^2 e^{bt_n}}{(e^{bt_n} - 1)^2}\right]$$

For a fixed k, let $c = t_k - t_{k-1}$ and $d = t_n$. Then the expression within the square bracket can be expressed as:

$$\frac{c^2 e^{bc}}{(e^{bc} - 1)^2} - \frac{d^2 e^{bd}}{(e^{bd} - 1)^2}$$

$$= \frac{c^2 e^{b(c+2d)} - 2c^2 e^{b(c+d)} + c^2 e^{bc} - d^2 e^{b(2c+d)} + 2d^2 e^{b(c+d)} - d^2 e^{bd}}{(e^{bc} - 1)^2 (e^{bd} - 1)^2}$$

$$= \frac{e^{b(c+d)}}{(e^{bc} - 1)^2 (e^{bd} - 1)^2}\left[c^2(e^{bd} + e^{-bd} - 2) + d^2(e^{bc} + e^{-bc} - 2)\right]. \quad (3.26)$$

The expression in the square bracket of equation (3.26) is same as the function $g_1(b)$ considered in Lemma 2, with $g_1(0) = 0$. Therefore, by Lemma 2 $h_1(b)$ is a decreasing function in b.

Since $h_1(b)$ is a decreasing function of b and $\lim_{b\to\infty} h_1(b) < 0$, the equation (3.21) would have a positive root if and only if $\lim_{b\to 0} h_1(b) > 0$ which in turn implies that (3.20) must be true. □

On making use of the results given in Theorem 2, we obtain the following results regarding MLEs.

LEMMA 4. *If $\hat{a}$ and $\hat{b}$ are finite positive roots of the ML equations (3.17) and (3.18) then they are unique.*

PROOF: It has already been proved that the function $h_1(b)$ is decreasing for all $b > 0$. Therefore, if the solution exists for b it will be unique. Again, using equation (3.17) we obtain the unique solution for a. □

3.3 YAMADA *et al* MODEL DEVELOPMENT

As in Goel-Okumoto model, let $\{N(t), t \geq 0\}$ be a counting process representing the cumulative number of failures up to time t. For any finite collection of times $t_1 < t_2 < ... < t_n$, the n random variables $N(t_1), N(t_2) - N(t_1), ...N(t_n) - N(t_{n-1})$ are assumed to be independently distributed.

Model Assumptions.

1. At the start of the debugging process the number of errors contained in a software system is a random variable.
2. Each time a failure occurs the bug which caused it is immediately removed.
3. The time between failures k-1 and k depends on the time to k-1 failures.

Let

$$m_2(t) = a[1 - (1 + bt)e^{-bt}], \qquad a, b > 0, \tag{3.27}$$

represent the expected number of software failures up to time t. Clearly, $m_2(t)$ is nondecreasing and satisfies the boundary conditions given by

$$m_2(t) = \begin{cases} 0, & \text{for } t = 0, \\ a, & \text{for } t \to \infty, \end{cases} \tag{3.28}$$

where a is the expected number of software errors to be eventually detected. The error detection rate $d(t)$ per error at time t is:

$$d_1(t) = \frac{m_2'(t)}{a - m_2(t)} = \frac{b^2 t}{1 + bt}. \tag{3.29}$$

The intensity function or the error detection rate, $\lambda_2(t)$, is obtained by differentiating equation (3.27) with respect to t, which yields

$$\lambda_2(t) = ab^2 t e^{-bt}.$$

Clearly, $\lambda_2(0) = 0$ and $\lambda_2(\infty) = 0$.

For $t \geq 0, N(t)$ has a Poisson distribution with expected value $m_2(t)$, i.e.

$$P\{N(t) = y\} = \frac{[m_2(t)]^y}{y!} e^{-m_2(t)}, \qquad y = 0, 1, \ldots. \tag{3.30}$$

Thus, the stochastic behavior of software failure phenomena is described through the $N(t)$ process. Equations (3.27) and (3.30) constitute the basic models.

3.4 ESTIMATION OF PARAMETERS OF YAMADA *et al* MODEL

Case of Interfailure Data.

Let S_k be the time to kth failure. Then the joint density function of $S_1, S_2, \ldots, S_n$ is given by

$$f_{S_1,\ldots,S_n}(s_1,\ldots,s_n) = a^n b^{2n} e^{-a[1-(1+bs_n)e^{-bs_n}]} \prod_{k=1}^{n} s_k e^{-bs_k},$$
$$\text{where} \qquad a > 0, b > 0. \tag{3.31}$$

Given the failure times $s = (s_1, \ldots, s_n)$, the log likelihood function is given by

$$L_3(a,b|s) = n\log a + 2n\log b - b\sum_{k=1}^{n} s_k - a[1-(1+bs_n)e^{-bs_n}], \tag{3.32}$$

and the ML equations are given by

$$\frac{n}{a} = 1 - (1+bs_n)e^{-bs_n}, \tag{3.33}$$

$$\frac{2n}{b} = \sum_{k=1}^{n} s_k + abs_n^2 e^{-bs_n}. \tag{3.34}$$

No work has been done regarding the existence of the solution of these equations. Here we are going to find a necessary and sufficient condition for the existence of a finite solution for these two equations. The steps will be to reduce the two equations into a single equation containing only one parameter and then find its necessary and sufficient condition for the existence of roots.

The following lemma is used to prove Theorem 3 and Lemma 6, concerning the solution of equations (3.33) and (3.34).

LEMMA 5. *The function*

$$g_3(b) = \frac{bs_n e^{bs_n} - e^{bs_n} + 1}{(e^{bs_n} - bs_n - 1)^2} - \frac{2}{(bs_n)^2} \tag{3.35}$$

is negative for all $b > 0$.

PROOF: Combining the terms on the right hand side of the equation (3.35) we have

$$g_3(b) = \frac{(bs_n e^{bs_n} - e^{bs_n} + 1)(bs_n)^2 - 2(e^{bs_n} - bs_n - 1)^2}{(e^{bs_n} - bs_n - 1)^2 (bs_n)^2}. \tag{3.36}$$

Let $g_4(b)$ denote the numerator of the equation (3.36). Then

$$g_4(b) = -2e^{2bs_n} + (bs_n)^3 e^{bs_n} - (bs_n)^2 e^{bs_n} + 4bs_n e^{bs_n} + 4e^{bs_n} - (bs_n)^2 - 4bs_n - 2$$

and it can be seen that $g_4(0) = 0$. Differentiating $g_4(b)$ with respect to b we have

$$g_4'(b) = -4s_n e^{2bs_n} + b^3 s_n^4 e^{bs_n} + 2b^2 s_n^3 e^{bs_n} + 2bs_n^2 e^{bs_n} + 8s_n e^{bs_n} - 2bs_n^2 - 4s_n$$

and $g_4'(0) = 0$. Again, differentiating $g_4'(b)$ with respect to b we have

$$g_4''(b) = -8s_n^2 e^{2bs_n} + b^3 s_n^5 e^{bs_n} + 5b^2 s_n^4 e^{bs_n} + 6bs_n^3 e^{bs_n} + 10s_n^2 e^{-bs_n} - 2s_n^2$$

and $g_4''(0) = 0$. Differentiating $g_4''(b)$ once again with respect to b we obtain

$$\begin{aligned}
g_4'''(b) &= -16s_n^3 e^{2bs_n} + b^3 s_n^6 e^{bs_n} + 8b^2 s_n^5 e^{bs_n} + 16bs_n^4 e^{bs_n} + 16s_n^3 e^{bs_n} \\
&= s_n^3 e^{bs_n} \left[-16e^{bs_n} + b^3 s_n^3 + 8b^2 s_n^2 + 16bs_n + 16\right] \\
&= s_n^3 e^{bs_n} \left[16 + 16bs_n + 8b^2 s_n^2 + b^3 s_n^3 - 16e^{bs_n}\right] \\
&\leq s_n^3 e^{bs_n} \left[(16 + 16bs_n + 8b^2 s_n^2 + b^3 s_n^3) - 16(1 + bs_n + \frac{b^2 s_n^2}{2!} + \frac{b^3 s_n^3}{3!})\right] \\
&= -\frac{5}{3} b^3 s_n^6 e^{bs_n}, \qquad \text{for all } b > 0, \qquad (3.37)
\end{aligned}$$

which is a negative quantity. This implies that $g_4''(b)$ is decreasing in b and as $g_4''(0) = 0$, and we can conclude that $g_4''(b)$ is negative for all $b > 0$. In a similar fashion, going backward step by step, it can be shown that $g_4(b)$ is also a negative function. □

In the following theorem we are going to prove the necessary and sufficient condition for finite positive MLE's.

THEOREM 3. *The necessary and sufficient condition for equations (3.33) and (3.34) to have a finite positive roots is*

$$s_n > \frac{3}{2n} \sum_{k=1}^{n} s_k. \tag{3.38}$$

PROOF: Equation (3.33) can be solved explicitly for a in terms of b, giving $a = n[1 - (1 + bs_n)e^{-bs_n}]^{-1}$. Clearly, a is positive and finite if $b > 0$. Substituting this value of a into (3.34), we have an equation in terms of b alone. Let $h_2(b)$ denote this new equation. Then,

$$h_2(b) = \frac{2n}{b} - \sum_{k=1}^{n} s_k - \frac{nbs_n^2}{e^{bs_n} - bs_n - 1} = 0. \tag{3.39}$$

Therefore, the proof of the theorem boils down to showing that the necessary and the sufficient condition for the existence of a positive root of (3.39) is

$s_n > \frac{3}{2n}\sum_{k=1}^{n} s_k$. Now $\lim_{b\to 0} h_2(b)$ and $\lim_{b\to\infty} h_2(b)$ are given by

$$\begin{aligned}
\lim_{b\to 0} h_2(b) &= \lim_{b\to 0}\left[\frac{2n}{b} - \sum_{k=1}^{n} s_k - \frac{nbs_n^2}{e^{bs_n} - bs_n - 1}\right] \\
&= \lim_{b\to 0}\left[\frac{2ne^{bs_n} - nb^2s_n^2 - 2nbs_n - 2n}{be^{bs_n} - b^2s_n - b}\right] - \sum_{1}^{n} s_k \\
&= \lim_{b\to 0}\left[\frac{2ns_ne^{bs_n} - 2nbs_n^2 - 2ns_n}{bs_ne^{bs_n} + e^{bs_n} - 2bs_n - 1}\right] - \sum_{1}^{n} s_k \\
&= \lim_{b\to 0}\left[\frac{2ns_n^2e^{bs_n} - 2ns_n^2}{2s_ne^{bs_n} + bs_n^2e^{bs_n} - 2s_n}\right] - \sum_{1}^{n} s_k \\
&= \lim_{b\to 0}\left[\frac{2ns_n^3e^{bs_n}}{3s_n^2e^{bs_n} + bs_n^3e^{bs_n}}\right] - \sum_{1}^{n} s_k \\
&= \frac{2n}{3}s_n - \sum_{1}^{n} s_k
\end{aligned}$$

and

$$\lim_{b\to\infty} h_2(b) = -\sum_{k=1}^{n} s_k < 0.$$

Therefore the sufficient condition for the existence of a positive root for (3.39) is

$$s_n > \frac{3}{2n}\sum_{k=1}^{n} s_k. \tag{3.40}$$

To prove the necessary part let us prove that the function $h_2(b)$ is decreasing in b. Differentiating $h_2(b)$ with respect to b gives

$$\begin{aligned} h_2'(b) &= -\frac{2n}{b^2} - \frac{ns_n^2(e^{bs_n} - bs_n - 1) - nbs_n^2(s_n e^{bs_n} - s_n)}{(e^{bs_n} - bs_n - 1)^2} \\ &= -\frac{2n}{b^2} - \frac{ns_n^2 e^{bs_n} - ns_n^2 - nbs_n^3 e^{bs_n}}{(e^{bs_n} - bs_n - 1)^2} \\ &= ns_n^2 \left[\frac{bs_n e^{bs_n} - e^{bs_n} + 1}{(e^{bs_n} - bs_n - 1)^2} - \frac{2}{(bs_n)^2} \right]. \end{aligned} \tag{3.41}$$

Note that the bracketed expression in (3.41) is the same as the function $g_3(b)$ considered in Lemma 5. Therefore, by Lemma 5, $h_2'(b)$ is negative and as such $h_2(b)$ is a decreasing function in b. Hence it follows that equation (3.39) has a positive root if (3.38) is true. □

On using some results of Theorem 3, we obtain the following lemma related to the MLE's.

LEMMA 6. *If the roots of equations (3.33) and (3.34) are positive and finite then they are unique.*

PROOF: We have proved that $h_2(b)$ is a decreasing function for all $b > 0$. Therefore, if a positive root b exists, it is unique. Equation (3.33) provides an explicit solution for a in terms of b and a is finite and positive if $b > 0$. As such the roots are unique, provided they are finite and positive. □

Case of Grouped Data.

Suppose that the data is obtained in pairs of the form (t_k, y_k), $k = 1, 2, \ldots, n$, where y_k is the number of software failures detected up to time t_k. Then the joint density function of the observed values is

$$P_r[N(t_1) = y_1, \ldots, N(t_n) = y_n]$$

$$= \prod_{k=1}^{n} P_r[N(t_k) - N(t_{k-1}) = y_k - y_{k-1}]$$

$$= e^{-m_2(t_n)} \prod_{1}^{n} \frac{[m_2(t_k) - m_2(t_{k-1})]^{y_k - y_{k-1}}}{(y_k - y_{k-1})!}$$

$$= e^{-a[1-(1+bt_n)e^{-bt_n}]} \prod_{1}^{n} \frac{[a\{(1+bt_{k-1})e^{-bt_{k-1}} - (1+bt_k)e^{-bt_k}\}]^{y_k - y_{k-1}}}{(y_k - y_{k-1})!},$$

where, $\quad a, b > 0.$

Note that the likelihood function for the parameters is nothing but the joint probability density function of the observed values $y_1, \ldots, y_n$ and is given by

$$L_4(a, b|\tilde{y}, \tilde{t}) = e^{-a[1-(1+bt_n)e^{-bt_n}]} \times$$

$$\prod_{1}^{n} \frac{[a\{(1+bt_{k-1})e^{-bt_{k-1}} - (1+bt_k)e^{-bt_k}\}]^{y_k - y_{k-1}}}{(y_k - y_{k-1})!}. \quad (3.42)$$

For brevity let us write L_4 for $L_4(a, b|\tilde{y}, \tilde{t})$. Taking the natural logarithm of equation (3.42) yields:

$$\log L_4 = -a[1-(1+bt_n)e^{-bt_n}] + \sum_1^n (y_k - y_{k-1}) \log a - \sum_1^n \log(y_k - y_{k-1})$$

$$+ \sum_1^n (y_k - y_{k-1}) \log[(1+bt_{k-1})e^{-bt_{k-1}} - (1+bt_k)e^{-bt_k}]. \qquad (3.43)$$

The ML equations are given by

$$y_n = a[1-(1+bt_n)e^{-bt_n}], \quad \text{since } y_0 = 0, \qquad (3.44)$$

$$abt_n^2 e^{-bt_n} = \sum_{k=1}^n (y_k - y_{k-1}) \frac{b(t_k^2 e^{-bt_k} - t_{k-1}^2 e^{-bt_{k-1}})}{[(1+bt_{k-1})e^{-bt_{k-1}} - (1+bt_k)e^{-bt_k}]} (3.45)$$

THEOREM 4. *The sufficient condition for the roots of equations (3.44) and (3.45) to be finite and positive is*

$$y_n t_n > \sum_1^n (y_k - y_{k-1}) \frac{(t_k^3 - t_{k-1}^3)}{(t_k^2 - t_{k-1}^2)}.$$

PROOF: From equation (3.44) we obtain the value of a, which is $y_n[1-(1+bt_n)e^{-bt_n}]^{-1}$. Substituting this into equation (3.45) yields

$$
\begin{aligned}
0 &= -\sum_{k=1}^{n}(y_k - y_{k-1})\frac{b(t_{k-1}^2 e^{-bt_{k-1}} - t_k^2 e^{-bt_k})}{[(1+bt_{k-1})e^{-bt_{k-1}} - (1+bt_k)e^{-bt_k}]} \\
&\quad - \frac{y_n bt_n^2 e^{-bt_n}}{[1-(1+bt_n)e^{-bt_n}]} \\
&= -\sum_{k=1}^{n}(y_k - y_{k-1})\times \\
&\quad \left[\frac{bt_{k-1}^2 e^{b(t_k - t_{k-1})} - bt_k^2}{(1+bt_{k-1})e^{b(t_k-t_{k-1})} - bt_k - 1} + \frac{bt_n^2}{e^{bt_n} - bt_n - 1}\right].
\end{aligned}
$$

Combining the terms within the square bracket we have

$$
0 = -\sum_{k=1}^{n}(y_k - y_{k-1})\left[\frac{h_4(b)}{h_5(b)}\right], \tag{3.46}
$$

where

$$
\begin{aligned}
h_4(b) &= bt_{k-1}^2 e^{b(t_n+t_k-t_{k-1})} + b^2(t_n^2 t_{k-1} - t_n t_{k-1}^2)e^{b(t_k-t_{k-1})} \\
&\quad + b(t_n^2 - t_{k-1}^2)e^{b(t_k-t_{k-1})} - b^2(t_n^2 t_k - t_n t_k^2) - bt_k^2 e^{bt_n} - b(t_n^2 - t_k^2),
\end{aligned}
$$

and

$$
\begin{aligned}
h_5(b) &= bt_{k-1}e^{b(t_n+t_k-t_{k-1})} + e^{b(t_n+t_k-t_{k-1})} - b^2 t_n t_{k-1} e^{b(t_k-t_{k-1})} + b(t_k + t_n) \\
&\quad - b(t_n + t_{k-1})e^{b(t_k-t_{k-1})} - e^{b(t_k-t_{k-1})} - bt_k e^{bt_n} - e^{bt_n} + b^2 t_k t_n + 1.
\end{aligned}
$$

The value of a, as in equation (3.44), is finite and positive iff b obtained from equation (3.46) is positive. Equivalently, the proof of this theorem reduces to proving that the sufficient condition for the root of equation (3.46) is positive. Let $g_5(b)$ denote the function on the right hand side of the equation (3.46). To find the sufficient condition for the root to be positive let us obtain $\lim_{b\to 0} g_5(b)$ and $\lim_{b\to\infty} g_5(b)$.

$$
\begin{aligned}
\lim_{b\to 0} h_4(b) = \lim_{b\to 0} \Big[& bt_{k-1}^2 e^{b(t_n+t_k-t_{k-1})} + b^2(t_n^2 t_{k-1} - t_n t_{k-1}^2)e^{b(t_k-t_{k-1})} \\
& + b(t_n^2 - t_{k-1}^2)e^{b(t_k-t_{k-1})} - b^2(t_n^2 t_k - t_n t_k^2) - bt_k^2 e^{bt_n} - b(t_n^2 - t_k^2)\Big] \\
= {} & 0.
\end{aligned}
$$

$$
\begin{aligned}
\lim_{b\to 0} h_5(b) = \lim_{b\to 0} \Big[& bt_{k-1} e^{b(t_n+t_k-t_{k-1})} + e^{b(t_n+t_k-t_{k-1})} - b^2 t_n t_{k-1} e^{b(t_k-t_{k-1})} \\
& - b(t_n + t_{k-1})e^{b(t_k-t_{k-1})} - e^{b(t_k-t_{k-1})} - bt_k e^{bt_n} - e^{bt_n} + b^2 t_k t_n \\
& + b(t_k + t_n) + 1\Big] \\
= {} & 1 - 1 - 1 + 1 \\
= {} & 0.
\end{aligned}
$$

$\implies \lim_{b\to 0} \dfrac{h_4(b)}{h_5(b)}$ is of the form zero over zero.

Using L'Hopital's rule we have,

$$\lim_{b\to 0} h_4'(b) = \lim_{b\to 0}\Big[t_{k-1}^2 e^{b(t_n+t_k-t_{k-1})} + bt_{k-1}^2(t_n+t_k-t_{k-1})e^{b(t_n+t_k-t_{k-1})}$$

$$+ b[2(t_n^2 t_{k-1} - t_n t_{k-1}^2) + (t_n^2 - t_{k-1}^2)(t_k - t_{k-1})]e^{b(t_k-t_{k-1})}$$

$$+ b^2(t_n^2 t_{k-1} - t_n t_{k-1}^2)(t_k - t_{k-1})e^{b(t_k-t_{k-1})} + (t_n^2 - t_{k-1}^2)e^{b(t_k-t_{k-1})}$$

$$- 2b(t_n^2 t_k - t_n t_k^2) - t_k^2 e^{bt_n} - bt_k^2 t_n e^{bt_n} - (t_n^2 - t_k^2)\Big]$$

$$= t_{k-1}^2 + (t_n^2 - t_{k-1}^2) - t_k^2 - (t_n^2 - t_k^2)$$

$$= 0.$$

$$\lim_{b\to 0} h_5'(b) = \lim_{b\to 0}\Big[bt_{k-1}(t_n+t_k-t_{k-1})e^{b(t_n+t_k-t_{k-1})} + (t_n+t_k)e^{b(t_n+t_k-t_{k-1})}$$

$$- b[2t_n t_{k-1} + (t_n + t_{k-1})(t_k - t_{k-1})]e^{b(t_k-t_{k-1})}$$

$$- b^2 t_n t_{k-1}(t_k - t_{k-1})e^{b(t_k-t_{k-1})} - (t_n+t_k)e^{b(t_k-t_{k-1})}$$

$$- (t_n+t_k)e^{bt_n} - bt_k t_n e^{bt_n} + 2bt_k t_n + (t_k+t_n)\Big]$$

$$= (t_n+t_k) - (t_n+t_k) - (t_n+t_k) + (t_k+t_n)$$

$$= 0.$$

$\Longrightarrow \lim_{b\to 0}\dfrac{h_4'(b)}{h_5'(b)}$ is of the form zero over zero.

Similarly, it can be shown that

$$\lim_{b\to 0}\frac{h_4''(b)}{h_5''(b)} \quad \text{and} \quad \lim_{b\to 0}\frac{h_4'''(b)}{h_5'''(b)}$$

are of the form zero over zero.

Once again using L'Hopital's rule, we obtain

$$\begin{aligned}\lim_{b\to 0} h_4^{iv}(b) &= 4t_{k-1}^2(t_n+t_k-t_{k-1})^3 - 12t_nt_{k-1}^2(t_k-t_{k-1})^2 - 4t_{k-1}^2\times \\ &\quad (t_k-t_{k-1})^3 - 4t_k^2t_n^3 + 4t_n^2(t_k-t_{k-1})^3 + 12t_n^2t_{k-1}(t_k-t_{k-1})^2 \\ &= 4t_n^2(t_k^3-t_{k-1}^3) - 4t_n(t_k^2-t_{k-1}^2).\end{aligned}$$

$$\begin{aligned}\lim_{b\to 0} h_5^{iv}(b) &= (t_n+t_k-t_{k-1})^4 - 4t_n(t_k-t_{k-1})^3 + 4t_{k-1}(t_n+t_k-t_{k-1})^3 \\ &\quad - 12t_nt_{k-1}(t_k-t_{k-1})^2 - t_k(t_k-t_{k-1})^3 - 3t_{k-1}(t_k-t_{k-1})^3 \\ &\quad - 4t_kt_n^3 - t_n^4 \\ &= 6t_n^2(t_k^2-t_{k-1}^2).\end{aligned}$$

$$\begin{aligned}\Longrightarrow \lim_{b\to 0} g_5(b) &= -\lim_{b\to 0}\sum_{k=1}^{n}(y_k-y_{k-1})\frac{h_4^{iv}(b)}{h_5^{iv}(b)} \\ &= -\sum_{k=1}^{n}(y_k-y_{k-1})\left[\frac{4t_n^2(t_k^3-t_{k-1}^3)-4t_n(t_k^2-t_{k-1}^2)}{6t_n^2(t_k^2-t_{k-1}^2)}\right] \\ &= \frac{2}{3}y_nt_n - \frac{2}{3}\sum_{1}^{n}(y_k-y_{k-1})\frac{(t_k^3-t_{k-1}^3)}{(t_k^2-t_{k-1}^2)}. \qquad (2.47)\end{aligned}$$

Again, from the first step of equation (3.46), let

$$\lim_{b\to 0} h_2(b) = \lim_{b\to 0}\left[\frac{bt_{k-1}^2 e^{b(t_k-t_{k-1})} - bt_k^2}{(1+bt_{k-1})e^{b(t_k-t_{k-1})} - bt_k - 1} + \frac{bt_n^2}{e^{bt_n} - bt_n - 1}\right]$$

On using L'Hospital's rule we have

$$\lim_{b\to 0} h_2(b) = \lim_{b\to 0}\left[\frac{t_{k-1}^2 e^{b(t_k-t_{k-1})} + bt_{k-1}^2(t_k - t_{k-1})e^{b(t_k-t_{k-1})} - t_k^2}{t_k(1+bt_{k-1})e^{b(t_k-t_{k-1})} - bt_{k-1}^2 e^{b(t_k-t_{k-1})} - t_k}\right.$$
$$\left. + \frac{t_n^2}{t_n e^{bt_n} - t_n}\right].$$

Using L'Hopital's rule once again and simplifying further yields

$$\lim_{b\to 0} h_2(b) = t_{k-1}.$$

$$\Longrightarrow \lim_{b\to\infty} g_5(b) = -\sum_{k=1}^{n}(y_k - y_{k-1})t_{k-1} \le 0. \tag{3.48}$$

Therefore, sufficient condition for the root to be positive is

$$\frac{2}{3}y_n t_n - \frac{2}{3}\sum_1^n (y_k - y_{k-1})\frac{(t_k^3 - t_{k-1}^3)}{(t_k^2 - t_{k-1}^2)} \ge 0,$$

i.e.,

$$y_n t_n \ge \sum_1^n (y_k - y_{k-1})\frac{(t_k^3 - t_{k-1}^3)}{(t_k^2 - t_{k-1}^2)}. \qquad \square$$

4. MODIFIED GOEL-OKUMOTO AND MODIFIED YAMADA *et al* MODELS

4.1 DEVELOPEMENT OF MODIFIED GOEL-OKUMOTO MODEL

STOCHASTIC ANALYSIS OF FAILURE PROCESS

The errors in a software system are encountered when a sequence of instructions is executed. Let $N(t)$ denote the number of these errors encountered up to time t and $t_1 < t_2 < \ldots < t_n$ denote a finite collection of times. Then it is plausible to assume that the n random variables $N(t_1), N(t_2) - N(t_1), \ldots, N(t_n) - N(t_{n-1})$ are independently distributed. To build our model we make the following assumptions:

Assumptions.

1. The initial error content of a software system is a random variable.
2. Each time a failure is encountered, it is removed immediately.
3. The time between $k-1$ and k failures depends on the time to k-1 failures.

Let $m_3(t)$ be the mean value function of the N(t) process, i.e.,

$$m_3(t) = E[N(t)]. \tag{4.1}$$

Since $m_3(t)$ represents the expected number of software failures by time t, it should be a non-decreasing function of t. If we assume a finite number of errors in the system then $m_3(t)$ should have the following boundary conditions

$$m_3(t) = \begin{cases} 0, & \text{when} \quad t = 0, \\ q, & \text{when} \quad t \to \infty, \end{cases} \tag{4.2}$$

where q is a finite quantity. Based on this boundary condition our suggested model is

$$m_3(t) = \log\left[\frac{e^a - c}{e^{ae^{-bt}} - c}\right], \qquad a, b > 0 \quad \text{and} \quad 0 \le c < 1. \tag{4.3}$$

Clearly,

$$m_3(t) = \begin{cases} 0, & \text{when} \quad t = 0, \\ \log(\frac{e^a - c}{1-c}), & \text{when} \quad t \to \infty. \end{cases} \tag{4.4}$$

Note that $\log(\frac{e^a-c}{1-c})$ is the expected number of faults to be eventually detected and is a finite quantity.

The rationale behind the suggested model is that when $c = 0$, it is the G-O model, and when $c = 1$, the corresponding probability density function of the failure time is a proper density. Note that for $0 \leq c < 1$, the probability density function of the first failure time is an improper density. So our endeavour will be to look for a c that will give a better estimate of the expected number of faults in the system than that of the Goel-Okumoto model.

FAILURE RATE

Let F be the life distribution of a unit. Then F is an Increasing Failure Rate (IFR) distribution or Decreasing Failure Rate (DFR) distribution depending on whether the function

$$\bar{F}(x|t) = \frac{\bar{F}(x+t)}{\bar{F}(t)}.$$

is decreasing or increasing in $0 < t < \infty$ for all $x > 0$, where $\bar{F} = 1 - F$.

Failure rate tells us about the number of failures occured per unit of time. Therefore, we would expect a DFR failure rate from a software reliability model. Following lemma proves that our model has DFR.

LEMMA 1. *The F corresponding to the model given by (4.3) is a DFR distribution.*

PROOF: The function $F(t)$ may be obtained from the relation $\bar{F}(t) =$

$e^{-m_3(t)}$, (cf. Barlow and Proschan (1975)). Therefore,

$$\begin{aligned}\bar{F}(x|t) &= \frac{\bar{F}(x+t)}{\bar{F}(t)} \\ &= \frac{e^{ae^{-b(t+x)}} - c}{e^{ae^{-bt}} - c}. \end{aligned} \tag{4.5}$$

To show that $\bar{F}(x|t)$ is increasing in t, let us show that $\bar{F}'(x|t)$, where prime stands for derivative with respect to t, is positive for all $x > 0$. Taking the derivative with respect to t of the natural log of the function given by (4.5), we have

$$\begin{aligned}\frac{d\log \bar{F}(x|t)}{dt} &= \frac{-abe^{-b(t+x)}e^{ae^{-b(t+x)}}}{e^{ae^{-b(t+x)}} - c} + \frac{abe^{-bt}e^{ae^{-bt}}}{e^{ae^{-bt}} - c} \\ &= abe^{-bt}\left[-\frac{e^{-bx}e^{ae^{-b(t+x)}}}{e^{ae^{-b(t+x)}} - c} + \frac{e^{ae^{-bt}}}{e^{ae^{-bt}} - c}\right].\end{aligned}$$

By taking derivative with respect to x it can easily be shown that the function $-\frac{e^{-bx}e^{ae^{-b(t+x)}}}{e^{ae^{-b(t+x)}} - c}$ is increasing in x. Therefore for all $x > 0$,

$$\frac{d\log \bar{F}(x|t)}{dt} \geq abe^{-bt}\left[-\frac{e^{ae^{-bt}}}{e^{ae^{-bt}} - c} + \frac{e^{ae^{-bt}}}{e^{ae^{-bt}} - c}\right] = 0$$

□

ERROR DETECTION RATE PER ERROR

The error detection rate per error in our suggested model is

$$
\begin{aligned}
d_3(t) &= \frac{m_3'(t)}{m_3(\infty) - m_3(t)} \\
&= \frac{abe^{-bt}e^{ae^{-bt}}}{(e^{ae^{-bt}} - c)\log(\frac{e^{ae^{-bt}}-c}{1-c})}. \qquad (4.6)
\end{aligned}
$$

Now we investigate the behavior of $m_3(t)$ in the following theorem.

THEOREM 1. *At the start $m_3(t)$ is a DEDR function and then gradually becomes an IEDR function based on the values of c.*

PROOF: Taking the natural logarithm of (4.6) and on further simplification after differentiating with respect to t, we have

$$
\begin{aligned}
\frac{d\log d(t)}{dt} &= -b - abe^{-bt} + \frac{abe^{-bt}e^{ae^{-bt}}}{(e^{ae^{-bt}} - c)} + \frac{abe^{-bt}e^{ae^{-bt}}}{(e^{ae^{-bt}} - c)\log(\frac{e^{ae^{-bt}}-c}{1-c})} \\
&= -b - abe^{-bt} + y(c) \qquad (4.7)
\end{aligned}
$$

Note that $\frac{d\log d(t)}{dt}\Big|_{c=0} = 0$. Now, we want to look at the domain of c to find where the function $\frac{d \log d(t)}{dt}$ is positive and where it is negative.

$$\Longrightarrow y'(c) = abe^{-bt}e^{ae^{-bt}}\left[\frac{1}{(e^{ae^{-bt}}-c)^2} - \frac{\frac{e^{ae^{-bt}}-1}{1-c} - \log(\frac{e^{ae^{-bt}}-c}{1-c})}{[(e^{ae^{-bt}}-c)\log(\frac{e^{ae^{-bt}}-c}{1-c})]^2}\right]$$

$$= \frac{abe^{-bt}e^{ae^{-bt}}}{[(e^{ae^{-bt}}-c)\log(\frac{e^{ae^{-bt}}-c}{1-c})]^2}\left[[\log(\frac{e^{ae^{-bt}}-c}{1-c})]^2 - \frac{e^{ae^{-bt}}-1}{1-c} + \log(\frac{e^{ae^{-bt}}-c}{1-c})\right]$$

$$= \frac{abe^{-bt}e^{ae^{-bt}}}{[(e^{ae^{-bt}}-c)\log(\frac{e^{ae^{-bt}}-c}{1-c})]^2}\left[[\log(\frac{e^{ae^{-bt}}-c}{1-c})]^2 + \log(\frac{e^{ae^{-bt}}-c}{1-c}) + 1 - \frac{e^{ae^{-bt}}-c}{1-c}\right].$$

Let $Q = \log(\frac{e^{ae^{-bt}}-c}{1-c})$. Then, the expression within the square bracket is

$$G(Q) = Q^2 + Q + 1 - e^Q.$$

It is easy to verify that the zero of $G(Q)$ lies in the interval $(1.7,1.8)$. Let r denote the exact zero of $G(Q)$. Therefore, $y'(c)$ is negative if $Q > r$. That is, $\frac{d \log d(t)}{dt}$ is negative if

$$\log(\frac{e^{ae^{-bt}}-c}{1-c}) > r.$$

$$\Longrightarrow \text{negative if} \quad c > \frac{e^r - e^{ae^{-bt}}}{e^r - 1}.$$

Clearly, as t increases the above inequality does not hold good because $0 \leq c < 1$ and as a result $\frac{d \log d(t)}{dt}$ becomes positive. Therefore, based on the value of c, the equation (4.7) is negative for small values of t and gradually becomes positive. □

SOFTWARE PERFORMANCE ANALYSIS

To use the model for predictive purposes we need to investigate the measures like the number of failures up to time t, expected faults remaining in the system at a future time, reliability during the mission, etc. In this section we develop models that can be used to estimate these measures.

1. NUMBER OF FAILURES UP TO TIME t.

Given the parameters a, b and c the distribution of $N(t)$ is Poisson with mean $m_3(t) = \log[\frac{e^a - c}{e^{ae^{-bt}} - c}]$, i.e.,

$$P[N(t) = y] = \frac{[m_3(t)]^y}{y!} e^{-m_3(t)}, \qquad y = 0, 1, 2, \ldots;$$

Clearly,

$$P[N(\infty) = y] = \frac{[m_3(\infty)]^y}{y!} e^{-m_3(\infty)}, \qquad y = 0, 1, 2, \ldots; \qquad (4.8)$$

which is the total number of failures encountered if the system is used indefinitely, is also a Poisson distribution with mean $m_3(\infty) = \log[\frac{e^a - c}{1 - c}]$.

2. REMAINING ERRORS AFTER DEBUGGING.

Let $\bar{N}(t)$ denote the number of errors remaining in the system at time t i.e.,

$$\bar{N}(t) = N(\infty) - N(t). \tag{4.9}$$

Then the expected number of faults remaining in the system at time t is

$$E\bar{N}(t) = \log\left(\frac{e^{ae^{-bt}} - c}{1 - c}\right). \tag{4.10}$$

Goel-Okumoto (1979), used the conditional distribution of $\bar{N}(t) = N(\infty) - N(t)$, given $N(t) = y$, to obtain expected number of faults remaining in the software system after y number of faults have been detected during the test period. The very concept of talking about the conditional distribution and the conditional expectation is erroneous. This is because in any finite collection of times $t_1, t_2, \ldots, t_n$ the n random variables $N(t_1), N(t_2) - N(t_1), \ldots, N(t_n) - N(t_{n-1})$ are independent. Therefore, all the inferences based on this conditional distribution of $\bar{N}(t)$ are erroneous since $\bar{N}(t)$ and $N(t)$ are independent. The equation (14) of Goel-Okumoto (1979), giving the conditional distribution of $\bar{N}(t)$, given $N(t)$, in in error, and this distribution is the same as the unconditional distribution of $\bar{N}(t)$ given below in (4.11).

Probability distribution of $\bar{N}(t)$, by definition of NHPP Barlow and Proschan (1975), is given by

$$P[\bar{N}(t) = y] = \frac{[m_3(\infty) - m_3(t)]^y}{y!} e^{-[m_3(\infty) - m_3(t)]}$$

$$= \frac{[\log \frac{e^{ae^{-bt}} - c}{1-c}]^y}{y!} \left(\frac{1-c}{e^{ae^{-bt}} - c}\right), \tag{4.11}$$

where $\quad k = 0, 1, 2, \ldots$

Knowing the distribution of remaining faults in the system is of utmost importance. Based on it we would decide whether the software system can be released or not.

CONDITIONAL RELIABILITY FUNCTION

The reliability of a fresh unit corresponding to a mission of duration x is, by definition, $1 - F(x)$, where F is the life distribution of the unit. The corresponding reliability of a unit of age x is

$$R(t|x) = \frac{1 - F(t + x)}{1 - F(x)}. \tag{4.12}$$

To find the conditional reliability function for our model let us first find out the life distribution function of a unit. This may be obtained from the relation $F(t) = 1 - e^{-m_3(t)}$ and is found to be

$$F(t) = 1 - \frac{e^{ae^{-bt}} - c}{e^a - c}. \tag{4.13}$$

Therefore using (4.13) in (4.12) we obtain the conditional reliability function given that the nth failure occured at time s_n, as

$$R_1(t|s_n = s) = \frac{e^{ae^{-b(s+t)}} - c}{e^{ae^{-bs}} - c}. \tag{4.14}$$

<u>JOINT DENSITY FUNCTION OF WAITING TIMES</u>

Let $\{X_k, \quad k = 1, 2, \ldots\}$ denote a sequence of times between software failures. Then

$$S_n = \sum_{k=1}^{n} X_k$$

is called the waiting time to the nth software failure.

THEOREM 2. *The joint density function of $S_1, S_2, \ldots, S_n$ is given by*

$$f_{S_1,\ldots,S_n}(s_1, \ldots, s_n) = \frac{a^n b^n e^{-b\sum_1^n s_k} e^{a\sum_1^n e^{-bs_k}}}{\prod_{k=0}^{n-1} (e^{ae^{-bs_k}} - c)}, \tag{4.15}$$

where $\quad a, b > 0, \quad 0 \le c < 1, \quad$ *and* $\quad 0 = s_0 < s_1 < \ldots < s_n < \infty$

PROOF: Differentiating $F(t)$, given by (4.13), with respect to t and noting that s_1 is the time to the first failure, we obtain the probability density function of S_1, given by

$$f_{(3)S_1}(s_1) = \frac{abe^{-bs_1}}{(e^a - c)} e^{ae^{-bs_1}}, \qquad 0 < s_1 < \infty \tag{4.16}$$

Conditional density function of a unit, given that it has survived up to time s may be obtained from the relation

$$f(x|s) = \frac{f(x+s)}{1 - F(s)}, \tag{4.17}$$

and for our model it is found to be

$$\begin{aligned} f_{S_2|S_1}(s_2|s_1) &= f_{S_2|S_1}(x_2 = s_2 - s_1|s_1) \\ &= \frac{abe^{-bs_2}}{(e^{ae^{-bs_1}} - c)} e^{ae^{-bs_2}}, \qquad 0 < s_1 < s_2 < \infty. \end{aligned}$$

Therefore,

$$\begin{aligned} f_{S_1,S_2}(s_1,s_2) &= f_{S_1}(s_1) f_{S_2|S_1}(s_2|s_1) \\ &= \frac{a^2b^2e^{-b(s_1+s_2)}}{(e^a - c)(e^{ae^{-bs_1}} - c)} e^{a(e^{-bs_1}+e^{-bs_2})}, \qquad 0 < s_1 < s_2 < \infty. \end{aligned}$$

To find the joint density function for s_1, s_2, s_3, first we use the relation (4.17) to obtain $f_{(3)S_3|S_2}(s_3|s_2)$. Then

$$\begin{aligned} f_{S_1,S_2,S_3}(s_1,s_2,s_3) &= f_{S_1}(s_1) f_{(3)S_2|S_1}(s_2|s_1) f_{S_3|S_1,S_2}(s_3|s_1,s_2) \\ &= \frac{a^3b^3e^{-b(s_1+s_2+s_3)}}{(e^a - c)(e^{ae^{-bs_1}} - c)(e^{ae^{-bs_2}} - c)} e^{a(e^{-bs_1}+e^{-bs_2}+e^{-bs_3})}, \\ &\qquad \text{where} \quad 0 < s_1 < s_2 < s_3 < \infty. \end{aligned}$$

Similarly, for a sample of size n, the joint density function is

$$f_{S_1,\ldots,S_n}(s_1,\ldots,s_n) = \frac{a^n b^n e^{-b\sum_1^n s_k} e^{a\sum_1^n e^{-bs_k}}}{\prod_{k=0}^{n-1}(e^{ae^{-bs_k}} - c)}. \tag{4.18}$$

□

JOINT DENSITY FUNCTION OF GROUPED DATA

Suppose the data is grouped for some collection of times $0 < t_1 < t_2 < \ldots < t_n$, then the joint density function of this grouped data is given by

$$P_3[N(t_1) = y_1, \ldots, N(t_n) = y_n]$$

$$= \prod_{k=1}^{n} P_3[N(t_k) - N(t_{k-1}) = y_k - y_{k-1}]$$

$$= e^{-m_3(t_n)} \prod_{k=1}^{n} \frac{[m_3(t_k) - m_3(t_{k-1})]^{y_k - y_{k-1}}}{(y_k - y_{k-1})!}$$

$$= (\frac{e^{ae^{-bt_n}} - c}{e^a - c}) \prod_{k=1}^{n} \frac{(\log \frac{e^{ae^{-bt_{k-1}}} - c}{e^{ae^{-bt_k}} - c})^{y_k - y_{k-1}}}{(y_k - y_{k-1})!} \tag{4.19}$$

where, y_k is the cumulative number of failures up to time t_k.

Once the joint density function is in hand we can obtain the log likelihood function of the observations and the ML equations. By solving the ML equations we will obtain the estimates for a b and c which will be used in the mean value function $m_3(t)$ to estimate the model.

4.2 ESTIMATION OF PARAMETERS

CASE OF INTER-FAILURE DATA

Given the failure times $s_1, s_2, \ldots, s_n$ and using equation (4.15), the likelihood function is given by

$$L_5(a,b,c|s_1,\ldots,s_n) = \frac{a^n b^n e^{-b\sum_1^n s_k} e^{a\sum_1^n e^{-bs_k}}}{\prod_{k=0}^{n-1}(e^{ae^{-bs_k}} - c)}. \tag{4.20}$$

We now discuss maximizing the likelihood function. Clearly, the likelihood function is increasing in c. Therefore $c = 1$ is the MLE. But when $c = 1$ the mean value function at infinity, $m_3(\infty)$, is infinite. That is why we are assuming $0 \leq c < 1$ and we shall look for a c which will be best in some other respect and the same is discussed later. To estimate a and b, let us differentiate (4.20) with respect to a and b separately and equate them with zero. The equations obtained after some simplification are as follows

$$\frac{d\log L_5(a,b,c)}{da} = \frac{n}{a} + e^{-bs_n} - \frac{e^a}{e^a - c} - c\sum_{k=1}^{n-1}\frac{e^{-bs_k}}{(e^{ae^{-bs_k}} - c)} = 0 \tag{4.21}$$

and

$$\frac{d\log L_5(a,b,c)}{db} = \frac{n}{b} - \sum_1^n s_k - as_n e^{-bs_n} + ac\sum_1^{n-1}\frac{s_k e^{-bs_k}}{(e^{ae^{-bs_k}} - c)} = 0 \tag{4.22}$$

We now need to solve these two equations for a and b which is discussed in Theorem 3. First, we prove the following lemmas which are used in proving Theorem 3.

LEMMA 2. *Let*

$$f_b(a) = \frac{n}{a} + e^{-bs_n} - \frac{e^a}{e^a - c} - c\sum_{k=1}^{n-1} \frac{e^{-bs_k}}{\left(e^{ae^{-bs_k}} - c\right)}.$$

Then, $f_b(a)$ is a decreasing function of a for given b.

PROOF: To prove $f_b(a)$ is decreasing in a, let us show that $f_b'(a)$ is negative. Differentiating $f_b(a)$ with respect to a we have

$$\begin{aligned}
f_b'(a) &= -\frac{n}{a^2} + \frac{ce^a}{(e^a - c)^2} + c\sum_{k=1}^{n-1} \frac{e^{-2bs_k} e^{ae^{-bs_k}}}{\left(e^{ae^{-bs_k}} - c\right)^2} \\
&= \left[-\frac{1}{a^2} + \frac{ce^a}{(e^a - c)^2}\right] + \sum_{k=1}^{n-1} \left[-\frac{1}{a^2} + \frac{ce^{-2bs_k} e^{ae^{-bs_k}}}{\left(e^{ae^{-bs_k}} - c\right)^2}\right] \\
&= \frac{ca^2 e^a - e^{2a} + 2ce^a - c^2}{a^2 (e^a - c)^2} + \\
&\quad \sum_{k=1}^{n-1} \frac{c\left(ae^{-bs_k}\right)^2 e^{ae^{-bs_k}} - e^{2ae^{-bs_k}} + 2ce^{ae^{-bs_k}} - c^2}{a^2\left(e^{ae^{-bs_k}} - c\right)^2}.
\end{aligned} \tag{4.23}$$

Clearly, the denominators of the right hand side expressions of (4.23) are positive. To show $f_b'(a)$ is negative let us consider the numerator of the second expresssion which is, say, $l_1(a)$. Then

$$l_1(a) = c(ae^{-bs_k})^2 e^{ae^{-bs_k}} - e^{2ae^{-bs_k}} + 2ce^{ae^{-bs_k}} - c^2,$$

giving

$$l_1(0) = -1 + 2c - c^2 = -(1-c)^2. \tag{4.24}$$

Now, differentiating $l_1(a)$ w.r.t. a, we get

$$l_1'(a) = 2ace^{-2bs_k} e^{ae^{-bs_k}} + a^2 e^{-3bs_k} e^{ae^{-bs_k}} - 2e^{-bs_k} e^{2ae^{-bs_k}}$$

$$+ 2ce^{-bs_k} e^{ae^{-bs_k}}, \tag{4.25}$$

giving

$$l_1'(0) = -2e^{-bs_k} + 2ce^{-bs_k} = -2(1-c)e^{-bs_k} \leq 0.$$

Again $\quad l_1'(a) = e^{ae^{-bs_k}} l_2(a),$

where $\quad l_2(a) = 2ace^{-2bs_k} + a^2 e^{-3bs_k} - 2e^{-bs_k} e^{ae^{-bs_k}} + 2ce^{-bs_k}.$

Note that

$$l_2(0) = -2e^{-bs_k} + 2ce^{-bs_k} = -2(1-c)e^{-bs_k} \leq 0. \tag{4.25}$$

and

$$\begin{aligned} l_2'(a) &= 2ce^{-2bs_k} + 2ae^{-3bs_k} - 2e^{-2bs_k} e^{ae^{-bs_k}} \\ &\leq 2ce^{-2bs_k} + 2ae^{-3bs_k} - \left[2e^{-2bs_k} + 2ae^{-3bs_k}\right] \\ &= -2(1-c)e^{-2bs_k} \\ &\leq 0. \end{aligned} \tag{4.27}$$

The inequalities (4.26) and (4.27) prove that $l_2(a)$ is negative, which in turn proves that $l_1'(a)$ is negative. Therefore, by (4.24)and (4.25), $l_1(a)$ is negative for all k. Note that the numerator of the first expression of (4.23) is the same as that of the second expression if we replace e^{-bs_k} of the second expression by 1. Therefore both of the numerators are negative, which proves that $f_b'(a)$ is negative, and this in turn proves that $f_b(a)$ is decreasing in a. □

LEMMA 3. *The upper and the lower bounds of a in the solution of equation $f_b(a) = 0$ is $n(1 - e^{-bs_n})^{-1}$ and 0 respectively i.e., the solution bounds are given by*

$$0 < a < n(1 - e^{-bs_n})^{-1}. \tag{4.28}$$

PROOF: Clearly,

$$\frac{e^a}{(e^a - c)} + c\sum_{k=1}^{n-1} \frac{e^{-bs_k}}{(e^{ae^{-bs_k}} - c)}$$

is decreasing in a and further more

$$\lim_{a\to\infty}\left[\frac{e^a}{(e^a - c)} + c\sum_{k=1}^{n-1} \frac{ce^{-bs_k}}{(e^{ae^{-bs_k}} - c)}\right] = 1$$

Therefore,

$$\begin{aligned} f_b(a) &\leq \frac{n}{a} + e^{-bs_n} - \lim_{a\to\infty}\left[\frac{e^a}{e^a - c} + c\sum_{k=1}^{n-1} \frac{e^{-bs_k}}{(e^{ae^{-bs_k}} - c)}\right] \\ &= \frac{n}{a} + e^{-bs_n} - 1 \end{aligned} \tag{4.29}$$

Equation (4.29) implies that

$$f_b(a) \leq 0 \quad \text{if} \quad \frac{n}{a} + e^{-bs_n} - 1 \leq 0$$

$$\Longrightarrow f_b(a) \leq 0 \quad \text{if} \quad a \geq n(1 - e^{-bs_n})^{-1} \tag{4.30}$$

Also we have

$$\lim_{a \to 0} f_b(a) = \lim_{a \to 0} \left[\frac{n}{a} + e^{-bs_n} - \frac{e^a}{e^a - c} - c \sum_{k=1}^{n-1} \frac{e^{-bs_k}}{\left(e^{ae^{-bs_k}} - c\right)} \right]$$

$$= \infty$$

$$\Longrightarrow f_b(a) \geq 0 \quad \text{as} \quad a \to 0. \tag{4.31}$$

The inequalities (4.30), (4.31) and Lemma 2 prove that for given b the root a of the equation $f_b(a) = 0$ lies in the interval 0 to $n(1 - e^{-bs_n})^{-1}$. □

THEOREM 3. *The sufficient condition for the equations (4.21) and (4.22) to have finite roots is*

$$s_n > \frac{2}{n} \sum_{k=1}^{n} s_k$$

PROOF: Let us denote the right hand side expression of equation (4.22) by $g_a(b)$. Therefore we have to determine the sufficient condition for $f_b(a)$ and $g_a(b)$ to have finite zeros. Clearly, $g_a(b)$ is decreasing in a. To get the sufficient condition for the existence of finite zeros, we need to determine $\inf_a\{g_a(b)\}$ and $\sup_a\{g_a(b)\}$.

$$\inf_a g_a(b) = \inf_a \left[\frac{n}{b} - \sum_1^n s_k - as_n e^{-bs_n} + ac\sum_1^{n-1} \frac{s_k e^{-bs_k}}{(e^{ae^{-bs_k}} - c)}\right]$$

$$\geq \frac{n}{b} - \sum_1^n s_k - \frac{ns_n}{e^{bs_n}-1} + c\sum_1^{n-1} s_k \frac{\frac{ne^{-bs_k}}{1-e^{-bs_n}}}{(e^{\frac{ne^{-bs_k}}{1-e^{-bs_n}}} - c)},$$

on using lemma 2. This implies that

$$\lim_{b\to 0}\inf_a g_a(b) \geq \lim_{b\to 0}\left[\frac{n}{b} - \sum_1^n s_k - \frac{ns_n}{e^{bs_n}-1} + c\sum_1^{n-1} s_k \frac{\frac{ne^{-bs_k}}{1-e^{-bs_n}}}{(e^{\frac{ne^{-bs_k}}{1-e^{-bs_n}}} - c)}\right]$$

$$= -\sum_1^n s_k + \lim_{b\to 0}\frac{ns_n e^{bs_n} - ns_n}{e^{bs_n} + bs_n e^{bs_n} - 1} +$$

$$c\lim_{b\to 0}\sum_1^{n-1} s_k \frac{\frac{d}{db}\frac{ne^{-bs_k}}{1-e^{-bs_n}}}{e^{\frac{ne^{-bs_k}}{1-e^{-bs_n}}}\frac{d}{db}\frac{ne^{-bs_k}}{1-e^{-bs_n}}}$$

$$= -\sum_1^n s_k + \lim_{b\to 0}\frac{ns_n^2 e^{bs_n}}{2s_n e^{bs_n} + bs_n^2 e^{bs_n}}$$

$$= -\sum_1^n s_k + n\frac{s_n}{2}. \tag{4.32}$$

Again, $$\sup_a g_a(b) \leq \frac{n}{b} - \sum_{k=1}^n s_k + c\lim_{a\to 0}\sum_{k=1}^{n-1} s_k \frac{ae^{-bs_k}}{e^{ae^{-bs_k}} - c}$$

$$= \frac{n}{b} - \sum_{k=1}^n s_k$$

Hence, $$\lim_{b\to\infty}\sup_a g_a(b) \leq -\sum_{k=1}^n s_k. \tag{4.33}$$

By (4.33) $\lim_{b\to\infty}\sup_a g_a(b)$ is negative, therefore we will have finite zeros if $\lim_{b\to 0}\inf_a g_a(b)$ is positive. By (4.32) $\lim_{b\to 0}\inf_a g_a(b) \geq -\sum_1^n s_k + n\frac{s_n}{2}$. Therefore equations (4.21) and (4.22) will have finite roots if $s_n > \frac{2}{n}\sum_{k=1}^n s_n$. Therefore for given c we can obtain the maximum likelihood estimates for a and b if the sufficient condition is satisfied. □

The method suggested to estimate c is to use the well known function $\frac{1}{n}\sum_{k=1}^n (O_k - E_k)^2$, where O_k and E_k stands for observed and expected time to failures, respectively . The estimated c is the minimum c for which the function is almost invariant with respect to any increase in c .

CASE OF GROUPED DATA

Given the ordered pairs $(y_k, t_k), k = 1, \ldots, n,$ of observations and using equation (4.19), the likelihood function for group data is given by

$$L_6\{a, b, c|\tilde{y}, \tilde{t}\} = \left(\frac{e^{ae^{-bt_n}} - c}{e^a - c}\right) \prod_{k=1}^n \frac{\left(\log \frac{e^{ae^{-bt_{k-1}}} - c}{e^{ae^{-bt_k}} - c}\right)^{y_k - y_{k-1}}}{(y_k - y_{k-1})!} \tag{4.34}$$

By taking the natural logarithm and then partial derivatives of equation (4.34) with respect to a, b, and c separately and equating them to zero, we obtain, after some simplification, the following equations:

$$
\begin{aligned}
\frac{d\log L_6(a,b,c)}{da} =& \frac{e^{-bt_n}e^{ae^{-bt_n}}}{e^{ae^{-bt_n}}-c} - \frac{e^a}{e^a-c} + \\
& \sum_{k=1}^{n}(y_k-y_{k-1})\frac{e^{-bt_{k-1}}e^{ae^{-bt_{k-1}}}}{(e^{ae^{-bt_{k-1}}}-c)\log\frac{e^{ae^{-bt_{k-1}}}-c}{e^{ae^{-bt_k}}-c}} - \\
& \sum_{k=1}^{n}(y_k-y_{k-1})\frac{e^{-bt_k}e^{ae^{-bt_k}}}{(e^{ae^{-bt_k}}-c)\log\frac{e^{ae^{-bt_{k-1}}}-c}{e^{ae^{-bt_k}}-c}} \\
=&0
\end{aligned}
\tag{4.35}
$$

$$
\begin{aligned}
\frac{d\log L_6(a,b,c)}{db} =& -at_n\frac{e^{-bt_n}e^{ae^{-bt_n}}}{e^{ae^{-bt_n}}-c} - \\
& \sum_{k=1}^{n}(y_k-y_{k-1})\frac{at_{k-1}e^{-bt_{k-1}}e^{ae^{-bt_{k-1}}}}{(e^{ae^{-bt_{k-1}}}-c)\log\frac{e^{ae^{-bt_{k-1}}}-c}{e^{ae^{-bt_k}}-c}} + \\
& \sum_{k=1}^{n}(y_k-y_{k-1})\frac{at_ke^{-bt_k}e^{ae^{-bt_k}}}{(e^{ae^{-bt_k}}-c)\log\frac{e^{ae^{-bt_{k-1}}}-c}{e^{ae^{-bt_k}}-c}} \\
=&0
\end{aligned}
\tag{4.36}
$$

and,

$$\frac{d\log L_6(a,b,c)}{dc} = -\frac{1}{e^{ae^{-bt_n}}-c}+\frac{1}{e^a-c}+$$

$$\sum_{k=1}^{n}\frac{(y_k-y_{k-1})}{(e^{ae^{-bt_k}}-c)\log(\frac{e^{ae^{-bt_{k-1}}}-c}{e^{ae^{-bt_k}}-c})}-$$

$$\sum_{k=1}^{n}\frac{(y_k-y_{k-1})}{(e^{ae^{-bt_{k-1}}}-c)\log(\frac{e^{ae^{-bt_{k-1}}}-c}{e^{ae^{-bt_k}}-c})}$$

$$=0 \tag{4.37}$$

As can easily be seen, the three equations do not yield simple analytical forms for the solutions of a, b, and c. Therefore we must resort to numerical methods for their solutions.

VARIANCE-COVARIANCE MATRIX

It is clear that the exact variance-covariance matrix for $(\hat{a},\hat{b},\hat{c})'$ is not obtainable because its true distribution is unknown. However, MLE's have a desirable property that they are asymptotically normally distributed. Asymptotic mean vector and covariance matrix are given by:

$$E\begin{pmatrix}\hat{a}\\ \hat{b}\\ \hat{c}\end{pmatrix}=\begin{pmatrix}a\\ b\\ c\end{pmatrix} \tag{4.38}$$

and variance-covariance matrix

$$\Sigma_{cov} = \begin{pmatrix} \mathrm{Var}(\hat{a}) & \mathrm{Cov}(\hat{a},\hat{b}) & \mathrm{Cov}(\hat{a},\hat{c}) \\ \mathrm{Cov}(\hat{b},\hat{a}) & \mathrm{Var}(\hat{b}) & \mathrm{Cov}(\hat{b},\hat{c}) \\ \mathrm{Cov}(\hat{c},\hat{a}) & \mathrm{Cov}(\hat{c},\hat{b}) & \mathrm{Var}(\hat{c}) \end{pmatrix}, \qquad (4.39)$$

given by

$$\Sigma_{cov} = \begin{pmatrix} \sigma_{aa} & \sigma_{ab} & \sigma_{ac} \\ \sigma_{ba} & \sigma_{bb} & \sigma_{bc} \\ \sigma_{ca} & \sigma_{cb} & \sigma_{cc} \end{pmatrix}^{-1}$$

where

$$\sigma_{ij} = -E\left(\frac{d^2 \log L_2}{di dj}\right), \qquad i,j = a,b,c. \qquad (4.40)$$

Taking the appropriate partial derivatives and using the relation (4.40) we obtain the variances and covariances as follows

$$\sigma_{aa} = \sum_{k=1}^{n} \left[\log \frac{e^{ae^{-bt_{k-1}}}}{e^{ae^{-bt_k}}} \right]^{-1} \times$$

$$\left[\frac{e^{-bt_k} e^{ae^{-bt_k}}}{e^{ae^{-bt_k}} - c} - \frac{e^{-bt_{k-1}} e^{ae^{-bt_{k-1}}}}{e^{ae^{-bt_{k-1}}} - c} \right]^2 \tag{4.41}$$

$$\sigma_{ab} = a \sum_{k=1}^{n} \left[\log \frac{e^{ae^{-bt_{k-1}}}}{e^{ae^{-bt_k}}} \right]^{-1} \times$$

$$\left[\frac{t_k e^{-bt_k} e^{ae^{-bt_k}}}{e^{ae^{-bt_k}} - c} - \frac{t_{k-1} e^{-bt_{k-1}} e^{ae^{-bt_{k-1}}}}{e^{ae^{-bt_{k-1}}} - c} \right] \times$$

$$\left[\frac{e^{-bt_{k-1}} e^{ae^{-bt_{k-1}}}}{e^{ae^{-bt_{k-1}}} - c} - \frac{e^{-bt_k} e^{ae^{-bt_k}}}{e^{ae^{-bt_k}} - c} \right] \tag{4.42}$$

$$\sigma_{ac} = - \sum_{k=1}^{n} \left[\log \frac{e^{ae^{-bt_{k-1}}}}{e^{ae^{-bt_k}}} \right]^{-1} \times$$

$$\left[\frac{e^{-bt_k} e^{ae^{-bt_k}}}{e^{ae^{-bt_k}} - c} - \frac{e^{-bt_{k-1}} e^{ae^{-bt_{k-1}}}}{e^{ae^{-bt_{k-1}}} - c} \right] \times$$

$$\left[\frac{1}{e^{ae^{-bt_k}} - c} - \frac{1}{e^{ae^{-bt_{k-1}}} - c} \right] \tag{4.43}$$

$$\sigma_{bb} = a^2 \sum_{k=1}^{n} \left[\log \frac{e^{ae^{-bt_{k-1}}}}{e^{ae^{-bt_k}}} \right]^{-1} \times$$

$$\left[\frac{t_k e^{-bt_k} e^{ae^{-bt_k}}}{e^{ae^{-bt_k}} - c} - \frac{t_{k-1} e^{-bt_{k-1}} e^{ae^{-bt_{k-1}}}}{e^{ae^{-bt_{k-1}}} - c} \right]^2 \tag{4.44}$$

$$\sigma_{bc} = a \sum_{k=1}^{n} \left[\log \frac{e^{ae^{-bt_{k-1}}}}{e^{ae^{-bt_k}}} \right]^{-1} \times$$

$$\left[\frac{t_k e^{-bt_k} e^{ae^{-bt_k}}}{e^{ae^{-bt_k}} - c} - \frac{t_{k-1} e^{-bt_{k-1}} e^{ae^{-bt_{k-1}}}}{e^{ae^{-bt_{k-1}}} - c} \right] \times$$

$$\left[\frac{1}{e^{ae^{-bt_k}} - c} - \frac{1}{e^{ae^{-bt_{k-1}}} - c} \right] \tag{4.45}$$

$$\sigma_{cc} = a \sum_{k=1}^{n} \left[\log \frac{e^{ae^{-bt_{k-1}}}}{e^{ae^{-bt_k}}} \right]^{-1} \times$$

$$\left[\frac{1}{e^{ae^{-bt_k}} - c} - \frac{1}{e^{ae^{-bt_{k-1}}} - c} \right]^2 \tag{4.46}$$

The invariance property of the MLE's can be used for estimating functions like the number of faults remaining in the system after time t and reliability functions.

4.3 DEVELOPMENT OF MODIFIED YAMADA *et al* MODEL

With the same assumptions as that of the model given by (4.3) and the boundary conditions given by (4.2), we suggest below a modified Yamada *et al* model with $m_4(t)$ given by:

$$m_4(t) = \log\left[\frac{e^a - c}{e^{a(1+bt)e^{-bt}} - c}\right], \qquad a, b > 0 \quad \text{and} \quad 0 \le c < 1. \tag{4.47}$$

Therefore,

$$m_4(t) = \begin{cases} 0, & \text{when} \quad t = 0, \\ \log(\frac{e^a - c}{1-c}), & \text{when} \quad t \to \infty. \end{cases} \tag{4.48}$$

Note that $\log(\dfrac{e^a - c}{1-c})$ is the expected number of faults to be eventually detected and is a finite quantity.

The logic behind introducing this model is when $c = 0$, it is the Yamada model, and when $c = 1$, the corresponding probability density function of the failure time is a proper density which is otherwise an improper one. So the Yamada model is a particular case of this model and we will be looking for estimates of a b, and c that will give us a better estimate of the expected number of faults in a software package under testing than the Yamada model.

FAILURE RATE

The Failure Rate function for this model is given by

$$\begin{aligned} r(t) &= \frac{f(t)}{\bar{F}(t)} \\ &= \frac{ab^2te^{-bt}e^{a(1+bt)e^{-bt}}}{e^{a(1+bt)e^{-bt}} - c}. \end{aligned} \tag{4.49}$$

ERROR DETECTION RATE PER ERROR

The error detection rate per error in model (4.38) is given by

$$\begin{aligned} d_3(t) &= \frac{m_4'(t)}{m_4(\infty) - m_4(t)} \\ &= \frac{ab^2te^{-bt}e^{a(1+bt)e^{-bt}}}{(e^{a(1+bt)e^{-bt}} - c)\log(\frac{e^{a(1+bt)e^{-bt}}-c}{1-c})}. \end{aligned} \tag{4.50}$$

SOFTWARE PERFORMANCE ANALYSIS

1. NUMBER OF FAILURES UP TO TIME t.

Given the parameters $a, b,$ and c the distribution of $N(t)$ is Poisson with mean $m_4(t) = \log[\frac{e^a - c}{e^{a(1+bt)e^{-bt}} - c}]$, i.e.,

$$P[N(t) = y] = \frac{(m(t))^y}{y!}(e^{-m(t)}), \qquad y = 0, 1, 2, \ldots$$

As t tends to infinity, we have

$$P[N(\infty) = y] = \frac{(m(\infty))^y}{y!}(e^{-m(\infty)}), \qquad y = 0, 1, 2, \ldots \tag{4.51}$$

which is the total number of faults to be eventually detected if the system is run indefinitely. It is obvious that $N(\infty)$ is also a Poisson random variable with mean $\log(\frac{e^a - c}{1-c})$.

2. REMAINING ERRORS AFTER DEBUGGING

The remaining number of errors in the system at time t is given by

$$\bar{N}(t) = N(\infty) - N(t). \tag{4.52}$$

Therefore the expected number of faults remaining in the system at time t is

$$\bar{m}(t) = E\bar{N}(t) = \log(\frac{e^{a(1+bt)e^{-bt}} - c}{1-c}). \tag{4.53}$$

Probability distribution of $\bar{N}(t)$ is given by

$$P[\bar{N}(t) = y] = \frac{[\bar{m}_4(t)]^y}{y!}e^{-\bar{m}_4(t)} \tag{4.54}$$

where $\quad y = 0, 1, 2, \ldots$

which is important in deciding whether to release the software package or not.

CONDITIONAL RELIABILITY FUNCTION

The conditional reliability function given that the nth failure occured at time s_n is given by

$$R_2(t|s_n = s) = \frac{e^{a(1+b(s+t))e^{-b(s+t)}} - c}{e^{a(1+bs)e^{-bs}} - c}. \tag{4.55}$$

JOINT DENSITY FUNCTION OF WAITING TIMES

As mentioned earlier, let $\{X_k, \quad k = 1, 2, \ldots\}$ denote a sequence of times between software failures with $S_n = \sum_{k=1}^{n} X_k$.

THEOREM 4. *The joint density function of waiting times $S_1, S_2, \ldots, S_n$ is given by*

$$U_{S_1 \ldots, S_n}(s_1, \ldots, s_n) = \frac{a^n b^{2n} \left(\prod_{k=1}^{n} s_k\right) e^{-b\sum_{1}^{n} s_k} e^{a\sum_{1}^{n}(1+bs_k)e^{-bs_k}}}{\prod_{k=0}^{n-1} \left(e^{a(1+bs_k)e^{-bs_k}} - c\right)}, \tag{4.56}$$

where $\quad a, b > 0, \quad 0 \leq c < 1 \quad$ *and* $\quad 0 = s_0 < s_1 < \ldots < s_n < \infty.$

PROOF: Differentiating $F(t)$, given by (4.13), with respect to t and noting that s_1 is time for the first failure to occur, we obtain the probability density function of S_1 given by

$$U_{S_1}(s_1) = \frac{ab^2 s_1 e^{-bs_1}}{(e^a - c)} e^{a(1+bs_1)e^{-bs_1}}, \qquad 0 < s_1 < \infty \tag{4.57}$$

Conditional density function of a unit, given that it has survived up to time s_1, is found to be

$$\begin{aligned} U_{S_2|S_1}(s_2|s_1) &= U_{S_2|S_1}(x_2 = (s_2 - s_1)|s_1) \\ &= \frac{ab^2 s_2 e^{-bs_2}}{\left(e^{a(1+bs_1)e^{-bs_1}} - c\right)} e^{a(1+bs_2)e^{-bs_2}}. \qquad 0 < s_1 < s_2 < \infty. \end{aligned}$$

Therefore,

$$\begin{aligned} U_{S_1,S_2}(s_1, s_2) &= U_{S_1}(s_1) U_{S_2|S_1}(s_2|s_1) \\ &= \frac{a^2 b^4 e^{-b(s_1+s_2)}}{(e^a - c)(e^{ae^{-bs_1}} - c)} e^{a((1+bs_1)e^{-bs_1} + (1+bs_2)e^{-bs_2})}, \\ &\qquad 0 < s_1 < s_2 < \infty. \end{aligned}$$

Similarly, it can be shown that for n observations $s_1, \ldots, s_n$ the joint density function is

$$U_{S_1,\ldots,S_n}(s_1, \ldots, s_n) = \frac{a^n b^{2n} \left(\prod_{k=1}^{n} s_k\right) e^{-b\sum_1^n s_k} e^{a\sum_1^n (1 + bs_k)e^{-bs_k}}}{\prod_{k=0}^{n-1} \left(e^{a(1+bs_k)e^{-bs_k}} - c\right)}, \tag{4.58}$$

where $a, b > 0$, $0 \le c < 1$ and $0 = s_0 < s_1 < \ldots < s_n < \infty$.□

JOINT DENSITY FUNCTION OF GROUPED DATA

Suppose that the data is obtained in ordered pairs like $(y_k, t_k), k = 1, 2, \ldots, n$; where y_k is the cumulative number of failures up to time t_k. Then the joint probability distribution function of grouped data is given by

$$P_4[N(t_1) = y_1, \ldots, N(t_n) = y_n]$$

$$= \prod_{k=1}^{n} P_4[N(t_k) - N(t_{k-1}) = y_k - y_{k-1}]$$

$$= e^{-m_4(t_n)} \prod_{k=1}^{n} \frac{[m_4(t_k) - m_4(t_{k-1})]^{y_k - y_{k-1}}}{(y_k - y_{k-1})!}$$

$$= \left(\frac{e^{a(1+bt_n)e^{-bt_n}} - c}{e^a - c}\right) \prod_{k=1}^{n} \frac{\left(\log \frac{e^{a(1+bt_{k-1})e^{-bt_{k-1}}} - c}{e^{a(1+bt_k)e^{-bt_k}} - c}\right)^{y_k - y_{k-1}}}{(y_k - y_{k-1})!} \tag{4.59}$$

where $a, b > 0$, $0 \leq c < 1$, and $y_0 = t_0 = 0$.

CASE OF INTER-FAILURE DATA

Given the failure times $s = (s_1, s_2, \ldots, s_n)$ and using equation (4.56), the likelihood function is given by

$$L_7(a,b,c|s) = \frac{a^n b^{2n} (\prod_{k=1}^{n} s_k) e^{-b\sum_{1}^{n} s_k} e^{a\sum_{1}^{n}(1+bs_k)e^{-bs_k}}}{\prod_{k=0}^{n-1} (e^{a(1+bs_k)e^{-bs_k}} - c)} \tag{4.59}$$

Now we want to maximize the likelihood function. It is clear from the functional form of the likelihood function that it is increasing in c. Therefore $\sup_c L_3(a,b,c|s)$ does not exist within the domain of parameter space of c.As such we shall look for a c that will be best with respect to some other criterion. To estimate a and b by the method of ML estimation, let us differentiate (4.59) with respect to a and b separately and equate them to zero. The equations obtained, after some simplification, are as follows

$$\frac{d \log L_7(a,b,c)}{da} = \frac{n}{a} + (1+bs_n)e^{-bs_n} - \frac{e^a}{e^a - c} - c\sum_{k=1}^{n-1} \frac{(1+bs_k)e^{-bs_k}}{(e^{a(1+bs_k)e^{-bs_k}} - c)} = 0 \tag{4.61}$$

and

$$\frac{d\log L_7(a,b,c)}{db} = \frac{2n}{b} - \sum_1^n s_k - abs_n^2 e^{-bs_n} + abc\sum_1^{n-1} \frac{s_k^2 e^{-bs_k}}{\left(e^{a(1+bs_k)e^{-bs_k}} - c\right)} = 0 \tag{4.62}$$

The solutions of these two equations are discussed in Theorem 5. Also the following lemmas are used in proving Theorem 5.

LEMMA 4. *Let*

$$V_b(a) = \frac{n}{a} + (1+bs_n)e^{-bs_n} - \frac{e^a}{e^a - c} - c\sum_{k=1}^{n-1} \frac{(1+bs_k)e^{-bs_k}}{\left(e^{a(1+bs_k)e^{-bs_k}} - c\right)} = 0.$$

Then, $V_b(a)$ is a decreasing function of a for given b.

PROOF: To prove $V_b(a)$ is decreasing in a, let us show that $V_b'(a)$ is negative. Differentiating $V_b(a)$ with respect to a we have

$$\begin{aligned} V_b'(a) &= -\frac{n}{a^2} + \frac{ce^a}{(e^a - c)^2} + c\sum_{k=1}^{n-1} \frac{(1+bs_k)^2 e^{-2bs_k} e^{ae^{-bs_k}}}{\left(e^{a(1+bs_k)e^{-bs_k}} - c\right)^2} \\ &= \left[-\frac{1}{a^2} + \frac{ce^a}{(e^a - c)^2}\right] + \sum_{k=1}^{n-1}\left[-\frac{1}{a^2} + \frac{c(1+bs_k)^2 e^{-2bs_k} e^{ae^{-bs_k}}}{\left(e^{a(1+bs_k)e^{-bs_k}} - c\right)^2}\right]. \end{aligned}$$

Combining the terms, we have

$$V_b'(a) = \frac{ca^2e^a - e^{2a} + 2ce^a - c^2}{a^2(e^a - c)^2} +$$

$$\sum_{k=1}^{n-1} \frac{c\left(a(1+bs_k)e^{-bs_k}\right)^2 e^{ae^{-bs_k}} - e^{2a(1+bs_k)e^{-bs_k}} + 2ce^{a(1+bs_k)e^{-bs_k}} - c^2}{a^2\left(e^{a(1+bs_k)e^{-bs_k}} - c\right)^2} \tag{4.63}$$

Equation (4.63) is similar to equation (4.23) of section 4.2, except that e^{-bs_k} of (4.23) is being replaced by $(1+bs_k)e^{-bs_k}$ in (4.63). But the aforesaid expression does not involve the parameter a. Therefore, on the same lines as in Lemma 2, we can prove that $V_b'(a)$ is negative. □

LEMMA 5. *The solution bounds for a in the equation $V_b(a) = 0$ are given by*

$$0 < a < n[1-(1+bs_n)e^{-bs_n}]^{-1}. \tag{4.64}$$

PROOF: Clearly,

$$\frac{e^a}{(e^a-c)} + c\sum_{k=1}^{n-1} \frac{e^{-bs_k}}{(e^{ae^{-bs_k}}-c)}$$

is decreasing in a and further more

$$\lim_{a\to\infty}\left[\frac{e^a}{(e^a-c)} + c\sum_{k=1}^{n-1} \frac{ce^{-bs_k}}{(e^{ae^{-bs_k}}-c)}\right] = 1.$$

Therefore,

$$V_b(a) \le \frac{n}{a} + (1+bs_n)e^{-bs_n} - \lim_{a\to\infty}\left[\frac{e^a}{e^a-c} + c\sum_{k=1}^{n-1} \frac{e^{-bs_k}}{(e^{ae^{-bs_k}}-c)}\right]$$

$$= \frac{n}{a} + (1+bs_n)e^{-bs_n} - 1. \tag{4.65}$$

Equation (4.65) implies that

$$V_b(a) \le 0 \quad \text{if} \quad \frac{n}{a} + (1 + bs_n)e^{-bs_n} - 1 \le 0,$$

$$\Longrightarrow V_b(a) \le 0 \quad \text{if} \quad a \ge n[1 - (1 + bs_n)e^{-bs_n}]^{-1}. \tag{4.66}$$

Also we have

$$\lim_{a\to 0} V_b(a) = \lim_{a\to 0}\left[\frac{n}{a} + (1 + bs_n)e^{-bs_n} - \frac{e^a}{e^a - c} - c\sum_{k=1}^{n-1}\frac{(1 + bs_k)e^{-bs_k}}{\left(e^{a(1+bs_k)e^{-bs_k}} - c\right)}\right]$$
$$= \infty.$$

$$\Longrightarrow V_b(a) \ge 0 \qquad \text{as} \qquad a \to 0. \tag{4.67}$$

The inequalities (4.66), (4.67), and Lemma 4 prove that for a given b the root a of the equation $V_b(a) = 0$ lies in the interval 0 to $n[1-(1+bs_n)e^{-bs_n}]^{-1}$ and it is unique. □

LEMMA 6. *Let*

$$\phi(b) = c\sum_{1}^{n-1}\frac{\frac{nbs_k^2e^{-bs_k}}{1-(1+bs_n)e^{-bs_n}}}{\left[e^{\frac{n(1+bs_k)e^{-bs_k}}{1-(1+bs_n)e^{-bs_n}}} - c\right]}.$$

Then $\lim_{b\to 0}\phi(b) = 0$.

PROOF: Let

$$\phi_1(b) = n(1+bs_k)e^{-bs_k}, \Longrightarrow \quad \phi_1(0) = n$$

then,

$$\phi_1'(b) = -nbs_k^2 e^{-bs_k}, \quad \Longrightarrow \phi_1'(0) = 0$$

$$\phi_1''(b) = -ns_k^2 e^{-bs_k} + nbs_k^3 e^{-bs_k} \quad \Longrightarrow \phi_1''(0) = -ns_k^2$$

$$\phi_1'''(b) = -nbs_k^4 e^{-bs_k} \quad \Longrightarrow \phi_1'''(0) = 0,$$

and,

$$\phi_2(b) = 1 - (1+bs_n)e^{-bs_n}, \Longrightarrow \quad \phi_2(0) = 0$$

then,

$$\phi_2'(b) = bs_n^2 e^{-bs_n}, \quad \Longrightarrow \phi_2'(0) = 0$$

$$\phi_2''(b) = s_n^2 e^{-bs_n} - bs_n^3 e^{-bs_n} \quad \Longrightarrow \phi_2''(0) = s_n^2.$$

$$\lim_{b\to 0}\phi(b) = \lim_{b\to 0}\frac{\frac{\phi_1'(b)}{\phi_2(b)}}{e^{\frac{\phi_1(b)}{\phi_2(b)}} - c}, \qquad \text{is of the form } \frac{0}{0}$$

$$= \lim_{b\to 0}\frac{\phi_1''(b)\phi_2(b) - \phi_1'(b)\phi_2'(b)}{(\phi_1'(b)\phi_2(b) - \phi_1(b)\phi_2'(b))e^{\frac{\phi_1(b)}{\phi_2(b)}}}, \qquad \text{is of the form } \frac{0}{0}$$

$$= \lim_{b\to 0}\frac{\phi_1''(b)\phi_2(b) - \phi_1'(b)\phi_2'(b)}{(\phi_1'(b)\phi_2(b) - \phi_1(b)\phi_2'(b))}\cdot\lim_{b\to 0} e^{\frac{\phi_1(b)}{\phi_2(b)}}$$

$$= \lim_{b\to 0}\frac{\phi_1'''(b)\phi_2(b) - \phi_1'(b)\phi_2''(b)}{(\phi_1''(b)\phi_2(b) - \phi_1(b)\phi_2''(b))}\cdot\lim_{b\to 0} e^{\frac{\phi_1(b)}{\phi_2(b)}}$$

$$= \frac{0}{-s_n^2}$$

$$= 0.$$

□

THEOREM 5. *The sufficient condition for the equations (4.61) and (4.62) to have finite positive roots is given by*

$$s_n > \frac{3}{2n}\sum_{k=1}^{n} s_k.$$

PROOF: Let us denote the right hand side expression of equation (4.62) by $W_a(b)$. Now we need to determine the sufficient condition for $V_b(a)$ and $W_a(b)$ to have finite zeros. Clearly, $W_a(b)$ is decreasing in a. To get the sufficient condition for the existence of finite zeros, we need to determine $\inf_a\{W_a(b)\}$ and $\sup_a\{W_a(b)\}$.

$$\inf_a W_a(b) = \inf_a \left[\frac{2n}{b} - \sum_1^n s_k - abs_n^2 e^{-bs_n} + abc\sum_1^{n-1} \frac{s_k^2 e^{-bs_k}}{\left(e^{a(1+bs_k)e^{-bs_k}} - c\right)}\right].$$

On using Lemma 4, we get

$$\inf_a W_a(b) > \frac{2n}{b} - \sum_1^n s_k - \frac{nbs_n^2}{e^{bs_n} - bs_n - 1} + c\sum_1^{n-1} \frac{\frac{nbs_k^2 e^{-bs_k}}{1-(1+bs_n)e^{-bs_n}}}{\left(e^{\frac{n(1+bs_k)e^{-bs_k}}{1-(1+bs_n)e^{-bs_n}}} - c\right)}$$

$$\Longrightarrow \lim_{b\to 0}\inf_a W_a(b) > \lim_{b\to 0}\left[\frac{2n}{b} - \sum_1^n s_k - \frac{nbs_n^2}{e^{bs_n} - bs_n - 1} + \right.$$
$$\left. c\sum_1^{n-1} \frac{\frac{nbs_k^2 e^{-bs_k}}{1-(1+bs_n)e^{-bs_n}}}{\left(e^{\frac{n(1+bs_k)e^{-bs_k}}{1-(1+bs_n)e^{-bs_n}}} - c\right)}\right].$$

After using Lemma 6, we get

$$\lim_{b\to 0}\inf_a W_a(b) > -\sum_1^n s_k + \lim_{b\to 0}\frac{2ns_ne^{bs_n} - 2ns_n - 2ns_n^2 b}{bs_ne^{bs_n} + e^{bs_n} - 2bs_n - 1} + 0$$
$$= -\sum_1^n s_k + \lim_{b\to 0}\frac{2ns_n^2e^{bs_n} - 2ns_n^2}{2s_ne^{bs_n} + bs_n^2e^{bs_n} - 2s_n}$$
$$= -\sum_1^n s_k + \frac{2}{3}ns_n. \tag{4.68}$$

$$\text{Again,}\quad \sup_a W_a(b) < \frac{2n}{b} - \sum_1^n s_k + bc\lim_{a\to 0}\sum_1^{n-1}\frac{as_k^2e^{-bs_k}}{\left(e^{a(1+bs_k)e^{-bs_k}} - c\right)}$$
$$= \frac{2n}{b} - \sum_1^n s_k$$
$$\Longrightarrow \lim_{b\to\infty}\sup_a W_a(b) < -\sum_{k=1}^n s_k, \tag{4.69}$$

Since $\lim\limits_{b\to\infty}\sup\limits_a g_a(b)$ is negative we will have finite zeros if $\lim\limits_{b\to 0}\inf\limits_a g_a(b)$ is positive. By (3.71), $\lim\limits_{b\to 0}\inf\limits_a g_a(b) > -\sum_1^n s_k + n\frac{s_n}{2}$. Therefore equations (4.60) and (4.61) will have finite roots if $s_n > \frac{3}{2n}\sum_{k=1}^n s_n$. □

Therefore, under the sufficient condition for given c, we can, for sure, obtain the maximum likelihood estimates of a and b. Since this is only the sufficient condition, we may expect to get the solutions without this condition being satisfied.

CASE OF GROUPED DATA

Let $(y_k, t_k), k = 1, \ldots, n$ be the observed ordered pairs of data. On using equation (4.60), the likelihood function for group data is found to be

$$L_8\{a, b, c|\tilde{y}, \tilde{t}\} = \left(\frac{e^{a(1+bt_n)e^{-bt_n}} - c}{e^a - c}\right)$$

$$\prod_{k=1}^{n} \frac{\left(\log \frac{e^{a(1+bt_{k-1})e^{-bt_{k-1}}} - c}{e^{a(1+bt_k)e^{-bt_k}} - c}\right)^{y_k - y_{k-1}}}{(y_k - y_{k-1})!}. \tag{4.70}$$

On taking partial derivatives of the log likelihood function given by equation (4.70) with respect to a, b, and c separately and equating them to zero, we obtain the following equations, after some simplification

$$
\begin{aligned}
\frac{d\log L_8(a,b,c)}{da} =& \frac{(1+bt_n)e^{-bt_n}e^{a(1+bt_n)e^{-bt_n}}}{e^{a(1+bt_n)e^{-bt_n}}-c} - \frac{e^a}{e^a-c} + \\
& \sum_{k=1}^{n} \frac{(y_k-y_{k-1})(1+bt_{k-1})e^{-bt_{k-1}}e^{a(1+bt_{k-1})e^{-bt_{k-1}}}}{\left(e^{a(1+bt_{k-1})e^{-bt_{k-1}}}-c\right)\log\frac{e^{a(1+bt_{k-1})e^{-bt_{k-1}}}-c}{e^{a(1+bt_k)e^{-bt_k}}-c}} - \\
& \sum_{k=1}^{n} \frac{(y_k-y_{k-1})(1+bt_k)e^{-bt_k}e^{a(1+bt_k)e^{-bt_k}}}{\left(e^{a(1+bt_k)e^{-bt_k}}-c\right)\log\frac{e^{a(1+bt_{k-1})e^{-bt_{k-1}}}-c}{e^{a(1+bt_k)e^{-bt_k}}-c}} \\
=&0,
\end{aligned}
\tag{4.71}
$$

$$
\begin{aligned}
\frac{d\log L_8(a,b,c)}{db} =& -abt_n^2\frac{e^{-bt_n}e^{a(1+bt_n)e^{-bt_n}}}{e^{a(1+bt_n)e^{-bt_n}}-c} - \\
& \sum_{k=1}^{n} \frac{(y_k-y_{k-1})abt_{k-1}^2e^{-bt_{k-1}}e^{a(1+bt_{k-1})e^{-bt_{k-1}}}}{\left(e^{a(1+btk-1)e^{-bt_{k-1}}}-c\right)\log\frac{e^{a(1+bt_{k-1})e^{-bt_{k-1}}}-c}{e^{a(1+bt_k)e^{-bt_k}}-c}} + \\
& \sum_{k=1}^{n} \frac{(y_k-y_{k-1})abt_k^2e^{-bt_k}e^{a(1+bt_k)e^{-bt_k}}}{\left(e^{a(1+bt_k)e^{-bt_k}}-c\right)\log\frac{e^{a(1+bt_{k-1})e^{-bt_{k-1}}}-c}{e^{a(1+bt_k)e^{-bt_k}}-c}} \\
=&0,
\end{aligned}
\tag{4.72}
$$

and

$$
\begin{aligned}
\frac{d\log L_8(a,b,c)}{dc} &= -\frac{1}{e^{a(1+bt_n)e^{-bt_n}}-c}+\frac{1}{e^{a}-c}-\\
&\sum_{k=1}^{n}\frac{(y_k-y_{k-1})}{\left(e^{a(1+bt_{k-1})e^{-bt_{k-1}}}-c\right)\log\left(\frac{e^{a(1+bt_{k-1})e^{-bt_{k-1}}}-c}{e^{a(1+bt_k)e^{-bt_k}}-c}\right)}+\\
&\sum_{k=1}^{n}\frac{(y_k-y_{k-1})}{\left(e^{a(1+bt_k)e^{-bt_k}}-c\right)\log\left(\frac{e^{a(1+bt_{k-1})e^{-bt_{k-1}}}-c}{e^{a(1+bt_k)e^{-bt_k}}-c}\right)}\\
&= 0. \qquad (4.73)
\end{aligned}
$$

It is clear that the three equations do not yield simple analytical forms for the solution of a, b and c. So we need to resort to numerical methods for their solution.

5. APPLICATION OF THE MODELS

5.1 Analysis of Failure Data from Naval Tactical Data System (NTDS) and Comparison between G-O and Modified G-O models.

In this section we analyze a set of data from the Naval Fleet Computer Programming Center. This data set has been used by several investigators for model validation purposes. It was extracted from information about errors in the development of software for the real-time, multi-computer complex which forms the core of the NTDS. The software consisted of some 38 different project schedules. Each module was supposed to follow three stages: the production phase, the test phase, and the user phase. The times (in days) between failures are shown in Table 5.1. Twenty-six software errors were found during the production phase and five additional errors during the test phase. The last error was found on January 4, 1971. One error was observed

Table No.5.1 **Software Failure Data from NTDS.**

Error No. k	Inter Failure Times x_k, (days)	Cumulative Time $s_k = \sum_{i=1}^{k} x_i$, (days)
Production Phase		
1	9	9
2	12	21
3	11	32
4	4	36
5	7	43
6	2	45
7	5	50
8	8	58
9	5	63
10	7	70
11	1	71
12	6	77
13	1	78
14	9	87
15	4	91
16	1	92
17	3	95
18	3	98
19	6	104
20	1	105
21	11	116
22	33	149
23	7	156
24	91	247
25	2	249
26	1	250
Test Phase		
27	87	337
28	47	384
29	12	396
30	9	405
31	135	540
User Phase		
32	258	798
Test Phase		
33	16	814
34	35	849

during the user phase in September, 1971. Again two errors were observed during the test phase in 1971.

1. DATA ANALYSIS

Our data is available as cumulative time to failures. Therefore, we will use the methods described in Section (4.2) and will consider only the first 26 observations of Table 5.1 to estimate the parameters.

Solving the simultaneous equations given by equations (4.20) and (4.21) for different values of c, we obtain the Table No.5.2. Using the criteria, as suggested in Section (4.2), the estimates are $\hat{a} = 33.951$, $\hat{b} = 0.005804$ and $\hat{c} = 0.99537$.

The fitted mean value function is

$$\hat{m}_3(t) = \log\left(\frac{e^{33.951} - .99537}{e^{33.951e^{-.005804t}} - .99537}\right), \qquad t \geq 0 \tag{5.1}$$

and is shown in Fig.5.2 along with the actual data. Therefore, the expected number of faults, $\hat{m}_3(\infty)$, to be eventually detected is 39.33.

Table No.5.2 Mean Sum of Squares of Deviations of Observed and Estimated Time to Failure for Different Values of c.

c	$\frac{1}{n}\sum_{k=1}^{n}(s_k - \hat{s}_k)^2$
0.0000000	533.58
0.8750000	532.74
0.9629630	532.66
0.9843750	532.63
0.9920000	532.63
0.9953704	532.62
0.9970845	532.62
0.9980469	532.62
0.9986283	532.62
0.9990000	532.62
0.9992487	532.62
0.9994213	532.62
0.9995448	532.62
0.9996356	532.62
0.9997037	532.62
0.9999987	532.62
0.9999987	532.62
0.9999988	532.62
0.9999988	532.62
0.9999988	532.62
0.9999989	532.62
0.9999989	532.62
0.9999989	532.62
0.9999990	532.62
1.0000000	532.62

2. TEST OF GOODNESS OF FIT

We will use the Kolmogorov-Smirnov goodness-of-fit test to check the adequacy of the fitted model. Before we do that, let us give the scenario of the test.

Suppose that $0 \leq S_1 \leq, \ldots, \leq S_n$ are the random times at which the first n events occur in a NHPP with an unknown mean value function $m(t)$. To test the simple hypothesis

$$H_0 : m(t) = m_0(t), \quad \text{for} \quad t \geq 0,$$

versus

$$H_1 : m(t) \neq m_0(t), \quad \text{for} \quad t \geq 0, \tag{5.2}$$

where, $m_0(t)$ stands for $\log(\frac{e^{a_0} - c_0}{e^{a_0 e^{-b_0 t}} - c_0})$, we need the joint conditional distribution of the failure times. The following theorem is used in deriving this distribution. We will give only the statement of the theorem. For proof see Cox and Lewis (1966).

THEOREM 5.1. *Given $s_n = t$, the $n-1$ failure times $0 \leq S_1 \leq, \ldots, \leq S_n$ have the same joint conditional distribution as the order statistics of a random sample of size $n-1$ from the distribution $G(x) = \dfrac{m(x)}{m(t)}, \quad 0 \leq x \leq t$.*

Therefore, the hypothesis boils down to testing the following

$$H_0 : G_0(x) = \frac{m_0(x)}{m_0(t)}, \qquad for \quad 0 \leq x \leq t. \tag{5.3}$$

Given the random sample $s_1, \ldots, s_n$ of size n, the cdf is defined by

$$H_{n-1}(x) = \begin{cases} 0, & \text{if } x < s_1 \\ \frac{k}{n-1}, & \text{if } s_{k-1} \leq x < s_k, \\ 1, & \text{if } x \geq s_{n-1}. \end{cases} \qquad k = 2, 3, \ldots, n-1. \tag{5.4}$$

Kolmogorov-Smirnov's D statistics is defined as

$$D = max_k\{D_k\}. \tag{5.5}$$

where

$$D_k = max\{|G_0(s_k) - \frac{k}{n-1}|, |G_0(s_k) - \frac{k-1}{n-1}|\}. \tag{5.6}$$

Now, if the calculated D is greater than or equal to the critical value $D_{n-1;\alpha}$, we either reject the null hypothesis or else we accept it.

We can also find the confidence limits for $G(x)$. The $100(1-\alpha)\%$ confidence limits for $G(x)$ are given by

$$H_{n-1}(x) - D_{n-1,\alpha} < G(x) < H_{n-1}(x) + D_{n-1,\alpha}. \tag{5.7}$$

Now let us perform the Kolmogorov-Smirnov goodness-of-fit test to test the adequacy of the fitted model for NTDS data. The test is based on 25 observations. The null hypothesis is

$$H_0 : G_0(x) = \frac{\log(e^{33.951} - .99537) - \log(e^{33.951e^{-.005804x}} - .99537)}{\log(e^{33.951} - .99537) - \log(e^{33.951e^{-.005804\times 250}} - .99537)} \tag{5.8}$$

and sample cdf is

$$H(x) = \begin{cases} 0, & \text{if } x < s_1 \\ \frac{k}{25}, & \text{if } s_{k-1} \leq x < s_k, \qquad k = 2, 3, \ldots, 25 \\ 1, & \text{if } x \geq s_{25}. \end{cases} \tag{5.9}$$

The necessary calculations for D statistic, for various values of s_k, are shown in Table No.5.3. D is found to be

$$D = 0.2040.$$

The critical value of $D_{25,0.05}$ at 5% level of significance is

$$D_{25,0.05} = 0.264.$$

Since $D < D_{25,0.05}$ we do not reject the null-hypothesis at 5% level of significance. Note that the $D = 0.2044$ for the G-O model, which is an indication that Modified G-O model fitted to the data better than the G-O model.

Table No. 5.3 **Kolmogorov-Smirnov Test For The NTDS Data Set.**

$H(s_k)$	$G_0(s_k)$	$\|H(s_k) - G_0(s_k)\|$	$\|H(s_{k-1}) - G_0(s_k)\|$
0.040	0.0665	0.0265	0.0665
0.080	0.1499	0.0699	0.1099
0.120	0.2214	0.1014	0.1414
0.160	0.2463	0.0863	0.1263
0.200	0.2885	0.0885	0.1285
0.240	0.3002	0.0602	0.1002
0.280	0.3290	0.0490	0.0890
0.320	0.3733	0.0533	0.0933
0.360	0.4000	0.0400	0.0800
0.400	0.4361	0.0361	0.0761
0.440	0.4411	0.0011	0.0411
0.480	0.4707	0.0093	0.0307
0.520	0.4755	0.0445	0.0045
0.560	0.5178	0.0422	0.0022
0.600	0.5359	0.0641	0.0241
0.640	0.5403	0.0997	0.0597
0.680	0.5536	0.1264	0.0864
0.720	0.5665	0.1535	0.1135
0.760	0.5919	0.1681	0.1281
0.800	0.5960	0.2040	0.1640
0.840	0.6399	0.2001	0.1601
0.880	0.7560	0.1240	0.0840
0.920	0.7779	0.1421	0.1021
0.960	0.9946	0.0346	0.0746
1.000	0.9982	0.0018	0.0382

The 95% confidence limits for $G(x)$ can now be calculated with $\alpha = .05$ and $D_{25,.05} = 0.264$. The lower and upper confidence bounds are

$$L(x) = max\{H(x) - 0.264, 0\}$$

and

$$U(x) = min\{H(x) - 0.264, 1\}. \tag{5.10}$$

The 95% confidence bounds for $G_0(x)$, and $G(x)$, are shown in Fig.5.1.

3. COMPARISON WITH G-O MODEL

The maximum likelihood estimates for the parameters of G-O model are $\hat{a} = 33.99$ and $\hat{b} = .00579$. The estimated expected number of errors by time t is

$$\hat{m}_1(t) = 33.99(1 - e^{-0.00579t}). \tag{5.11}$$

Therefore an estimate of the expected number of faults to be eventually detected is 33.99 and that of the modified model is 39.33.

The mean time to failure (MTTF) for the processes do not exist since the inter-failure times have improper density function. Therefore we use the inverse transformation of the mean value function to get the

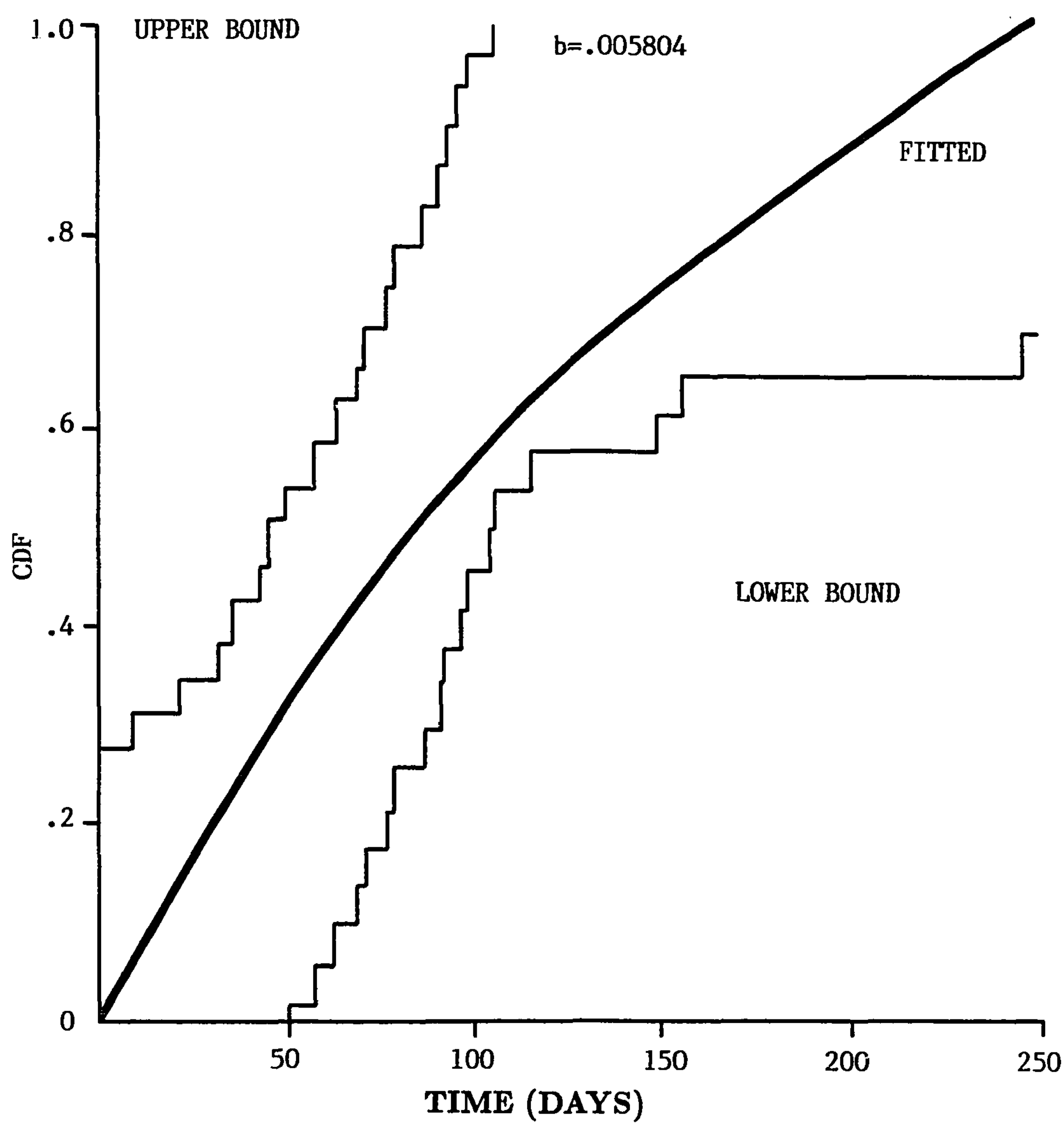

Fig. 5.1 95% Confidence Bounds for CDF of G(x)

and Fitted CDF Curve.

estimate of time to kth failure. The inverse functions for G-O and Modified G-O are respectively

$$\begin{aligned}\hat{s}_k &= -\frac{1}{\hat{b}}\log(1-\frac{k}{\hat{a}})\\ &= -\frac{1}{.00579}\log(1-\frac{k}{33.99}) \quad \text{and} \qquad (5.12)\\ \hat{s}_k &= -\frac{1}{\hat{b}}\log\frac{1}{\hat{a}}\log[(e^{\hat{a}}-\hat{c})e^{-k}+\hat{c}]\\ &= -\frac{1}{.005804}\log\frac{1}{33.951}\log[(e^{33.951}-.99537)e^{-k}+.99537]. \qquad (5.13)\end{aligned}$$

The observed and estimated s_k's are given in Table No.5.4. The criteria for comparing the results of the two models will be the sum of squares of differences (SSD) between the actual and the estimated s_k's. Thus, we have

$$Fit(SSD) = \sum_{k=1}^{26}(s_k - \hat{s}_k)^2 \qquad (5.14)$$

$$\text{and} \quad Prediction(SSD) = \sum_{k=27}^{34}(s_k - \hat{s}_k)^2. \qquad (5.15)$$

For G-O model the Fit (SSD)=13873.10 and Prediction (SSD)=4277.24. In the case of Modified G-O model the Fit(SSD)=13848.22 and Prediction(SSD)=6923.05. Clearly, Modified shows a better fit. For prediction G-O model seems to be doing better in this example.

4. DISTRIBUTION OF FAULTS REMAINING AFTER DEBUGGING

Using the probability distribution of the remaining faults, as described

Table No.5.4 Comparison of Results Based on the Goel-Okumoto and Modified Goel-Okumoto models.

Error No. k	Actual Failure Time s_k, (days)	Estimated Failure Time $\hat{s}_k$ G-O	Modified G-O
1	9	5.16	5.15
2	21	10.47	10.46
3	32	15.96	15.94
4	36	21.62	21.60
5	43	27.48	27.44
6	45	33.54	33.51
7	50	39.82	39.78
8	58	46.34	46.30
9	63	53.12	53.07
10	70	60.17	60.12
11	71	67.52	67.46
12	77	75.20	75.14
13	78	83.24	83.17
14	87	91.67	91.60
15	91	100.53	100.46
16	92	109.87	109.80
17	95	119.74	119.68
18	98	130.22	130.15
19	104	141.37	141.31
20	105	153.29	153.24
21	116	166.10	166.05
22	149	179.93	179.90
23	156	194.96	194.95
24	247	211.43	211.45
25	249	229.64	229.70
26	250	250.00	250.10

in Section (4.1) by equation (4.10), we have

$$P_3[\bar{N}(t) = k] = \frac{\left[ln \frac{e^{33.951e^{-.005804t}} - .99537}{1-.99537}\right]^k}{k!} \left(\frac{1 - .99537}{e^{33.951e^{-.005804t}} - .99537} \right),$$

where $k = 0, 1, 2, \ldots$ (5.16)

Table No.5.5 Distribution of Remaining Errors After Debugging.

k	$P_3[\bar{N}(250) \le k]$
1	0.00002
2	0.00017
3	0.00081
4	0.00295
5	0.00864
6	0.02130
7	0.04540
8	0.08557
9	0.14506
10	0.22436
11	0.32047
12	0.42724
13	0.53673
14	0.64098
15	0.73363
16	0.81083
17	0.87136
18	0.91619
19	0.94764
20	0.96861
21	0.98192
22	0.98998
23	0.99466
24	0.99725
25	0.99864

From this table it is clear that the $P_3(\bar{N}(250) \le 23) = 0.9947$. Therefore, the probability that twenty-three or less number of faults remained in the system after observing for 250 days is about 0.99.

by equation (4.10) in Section (4.1), we have

$$P_3[\bar{N}(t) = k] = \frac{[\log \frac{e^{33.951e^{-.005804t}} - .99537}{1-.99537}]^k}{k!}\left(\frac{1-.99537}{e^{33.951e^{-.005804t}} - .99537}\right),$$

where $k = 0, 1, 2, \ldots$ (5.16)

An estimate of the cumulative distribution function of $\bar{N}(t)$ for the NTDS example is given below in Table No.5.5.

k	1	2	3	4	5	6	7	8
$P_r[\bar{N}(t) \leq k]$	.0055	.0233	.0667	.1461	.2622	.4038	.5518	.6871

k	9	10	11	12	13	14	15	16
$P_r[\bar{N}(t) \leq k]$	.7971	.8776	.9311	.9637	.9821	.9917	.9963	.9984

From this table it is clear that the $P_3(\bar{N}(250) \leq 23) = 0.9947$. Therefore, the probability that twenty-three or a lower number of faults remained in the system after being observed for 250 days is about 0.99.

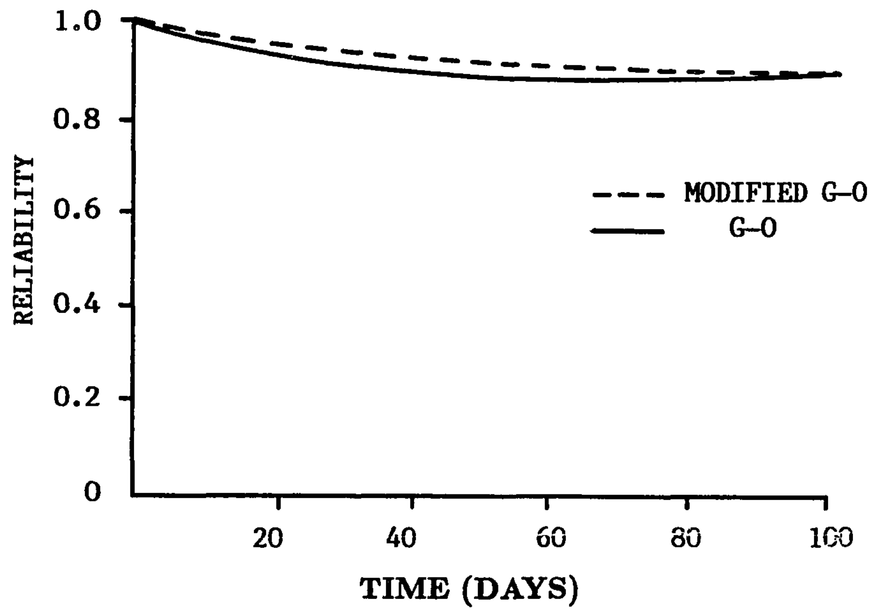

Fig.5.3 Plots of Conditional Reliability Functions of G-O & Modified G-O Models.

The estimated conditional reliability function for the Modified G-O model is

$$\hat{R}_1(t|s_{26} = 250) = \frac{e^{33.951e^{-.005804(250+t)}} - .99537}{e^{33.951e^{-.005804\times 250}} - .99537}. \tag{5.17}$$

and that of the G-O model is

$$\hat{R}(t|s_{26} = 250) = e^{33.99(e^{-.00579\times 250} - e^{-.00579(250+t)})}. \tag{5.18}$$

The plot of these reliability functions versus time are shown in Fig.5.3 along with the reliability function after $n = 31$ errors. As expected the reliability after $n = 31$ is monotonically higher than that after $n = 26$.

5.2 Analysis of Failure Data from Project 1, Musa (1975).

In this section we analyze a set of data from Project 1, Musa (1975). This data set has been used by several investigators for model validation purposes. The set has 136 observations with execution times between successive failures in seconds.

1. <u>DATA ANALYSIS</u>

We will consider all of the 136 observations of Table 5.6 to estimate the parameters. Solving the simultaneous equations given by equations (4.20) and (4.21) for different values of c, we obtain the Table No.5.7. The estimates of the parameters are $\hat{a} = 142.815$, $\hat{b} = 0.000034$, and $\hat{c} = 0.999997$.

The fitted mean value function is

$$\hat{m}_3(t) = \log\left(\frac{e^{142.815} - .999997}{e^{142.815e^{-.000034t}} - .999997}\right), \qquad t \geq 0, \tag{5.19}$$

and is shown in Fig.5.5 along with the actual data. The expected number of faults, $\hat{m}_3(\infty)$, to be eventually detected is 155.52.

Table No.5.6 Software Failure Data from Project 1,. Musa (1975).

Error No. k	Inter Failure Times x_k, (days)	Cumulative Time $s_k = \sum_{i=1}^{k} x_i$, (days)
1	3	3
2	30	33
3	113	146
4	81	227
5	115	342
6	9	351
7	2	353
8	91	444
9	112	556
10	15	571
11	138	709
12	50	759
13	77	836
14	24	860
15	108	968
16	33	1001
17	670	1671
18	120	1791
19	26	1817
20	114	1931
21	325	2256
22	55	2311
23	242	2553
24	68	2621
25	422	3043
26	180	3223
27	10	3233
28	1146	4379
29	600	4979
30	15	4994
31	36	5030
32	4	5034
33	0	5034
34	8	5042
35	227	5269
36	65	5334
37	176	5510
38	58	5568

Table No.5.6 continued.

Error No. k	Inter Failure Times x_k, (days)	Cumulative Time $s_k = \sum_{i=1}^{k} x_i$, (days)
39	457	6025
40	300	6325
41	97	6422
42	263	6685
43	452	7137
44	255	7392
45	197	7589
46	193	7782
47	6	7788
48	79	7867
49	816	8683
50	1351	10034
51	143	10177
52	21	10198
53	233	10431
54	134	10565
55	357	10922
56	193	11115
57	236	11351
58	31	11382
59	369	11751
60	748	12499
61	0	12499
62	232	12731
63	330	13061
64	365	13426
65	1222	14648
66	543	15191
67	10	15201
68	16	15217
69	529	15746
70	379	16125
71	44	16169
72	129	16298
73	810	17108
74	290	17398
75	300	17698
76	529	18227

Table No.5.6 continued.

Error No. k	Inter Failure Times x_k, (days)	Cumulative Time $s_k = \sum_{i=1}^{k} x_i$, (days)
77	281	18508
78	160	18668
79	828	19496
80	1011	20507
81	445	20952
82	296	21248
83	1755	23003
84	1064	24067
85	1783	25850
86	860	26710
87	983	27693
88	707	28400
89	33	28433
90	868	29301
91	724	30025
92	2323	32348
93	2930	35278
94	1461	36739
95	843	37582
96	12	37594
97	261	37855
98	1800	39655
99	865	40520
100	1435	41955
101	30	41985
102	143	42128
103	108	42236
104	0	42236
105	3110	45346
106	1247	46593
107	943	47536
108	700	48236
109	875	49111
110	245	49356
111	729	50085
112	1897	51982
113	447	52429
114	386	52815

Table No.5.6 continued.

Error No. k	Inter Failure Times x_k, (days)	Cumulative Time $s_k = \sum_{i=1}^{k} x_i$, (days)
115	446	53261
116	122	53383
117	990	54373
118	948	55321
119	1082	56403
120	22	56425
121	75	56500
122	482	56982
123	5509	62491
124	100	62591
125	10	62601
126	1071	63672
127	371	64043
128	790	64833
129	6150	70983
130	3321	74304
131	1045	75349
132	648	75997
133	5485	81482
134	1160	82642
135	1864	84506
136	4116	88622

Table No.5.7 Mean Sum of Squares of Deviations of Observed and Estimated Time to Failure for Different Values of c.

c	$\frac{1}{n}\sum_{k=1}^{n}(s_k - \hat{s}_k)^2$
0.0000000	13210573.07
0.9920000	13208540.07
0.9986283	13208526.51
0.9995448	13208524.64
0.9997965	13208524.12
0.9998920	13208523.93
0.9999360	13208523.84
0.9999590	13208523.79
0.9999722	13208523.76
0.9999803	13208523.75
0.9999855	13208523.74
0.9999890	13208523.73
0.9999915	13208523.73
0.9999933	13208523.72
0.9999946	13208523.72
0.9999956	13208523.72
0.9999964	13208523.72
0.9999970	13208523.71
0.9999974	13208523.71
0.9999978	13208523.71
0.9999981	13208523.71
0.9999984	13208523.71
0.9999986	13208523.71
0.9999988	13208523.71
1.0000000	13208523.71

2. TEST OF GOODNESS OF FIT

We will use the Kolmogorov-Smirnov goodness-of-fit test to check the adequacy of the fitted model. As described in Section (5.1), the test is based on 135 observations. The null hypothesis is

$$H_0 : G_0(x) = \frac{\log(e^{142.815} - .999997) - \log(e^{142.815e^{-.000034x}} - .999997)}{\log(e^{142.815} - .999997) - \log(e^{155.52e^{-.000034\times 88622}} - .999997)} \tag{5.20}$$

and the sample cdf is

$$H(x) = \begin{cases} 0, & \text{if } x < s_1 \\ \frac{k}{135}, & \text{if } s_{k-1} \leq x < s_k, \qquad k = 2, 3, \ldots, 135 \\ 1, & \text{if } x \geq s_{135}. \end{cases} \tag{5.21}$$

The necessary calculations for D statistic defined in (5.5), for various values of s_k, are shown in Table 5.8 and its value is given by $D = 0.1087$. The critical value, $D_{135,0.05}$, at 5% level of significance is given by

$$D_{135,0.05} = 0.11705.$$

Since $D < D_{135,0.2}$ we do not reject the null- hypothesis. Note that in the case of the G-O model D is found to be $D = 0.1087$.

The 95% confidence limits for $G(x)$ can now be calculated with $\alpha = .05$ and $D_{135,.05} = 0.11705$. The lower and upper confidence bounds are

Table No. 5.8 **Kolmogorov-Smirnov Test For Data Set of Project 1.**

$H(s_k)$	$G_O(s_k)$	$\lvert H(s_k) - G_C(s_k)\rvert$	$\lvert H(s_{k-1}) - G_O(s_k)\rvert$
0.007	0.0001	0.0073	0.0001
0.015	0.0012	0.0136	0.0062
0.022	0.0052	0.0170	0.0096
0.030	0.0081	0.0215	0.0141
0.037	0.0122	0.0249	0.0175
0.044	0.0125	0.0320	0.0246
0.052	0.0125	0.0393	0.0319
0.059	0.0158	0.0435	0.0361
0.067	0.0197	0.0470	0.0396
0.074	0.0202	0.0539	0.0464
0.081	0.0250	0.0564	0.0490
0.089	0.0268	0.0621	0.0547
0.096	0.0295	0.0668	0.0594
0.104	0.0303	0.0734	0.0660
0.111	0.0340	0.0771	0.0697
0.119	0.0352	0.0833	0.0759
0.126	0.0581	0.0678	0.0604
0.133	0.0621	0.0712	0.0638
0.141	0.0630	0.0777	0.0703
0.148	0.0668	0.0813	0.0739
0.156	0.0777	0.0779	0.0705
0.163	0.0795	0.0835	0.0761
0.170	0.0874	0.0829	0.0755
0.178	0.0897	0.0881	0.0807
0.185	0.1034	0.0818	0.0744
0.193	0.1092	0.0834	0.0760
0.200	0.1095	0.0905	0.0831
0.207	0.1455	0.0619	0.0545
0.215	0.1638	0.0510	0.0436
0.222	0.1642	0.0580	0.0506
0.230	0.1653	0.0643	0.0569
0.237	0.1654	0.0716	0.0642
0.244	0.1654	0.0790	0.0716
0.252	0.1657	0.0862	0.0788

Table No. 5.8 continued.

$H(s_k)$	$G_0(s_k)$	$\|H(s_k) - G_0(s_k)\|$	$\|H(s_{k-1}) - G_0(s_k)\|$
0.259	0.1725	0.0868	0.0794
0.267	0.1744	0.0922	0.0848
0.274	0.1797	0.0944	0.0870
0.281	0.1814	0.1001	0.0927
0.289	0.1948	0.0941	0.0867
0.296	0.2035	0.0928	0.0854
0.304	0.2063	0.0974	0.0900
0.311	0.2138	0.0973	0.0899
0.319	0.2266	0.0919	0.0845
0.326	0.2337	0.0922	0.0848
0.333	0.2392	0.0942	0.0867
0.341	0.2445	0.0963	0.0888
0.348	0.2447	0.1035	0.0961
0.356	0.2468	0.1087	0.1013
0.363	0.2688	0.0941	0.0867
0.370	0.3040	0.0664	0.0590
0.378	0.3076	0.0702	0.0628
0.385	0.3081	0.0770	0.0696
0.393	0.3140	0.0786	0.0712
0.400	0.3174	0.0826	0.0752
0.407	0.3262	0.0812	0.0738
0.415	0.3310	0.0838	0.0764
0.422	0.3367	0.0855	0.0781
0.430	0.3375	0.0921	0.0847
0.437	0.3464	0.0907	0.0832
0.444	0.3641	0.0803	0.0729
0.452	0.3641	0.0878	0.0803
0.459	0.3695	0.0898	0.0824
0.467	0.3771	0.0896	0.0821
0.474	0.3854	0.0886	0.0812
0.481	0.4125	0.0689	0.0615
0.489	0.4242	0.0647	0.0572
0.496	0.4244	0.0718	0.0644
0.504	0.4248	0.0789	0.0715

Table No. 5.8 continued.

$H(s_k)$	$G_0(s_k)$	$\lvert H(s_k) - G_0(s_k)\rvert$	$\lvert H(s_{k-1}) - G_0(s_k)\rvert$
0.511	0.4360	0.0751	0.0677
0.519	0.4438	0.0747	0.0673
0.526	0.4448	0.0812	0.0738
0.533	0.4474	0.0859	0.0785
0.541	0.4638	0.0769	0.0695
0.548	0.4696	0.0786	0.0711
0.556	0.4755	0.0801	0.0726
0.563	0.4858	0.0772	0.0698
0.570	0.4912	0.0792	0.0718
0.578	0.4942	0.0836	0.0762
0.585	0.5097	0.0755	0.0681
0.593	0.5280	0.0646	0.0572
0.600	0.5358	0.0642	0.0567
0.607	0.5410	0.0664	0.0590
0.615	0.5706	0.0442	0.0368
0.622	0.5877	0.0345	0.0271
0.630	0.6150	0.0147	0.0072
0.637	0.6276	0.0095	0.0021
0.644	0.6415	0.0029	0.0045
0.652	0.6512	0.0006	0.0068
0.659	0.6517	0.0076	0.0002
0.667	0.6633	0.0033	0.0041
0.674	0.6728	0.0013	0.0061
0.681	0.7015	0.0201	0.0275
0.689	0.7347	0.0459	0.0533
0.696	0.7501	0.0538	0.0612
0.704	0.7586	0.0549	0.0623
0.711	0.7587	0.0476	0.0550
0.719	0.7613	0.0428	0.0502
0.726	0.7786	0.0526	0.0600
0.733	0.7865	0.0531	0.0606
0.741	0.7991	0.0584	0.0658
0.748	0.7994	0.0512	0.0586
0.756	0.8006	0.0450	0.0524

Table No. 5.8 continued.

$H(s_k)$	$G_0(s_k)$	$\lvert H(s_k) - G_0(s_k)\rvert$	$\lvert H(s_{k-1}) - G_0(s_k)\rvert$
0.763	0.8015	0.0385	0.0460
0.770	0.8015	0.0311	0.0385
0.778	0.8266	0.0488	0.0562
0.785	0.8360	0.0508	0.0582
0.793	0.8428	0.0502	0.0576
0.800	0.8477	0.0477	0.0551
0.807	0.8537	0.0462	0.0537
0.815	0.8553	0.0405	0.0479
0.822	0.8601	0.0379	0.0453
0.830	0.8721	0.0424	0.0498
0.837	0.8748	0.0377	0.0451
0.844	0.8771	0.0326	0.0400
0.852	0.8797	0.0279	0.0353
0.859	0.8804	0.0212	0.0286
0.867	0.8861	0.0194	0.0268
0.874	0.8913	0.0173	0.0247
0.881	0.8971	0.0157	0.0231
0.889	0.8973	0.0084	0.0158
0.896	0.8976	0.0013	0.0088
0.904	0.9001	0.0036	0.0039
0.911	0.9260	0.0149	0.0223
0.919	0.9265	0.0079	0.0153
0.926	0.9265	0.0006	0.0080
0.933	0.9310	0.0024	0.0050
0.941	0.9325	0.0083	0.0008
0.948	0.9356	0.0125	0.0051
0.956	0.9575	0.0020	0.0094
0.963	0.9676	0.0046	0.0120
0.970	0.9705	0.0002	0.0076
0.978	0.9723	0.0055	0.0019
0.985	0.9858	0.0006	0.0080
0.993	0.9883	0.0042	0.0032
1.000	0.9922	0.0078	0.0004

$$L(x) = max\{H(x) - 0.11705, 0\}$$

and

$$U(x) = min\{H(x) - 0.11705, 1\}. \tag{5.22}$$

The 95% confidence bounds for $G_0(x)$, and $G(x)$, are shown in Fig.5.3.

3. COMPARISON WITH G-O MODEL

The maximum likelihood estimates for the parameters of the G-O model are $\hat{a} = 142.83$ and $\hat{b} = .000034$. Therefore the estimated expected number of errors by time t is

$$\hat{m}_1(t) = 142.83(1 - e^{-0.000034t}). \tag{5.23}$$

Hence the expected number of faults to be eventually detected is 142.83 and for the modified model it is 155.52.

The inverse functions, to estimate mean time to failure for G-O and modified G-O, are, respectively,

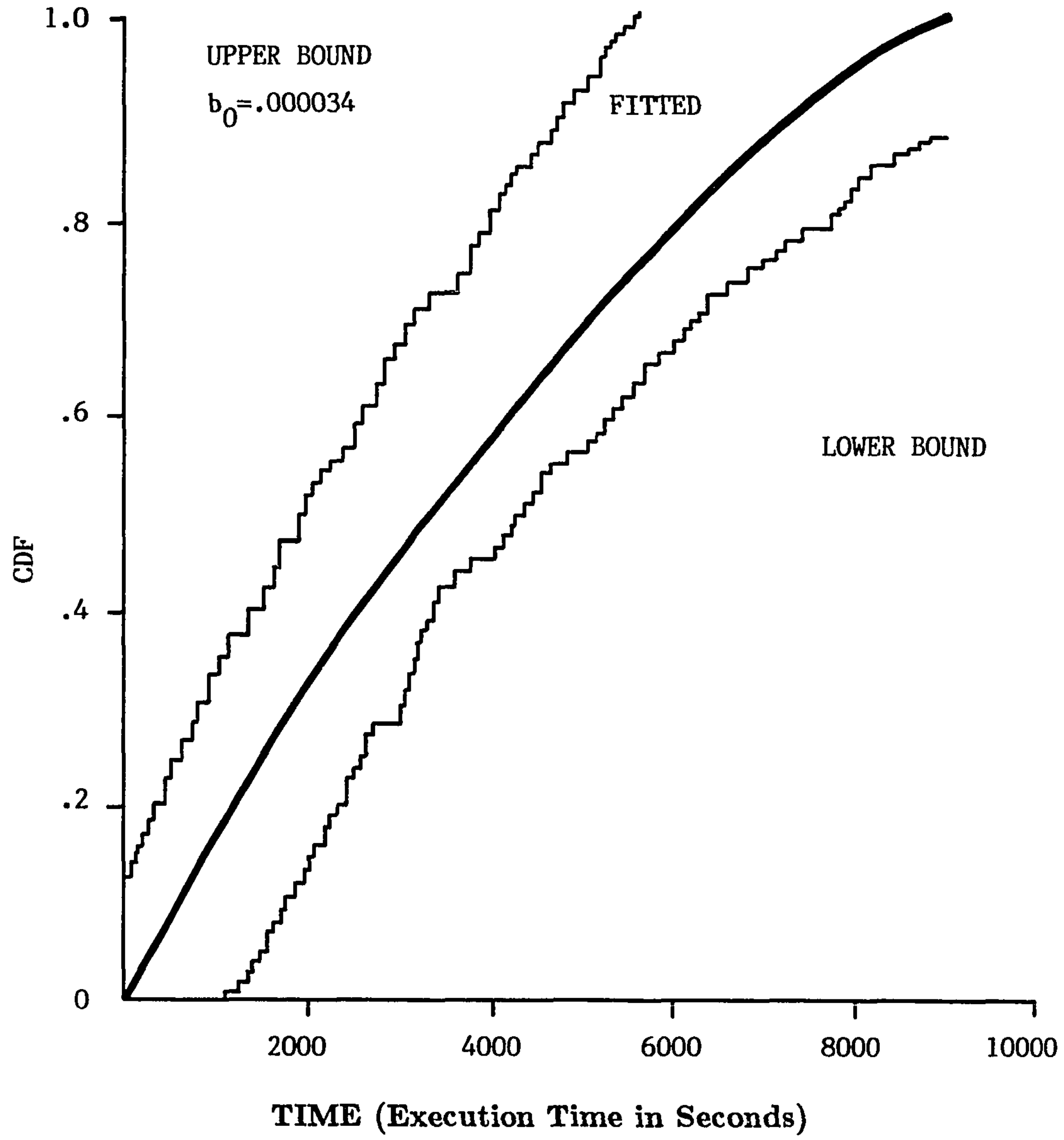

Fig.5.3 95% Confidence Bounds for CDF of

$G(x)$ and Fitted CDF Curve.

$$\hat{s}_k = -\frac{1}{\hat{b}}\log(1-\frac{k}{\hat{a}})$$

$$= -\frac{1}{.000034}\log(1-\frac{k}{142.83}) \quad \text{and} \tag{5.24}$$

$$\hat{s}_k = -\frac{1}{\hat{b}}\log\frac{1}{\hat{a}}\log[(e^{\tilde{a}}-\tilde{c})e^{-k}+\tilde{c}]$$

$$= -\frac{1}{.000034}\log\frac{1}{142.815}\log[(e^{142.815}-.999997)e^{-k}+.999997] \tag{5.25}$$

The observed and the estimated s_k's for both the models are given in Table 5.9. For the G-O model the Fit (SSD)= $\sum_{k=1}^{136}(s_k-\hat{s}_k)^2$ =1796637938.1 and for the modified G-O model it is 1796359225.1. Clearly, Modified shows a better fit.

4. DISTRIBUTION OF FAULTS REMAINING AFTER DEBUGGING

Using the probability distribution of the remaining faults, as described in Section 4.1 by equation (4.10), we have

$$P_3[\bar{N}(t)=k] = \frac{[\log\frac{e^{ae^{-bt}}-c}{1-c}]^k}{k!}\left(\frac{1-c}{e^{ae^{-bt}}-c}\right)$$

$$= \frac{[\log\frac{e^{142.815e^{-.000034t}}-.999997}{1-.999997}]^k}{k!}\left(\frac{1-.999997}{e^{142.815e^{-.000034t}}-.999997}\right),$$

$$\text{where} \quad k=0,1,2,\ldots \tag{5.26}$$

The distribution of $\bar{N}(t)$ is appended below in Table No.5.10.

Table No.5.9 Comparison of Results Based on the Goel-Okumoto and Modified Goel-Okumoto models.

Error No. k	Actual Failure Time s_k,	Estimated Failure Time $\hat{s}_k$ G-O	Modified G-O
1	3	204.79	204.75
2	33	411.03	410.95
3	146	618.74	618.62
4	227	827.94	827.77
5	342	1038.66	1038.44
6	351	1250.90	1250.65
7	353	1464.71	1464.41
8	444	1680.09	1679.75
9	556	1897.08	1896.70
10	571	2115.70	2115.27
11	709	2335.96	2335.49
12	759	2557.91	2557.39
13	836	2781.56	2781.00
14	860	3006.93	3006.33
15	968	3234.07	3233.42
16	1001	3462.98	3462.29
17	1671	3693.71	3692.98
18	1791	3926.28	3925.50
19	1817	4160.72	4159.90
20	1931	4397.07	4396.20
21	2256	4635.34	4634.42
22	2311	4875.58	4874.62
23	2553	5117.81	5116.81
24	2621	5362.08	5361.02
25	3043	5608.40	5607.31
26	3223	5856.83	5855.69
27	3233	6107.40	6106.21
28	4379	6360.13	6358.90
29	4979	6615.08	6613.80
30	4994	6872.27	6870.95
31	5030	7131.76	7130.39
32	5034	7393.58	7392.16
33	5034	7657.77	7656.30
34	5042	7924.37	7922.86
35	5269	8193.44	8191.88
36	5334	8465.02	8463.41
37	5510	8739.15	8737.49
38	5568	9015.88	9014.18
39	6025	9295.26	9293.51
40	6325	9577.35	9575.55

Table No.5.9 continued.

Error No. k	Actual Failure Time s_k, (days)	Estimated Failure Time $\hat{s}_k$ G-O	Modified G-O
41	6422	9862.19	9860.35
42	6685	10149.85	10147.96
43	7137	10440.37	10438.43
44	7392	10733.82	10731.83
45	7589	11030.25	11028.22
46	7782	11329.73	11327.65
47	7788	11632.32	11630.19
48	7867	11938.08	11935.90
49	8683	12247.08	12244.85
50	10034	12559.39	12557.12
51	10177	12875.09	12872.77
52	10198	13194.24	13191.87
53	10431	13516.93	13514.51
54	10565	13843.23	13840.76
55	10922	14173.22	14170.71
56	11115	14506.99	14504.43
57	11351	14844.63	14842.02
58	11382	15186.22	15183.57
59	11751	15531.87	15529.17
60	12499	15881.66	15878.91
61	12499	16235.71	16232.91
62	12731	16594.10	16591.26
63	13061	16956.96	16954.07
64	13426	17324.39	17321.45
65	14648	17696.51	17693.53
66	15191	18073.45	18070.42
67	15201	18455.32	18452.25
68	15217	18842.27	18839.14
69	15746	19234.42	19231.25
70	16125	19631.91	19628.70
71	16169	20034.91	20031.65
72	16298	20443.55	20440.25
73	17108	20858.00	20854.66
74	17398	21278.43	21275.05
75	17698	21705.02	21701.59
76	18227	22137.94	22134.47
77	18508	22577.39	22573.88
78	18668	23023.56	23020.02
79	19496	23476.67	23473.09
80	20507	23936.94	23933.32

Table No.5.9 continued.

Error No. k	Actual Failure Time s_k, (days)	Estimated Failure Time $\hat{s}_k$ G-O	Modified G-O
81	20952	24404.59	24400.93
82	21248	24879.87	24876.17
83	23003	25363.02	25359.29
84	24067	25854.32	25850.56
85	25850	26354.04	26350.25
86	26710	26862.48	26858.66
87	27693	27379.95	27376.10
88	28400	27906.77	27902.89
89	28433	28443.29	28439.39
90	29301	28989.86	28985.94
91	30025	29546.89	29542.95
92	32348	30114.76	30110.81
93	35278	30693.92	30689.95
94	36739	31284.82	31280.84
95	37582	31887.95	31883.96
96	37594	32503.83	32499.83
97	37855	33132.99	33129.00
98	39655	33776.04	33772.05
99	40520	34433.60	34429.62
100	41955	35106.33	35102.37
101	41985	35794.96	35791.01
102	42128	36500.25	36496.33
103	42236	37223.03	37219.15
104	42236	37964.19	37960.35
105	45346	38724.69	38720.90
106	46593	39505.57	39501.83
107	47536	40307.94	40304.27
108	48236	41133.03	41129.44
109	49111	41982.15	41978.65
110	49356	42856.76	42853.36
111	50085	43758.42	43755.14
112	51982	44688.87	44685.73
113	52429	45650.00	45647.01
114	52815	46643.91	46641.10
115	53261	47672.91	47670.30
116	53383	48739.57	48737.19
117	54373	49846.75	49844.63
118	55321	50997.66	50995.82
119	56403	52195.88	52194.38
120	56425	53445.48	53444.36

Table No.5.9 continued.

Error No. k	Actual Failure Time s_k, (days)	Estimated Failure Time $\hat{s}_k$	
		G-O	Modified G-O
121	56500	54751.07	54750.37
122	56982	56117.88	56117.67
123	62491	57551.96	57552.31
124	62591	59060.26	59061.25
125	62601	60650.89	60652.63
126	63672	62333.36	62335.96
127	64043	64118.92	64122.52
128	64833	66021.05	66025.83
129	70983	68056.03	68062.19
130	74304	70243.82	70251.64
131	75349	72609.24	72619.04
132	75997	75183.72	75195.92
133	81482	78007.81	78022.91
134	82642	81135.17	81153.74
135	84506	84638.86	84661.32
136	88622	88622.00	88647.77

From this table it is clear that the $P_3(\bar{N}(88622) \leq 31) = 0.9932$. Therefore the probability that thirtyone or a lesser number of faults remained in the system after being observed for 88622 execution times, in seconds, is about 0.99.

Table No.10 Distribution of Remaining. Errors After Debugging.

k	$P_3[\bar{N}(88622) \leq k]$
2	0.000001
3	0.000004
4	0.000021
5	0.000088
6	0.000309
7	0.000931
8	0.002465
9	0.005830
10	0.012469
11	0.024380
12	0.043967
13	0.073700
14	0.115608
15	0.170742
16	0.238740
17	0.317672
18	0.404206
19	0.494080
20	0.582756
21	0.666084
22	0.740827
23	0.804955
24	0.857683
25	0.899303
26	0.930892
27	0.953980
28	0.970251
29	0.981323
30	0.988606
31	0.993242
32	0.996100

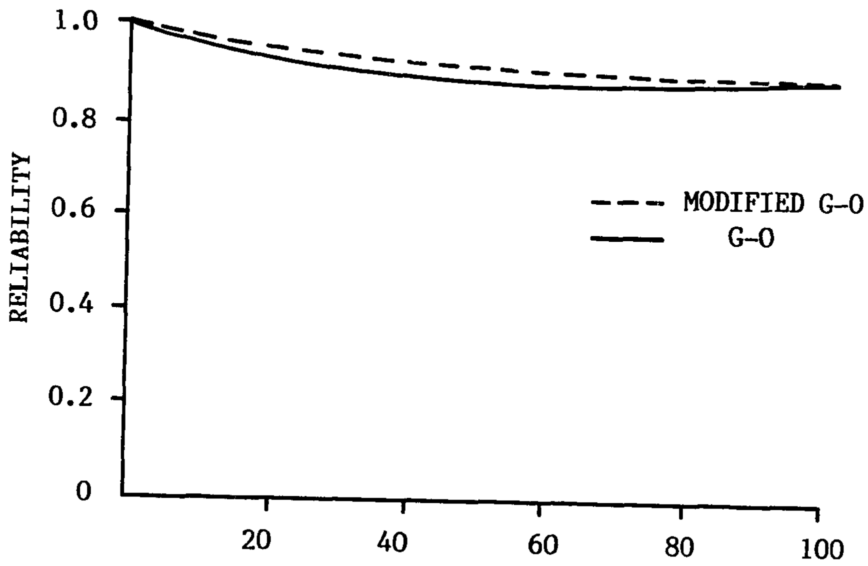

Fig.5.4 Plots of Conditional Reliability Functions

of Modified G-O & G-O Models.

The estimated conditional reliability function for the modified G-O model is

$$\hat{R}_1(t|s_{136} = 88622) = \frac{e^{142.815e^{-.000034(88622+t)}} - .999997}{e^{142.815e^{-.000034\times 88622}} - .999997} \tag{5.27}$$

and that of G-O model is

$$\hat{R}(t|s_{136} = 88622) = e^{142.83(e^{-.000034\times 88622} - e^{-.000034(88622+t)})}. \tag{5.28}$$

The plot of these reliability functions versus time are shown in Fig.5.6.

5.3 Comparison between Yamada *et al* and Modified Yamada *et al* Models based on Failure Data from NTDS.

In this section we will once again analyze the NTDS data to compare the performances of the Modified Yamada *et al* and Yamada *et al* models.

1. DATA ANALYSIS

As before, we will consider the 26 observations of production phase as shown in Table 5.8 to estimate our parameters. Solving the equations (4.20) and (4.21) for different values of c, we obtain Table 5.11. The estimates of the parameters are $\hat{a} = 24.7429$, $\hat{b} = 0.023069$, and $\hat{c} = 0.99996$.

The fitted mean value function is

$$\hat{m}_4(t) = \log\left(\frac{e^{24.7429} - .99996}{e^{24.7429(1+.023069t)e^{-.023069t}} - .99996}\right), \qquad t \geq 0, \qquad (5.29)$$

and is shown in Fig.5.8 along with the actual data. Note that the expected number of faults, $\hat{m}_4(\infty)$, to be eventually detected is 34.84.

Table No.5.11 Mean Sum of Squares of Deviations of Observed and Estimated Time to Failure for Different Values of c.

c	$\frac{1}{n}\sum_{k=1}^{n}(s_k - \hat{s}_k)^2$
0.0000000	330.96
0.9920000	241.44
0.9986283	241.01
0.9995448	240.95
0.9997965	240.93
0.9998920	240.93
0.9999360	240.93
0.9999590	240.92
0.9999722	240.92
0.9999803	240.92
0.9999855	240.92
0.9999890	240.92
0.9999915	240.92
0.9999933	240.92
0.9999946	240.92
0.9999956	240.92
0.9999964	240.92
0.9999970	240.92
0.9999974	240.92
0.9999978	240.92
0.9999981	240.92
0.9999984	240.92
0.9999986	240.92
0.9999988	240.92
1.0000000	240.92

2. TEST OF GOODNESS OF FIT

We will use Kolmogorov-Smirnov goodness-of-fit test to check the fitness of the model. As described in Section 5.1, the test is based on 25 observations. The null hypothesis is

H_0 :

$$G_0(x) = \frac{\log(e^{24.743} - .99996) - \log(e^{24.743(1+.02307x)e^{-.02307x}} - .99996)}{\log(e^{24.743} - .99996) - \log(e^{24.743(1+.02307\times 250)e^{-.02307\times 250}} - .99996)} \tag{5.30}$$

and the sample cdf is

$$H(x) = \begin{cases} 0, & \text{if } x < s_1 \\ \frac{k}{25}, & \text{if } s_{k-1} \leq x < s_k, \qquad k = 2, 3, \ldots, 25. \\ 1, & \text{if } x \geq s_{25}. \end{cases} \tag{5.31}$$

The necessary calculations for D statistic for various values of s_k, are shown in Table No.5.12, given by $D = 0.1140$. The critical value of $D_{25,0.05}$ at 5% level of significance is

$$D_{25,0.05} = 0.264.$$

Since $D < D_{25,0.05}$ we accept the null- hypothesis. Note that the $D = 0.1862$ in the case of Yamada *et al* model indicates a better fit infavor of modified Yamada *et al.*

The 95% confidence limits for $G(x)$ can now be calculated with $\alpha = .05$ and $D_{25,.05} = 0.264$. The lower and upper confidence bounds are

Table No. 5.12 **Kolmogorov-Smirnov Test For The NTDS Data Set.**

$H(s_k)$	$G_0(s_k)$	$\lvert H(s_k) - G_0(s_k)\rvert$	$\lvert H(s_{k-1}) - G_0(s_k)\rvert$
0.040	0.0185	0.0215	0.0185
0.080	0.0843	0.0043	0.0443
0.120	0.1667	0.0467	0.0867
0.160	0.1992	0.0392	0.0792
0.200	0.2574	0.0574	0.0974
0.240	0.2741	0.0341	0.0741
0.280	0.3157	0.0357	0.0757
0.320	0.3808	0.0608	0.1008
0.360	0.4201	0.0601	0.1001
0.400	0.4727	0.0727	0.1127
0.440	0.4800	0.0400	0.0800
0.480	0.5222	0.0422	0.0822
0.520	0.5290	0.0090	0.0490
0.560	0.5870	0.0270	0.0670
0.600	0.6110	0.0110	0.0510
0.640	0.6168	0.0232	0.0168
0.680	0.6338	0.0462	0.0062
0.720	0.6502	0.0698	0.0298
0.760	0.6811	0.0789	0.0389
0.800	0.6860	0.1140	0.0740
0.840	0.7359	0.1041	0.0641
0.880	0.8458	0.0342	0.0058
0.920	0.8630	0.0570	0.0170
0.960	0.9970	0.0370	0.0770
1.000	0.9990	0.0010	0.0390

$$L(x) = max\{H(x) - 0.264, 0\}$$

and

$$U(x) = min\{H(x) - 0.264, 1\}. \tag{5.32}$$

The 95% confidence bounds for $G_0(x)$ and $G(x)$ are shown in Fig.5.5.

3. COMPARISON WITH YAMADA *et al* MODEL

The maximum likelihood estimates for the parameters of the Yamada model are obtained to be $\hat{a} = 27.4915$ and $\hat{b} = .018579$. Therefore the estimated expected number of errors by time t is

$$\hat{m}_2(t) = 27.4915[1 - (1 + 0.01858t)e^{-0.01858t}]. \tag{5.33}$$

Hence the expected number of faults to be eventually detected is 27.49 and for that of the modified model it is 34.84.

The inverse functions to estimate the mean time to failure for modified Yamada and Yamada, are solutions of equations (5.34) and (5.35) for s_k, respectively.

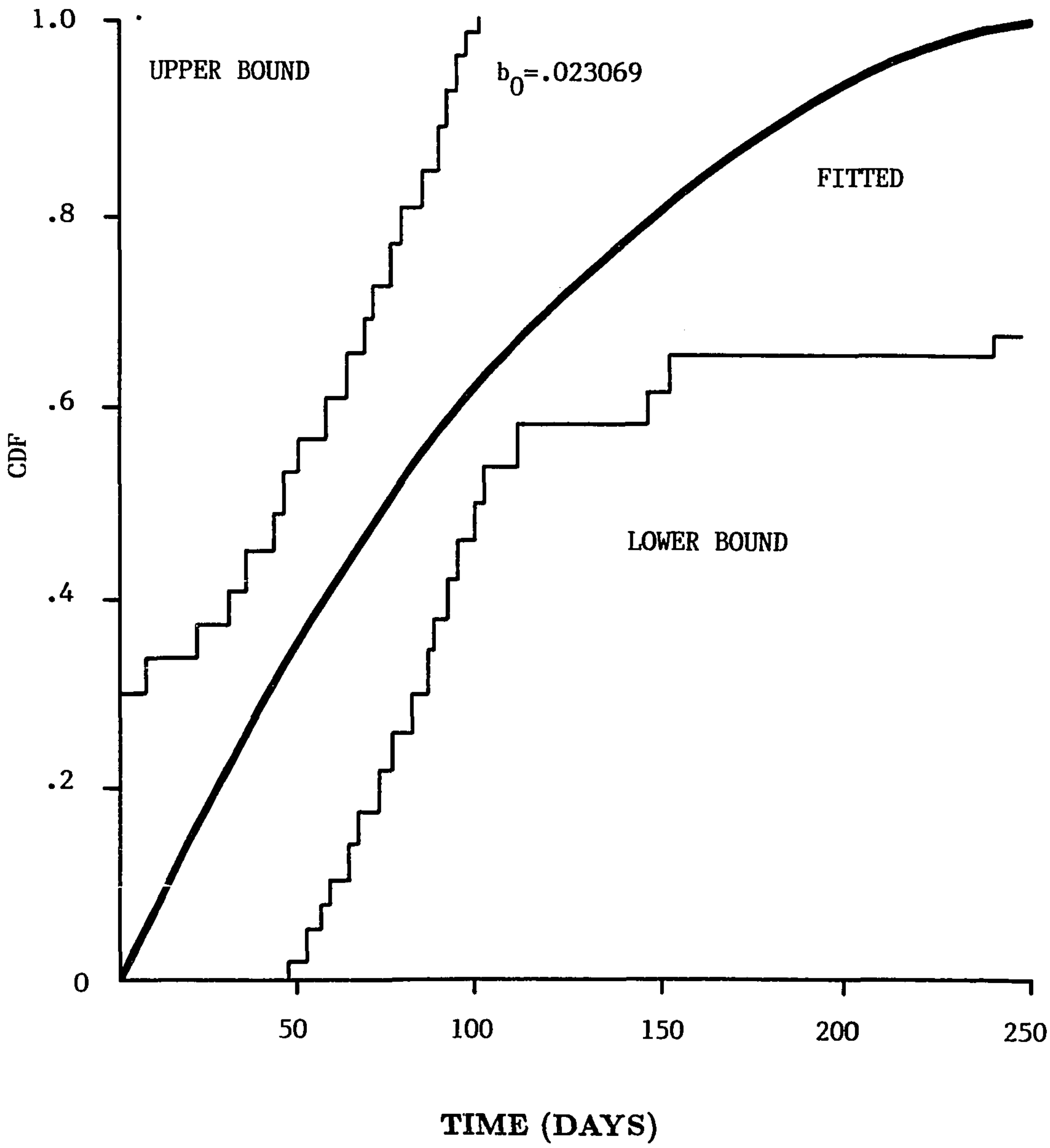

Fig. 5.5 95% Confidence Bounds for CDF of G(x)

and Fitted CDF Curve.

$$(1+\hat{b}s_k)e^{-\hat{b}s_k} = \frac{1}{\hat{a}}\log[(e^{\hat{a}}-\hat{c})e^{-k}+\hat{c}], \quad \text{therefore,}$$

$$(1+.023067s_k)e^{-.023067s_k} = \frac{1}{24.7429}\log[(e^{24.7429}-.99996)e^{-k}+.99996], \tag{5.34}$$

$$\text{and} \quad (1+\hat{b}s_k)e^{-\hat{b}s_k} = 1-\frac{k}{\hat{a}}$$

$$\implies (1+.01858s_k)e^{-.01858s_k} = 1-\frac{k}{27.4915}. \tag{5.35}$$

The observed s_k's and the estimated s_k's for both the models are given in Table No.5.13. For the Yamada *et al* model the Fit (SSD)= $\sum_{k=1}^{26}(s_k-\hat{s}_k)^2$ =8604.96, and for the modified Yamada *et al* model it is 6263.92. Clearly, odified model shows a better fit.

4. DISTRIBUTION OF FAULTS REMAINING AFTER DEBUGGING

Using the probability distribution of the remaining faults, as described in Section 4.1 by equation (4.10), on using $\bar{a}$, $\bar{b}$ and $\bar{c}$ we have

$$P_4[\bar{N}(t)=k] = \frac{[\log\frac{e^{a(1+bt)e^{-bt}}-c}{1-c}]^k}{k!}\left(\frac{1-c}{e^{a(1+bt)e^{-bt}}-c}\right)$$

$$= \frac{[\log\frac{e^{24.7429(1+.023069t)e^{-.023069t}}-.99996}{1-.99996}]^k}{k!}\times$$

$$= \left(\frac{1-.99996}{e^{24.7429(1+.023069t)e^{-.023069t}}-.99996}\right),$$

$$\text{where} \quad k=0,1,2,\ldots \tag{5.36}$$

Table No.5.13 **Comparison of Results Based on the Yamada *et al* and Modified Yamada *et al* models.**

Error No. k	Actual Failure Time s_k, (days)	Estimated Failure Time $\hat{s}_k$ Yamada *et al*	Modified Yamada *et al*
1	9	16.01	13.67
2	21	23.71	20.31
3	32	[illegible]	25.92
4	36	36.07	31.07
5	43	41.66	35.99
6	45	47.08	40.78
7	50	52.42	45.55
8	58	57.74	50.34
9	63	63.10	55.20
10	70	68.54	60.19
11	71	74.10	65.33
12	77	79.83	70.70
13	78	85.75	76.33
14	87	91.93	82.30
15	91	98.42	88.68
16	92	105.29	95.58
17	95	112.61	103.14
18	98	120.49	111.55
19	104	129.07	121.08
20	105	138.53	132.14
21	116	149.12	145.37
22	149	161.25	161.74
23	156	175.53	182.77
24	247	193.05	210.29
25	249	215.99	245.45
26	250	250.00	287.19

The distribution of $\bar{N}(t)$ is appended below in Table No.5.14. From this table it is clear that the $P_4(\bar{N}(250) \leq 18) = 0.9944$. Therefore, the probability that eighteen or less number of faults remained in the system after observing for 250 execution time in seconds is about 0.99.

Table No.5.14 Distribution of Remaining. Faults After Debugging.

k	$P_4[\bar{N}(250) \leq k]$
1	0.00062
2	0.00339
3	0.01237
4	0.03428
5	0.07701
6	0.14647
7	0.24326
8	0.36126
9	0.48914
10	0.61386
11	0.72445
12	0.81434
13	0.88177
14	0.92875
15	0.95930
16	0.97792
17	0.98861
18	0.99440
19	0.99737
20	0.99882

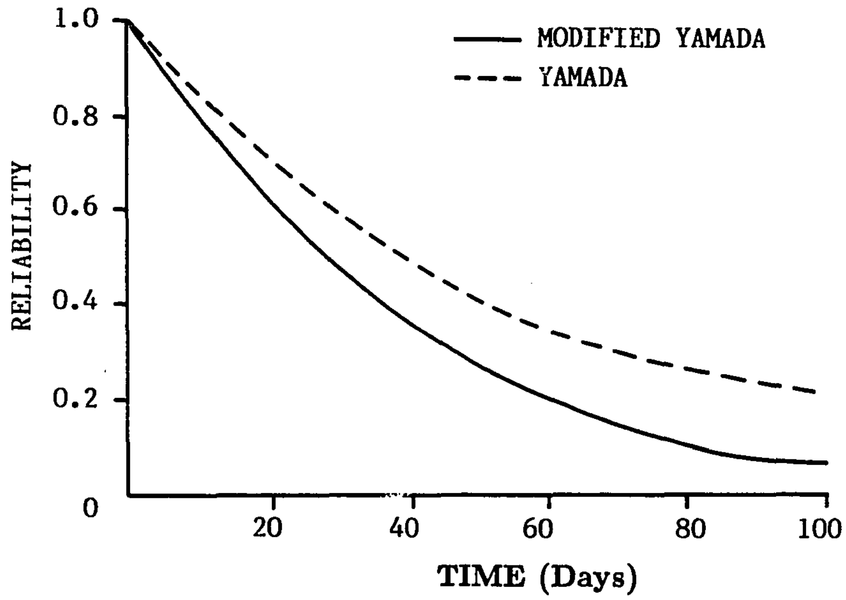

Fig.5.6 Plots of Conditional Reliability Functions of Modified Yamada & Yamada Models.

The estimated conditional reliability function for the modified Yamada *et al* model is

$$\hat{R}_2(t|s_{26} = 250) = \frac{e^{24.7429[1+.023069(250+t)]e^{-.023069(250+t)}} - .99996}{e^{24.7429(1+.023069\times 250)e^{-.023069\times 250}} - .99996}, \quad (5.37)$$

and that of the G-O model it is

$$\hat{R}(t|s_{26} = 250) = e^{27.4915(e^{-.01858\times 250} - e^{-.01858(250+t)})}. \quad (5.38)$$

The plot of these reliability functions versus time are shown in Fig.5.9.

5.4 Analysis of Failure Data Given in Execution Time and Comparison between Yamada *et al* and Modified Yamada *et al* Models.

In this section we will analyze a set of data extracted from Abdel-Ghaly *et al* (1986), to compare the performances of the Modified Yamada *et al* and Yamada *et al* models.

1. <u>DATA ANALYSIS</u>

All the 86 observations of Table 5.15 will be used to estimate the parameters. Solving the equations (4.20) and (4.21) for different values of c, we obtain the Table No.5.16. The estimates of the parameters are $\hat{a} = 89.691$, $\hat{b} = 0.000048$, and $\hat{c} = 0.9999999976$.

The fitted mean value function is

$$\hat{m}_4(t) = \log\left(\frac{e^{89.691} - .9999999976}{e^{89.691(1+.000048t)e^{-.000048t}} - .9999999976}\right), \qquad t \geq 0, \tag{5.39}$$

and is shown in Fig.5.11 along with the actual data. Note that the expected number of faults, $\hat{m}_4(\infty)$, to be eventually detected is 109.54.

Table No.5.15 **Software Failure Data.**

Error No. k	Inter Failure Times x_k, (CPU)	Cumulative Time $s_k = \sum_{i=1}^{k} x_i$, (CPU)
1	479	479
2	266	745
3	277	1022
4	554	1576
5	1034	2610
6	249	2859
7	693	3552
8	597	4149
9	117	4266
10	170	4436
11	117	4553
12	1274	5827
13	469	6296
14	1174	7470
15	693	8163
16	1908	10071
17	135	10206
18	277	10483
19	596	11079
20	757	11836
21	437	12273
22	2230	14503
23	437	14940
24	340	15280
25	405	15685
26	575	16260
27	277	16537
28	363	16900
29	522	17422
30	613	18035
31	277	18312
32	1300	19612
33	821	20433
34	213	20646
35	1620	22266
36	1601	23867

Table No.5.15 continued.

Error No. k	Inter Failure Times x_k, (CPU)	Cumulative Time $s_k = \sum_{i=1}^{k} x_i$, (CPU)
37	298	24165
38	874	25039
39	618	25657
40	2640	28297
41	5	28302
42	149	28451
43	1034	29485
44	2441	31926
45	460	32386
46	565	32951
47	1119	34070
48	437	34507
49	927	35434
50	4462	39896
51	714	40610
52	181	40791
53	1485	42276
54	757	43033
55	3154	46187
56	2445	48632
57	884	49516
58	2037	51553
59	1481	53034
60	559	53593
61	490	54083
62	593	54676
63	1769	56445
64	85	56530
65	2836	59366
66	213	59579
67	1866	61445
68	490	61935
69	1437	63372
70	4322	67694
71	1418	69112
72	1023	70135

Table No.5.15 continued.

Error No. k	Inter Failure Times x_k, (CPU)	Cumulative Time $s_k = \sum_{i=1}^{k} x_i$, (CPU)
73	5490	75625
74	1520	77145
75	3281	80426
76	2716	83142
77	2175	85317
78	3505	88822
79	725	89547
80	1963	91510
81	3979	95489
82	1090	96579
83	245	96824
84	1194	98018
85	994	99012
86	3902	102914

Table No.5.16 Mean Sum of Squares of. Deviations of Observed and Estimated. Time to Failure for Different Values of c.

c	$\frac{1}{n}\sum_{k=1}^{n}(s_k - \hat{s}_k)^2$
0.000000000000	45073482.64
0.999680000000	44062324.54
0.999983064912	44061975.70
0.999997306709	44061959.27
0.999999295704	44061957.04
0.999999755148	44061956.74
0.999999897600	44061956.66
0.999999951246	44061956.63
0.999999974448	44061956.61
0.999999985579	44061956.61
0.999999991369	44061956.60
0.999999994581	44061956.60
0.999999996460	44061956.60
0.999999997609	44061956.59
0.999999998338	44061956.59
0.999999998816	44061956.59
0.999999999138	44061956.59
0.999999999990	44061956.59
0.999999999991	44061956.59
0.999999999992	44061956.59
0.999999999993	44061956.59
0.999999999994	44061956.59
0.999999999994	44061956.59
0.999999999997	44061956.59
1.000000000000	44061956.59

2. TEST OF GOODNESS OF FIT

As before we will use the Kolmogorov-Smirnov goodness-of-fit test to check the fitness of the model. The test is based on 85 observations. The null hypothesis is

$$H_0: \quad G_0(x) = \frac{\log\left[\frac{e^{89.69} - .9999999976}{e^{89.69(1+.000048x)e^{-.000048x}} - .9999999976}\right]}{\log\left[\frac{e^{89.69} - .9999999976}{e^{89.69(1+.000048\times 102914)e^{-.000048\times 102914}} - .9999999976}\right]} \tag{5.40}$$

and the sample cdf is

$$H(x) = \begin{cases} 0, & \text{if } x < s_1 \\ \frac{k}{85}, & \text{if } s_{k-1} \leq x < s_k, \qquad k = 2, 3, \ldots, 85 \\ 1, & \text{if } x \geq s_{85}. \end{cases} \tag{5.41}$$

The necessary calculations for D statistic for various values of s_k are shown in Table No.5.17. The estimate is $D = 0.1352$. The critical value of $D_{85,0.05}$ at a 5% level of significance is

$$D_{85,0.05} = 0.1475.$$

Since $D < D_{85,0.05}$, we accept the null- hypothesis. Note that the $D = 0.1412$ in the case of the Yamada *et al* model indicating a better fit in favor of the modified Yamada *et al.*

The 95% confidence limits for $G(x)$ can now be calculated with $\alpha = .05$ and $D_{85,.05} = 0.1475$. The lower and upper confidence bounds are

Table No. 5.17 **Kolmogorov-Smirnov Test For The Data Set.**

$H(s_k)$	$G_0(s_k)$	$\lvert H(s_k) - G_0(s_k)\rvert$	$\lvert H(s_{k-1}) - G_0(s_k)\rvert$
0.012	0.0003	0.0115	0.0003
0.024	0.0007	0.0229	0.0111
0.035	0.0012	0.0341	0.0223
0.047	0.0028	0.0442	0.0325
0.059	0.0075	0.0513	0.0395
0.071	0.0090	0.0616	0.0498
0.082	0.0136	0.0688	0.0570
0.094	0.0182	0.0760	0.0642
0.106	0.0191	0.0868	0.0750
0.118	0.0206	0.0971	0.0853
0.129	0.0216	0.1078	0.0961
0.141	0.0340	0.1072	0.0954
0.153	0.0391	0.1139	0.1021
0.165	0.0530	0.1117	0.0999
0.176	0.0620	0.1145	0.1027
0.188	0.0890	0.0993	0.0875
0.200	0.0910	0.1090	0.0972
0.212	0.0952	0.1166	0.1048
0.224	0.1044	0.1191	0.1074
0.235	0.1164	0.1189	0.1071
0.247	0.1235	0.1235	0.1117
0.259	0.1613	0.0975	0.0858
0.271	0.1689	0.1017	0.0899
0.282	0.1749	0.1075	0.0957
0.294	0.1821	0.1120	0.1003
0.306	0.1923	0.1135	0.1018
0.318	0.1973	0.1203	0.1086
0.329	0.2039	0.1256	0.1138
0.341	0.2133	0.1279	0.1161
0.353	0.2245	0.1285	0.1167
0.365	0.2295	0.1352	0.1234
0.376	0.2534	0.1231	0.1113
0.388	0.2685	0.1197	0.1080
0.400	0.2724	0.1276	0.1158
0.412	0.3023	0.1095	0.0977
0.424	0.3316	0.0919	0.0801

Table No. 5.17 continued.

$H(s_k)$	$G_0(s_k)$	$\|H(s_k) - G_0(s_k)\|$	$\|H(s_{k-1}) - G_0(s_k)\|$
0.435	0.3371	0.0982	0.0864
0.447	0.3530	0.0941	0.0823
0.459	0.3641	0.0947	0.0829
0.471	0.4110	0.0595	0.0478
0.482	0.4111	0.0712	0.0595
0.494	0.4137	0.0804	0.0686
0.506	0.4317	0.0742	0.0624
0.518	0.4730	0.0447	0.0329
0.529	0.4806	0.0489	0.0371
0.541	0.4898	0.0514	0.0396
0.553	0.5079	0.0451	0.0333
0.565	0.5148	0.0499	0.0381
0.576	0.5294	0.0471	0.0353
0.588	0.5957	0.0074	0.0192
0.600	0.6057	0.0057	0.0174
0.612	0.6082	0.0036	0.0082
0.624	0.6284	0.0049	0.0166
0.635	0.6384	0.0031	0.0149
0.647	0.6782	0.0311	0.0429
0.659	0.7069	0.0480	0.0598
0.671	0.7168	0.0462	0.0579
0.682	0.7386	0.0563	0.0681
0.694	0.7538	0.0597	0.0714
0.706	0.7593	0.0535	0.0652
0.718	0.7641	0.0465	0.0582
0.729	0.7698	0.0404	0.0522
0.741	0.7863	0.0451	0.0568
0.753	0.7870	0.0341	0.0458
0.765	0.8115	0.0468	0.0586
0.776	0.8133	0.0368	0.0486
0.788	0.8282	0.0399	0.0517
0.800	0.8319	0.0319	0.0437
0.812	0.8426	0.0309	0.0426
0.824	0.8720	0.0484	0.0602
0.835	0.8807	0.0454	0.0572
0.847	0.8868	0.0397	0.0515

Table No. 5.17 continued.

$H(s_k)$	$G_0(s_k)$	$\|H(s_k) - G_0(s_k)\|$	$\|H(s_{k-1}) - G_0(s_k)\|$
0.859	0.9159	0.0571	0.0689
0.871	0.9231	0.0525	0.0642
0.882	0.9372	0.0549	0.0667
0.894	0.9478	0.0537	0.0654
0.906	0.9555	0.0497	0.0614
0.918	0.9668	0.0491	0.0609
0.929	0.9689	0.0395	0.0513
0.941	0.9745	0.0333	0.0451
0.953	0.9846	0.0317	0.0434
0.965	0.9871	0.0224	0.0342
0.976	0.9877	0.0112	0.0230
0.988	0.9903	0.0021	0.0138
1.000	0.9924	0.0076	0.0042

$$L(x) = max\{H(x) - 0.1475, 0\}$$

and

$$U(x) = min\{H(x) - 0.1475, 1\} \tag{5.42}$$

The 95% confidence bounds for $G_0(x)$ and $G(x)$ are shown in Fig.5.7.

3. COMPARISON WITH YAMADA *et al* MODEL

The maximum likelihood estimates for the parameters of the Yamada model *et al* are obtained to be $\hat{a} = 90.0624$ and $\hat{b} = .000047$. Therefore the estimated expected number of errors by time t is

$$\hat{m}_2(t) = 90.0624[1 - (1 + 0.000047t)e^{-0.000047t}] \tag{5.43}$$

Hence the expected number of faults to be eventually detected is 90.06, and is 109.54 for the modified model.

The inverse functions to estimate expected time to failure for modified Yamada *et al* and Yamada *et al* are, respectively, solutions of equations (5.44) and (5.45) for s_k.

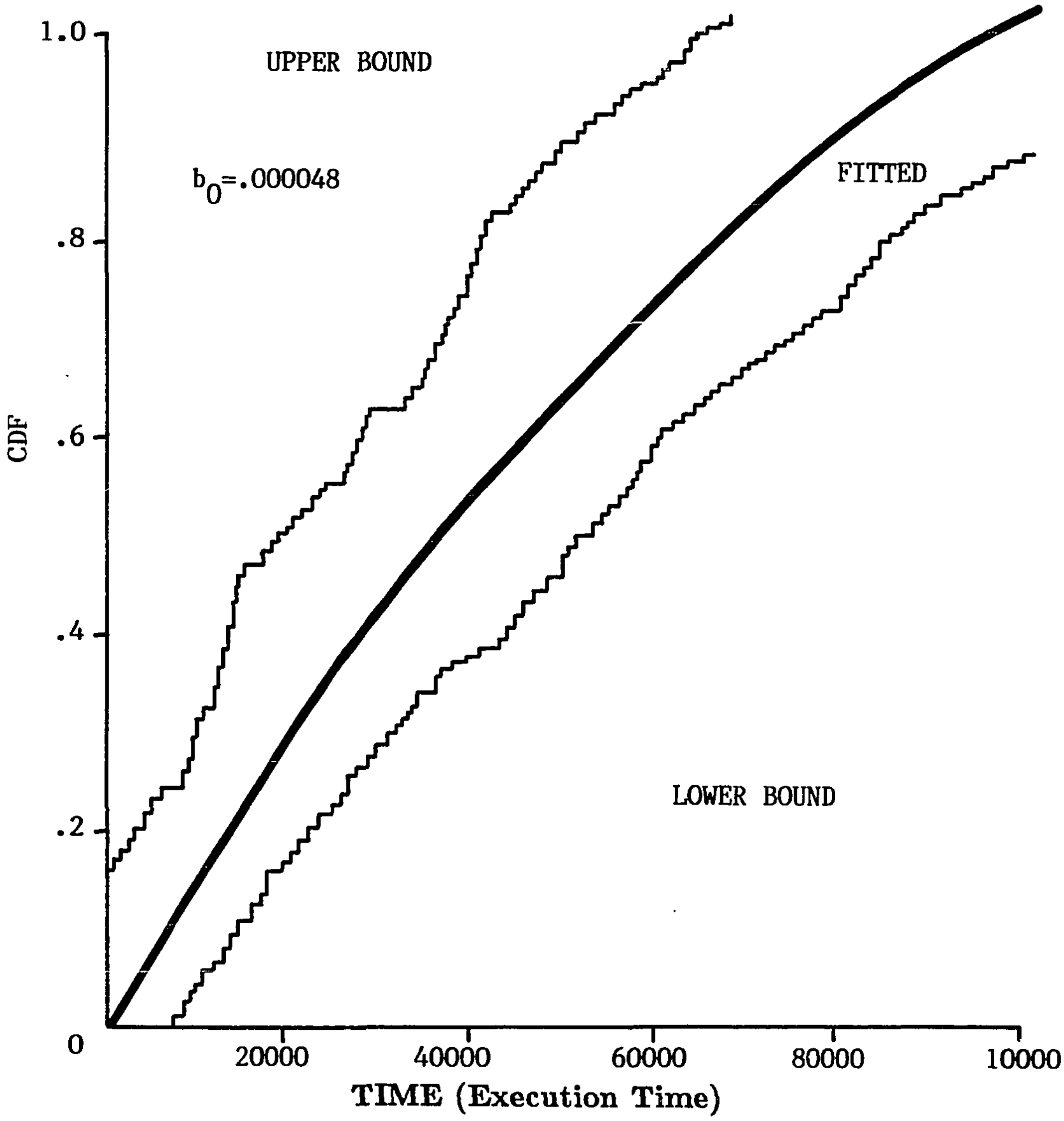

Fig. 5.7 95% Confidence Bounds for CDF of G(x)

and Fitted CDF Curve.

$$(1+\hat{b}s_k)e^{-\hat{b}s_k} = \frac{1}{\hat{a}}\log[(e^{\hat{a}}-\hat{c})e^{-k}+\hat{c}], \quad \text{therefore,}$$

$$(1+.000048s_k)e^{-.000048s_k} = \frac{1}{89.69}\log[(e^{89.69}-.9999999976)e^{-k}+ .9999999976], \tag{5.44}$$

$$\text{and} \quad (1+\hat{b}s_k)e^{-\hat{b}s_k} = 1-\frac{k}{\hat{a}}$$

$$\Longrightarrow (1+.000047s_k)e^{-.000047s_k} = 1-\frac{k}{90.0624}. \tag{5.45}$$

The observed and the estimated s_k's for both of the models are given in Table No.5.13. For the Yamada the Fit (SSD)= $\sum_{k=1}^{86}(s_k-\hat{s}_k)^2$ =3876319507 and for the modified Yamada it is 3789328267. Clearly, the Modified shows a better fit.

4. DISTRIBUTION OF FAULTS REMAINING AFTER DEBUGGING

$$P_4[\bar{N}(t)=k] = \frac{[\log\frac{e^{a(1+bt)e^{-bt}}-c}{1-c}]^k}{k!}\left(\frac{1-c}{e^{a(1+bt)e^{-bt}}-c}\right),$$

$$= \frac{[\log\frac{e^{89.69(1+.000048t)e^{-.000048t}}-.9999999976}{1-.9999999976}]^k}{k!}\times \left(\frac{1-.9999999976}{e^{89.69(1+.000048t)e^{-.000048t}}-.9999999976}\right),$$

$$\text{where} \quad k=0,1,2,\ldots \tag{5.46}$$

Table No.5.18 Comparison of Results Based on the Yamada *et al* and Modified Yamada *et al* models.

Error No. k	Actual Failure Time s_k, (days)	Estimated Failure Time $\hat{s}_k$ Yamada *et al*	Modified Yamada *et al*
1	479	3318.29	3298.28
2	745	4802.04	4773.33
3	1022	5990.06	5954.51
4	1576	7027.82	6986.39
5	2610	7971.56	7924.85
6	2859	8850.04	8798.50
7	3552	9680.35	9624.30
8	4149	10473.62	10413.32
9	4266	11237.61	11173.27
10	4436	11977.95	11909.76
11	4553	12698.93	12627.03
12	5827	13403.90	13328.43
13	6296	14095.52	14016.61
14	7470	14776.01	14693.75
15	8163	15447.17	15361.66
16	10071	16110.56	16021.89
17	10206	16767.50	16675.75
18	10483	17419.15	17324.40
19	11079	18066.54	17968.85
20	11836	18710.56	18610.01
21	12273	19352.04	19248.68
22	14503	19991.72	19885.63
23	14940	20630.29	20521.52
24	15280	21268.39	21156.99
25	15685	21906.59	21792.63
26	16260	22545.47	22429.00
27	16537	23185.55	23066.62
28	16900	23827.34	23706.02
29	17422	24471.34	24347.68
30	18035	25118.02	24992.07
31	18312	25767.85	25639.67
32	19612	26421.28	26290.93
33	20433	27078.78	26946.32
34	20646	27740.79	27606.28
35	22266	28407.76	28271.28
36	23867	29080.17	28941.77

Table No.5.18 continued.

Error No. k	Actual Failure Time s_k, (days)	Estimated Failure Time $\hat{s}_k$ Yamada *et al*	Modified Yamada *et al*
37	24165	29758.47	29618.23
38	25039	30443.13	30301.11
39	25657	31134.33	30990.92
40	28297	31833.47	31688.15
41	28302	32540.16	32393.31
42	28451	33255.22	33106.94
43	29485	33979.21	33829.59
44	31926	34712.69	34561.84
45	32386	35456.25	35304.28
46	32951	36210.53	36057.56
47	34070	36976.18	36822.34
48	34507	37753.90	37599.33
49	35434	38544.41	38389.26
50	39896	39348.50	39192.94
51	40610	40167.00	40011.20
52	40791	41000.79	40844.96
53	42276	41850.83	41695.17
54	43033	42718.12	42562.87
55	46187	43603.77	43449.19
56	48632	44508.96	44355.33
57	49516	45434.97	45282.60
58	51553	46383.19	46232.43
59	53034	47355.14	47206.38
60	53593	48352.48	48206.15
61	54083	49377.03	49233.61
62	54676	50430.79	50290.81
63	56445	51515.97	51380.04
64	56530	52635.03	52503.84
65	59366	53790.70	53665.03
66	59579	54986.04	54866.78
67	61445	56224.49	56112.66
68	61935	57509.95	57406.72
69	63372	58846.83	58753.55
70	67694	60240.18	60158.44
71	69112	61695.83	61627.47
72	70135	63220.51	63167.73

Table No.5.18 continued.

Error No. k	Actual Failure Time s_k, (days)	Estimated Failure Time $\hat{s}_k$ Yamada *et al*	Modified Yamada *et al*
73	75625	64822.10	64787.52
74	77145	66509.89	66496.64
75	80426	68294.93	68306.84
76	83142	70190.55	70232.31
77	85317	72212.96	72290.42
78	88822	74382.23	74502.81
79	89547	76723.59	76896.83
80	91510	79269.39	79507.81
81	95489	82062.04	82382.45
82	96579	85158.74	85584.21
83	96824	88639.27	89202.04
84	98018	92619.81	93365.28
85	99012	97279.13	98269.07
86	102914	102914.00	104217.94

From Table 5.19 it is clear that the $P_4(\bar{N}(102914) \leq 36) = 0.9934$. Therefore the probability that thirty six or fewer faults remained in the system after observing for 102914 execution time in seconds is about 0.99.

Table No.5.19 Distribution of Remaining. Faults After Debugging.

k	$P_4[\bar{N}(102914) \leq k]$
5	0.00000
6	0.00002
7	0.00006
8	0.00019
9	0.00054
10	0.00135
11	0.00310
12	0.00655
13	0.01281
14	0.02339
15	0.04005
16	0.06468
17	0.09892
18	0.14388
19	0.19982
20	0.26593
21	0.34034
22	0.42029
23	0.50246
24	0.58339
25	0.65990
26	0.72947
27	0.79037
28	0.84178
29	0.88369
30	0.91671
31	0.94188
32	0.96048
33	0.97380
34	0.98306
35	0.98932
36	0.99342
37	0.99605

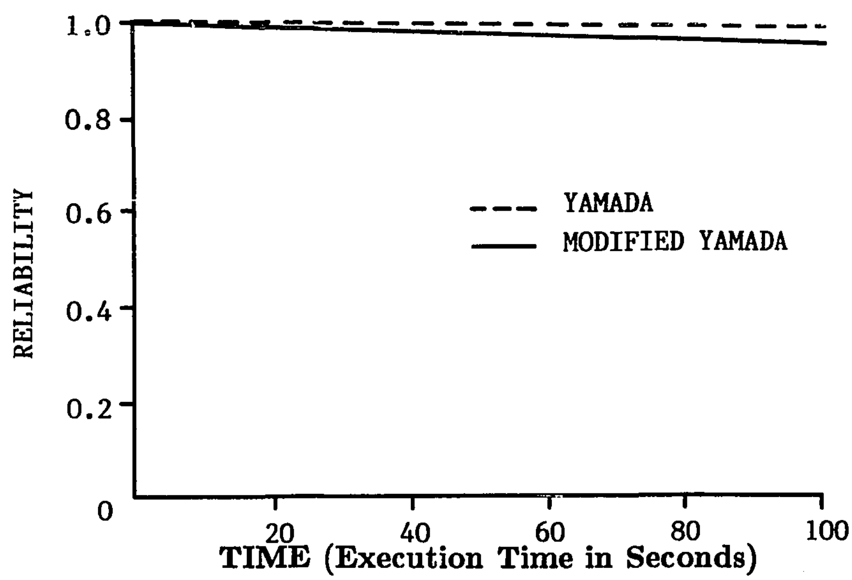

Fig.5.8 Plots of Conditional Reliability Functions of Modified Yamada *et al* & Yamada *et al* Models.

The estimated conditional reliability function for the modified Yamada *et al* model given $s_{86} = 102914$ is

$$\hat{R}_2(t|s_{86}) = \frac{e^{89.69[1+.000048(102914+t)]e^{-.000048(102914+t)}} - .9999999976}{e^{89.69(1+.000048\times 102914)e^{-.000048\times 102914}} - .9999999976} \tag{5.47}$$

and that of the G-O model is

$$\hat{R}(t|s_{86}) = e^{90.06(e^{-.000047\times 102914} - e^{-.000047(102914+t)})}. \tag{5.48}$$

The plot of these reliability functions versus time is shown in Fig.5.12.

REFERENCES

[1] Abdel-Ghaly, Abdalla A., Chan, P. Y., and Littlewood, B., (1986). Evaluation of competing software reliability predictions. *IEEE Trans. On Software Eng.*, vol. SE-12, No.9, pp. 950-967.

[2] Barlow, R. E. and Proschan, F., (1975). Statistical Theory of Reliability and Life Testing. *Hold, Rinehart and Winston Inc.*

[3] Cox, D. R. and Lewis, P. A. W., (1966). The Statistical Analysis of Series of Events. Methuen, London.

[4] Dale, C. J., (1982). Software reliability evaluation methods. *British Aerospace Dynamics Group, Rep. ST-26750.*

[5] Foreman, E. H. and Singapurwalla, N. D., (1977). An empirical stopping rule for debugging and testing computing software. *J. Amer. Statist. Ass.*, vol. 72, pp. 750-757.

[6] Goel, Amrit L., (1985). Software reliability models: Assumptions, limitations, and applicability. *IEEE Trans. On Software Engg.*, vol. SE-11, No.12, pp. 1411-1423.

[7] Goel, A. L., Okumoto, K., (1978). An analysis of recurrent software errors in real-time control system. *Proc. ACM Conf.* pp. 496-501.

[8] Goel, A. L. and Okumoto, K., (1979). Time-dependent error-detection rate model for software reliability and other performance measures. *IEEE Trans. On Rel.,* vol. R-28, pp. 206-211.

[9] Jelinski, Z. and Moranda, P. B., (1972). Software reliability research. *Statistical Computer Performance Evaluation,* W. Freiberger, Ed. New York: Academic, pp.465-484.

[10] Joe, H. and Reid, N., (1985a). Estimating the number of faults in a system. *Journal of American Statist. Ass.,* vol. 80, pp. 222-226.

[11] Joe, H. and Reid, N., (1985b). On the Software Reliability Models of Jelinski-Moranda and Littlewood. *IEEE Trans. On Rel.,* vol. R-34, pp. 216-218.

[12] Littlewood, B., (1981). Stochastic reliability growth: A model for fault-removal in computer programs and hardware designs. *IEEE Trans. On Rel.,* vol. R-30, pp. 313-320.

[13] Littlewood, B. and Verrall, J. L., (1973). A Bayesian reliability growth model for computer software. *IEEE Symposium on Computer Software Reliability.* New York, Apr. 30- May 2, 1973, pp.70-76.

[14] Littlewood, B. and Verrall, J. L., (1974). A Bayesian reliability model with a stochastically monotone failure rate. *IEEE Trans. Relaibility,* vol. R-23, pp. 108-114.

[15] Littlewood, B. and Verrall, J. L., (1981). Likelihood function of a debugging model for computer software reliability. *IEEE Trans. On Rel.*, vol. R-30, pp. 145-148.

[16] Musa, J. D., (1975). A theory of software reliability and its application. *IEEE Trans. On Software Eng.*, vol. SE-1, pp. 312-327.

[17] Musa, J. D., and Okumoto, k., (1983). Software Reliability Models: Concepts, Classification, Comparisons, and Practice. *Electronic Systems Effectiveness and Life Cycle Costing.* NATO ASI series, F3, Springer-Verlag, Heidelberg, pp.395-424.

[18] Musa, J. D., and Okumoto, k., (1984). A Logarithmic Poisson Execution Time Model for Software Reliability Measurement. *Proc. Seventh International Conference on Software Eng.*, Orlando, pp. 230-238.

[19] Musa, J. D., Iannino, A. and Okumoto, K., (1987). Software Reliability: Measurement, Prediction, Application. *McGraw-Hill Series in Software Engineering and Technology.*

[20] Ohba, M., (1981). Software reliability analysis models. *IBM J. Res.*

Develop., vol. 28, no.4, pp. 428-443.

[21] Rohatgi, V. K., (1976). An Introduction to Probability Theory and Mathematical Statistics, Wiley.

[22] Ross, S. M., (1983). *Stochastic Processes.* New York: Wiley.

[23] Schick, Goerge J., and Wolverton, Ray W., (1973). Assesment of software reliability. *Proceedings of Operations Research,* Physica-Verlag, Wurzburg-Wien, pp. 395-422.

[24] Schick, Goerge J., and Wolverton, Ray W., (1978). An analysis of Competing software reliability models. *IEEE Trans. On Software Engg.*, vol. SE-4, No.2, pp. 104-120.

[25] Shooman, M. (1972). Probabilistic Models for Software Reliability Prediction. *Statistical Computer Performance Evaluation,* Academic, New York, pp. 485-502.

[26] Yamada S. and Osaki, S., (1985). Software reliability growth modeling: Models and applications. *IEEE Trans. On Software Engg.*, vol. SE-11, NO.12, pp. 1431-1437.

[27] Yamada, S., Ohba S. and Osaki, S., (1983). S-shaped reliability growth modeling for software error detection. *IEEE Trans. On Rel.*, vol. R-32, pp. 475-478,484.