Inference in nonhomogeneous Poisson process models, with applications to software reliability

by

Jacinte Jean

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Statistics

Waterloo, Ontario, Canada, 1998

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

Nonhomogeneous Poisson Process (NHPP) models are commonly used to model recurrent events (failures or repairs) in repairable systems which fail or break down many times during their lifetime. NHPP models having an intensity of the form $\lambda(t;\boldsymbol{\eta}) = v\,\lambda_0(t;\boldsymbol{\beta})$, for scalar $v > 0$ and each component of the vector $\boldsymbol{\beta}$ positive-valued, are widely used in modelling the times of occurrences of failures in the debugging phase of software development. In software system reliability applications, these models are used to predict future behaviour of the occurrence of failures and to provide information for making decisions on when to stop testing. In this thesis, we have addressed statistical issues pertaining to parameter estimation, model verification, and interval prediction for NHPP models having an intensity of the above form. In Chapter 2, we assess the maximum likelihood estimation procedure used to obtain estimates for v and $\boldsymbol{\beta}$ for specific models belonging to the above general family of NHPP models. In particular, we study conditions under which a finite, positive-valued maximum likelihood estimate for v is obtained, consider choices of parameterization to facilitate estimation, and consider the effects of total test time on these matters. In Chapter 3, we propose a new approach for testing the goodness of fit of NHPP models of the above general form. We also suggest two alternative models that include specific software reliability models of interest. These models are of use for testing the goodness of fit of their submodels. In Chapter 4, we propose a frequentist approach for providing approximate interval predictors of $N2 = N(T1, T2]$, the number of events in the future time interval $(T1, T2]$, based on the observed data up to time $T1$. We also use this method to assess the effect

of data accumulation on prediction of $N3 = N(T1, \infty]$, the number of remaining events to be eventually observed given data has been observed up to time $T1$. We also discuss how to obtain Bayesian prediction intervals and compare them with the frequentist-based prediction intervals for $N3 = N(T1, \infty]$ in some examples. In Chapter 5, we discuss research areas to be investigated further. The problems presented here are not unique to the software reliability context. In fact, the results of this thesis may be extended to various reliability applications in which NHPP models of the above form are of use.

Acknowledgements

I thank all my friends and colleagues in the Department of Statistics and Actuarial Science, as well as to those living in Minota Hagey Residence. My life spent in Waterloo has been enriched greatly by the thoughtful, caring, generous people with whom I have worked and lived all these years. Also, I thank my friends and family living afar for their emotional support "across the miles".

I give special thanks and appreciation to the following individuals, without whom I would not have finished this goal:

- God, for sending me the following earthly angels in this time of challenge.
- Jerry Lawless, my thesis supervisor, for his great insight, generosity, patience, and determination. I thank him for giving me the opportunity to pursue this goal to completion.
- Prasad Gudem, my fiancé, for his great love, generosity, insight, and encouragement. I thank him for the hours spent comforting me in times of great stress and despair. In addition, I appreciate greatly our numerous adventurous journeys taken together throughout these years, especially that unforgettable weekend in Hawaii!
- Mary Thompson, David Matthews, and Jock MacKay, my thesis committee members, for their input. I appreciate their time spent in reading and commenting on my thesis.

- Hugh Chipman, Daniel Fong, Karen Kopciuk, Ronnie Lee, Rhonda Rosychuk, Lara Wolfson, and Min Zhan, my friends and colleagues in department, for their encouragement and time spent in statistical discussions.

- Darlene and Henry Jean, my parents, for encouraging a sense of of adventure, independence, determination, and humour in their children. I appreciate greatly their faith in my ability to achieve this goal.

- Anisa Yasmin, my dear friend and social activities partner, for the many hours of encouragement, comfort, laughter and fun throughout these years. We did it!

- Audrey Jahrig and Anne Murland, my dear friends in Alberta, for their many letters, cards, gifts and phone calls filled with numerous words of encouragement, comfort and laughter. I will always treasure the many hours of joy that their friendship has brought to my life.

- All my teachers at the Ashmont Community School for their encouragement throughout the years. In particular, Rod Ayres, for sparking my interest in mathematics by teaching me the Pythagoras theorem while I was yet in elementary school.

In memory of my grandparents, Gerda & Carl Smith, and Rose & Ovila Jean.

Contents

List of Tables

List of Figures

Chapter 1

Data, model, and thesis information

1.1 Introduction

Nonhomogeneous Poisson Process (NHPP) models are commonly used to model recurrent events (failures or repairs) in repairable systems which fail or break down many times during their lifetime. Such models are justified when a system has many components that may fail and repairing a failed component has little effect on the overall system reliability. Many examples and references are given in [8], [35], [27], [30], and [26]. The most widely used and discussed NHPP model is the power law process (also known as the "Weibull process" [36]). Its popularity is not only due to its usefulness for modelling systems which are either deteriorating (times between events are getting shorter) or improving (times between events are getting longer), but for the simplicity of statistical inferences associated with the model

relative to other NHPP models. In particular, a combination of the existence of a closed form maximum likelihood estimation (m. l. e.) solution for this model and the inherent form of its intensity function has lead to nice distributional results for its estimated parameters and their functionals. Constructing confidence intervals and tests of hypotheses for the model's parameters ([9], [28], [29], [39], [42], [70], [79]), prediction intervals ([38]), and goodness of fit tests ([10], [11], [63], [69], [87], [92], [93]) have been discussed at length in the literature. Although it is appropriate for many reliability applications, the power law process may be an inadequate NHPP model in other situations, such as software system reliability.

There has been a rising interest in modelling software system reliability over the last twenty-five years (e. g. of review papers include [16], [47], [85], [96], [98], [99], [100]). A vast literature exists on the modelling strategies used in the various phases of software system development and management. We focus on the testing and repair (debugging) phases of software program development. During this phase, the fundamental objective of assessing the software system's reliability is to predict future behaviour of the occurrence of failures and to provide information for making decisions on when to stop testing. It is during this phase that General Order Statistics (GOS) and NHPP families of models are often used to model the times of occurrences of failures. Although the GOS family has been the most widely used throughout the years, the NHPP family of models has gained popularity in the last ten years, due in part to the existence of an equivalence relationship between the two models and the relative ease of application of the NHPP models compared to those of their GOS counterparts. Regardless of the reasons for choosing the NHPP

family over the GOS family, the increased usage of NHPP models in modelling failures in the debugging phase of software development provides an incentive for obtaining information about how well these models perform in terms of predicting future behaviour of events and providing information for making decisions on when to stop testing.

In this chapter, we first discuss briefly the nature of software reliability data. Secondly, we introduce the GOS and NHPP families of models often used in the analysis of this data followed by a brief discussion on their equivalence. Next, we discuss the ways in which these models are used to predict future behaviour of the occurrence of failures and to provide information for making decisions on when to stop testing. Lastly, we motivate and provide an outline of the research areas covered in this thesis.

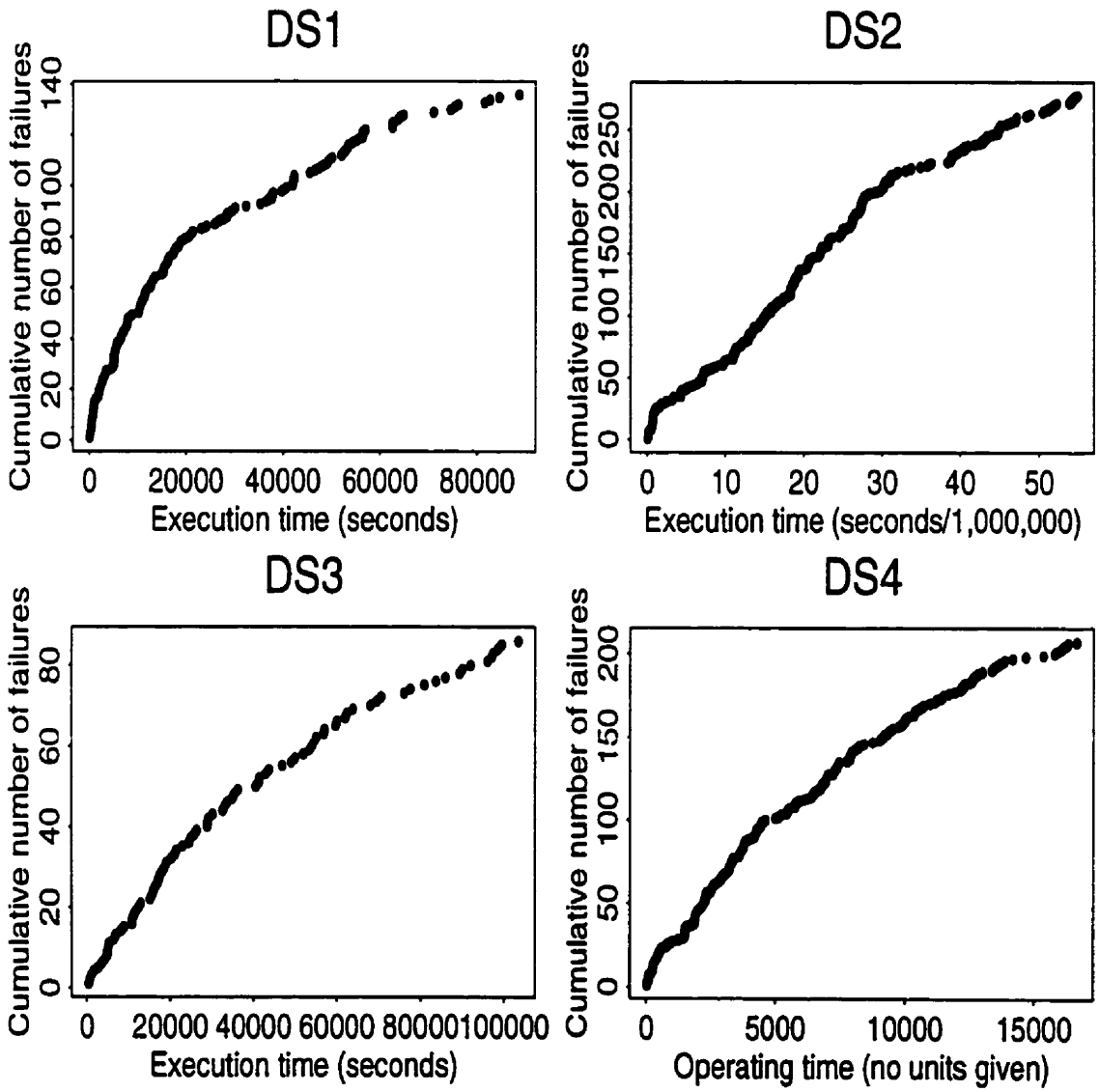
1.2 Software reliability data

A brief overview of the basic software reliability modelling definitions and concepts is given in this section. A thorough discussion may be found in [83] and [96]. In this context, a *program* refers to a particular block (unit) or structure of blocks (module) of software code. Each block or structure of blocks is modelled separately. Although software failures in the program may not be generated at random, it is assumed that the underlying testing process that detects the failures is random. A *random testing strategy* is one whereby inputs are chosen from the set of all possible inputs in a random manner. The specification of input variables used to test the program and their relative frequencies is called the program's *operational profile*. A program

test *run* maps the input variables from the operational profile to a set of output variables in a certain amount of time. One or more *failures* occur in a run when output variables do not conform to prespecified requirements. Time is often given in units of *execution* (operation) or *calendar* time. The former measure of time is generally considered (e. g. [83]) to be the most accurate reflection of software stress and reliability; however, calendar time is used by some to avoid the difficulties in measuring execution time. Numerous methods of measuring execution time exist. Few authors provide detailed remarks on the procedures used to measure execution time.

As mentioned earlier, the GOS and NHPP models are commonly used for modelling the occurrence of failures in the debugging phase of software program development. Although they are not highly realistic in this context, these models have their uses. In Chapter 5, we discuss further improvements to modelling. These models are justified when a software system has many components that may fail and repairing a failed component has little effect on the overall system reliability. The program failures are assumed to be of a similar type and the interaction of software and hardware failures in the computer system is ignored in these models. The data is given in the form of a list of software program failure or interfailure detection times (ungrouped data) or as the number of failures in a given interval (grouped data). There are numerous grouped and ungrouped data sets available in the literature, few of which include supplementary information to the failure times. Both *time-truncated* and *failure-truncated* data sets are also available. In time-truncated data sets, the data is collected over a prespecified time interval, whereas

Figure 1.1: Plots of cumulative number of failures versus time for DS1-4.



in failure-truncated data sets, the data is collected until a prespecified number of events have been observed. A description of four typical (ungrouped) data sets that we consider throughout the thesis is given below. Plots of the cumulative number of failures versus time for each of the data sets considered is given in Figure 1.1. This thesis will focus on the ungrouped data case for the most part.

1.2.1 Specific data sets

DS1

The first data set, henceforth denoted as DS1, consists of interfailure time data collected over the complete system test and operational phases of a real-time command and control software system. This well-known data set, originally collected and discussed in [81, system 1], consists of a set of 136 interfailure times, each listed with the corresponding number of working days from the start of the system test until that particular failure. The interfailure times are in seconds, measured in *running clock time*, which is proportional to execution time. This time-truncated data set was observed until time $T1 = 91208$ seconds. It is also known that nine programmers worked on the 21,700 delivered object instructions of code. This data set has been analyzed and discussed by many other authors, including [2] and [21].

DS2

The second data set, DS2, consists of interfailure data collected over part of the operational phase of a word processing software system. This phase of the system was still underway at the time of the report originally made by [81, sub-system

3]. The interfailure times are in seconds, measured in execution time. This time-truncated data set was observed until time $T1 = 54.93336 \times 10^6$ seconds. As far as we know, subsequent failure data on this system has not been released. DS2 consists of a set of 278 interfailure times, also listed with the corresponding number of working days from the start of the operating phase until that particular failure. An unknown number of programmers have worked on the hundreds of thousands of delivered object instructions of code in this system, and it is noted in [81] that the system was subjected to small design changes throughout the period the data was collected. Regardless of these design changes, the system was noted to be more stable than other similar systems considered in that report. Numerous other authors have considered this data, including [21] and [22].

DS3 and DS4

We also considered two lesser-known data sets, DS3 and DS4, both analyzed and discussed in [2]. DS3 consists of 86 interfailure times, measured in seconds of execution time, whereas DS4 consists of 207 interfailure times, measured in operating time (no units given). These failure-truncated data sets were observed until time $T1 = 103,334$ seconds and $T1 = 16648$ units, respectively. The unique feature in DS4 is that its failures are due to software faults *and* hardware design faults. DS4 was also analyzed in [55].

1.3 Software model information

1.3.1 GOS family

The original software reliability models that modelled failure or interfailure times (e.g. Jelinski-Moranda [54], Littlewood [73] and [74], Musa [80]) belong to the GOS family of models. Under this family, the initial number of failures in the software is assumed to be an unknown fixed integer N. It is supposed further that the times to detecting each software failure are considered order statistics from a sample of N independent failure times that belong to a parametric family, $p(t;\psi)$, where ψ is a parameter vector.

We start with N software failure times and we observe the n smallest over $(0, T1]$ at times $t_1 < t_2 < \ldots < t_n$. With n out of N bugs detected by time $T1$, we can consider the software system as having n "failures" and $N-n$ "survivors" by time $T1$. It then follows from general results on order statistics or on censored samples that the joint probability distribution function (and hence likelihood function) of n and these first n failure times is

$$L(N,\psi) \;=\; N^{(n)} \left(\prod_{i=1}^{n} p(t_i;\psi)\right) \{S(T1;\psi)\}^{(N-n)},$$

where $S(T1;\psi) = \int_{T1}^{\infty} p(y;\psi)dy$ is the survival function corresponding to $p(t;\psi)$ and $N^{(n)} = N(N-1)\cdots(N-n+1)$ denotes the possible number of ordered samples of size n from N. Since it is usually more convenient to work with the log-likelihood, we can take logs (here, $log = log_e$, unless otherwise noted) on both

sides of the above equation to obtain

$$l(N,\psi) = \sum_{i=1}^{n}\log(N-i+1) + \sum_{i=1}^{n}\log\{p(t_i;\psi)\} + (N-n)\log\{S(T1;\psi)\}.$$

Note that $N \geq n$ and ψ are to be estimated here.

1.3.2 NHPP family

The NHPP family of models (e. g. Goel-Okumoto [46], Musa-Okumoto [84], Schneidewind [94], Yamada et. al. [102]) have recently become more popular in modelling software reliability. Let $\{N(t); t \geq 0\}$ be a NHPP with intensity function $\lambda(t;\eta)$ that represents the cumulative number of software failures found by time t. We observe a single sample path of the process over $(0, T1]$, with n error detections (events) occurring at times $t_1 < t_2 < \ldots < t_n$. The probability of observing no events in $(0, t_1)$, one event in $(t_1, t_1 + \delta t_1)$, no events in $(t_1 + \delta t_1, t_2)$, one event in $(t_2, t_2 + \delta t_2)$, and so on up to no events in $(t_n + \delta t_n, T1)$ is for small $\delta t_1, \delta t_2, \ldots, \delta t_n$,

$$\begin{aligned} L(\eta) = {} & \exp\left(-\int_0^{t_1}\lambda(y;\eta)dy\right)\lambda(t_1;\eta)\delta t_1 \exp\left(-\int_{t_1+\delta t_1}^{t_2}\lambda(y;\eta)dy\right)\lambda(t_2;\eta)\delta t_2 \\ & \cdots \exp\left(-\int_{t_n+\delta t_n}^{T1}\lambda(y;\eta)dy\right). \end{aligned}$$

Dividing through by $\delta t_1, \delta t_2, \ldots, \delta t_n$ and letting $\delta t_i \to 0$ gives us,

$$L(\eta) = \prod_{i=1}^{n}\lambda(t_i;\eta)\exp(-\int_0^{T1}\lambda(y;\eta)dy). \tag{1.1}$$

Since we know that the mean value function is simply $E\{N(t)\} = \Lambda(t;\eta) = \int_0^t \lambda(y;\eta)dy$, we can substitute this into the above to obtain the following log-likelihood,

$$l(\eta) = \sum_{i=1}^{n} \log\{\lambda(t_i;\eta)\} - \Lambda(T1;\eta). \tag{1.2}$$

Note that the vector of parameters η is to be estimated here.

1.3.3 Equivalence of GOS and NHPP models

Although there are fundamental underlying differences in the motivation behind each of the two families, an equivalence relationship exists between the GOS and NHPP models, as shown in general in [77]. For every specified GOS family, one can obtain an "equivalent" NHPP family by relaxing the assumption that N is fixed and assuming further that it is a Poisson random variable with finite mean. For instance, under an exponential order statistics (EOS) model, N is an unknown fixed constant and $p(t;\psi) = \beta \exp(-\beta t)$. If we assume further that N is distributed as a Poisson random variate with finite mean v, we obtain the equivalent EOSN (my notation) NHPP model with intensity function $\lambda(t;\psi) = v\beta \exp(-\beta t)$. On the other hand, for any NHPP model that has $\Lambda(\infty;\eta) < \infty$, one can obtain an "equivalent" GOS family by conditioning on the value of $N = N(\infty)$. In fact, for a specified NHPP model with intensity function $\lambda(t;\eta)$, the corresponding GOS model will have

$$p(t;\eta) = \frac{\lambda(t;\eta)}{\Lambda(\infty;\eta)}.$$

Table 1.1: A list of NHPP models used in modelling software reliability.

Model	$\lambda(t;\boldsymbol{\eta})$	$\Lambda(t;\boldsymbol{\eta})$
$\Lambda(\infty)$ *finite*		
EOSN	$v\beta\,\exp(-\beta\,t)$, $v,\beta>0$	$v[1-\exp(-\beta\,t)]$
GAMN	$v\beta^2\,t\,\exp(-\beta\,t)$, $v,\beta>0$	$v[1-(1+\beta\,t)\,\exp(-\beta\,t)]$
WOSN	$v\beta_1\beta_2\,t^{\beta_2-1}\,\exp(-\beta_1\,t^{\beta_2})$, $v,\beta_1,\beta_2>0$	$v[1-\exp(-\beta_1\,t^{\beta_2})]$
POSN	$\frac{v\beta_2\,\beta_1^{\beta_2}}{(\beta_1+t)^{(\beta_2+1)}}$ $v,\beta_1,\beta_2>0$	$v[1-\left(\frac{\beta_1}{\beta_1+t}\right)^{\beta_2}]$
$\Lambda(\infty)$ *infinite*		
LOGN	$\frac{v\beta}{\beta\,t+1}$, $v,\beta>0$	$v\,log(1+\beta\,t)$
POWN	$v\beta\,t^{\beta-1}$, $v,\beta>0$	$v\,t^{\beta}$

For instance, if we assume a NHPP model with $\lambda(t;\boldsymbol{\psi}) = v\beta_1\beta_2\,t^{\beta_2-1}\,\exp(-\beta_1\,t^{\beta_2})$, the corresponding GOS model will have $p(t_i;\boldsymbol{\eta}) = \beta_1\beta_2\,t^{\beta_2-1}\,\exp(-\beta_1\,t^{\beta_2})$. This is simply the Weibull order statistics (WOS) model. At this point, we want to stress that it is not possible to distinguish between a family of NHPP models and their related family of GOS models without proper replication of a recurrent event process. A thorough discussion on this matter is given in [77].

1.3.4 Specific models

Table 1.1 gives a list of the most popular NHPP models used in modelling the times of occurrences of failures in the debugging phase of software development. As noted in [83], the NHPP models widely used in software reliability belong to a family of NHPP models having intensity functions of the particular form, $\lambda(t;\boldsymbol{\eta}) = v\,\lambda_0(t;\boldsymbol{\beta})$, for scalar $v > 0$ and each component of the vector $\boldsymbol{\beta}$ positive-valued (henceforth denoted $\boldsymbol{\beta} > 0$). This family is split further into two subfamilies based on the limiting behaviour of a model's mean value function $v\,\Lambda_0(t;\boldsymbol{\beta})$, where $\Lambda_0(t;\boldsymbol{\beta}) =$

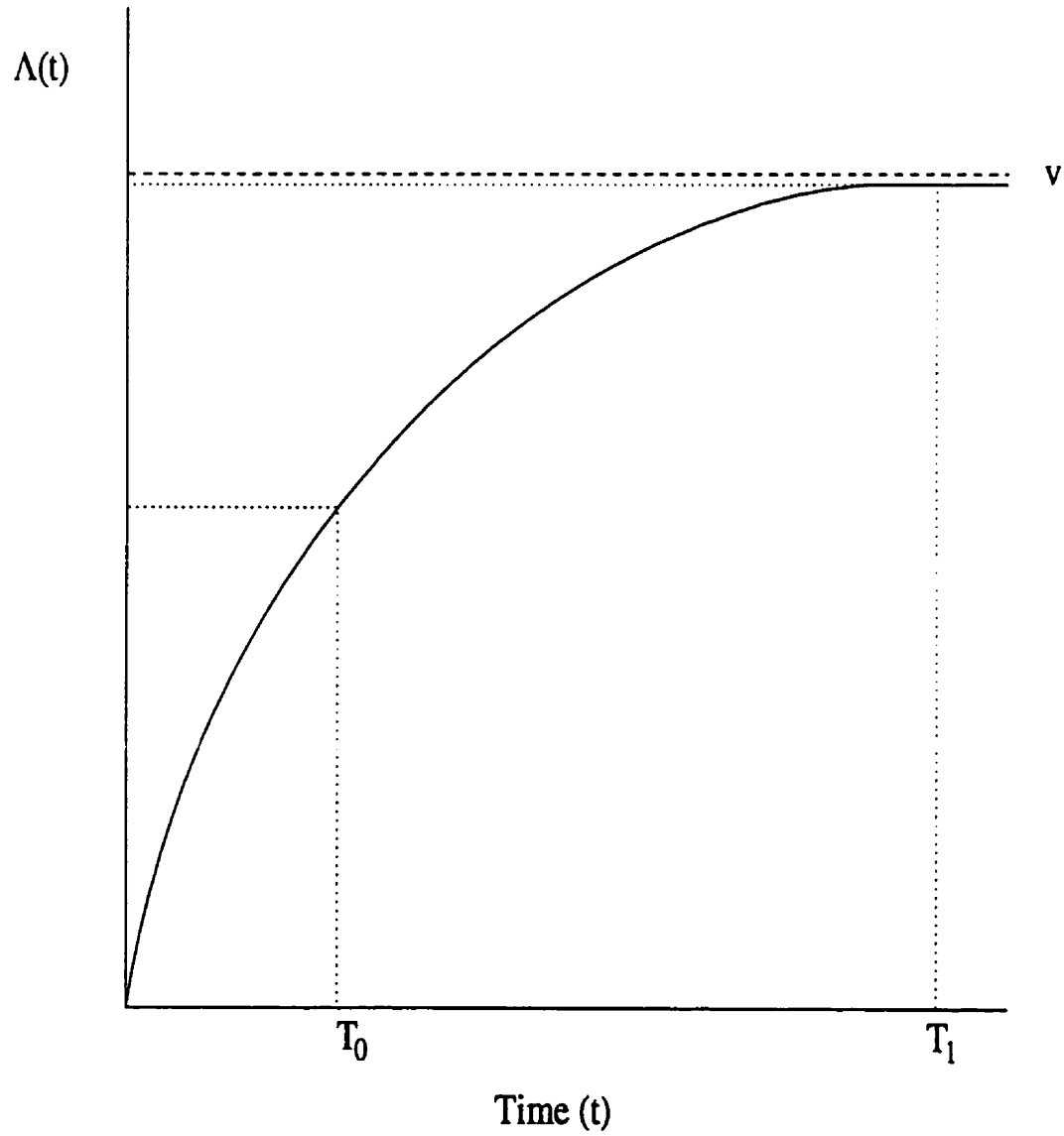
$\int_0^t \lambda_0(s;\boldsymbol{\beta})\,ds$. The *finite* subfamily consists of those models which have a mean value function with a finite limit when $t \to \infty$. In particular, for these models, as $t \to \infty$, $\Lambda(t; v, \boldsymbol{\beta}) \to v$ because $\Lambda_0(t;\boldsymbol{\beta}) \to 1$. For this to occur, we note that $\lambda_0(t;\boldsymbol{\beta})$ has a probability density functional form with corresponding cumulative distribution functional form $\Lambda_0(T1;\boldsymbol{\beta})$. Referring to Table 1.1, we note that the finite NHPP models are named according to the probability density functional form $\lambda_0(t;\boldsymbol{\beta})$. In other words, the models are named according to the particular $p(t;\boldsymbol{\eta})$ that would be assumed in its GOS counterpart. The parameter v in these models represents the expected number of errors to be eventually detected. Since the fundamental objective of assessing the software system's reliability during the debugging phase is to predict the future behaviour of the occurrence of failures and to provide information for making decisions on when to stop testing, this parameter is of special practical interest. In contrast, the *infinite* subfamily is one which has a mean value function with infinite limit when $T1 \to \infty$. The POWN and LOGN infinite NHPP models denote power law and logarithmic Poisson processes, respectively. Although it is rarely appropriate in this context, the POWN model is included because of its popularity in general reliability applications. In software reliability literature, the EOSN model with the parameterization given in Table 1.1 is widely referred to as the Goel-Okumoto [46] model, and with parameterization $p = v\beta$ and $\beta = \beta$ is known as the Schneidewind [94] model. The GAMN model is widely known as the "delayed s-shaped" model by Yamada et. al. [102], whereas the LOGN model is widely known as the Musa-Okumoto [84] model. The POSN model is the NHPP equivalent of the Littlewood [73] POS model.

1.4 Software reliability quantities of interest

Recall that the fundamental reason for assessing a software system's reliability during the debugging phases of software program development is to predict future behaviour of the failures and to provide information about when to stop testing or to release the software. An immediate question arises as to what particular quantities are of interest. For repairable systems, such as a software system, it is appropriate to assess the *number* and *pattern* of successive failures rather than to consider the time to any particular failure. In other words, it is of interest to assess the behaviour of the entire process as a whole, as opposed to assessing the distribution of time to any particular failure. For a thorough argument on this issue, please refer to [8]. With this in mind, we believe that appropriate measures of interest for predicting future behaviour of the failure process and for providing information about when to stop testing are a) $N2 = N(T1, T2]$, the number of events in the future time interval $(T1, T2]$, and; b) $N3 = N(T1, \infty]$, the number of remaining events to be eventually observed, both measures predicted from observed data up to time $T1$.

Also recall that for the *finite* subfamily of NHPP models, v represents the expected number of errors to be eventually detected. This quantity is also of special interest because it can provide information about when to stop testing. An illustrative plot of $\Lambda(t)$ versus t is given in Figure 1.2. The data in $(0, T_0]$ generally will not lead to a very precise estimate of v, whereas that in $(0, T_1]$ will. In other words, the amount of information on v depends on what fraction of the eventual faults has been observed. Since a "good" (defined as having low variability and

Figure 1.2: Illustrative plot of $\Lambda(t)$ versus t.



low bias) estimate for v is needed to accurately predict $N2$ and $N3$, an important design issue is to ensure that $T1$ is large enough to obtain good predictions of these quantities.

1.5 Thesis motivation and outline

As mentioned earlier, the main reason for assessing a software system's reliability during the debugging phases of software program development is to predict future behaviour of the failures and to provide information about when to stop testing. Since NHPP models are used to model the occurrence of failures during this phase of development, it is of particular interest to a) ensure there are adequate methods available for using NHPP models to perform these tasks, and; b) determine how well NHPP models perform these tasks. In this section, we will motivate the areas of interest in this thesis and mention how they relate to the above objective in software reliability. We follow with an outline of the thesis.

We argue that since NHPP models are used to predict future behaviour of the failure process to provide information about when to stop testing, it is of prime interest to ensure that there is an adequate method for predicting quantities of interest under an assumed NHPP model. As pointed out in Section 1.4, we are particularly interested in obtaining interval predictors for $N2 = N(T1, T2]$, the number of events in the future time interval $(T1, T2]$, based on the observed data up to time $T1$. Since there is not an "exact" pivotal available for predicting such quantities in the NHPP scenario, we examined an approach based on approximate pivotals that may be used for all NHPP models. We also used this method to assess

the effect of data accumulation on prediction of $N3 = N(T1, \infty]$, the number of remaining events to be eventually observed given data has been observed up to time $T1$. We also obtained Bayesian prediction intervals and compared them with the frequentist-based prediction intervals for $N3 = N(T1, \infty]$ in some examples.

With a method available for predictive purposes, it is of interest to assess and reduce the sources of error in prediction. In general, there are three sources of error involved when predicting an unknown quantity of interest. First, there will always be *inherent variation* in the true process. In other words, even if we assume that the process $\{N(t); t \geq 0\}$ is an NHPP with *known* intensity function $\lambda(t; \eta) = v\,\lambda_0(t; \boldsymbol{\beta})$, for scalar $v > 0$ and each component of the vector $\boldsymbol{\beta}$ is finite valued (henceforth denoted $\boldsymbol{\beta} > 0$), we can not predict exactly the quantity of interest simply due to random variability in the NHPP itself. Essentially, this is an uncontrollable source of variability. Second, another source of error in predictions is due to *sampling variation*. This arises from having to estimate the parameters in the assumed model based on the observed process. As a result, it is important to be able to assess how well the chosen estimation method performs. In our situation, we estimate v and $\boldsymbol{\beta}$ in the intensity $\lambda(t; \eta) = v\,\lambda_0(t; \boldsymbol{\beta})$ using maximum likelihood estimation. It is important to assess how well this estimation procedure performs and to make improvements on the estimation process, if necessary. Lastly, the largest potential source of error in predictions is due to *model bias*. This may arise in two ways. For one thing, we may have assumed an incorrect model in making the predictions. As a consequence, it is very important to be able to adequately assess the goodness of fit of the assumed model. In our situation, we are particularly interested in having

adequate measures of goodness of fit of NHPP models having an intensity of the form $\lambda(t;\boldsymbol{\eta}) = v\,\lambda_0(t;\boldsymbol{\beta})$. Even though the assumed model does fit the observed data well, another source of model bias may arise when the same model may not hold for the future unobserved process. Unfortunately, there is not much we can do to control this source of error.

A brief outline of the thesis is as follows. In Chapter 2, we assess the maximum likelihood estimation procedure used to obtain estimates for v and $\boldsymbol{\beta}$ for specific models belonging to the general family of NHPP models with intensity $\lambda(t;\boldsymbol{\eta}) = v\,\lambda_0(t;\boldsymbol{\beta})$. In particular, we study conditions under which a finite, positive-valued maximum likelihood estimate for v is obtained, consider choices of parameterization to facilitate estimation, and consider the effects of total test time on these matters. In Chapter 3, we propose a new approach for testing the goodness of fit of NHPP models of the above general form. We also suggest two alternative models that include specific software reliability models of interest. These models are of use for testing the goodness of fit of their submodels. In Chapter 4, we propose a frequentist approach for providing interval predictors of $N2 = N(T1, T2]$, the number of events in the future time interval $(T1, T2]$, based on the observed data up to time $T1$. We also use this method to assess the effect of data accumulation on prediction of $N3 = N(T1, \infty]$ the number of remaining events to be eventually observed given data has been observed up to time $T1$. We also discuss how to obtain Bayesian prediction intervals and compare them with the frequentist-based prediction intervals for $N3$ in some examples. In Chapter 5, we discuss research areas to be investigated further.

Chapter 2

Maximum likelihood estimation

2.1 Introduction

To make inferences about the true underlying failure process based on the observed data in $(0, T1]$, we need to estimate the parameters in a chosen NHPP model based on the available data. Although other methods are available (such as the method of least squares), maximum likelihood estimation (m.l.e.) is often used for this purpose. As mentioned in Section 1.5, we want to assess how well (m.l.e.) performs as well as improve the estimation method. In particular, we are interested in a) determining conditions for a positive, finite $\hat{v} = \hat{\Lambda}(\infty)$, and; b) determining the choice of model parameterization for estimation. In this section, we motivate our interest in the above topics. We follow with a chapter outline.

2.1.1 Determining conditions for positive, finite $\hat{\Lambda}(\infty)$

Estimating the parameter $v = \Lambda(\infty)$ in *finite* mean value function models is of particular interest because it represents the total number of expected failures to be eventually discovered. For the *infinite* POWN model in Table 1.1, a *closed form* solution for v and β always exists in the ungrouped data case. In particular, in the time-truncated scenario, the m.l.e.'s for this model are $\hat{\beta} = \frac{n}{\sum_i^n log(T1/t_i)} > 0$ and $\hat{v} = (n/T1^{\hat{\beta}}) > 0$. In contrast, a closed-form solution for v and β does not exist for the other NHPP models given in Table 1.1. Furthermore, a finite, positive-valued unrestricted solution for v does *not* always exist for the finite NHPP models.

As an example, let us consider the DS2 data set. A plot of the cumulative number of failures versus cumulative time for this data set is shown in Figure 1.1. For illustrative purposes, we arbitrarily truncate the data at $T = 0.84$ (x 10^6) seconds to obtain a subset of the complete data set. If we model both the truncated ($n = 22$) and complete data sets ($n = 278$) using the EOSN model, we need to estimate the model parameters. Using maximum likelihood estimation and the log-linear parameterization of the model (to be discussed in Section 2.4), we obtain unrestricted $\hat{\beta} = -1.84$ with a 95% confidence interval $(-3.67, -0.01)$ for the truncated data set, and $\hat{\beta} = 0.02$ with 95% confidence interval $(0.01, 0.03)$ for the entire data set. The corresponding unrestricted estimates of v (and 95% confidence intervals) for the truncated and entire data sets are $\hat{v} = -6.00$ $(-17.03, -2.11)$ and $\hat{v} = 402.05$ $(326, 494)$, respectively. Neither set of confidence intervals include zero, which imply that there is no strong evidence at level $\alpha = 0.05$ that $\beta > 0$ (and $v > 0$) using the truncated data, or that $\beta < 0$ (and $v < 0$) using the entire

data. We note that when $\beta < 0$, $v < 0$ and v *no longer* represents $\Lambda(\infty)$. In fact, $\hat{\Lambda}(\infty) = \infty$ here. To see this, recall that for $v = \Lambda(\infty)$ we need $\lambda_0(t;\boldsymbol{\beta})$ to be a probability density function integrating to one. The negative estimate for v here is due to the fact that testing has not gone long enough to provide much information about $\Lambda(\infty)$.

To make some decisions in advance on how long one must test to obtain a reliable estimate of v, it is of interest to obtain conditions for which $\hat{v} = \hat{\Lambda}(\infty)$ is positive and finite for the models given in Table 1.1. Although work has been published on the conditions under which the m.l.e. does not yield finite positive-valued estimates for N for particular GOS models (EOS:[43], [56], [75]; WOS:[55]; and, POS:[12], [78]), comparatively less work has been done for obtaining finite positive-valued estimates for v in the corresponding NHPP models (EOSN:[78], [105]; WOSN:[55]; and, GAMN:[53], [105]). With the exception of [53], no one has obtained appropriate conditions under a grouped data situation.

2.1.2 Choice of parameterization for estimation

To motivate the need for considering different parameterizations of a model, we first provide a brief discussion on the nature of asymptotics applicable to NHPP (and GOS) families of models and then we discuss how this relates to the problem on hand. A thorough treatment of the nature of the asymptotics applicable in this case is given in [55], [72], and [97]. Our emphasis is on the results provided in [97]. As mentioned on [97, page 236], situations arise when it does not make practical sense to have the time variable or the sample size increased in order to

use asymptotic results. To accommodate situations when it makes sense to do so, van Pul ([97]) developed a new asymptotic approach that involves looking at the case when one of the model parameters, rather than time or sample size, is considered large. The asymptotics are in a sense artificial but they do allow the possibility of using standard asymptotic properties (e.g. consistency, asymptotic normality, efficiency) of maximum likelihood estimation to provide standard error approximations for the parameters. For a general class of parametric counting process models, he derived conditions on the intensity function that are sufficient for these asymptotic properties to hold.

To begin with, counting processes are assumed to have an intensity function of the form $\lambda(t; v, \boldsymbol{\beta})$, for scalar parameter v and vector of parameters $\boldsymbol{\beta}$. Interest is in estimating the true parameters v_O and $\boldsymbol{\beta}_0$ as $v_O \to \infty$. v is assumed to represent the "scale or size of the problem". Next, to consider parameter estimation when v_O is large, a reparameterized series of counting processes $N_\kappa(t)$, for $\kappa = 1, 2, \ldots$ is introduced with associated intensity functions given by $\lambda_\kappa(t; \gamma, \boldsymbol{\beta}) = \lambda(t; \kappa\gamma, \boldsymbol{\beta})$. With $v = v_\kappa = \kappa\gamma$, we now consider the estimation of γ and $\boldsymbol{\beta}$ as $\kappa \to \infty$. If the associated intensities are assumed further to have the form $\lambda_\kappa(t; \gamma, \boldsymbol{\beta}) = \kappa\,\lambda_O(t; \gamma, \boldsymbol{\beta})$, for an arbitrary non-negative function λ_O, it is then shown in [97] that these conditions on the intensity are sufficient for the standard asymptotic properties of m.l.e. to hold. It is easily seen that all of the models in Table 1.1 satisfy these conditions, with $v = \kappa\gamma$ and $\lambda_O(t; \gamma, \boldsymbol{\beta}) = \kappa\gamma\,\lambda_0(t; \boldsymbol{\beta})$.

With standard asymptotic properties of m.l.e. at our disposal when v is considered large, one can then obtain likelihood-ratio (LR), Wald or score-based ap-

proximate asymptotic confidence intervals for the parameters. Nevertheless, it is always desirable to check the adequacy of these approximations for small to medium sample sizes. Plotting constant likelihood ratio contours may be of use here. Comparisons may also be made with appropriate plots of Wald and score-based contours. Non-elliptical (in the extreme case, banana-shaped) likelihood contours reflect inadequacy of quadratic approximations. This implies that the shape of the contours may alert one to problems with the performance of the likelihood maximization algorithm to be used and the accuracy of asymptotic likelihood inference. Since the most used optimization algorithms are adaptations of the classical Newton method, it makes sense that they would work best on approximately quadratic functions. Similarly, because the normal approximation is achieved by ignoring the cubic and higher order terms in a Taylor's series expansion of the log-likelihood at the maximum likelihood estimate, its accuracy relies on the size of the cubic or higher terms relative to the quadratic term. When the contours of constant likelihood are roughly elliptical and centered at the m.l.e, the normal approximation is sufficiently accurate. As well, all of the information concerning the parameters is summarized accurately in the point estimate and its measure of precision, the observed or expected Fisher information matrix (to be discussed later). On the other hand, if the likelihood contour plots are not roughly elliptical, approximate inferences based on the asymptotic normality of the estimates, including confidence intervals, will be misleading.

Although likelihood functions are invariant under one-to-one parameter transformations their normal approximations are not invariant. For this reason, it may be

possible to obtain greater accuracy of the normal approximation by reparametrizing the model before taking approximations [58, page 43]. At the same time, there may be a noticeable improvement in the performance of the likelihood maximization algorithm used [95, page 95]. We can then recover approximate confidence intervals for the old parameters by making use of the invariance property in transforming those obtained for the new parameters. It then follows that it is beneficial to transform the parameters in the model if the LR- or score-based contours are not elliptical under the current parametrization.

As an example, we obtained contour plots based on constant relative LR and Wald test values (variance estimates based on the expected information matrix) under the EOSN model for data sets DS1,3. The contour plots are given in Figure 2.1. The m.l.e. is denoted by an asterisk. Similar plots based on the score statistic were also obtained but are not provided, as they are similar to those based on the Wald statistic. The "banana-shaped" curves in the LR plot for DS3 indicate clearly that approximate inferences based on the asymptotic normality of the parameters will be misleading for that data set; hence, it is of interest to reparameterize the original EOSN model in order to improve the accuracy of a normal approximation. In addition, we calculated the LR- and Wald-based confidence intervals for the parameters v and β and $u3 = E(N3) = E(N(T1, \infty))$ for DS1 and DS3. The results are given in Table 2.1. Although they are equal for β in both data sets, the LR- and Wald-based confidence intervals for v (and the resulting $u3$) differ in DS3. This difference in the two confidence intervals reflects further the difference in the shapes of the LR and Wald contours for DS3. If the shape of the LR

Table 2.1: LR and Wald confidence intervals for β, v, and $u3$ under the EOSN model and DS1,3.

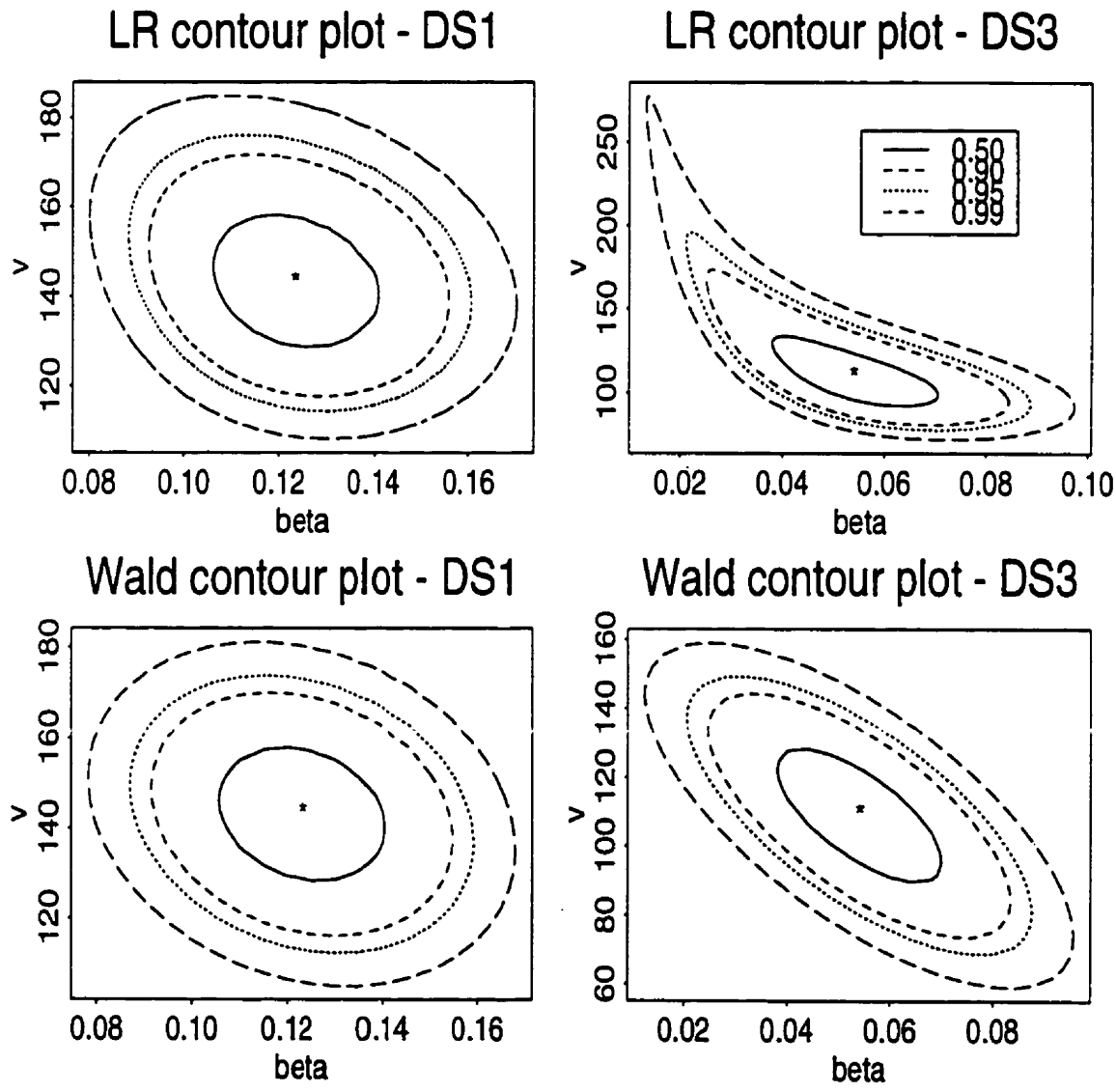
parameter \ *CI*	*LR*	*Wald*
DS1: $\hat{\beta} = 0.12$, $\hat{v} = 142$, $T1 = 25.34$		
β	$(0.09, 0.16)$	$(0.09, 0.16)$
v	$(115, 178)$	$(115, 172)$
$u3$	$(3, 12)$	$(3, 12)$
DS3: $\hat{\beta} = 0.05$, $\hat{v} = 109$, $T1 = 28.70$		
β	$(0.02, 0.09)$	$(0.02, 0.09)$
v	$(75, 198)$	$(75, 150)$
$u3$	$(15, 42)$	$(11, 42)$

contours were more elliptical (as is the case for DS1), then the Wald and LR-based confidence intervals would be the same. It is also of interest to examine LR plots of other models given in Table 1.1 for similar plots. If so, we would be interested in proposing and assessing different parameterizations of these models to be used for estimation purposes.

2.1.3 Outline of chapter

In Sections 2.2 and 2.3, we discuss an approach used to obtain m.l.e.'s for the general class of NHPP software models under the ungrouped and grouped scenarios, respectively. For the ungrouped case, we have provided the appropriate m.l.e. equations for the models given in Table 1.1. For the EOSN, GAMN and LOGN models, we have determined necessary and sufficient conditions for which the m.l.e. solution for v is positive and finite-valued. In addition, we have provided results on simulation studies conducted to obtain probability statements on how often these

Figure 2.1: Likelihood ratio and Wald contour plots for EOSN and DS1,3.



conditions do not hold. Although not essential, we have also obtained bounds on the m.l.e. solution for β (and hence, v). They are an asset when one needs an initial first estimate at a solution in the numerical routine used for optimization. We have not been able to obtain complete conditions for the WOSN and POSN models, but we give some partial results for these two models. For the grouped case, we have determined necessary and sufficient conditions for which the m.l.e. solution for v is positive and finite-valued for the EOSN and POWN models. In Section 2.4, we suggest two reparameterizations of the models in Table 1.1 that are useful for estimation purposes. Our conclusions are given in Section 2.5.

2.2 Conditions for positive, finite $\hat{\Lambda}(\infty)$ — ungrouped case

2.2.1 General case

The log-likelihood function for a single realization (ungrouped) of a time-truncated NHPP with a general intensity function $\lambda(t;\eta)$ was derived in Section 1.3.2. With an intensity function of the form $\lambda(t;\eta) = v\,\lambda_0(t;\beta)$ and corresponding mean value function $\Lambda(T1;\eta) = v\,\Lambda_0(T1;\beta)$, for scalar $v > 0$ and vector $\beta > 0$, (1.2) becomes

$$l(\eta) \;=\; n\log(v) + \sum_{i=1}^{n}\log\{\lambda_0(t_i;\beta)\} - v\,\Lambda_0(T1;\beta). \qquad (2.1)$$

Differentiating the above with respect to parameters $(v, \boldsymbol{\beta})$ yields the following set of $J+1$ (score) equations to be solved for the maximum likelihood estimators:

$$\frac{\partial l(\boldsymbol{\eta})}{\partial v} = \frac{n}{v} - \Lambda_0(T1;\boldsymbol{\beta})$$

$$\frac{\partial l(\boldsymbol{\eta})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial}{\partial \beta_j} \log\{\lambda_0(t_i;\boldsymbol{\beta})\} - v \frac{\partial}{\partial \beta_j} \Lambda_0(T1;\boldsymbol{\beta}), \quad j = 1, \ldots, J.$$

Solving for v in the first equation yields

$$\hat{v} = \frac{n}{\Lambda_0(T1;\hat{\boldsymbol{\beta}})}. \tag{2.2}$$

This expression for $\hat{v}$ can then be substituted into the other J score functions to obtain

$$\sum_{i=1}^{n} \left\{ \frac{\dot{\lambda}_0(t_i;\boldsymbol{\beta})}{\lambda_0(t_i;\boldsymbol{\beta})} - \frac{\dot{\Lambda}_0(T1;\boldsymbol{\beta})}{\Lambda_0(T1;\boldsymbol{\beta})} \right\} = 0, \tag{2.3}$$

for $\dot{f} = \frac{\partial}{\partial \beta_j} f$, $j = 1, \ldots, J$. With v eliminated, we can solve the above equation(s) for $\hat{\boldsymbol{\beta}}$ and then substitute into the expression for $\hat{v}$. Since a closed form solution for $\hat{\boldsymbol{\beta}}$ rarely exists, numerical procedures are used to obtain the solution to the above equations for $\hat{\boldsymbol{\beta}}$. Of the models we have considered (c.f. Table 1.1) a closed form solution is available only for the POWN (power law) model. If the expected total number of events is large and appropriate conditions on the model hold, we know that $\hat{\boldsymbol{\eta}}$ is asymptotically normally distributed with mean $\boldsymbol{\eta}$ and covariance matrix

$I(\boldsymbol{\eta})^{-1}$. The Fisher (or expected) information matrix $I(\boldsymbol{\eta})$ has entries

$$I_{ij}(\boldsymbol{\eta}) = E\left(\frac{-\partial^2 l(\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j}\right), \quad i,j = 1,...,J+1.$$

We can estimate the above with $I(\hat{\boldsymbol{\eta}})$, or with the observed information matrix I_O, which has entries

$$I_{O,ij} = \left(\frac{-\partial^2 l(\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j}\bigg|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}}\right), \quad i,j = 1,...,J+1.$$

In some cases, equation (2.3) will not have an "admissible" solution $\hat{\boldsymbol{\beta}} > 0$, but there may be an "unrestricted" solution with some parameter estimates negative. This corresponds to $l(\boldsymbol{\eta})$ being maximized at a boundary point for the restricted parameter space with $\boldsymbol{\beta} > 0$ and $v > 0$. For the models we consider, it is possible to extend the parameter space for $\boldsymbol{\beta}$ and v and discuss the solution of equation (2.3). We do this below.

2.2.2 Specific cases

EOSN

Introduction

A necessary and sufficient condition for obtaining an unique positive, finite-valued m.l.e. solution under the ungrouped EOSN (Goel-Okumoto) model is derived in a slightly more difficult manner in [78]. Rather than solving for $\hat{\boldsymbol{\beta}}$ and then substituting in for $\hat{v}$, [78] numerically solves an equation for $\hat{v}$ and then substitutes in its m.l.e. to obtain that for $\hat{\boldsymbol{\beta}}$. It may seem like a slight difference, but it turns

out that the function of $\hat{v}$ to be solved is more complicated than our function of β; hence, finding the necessary and sufficient conditions is more tedious in that situation. In addition, our approach may be applied to all NHPP models, whereas their approach may not. Although upper and lower bounds on the solution are also given in that paper, we have determined a tighter interval in which the solution lies. We also conducted a small simulation study to give some probability statements on how often the necessary and sufficient condition is not obtained. Some comparisons with the results for the EOS model are also given.

Necessary and sufficient condition for a positive $\hat{\beta}$

For the EOSN model, $\lambda(t;\eta) = v\beta \exp(-\beta t)$ with corresponding $\Lambda(T1;\eta) = v[1 - \exp(-\beta T1)]$. In this case, (2.2) is $\hat{v} = \frac{n}{1-\exp(-\hat{\beta} T1)}$. Clearly, $\hat{v}$ is finite and positive if and only if $\hat{\beta} > 0$. After substituting the required derivatives into (2.3), and rearranging some of the terms, we obtain

$$\frac{1}{\beta T1} - \frac{1}{\exp(\beta T1) - 1} = \frac{1}{n}\frac{\sum_{i=1}^{n} t_i}{T1}. \tag{2.4}$$

If we let $g(\beta)$ denote the left-hand side of (2.4), a Taylor expansion of the function $\exp(\beta T1)$ shows that $\lim_{\beta\to 0} g(\beta) = 1/2$ and $\lim_{\beta\to\infty} g(\beta) = 0$. Furthermore, since $\frac{\partial g(\beta)}{\partial \beta} < 0$, for all $\beta > 0$, $g(\beta)$ is monotonically decreasing. Since the right-hand side of (2.4) is always greater than zero, a sufficient condition for the existence of a positive root for (2.4) is

$$\frac{1}{n}\frac{\sum_{i=1}^{n} t_i}{T1} < 0.5. \tag{2.5}$$

In addition, since $g(\beta)$ is monotonically decreasing, condition (2.5) is also necessary for (2.4) to have an unique finite positive root.

Bounds on solution

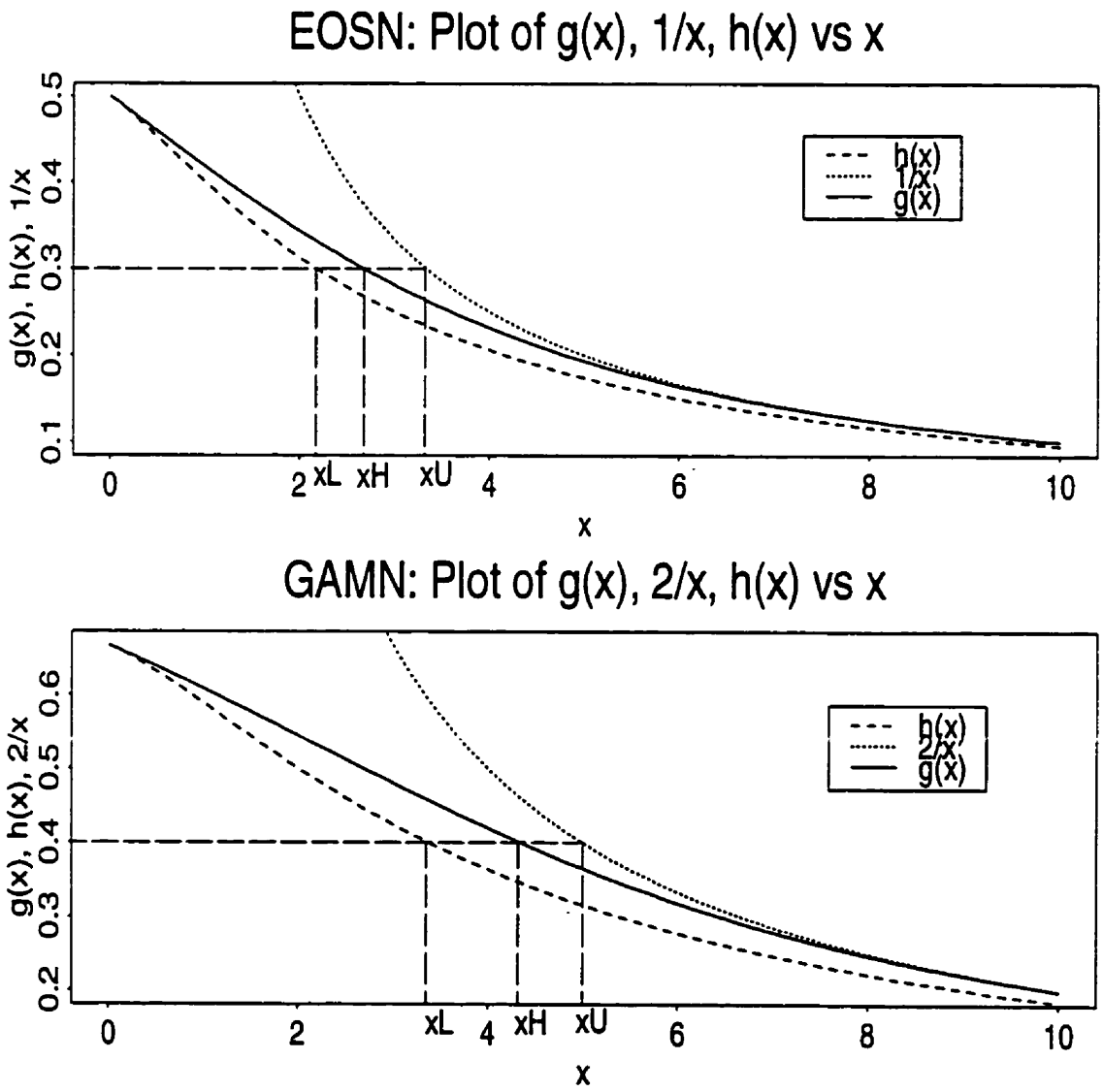
An ad hoc approach was taken to obtain bounds on the solution. Let us first look for an upper bound on the m.l.e. $\hat{\beta}$, the value of β such that (2.4) holds. Figure 2.2 illustrates that $\frac{1}{\beta T1} > g(\beta T1)$, for all $x = \beta T1$, and that both functions are monotonically decreasing. For a particular value of $T1_{AVE} = \frac{1}{n}\sum_{i=1}^{n} t_i/T1$, it follows that $\frac{1}{\beta T1} = T1_{AVE}$ will occur at a value of $\beta > \hat{\beta}$. Hence, this value of β, say β_U, will be an upper bound. Since $\hat{v} = \frac{n}{1-\exp(-\hat{\beta}T1)}$ is a monotonically decreasing function of $\hat{\beta}$, a lower bound on $\hat{v}$ will be $v_L = \frac{n}{1-\exp(-\beta_U T1)}$, where $\beta_U = \frac{n}{\sum_{i=1}^{n} t_i}$.

To obtain a lower bound on $\hat{\beta}$, we found a monotonically decreasing function $h(x)$ that is always smaller than $g(x)$. The function was obtained by first taking a Taylor series expansion of $\exp(x)$ in the denominator of the second term of $g(x)$ and then ignoring the fourth order and higher terms. It can be expressed as

$$h(x) = \frac{\frac{1}{2} + \frac{x}{6}}{1 + \frac{x}{2} + \frac{x^2}{6}}.$$

Figure 2.2 shows that $h(\beta T1) < g(\beta T1)$, for all $x = \beta T1$ and that both functions are monotonically decreasing in $\beta T1$. We also see that $h(\beta T) = T_{AVE}$ will occur at a value of $\beta < \hat{\beta}$. Hence, the lower bound on $\hat{\beta}$ is the value of $\beta_L > 0$ that solves the quadratic equation $h(\beta_L T) = T_{AVE}$. Substituting this lower bound for $\hat{\beta}$ into the equation for $\hat{v}$ yields the upper bound on $\hat{v}$, $v_U = \frac{n}{1-\exp(-\beta_L T1)}$. Our bounds for $\hat{v}$, $(\frac{n}{1-\exp(-\beta_U T1)}, \frac{n}{1-\exp(-\beta_L T1)})$, are tighter than those found in [78], which are $(n, \frac{(T1_{AVE}-1)^2}{(1-2T1_{AVE})}]$.

Figure 2.2: Maximum likelihood estimation plots for EOSN and GAMN.



Simulation study

A small simulation study was performed to determine how often a positive, finite-valued m. l. e. solution is not available for the EOSN and EOS models. We generated samples from the EOSN model with $\beta = 1$, $v = 15, 25, 50, 100, 200$, and for two different values of $T1$. The $T1$ were chosen such that roughly 50% and 90% of the expected number of errors to be eventually detected are in fact discovered by $T1$. In other words, $T1$ is determined such that $E\{N(0, T1)\} = v(1 - \exp(-\beta T1)) = av$, for $a = 0.5, 0.9$ and given β. This implies that $T1 = \frac{-\log(1-a)}{\beta}$. In fact, $T1 = 0.6931, 2.3026$ for $\beta = 1$ and $a = 0.5, 0.9$, respectively.

In particular, 2000 samples were obtained for each combination of v and $T1$ in the following manner:

1. Simulate N, a Poisson random variable with mean $v = E\{N(0, \infty)\}$.

2. Simulate N standard exponential random variables. Keep those simulated values that are $\leq T1$. Sort these remaining $N1$ values in ascending order. This now constitutes a sample of $N1$ ordered failures times $t_1^*, \cdots, t_{N1}^*$ for given $T1$ and v.

For each sample generated, we calculated (2.5), the condition for which an unique, finite, positive-valued m. l. e. solution is available for the EOSN model. The percentages of the ten sets of simulated data which did not satisfy condition (2.5) for the EOSN model are given in Table 2.2, which is a subset of the results given later in Table 2.3. We clearly see that as v increases, these percentages decrease. In addition, we see that these percentages are substantially lower when we have already observed a large percentage of the expected errors.

Table 2.2: Percentages of the ten sets of simulated data that did not satisfy condition (2.5).

$a \setminus v$	15	25	50	100	200
0.5	28.90	26.15	16.70	7.90	2.15
0.9	1.20	0.20	0.00	0.00	0.00

In addition, for each of the above simulated samples, we calculated the condition for which an unique, finite, positive-valued m.l.e. solution is available for the EOS model. This condition, derived by [78], is given by

$$\frac{n * (T1 - t_n) + \sum_{i=1}^{n} (i-1)\, x_i}{T1} > \frac{n-1}{2}, \tag{2.6}$$

where x_i is the ith interfailure time; that is, x_i is the time between the $(i-1)$th and ith failure. The results for the entire study are given in Table 2.3. As expected, the results improve as a and v increases. The percentage of the simulated data which did not satisfy condition (2.6) for the EOS model is lower than that for the EOSN model when $a = 0.5$ for all values of v. This indicates that there is a larger chance of obtaining an infinite estimate for $\hat{v}$ using the EOSN model than obtaining an infinite estimate for $\hat{N}$ using its equivalent EOS model when $a = 0.5$.

We can show further that since $\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} (n - i + 1) x_i$, we can rewrite equation (2.6) as

$$\frac{1}{n} \frac{\sum_{i=1}^{n} t_i}{T1} < 0.5 + \frac{1}{2n}. \tag{2.7}$$

Comparing the above equation with equation (2.5), we see that if (2.5) holds, (2.6)

Table 2.3: Classification of the $B = 2000$ simulations according to the finiteness of $\hat{v}$ in the EOSN model and $\hat{N}$ in the EOS model.

Model attributes			$a = 0.5$			$a = 0.9$		
				EOS			EOS	
			finite	*infinite*	*totals*	*finite*	*infinite*	*totals*
$v = 15$		*finite*	1422	0	1422	1976	0	1976
	EOSN	*infinite*	351	227	578	19	5	24
		totals	1773	227	2000	1995	5	2000
$v = 25$		*finite*	1477	0	1477	1996	0	1996
	EOSN	*infinite*	287	236	523	4	0	4
		totals	1764	236	2000	2000	0	2000
$v = 50$		*finite*	1666	0	1666	2000	0	2000
	EOSN	*infinite*	143	191	334	0	0	0
		totals	1809	191	2000	2000	0	2000
$v = 100$		*finite*	1842	0	1842	2000	0	0
	EOSN	*infinite*	52	106	158	0	0	0
		totals	1894	106	2000	2000	0	2000
$v = 200$		*finite*	1957	0	1957	2000	0	0
	EOSN	*infinite*	9	34	43	0	0	0
		totals	1966	34	2000	2000	0	2000

also holds, but not vice versa. We note further that a comparison of these equations indicates that as the observed sample size n increases, the second term in the above equation goes to zero and the two conditions become equal. The simulation results in Table 2.3 confirm further these mathematical results.

GAMN

Introduction

A necessary and sufficient condition for obtaining an unique positive finite-valued m. l. e. solution under the GAMN (Yamada et. al. S-shaped) model as well as bounds on the solution were obtained in a similar manner to that done for the EOSN model. Although [53] derived the same conditions, bounds on the solution were not given.

We also conducted a small simulation study to give some probability statements on how often the necessary and sufficient condition is not obtained.

Necessary and sufficient condition for a positive $\hat{\beta}$

For the GAMN model $\lambda(t;\eta) = v\beta^2 t \exp(-\beta t)$ with corresponding $\Lambda(T1;\eta) = v[1-(1+\beta T1)\exp(-\beta T1)]$. In this case, (2.2) is $\hat{v} = \frac{n}{1-(1+\hat{\beta}T1)\exp(-\hat{\beta}T1)}$. Clearly, $\hat{v}$ is finite and positive if $\hat{\beta} > 0$. When $\hat{\beta}T1 = -1$, $\hat{v} = n$ is also finite and positive; however, this event occurs with probability zero.

After substituting the required derivatives into the equation for β, and rearranging some of the terms, we obtain:

$$\frac{2}{\beta T1} - \frac{\beta T1}{\exp(\beta T1) - (1+\beta T1)} = \frac{1}{n}\frac{\sum_{i=1}^{n} t_i}{T1} \tag{2.8}$$

If we let $g(\beta)$ denote the left-hand side of (2.8), a Taylor expansion of the function $\exp(\beta T1)$ shows that $\lim_{\beta\to 0} g(\beta) = 2/3$ and $\lim_{\beta\to\infty} g(\beta) = 0$. Furthermore, since $\frac{\partial g(\beta)}{\partial \beta} < 0$, for all $\beta > 0$, $g(\beta)$ is monotonically decreasing. Since the right-hand side of (2.8) is always greater than zero, a sufficient condition for the existence of a positive root for (2.8) is

$$\frac{1}{n}\frac{\sum_{i=1}^{n} t_i}{T1} < \frac{2}{3}. \tag{2.9}$$

In addition, since $g(\beta)$ is monotonically decreasing, condition (2.9) is also necessary for (2.8) to have an unique finite positive root.

Bounds on solution

A similar approach to that taken for the EOSN model was taken to obtain bounds

on this solution. Figure 2.2 shows that $\frac{2}{\beta T1} > g(\beta T1)$, for all $x = \beta T1$, and that both functions are monotonically decreasing in $\beta T1$. For a particular value of $T1_{AVE} = \frac{1}{n}\sum_{i=1}^{n} t_i/T1$, it follows that $\frac{2}{\beta T1} = T1_{AVE}$ will occur at a value of $\beta > \hat{\beta}$. Hence, this value of β, say β_U, will be an upper bound. Since $\hat{v} = \frac{n}{1-(1+\hat{\beta}T1)\exp(-\hat{\beta}T1)}$ is a monotonically decreasing function of $\hat{\beta}$, a lower bound on $\hat{v}$ will be $v_L = \frac{n}{1-(1+\beta_U T1)\exp(-\beta_U T1)}$, where $\beta_U = \frac{2n}{\sum_{i=1}^{n} t_i}$.

Similarly, to obtain a lower bound on $\hat{\beta}$, we found a monotonically decreasing function $h(x)$ that is always smaller than $g(x)$. This function was also obtained by first taking a Taylor series expansion of $\exp(x)$ in the denominator of the second term of $g(x)$, this time ignoring the fifth order and higher terms. It can be expressed as

$$h(x) = \frac{\frac{1}{3} + \frac{x}{12}}{\frac{1}{2} + \frac{x}{6} + \frac{x^2}{24}}.$$

Figure 2.2 shows that $h(\beta T1) < g(\beta T1)$, for all $x = \beta T1$ and that both functions are monotonically decreasing in $\beta T1$. We also see that $h(\beta T1) = T1_{AVE}$ will occur at a value of $\beta < \hat{\beta}$. Hence, the lower bound on $\hat{\beta}$ is the value of $\beta_L > 0$ that solves the quadratic equation $h(\beta_L T1) = T1_{AVE}$. Substituting this lower bound for $\hat{\beta}$ into the equation for $\hat{v}$ yields the upper bound on $\hat{v}$, $v_U = \frac{n}{1-(1+\beta_L T1)\exp(-\beta_L T1)}$.

Simulation study

A small simulation study was also performed to determine how often a finite, positive-valued m. l. e. solution is not available for the GAMN model. As for the EOSN model, we generated GAMN samples with $\beta = 1$, $v = 15, 25, 50, 100, 200$, and for two different values of $T1$. In this particular case, $T1$ was determined such

that $E\{N(0,T1)\} = v(1-(1-\beta T1)\exp(-\beta T1)) = av$, for $a = 0.5, 0.9$. Although $T1$ and a are one-to-one functions, an analytical expression for $T1$ as a function of a is unavailable. A simple numerical method, such as the bisection or Newton-Raphson methods, may be used to obtain $T1$. We used the bisection method to obtain $T1 = 0.6931, 2.3026$ for $a = 0.5, 0.9$, respectively. In particular, 2000 samples were obtained for each combination of v and $T1$ in the following manner:

1. Simulate N, a Poisson random variable with mean $v = E\{N(0,\infty)\}$.

2. Simulate N uniform $(0,1)$ random variables, u_i's. A bisection method was used to obtain the t_i for each u_i, where $u_i = 1-(1-\beta t_i)\exp(-\beta t_i)$. Then, keep those simulated t_i's that are $\leq T1$. Sort these remaining $N1$ values in ascending order. This now constitutes a sample of $N1$ ordered failures times $t_1^*, \cdots, t_{N1}$ for given $T1$ and v.

The percentages of the ten set of simulated data which did not satisfy condition (2.9) are given in Table 2.4. The results are similar to those obtained for the EOSN model. The only difference is that these percentages are smaller than for the EOSN model. That is, given the same v and a, there is less chance of obtaining negative-valued or infinite m.l.e.'s under the GAMN model than there is for the EOSN model.

LOGN

Introduction

As far as we know, results on the conditions for which a m.l.e. solution is unique, positive and finite-valued for the LOGN (Musa-Okumoto) model have not been

Table 2.4: Percentages of the ten sets of simulated data that did not satisfy condition (2.9).

$a \setminus v$	15	25	50	100	200
0.5	14.10	8.30	2.70	0.15	0.00
0.9	0.10	0.00	0.00	0.00	0.00

published. We have derived such a condition. We also conducted a small simulation study to give some probability statements on how often the necessary and sufficient condition is not obtained.

Necessary and sufficient condition for a positive $\hat{\beta}$

For the LOGN model, $\lambda(t;\eta) = v\beta/(1+\beta t)$ and $\Lambda(T1;\eta) = v\,log(1+\beta T1)$. In this case, (2.2) is $\hat{v} = \frac{n}{log(1+\hat{\beta} T1)}$. After substituting the required derivatives into (2.3), and rearranging some of the terms, we obtain

$$\frac{\beta T1}{(1+\beta T1)\,log(1+\beta T1)} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\beta t_i + 1}. \tag{2.10}$$

Let $l(\beta)$ and $r(\beta)$ denote the left-hand and right-hand sides of the above equation. Simple calculations show that both $l(\beta)$ and $r(\beta)$ are monotonically decreasing functions because $\frac{\partial r(\beta)}{\partial \beta} < 0$ and $\frac{\partial l(\beta)}{\partial \beta} < 0$. Furthermore, their derivatives are decreasing at an increasing rate since $\frac{\partial^2 r(\beta)}{\partial \beta^2} > 0$ and $\frac{\partial^2 l(\beta)}{\partial \beta^2} > 0$. This implies that both functions will always be concave up.

At this point, it is useful to obtain some information on the behaviour of these functions for large and small values of β. First, let's consider the case when β is small. Simple calculations show that $\lim_{\beta\to 0} l(\beta) = 1$ and $\lim_{\beta\to 0} r(\beta) = 1$. This

implies that both functions $l(\beta)$ and $r(\beta)$ "start off together"; hence, $\beta = 0$ is always a solution to these equations. In addition, $\lim_{\beta \to 0} \frac{\partial l(\beta)}{\partial \beta} = -0.5\,T1$ and $\lim_{\beta \to 0} \frac{\partial r(\beta)}{\partial \beta} = -\frac{1}{n}\sum_{i=1}^{n} t_i$. Since $|\frac{1}{n}\sum_{i=1}^{n} t_i| < T1$, we know further that $0 < |\lim_{\beta \to 0} \frac{\partial r(\beta)}{\partial \beta}| < T1$. Next, let's consider the case when β is large. Simple calculations show that $\lim_{\beta \to \infty} l(\beta) = 0$ and $\lim_{\beta \to \infty} r(\beta) = 0$. In addition, when β is large,

$$l(\beta) \approx \frac{1}{log(1+\beta\,T1)} \quad \text{and} \quad r(\beta) \approx \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\beta t_i} = \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{t_i}}{\beta}.$$

Since $log(1+\beta\,T1) < \beta$, it then follows that

$$\frac{1}{log(1+\beta\,T1)} > \frac{1}{\beta} \Longrightarrow \frac{1}{log(1+\beta\,T1)} > \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{t_i}}{\beta} \Longrightarrow l(\beta) > r(\beta),$$

for large β.

We are now ready to put the above information together. As mentioned earlier, we know that both functions $l(\beta)$ and $r(\beta)$ "start off together", since $\lim_{\beta \to 0} l(\beta) = \lim_{\beta \to 0} r(\beta) = 1$. We know further that these functions will initially proceed at different slopes because $|\lim_{\beta \to 0} \frac{\partial l(\beta)}{\partial \beta}| = 0.5\,T1$ and $0 < |\lim_{\beta \to 0} \frac{\partial r(\beta)}{\partial \beta}| < T1$. In particular, one of the following two situations may arise:

$$|\lim_{\beta \to 0} \frac{\partial l(\beta)}{\partial \beta}| = 0.5\,T1 \quad \text{and} \quad 0 < |\lim_{\beta \to 0} \frac{\partial r(\beta)}{\partial \beta}| < 0.5\,T1$$

or

$$|\lim_{\beta \to 0} \frac{\partial l(\beta)}{\partial \beta}| = 0.5\,T1 \quad \text{and} \quad 0.5\,T1 < |\lim_{\beta \to 0} \frac{\partial r(\beta)}{\partial \beta}| < 1.0\,T1.$$

In the first case, the slope of $r(\beta)$ will initially be greater than that of $l(\beta)$, whereas in the second case, $r(\beta)$ will initially be less than that of $l(\beta)$. This implies that $r(\beta) > l(\beta)$ for small β in the first case, whereas $r(\beta) < l(\beta)$ for small β in the latter case. Regardless of the differences in the initial slopes of these functions, we know that for large β, $l(\beta) > r(\beta)$. It then follows that $r(\beta) > l(\beta)$ for small β and $r(\beta) < l(\beta)$ for large β in the first scenario, while $r(\beta) < l(\beta)$ for both small β and large β in the second scenario. Since $l(\beta)$ and $r(\beta)$ are always concave up, they will cross one another at most once. As shown above, these functions only cross under the first scenario listed above. Since $\lim_{\beta \to 0} \frac{\partial r(\beta)}{\partial \beta} = -\frac{1}{n}\sum_{i=1}^{n} t_i$, it follows that a necessary and sufficient condition for an unique, positive, finite root is

$$\frac{1}{n}\frac{\sum_{i=1}^{n} t_i}{T1} < 0.5. \tag{2.11}$$

Simulation study

A small simulation study was performed to determine how often condition (2.11) does not hold. We generated samples from the LOGN model with $\beta = 1$, $v = 15, 25, 50, 100, 200$, and for two different values of $T1$. For this infinite mean value model, we are not able to simulate samples whereby $T1$ is chosen such that a certain percentage of the total expected errors are observed because $a \infty = \infty$. An alternative would be to choose $T1$ such that roughly a certain percentage of the total expected errors under the EOSN model is obtained. The reason for using the EOSN model in this manner is because, for small β, these models are approximately equal. We can easily see this if we compare the intensities for these two models, as

Table 2.5: Percentages of the ten sets of simulated data that did not satisfy condition (2.11)

$a \setminus v$	15	25	50	100	200
0.5	35.15	29.15	21.95	13.90	6.15
0.9	7.80	3.50	0.50	0.00	0.00

is done below.

$$\begin{aligned}
\lambda(t_i; v, \beta)_{EOSN} &= v\beta \exp(-\beta t_i) \\
&= \frac{v\beta}{1 + \beta t_i + (\beta t_i)^2/2 + (\beta t_i)^3/3! + \ldots} \\
&\approx \frac{v\beta}{1 + \beta t_i}, \quad \text{for } \beta \text{ small} \\
&= \lambda(t_i; v, \beta)_{LOGN}
\end{aligned}$$

In particular, 2000 samples were obtained for each combination of v and $T1$ in the following manner:

1. Simulate n, a Poisson random variable with mean $v\,log(1 + \beta T1)$.

2. Simulate n uniform $(0, 1)$ random variables, u_i's, and sort them. The corresponding t_i's are obtained based on the formula

$$u_i = \frac{log(1 + \beta t_i)}{log(1 + \beta T1)}.$$

The percentages of the ten sets of simulated data which did not satisfy (2.11) are given in Table 2.5. As expected, these percentages are comparable to those given for the EOSN model.

WOSN

Introduction

It was stated in [55, page 1488] that there is a possibility that the likelihood under this model "may be maximized at $v = \infty$", but no proof was given to substantiate this claim. At this time, we have not been able to obtain conditions for obtaining a positive, finite-valued m. l. e. solution for this model; however, we have worked on bounds to the solution which we anticipate will lead to the required conditions.

Maximum likehood estimation equations

For the WOS model, $\lambda(t;\eta) = v\,\beta_2\,\beta_1\,t^{\beta_2-1}\exp(-\beta_1\,t^{\beta_2})$ and $\Lambda(T1;\eta) = v[1 - \exp(-\beta_1\,T1^{\beta_2})]$. In this case, (2.2) is $\hat{v} = \frac{n}{1-\exp(-\hat{\beta}_1\,T1^{\hat{\beta}_2})}$. After substituting the required derivatives into the equations for β, and rearranging some of the terms, we obtain the following equations:

$$\frac{1}{\beta_1\,T1^{\beta_2}} - \frac{1}{\exp(\beta_1 T1^{\beta_2}) - 1} = \frac{1}{n}\frac{\sum_{i=1}^{n} t_i^{\beta_2}}{T1^{\beta_2}} \tag{2.12}$$

$$\frac{1}{\beta_1\,T1^{\beta_2}\,\beta_2\,logT1} - \frac{1}{\exp(\beta_1\,T1^{\beta_2}) - 1} = \frac{1}{n}\frac{1}{T1^{\beta_2}}\sum_{i=1}^{n}(t_i^{\beta_2} - \beta_1^{-1})\frac{log\,t_i}{logT1} \tag{2.13}$$

Since both of these equations share a term we can subtract one from the other to obtain:

$$\frac{1}{\beta_1\,T1^{\beta_2}} - \frac{1}{\beta_1\,T1^{\beta_2}(\beta_2\,logT1)} = \frac{1}{n}\frac{1}{T1^{\beta_2}}\sum_{i=1}^{n}\left\{t_i^{\beta_2}\left(1 - \frac{log\,t_i}{logT1}\right) + \frac{log\,t_i}{\beta_1\,logT1}\right\}.$$

After multiplying both sides by $\beta_1\,T1^{\beta_2}$ and rearranging some of the terms, we

obtain an expression for β_1,

$$\beta_1 = \frac{\left[logT1 - \frac{1}{\beta_2} - \frac{1}{n}\sum_{i=1}^{n} log\, t_i\right]}{\frac{1}{n}\sum_{i=1}^{n} t_i^{\beta_2}(logT1 - log\, t_i)} \tag{2.14}$$

Equations (2.12) and (2.14) can then be used to solve for the m. l. e.'s of β_1 and β_2.

Lower bounds on solution

To obtain bounds on the solution, it is useful to scale the failure times so that $T1 \equiv 1$. We can then write equation (2.12) as

$$\frac{1}{\beta_1} - \frac{1}{\exp(\beta_1) - 1} = \frac{1}{n}\sum_{i=1}^{n} t_i^{\beta_2}, \tag{2.15}$$

and equation (2.14) as

$$\beta_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}[-log(t_i)] - \frac{1}{\beta_2}}{\frac{1}{n}\sum_{i=1}^{n} t_i^{\beta_2}[-log(t_i)]}. \tag{2.16}$$

The argument for obtaining preliminary bounds on the solution is as follows. First, note that the left-hand side of (2.15) is the same as that in (2.4) with $T1 \equiv 1$. The same reasoning used in obtaining a sufficient condition for the existence of a positive root for (2.4) can then be used to obtain a boundary condition for a positive root for (2.15). Since the left-hand side of (2.15) is largest at 0.5, we know that

$$0.5 > \frac{1}{n}\sum_{i=1}^{n} t_i^{\beta_2}.$$

After expanding the right-hand side of the inequality using $t_i^{\beta_2} = exp(\beta_2 \log(t_i))$, disregarding second-order and higher terms, and rearranging the inequality, we

obtain

$$\beta_2 > \frac{0.5}{\frac{1}{n}\sum_{i=1}^{n}[-log(t_i)]}. \tag{2.17}$$

Another bound on a solution is obtained by considering (2.16). We note that for $\beta_1 > 0$, the numerator in (2.16) needs to be positive. This occurs when

$$\beta_2 > \frac{1}{\frac{1}{n}\sum_{i=1}^{n}[-log(t_i)]}. \tag{2.18}$$

Since the right-hand side of (2.17) is always smaller than that of (2.18), the lower bound on β_2 is given by (2.17); however, for $\beta_1 > 0$, we require (2.18) to hold. In other words, (2.18) is the lower bound on β_2.

POSN

Introduction

As far as we know, results on the conditions for which a positive, finite-valued unique m.l.e. solution is obtained for the POSN model have not been published. Just as multiple roots have been observed in some instances for the equivalent POS (Littlewood) model [12], we have observed multiple roots for some data sets under the POSN model. At this time, we have not been able to obtain conditions for obtaining finite m.l.e. solution(s) for this model. We have provided some results for obtaining bounds on the solution.

Maximum likelihood estimation equations

For the POSN model, $\lambda(t;\eta) = \frac{v\beta_2\beta_1^{\beta_2}}{(\beta_1+t)^{(\beta_2+1)}}$ and $\Lambda(T1;\eta) = v[1 - \left(\frac{\beta_1}{\beta_1+T1}\right)^{\beta_2}]$. After

substituting the required derivatives into the equations for β, and rearranging some of the terms, we obtain the following equations:

$$\frac{1}{n}\sum_{i=1}^{n} log\left(1+\frac{t_i}{\beta_1}\right) - \frac{1}{\beta_2} = -\frac{log(1+\frac{T1}{\beta_1})}{(1+\frac{T1}{\beta_1})^{\beta_2}-1} \tag{2.19}$$

$$\frac{1}{\beta_2}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\frac{t_i}{\beta_1}}\right] - \frac{1}{n}\sum_{i=1}^{n}\frac{\frac{t_i}{\beta_1}}{1+\frac{t_i}{\beta_1}} = \frac{\frac{T1}{\beta_1}/(1+\frac{T1}{\beta_1})}{(1+\frac{T1}{\beta_1})^{\beta_2}-1}. \tag{2.20}$$

Since the denominator of the expression on each of the right-hand sides is shared by both equations, we can eliminate this term by rearranging each equation and then subtracting one from the other. We then obtain

$$\frac{1}{\beta_2}\left[\left(\frac{1+\frac{T1}{\beta_1}}{\frac{T1}{\beta_1}}\right)\left(\sum_{i=1}^{n}\frac{1}{1+\frac{t_i}{\beta_1}}\right) - \frac{1}{log(1+\frac{T1}{\beta_1})}\right] - \left(\frac{1+\frac{T1}{\beta_1}}{\frac{T1}{\beta_1}}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{t_i}{\beta_1}}{1+\frac{t_i}{\beta_1}}\right) + \frac{1}{n}\frac{\sum_{i=1}^{n} log(1+\frac{t_i}{\beta_1})}{log(1+\frac{T1}{\beta_1})} = 0.$$

Rearranging the above equation yields an expression for β_2,

$$\beta_2 = \frac{\left(\frac{1+\frac{T1}{\beta_1}}{\frac{T1}{\beta_1}}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\frac{t_i}{\beta_1}}\right) - \frac{1}{log(1+\frac{T1}{\beta_1})}}{\left(\frac{1+\frac{T1}{\beta_1}}{\frac{T1}{\beta_1}}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{t_i}{\beta_1}}{1+\frac{t_i}{\beta_1}}\right) - \frac{1}{n}\frac{\sum_{i=1}^{n} log(1+\frac{t_i}{\beta_1})}{log(1+\frac{T1}{\beta_1})}}. \tag{2.21}$$

Equations (2.19) and (2.21) can then be used to solve for the m.l.e.'s of β_1 and β_2.

Bounds on solution

Once again, we found it useful to scale the failure times so that $T1 \equiv 1$. We can

then write equation (2.19) as

$$\frac{1}{n}\sum_{i=1}^{n} log\left(1+\frac{t_i}{\beta_1}\right) = \frac{1}{\beta_2} - \frac{log(1+\frac{1}{\beta_1})}{(1+\frac{1}{\beta_1})^{\beta_2}-1}, \tag{2.22}$$

and equation (2.21) as

$$\beta_2 = \frac{(\beta_1+1)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\frac{t_i}{\beta_1}}\right) - \frac{1}{log(1+\frac{1}{\beta_1})}}{(\beta_1+1)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{t_i}{\beta_1}}{1+\frac{t_i}{\beta_1}}\right) - \frac{1}{n}\frac{\sum_{i=1}^{n} log(1+\frac{t_i}{\beta_1})}{log(1+\frac{1}{\beta_1})}}. \tag{2.23}$$

The argument for obtaining preliminary bounds on the solution is as follows. If we let $g(x,d)$, for $x = \beta_2$ and $d = log(1+\frac{1}{\beta_1})$, denote the right-hand side of (2.22), a Taylor expansion of the function $\exp(x)$ shows that $\lim_{x\to 0} g(x,d) = d/2$ and $\lim_{x\to\infty} g(x,d) = 0$. Since the left-hand side is always greater than zero, a sufficient condition for a positive root for (2.22) is

$$\frac{\frac{1}{n}\sum_{i=1}^{n} log\left(1+\frac{t_i}{\beta_1}\right)}{log(1+\frac{1}{\beta_1})} < \frac{1}{2}. \tag{2.24}$$

If (2.24) does not hold, there will not be a solution to (2.15). If this condition does hold, a solution may exist. Setting $t_i = t_1$ for all i yields the lowest value for the left-hand side of (2.24). If we let $g(x,d)$, for $x = \beta_1$ and $d = t_1$, denote the left-hand side, we can easily show that $\lim_{x\to 0} g(x,d) = 1$, $\lim_{x\to\infty} g(x,d) = d$ and $g(x,d)$ is monotonically decreasing in x. A lower bound for β_1 can be obtained if $d < 0.5$. In this case, we set $g(x,d) = 0.5$ and solve for the positive root for x.

Another bound on a solution is obtained by looking at (2.23). Note that for

$\beta_2 > 0$, the numerator in (2.23) needs to be positive. This occurs when

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\frac{t_i}{\beta_1}} > \frac{1}{(\beta_1+1)\,log(1+\frac{1}{\beta_1})}. \tag{2.25}$$

Setting $t_i = t_1$ for all i yields the largest value for the left-hand side of (2.25). Similarly, setting $t_i = t_n$ yields the smallest value. If we let $g(x,d)$, for $x = \beta_1$ and $d = t_n$ or t_1, denote the left-hand side, we can easily show that $\lim_{x\to 0} g(x,d) = 0$ and $\lim_{x\to\infty} g(x,d) = 1$, regardless of the value of d. Unfortunately, $g(x,d)$ is *not* monotonically increasing in x for all values of d.

2.3 Conditions for positive, finite $\hat{\Lambda}(\infty)$ – grouped case

2.3.1 General case

The log-likelihood function for a grouped realization of a time-truncated NHPP with the intensity function of the form $\lambda(t;\boldsymbol{\eta}) = v\,\lambda_0(t;\boldsymbol{\beta})$ and corresponding mean value function $\Lambda(T1;\boldsymbol{\eta}) = v\,\Lambda_0(T1;\boldsymbol{\beta})$, for scalar $v > 0$ and vector $\boldsymbol{\beta} > 0$, is

$$l(\boldsymbol{\eta}) = n\log(v) + \sum_{k=1}^{K}\{n_k\,\log[\Lambda_0(t_k;\boldsymbol{\beta}) - \Lambda_0(t_{k-1};\boldsymbol{\beta})] - log(n_k!)\} - v\,\Lambda_0(T1;\boldsymbol{\beta}), \tag{2.26}$$

where $n_1,\ldots,n_k$ are the numbers of failures in time intervals (t_{k-1},t_k), $k = 1,\ldots,K$ and $0 = t_0 < t_1 < \ldots < t_k = T1$. Differentiating the above with respect to parameters $(v,\boldsymbol{\beta})$ yields the following set of $J+1$ (score) equations to be solved for

the maximum likelihood estimators:

$$\frac{\partial l(\boldsymbol{\eta})}{\partial v} = \frac{n}{v} - \Lambda_0(T1;\boldsymbol{\beta}), \qquad \sum_{k=1}^{K} n_k = n$$

$$\frac{\partial l(\boldsymbol{\eta})}{\partial \beta_j} = \sum_{k=1}^{K} \frac{\partial}{\partial \beta_j} \{n_k \log[\Lambda_0(t_k;\boldsymbol{\beta}) - \Lambda_0(t_{k-1};\boldsymbol{\beta})]\} - v \frac{\partial}{\partial \beta_j} \Lambda_0(T1;\boldsymbol{\beta}),$$
$$j = 1, \ldots, J.$$

We first note that solving for v in the first equation yields the same m.l.e. as that given by (2.2) for the ungrouped scenario. This expression for $\hat{v}$ can then be substituted into the other J score functions to obtain

$$\sum_{k=1}^{K} n_k \left\{ \frac{\dot{\Lambda}_0(t_k;\boldsymbol{\beta}) - \dot{\Lambda}_0(t_{k-1};\boldsymbol{\beta})}{\Lambda_0(t_k;\boldsymbol{\beta}) - \Lambda_0(t_{k-1};\boldsymbol{\beta})} - \frac{\dot{\Lambda}_0(T1;\boldsymbol{\beta})}{\Lambda_0(T1;\boldsymbol{\beta})} \right\} = 0, \tag{2.27}$$

for $\dot{f} = \frac{\partial}{\partial \beta_j} f$, $j = 1, \ldots, J$. With v eliminated, we can solve the above equation(s) for $\hat{\boldsymbol{\beta}}$ and then substitute into the expression for $\hat{v}$. Since a closed form solution for $\hat{\boldsymbol{\beta}}$ does not exist, numerical procedures are used to obtain the solution to the above equations for $\hat{\boldsymbol{\beta}}$. Note that a closed form solution is not available for the POWN model in this scenario. As in the ungrouped case, if the expected total number of events is large and appropriate conditions on the model hold, we know that $\hat{\boldsymbol{\eta}}$ is asymptotically normally distributed with mean $\boldsymbol{\eta}$ and covariance matrix $I(\boldsymbol{\eta})^{-1}$. As in the earlier case, the Fisher (or expected) information matrix $I(\boldsymbol{\eta})$ has entries

$$I_{ij}(\boldsymbol{\eta}) = E\left(\frac{-\partial^2 l(\boldsymbol{\eta})}{\partial \eta_i \partial \eta_j} \right), \quad i, j = 1, \ldots, J+1.$$

We can estimate the above with $I(\hat{\eta})$, or with the observed information matrix I_O. which has entries

$$I_{O,ij} = \left(\frac{-\partial^2 l(\eta)}{\partial \eta_i \partial \eta_j} \Big|_{\eta=\hat{\eta}} \right), \quad i,j = 1,...,J+1.$$

As in the ungrouped case, in some cases, equation (2.27) will not have an "admissible" solution $\hat{\beta} > 0$, but there may be an "unrestricted" solution with some parameter estimates negative. This corresponds to $l(\eta)$ being maximized at a boundary point for the restricted parameter space with $\beta > 0$ and $v > 0$. For the models we consider, it is possible to extend the parameter space for β and v and discuss the solution of equation (2.3). We do this below.

2.3.2 Specific cases

EOSN

Introduction

As far as we know, a necessary and sufficient condition for which a m.l.e. solution is unique, positive and finite-valued for the grouped EOSN model has not been published. We have derived such a condition, along with bounds on the solution. We also conducted a small simulation study to give some probability statements on how often the necessary and sufficient condition is not obtained.

Necessary and sufficient condition for a positive $\hat{\beta}$

For the EOSN model, $\lambda(t;\eta) = v\beta \exp(-\beta t)$ with corresponding $\Lambda(t;\eta) = v[1 - \exp(-\beta t)]$. After substitute the required derivatives into (2.27), and rearranging

some of the terms, we obtain

$$0 = \sum_{k=1}^{K} n_k \left\{ \frac{(\exp(\beta T1) - 1)(t_k \exp(\beta(t_{k-1} - t_k)) - t_{k-1})}{(1 - \exp(\beta(t_{k-1} - t_k)))(\exp(\beta T1) - 1)} - \frac{T1(1 - \exp(\beta(t_{k-1} - t_k)))}{(1 - \exp(\beta(t_{k-1} - t_k)))(\exp(\beta T1) - 1)} \right\}. \quad (2.28)$$

We let $g(\beta)$ denote the bracketed term on the right-hand side of (2.28). Using L'Hôpital's rule, we see that $\lim_{\beta \to \infty} g(\beta) = \sum_{k=1}^{K} n_k(-t_{k-1}) < 0$. A Taylor expansion of the function $\exp(\cdot)$ shows that $\lim_{\beta \to 0} g(\beta) = \frac{1}{2}[nT1 - \sum_{k=1}^{K} n_k(t_k + t_{k-1})]$. Furthermore, $g(\beta)$ is monotonically decreasing because we can show that

$$\frac{\partial g(\beta)}{\partial \beta} = -\exp(\beta(t_{k-1} - t_k)) \left\{ \frac{(t_k^2 + t_{k-1}^2)(1 + \exp(2\beta)) + \exp(\beta)(4t_k t_{k-1} + 2))}{(-1 + \exp(\beta(t_{k-1} - t_k)))^2(\exp(\beta) - 1)^2} - \frac{(2t_k t_{k-1})(1 + \exp(2\beta))}{(-1 + \exp(\beta(t_{k-1} - t_k)))^2(\exp(\beta) - 1)^2} - \frac{\exp(\beta)(2(t_k^2 + t_{k-1}^2) + \exp(\beta(t_{k-1} - t_k)) + \exp(\beta(t_k - t_{k-1})))}{(-1 + \exp(\beta(t_{k-1} - t_k)))^2(\exp(\beta) - 1)^2} \right\} < 0,$$

for $\beta > 0$. Since the function is monotonically decreasing and its lower limit is negative, it follows that a necessary and sufficient condition for a positive finite solution to (2.28) is for the upper limit to be positive. This occurs when

$$\frac{1}{n}\frac{1}{T1}\sum_{k=1}^{K} n_k(t_k + t_{k-1}) < 1. \quad (2.29)$$

Note that if we substitute $n_k = 1$ for all k, we obtain $K = n$ and this condition reduces to that given for the ungrouped EOSN model, given by (2.5).

Simulation study

A small simulation study was performed to determine how often a finite positive-

Table 2.6: Percentages of the thirty sets of 2000 simulated data sets that did not satisfy condition (2.29).

$a \setminus v$	15	25	50	100	200
$K = 5$					
0.5	32.25	26.90	17.20	9.80	2.90
0.9	1.60	0.35	0.00	0.00	0.00
$K = 10$					
0.5	30.50	24.50	16.45	7.20	2.20
0.9	1.15	0.30	0.00	0.00	0.00
$K = 15$					
0.5	30.30	25.40	16.55	8.50	2.50
0.9	1.50	0.20	0.00	0.00	0.00

valued m.l.e. solution is not available for grouped data under the EOSN model. We generated samples from the EOSN model with $\beta = 1$, $v = 15, 25, 50, 100, 200$, $K = 5, 10, 15$, and for $T1 = 0.6931, 2.3026$. In particular, 2000 samples were obtained for each combination of v and $T1$ in the following manner:

1. Simulate $K - 1$ uniform $(0, T1)$ random variables. The sorted values are the interval endpoints, $t_1, \ldots, t_k$. I have set $t_0 = 0$ and $t_K = T1$.

2. Simulate n_k Poisson random variables with mean given by $v(\exp(-\beta\, t_{k-1}) - \exp(-\beta\, t_k))$, for $k = 1, \ldots, K$.

The percentages of the thirty sets of simulated data which did not satisfy condition (2.29) are given in Table 2.6. The results are similar to those obtained for the ungrouped case. The only difference is that these percentages are slightly larger in size.

POWN

Introduction

As indicated in the introduction to this chapter, a finite, positive-valued closed form solution for β (and hence, v) always exists in the ungrouped case. Since a closed form solution is not available for the grouped case, it is of interest to obtain a necessary and sufficient condition for which a finite, positive-valued solution for β exists for the grouped POWN model. We have shown that an unique, positive solution for β *always* exists for this model. We have also obtained bounds on the solution.

Necessary and sufficient condition for a positive $\hat{\beta}$

For the POWN model, $\lambda(t;\eta) = v\beta t^{\beta-1}$ with corresponding $\Lambda(t;\eta) = vt^{\beta}$. After substituting the required derivatives into (2.27), and rearranging some of the terms, we obtain

$$0 = \sum_{k=1}^{K} n_k \left\{ \frac{log(t_k) - log(t_{k-1})(t_{k-1}/t_k)^{\beta} - log(T1)(1-(t_{k-1}/t_k)^{\beta})}{1-(t_{k-1}/t_k)^{\beta}} \right\} \quad (2.30)$$

A Taylor expansion of the right-hand side, $g(\beta)$, shows that $\lim_{\beta\to 0} g(\beta) = \infty$ and $\lim_{\beta\to\infty} g(\beta) = \sum_{k=1}^{K} n_k log(t_k/T1) < 0$. In addition, $g(\beta)$ is monotonically decreasing because we can easily show that

$$\frac{\partial g(\beta)}{\partial \beta} = \frac{(t_{k-1}/t_k)^{\beta}\log(t_{k-1}/t_k)\log(t_k/t_{k-1})}{(-1+(t_{k-1}/t_k)^{\beta})^2} < 0,$$

for all β. Since the function is monotonically decreasing and its upper and lower limit is negative and then positive, respectively, it follows that there will always be

an unique positive solution to (2.30).

Bounds on solution

An ad hoc approach was taken to obtain bounds on the solution. First, we will obtain a lower bound on the solution. For convenience, we set $T1 \equiv 1$. After setting $T1 \equiv 1$, (2.30) becomes

$$0 = \sum_{k=1}^{K} n_k \left\{ \frac{log(t_k) - log(t_{k-1})(t_{k-1}/t_k)^{\beta}}{1 - (t_{k-1}/t_k)^{\beta}} \right\}. \tag{2.31}$$

Then, rewriting $\frac{t_{k-1}}{t_k} = 1 + (-\frac{t_k - t_{k-1}}{t_k})$, we expand $(\frac{t_{k-1}}{t_k})^{\beta}$ using a binomial expansion. Ignoring the third order and higher terms of this expansion, and rearranging (2.31) yields

$$\sum_{k=1}^{K} n_k \left\{ \frac{log(t_k) - log(t_{k-1})(\frac{t_{k-1}}{t_k})^{\beta}}{\beta(\frac{t_k - t_{k-1}}{t_k})} \right\} < 0.$$

After re-writing and expanding $(\frac{t_{k-1}}{t_k})^{\beta}$ once again, we obtain

$$\sum_{k=1}^{K} n_k \left\{ \frac{log(\frac{t_k}{t_{k-1}}) - \beta(\frac{t_k - t_{k-1}}{t_k})[-log(t_{k-1})]}{\frac{t_k - t_{k-1}}{t_k}} \right\} < 0.$$

A simple rearrangement of the above equation yields a lower bound for $\hat{\beta}$,

$$\hat{\beta} > \frac{\sum_{k=1}^{K} n_k (\frac{t_k}{t_{k-1} - t_k}) log(\frac{t_k}{t_{k-1}})}{\sum_{k=1}^{K} n_k (-log(t_{k-1}))}.$$

An upper bound is obtained in a similar manner. In this case, we rewrite (2.31) as

$$\sum_{k=1}^{K} n_k \left\{ \frac{(t_k/t_{k-1})^{\beta} log(t_k) - log(t_{k-1})}{1 - (t_{k-1}/t_k)^{\beta}} \right\}. \tag{2.32}$$

Then, rewriting $\frac{t_k}{t_{k-1}} = 1 + (\frac{t_k - t_{k-1}}{t_k})$, we expand $(\frac{t_k}{t_{k-1}})^\beta$. Ignoring the third order and higher terms of this expansion, and rearranging (2.32) yields

$$\sum_{k=1}^{K} n_k \left\{ \frac{(\frac{t_k}{t_{k-1}})^\beta log(t_k) - log(t_{k-1})}{\beta(\frac{t_k - t_{k-1}}{t_k})} \right\} > 0.$$

After re-writing and expanding $(\frac{t_k}{t_{k-1}})^\beta$ once again, we obtain

$$\sum_{k=1}^{K} n_k \left\{ \frac{log(\frac{t_k}{t_{k-1}}) + \beta(\frac{t_k - t_{k-1}}{t_{k-1}}) log(t_k)}{\frac{t_k - t_{k-1}}{t_{k-1}}} \right\} > 0.$$

A simple rearrangement of the above equation yields an upper bound for $\hat{\beta}$.

$$\hat{\beta} < \frac{\sum_{k=1}^{K} n_k (\frac{t_{k-1}}{t_k - t_{k-1}}) log(\frac{t_k}{t_{k-1}})}{\sum_{k=1}^{K} n_k (-log(t_k))}.$$

2.4 Alternative parameterizations

If there is some evidence that the approximate asymptotic inferences based on an original parameterization of a model may be misleading, it is of interest to consider a reparameterization of the model. We examined a pair of parameterizations as alternatives to those in Table 1.1. We first considered the EOSN model. We compared $\lambda(t;\eta) = \rho \exp(-\beta t)$, for $\rho = v\beta > 0$, with $\lambda(t;\eta) = \exp(\alpha_1 + \alpha_2 t)$, for $-\infty < \alpha_1 = log(v\beta) < \infty$ and $\alpha_2 = -\beta < 0$. We denote the first one a "Raftery" parameterization because it was first used by him in [90], whereas we denote the latter the "log-linear" parameterization because it is of log-linear form [27]. Preliminary likelihood ratio (LR) plots based on all three parameterizations of the EOSN model for DS1 and DS3 indicate that there is a noticeable improvement

Table 2.7: A list of the intensity functions of common software reliability NHPP models under three different parameterizations.

Model	*original*	*Raftery*	*log-linear*
$\Lambda(\infty)$ *finite*			
EOSN	$v\beta\, e^{(-\beta t)}$	$\rho\, e^{(-\beta t)}$	$e^{(\alpha_1+\alpha_2 t)}$
GAMN	$v\beta^2 t\, e^{(-\beta t)}$	$\rho\, t\, e^{(-\beta t)}$	$e^{(\alpha_1+log(t)+\alpha_2 t)}$
WOSN	$v\beta_1\beta_2\, t^{\beta_2-1}\, e^{(-\beta_1 t^{\beta_2})}$	$\rho\, t^{\beta_2-1}\, e^{(-\beta_1 t^{\beta_2})}$	$e^{(\alpha_1+(\alpha_2-1)\,log(t)+\alpha_3 t^{\alpha_2})}$
POSN	$\frac{v\beta_2\,\beta_1^{\beta_2}}{(\beta_1+t)^{(\beta_2+1)}}$	$\frac{\rho\,\beta_1^{\beta_2}}{(\beta_1+t)^{(\beta_2+1)}}$	$e^{(\alpha_1+\alpha_2\, log(\alpha_3+t))}$
$\Lambda(\infty)$ *infinite*			
LOGN	$\frac{v\beta}{\beta t+1}$	$\frac{\rho}{\beta t+1}$	$e^{(\alpha_1-log(\alpha_2 t+1))}$
POWN	$v\beta\, t^{\beta-1}$	$\rho\, t^{\beta-1}$	$e^{(\alpha_1+\alpha_2\, log(t))}$

in the shape of the contours under the Raftery and the log-linear parameterizations for DS3.

We can obtain similar "Raftery" and "non-linear" reparameterizations for the other models listed in Table 1.1. Table 2.7 provides a list of the intensity functions for these NHPP models under all three parameterizations. Likelihood ratio contour plots based on the above three parameterizations of the EOSN, GAMN, LOGN, and POWN models are given in Figures 2.3 - 2.6, respectively. Since the WOSN model has three parameters, *profile* likelihood ratio contour plots were obtained and given in Figures 2.7 and 2.8. For instance, the profile likelihood contour plot for β_1 was found by 1) setting $\beta_1 = \hat{\beta}_1$, where $\hat{\beta}_1$ was obtained from maximizing the log-likelihood over all parameters; 2) maximizing the log-likelihood over the other two parameters, and; 3) substituting the particular value for β_1 and the maximized values for the other parameters into the log-likelihood. A noticeable improvement under the two alternative parameterizations is shown in the plotted

Figure 2.3: LR contour plots using three parameterizations for EOSN and DS1.3.

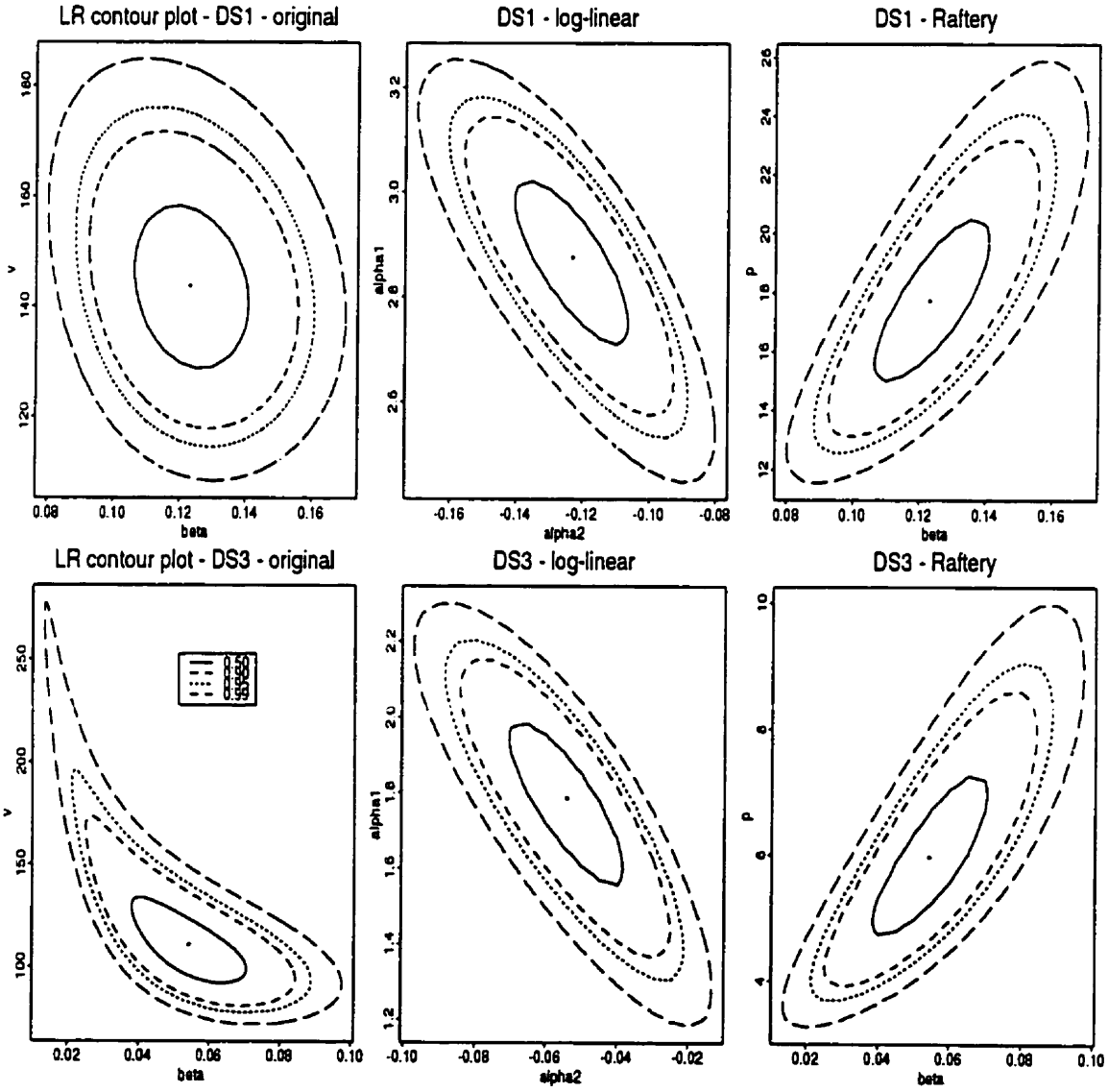


Figure 2.4: LR contour plots using three parameterizations for GAMN and DS1,3.

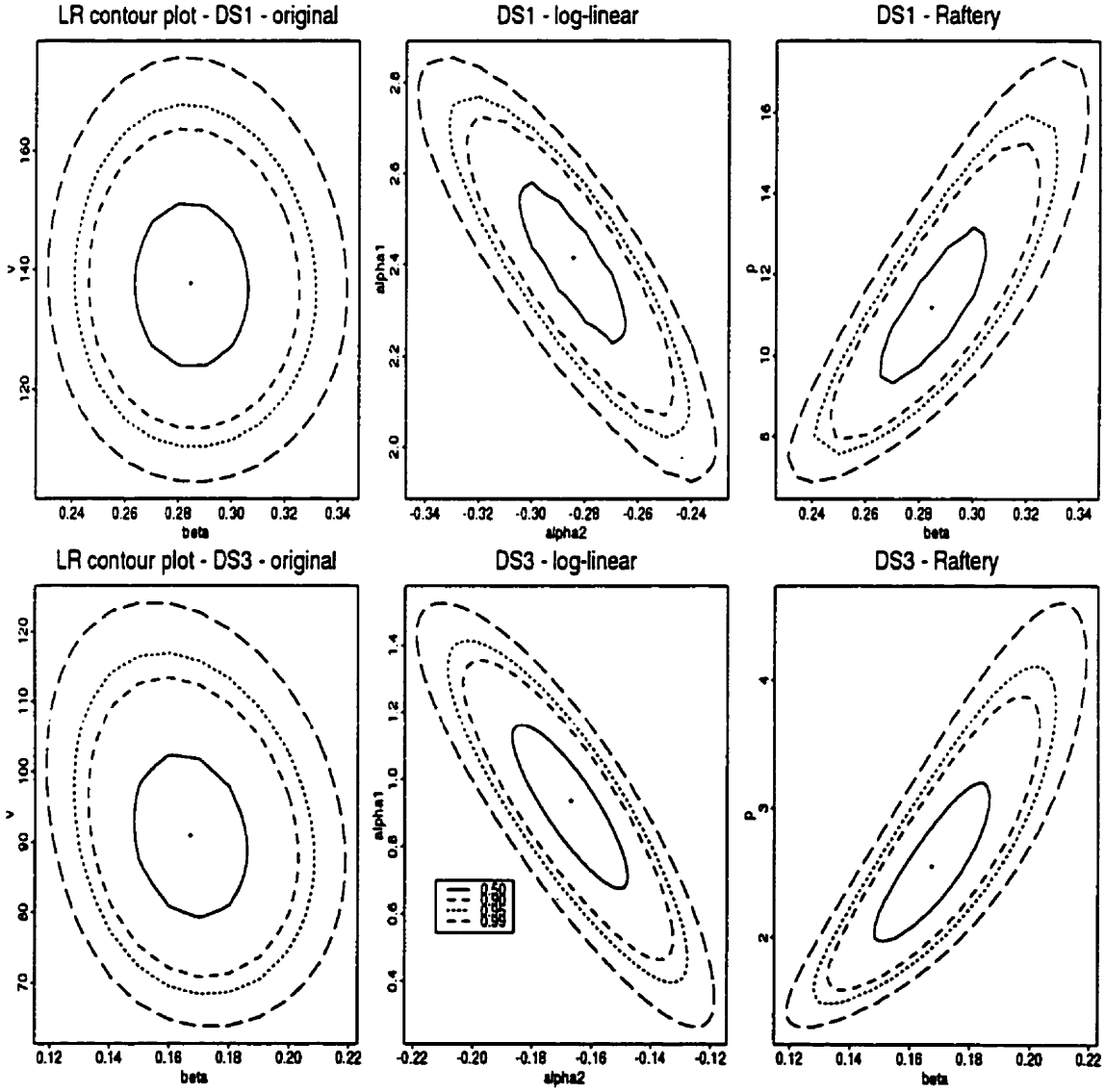


Figure 2.5: LR contour plots using three parameterizations for LOGN and DS1.3.

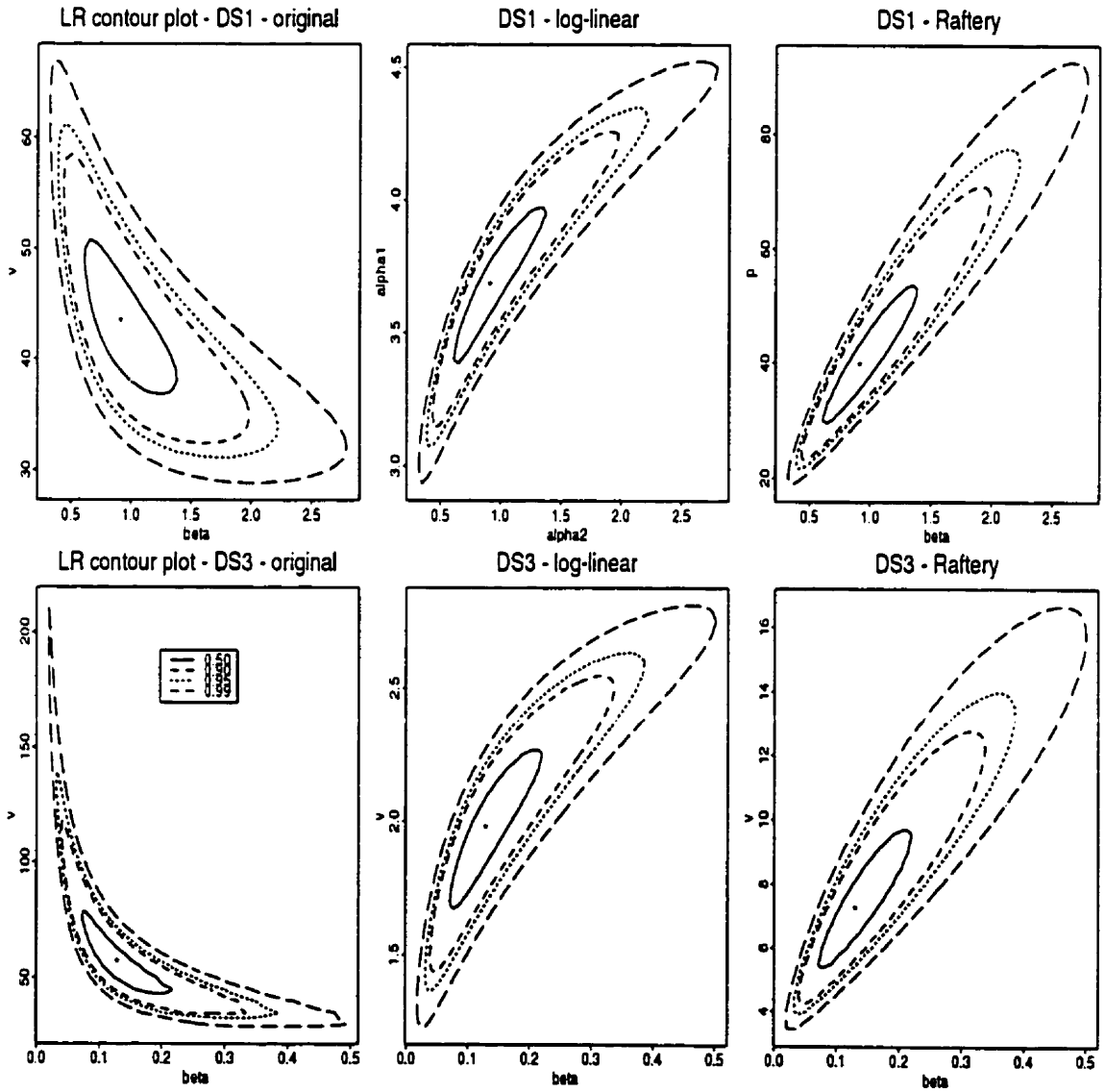


Figure 2.6: LR contour plots using three parameterizations for POWN and DS1,3.

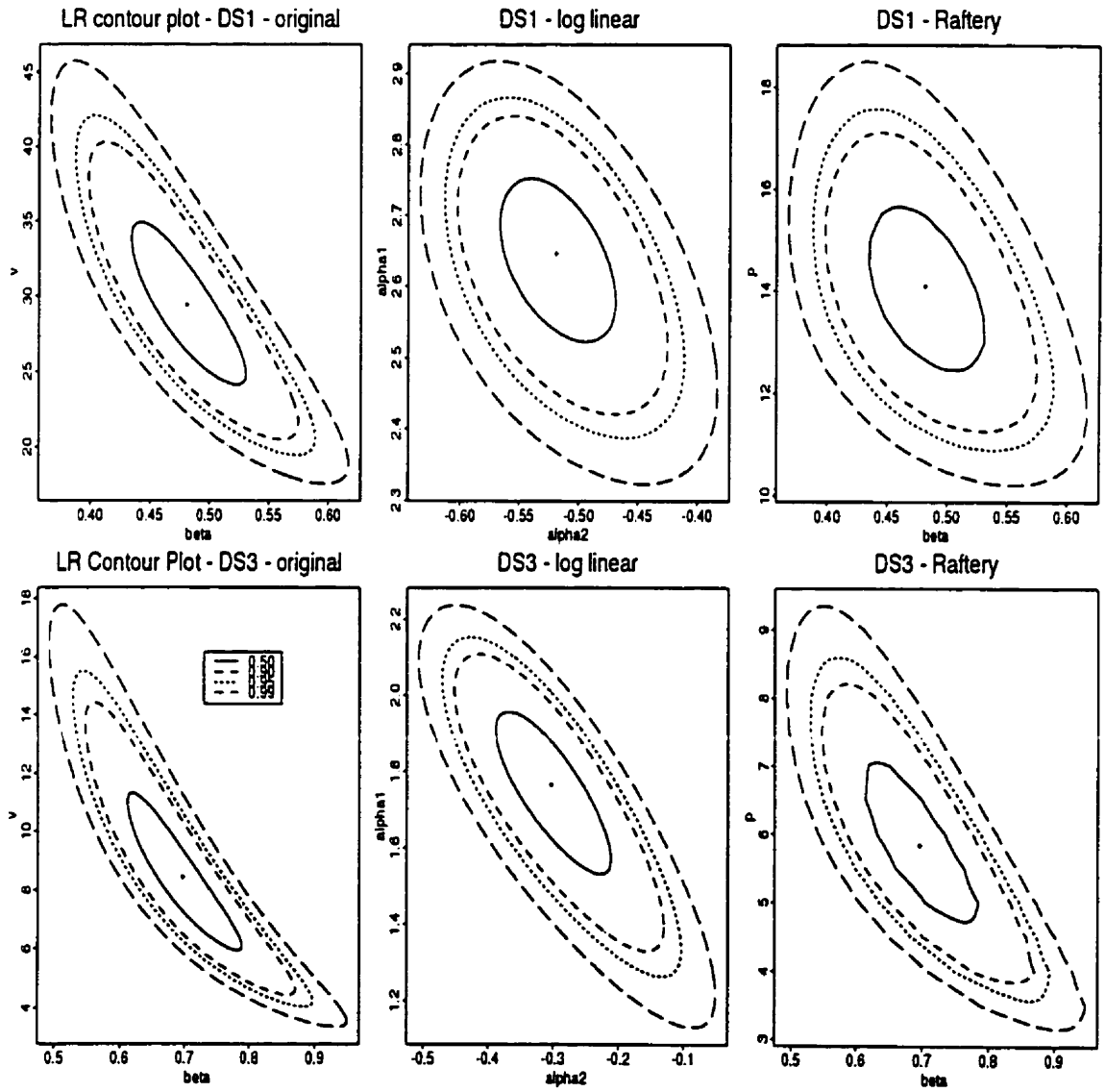


Figure 2.7: Profile LR contour plots using three parameterizations for WOSN and DS1.

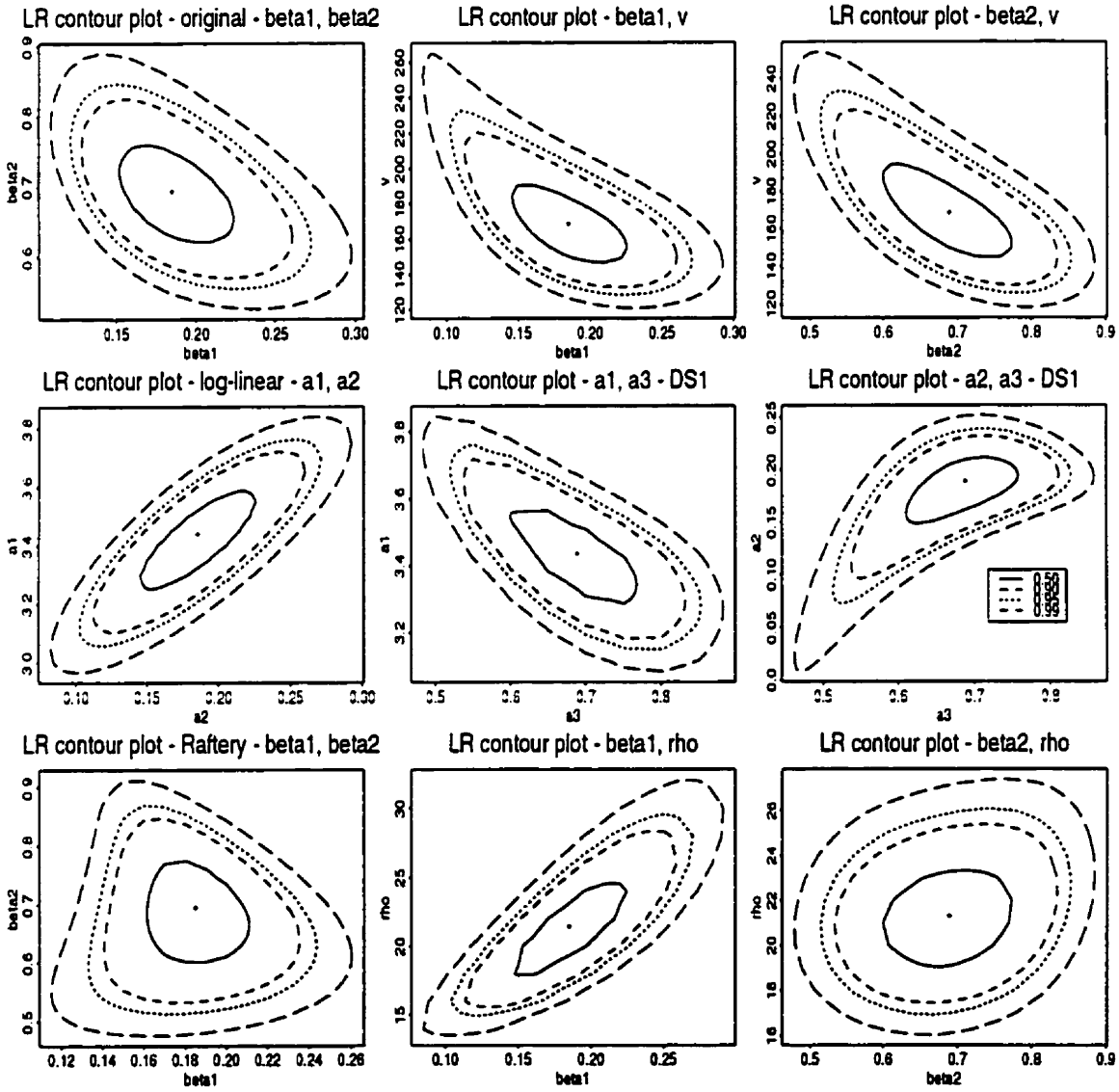
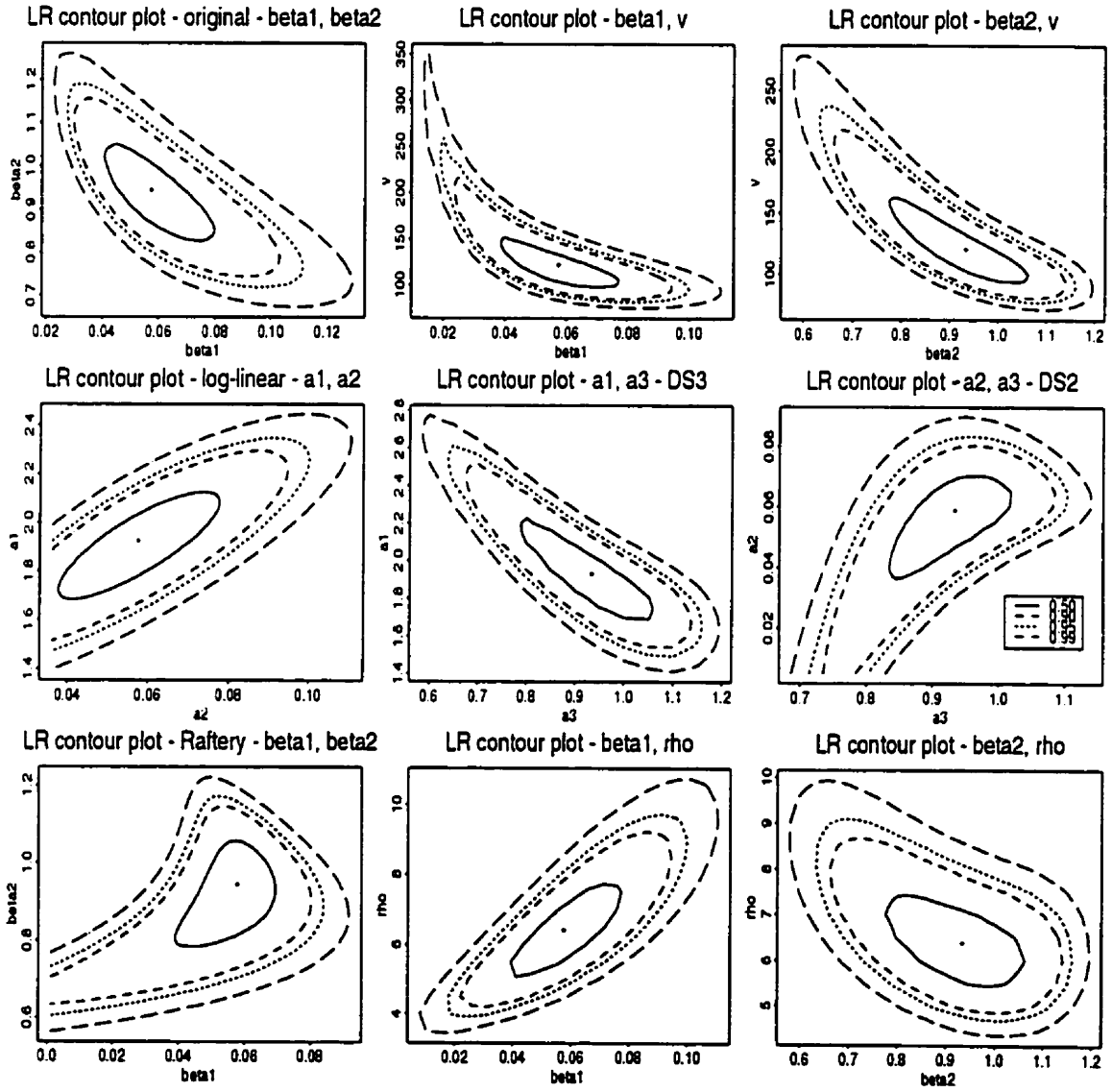


Figure 2.8: Profile LR contour plots using three parameterizations for WOSN and DS3.



likelihood contours for DS3 for the EOSN, LOGN, POWN and WOSN models. Raftery's reparameterization seems better under the LOGN and WOSN models, but the log-linear reparameterization seems better for the EOSN and POWN models. The original parameterization is best for the GAMN model.

2.5 Conclusions

With the exception of the POWN model, the maximum likelihood estimation solution for the models given in Table 1.1 is not of closed form. This implies that numerical methods, such as the Newton-Raphson or bisection methods, are needed to obtain the m.l.e.'s. Obtaining the m.l.e.'s is not difficult for these models, with the exception of the POSN model. Further investigation into the possibility of multiple roots is required for this model.

We have determined conditions for which $\hat{v} = \hat{\Lambda}(\infty)$ is finite and positive-valued for the EOSN, GAMN and LOGN models under the ungrouped data scenario, and for the EOSN model under the grouped data scenario. We have shown that an unique, positive solution for β always exists for the POWN model under the grouped data case. In general, the probability statements on how often these conditions do not hold for the EOSN, GAMN and LOGN models will depend on β and $T1$ only through the combination $\beta T1$ because β is a scale parameter in these models. The simulation studies for these models indicate that when only roughly 50% of the total expected number of failures have been observed by time $T1$ (for given β), there is a substantial probability that we do not obtain positive, finite estimates for v. This implies that, in some situations, there is a substantial probability that

there will be insufficient data available to make any valid inferential statements about a finite v. Without good estimates for v, we can not obtain good predictions of quantities of interest that depend on v, such as $N2$ and $N3$. Consequently, for those situations whereby we do not obtain finite, positive-valued estimates for v based on data observed over $(0, T1]$, it is best to continue observing the process for a longer period of time before making any inferences about the process.

In order to improve the adequacy of the asymptotic approximations for obtaining confidence (and prediction) intervals of the quantities of interest, we suggest using either the log-linear or Raftery parameterizations of the model of interest for estimation purposes. There is clearly a noticeable improvement in the shape of the likelihood ratio contours under these parameterizations, with the exception of the GAMN model. Maximum likelihood estimates are easier to obtain iteratively when the parameterizations giving approximately quadratic log-likelihoods are used, and Wald-type confidence interval procedures are more accurate. When there is only a single β parameter, we need only one-dimensional iteration to maximize $l(\hat{v}(\beta), \beta)$, so this point about normal approximations is less important for these models. Of course, the plots in this chapter are based on two specific data sets, but they illustrate general features. In practice, we recommend that log-likelihood contours be routinely plotted and examined in order to select good parameterizations.

Chapter 3

Goodness of Fit

3.1 Introduction

There are a variety of goodness of fit methods for assessing the adequacy of a particular NHPP family of models based on one observed realization of a recurrent event process. In this section, we address separately each method and motivate the need for our proposed goodness of fit technique. We follow with a chapter outline.

There are various methods available for testing the validity of the homogeneous (intensity constant over time) Poisson process or renewal process (a nonterminating sequence of independent, identically distributed interarrival times) against specific nonhomogeneous alternatives. Informal methods include an "eye-ball analysis" of an index plot of the interfailure times and a lag one plot of the interfailure times under a log or square root transformation to check for dependence among the interfailure times. For example, Figure 3.1 indicates that there is a gradual increase in the length of the interfailure times over time for DS1-4. The nonhomogeneity

in the interfailure times implies that an HPP or renewal assumption would be inappropriate. Formal methods include the tests for a monotone trend discussed in [8, chapter 3], [10], and [27, chapter 5], such as the Laplace test optimal against the EOSN model and the $\mathcal{U}$ test (same notation used in [8, page 79]) optimal against the POWN model. The point process model, considered by [68], that incorporates *both* renewal and time trends may also be used to test renewal behaviour against specific forms of time nonhomogeneity. Although they are useful for many reliability applications, the above methods are of limited use in other situations where there is no question of time nonhomogeneity in the process, such as software system reliability. In the situation where we anticipate other patterns of behaviour in addition to time nonhomogeneity in the process, it may be of interest to compare NHPP models with other nonhomogeneous point process models, such as cluster or change-point processes. This can be done by model expansion or nesting (e.g. [68]). Further discussion on change-point processes is given in Chapter 5.

The most common technique used to check the validity of the form of any parametric NHPP model is to "eyeball" the plot of the cumulative number of failures and the estimated model-based mean value function against time. Such plots are given in Figure 3.2. These plots indicate clearly that the estimated mean value functions based on the POWN and GAMN models do *not* fit the observed cumulative number of failures in DS1-4. In addition, the plots indicate that all of the EOSN, WOSN, and LOGN models seem to fit DS3-4 well. In contrast, there is an apparent lack of fit between the EOSN, WOSN and LOGN models and the observed data sets DS1,2 due to the presence of change-points in the failure behaviour over

Figure 3.1: Plots of the ordered interfailure times for DS1-4.

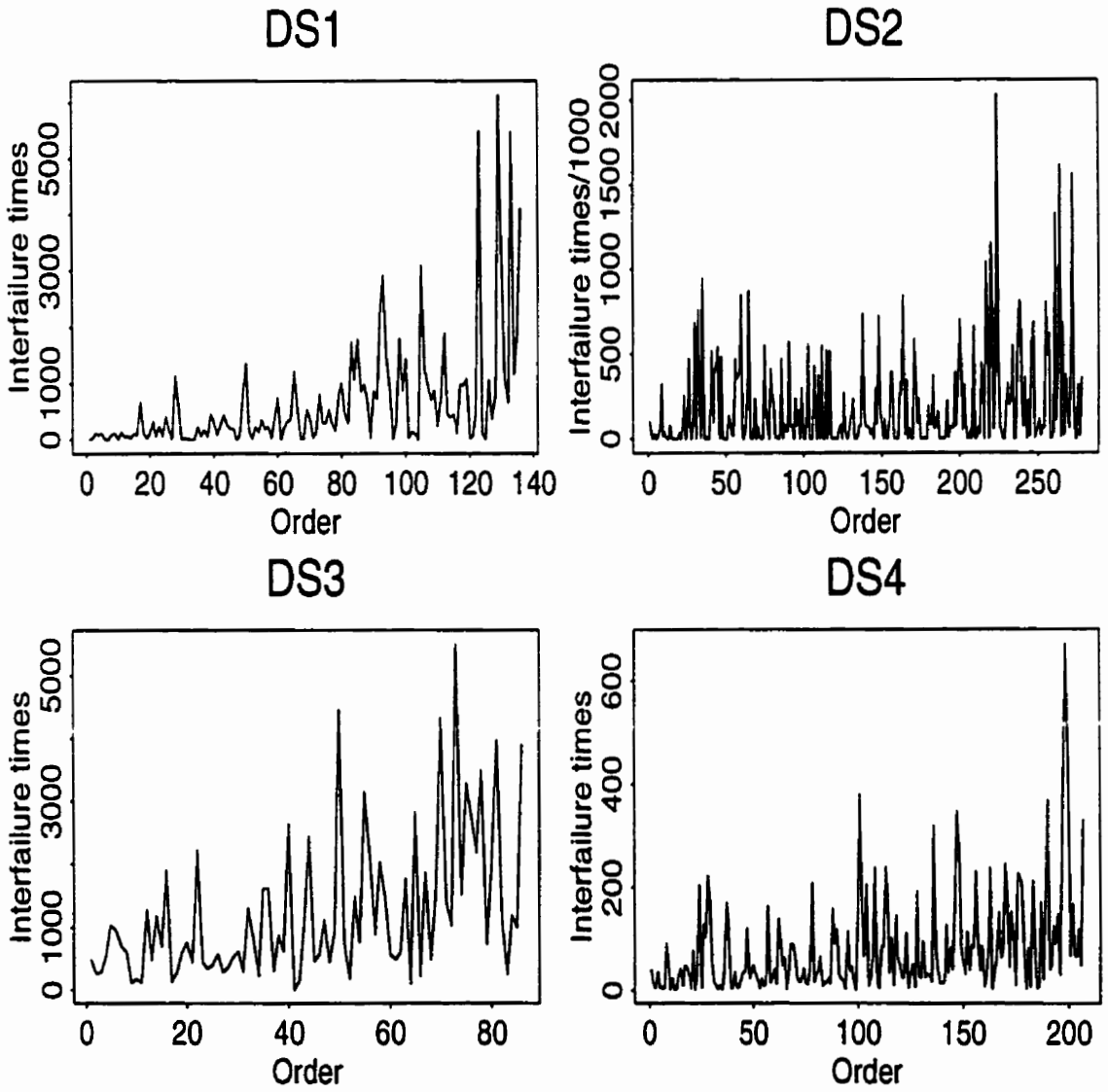
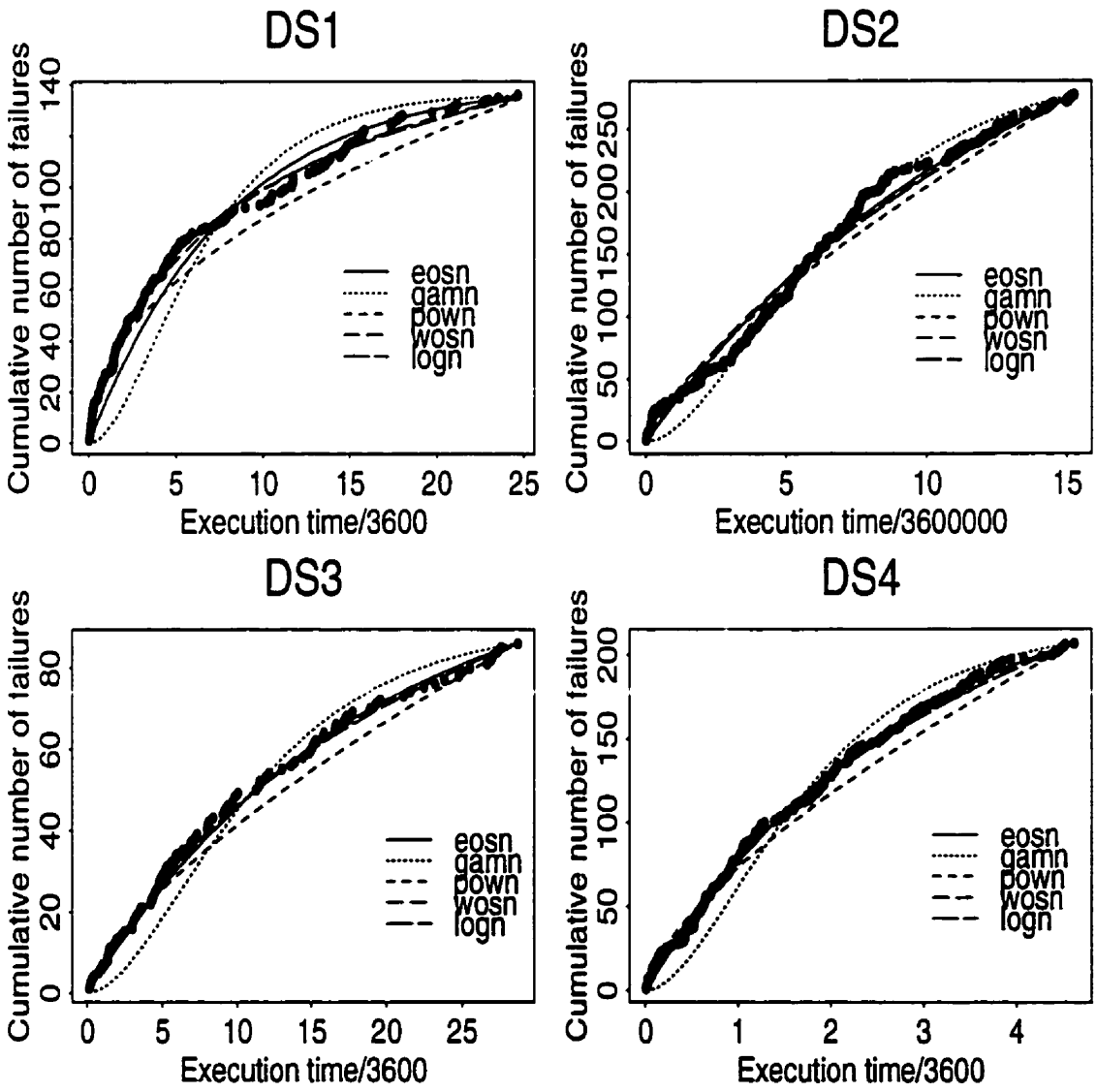


Figure 3.2: $\Lambda(t;\hat{\eta})$ versus t for specific models in Table 1.1 and DS1-4. Time is measured in hours.



time. Another informal method that is useful is to graph the generalized residuals.

$$\hat{e}_i = \int_{t_{i-1}}^{t_i} \lambda(u;\hat{\eta})\,du, \quad \text{for } i = 1,\ldots,n.$$

If the assumed NHPP model is satisfactory, the residuals should roughly look like a sample of i. i. d. standard exponential random variables. Common graphical checks include exponential probability plots of the $\hat{e}_i$ and plots of $\hat{e}_i$ vs. i or $log\,\hat{e}_i$ vs. i.

A formal technique that may be used for testing the goodness of fit of any parametric NHPP model is to split the interval $(0, T1]$ into a number of nonoverlapping intervals $A_k = (a_{k-1}, a_k]$ in order to check the assumption that $N(A_k)$ is distributed as a Poisson random variable with mean $[\Lambda(a_k;\eta) - \Lambda(a_{k-1};\eta)]$ using an appropriate LR-based or Pearson's χ^2 test. Unfortunately, choosing the most appropriate $A_k's$ is problematic. In addition, this type of test does not use all of the information when ungrouped data is available.

Another set of formal techniques available for testing the parametric form of all NHPP models involves transforming the original data $t_1 < t_2 < \ldots < t_n$ before applying an empirical cumulative distribution function (e. c. d. f.)-type statistic, such as the Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling, to test the goodness of fit of the specified model against a general alternative. For instance, one type involves transforming the original data into $u_i = \frac{\Lambda_0(t_i;\beta)}{\Lambda_0(T1;\beta)}$ (assuming β is known for now). If the model is correct, the $u_i's$ are order statistics from an uniform $(0,1)$ distribution. This transformation was used in [11], [28], [87], [92], and [93] to test the POWN model. The second type involves transforming further the $u_i's$ into $w_i's$, where $w_i = \frac{-log(1-u_i)}{\beta}$. If the model is correct, the $u_i's$ are order statistics from

an exponential distribution with parameter β. This transformation was used in [63], [71], [79], [92], and [93] to test the POWN model, and in [17] to test a general NHPP model. Since β is unknown and must be estimated, the above transformed variables are modified by replacing β with an estimate, such as its m.l.e. or its unbiased estimator (available for POWN model). In general, when the parameter values are replaced by their estimates, the distribution of the resulting statistic still depends on the population values; however, for the POWN model, the nature of the closed form m.l.e. solution for the parameters enables one to obtain *exact* distributional results on the estimates. For other NHPP models, we are not able to determine exact distributional results on the parameters because we lack a closed form solution under these models. Consequently, we are unable to obtain exact tests even though we are still able to transform the data.

Since all NHPP models satisfy Aalen's multiplicative intensity model of the form $\lambda(t) = \alpha(t)\, Y(t)$, for a deterministic function $\alpha(t)$ and an observable indicator process $Y(t) = I(t \leq T1)$, informal and formal asymptotically distribution-free goodness of fit methods available for testing the parametric form of this general model may also be considered for use in the NHPP special case. Many of these methods are discussed in [5, chapter VI]. For instance, we may use the total time on test (TTT) plot and related cumulative TTT statistic defined as $N(T)$ times the area under the TTT plot. The TTT plot was introduced to failure time data by [14] and extensions were made to Aalen's multiplicative intensity counting processes by [1] and [45]. An application of this method to the POWN model is given in [63]. Another possibility is to consider goodness-of-fit statistics based on a distance

measure between the Nelson-Aalen estimator $\hat{A}(t) = \int_0^t \{J(s)/Y(s)\}\, dN(s)$ and the estimated parametric function $A^*(t;\hat{\boldsymbol{\theta}}) = \int_0^t J(s)\alpha(s;\hat{\boldsymbol{\theta}})\, ds$, for the indicator function $J(t) = I(Y(t) > 0)$. First considered by [52], these goodness of fit statistics are of the form

$$\hat{A}(t) - A^*(t;\hat{\boldsymbol{\theta}}) \;=\; \int_0^t \frac{J(s)}{Y(s)}\{dN(s) - \alpha(s;\hat{\boldsymbol{\theta}})\, Y(s)\, ds\}.$$

In addition, we may use an idea by [59] and [60] to transform the above goodness of fit process into a process having a nice limiting distribution, such as the process obtained by asymptotically replacing the process with the process minus its mean value function (compensator). This idea was used to test the Jelinski-Moranda (EOS) model by [44].

Another formal goodness of fit technique that is useful in some situations is to nest the model of interest in a larger parametric family and use any of the three asymptotically equivalent test statistics (LR, score or Wald) to check the significance of the reduction to the actual model of interest. For instance, the WOSN model may be used to test the EOSN model, because the EOSN model is nested within the WOSN model when $\beta_2 = 1$. Similarly, the "inflection s-shaped" (SEOSN, my notation) model proposed by [103], with

$$\lambda(t;\boldsymbol{\eta}) \;=\; \frac{v\,\beta_1(1+\beta_2)\exp(-\beta_1 t)}{(1+\beta_2\exp(-\beta_1 t))^2},$$

for $v, \beta_1, \beta_2 > 0$ may also be used to test the EOSN model (when $\beta_2 = 0$). Although we have not seen it used in the software reliability literature, another model that

may be useful in this situation is the one introduced by [69] (LEEN, my notation), with

$$\lambda(t;\boldsymbol{\eta}) \;=\; \alpha_1\,\alpha_2\,t^{\alpha_2-1}\exp(-\alpha_3\,t),$$

for $\alpha_1, \alpha_2 > 0$ and $-\infty < \alpha_3 < \infty$. The LEEN model reduces to the EOSN (when $\alpha_2 = 1$), GAMN (when $\alpha_2 = 2$), and POWN (when $\alpha_3 = 0$). In contrast, the "k-stage Erlangian" model proposed by [61] with

$$\lambda(t;\boldsymbol{\eta}) \;=\; v\,\beta\,\exp(-\beta\,t)\left(\frac{(\beta\,t)^{k-1}}{(k-1)!}\right),$$

for $v, \beta > 0$ and k a small integer, can *not* be used to test the EOSN and GAMN models in this manner, even though the model reduces to EOSN (when $k = 0$) and the GAMN (when $k = 1$). This is due to the fact that k is assumed *known* before estimation begins ([61],[62],[105]). In the software reliability context, either of the WOSN, SEOSN or LEEN models may be used to test smaller models. One concern is that they may be too restrictive in that all three models nest only one model (EOSN) that seems to fit software reliability data overall. Although the GAMN model is an adequate fit in some situations, the POWN model is generally inadequate in the software reliability context. Consequently, it is of interest to obtain an NHPP model that would nest more potential models applicable in software reliability applications.

In this chapter, we propose two new methods for testing the goodness of fit of various NHPP models. In Section 3.2, we propose a goodness of fit approach

applicable for testing any NHPP model having an intensity of the form $\lambda(t;\eta) = v\,\lambda_0(t;\beta)$. As an example, we conducted a small simulation study using the EOSN model to calculate upper tail critical values. In Section 3.3, we propose a four-parameter NHPP model that nests a larger subset of models — the WOSN, LEEN and EOSN models. We discuss model characteristics and maximum likelihood estimation based on the model. We applied both goodness of fit approaches to DS1,2,4 and compared the results. Our conclusions are given in Section 3.4.

3.2 Proposed general goodness of fit method

In our situation, recall that $\{N(t): t \geq 0\}$ is assumed to be an NHPP with intensity function $\lambda(t;\eta)$. We observe a single sample path of the process over $(0, T1]$ with n error detections (events) occurring at times $t_1 < t_2 < \ldots < t_n$. We assume further that the intensity is of the form $\lambda(t;\eta) = v\,\lambda_0(t;\beta)$, for scalar $v > 0$ and vector $\beta > 0$. The corresponding mean value function is defined as $\Lambda(T1;\eta) = v\,\Lambda_0(T1;\beta)$. In the spirit of [52], we are specifically interested in the goodness of fit measure of the form

$$H_{n,\hat{\beta}} = \int_0^{T1} \{\hat{\Lambda}_{NA}(t) - \hat{v}\Lambda_0(t;\hat{\beta})\}^2 d(\hat{v}\Lambda_0(t;\hat{\beta})), \tag{3.1}$$

where, in general, $H_{n,\hat{\beta}}$ depends on n and $\hat{\beta}$; hence, the notation. As well, $\hat{\Lambda}_{NA}(t)$ is the Nelson-Aalen estimator [5], which is just $\hat{\Lambda}_{NA}(t) = N(t)$ for NHPP models. In other words, we can write $\hat{\Lambda}_{NA}(t) = n\,F_n(t)$, where $F_n(t)$ denotes the e. c. d. f, $\frac{N(t)}{n}$, for $0 \leq t \leq T1$. Combining this information with the fact that $\hat{v} = \frac{n}{\Lambda_0(T1;\hat{\beta})}$

for these models, we can rewrite equation (3.1) as

$$H_{n,\hat{\beta}} = \int_0^{T1} \left\{ n\, F_n(t) - n\frac{\Lambda_0(t;\hat{\beta})}{\Lambda_0(T1;\hat{\beta})} \right\}^2 \hat{v}\frac{d\,\Lambda_0(t;\hat{\beta})}{\Lambda_0(T1;\hat{\beta})}\,\Lambda_0(T1;\hat{\beta}).$$

Recall that the quantity $F_0(t;\beta) = \frac{\Lambda_0(t;\beta)}{\Lambda_0(T1;\beta)}$ is a cumulative distribution function defined over $(0, T1]$. With this information, we can rewrite further the above equation to obtain

$$\begin{aligned} H_{n,\hat{\beta}} &= n^2\,\hat{v}\,\Lambda_0(T1;\hat{\beta}) \int_0^{T1} \{F_n(t) - F_0(t;\hat{\beta})\}^2\, d\,F_0(t;\hat{\beta}) \\ &= n^3 \int_0^{T1} \{F_n(t) - F_0(t;\hat{\beta})\}^2\, d\,F_0(t;\hat{\beta}). \end{aligned} \tag{3.2}$$

The above expression implies that essentially $H_{n,\hat{\beta}}$ is an e. c. d. f. -type goodness of fit statistic for the distribution $F_0(t;\beta)$ that is defined over $(0, T1]$, except that there is one complication. In particular, for the NHPP time-truncated observed process, $T1$ is fixed and n is a random variable rather than a fixed sample size. Consequently, to interpret (3.2) in the usual way as an e. c. d. f. test of $F_0(t;\beta)$, we must condition on $N(T1) = n$. This implies that the size or power of this test may be calculated *conditionally* for values of n in the usual way, or *unconditionally* by averaging over the distribution of $n = N(T1)$. Furthermore, $H_{n,\hat{\beta}} = n^2\, W^2_{n,\hat{\beta}}$, where $W^2_{n,\hat{\beta}}$ is a Cramér-von Mises e. c. d. f. -type goodness of fit statistic. For evaluation of $W^2_{n,\hat{\beta}}$, we can use the well-known expression

$$W^2_{n,\hat{\beta}} = \frac{1}{12\,n} + \sum_{i=1}^{n} \left\{ F_0(t_i;\hat{\beta}) - \frac{i - 0.5}{n} \right\}^2.$$

The null hypothesis is rejected in favour of the alternative for large values of $W^2_{n,\hat{\beta}}$.

3.2.1 An application

Model and data information

We simulated critical values of $W^2_{n,\hat{\beta}}$ and $W^2_{n,\beta}$ (where we have used the assumed value for β) for the EOSN model. We calculated both expressions in order to obtain an estimate of the effect of estimating β. First, we conducted a simulation study to determine how the critical values of $W^2_{n,\beta}$ and $W^2_{n,\hat{\beta}}$ change for increasing values of n, β, and a, the proportion of the expected number of errors to be eventually detected that are in fact discovered by $T1$. In particular, we generated samples from the EOSN model with $n = 10, 20, 40, 80, 100$ and $T1$ determined such that $E\{N(0, T1)\} = v(1 - \exp\{-\beta T1\}) = av$, for $a = 0.5, 0.9$ and $\beta = 0.2, 0.5, 1$. This implies that $T1 = \frac{-\log(1-a)}{\beta}$. For example, $T1 = 3.4657, 11.5129$, when $\beta = 0.2$ and $a = 0.5, 0.9$, respectively. Similarly, $T1 = 0.6931, 2.3026$, when $\beta = 1$ and $a = 0.5, 0.9$, respectively. The simulation study was conducted in the following manner with $B = 2000$:

1. Repeat the following process $B = 2000$ times:

   (a) Simulate a sample of n uniform $(0, 1)$ random variables, u_i, and sort them.

   (b) Obtain the ordered t_i using $t_i = -\frac{log(1-u_i*(1-exp(-\beta*T1)))}{\beta}$.

   (c) Obtain the m.l.e. $\hat{\beta}$ based on this simulated sample. Note that negative values for $\hat{\beta}$ are *not* to be discarded.

(d) Obtain the $\hat{u}_i$'s using $\hat{\beta}$.

(e) Calculate $W^2_{n,\beta}$ using the u_i's and $W^2_{n,\hat{\beta}}$ using the $\hat{u}_i$'s.

2. Sort separately the sets of B W^{β}_n and $W^{\hat{\beta}}_n$ values obtained above. Obtain the 80%, 85%, 90%, 95%, 98%, 99% empirical percentiles of their e. c. d. f.

We then repeated the entire process ten times to obtain the mean and standard deviation of each of the above critical values, as was done in [87]. Next, we conducted similar simulation studies to determine the critical values of $W^2_{n,\beta}$ for a given combination of $\hat{\beta}$ and $T1$ values corresponding to those observed for data sets DS1-4. We then calculated the observed value for $W^2_{n,\hat{\beta}}$ and compared it to the appropriate critical values to determine whether the EOSN model fits that particular data set.

Results

First, we discuss the results for the larger simulation study. The 0.20, 0.15, 0.10, 0.05, 0.02, and 0.01 upper critical values of $W^2_{n,\hat{\beta}}$ and $W^2_{n,\beta}$ were obtained for each combination of n, β, and a (and hence, $T1$). The *same* critical values were generated for a given combination of n and a (and hence, $\beta T1$ combination). The calculated mean critical values of the ten sets of simulations are given in Table 3.1. The calculated standard deviation for each set of simulations are not given, but the approximate range of values is $0.0004 - 0.0035$ for the estimated critical values based on $W^2_{n,\hat{\beta}}$, and $0.0013 - 0.0155$ for those based on $W^2_{n,\beta}$. Referring to the table, we note that the corresponding critical values for $W^2_{n,\hat{\beta}}$ and $W^2_{n,\beta}$ differ substantially. It is obvious that we can not ignore the effect of estimating β. We also note that, for given a, the critical values for $W^2_{n,\hat{\beta}}$ only differ in the third decimal place as

Table 3.1: Upper critical values of $W^2_{n,\hat{\beta}}$ and $W^2_{n,\beta}$ based on EOSN.

Data attributes	0.20	0.15	0.10	0.05	0.02	0.01
			$W^2_{n,\hat{\beta}}$			
$a = 0.50$						
$n = 10$	0.097	0.109	0.126	0.156	0.194	0.225
$n = 20$	0.094	0.106	0.123	0.153	0.193	0.223
$n = 40$	0.094	0.105	0.122	0.151	0.193	0.222
$n = 80$	0.093	0.104	0.121	0.149	0.191	0.223
$n = 100$	0.093	0.104	0.121	0.151	0.190	0.219
$a = 0.90$						
$n = 10$	0.106	0.120	0.139	0.173	0.217	0.249
$n = 20$	0.104	0.118	0.137	0.172	0.218	0.254
$n = 40$	0.104	0.117	0.137	0.171	0.218	0.252
$n = 80$	0.103	0.116	0.135	0.170	0.218	0.254
$n = 100$	0.103	0.116	0.135	0.171	0.218	0.254
			$W^2_{n,\beta}$			
for all a						
$n = 10$	0.242	0.283	0.345	0.455	0.599	0.718
$n = 20$	0.244	0.285	0.352	0.459	0.603	0.738
$n = 40$	0.241	0.284	0.346	0.453	0.616	0.744
$n = 80$	0.242	0.284	0.347	0.455	0.611	0.735
$n = 100$	0.240	0.280	0.342	0.455	0.610	0.735

n increases. In other words, to two decimal places, the critical values are roughly the same for $n > 20$. A similar result was found for the calculated Cramér-von Mises e. d. f-type goodness fit statistic for the POWN model obtained in [87]. It also appears that, as a increases from 0.5 to 0.90, the 0.20 critical value increases slightly, whereas the 0.01 critical value increases relatively more.

Next, we calculated the observed $W^2_{n,\hat{\beta}}$ values based on the estimated $\hat{\beta}$ and $T1$ values for the EOSN model for DS1-4 and compared these values to the simulated

Table 3.2: Critical values of $W^2_{n,\hat{\beta}}$ and $W^2_{n,\beta}$ based on EOSN using $\hat{\beta}$ and $T1$ from DS1-4.

$W^2_{n,\hat{\beta}}$	0.80	0.85	0.90	0.95	0.98	0.99
DS1	0.112	0.126	0.145	0.185	0.238	0.286
DS2	0.095	0.108	0.125	0.155	0.195	0.227
DS3	0.097	0.110	0.129	0.159	0.202	0.236
DS4	0.097	0.110	0.130	0.163	0.209	0.239

critical values based on the same values for $\hat{\beta}$ and $T1$. The simulated critical values of $W^2_{n,\hat{\beta}}$ are given in Table 3.2 and the observed values for $W^2_{n,\hat{\beta}}$ are given in Table 3.3. We reject the null hypothesis that the EOSN model is an appropriate model at the α level if the observed statistic is greater than the simulated $1-\alpha$ critical value. For example, the observed value for $W^2_{n,\hat{\beta}}$ is 0.423 for DS1. Since this value is greater than the 95% simulated critical value, 0.185, we have strong evidence to reject the null hypothesis that the EOSN model is an adequate model. After performing the same tests for the other data sets, we observe that we also reject the EOSN model for DS2, but we do not have evidence to reject the EOSN model for DS3 and DS4.

Overall, we believe the $W^2_{n,\hat{\beta}}$ statistic is useful for testing the goodness of fit of the EOSN model. The parameter β is a scale parameter and it may be shown that the critical values of the distribution of $W^2_{n,\hat{\beta}}$ depend on β and $T1$ only through the combination of $\beta T1$ (and hence, a). Our results indicate that the difference for varying $\beta T1$ (or a) is in the first decimal place - even for large n. Consequently, it seems that the asymptotic distribution of the $W^2_{n,\hat{\beta}}$ statistic depends slightly on the $\beta T1$ combination. Although we chose to calculate the critical values of $W^2_{n,\hat{\beta}}$

Table 3.3: Observed values of $W^2_{n,\hat{\beta}}$ based on EOSN for DS1-4.

DS1	*DS2*	*DS3*	*DS4*
0.423	0.318	0.055	0.050

conditional on $N(T1) = n$, unconditional results may be obtained by averaging over the distribution of $n = N(T1)$. Since our results indicate that the critical values are essentially the same (up to two decimal places) for $n > 20$, we believe that the averaged unconditional results will be similar to the conditional results. Further discussion is provided in Section 3.4.

3.3 Nested models

To obtain an NHPP model that nests the LEEN and WOSN models, we compared their corresponding intensity functions. Recall that for the WOSN model, $\lambda(t;\eta) = v\beta_1\beta_2\, t^{\beta_2-1}\, \exp(-\beta_1\, t^{\beta_2})$, and, for the LEEN model $\lambda(t;\eta) = \alpha_1\, \alpha_2\, t^{\alpha_2-1}\, \exp(-\alpha_3\, t)$. Note that the essential difference is in the exponents. The WOSN model has t raised to a power in the exponent, while the LEEN model does not. Thus, we propose an NHPP model with intensity $\lambda(t;\eta) = \alpha_1\, \alpha_2\, t^{\alpha_2-1}\, \exp(-\alpha_3\, t^{\alpha_4})$, for $\alpha_1,\alpha_2,\alpha_3,\alpha_4 > 0$. The corresponding mean value function is

$$\Lambda(t;\eta) \;=\; \begin{cases} \alpha_1\, t^{\alpha_2} & \text{if } \alpha_3 = 0 \\ \alpha_1(\frac{\alpha_2}{\alpha_4})\alpha_3^{(-\frac{\alpha_2}{\alpha_4})} \int_0^{\alpha_3\, t^{\alpha_4}} u^{(\frac{\alpha_2}{\alpha_4}-1)} \exp(-u)\, du & \text{otherwise} \end{cases}$$

First, we notice that when $\alpha_3 \approx 0$, regardless of the value of α_4, the exponent term in $\lambda(t;\eta)$ is essentially equal to one; hence, we obtain the POWN model. Furthermore,

this new general model "NEWN" reduces to the LEEN model when $\alpha_4 = 1$, to the WOSN model when $\alpha_2 = \alpha_4$, and to the EOSN model when $\alpha_2 = \alpha_4 = 1$.

Next, we note that the above model is **not** of the form $\lambda(t;\boldsymbol{\eta}) = v\,\lambda_0(t;\boldsymbol{\beta})$ with corresponding mean value function $\Lambda(T1;\boldsymbol{\eta}) = v\,\Lambda_0(T1;\boldsymbol{\beta})$, for scalar $v > 0$ and vector $\boldsymbol{\beta} > 0$. To obtain a reparameterization of the model that will have an intensity of this form we altered the mean value function, and then obtained the corresponding intensity from that. In particular, we note that, when $\alpha_3 \neq 0$, we can write $\Lambda(t;\boldsymbol{\eta}) = \alpha_1\,\gamma\,\alpha_3^{-\gamma}\Gamma(\gamma) \cdot I\,G(\alpha_3\,t^{\alpha_4},\gamma)$, where $\gamma = \frac{\alpha_2}{\alpha_4}$ and $I\,G(x,d) = \frac{1}{\Gamma(d)}\int_0^x u^{d-1}\,e^{-u}\,du$ is the incomplete gamma function. Since the $IG(\cdot)$ function is of the form of a truncated cumulative distribution function, we need to set $v = \alpha_1\,\gamma\,\alpha_3^{-\gamma}\,\Gamma(\gamma)$ to obtain $\Lambda(t;\boldsymbol{\eta}) = v\,I\,G(\alpha_3\,t^{\alpha_4},\gamma)$. The resulting mean value function is of the form $\Lambda(t;\boldsymbol{\eta}) = v\,\Lambda_0(t;\boldsymbol{\beta})$. The corresponding intensity is then

$$\lambda(t;\boldsymbol{\eta}) \;=\; \frac{v\,\alpha_4\,\alpha_3^{\gamma}}{\Gamma(\gamma)}\,t^{\gamma\alpha_4-1}\exp(-\alpha_3\,t^{\alpha_4}),$$

for $v, \gamma, \alpha_3, \alpha_4 > 0$. Under this new parameterization, $\Lambda(t;\boldsymbol{\eta})$ looks like v times the c. d. f. for a lifetime model where T^{α_4} is distributed as a two parameter Gamma with shape parameter γ and scale parameter α_3^{-1}. When $\gamma = 1$, the NEWN model reduces to the WOSN model. In this case, T^{α_4} is distributed as a two parameter Gamma with shape parameter one and scale parameter α_3^{-1}. When $\alpha_4 = 1$, the NEWN model reduces to the LEEN model. In this case, $\Lambda(t;\boldsymbol{\eta})$ looks like v times the c. d. f. for a lifetime model where T is distributed as a two parameter Gamma with shape parameter γ and scale parameter α_3^{-1}. When $\gamma \longrightarrow \infty$, the shape of the p. d. f. for this distribution (and hence, the shape of $\lambda_0(t;\boldsymbol{\eta})$) becomes similar

to the normal p. d. f. As the p. d. f. tends to that of the normal, the resulting c. d. f. (and hence, $\Lambda_0(t;\boldsymbol{\eta})$) will become s-shaped. When $\gamma \longrightarrow \infty$ and $\alpha_3^{-1} \longrightarrow \infty$, the p. d. f. becomes similar to a very spread out normal density. When $\gamma = \alpha_4 = 1$, the NEWN model reduces to the EOSN model.

3.3.1 Maximum likelihood estimation

For this model, we have $\lambda(t;\boldsymbol{\eta}) = \frac{v\alpha_4\alpha_3^{\gamma}}{\Gamma(\gamma)} t^{\gamma\alpha_4-1}\exp(-\alpha_3 t^{\alpha_4})$. In this case (2.2) is $\hat{v} = \frac{n}{IG(\alpha_3 t^{\alpha_4},\gamma)}$. After plugging in the required derivatives into (2.3), and rearranging some of the terms, we obtain

$$\begin{aligned}
\frac{\int_0^{\alpha_3 T1^{\alpha_4}} u^{\gamma-1} log(u)\exp(-u)\,du}{IG(\alpha_3 T1^{\alpha_4},\gamma)} &= log(\alpha_3) + \frac{\alpha_4\sum_{i=1}^n log(t_i)}{n} \\
(\alpha_3 T1^{\alpha_4})^{\gamma}\exp(-\alpha_3 T1^{\alpha_4}) &= \gamma - \alpha_3\frac{\sum_{i=1}^n t_i^{\alpha_4}}{n} \qquad (3.3)\\
(\alpha_3 T1^{\alpha_4})^{\gamma}\exp(-\alpha_3 T1^{\alpha_4}) &= \frac{1}{\alpha_4 log(T1)} + \frac{\gamma}{n\,log(T1)}\sum_{i=1}^n log(t_i) \\
&\quad -\alpha_3\frac{1}{n\,log(T1)}\sum_{i=1}^n log(t_i)t_i^{\alpha_4}
\end{aligned}$$

Since the left-hand side of the two latter equations are equal, we can equate and rearrange these equations to obtain

$$\alpha_3 = \frac{n + \alpha_4\gamma\sum_{i=1}^n log(t_i/T1)}{\alpha_4\sum_{i=1}^n t_i^{\alpha_4} log(t_i/T1)}. \qquad (3.4)$$

Equation (3.4) and any two equations in (3.3) can then be used to solve for the m. l. e.'s of γ, α_3, and α_4.

Figure 3.3: Profile LR contour plots of α_1, α_2, α_3, and α_4 for the NEWN model and DS1.

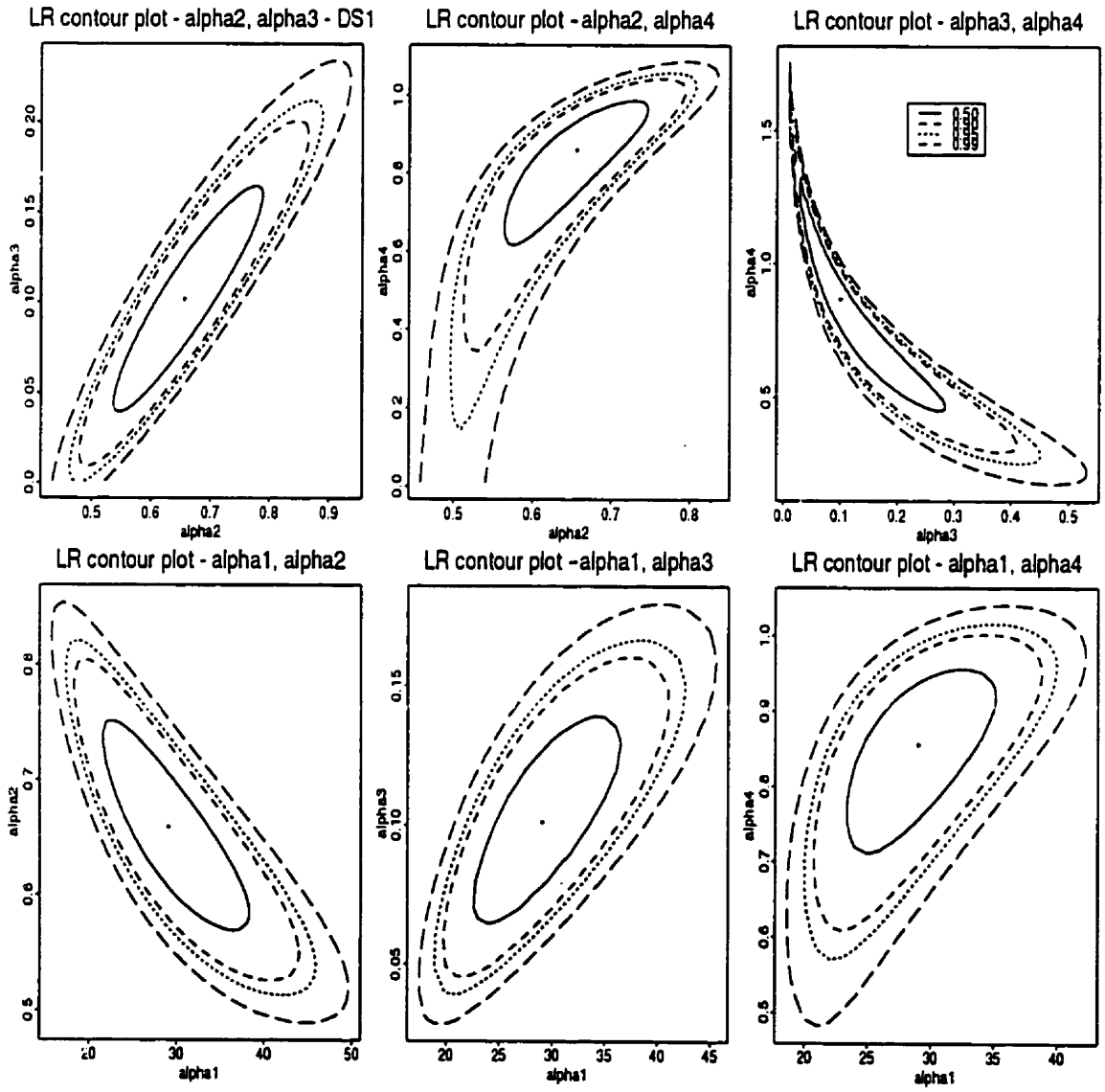


Figure 3.4: Profile LR contour plots of α_1, α_2, α_3, and α_4 for the NEWN model under the original parameterization and DS4.

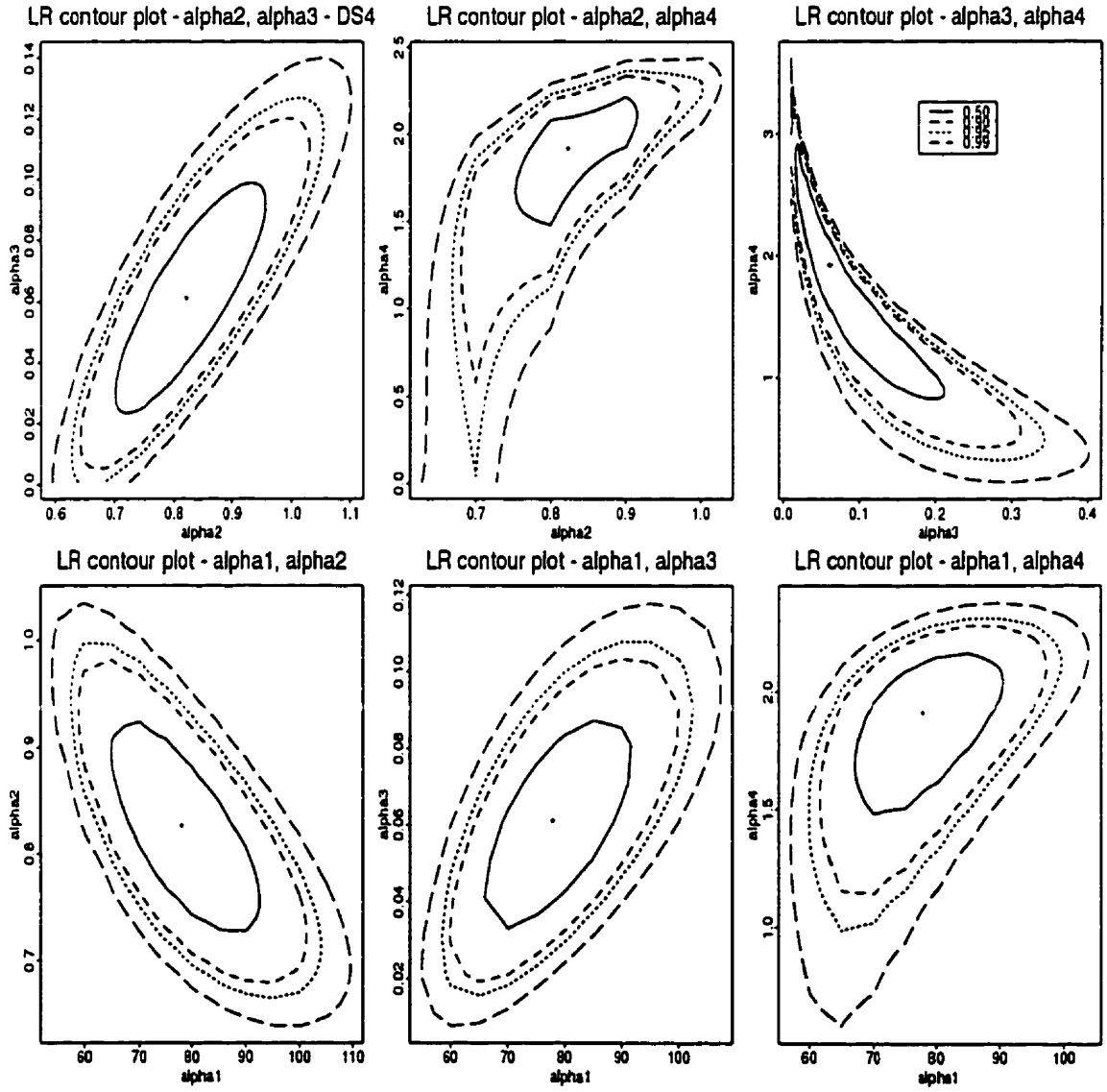


Figure 3.5: Profile LR contour plots of v, γ, α_3, and α_4 for the NEWN model and DS1.

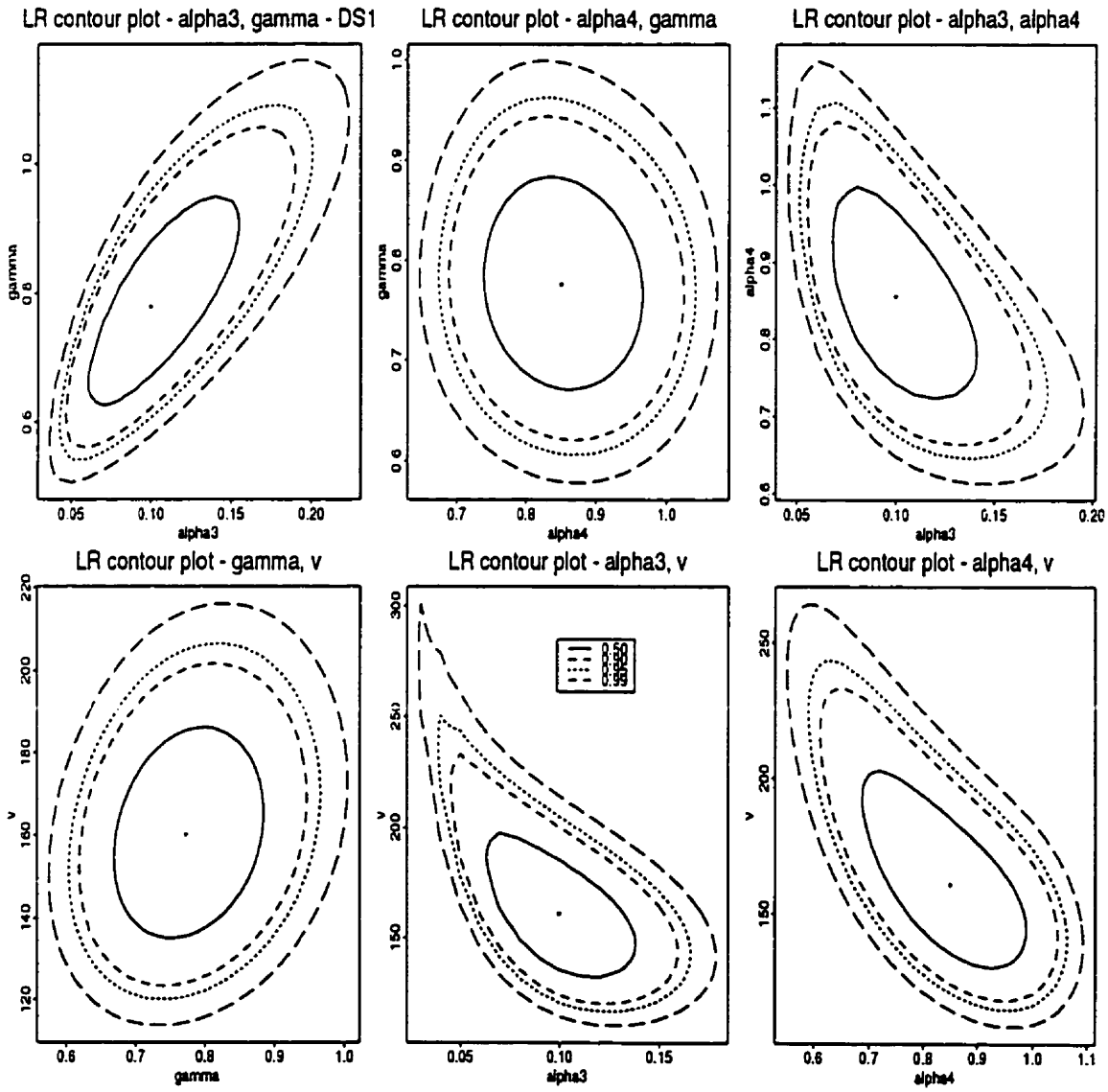


Figure 3.6: Profile LR contour plots of v, γ, α_3, and α_4 for the NEWN model and DS4.

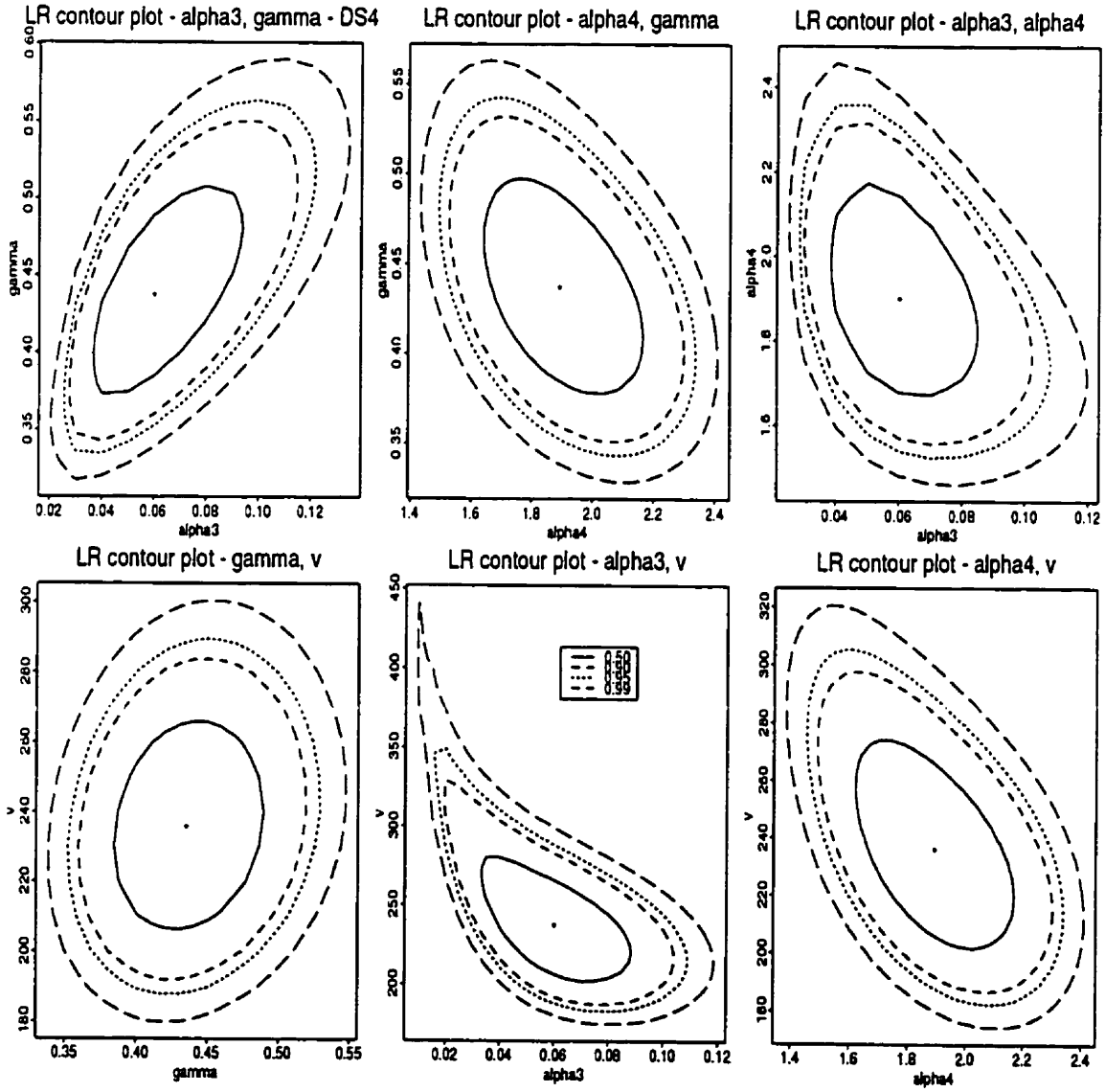


Figure 3.7: Profile likelihood plots of α_4 and γ for DS1,DS4.

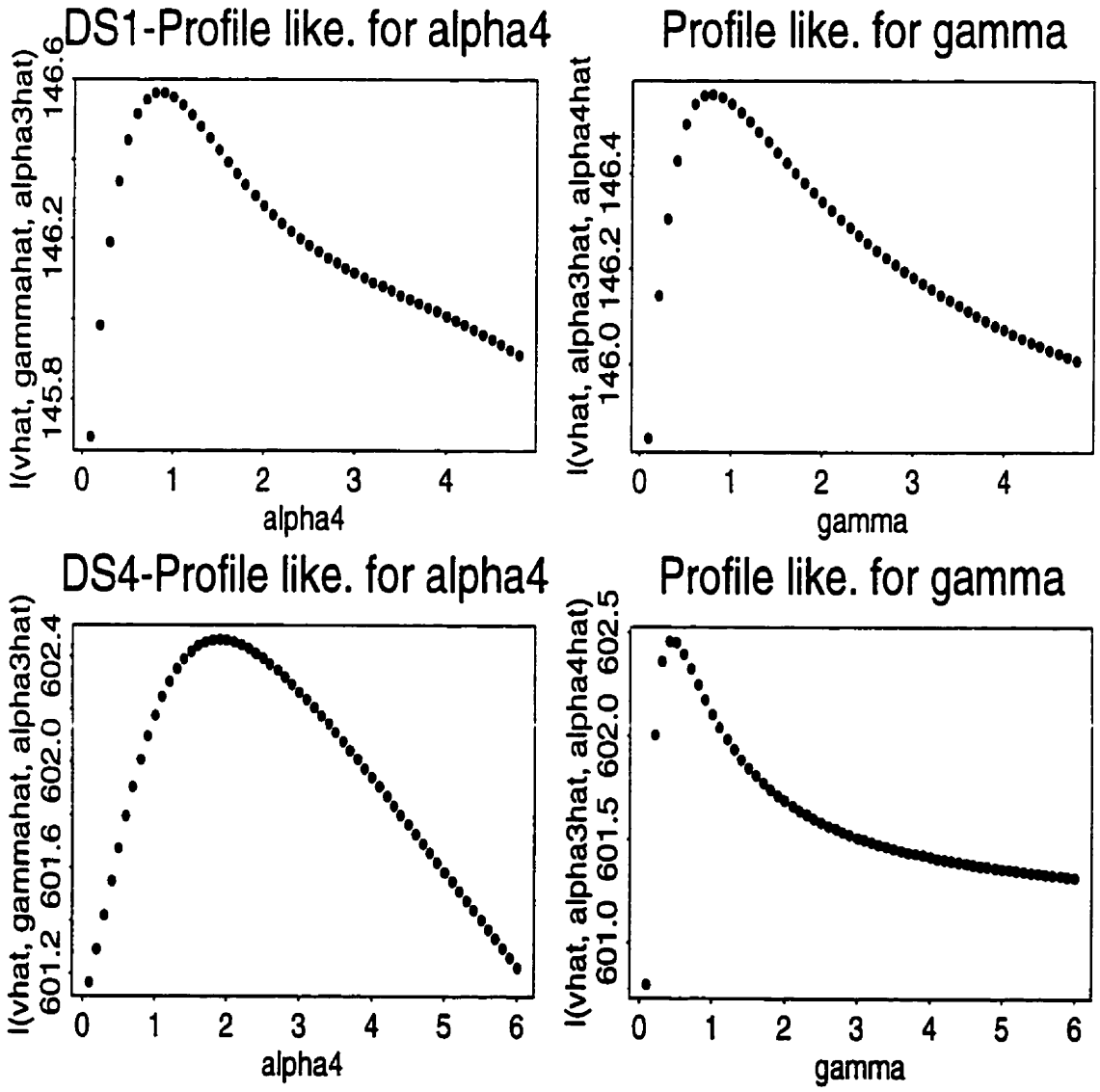
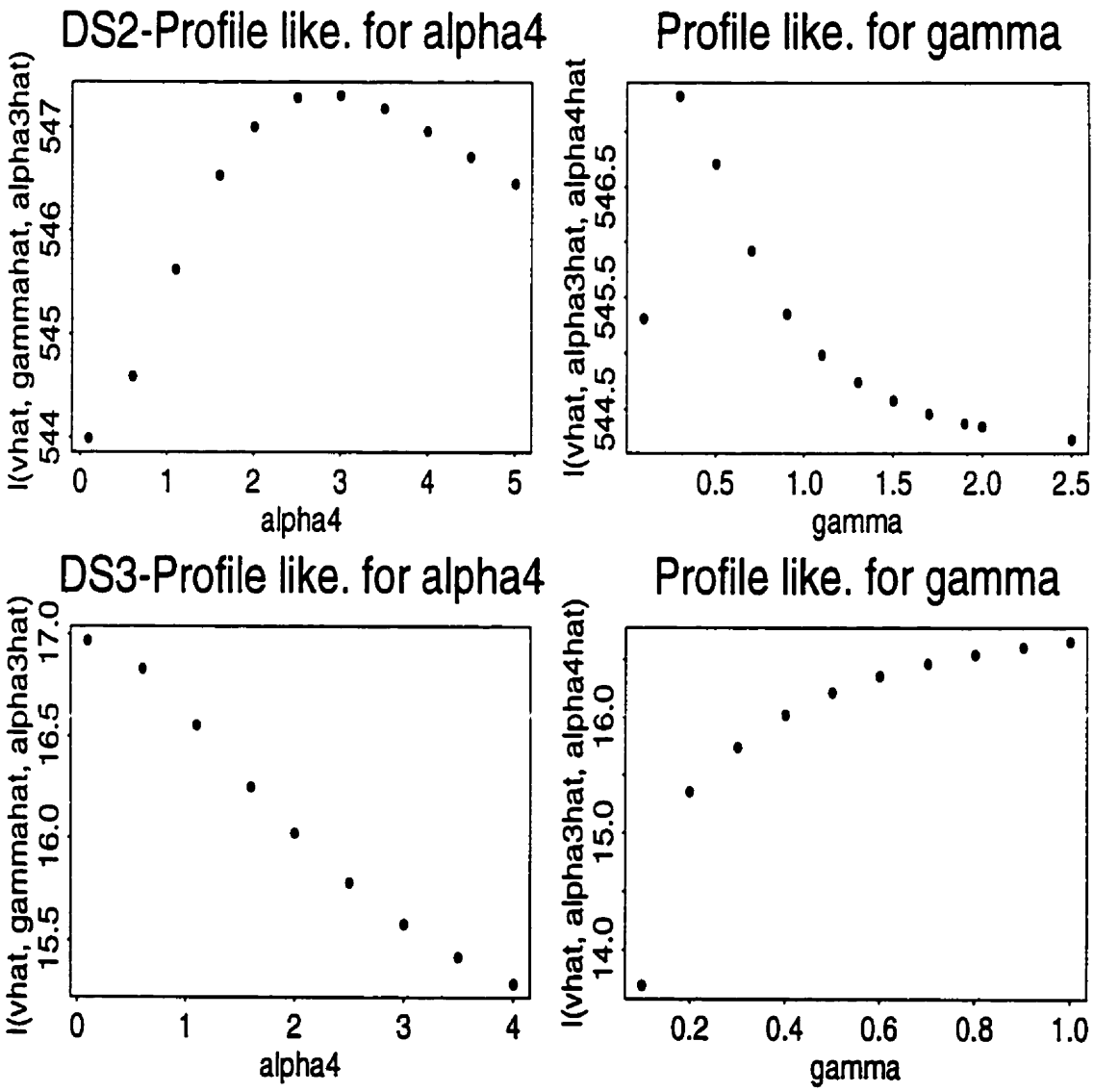


Figure 3.8: Profile likelihood plots of α_4 and γ for DS2,DS3.



It is also of interest to determine which parameterization is best to use for the NEWN model. We fitted the NEWN model to DS1-4 using both parameterizations. We were not able to obtain m.l.e.'s for DS3 (see discussion below). Profile likelihood ratio contour plots of α_1, α_2, α_3, and α_4 under the original parameterization for DS1,4 are given in Figures 3.3 and 3.4, respectively. The corresponding plots of v, γ, α_3, and α_4 under the reparameterized model are given in Figures 3.5 and 3.6. A comparison of the plots shows that the reparameterized model is preferable because the two-by-two contours are less banana-shaped than that for the original parameterization. Profile likelihood plots of α_4 and γ for DS1-4 are given in Figures 3.7 and 3.8. For instance, the profile likelihood for α_4 was found by 1) setting a particular value for α_4 and maximizing the log-likelihood over the other parameters, and; 2) plugging in the particular value for α_4 and the maximized values for the other parameters into the log-likelihood. The process is repeated for various values of α_4. These plots indicate that there is only one maximum for α_4 and γ. as a function of the other parameters. Unfortunately, in some situations (such as for DS3), α_4 is maximized at zero. This implies further that $\gamma \longrightarrow \infty$ because $\gamma = \frac{\alpha_2}{\alpha_4}$ in this parameterization. The fact that these plots indicate that one parameter is tending to zero and another parameter is tending to infinity for this data set may explain the reason we were not able to obtain a m.l.e. for the NEWN model for DS3. In general, further investigation is needed to obtain a good feel for the case where we can not obtain m.l.e.'s.

3.3.2 An application to goodness of fit

We fitted the WOSN, SEOSN, LEEN and NEWN models to DS1-4 and tested their submodels using the *generalized* likelihood ratio test statistic, $LR_g = -2[l_R(\hat{\eta}) - l_U(\hat{\eta})]/r$, where l_R refers to the restricted likelihood under the null hypothesis, l_U refers to the unrestricted likelihood under the alternative hypothesis, and r is the number of restrictions made on the parameters. Under the null hypothesis, the LR_g statistic is asymptotically distributed as a χ^2_r random variable. For comparative purposes, we also calculated the Akaike information criterion (AIC) for each data set. Although it has been used in many other fields, this model selection criterion has only been recently used for selecting NHPP models by [62]. It is defined to be $2\,l(\hat{\eta}) - 2\,(J+1)$, where $l(\cdot)$ is the log-likelihood evaluated at its m.l.e. and $J+1$ are the number of parameters in η. An advantage of using the AIC is that the models do not need to be subsets of one another, whereas for the generalized LR statistic, the model under the null hypothesis must be a subset of the model given under the alternative.

Results

The observed generalized LR statistic and corresponding p-value (given in parentheses) are given in Table 3.4. The starred values are statistically significant at $\alpha = 0.05$. No results are available for testing the NEWN model with its subsets for DS3 because we were not able to obtain m.l.e's in this case. In other words, we reject the null hypothesis that the nested (restricted) model is appropriate when the observed p-value is less than $\alpha = 0.05$. Similar results were obtained for test-

Figure 3.9: $\Lambda(t;\hat{\eta})$ versus t for the NEWN, LEEN, and WOSN models and DS1-4.

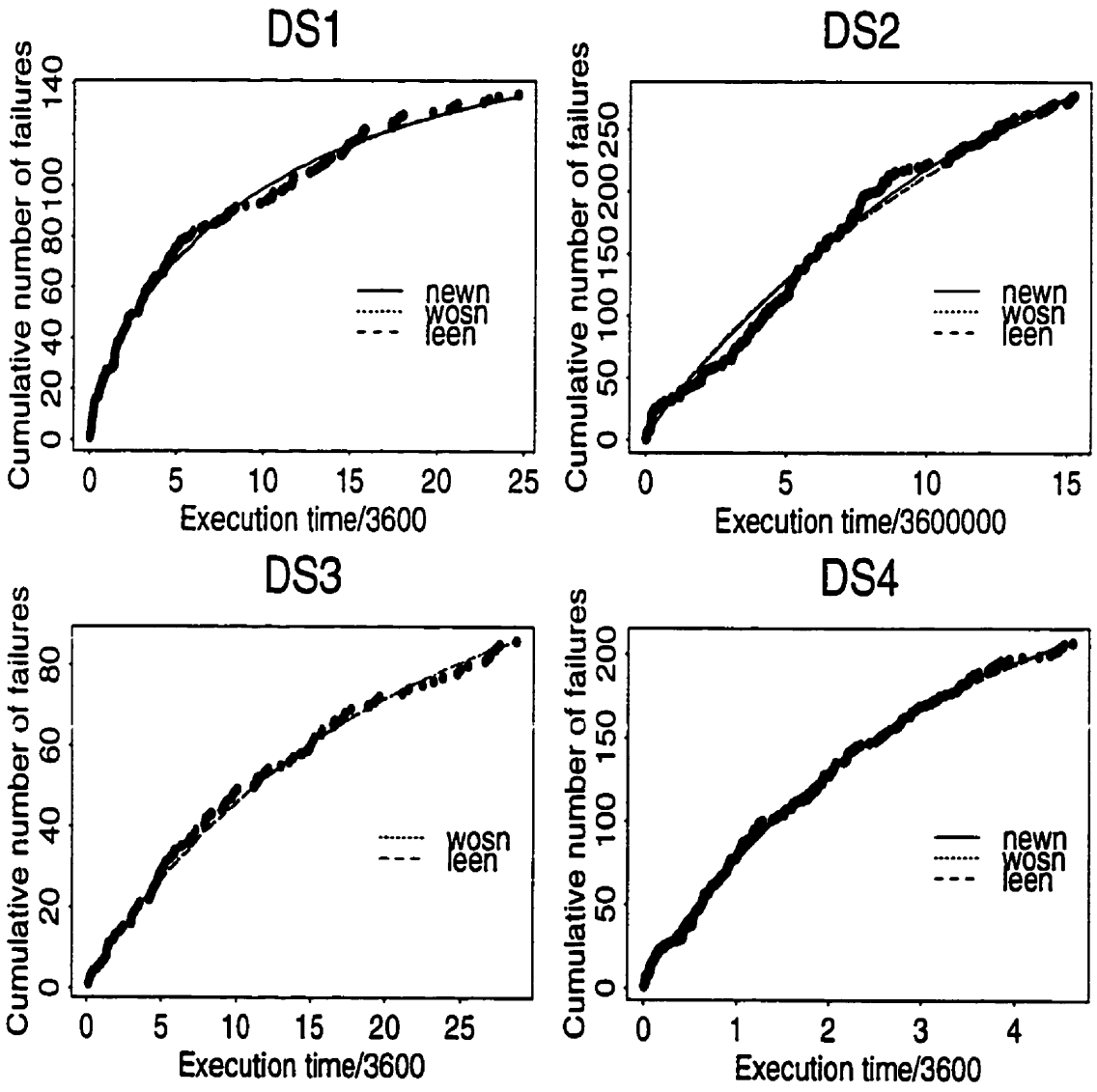


Table 3.4: A comparison of the NEWN, LEEN and WOSN models with their subsets for DS1-4.

Model	*DS1*	*DS2*	*DS3*	*DS4*
WOSN vs EOSN	17.45 (0.00)*	1.58 (0.21)	0.24 (0.62)	0.92 (0.34)
SEOSN vs EOSN	8.30 (0.00)*	1.32 (0.25)	0.57 (0.45)	0.07 (0.78)
LEEN vs EOSN	16.51 (0.00)*	2.11 (0.15)	0.17 (0.68)	1.07 (0.30)
LEEN vs POWN	9.493 (0.00)*	3.68 (0.05)	3.73 (0.05)	8.49 (0.00)*
LEEN vs GAMN	137.25 (0.00)*	92.81 (0.00)*	25.45 (0.00)*	70.47 (0.00)*
NEWN vs WOSN	0.04 (0.84)	4.35 (0.04)*	—	0.72 (0.40)
NEWN vs LEEN	0.03 (0.87)	3.82 (0.05)	—	0.57 (0.45)
NEWN vs EOSN	16.54 (0.00)*	5.93 (0.05)	—	1.64 (0.44)
NEWN vs GAMN	137.27 (0.00)*	96.63 (0.00)*	—	71.04 (0.00)*

ing the EOSN model against the alternative WOSN, SEOSN, LEEN and NEWN models. Although we can reject it for DS1, we can not reject the EOSN model in favour of the alternative model (WOSN, SEOSN, LEEN or NEWN) for DS2-4. As indicated by Figure 3.2, the GAMN model is rejected clearly in favour of the LEEN and NEWN alternatives for all data sets. Although it is rejected clearly for DS1 and DS4, we have some reservations about rejecting the POWN model in favour of the LEEN model for DS2 and DS3 since the observed p-value, 0.05, is on the borderline. This result is unexpected because Figure 3.2 indicates that the POWN model does not fit DS2 and DS3. Neither the WOSN nor the LEEN model was rejected in favour of the NEWN model for DS1 and DS4. On the other hand, there is slight evidence for rejecting the WOSN model and slight evidence for *not* rejecting the LEEN model in favour of the NEWN model for DS2. These results verify those indicated by Figure 3.9, a plot of the cumulative number of failures and the estimated model-based mean value function against time for the NEWN,

Table 3.5: AIC values for various models fitted to DS1-4.

Model	*DS1*	*DS2*	*DS3*	*DS4*
EOSN	272.60	1084.72	29.06	1199.29*
GAMN	151.86	994.01	3.78	1129.89
WOSN	287.09	1084.30	27.30	1198.21
POWN	279.62	1083.14	25.49	1191.86
LOGN	285.42	1082.64	29.73*	1198.61
LEEN	287.11*	1084.83	27.23	1198.35
NEWN	285.13	1086.65*	-	1196.93
SEOSN	278.89	1084.04	27.64	1197.37

LEEN and WOSN models. Although the fit is very similar for DS1 and DS4, there is a slight difference among the fit of the NEWN, LEEN and WOSN models to DS2. The generalized LR test picked up this slight difference.

The observed AIC values are given in Table 3.5. The starred values denote the model with the highest AIC and hence the "best" model for the data set according to this criterion. The table indicates that the LOGN and EOSN models are considered "best" for DS3 and DS4, whereas the LEEN and NEWN models are considered "best" for DS1 and DS2. respectively. Both the generalized LR statistic and the AIC criterion indicate a slight difference among the fit of the NEWN, LEEN and WOSN models in DS2. Although the generalized LR statistic was not able to do so, the AIC indicates a slight difference in the LEEN and WOSN models for DS1. In addition, the AIC indicates a slight differences in the LOGN and EOSN models for DS3.

3.4 Conclusions

We believe that the easiest, if not the best, informal technique for checking the validity of the parametric form of any NHPP model is to "eyeball" the plot of the cumulative number of failures and the estimated model-based mean value function against time. Any departures in the model may be detected easily from this plot.

As for formal techniques of goodness of fit, we believe $W^2_{n,\hat{\beta}}$ is useful for testing any NHPP model having an intensity of the general form $\lambda(t;\boldsymbol{\eta}) = v\,\lambda_0(t;\boldsymbol{\beta})$. In general, the critical values of its distribution will depend on the vector n, $\boldsymbol{\beta}$ and $T1$. Simulation of these critical values for given n, $\boldsymbol{\beta}$ and $T1$ is relatively easy to implement. Although we chose to calculate the critical values of $W^2_{n,\hat{\beta}}$ conditional on $N(T1) = n$, unconditional results may be obtained by averaging over the distribution of $n = N(T1)$. Since our results under the EOSN model indicate that the critical values are essentially the same (up to two decimal places) for $n > 20$, we believe that the averaged unconditional results will be similar to the conditional results. For those models which have a single β scale parameter, such as the GAMN and EOSN model, the critical values will depend on β and $T1$ only through the combination of $\beta T1$ (and hence, a). Our results for the EOSN model indicate that the difference for varying $\beta T1$ is in the first decimal place - even for large n. Consequently, it seems that the asymptotic distribution of the $W^2_{n,\hat{\beta}}$ statistic depends slightly on this combination $\beta T1$. We expect that for NHPP models with unknown shape parameters (such as the WOSN, POSN, LEEN and NEWN models), the distribution of the $W^2_{n,\hat{\beta}}$ statistic will depend on the underlying true values of these parameters (even in the asymptotic case). This makes sense

given the known e. c. d. f. goodness of fit results with unknown shape parameters for the censored and uncensored data situations, as discussed at length in [31]. We also believe that the above conditional results will hold for testing the goodness of fit of truncated data from a particular distribution.

We also suggest nesting specific models of interest within larger families in order to test whether the submodel is an appropriate model. The WOSN, SEOSN, and the LEEN models are useful in these situations. The NEWN model may also be useful, though we do not yet have a good feel for the case where we can not obtain m. l. e.'s. In addition, since the NEWN model is a four-parameter model, we expect that a large amount of data is required to differentiate among it, the LEEN, and the WOSN models. For these models, we expect further that $T1$ must also be large enough so that a substantial amount of the total expected number of failures will have already been observed by $T1$.

Chapter 4

Prediction

4.1 Introduction

As we argued in Section 1.4, we are specifically interested in obtaining interval predictors for $N2 = N(T1, T2]$, the number of events in the future time interval $(T1, T2]$ based on the observed data up to time $T1$, for the general NHPP model under the time-truncated sampling scheme. In addition, for the *finite* subfamily of NHPP models, we are interested in assessing the effect of data accumulation on the prediction of the number of remaining events to be eventually observed given data has been observed up to time $T1$. This quantity is $N3 = N(T1, \infty]$, where $N3$ is $N2$ with $T2 \equiv \infty$. In general, there are Bayesian and frequentist approaches to prediction. In this section, we address separately each approach. We argue that there is a need for our frequentist-based approach to predicting $N2$ and $N3$ for the general NHPP family of models. We follow with a chapter outline.

4.1.1 Bayesian approach

In principle, the Bayesian approach to prediction, regardless of whether an underlying NHPP model is used, is well-defined and relatively straightforward. To obtain the appropriate predictive distribution for $N2$ and $N3$ we need to make some assumptions about the prior density of the parameters as well as about the model. First, as in the frequentist scenario, we assume a particular NHPP family model for the data. We denote the general likelihood given by equation (1.1) as $L(D_{0,T1}|\boldsymbol{\eta})$, where the data set $D_{0,T1}$ consists of the data observed in $(0, T1]$. Second, we formulate a joint prior density for the parameters, $p(\boldsymbol{\eta})$. Once the likelihood and the prior have been specified, the posterior density, $p(\boldsymbol{\eta}|D_{0,T1})$, is defined as being proportional to the product of the likelihood and the joint prior. Furthermore, since our focus is on predicting the number of events in a future time interval, as opposed to drawing inferences about the parameters, we also need to specify $p(N2|\boldsymbol{\eta})$. In the NHPP scenario, it turns out that once the model has been chosen, $p(N2|\boldsymbol{\eta})$ is automatically defined to be the Poisson distribution with mean $u(\boldsymbol{\eta}) = \int_{T1}^{T2} \lambda(t;\boldsymbol{\eta})dt$. It immediately follows that $p(N3|\boldsymbol{\eta})$ is the Poisson distribution with mean $u(\boldsymbol{\eta}) = \int_{T1}^{\infty} \lambda(t;\boldsymbol{\eta})dt$. With all this information on hand, the resulting predictive distribution for $N2$ is given by

$$p(N2 = r|D_{0,T1}) \;=\; \int_{\boldsymbol{\eta}} p(N2 = r|\boldsymbol{\eta})\, p(\boldsymbol{\eta}|D_{0,T1})\, d\boldsymbol{\eta}, \tag{4.1}$$

for $r = 0, 1, \ldots, \infty$, and integration is over the parameter vector $\boldsymbol{\eta}$. It follows that $p(N3 = r|D_{0,T1})$ is similarly defined with $N2$ replaced with $N3$ in the above equation. Point estimates of either $N2$ or $N3$ may be obtained from (4.1) by

combining it with an appropriate loss function. Interval estimates of $N2$ or $N3$, such as highest density regions or one-sided intervals may also be obtained from (4.1).

Since the method of prediction under the Bayesian framework is relatively straightforward once a model for the data and a prior for the model parameters have been chosen, it makes sense that the Bayesian approach to prediction has been considered by many authors in the software reliability literature ([2], [13], [21], [22], [23], [24], [25], [48], [51], [64], [65], [66], [90], [91]). With the exception of [2], [21], and [22], the above authors have considered the NHPP framework. While most have looked at one NHPP family in particular (LOGN: [24], [25]; and, POWN:[13], [23], [48], [51], [66]), a few have looked at approaches for the general NHPP ([64], [65], [90], [91]). The main focus in most of the above work is on the prediction of T_{n+k}, the time until the kth future event, based on the n events observed under an event-truncated sampling scheme. The exceptions are [24], [25], [13] and [66], who have also considered prediction of $N2$ and $N3$. Since the Bayesian approach to prediction has been considered by many authors, we are not particularly interested in developing any new methodology in this situation. Instead, we provide some general comments about the choice of prior and numerical integration techniques for evaluating the integrals of interest.

Choosing a prior

There are many approaches available for choosing a prior density for the parameters, as discussed in introductory texts in Bayesian analysis (e.g. [19], [86]). Within the NHPP model framework where the intensity function is of the form

$\lambda(t;\eta) = v\,\lambda_0(t;\beta)$, for scalar $v > 0$ and vector $\beta > 0$, most authors have assumed the parameters v and β are jointly independent *apriori* and have selected marginal priors for each of the model parameters. The priors are often chosen for mathematical convenience. In fact, the uniform, gamma, or a combination of both families is often chosen for the marginal priors. The joint prior for the parameters is then simply the product of the marginal priors. The exceptions are ([24], [25]), who consider elicitation of expert opinion to obtain the joint prior, and ([89], [90], [91]), who consider a joint conjugate prior under the "Raftery" parameterization.

The likelihood based on the original parameterization of the NHPP models given in Table 1.1 is not of exponential family form; however, reparameterizing the model using either the Raftery or log-linear parameterizations given in Table 2.7 *may* yield a likelihood of exponential family form. In fact, this occurs for the EOSN, POWN, and GAMN models. For these models, a natural family of conjugate prior densities exists. In contrast, the likelihoods based on these parameterizations for the WOSN and POSN models may be considered "conditionally" exponential in that they are of exponential family form given α_3 is known. In these cases, a natural family of conjugate prior densities exists for the joint density of the parameters α_1 and α_2 given α_3. At this point, for mathematical convenience, one may assume that the joint prior for α_1 and α_2 is independent of the marginal prior for α_3; however, this may not be appropriate, as suggested by the plots given in Figure 2.8 on page 61.

Choosing a numerical integration technique

There are many techniques and approaches available for numerically approximating integrals, and it is often difficult to choose which ones are the best to use in any

given situation (e. g. [40], [41], [49]). Since the choice of the technique is dependent on the form of the integral to be evaluated, using only one particular technique for evaluating all of the required integrals is often not possible. In our situation, it so happens that we do not have too many integrals to evaluate numerically. In particular, to obtain the constant of proportionality for the posterior density $p(\eta|D_{0,T1})$ and to obtain the predictive distribution $p(N3|D_{0,T1})$, we need to integrate over the parameter vector η. It turns out that integrating over the scale parameter v (p and α_1 in the Raftery and log-linear parameterizations, respectively) may usually be done analytically, so the above two multiple integrals of dimension $J+1$ is reduced by one. For those models given in Table 1.1, the above implies that we will have at most a two-dimensional integral to approximate numerically. As recommended in [40] and [41], asymptotic methods, importance (adaptive) sampling, multiple quadrature, and subregion adaptive integration are methods of choice for lower-dimensional integrals. Simulation methods, such as Markov chain Monte Carlo (MCMC) methods, may also be used (e. g. [64] and [65]). An illustration of the Bayesian approach is given in Section 4.3.

4.1.2 Frequentist approach

In general, based on an observed random vector $\boldsymbol{Y}$, a frequentist $100(1-\alpha)\%$ *prediction interval* is defined to be a random set $(L,U) = \{L(\boldsymbol{Y}), U(\boldsymbol{Y})\}$ which covers a value of an as yet unobserved random scalar Y^* with *coverage probability* $CP(L,U) = Pr\{L \leq Y^* \leq U\} = 1-\alpha$, for prespecified *nominal* probability α. The distribution of Y^* is assumed to be related to that of $\boldsymbol{Y}$ in the sense that

both depend on an unknown parameter vector $\boldsymbol{\eta}$. There are numerous methods of constructing prediction intervals, many of which are discussed in the review papers [50] and [88]. The most common approach involves inverting *pivotal* quantities. The pivotal $Q(\boldsymbol{Y}, Y^*)$ is defined to be a function of statistics based on the past and future samples whose distribution does not depend on the parameter vector $\boldsymbol{\eta}$.

There is limited statistical literature available on the frequentist approach to prediction in the NHPP framework ([38], [70], [94]). While [38] and [70] consider the POWN model, [94] considers the EOSN model. With the exception of [94], the main focus in their work is on the prediction of T_{n+k}, the time until the kth future event, based on the n events observed under an event-truncated sampling scheme. We consider here some useful methods for predicting $N2$.

4.1.3 Outline of chapter

In Section 4.2, we discuss our strategy for constructing and assessing frequentist-based interval predictors for $N2$ (and hence, $N3$). In particular, we have obtained a set of five interval predictors based on ad hoc *approximate* pivotals obtained from well-known NHPP results. In Section 4.3, we discuss both the Bayesian and frequentist strategy for assessing the effect of data accumulation on $N3$. We follow each section with an application using the EOSN model. Our conclusions are given in Section 4.4.

4.2 Proposed frequentist interval predictors

Recall that, in our situation, $\{N(t); t \geq 0\}$ is assumed to be an NHPP with intensity function $\lambda(t;\boldsymbol{\eta})$. We observe a single sample path of the process over $(0, T]$, with n error detections (events) occurring at times $t_1 < t_2 < \ldots < t_n$. We are interested in obtaining prediction intervals for the future value of $N2$ applicable to the entire family of NHPP models. Since we have not found any exact pivotals, we use some ad hoc *approximate* pivotals obtained from well-known NHPP results to construct the predictors to be considered. First, we consider pivotals based on the distribution of $N2$. Since the counts in non-overlapping intervals are independent in a NHPP, we know that $N2$ will be independent of $N1 = N(0, T1]$. We know further that $N2$ is Poisson distributed with mean $u(\boldsymbol{\eta}) = \int_{T1}^{T2} \lambda(t;\boldsymbol{\eta})dt$. If $\boldsymbol{\eta}$ was known, the distribution of $N2$ would be known. For convenience, the approximate pivotals based on this distribution can easily be found. For instance, for large mean $u(\boldsymbol{\eta})$, it is well-known that $[N2 - u(\boldsymbol{\eta})]/[\sqrt{u(\boldsymbol{\eta})}]$ is approximately distributed as a standard normal random variable. We can then simply invert this pivotal into a prediction limit for $N2$. The crux is that $\boldsymbol{\eta}$ will not be known in practice. One will need to obtain $\hat{\boldsymbol{\eta}}$ based on the observed data, and then use it somehow to estimate $N2$.

The simplest method of incorporating the observed data would be to assume that $N2$ is Poisson distributed with mean (and hence, variance) $u(\hat{\boldsymbol{\eta}})$. One could then obtain prediction limits for $N2$ by simply inverting the standardized approximate pivotal $[N2 - u(\hat{\boldsymbol{\eta}})]/[\sqrt{u(\hat{\boldsymbol{\eta}})}]$. Unfortunately, using $u(\hat{\boldsymbol{\eta}})$ as the estimated variance for $N2 - u(\hat{\boldsymbol{\eta}})$ would underestimate the variation in the process. Note that

we can write

$$N2 - u(\hat{\eta}) \;=\; [N2 - u(\eta)] - [u(\hat{\eta}) - u(\eta)].$$

Since the two bracketed quantities are independent for NHPP models, we obtain the *exact* result

$$Var[N2 - u(\hat{\eta})] \;=\; u(\eta) + Var[u(\hat{\eta})].$$

It then follows that one way of capturing the added uncertainty in estimating η would be to "improve" the estimated variance by estimating the above quantity with

$$\hat{V}ar[N2 - u(\hat{\eta})] \;=\; u(\hat{\eta}) + \hat{V}ar_{\hat{\eta}}[u(\hat{\eta})].$$

Note that an exact expression for $Var_{\hat{\eta}}[u(\hat{\eta})]$ may be difficult to obtain in practice and approximations to its value may be more feasible. For instance, one may consider using the well-known delta method (e.g. [67, appendix C]). A reasonable standardized approximate pivotal that incorporates this "improved" estimated variance formula would then be

$$\mathcal{R}_1 \;=\; \frac{N2 - u(\hat{\eta})}{\sqrt{u(\hat{\eta}) + \hat{V}ar_{\hat{\eta}}[u(\hat{\eta})]}}.$$

A one-sided upper $(1-\alpha)100\%$ prediction interval generated from $\mathcal{R}_1$ is of the form

$$\mathcal{D}_1 = \left[0, u(\hat{\eta}) + r^*_{1,\alpha}\sqrt{u(\hat{\eta}) + \hat{V}ar_{\hat{\eta}}[u(\hat{\eta})]}\right],$$

where $r^*_{1,\alpha}$ is the $(1-\alpha)$ quantile of the distribution of $\mathcal{R}_1$.

For comparative purposes, we next considered approximate pivotals based on well-known variance stabilizing and normalizing transformations of the Poisson family. Although there are many available [57, page 162], only those that can be easily inverted into prediction intervals for $N2$ are suitable for our purposes. Two well-known variance stabilizers for the Poisson family, first given in [15], are $\sqrt{N2}$ and $\sqrt{N2+1/2}$. They are both known to be approximately normally distributed with variance $1/4$ and mean $\sqrt{u(\eta)}$ and $\sqrt{u(\eta)+1/2}$, respectively. Although they are very similar, the latter transformation is often preferable because $u(\eta)$ need not be large for this transformed variable to be approximately normally distributed. A better variance stabilizer, originally given in [6], is the transformation $\sqrt{N2+3/8}$. Regardless of the size of $u(\eta)$, it is known to be approximately normally distributed with variance $1/4$ and mean $\sqrt{u(\eta)+3/8}$. Note that all of the above variance transformations are of the form $\sqrt{N2+c}$, for $c = 0, 3/8, 1/2$, respectively.

As suggested earlier, one could simply consider using a *set* of standardized approximate pivotals based on "plugged-in" estimates of η. These would be of the form $2\left\{\sqrt{N2+c} - \sqrt{u(\hat{\eta})+c}\right\}$, for each c considered above. For the same reasons given earlier, it would make sense to incorporate the uncertainty of estimating η into the variance estimator. In this case, the set of "improved" standardized

pivotals would be of the form,

$$\begin{aligned}\mathcal{R}_2 &= \frac{\sqrt{N2+c}-\sqrt{u(\hat{\eta})+c}}{\sqrt{\hat{V}ar(\sqrt{N2+c}-\sqrt{u(\hat{\eta})+c})}} \\ &\approx \frac{\sqrt{N2+c}-\sqrt{u(\hat{\eta})+c}}{\sqrt{1/4+Var_{\hat{\eta}}[\sqrt{u(\hat{\eta})+c}]}}\end{aligned}$$

for each $c = 0, 3/8, 1/2$. As in the earlier case, $Var_{\hat{\eta}}[\sqrt{u(\hat{\eta})+c}]$ will need to be approximated. The one-sided upper $(1-\alpha) * 100\%$ prediction interval generated from $\mathcal{R}_2$ will be of the form

$$\mathcal{D}_2 = \left[0, \left\{\sqrt{u(\hat{\eta})+c} + r^*_{2,\alpha}\sqrt{1/4+\hat{V}ar_{\hat{\eta}}[\sqrt{u(\hat{\eta})+c}]}\right\}^2 - c\right].$$

where $r^*_{2,\alpha}$ is the $(1-\alpha)$ quantile of the distribution of $\mathcal{R}_2$.

An excellent normalizer for the Poisson family, originally given in [7], is $[N2]^{2/3}$. It is shown to be approximately normally distributed with mean $[u(\eta)]^{2/3}$ and variance $\frac{4}{9}[u(\eta)]^{1/3}$. An approximate pivotal that incorporates the uncertainty of estimating η is

$$\begin{aligned}\mathcal{R}_3 &= \frac{[N2]^{2/3}-[u(\hat{\eta})]^{2/3}}{\sqrt{\hat{V}ar[N2^{2/3}-u(\hat{\eta})^{2/3}]}} \\ &\approx \frac{[N2]^{2/3}-[u(\hat{\eta})]^{2/3}}{\sqrt{\frac{4}{9}[u(\hat{\eta})]^{1/3}+Var_{\hat{\eta}}\{[u(\hat{\eta})]^{2/3}\}}}.\end{aligned}$$

Once again, $Var_{\hat{\eta}}\{[u(\hat{\eta})]^{2/3}\}$ will need to be approximated. The one-sided upper

$(1-\alpha)*100\%$ prediction interval generated from $\mathcal{R}_3$ is of the form

$$\mathcal{D}_3 = \left[0, \left\{[u(\hat{\eta})]^{2/3} + r^*_{3,\alpha}\sqrt{\frac{4}{9}[u(\hat{\eta})]^{1/3} + \hat{V}ar_{\hat{\eta}}\{[u(\hat{\eta})]^{2/3}\}}\right\}^{3/2}\right],$$

where $r^*_{3,\alpha}$ is the $(1-\alpha)$ quantile of the distribution of $\mathcal{R}_3$.

4.2.1 Approximating the distributions of $\mathcal{R}_i$

As indicated in the previous section, one needs to determine the appropriate α (and possibly $1-\alpha$) quantiles of the distribution for each $\mathcal{R}_i$ in order to calculate the prediction intervals. Ideally one would like to obtain exact distributions for each $\mathcal{R}_i$; however, realistically this is not possible. In fact, there is no single distribution in general because these will depend on η. It seems more reasonable to approximate their distributions for given η. The required quantiles are then obtained from this approximated distribution. We consider two methods of approximating these distributions.

First, we consider the standard normal distribution as an approximation to those of the $\mathcal{R}_i's$. It seems reasonable to do this because of the way the $\mathcal{R}_i's$ were generated. We indicated in the previous section that if η were completely known and certain specific conditions were met (depending on the $\mathcal{R}_i$ considered, $u(\eta)$ large or n large), then the distribution of that particular $\mathcal{R}_i$ would be approximately standard normal.

Next, we consider the parametric bootstrap-t (some prefer, bootstrap critical value) method (e.g. [33], [34]), to obtain a bootstrap distribution of each $\mathcal{R}_i$. The empirical quantiles of this distribution are then used in obtaining the required

prediction intervals. Note that B, the number of independent parametric bootstrap samples to be generated, is to be at least 1000, and possibly more like 2000, as mentioned in [33] and [34, section 12.5]. We found $B = 1000$ adequate for the simulation study discussed below. In particular, the following algorithm may be used:

1. Calculate the maximum likelihood estimates, $\hat{\eta}$, based on the data set of interest.

2. Repeat the following process B times:

   (a) Simulate $N1^*$, a Poisson random variable with mean $\Lambda(T1; \hat{\eta})$.

   (b) Simulate $N2^*$, a Poisson random variable with mean $u(\hat{\eta})$.

   (c) Simulate $N1^*$ random variables with p.d.f. $f(t; \hat{\eta}) = \lambda(t; \hat{\eta})/\Lambda(T1; \hat{\eta})$, for $0 \leq t \leq T1$. Sort these values in ascending order. This now constitutes a new simulated sample of $N1^*$ ordered failure times $t_1^*, \ldots, t_{N1^*}^*$.

   (d) Calculate the maximum likelihood estimates, $\hat{\eta}^*$, based on this new simulated sample (keep those parameter estimates which are outside allowable range).

   (e) Calculate $\mathcal{R}_i^*$, for each i, based on $N2^*$ and the $\hat{\eta}^*$ for this new simulated sample.

3. Sort the B $\mathcal{R}_i^*$ values in ascending order, for each i.

4. Obtain the empirical quantile $r_{i,\alpha}^*$ based on this ordered set of $\mathcal{R}_i^*$ values.

4.2.2 Assessing interval predictors

The *conditional* coverage probability of a prediction interval (L, U) for the random quantity Y^*, based on the observed random vector $\boldsymbol{Y}$, is defined as $CP_C(L,U) = Pr\{L \leq Y^* \leq U|\boldsymbol{Y}\}$. As defined earlier, the corresponding coverage probability of the prediction interval is defined as $CP(L,U) = Pr\{L \leq Y^* \leq U\}$. Although some authors consider both (e.g. [18], [20]), we will focus on the latter measure of interval coverage. In general, to assess the accuracy of interval predictors, various summary measures of the closeness of the observed coverage probabilities to the prespecified nominal level α are calculated and comparisons are made of their expected values for different predictors of interest. The mean, standard deviation and mean square error of these coverage probabilities are examples of measures considered. Since many interval predictors may have close-to-nominal coverage, a comparison of the narrowness of the interval lengths is also important. Various measures of this narrowness can be used, such as average length. When exact expressions are difficult to obtain, approximate estimates of these measures, such as those based on asymptotic or simulation results, are compared. In our particular case, we will obtain approximate coverage probabilities and average lengths of the intervals based on simulation results for the five predictors discussed earlier. Nominal levels will be set at $\alpha = 0.05, 0.10$. Comparisons of the approximate coverage probabilities to the nominal levels and the average lengths will be made among the five interval predictors of $Y^* = N2$ for each NHPP model given in Table 1.1.

4.2.3 An application

Model and data information

We obtained one- and two-sided prediction intervals for $N2$ under the EOSN model. Rather than maximizing the log-likelihood based on the original parameterization, we maximized the log-likelihood based on the log-linear reparameterization given in Table 2.7, for reasons given in Section 2.4. We did *not* ignore those cases for which $\alpha_2 > 0$ (and hence, $\beta, v < 0$). Under this parameterization, $u(\eta) = \frac{\exp(\alpha_1)}{\alpha_2}[\exp(\alpha_2 T2) - \exp(\alpha_2 T1)]$. The variance estimates for the parameters α_1 and α_2 were obtained using the expected information matrix under this parameterization and $Var_{\hat{\eta}}[u(\hat{\eta})]$ was obtained using the delta method.

First, we obtained one- and two-sided 95% prediction intervals based on the EOSN model for DS3 and DS4. We chose these particular data sets as examples because the EOSN model fits well for both models based on evidence gathered in Chapter 3. We split the two data sets at $T1 = 15.00$ (hours) and $T1 = 2.50$ (hours), respectively. We then used the observed data over $(0, T1)$ to obtain predictions of $N2$ over $(T1, T2]$, where $T2$ is the actual truncation time for these data sets. In particular, $T2 = 28.71$ (hours) and $T2 = 4.62$ (hours) for the respective data sets. The resulting prediction intervals are given in Table 4.1. The actual $N2$ for both truncated data sets is included in all the prediction intervals. For both data sets, we see that $\mathcal{D}_{21}^*$, $\mathcal{D}_{22}^*$, and $\mathcal{D}_{23}^*$ are the same and that they differ slightly from $\mathcal{D}_1^*$ and $\mathcal{D}_3^*$. Overall, the normal-based prediction intervals are shorter than those based on the bootstrap distribution. We believe that bigger differences in the prediction intervals might be seen if $u(\hat{\eta})$ were smaller.

Next, we obtained prediction intervals for twelve simulated data sets. We gener-

Table 4.1: One- and two-sided 95% prediction intervals of N2 based on EOSN for DS3-4. Time is measured in hours.

Data			$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_{22}$	$\mathcal{D}_{23}$	$\mathcal{D}_3$
DS3							
$T1 = 15.00$ $T2 = 28.71, N2 = 26, u(\hat{\eta}) = 19$							
	two-sided						
		bootstrap	[6,52]	[6,50]	[6,50]	[6,50]	[6,51]
		normal	[0,38]	[4,43]	[4,43]	[4,43]	[3,41]
	one-sided						
		bootstrap	[0,44]	[0,43]	[0,43]	[0,43]	[0,43]
		normal	[0,35]	[0,38]	[0,38]	[0,38]	[0,37]
DS4							
$T1 = 2.50$ $T2 = 4.62, N2 = 60, u(\hat{\eta}) = 45$							
	two-sided						
		bootstrap	[22,86]	[23,85]	[23,85]	[23,85]	[23,85]
		normal	[16,74]	[20,78]	[20,78]	[20,78]	[19,76]
	one-sided						
		bootstrap	[0,76]	[0,76]	[0,76]	[0,76]	[0,76]
		normal	[0,69]	[0,72]	[0,72]	[0,72]	[0,71]

ated samples from the EOSN model with $\beta = 1$, $v = 50, 100, 200$, and for two different values of $T1$ and $T2$. The $T1$ were chosen such that roughly 50% and 90% of the expected number of errors to be eventually detected are in fact discovered by $T1$. In other words, $T1$ is determined such that $E\{N(0, T1)\} = v(1 - \exp\{-\beta T1\}) = av$, for $a = 0.5, 0.9$. Two values of $T2$, $T2 = 2\,T1$ and $T2 = 5\,T1$, were also chosen. In general, the results will depend on the product $\beta T1$ and $\beta T2$. In this case, since $T2$ is a function of $T1$ and $T1$ is a function of a and we chose $\beta = 1$, the results will hold for $\beta = 1$ and the two values of a. The twelve samples were simulated for each combination of $T1$, $T2$ and v similar to that that done in Section (2.2.2). In particular, we did the following:

1. Simulate N, a Poisson random variable with mean $v = E\{N(0, \infty)\}$.

2. Simulate N standard exponential random variables. Keep those simulated values that are $\leq T1$. Sort these remaining $N1$ values in ascending order. This now constitutes the sample of $N1$ ordered failures times $t_1^*, \cdots, t_{N1}$ for given $T1$ and v.

3. The number of left-over simulated values that are $\leq T2$ constitute the sample of $N2$ failure times.

Results of simulation study

Introduction

For each of the twelve samples that were generated, we obtained two-sided 90% and 95% prediction intervals for $N2$ based on the various $\mathcal{R}_i's$. Both the normal distribution and the parametric bootstrap-t distribution (with $B = 1000$) were used as approximates to the distributions of the $\mathcal{R}_i's$. The above process was repeated 1000 times to obtain coverage probabilities (CP's) and average lengths (AL's) of the prediction intervals. That is, prediction intervals were calculated for 1000 sets of twelve samples, and coverage probabilities and average lengths were calculated using these 1000 prediction intervals.

Comparison of the $\mathcal{R}_i^*$

We first compared the distributions of the simulated parametric bootstrap samples for each of the five $\mathcal{R}_i^*$, based on $B = 1000$ and various sets of values for v, $T1$ and $T2$. The sample mean ($\bar{x}$), variance ($\hat{\sigma}^2$), skewness ($\hat{\gamma}_1$) and kurtosis ($\hat{\gamma}_2$) coefficients calculated for each sample in one set of simulations are given in Table 4.2. Interaction plots for these quantities are also given in Figures 4.1-4.4. Normal probability plots were also obtained but are not provided in the interest of space. To begin with, the table values and plots indicate that the generated bootstrap

Table 4.2: Sample means ($\bar{x}$), variances ($\hat{\sigma}^2$), skewness ($\hat{\gamma}_1$) and kurtosis ($\hat{\gamma}_2$) coefficients for one set of bootstrap distributions based on EOSN.

Data		$\mathcal{R}_1$	$\mathcal{R}_{21}$	$\mathcal{R}_{22}$	$\mathcal{R}_{23}$	$\mathcal{R}_3$
$v = 50$						
T1=0.69 T2=2T1	$\bar{x}/\hat{\sigma}^2$	0.37/2.08	0.11/1.08	0.11/1.08	0.11/1.08	0.19/1.25
	$\hat{\gamma}_1/\hat{\gamma}_2$	2.71/14.89	0.66/0.35	0.69/0.41	0.69/0.41	1.13/1.96
T2=5T1	$\bar{x}/\hat{\sigma}^2$	0.73/4.59	0.32/1.20	0.33/1.24	0.33/1.24	0.43/1.67
	$\hat{\gamma}_1/\hat{\gamma}_2$	5.49/49.31	1.49/2.89	1.53/3.11	1.53/3.10	2.28/8.51
T1=2.30 T2=2T1	$\bar{x}/\hat{\sigma}^2$	1.33/15.60	0.56/3.84	0.57/3.89	0.57/3.90	0.75/5.64
	$\hat{\gamma}_1/\hat{\gamma}_2$	3.77/20.73	1.69/2.91	1.73/3.08	1.74/3.12	2.22/5.67
T2=5T1	$\bar{x}/\hat{\sigma}^2$	1.15/12.28	0.50/2.88	0.51/2.92	0.51/2.94	0.67/4.26
	$\hat{\gamma}_1/\hat{\gamma}_2$	4.59/30.87	1.89/4.54	1.94/4.79	1.95/4.85	2.59/8.94
$v = 100$						
T1=0.69 T2=2T1	$\bar{x}/\hat{\sigma}^2$	0.27/1.41	0.12/1.05	0.12/1.05	0.12/1.05	0.17/1.13
	$\hat{\gamma}_1/\hat{\gamma}_2$	1.45/3.15	0.53/0.37	0.54/0.39	0.55/0.40	0.82/0.95
T2=5T1	$\bar{x}/\hat{\sigma}^2$	0.52/2.15	0.29/1.14	0.29/1.15	0.29/1.15	0.36/1.36
	$\hat{\gamma}_1/\hat{\gamma}_2$	2.55/9.21	1.31/2.12	1.32/2.19	1.33/2.21	1.67/3.80
T1=2.30 T2=2T1	$\bar{x}/\hat{\sigma}^2$	0.17/1.29	0.03/1.04	0.03/1.03	0.03/1.02	0.07/1.09
	$\hat{\gamma}_1/\hat{\gamma}_2$	1.43/5.78	0.27/0.09	0.28/0.09	0.28/0.09	0.57/0.86
T2=5T1	$\bar{x}/\hat{\sigma}^2$	0.22/1.39	0.06/1.03	0.06/1.02	0.06/1.01	0.11/1.10
	$\hat{\gamma}_1/\hat{\gamma}_2$	1.81/7.19	0.50/0.48	0.50/0.50	0.50/0.50	0.85/1.60
$v = 200$						
T1=0.69 T2=2T1	$\bar{x}/\hat{\sigma}^2$	0.18/1.20	0.07/1.04	0.07/1.04	0.07/1.04	0.11/1.08
	$\hat{\gamma}_1/\hat{\gamma}_2$	0.96/1.29	0.36/0.03	0.36/0.03	0.36/0.03	0.55/0.30
T2=5T1	$\bar{x}/\hat{\sigma}^2$	0.37/1.58	0.20/1.10	0.20/1.11	0.20/1.11	0.26/1.22
	$\hat{\gamma}_1/\hat{\gamma}_2$	1.71/3.73	0.96/0.77	0.97/0.78	0.97/0.79	1.20/1.52
T1=2.30 T2=2T1	$\bar{x}/\hat{\sigma}^2$	0.12/1.19	0.03/1.07	0.03/1.06	0.03/1.06	0.06/1.10
	$\hat{\gamma}_1/\hat{\gamma}_2$	0.81/0.78	0.32/-0.08	0.32/-0.08	0.32/-0.08	0.48/0.11
T2=5T1	$\bar{x}/\hat{\sigma}^2$	0.17/1.17	0.07/1.03	0.07/1.03	0.07/1.03	0.10/1.06
	$\hat{\gamma}_1/\hat{\gamma}_2$	0.87/0.69	0.39/-0.19	0.40/-0.18	0.40/-0.18	0.55/0.02

Figure 4.1: Plots of the sample mean ($\bar{x}$) as $T2$ increases from $2 * T1$ to $5 * T1$ and v increases from 50 to 200 for one set of bootstrap distributions based on EOSN.

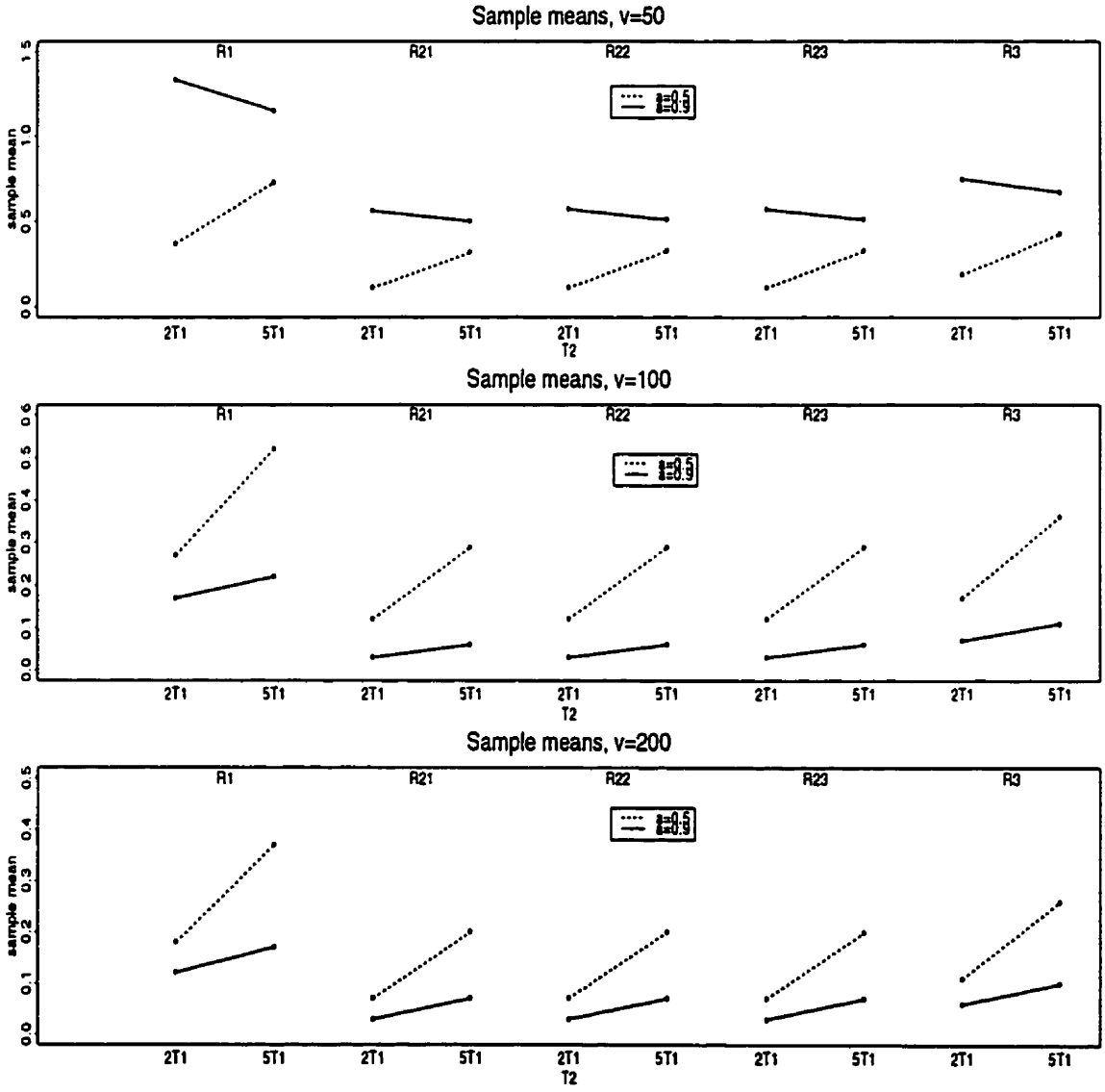


Figure 4.2: Plots of the sample variance ($\hat{\sigma}^2$) as $T2$ increases from $2 * T1$ to $5 * T1$ and v increases from 50 to 200 for one set of bootstrap distributions based on EOSN.

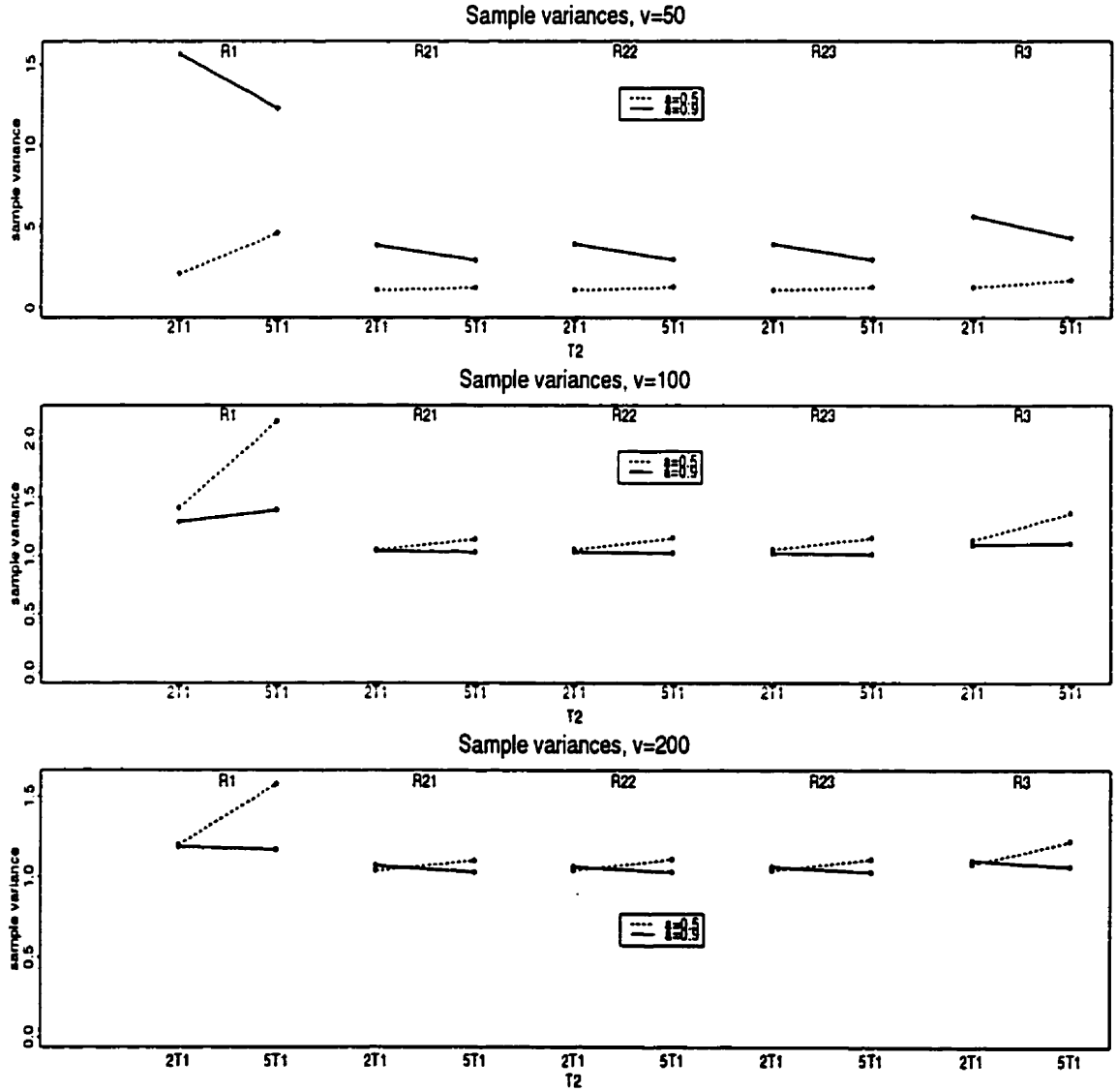


Figure 4.3: Plots of the sample skewness ($\hat{\gamma}_1$) as $T2$ increases from $2 * T1$ to $5 * T1$ and v increases from 50 to 200 for one set of bootstrap distributions based on EOSN.

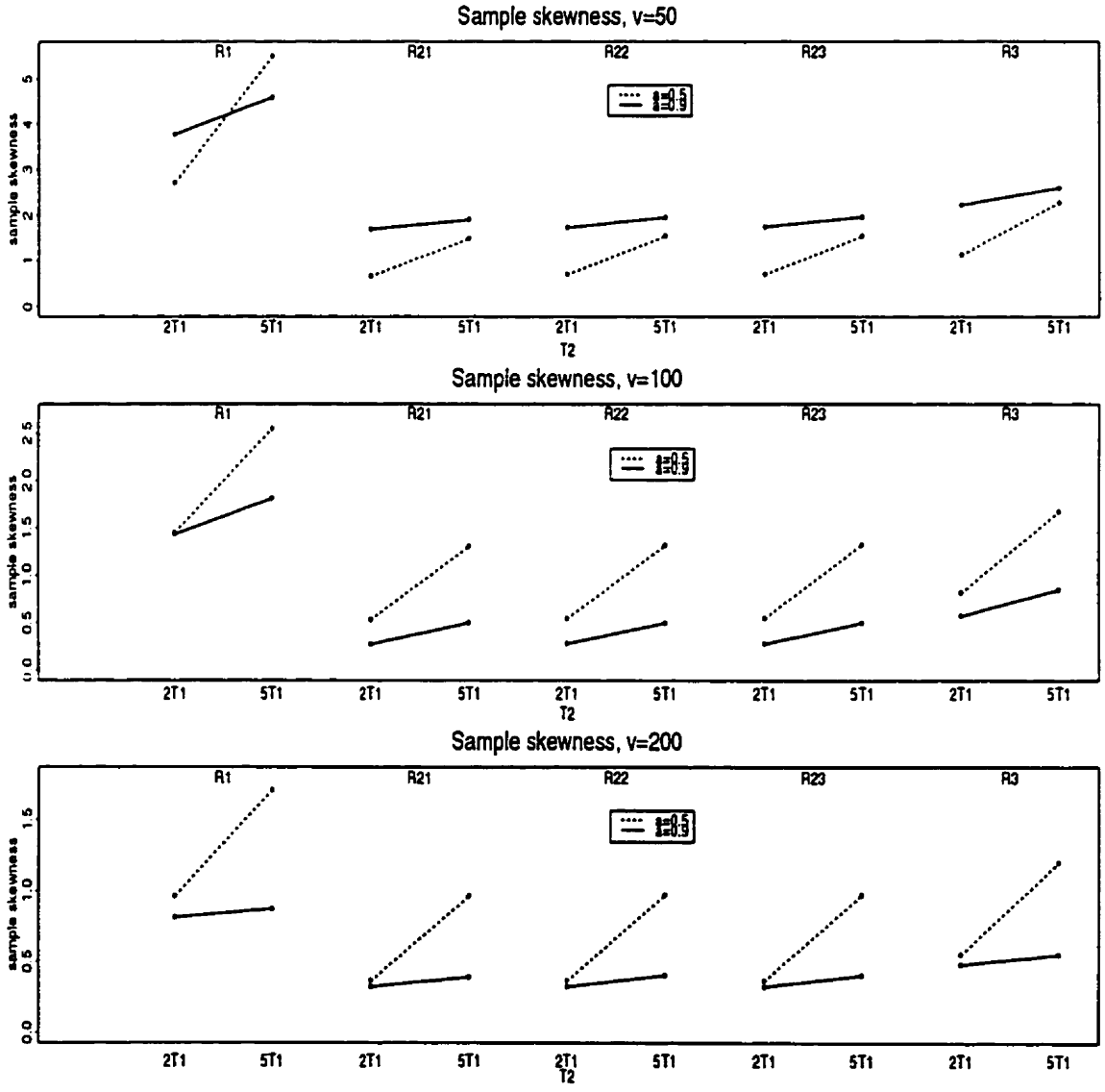


Figure 4.4: Plots of the sample kurtosis ($\hat{\gamma}_2$) as $T2$ increases from $2*T1$ to $5*T1$ and v increases from 50 to 200 for one set of bootstrap distributions based on EOSN.

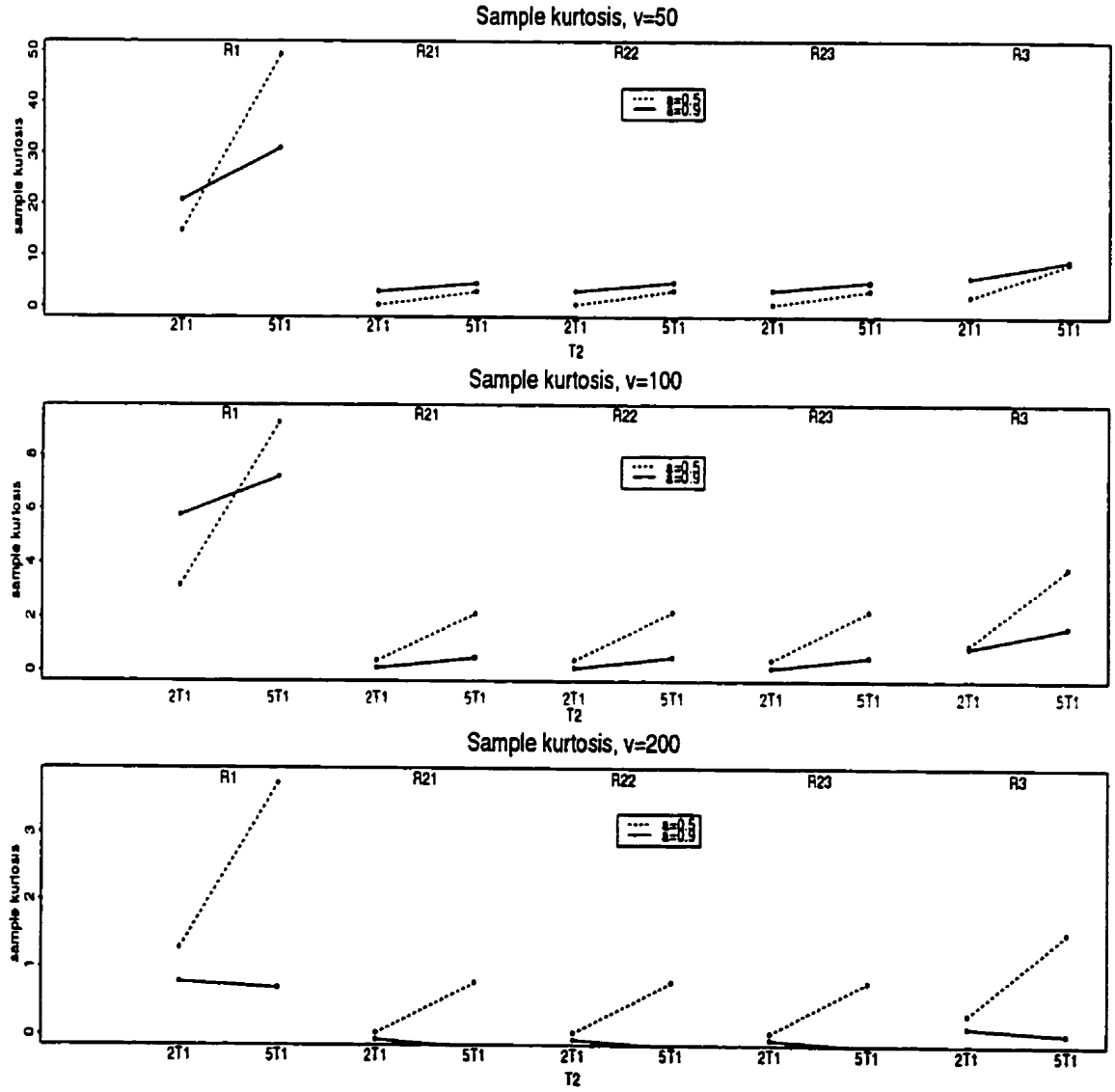


Table 4.3: Two-sided 95% prediction intervals of N2 - for one set of twelve simulated samples from the EOSN model.

Data		$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_{22}$	$\mathcal{D}_{23}$	$\mathcal{D}_3$
$v = 50$						
T1=0.69 T2=2T1, N2=11, $u(\hat{\eta}) = 10$	boot	[1,42]	[1,39]	[1,39]	[1,39]	[1,40]
	normal	[0,25]	[0,31]	[0,31]	[0,31]	[0,29]
T2=5T1, N2=22, $u(\hat{\eta}) = 16$	boot	[5,116]	[4,115]	[4,117]	[4,117]	[4,114]
	normal	[0,50]	[0,69]	[0,69]	[0,69]	[0,61]
T1=2.30 T2=2T1, N2=5, $u(\hat{\eta}) = 8$	boot	[1,74]	[1,69]	[1,70]	[1,70]	[1,71]
	normal	[0,18]	[1,21]	[0,21]	[0,21]	[0,19]
T2=5T1, N2=5, $u(\hat{\eta}) = 9$	boot	[2,84]	[2,81]	[2,81]	[2,81]	[2,82]
	normal	[0,22]	[0,27]	[0,27]	[0,27]	[0,25]
$v = 100$						
T1=0.69 T2=2T1, N2=24, $u(\hat{\eta}) = 25$	boot	[7,69]	[8,67]	[8,67]	[8,67]	[8,68]
	normal	[0,51]	[5,57]	[5,57]	[5,57]	[4,55]
T2=5T1, $N2 = 45, u(\hat{\eta}) = 43$	boot	[17,193]	[15,204]	[14,204]	[14,204]	[16,199]
	normal	[0,110]	[2,136]	[2,136]	[1,136]	[0,125]
T1=2.30 T2=2T1, N2=9, $u(\hat{\eta}) = 10$	boot	[2,24]	[2,23]	[2,23]	[2,23]	[2,24]
	normal	[0,20]	[2,23]	[2,23]	[2,23]	[1,22]
T2=5T1, N2=10, $u(\hat{\eta}) = 12$	bootstrap	[3,29]	[3,28]	[3,28]	[3,28]	[3,29]
	normal	[0,23]	[2,26]	[2,26]	[2,26]	[1,25]
$v = 200$						
T1=0.69 T2=2T1, N2=48, $u(\hat{\eta}) = 44$	boot	[18,93]	[18,91]	[18,91]	[18,91]	[18,92]
	normal	[10,78]	[16,84]	[16,84]	[16,84]	[14,82]
T2=5T1, N2=92, $u(\hat{\eta}) = 75$	boot	[33,246]	[30,252]	[30,252]	[30,252]	[31,250]
	normal	[0,162]	[13,187]	[13,187]	[13,187]	[8,177]
T1=2.30 T2=2T1, N2=17, $u(\hat{\eta}) = 32$	bootstrap	[15,59]	[15,59]	[15,59]	[15,59]	[15,59]
	normal	[11,52]	[14,55]	[14,55]	[14,55]	[13,54]
T2=5T1, N2=19, $u(\hat{\eta}) = 38$	boot	[18,75]	[18,74]	[18,74]	[18,74]	[18,74]
	normal	[10,65]	[15,70]	[15,70]	[15,70]	[14,68]

Table 4.4: Coverage probabilities and average lengths of two-sided 95% prediction intervals of N2 based on EOSN.

Data		$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_{22}$	$\mathcal{D}_{23}$	$\mathcal{D}_3$
$v = 50$						
T1=0.69 T2=2T1	boot	98.2 (49)	96.8 (44)	96.6 (45)	96.6 (45)	97.3 (46)
	normal	88.7 (29)	94.3 (36)	94.2 (36)	94.2 (36)	92.5 (33)
T2=5T1	boot	89.9 (348)	91.8 (415)	91.9 (427)	91.6 (431)	91.9 (384)
	normal	82.0 (98)	80.8 (144)	81.0 (144)	80.9 (144)	85.9 (125)
T1=2.30 T2=2T1	boot	98.4 (19)	97.9 (17)	97.7 (17)	97.6 (17)	98.3 (18)
	normal	92.4 (12)	96.3 (15)	96.3 (15)	96.4 (15)	94.9 (14)
T2=5T1	boot	93.9 (25)	95.0 (23)	95.2 (23)	95.1 (23)	94.6 (24)
	normal	92.2 (15)	95.8 (19)	96.1 (19)	96.3 (19)	94.8 (18)
$v = 100$						
T1=0.69 T2=2T1	boot	96.6 (61)	96.0 (59)	96.0 (59)	96.0 (59)	96.4 (60)
	normal	91.6 (50)	95.1 (53)	95.1 (53)	95.1 (53)	93.4 (52)
T2=5T1	boot	88.2 (332)	92.5 (375)	92.5 (377)	92.5 (378)	91.6 (357)
	normal	84.8 (155)	89.6 (204)	89.6 (204)	89.6 (204)	87.8 (183)
T1=2.30 T2=2T1	boot	98.2 (23)	97.8 (22)	97.9 (22)	97.9 (22)	97.9 (22)
	normal	94.4 (19)	96.2 (20)	96.3 (20)	96.3 (20)	96.0 (20)
T2=5T1	boot	97.3 (28)	97.5 (27)	97.5 (27)	97.5 (27)	97.5 (28)
	normal	93.7 (23)	96.6 (25)	96.8 (25)	96.8 (25)	95.8 (24)
$v = 200$						
T1=0.69 T2=2T1	boot	96.5 (83)	96.3 (82)	96.2 (82)	96.2 (82)	96.3 (82)
	normal	93.2 (78)	95.3 (78)	95.3 (78)	95.3 (78)	94.5 (77)
T2=5T1	boot	91.1 (369)	93.8 (397)	93.8 (398)	93.8 (398)	93.4 (386)
	normal	89.0 (252)	92.4 (288)	92.3 (288)	92.3 (288)	91.2 (275)
T1=2.30 T2=2T1	boot	96.1 (29)	95.3 (28)	95.3 (28)	95.3 (28)	95.5 (28)
	normal	94.2 (28)	95.5 (28)	95.7 (28)	95.8 (28)	95.0 (27)
T2=5T1	boot	95.5 (29)	95.2 (28)	95.2 (28)	95.2 (28)	95.3 (28)
	normal	94.1 (28)	95.7 (28)	95.7 (28)	95.7 (28)	94.9 (27)

Figure 4.5: Plots of the coverage probabilities of the 95% prediction intervals as $T2$ increases from $2 * T1$ to $5 * T1$ and v increases from 50 to 200.

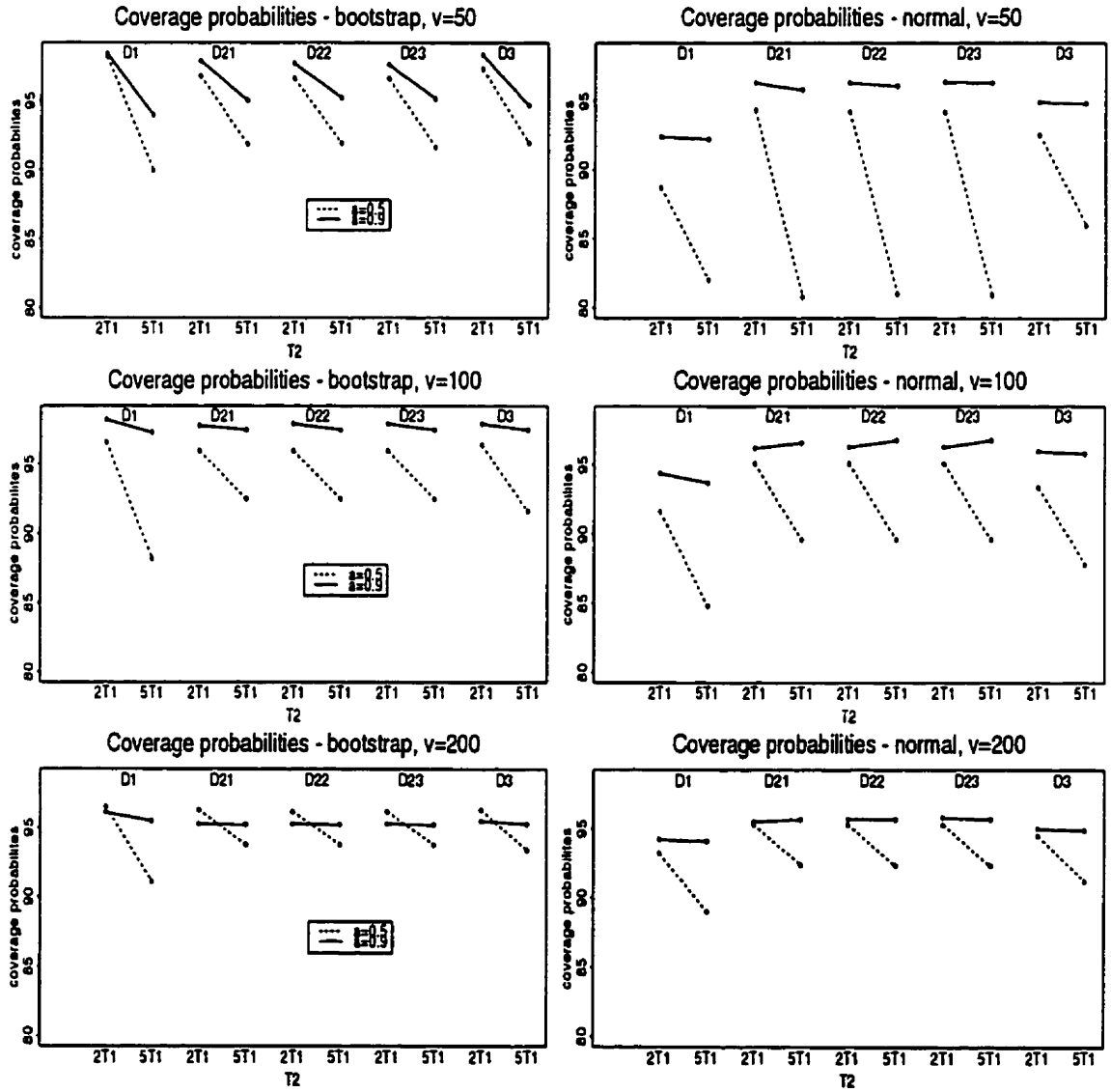


Figure 4.6: Plots of the average lengths of the 95% prediction intervals as $T2$ increases from $2 * T1$ to $5 * T1$ and v increases from 50 to 200.

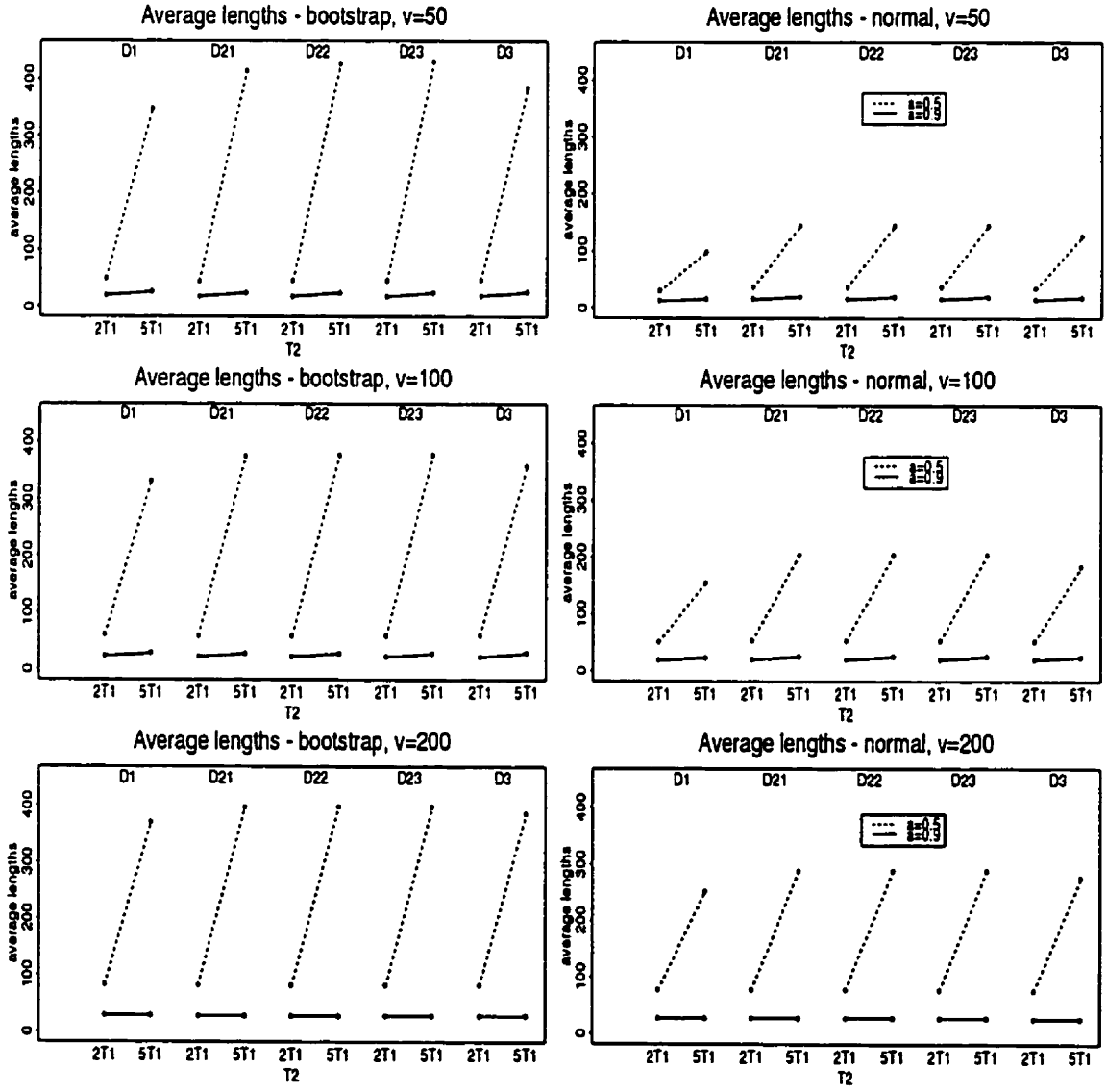


Table 4.5: Coverage probabilities and average lengths of two-sided 90% prediction intervals of N2 based on EOSN.

Data		$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_{22}$	$\mathcal{D}_{23}$	$\mathcal{D}_3$
$v = 50$						
T1=0.69 T2=2T1						
	boot	95.4 (37)	93.9 (35)	93.7 (35)	93.7 (35)	94.5 (36)
	normal	85.1 (26)	91.4 (30)	91.1 (30)	91.1 (30)	89.2 (29)
T2=5T1						
	boot	84.0 (242)	87.7 (284)	87.5 (289)	87.4 (291)	86.9 (265)
	normal	79.4 (86)	82.8 (121)	82.9 (121)	82.7 (121)	82.7 (106)
T1=2.30 T2=2T1						
	boot	95.6 (14)	94.0 (14)	94.2 (14)	93.9 (13)	94.6 (14)
	normal	90.0 (11)	92.6 (12)	92.8 (12)	93.2 (12)	92.1 (12)
T2=5T1						
	boot	90.7 (19)	90.5 (18)	90.6 (18)	90.4 (18)	90.7 (19)
	normal	89.5 (14)	93.6 (16)	94.0 (16)	94.0 (16)	92.7 (15)
$v = 100$						
T1=0.69 T2=2T1						
	boot	93.9 (49)	92.2 (48)	92.2 (48)	92.2 (48)	92.8 (48)
	normal	87.5 (44)	91.5 (44)	91.5 (45)	91.4 (45)	90.9 (44)
T2=5T1						
	boot	84.2 (247)	88.3 (273)	88.3 (274)	88.3 (274)	86.8 (262)
	normal	81.6 (139)	86.5 (172)	86.5 (172)	86.5 (172)	84.8 (159)
T1=2.30 T2=2T1						
	boot	96.3 (18)	95.3 (18)	95.3 (18)	95.3 (18)	95.5 (18)
	normal	90.9 (17)	93.1 (17)	93.3 (17)	93.4 (17)	92.3 (17)
T2=5T1						
	boot	94.2 (23)	94.1 (22)	94.2 (22)	94.3 (22)	94.1 (22)
	normal	91.1 (21)	93.2 (21)	93.2 (21)	93.3 (21)	92.4 (21)
$v = 200$						
T1=0.69 T2=2T1						
	boot	92.1 (68)	91.6 (68)	91.6 (68)	91.6 (68)	91.6 (68)
	normal	89.1 (65)	91.6 (65)	91.5 (65)	91.5 (65)	90.7 (65)
T2=5T1						
	boot	85.7 (290)	89.2 (307)	89.2 (308)	89.2 (309)	88.2 (301)
	normal	85.5 (227)	89.4 (242)	89.4 (242)	89.4 (242)	88.6 (301)
T1=2.30 T2=2T1						
	boot	91.6 (24)	91.3 (24)	91.2 (23)	91.2 (23)	91.4 (24)
	normal	90.8 (23)	91.8 (23)	91.8 (23)	91.8 (23)	91.1 (23)
T2=5T1						
	boot	90.3 (28)	90.2 (28)	90.2 (28)	90.2 (28)	90.3 (28)
	normal	91.3 (28)	91.5 (28)	91.6 (28)	91.6 (28)	91.6 (28)

empirical distributions for all the $\mathcal{R}_i^*$'s are positively skewed ($\hat{\gamma}_1 >> 0$), thick-tailed ($\hat{\gamma}_2 >> 0$) and non-normal when $v = 50$ and $T2 = 5\,T1$. For the larger values of v and same $T2$, the distributions seem approximately normal, except for those based on the $\mathcal{R}_1^*$. We also see that as v increases, the distributions within each set of $\mathcal{R}_i^*$'s become closer to normally distributed. The samples based on $\mathcal{R}_1^*$ are relatively more positively skewed, thick-tailed and non-standard normal than those based on the other $\mathcal{R}_i^*$'s, regardless of the values of v, $T1$ and $T2$. Similar behaviour occurs for $\mathcal{R}_3^*$, but not as pronounced as that for $\mathcal{R}_1^*$. As for the samples based on the $\mathcal{R}_2^*$'s, the ones for $\mathcal{R}_{22}^*$ and $\mathcal{R}_{23}^*$ within each sample have almost identical table values and plots. Overall, the samples based on these two $\mathcal{R}_i^*$'s are approximately normal, with the exceptions cited earlier. Lastly, the distributions based on $\mathcal{R}_{21}^*$ exhibit similar behaviour to those of the other $\mathcal{R}_2^*$'s when v is large, but are slightly less normal when $v = 50$.

Comparison of the $\mathcal{D}_i^*$

We next compared the 95% prediction intervals for various sets of twelve simulated samples. The results for one set of simulations are given in Table 4.3. When $T1 = 0.6931$ and $T2 = 5\,T1$, regardless of the size of v, the prediction intervals for all the $\mathcal{R}_i^*$'s are extremely long. This behaviour makes sense because if one has only observed roughly 50% of the total expected failures, one would anticipate that there would be more uncertainty attached to predicting the expected number in a rather large future time interval. Regardless of the size of v, we see that $\mathcal{D}_{21}^*$, $\mathcal{D}_{22}^*$, and $\mathcal{D}_{23}^*$ are virtually the same, except that the limits may differ by one or two when $v = 50$. This is not surprising, given the results for their respective $\mathcal{R}_i^*$'s. The $\mathcal{D}_3^*$'s also agree well with those given by the above three $\mathcal{D}_i^*$'s. There are some slight differences, but none that are appreciably large. Overall, the normal-based prediction intervals are shorter than those based on the bootstrap distribution. In

those situations for which the $\mathcal{R}_i^*$'s are non-normal in shape, there are relatively large differences in the prediction limits of the bootstrap intervals compared to those based on the normal-based ones. One would not want to obtain prediction intervals based on the normal approximation to the $\mathcal{R}_i^*$'s in these situations.

Comparison of the CP and AL

Both 95% and 90% prediction intervals were calculated for 1000 sets of twelve samples, and coverage probabilities and average lengths were calculated using these prediction intervals. The results are given in Tables 4.4 and 4.5. The CP are recorded in percentages and the AL are given in parentheses in both tables. Interaction plots for the CP and AL of the 95% prediction intervals are also given in Figures 4.5-4.6. The relatively large difference in the AL's between the 95% and 90% prediction intervals indicate that the distribution of the approximate pivotals has a long tail. Since the results are similar for both, the remaining discussion is based on those for the 95% prediction intervals. When $T1 = 0.6931$ and $T2 = 5\,T1$, regardless of the value of v and approximation used, the calculated CP's for all the $\mathcal{D}_i^*$'s are slightly smaller than the nominal value and the AL's are extremely long. This agrees with the results given above for the respective $\mathcal{R}_i^*$'s and $\mathcal{D}_i^*$'s for one set of simulations. For a given sample and approximation, the CP's and AL's for all the $\mathcal{D}_2$'s are equal. The CP's and AL's obtained using the normal approximation appear smaller than those based on the bootstrap.

Overall, for the EOSN model, both the normal- and the bootstrap-based prediction intervals have coverage probabilities close to the nominal values (either 0.95 or 0.90). The exception is when roughly 50% of the total number of expected failures v have been observed by $T1$. The normal-based prediction intervals are relatively shorter and have smaller coverage probabilities than those based on the bootstrap-t, regardless of which $\mathcal{R}_i^*$ is used. A conservative, but more computer-intensive,

approach would be to use the bootstrap approximation all the time, or at least when v is expected to be small. Nonetheless, for v known to be large, the quick and easy normal-based approximation seems appropriate for those prediction intervals based on any of the $\mathcal{R}_2$'s and possibly the $\mathcal{R}_3$. We do not recommend using the normal-based approximation for those based on $\mathcal{R}_1$, unless v is known to be very large. Regardless of the $\mathcal{R}_i$ used and the value of v, there is a lot of uncertainty observed for predicting $N2$ for large $T2$ when roughly 50% of the total number of expected failures v have been observed by $T1$. This behaviour is reflected in the extreme average lengths and the less-than-nominal coverage probabilities of all the prediction intervals for $N2$ based on both approximations to the $\mathcal{R}_i$'s. Further discussion is provided in Section 4.4.

4.3 Prediction with the accumulation of data

4.3.1 Introduction

As more information becomes available - that is, as $T1$ increases - $N3$ decreases and prediction of $N3 = N(T1, \infty]$ becomes more precise. This information may be useful for design purposes and for determining when to stop testing. In this section, we first discuss the frequentist and Bayesian approaches to repeated predictions. We follow with an application using the EOSN model.

4.3.2 Bayesian approach

To assess the effect of data accumulation on prediction, it may be easiest to think in terms of the Bayesian framework. Updating the appropriate predictive distribution is usually relatively easy to implement. Since we are interested particularly in

assessing the effect of data accumulation on prediction of $N3$, we need to update the predictive distribution as more data becomes available. For illustrative purposes, let us assume that we observe another time-truncated sample $D_{T1,T2}$ over the time interval $(T1, T2]$, where $0 < T1 < T2 < \infty$.

$$\begin{aligned} p(\eta|D_{0,T1}, D_{T1,T2}) &\propto p(\eta)\, L(D_{0,T1}, D_{T1,T2}|\eta) \\ &\propto p(\eta)\, p(D_{0,T1}|\eta)\, p(D_{T1,T2}|\eta) \\ &\propto p(\eta|D_{0,T1})\, p(D_{T1,T2}|\eta). \end{aligned}$$

Note that $p(\eta|D_{0,T1})$ effectively becomes the new prior density. It follows that the updated predictive distribution for $N3$ is given by

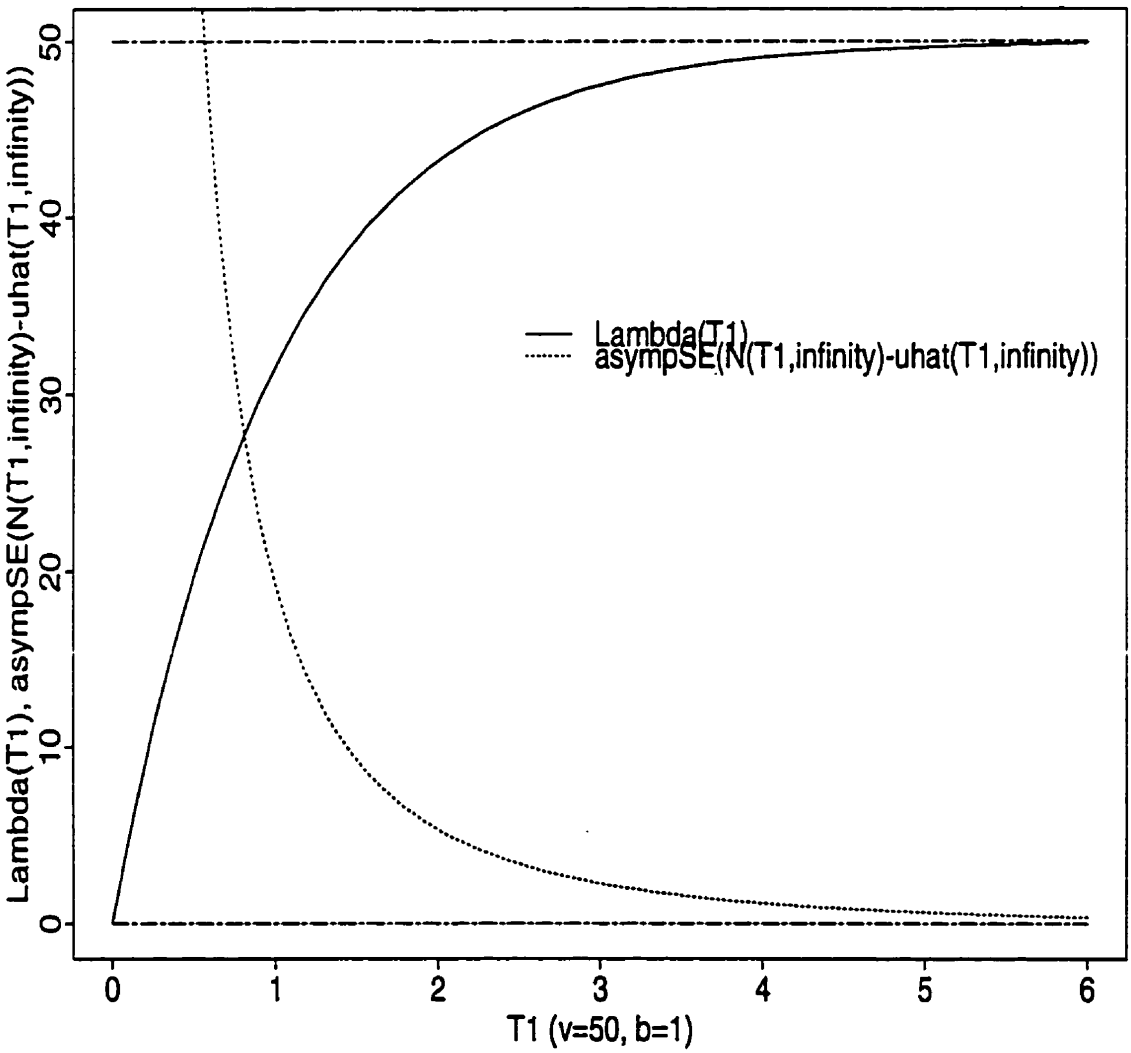
$$p(N3 = r|D_{0,T1}, D_{T1,T2}) = \int_{\eta} p(N3 = r|\eta)\, p(\eta|D_{0,T1}, D_{T1,T2})\, d\eta,$$

for $r = 0, 1, \ldots, \infty$. The updated point and interval estimates are then obtained using this updated predictive distribution. In the example below, we calculate 95% one- and two-sided highest density regions (HDR's) for the predictive distribution for $N3$. In this situation, the $100\alpha\%$ HDR for $N3$ is defined to be $I = \{N3 : p(N3|D_{(\cdot)}) \geq c\}$, where c is chosen such that $\sum_I p(N3|D_{(\cdot)}) = \alpha$.

4.3.3 Frequentist approach

In addition to the Bayesian approach, we consider a frequentist approach to assess the effect of data accumulation on the prediction of $N3 = N(T1, \infty)$. Although probability calculations are clear, there is no straightforward approach to assessing joint confidence levels in the frequentist scenario. Our approach is to simply give repeated 95% (say) prediction intervals. The frequentist prediction intervals are then

Figure 4.7: A plot of $\Lambda(T1;\hat{\eta})$ vs. $T1$ overlaid with a plot of $\sqrt{\hat{V}ar[N3-u(\hat{\eta})]}$ vs. $T1$ for the EOSN model and simulated data.



compared to the appropriate Bayesian HDR's as $T1$ increases. We also consider plots that show how $\sqrt{Var[N3 - u(\hat{\eta})]}$ changes as more data becomes available. For example, we may look at a plot of $\Lambda(T1; \hat{\eta})$ versus $T1$ overlaid with a plot of $\sqrt{Var[N3 - u(\hat{\eta})]}$ versus $T1$ to see how these curves relate to one another as $T1$ increases. As an illustration, we generated some data from the EOSN model with $\beta = 1$ and $v = 50$. We calculated and plotted the relevant quantities. The plot is given in Figure 4.7. We notice immediately that in the region where there is not much information about $v = \Lambda(\infty)$ (as discussed in Section 1.4), $\sqrt{Var[N2 - u(\hat{\eta})]}$ is relatively large. As more information becomes available, the asymptotic standard error decreases to zero.

4.3.4 An application

Model and data information

We obtained Bayesian and frequentist prediction intervals for the EOSN model. For reasons given in Section 2.4, we used the log-linear parameterization of the model. Under this parameterization, $u3(\eta) = \frac{-\exp(\alpha_1)}{\alpha_2}[\exp(\alpha_2 T1)]$. The frequentist prediction intervals for $N3$ were calculated in the same manner as those for $N2$, as described in Section 4.2. The relevant quantities used to calculate the Bayesian HDR's are given below.

Bayesian model information

For the log-linear parameterization,

$$L(D_{0,T1}|\alpha_1, \alpha_2) \propto \exp\left\{n\,\alpha_1 + \alpha_2\, S - \frac{\exp(\alpha_1)}{\alpha_2}[\exp(\alpha_2 T1) - 1]\right\}, \quad (4.2)$$

where $S = \sum_{i=1}^{n} t_i$, $-\infty < \alpha_1 < \infty$, and $\alpha_2 < 0$. We chose to use a joint conjugate

prior. In particular, we used the natural family of conjugate joint priors of the form

$$p(\alpha_1, \alpha_2) \propto \exp\left\{k1\,\alpha_1 + k2\,\alpha_2 + k3\frac{\exp(\alpha_1)}{\alpha_2}[\exp(\alpha_2 T1) - 1]\right\},$$

where $k1$, $k2$ and $k3$ are the hyper parameters that need to be specified. In particular, we chose an improper prior of the form

$$p(\alpha_1, \alpha_2) \propto \exp\{k1\,\alpha_1\}, \quad \text{for } k1 \leq 0, \tag{4.3}$$

by setting $k3 \equiv 0$ because we want the prior to be independent of the stopping time $T1$, and $k2 \equiv 0$ because this was done in [90] and [91]. For the comparative study, we preassigned $k1 = 0, -1$. We chose $k1 = -1$ because it corresponds to the noninformative conjugate prior under the "Raftery" parameterization used by [90] and [91] in their analysis. This prior is *not* invariant to changes in the time variable because both α_1 and α_2 are scale parameters under this parameterization. We also considered the time invariant conjugate prior with $k1 = 0$.

Multiplying equations (4.2) and (4.3) yields the following posterior density

$$p(\alpha_1, \alpha_2 | D_{0,T1}) = c^{-1} \exp\left\{(k1 + n)\alpha_1 + S\,\alpha_2 - \frac{\exp(\alpha_1)}{\alpha_2}[\exp(\alpha_2 T1) - 1]\right\},$$

where the constant c is obtained by integrating the above equation with respect to both parameters. In particular,

$$c = \Gamma(n + k1)\int_0^\infty \exp[-S\beta]\left(\frac{\beta}{1 - exp(-\beta T1)}\right)^{n+k1} d\beta.$$

Evaluation of the above integral requires numerical approximation methods. Since it is not a multiple integral, simple quadrature integration techniques, such as the

trapezoidal rule, may be implemented to approximate the integral with a high degree of accuracy. Before implementing the numerical integration method, it is useful to plot the integrand for the specific case on hand. Since the integrand is bell-shaped with rapidly decreasing tails in most cases, we can integrate over a closed region, say (c_L, c_U), instead of integrating over $(0, \infty)$. The values for c_L and c_U are chosen such that area under the curve is essentially zero outside this closed region. Trial and error is usually needed to ensure that you have picked appropriate values for c_L and c_U.

With $u3(\eta)$ given above, the predictive distribution for $N3$ is given by

$$\begin{aligned} p(N3 = r|D_{0,T1}) &= \int_{-\infty}^{0}\int_{-\infty}^{\infty} p(r|\alpha_1,\alpha_2)\, p(\alpha_1,\alpha_2|D_{0,T1})\, d\alpha_1 d\alpha_2 \\ &\propto \frac{\Gamma(n+k1+r)\,\Gamma(n+k1+1)}{r!\,c}\left(\frac{1}{T1\,r+S}\right)^{n+k1+1} \\ &\propto (n+k1)\prod_{i=-k1}^{n-2}(r+n-i-1)\left(\frac{1}{T1\,r+S}\right)^{n+k1+1} \end{aligned}$$

where the constant of proportionality c is given above, and $r = 0, 1, \ldots, \infty$. Numerical evaluation of integrals is not required in this case. In fact, numerically integrating the integral for c is not required as long as the calculated values for $p(N3 = r|D_{0,T1})$ are scaled so that $\sum_{r=0}^{\infty} p(N3 = r|D_{0,T1}) = 1$. We verified for specific cases that the infinite series $\sum_{r=0}^{\infty} p(N3 = r|D_{0,T1})$ converges for $k1 \leq 0$ using the integral test for the region of the series for which $p(N3 = r|D_{0,T1})$ is a decreasing sequence of values as r increases.

Data information

First, we calculated one- and two-sided 95% frequentist and Bayesian prediction intervals for $N3$ for DS3-4. These data sets were chosen for the same reasons given in the last section. As was done in the last section, we split the two data sets at

$T1 = 15.00$ and $T1 = 2.50$, respectively. $T2$ was chosen to be the actual truncation times for the data sets. We then used the observed data over $(0, T1)$ to obtain estimates of $N3$ over $(T1, \infty)$ and then used the entire data set $(0, T2)$ to obtain estimates of $N3$ over $(T2, \infty)$.

Next, we calculated one- and two-sided 95% frequentist and Bayesian prediction intervals for $N3$ for a set of three simulated samples from the EOSN model. Three samples from the EOSN model were generated for $\beta = 1$ and $v = 50, 100, 200$. The updated truncated times $T1$ and $T2$ were determined such that roughly 50% and 90% of the expected number of errors to be eventually detected are in fact discovered by these times, respectively. An outline of the simulation study is as follows:

1. Simulate N, a Poisson random variable with mean $v = E\{N(0, \infty)\}$.
2. Simulate N standard exponential random variables and sort them.
3. Segment the data set into two separate sections. In particular, $D_{0,T1}$ consists of the data in $(0, T1]$ with corresponding $N3$ given by the number of remaining simulated values; and, $D_{T1,T2}$ consists of the data in $(T1, T2]$, with corresponding $N3$ given by the number of left-over simulated values.
4. Obtain the Bayesian and frequentist prediction intervals of $N3$ based on D_{T1} and $D_{T2} = D_{T1} \bigcup D_{T1,T2}$.

Results

First, we discuss the results for DS3-4. Plots of the Bayesian predictive distributions for $N3$ based on the two prior densities are given in Figures 4.8 and 4.9. Regardless of the prior used, the predictive distributions are skewed. As $T1$ is increased,

Figure 4.8: Plots of the predictive distribution with $k1 = -1$ for N3, DS3,4 and two values of T1.

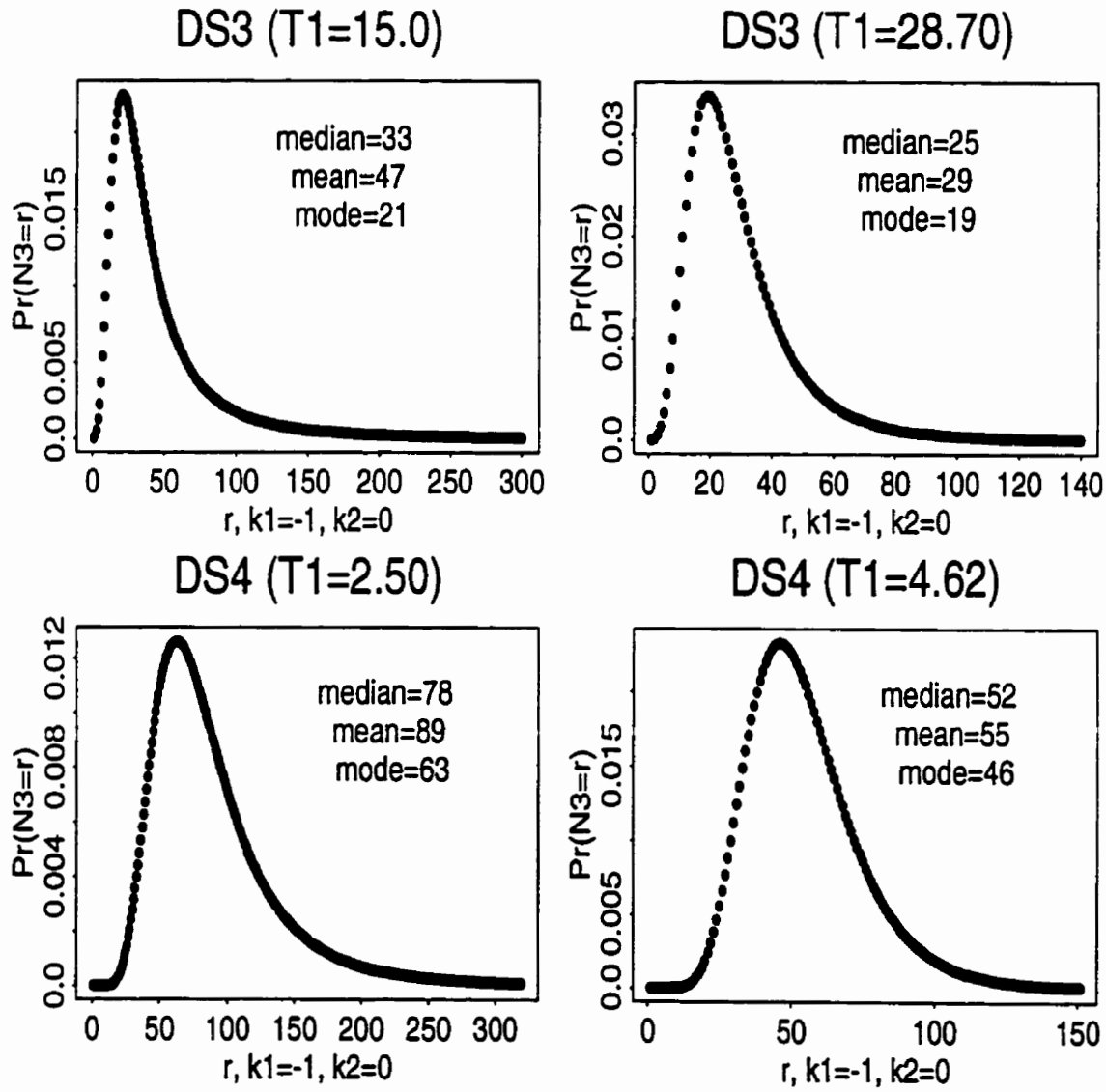


Figure 4.9: Plots of the predictive distribution with $k1 = 0$ for N3, DS3,4 and two values of T1.

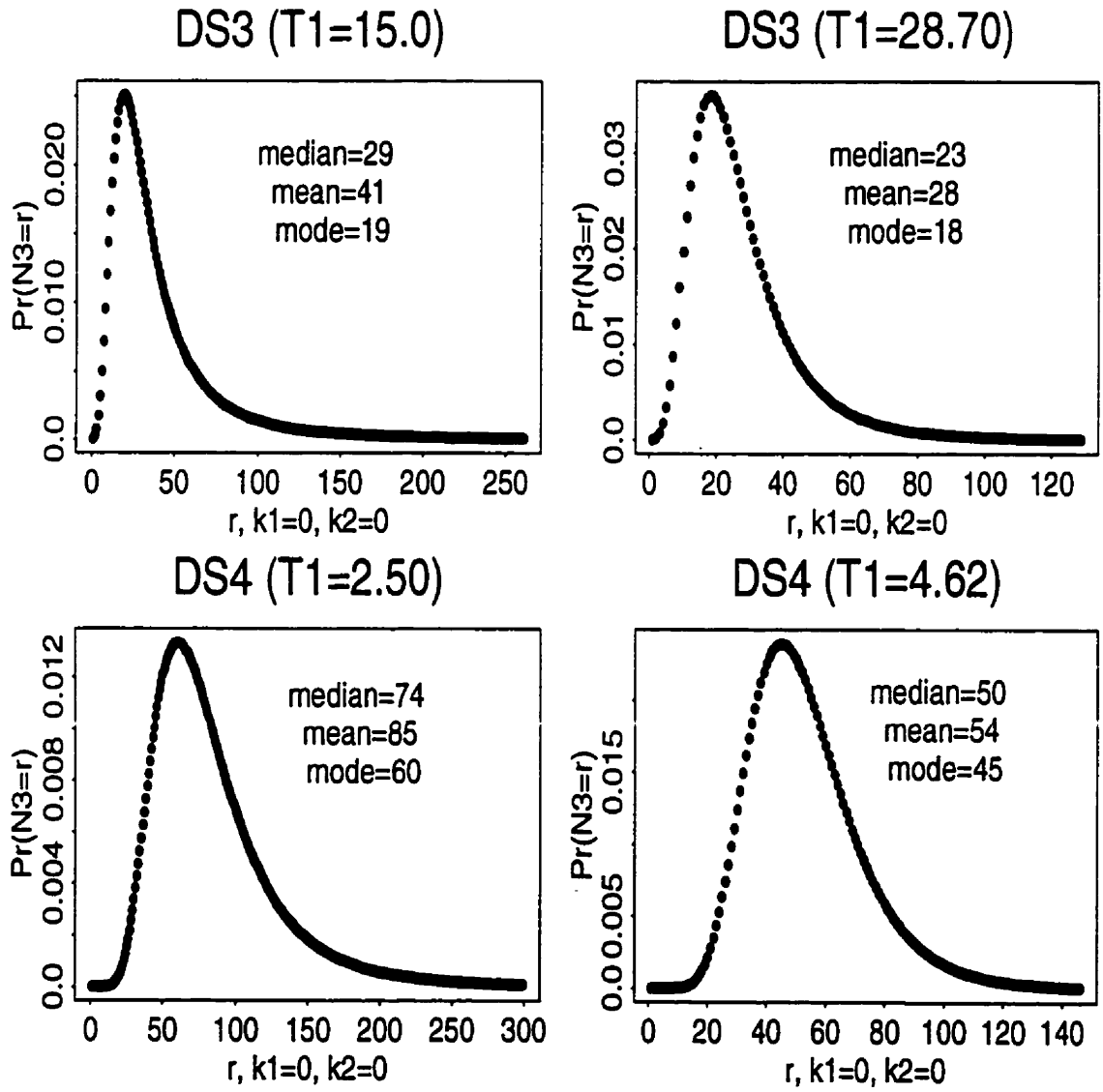


Table 4.6: Two-sided 95% prediction intervals of N3 based on EOSN for DS3-4 and two values of $T1$. Time is measured in hours.

Data attributes		*prediction intervals*		
DS3				
$T1 = 15.0$	$u(\hat{\eta}) = 30, \hat{V}(\hat{u}) = 459.30$			
	frequentist	$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
	bootstrap	[14,137]	[13,137]	[13,137]
	normal	[0,73]	[2,89]	[0,82]
	Bayesian			
	$k1 = -1, k2 = 0$		[2,132]	
	$k1 = k2 = 0$		[3,112]	
$T1 = 28.70$	$u(\hat{\eta}) = 23, \hat{V}(\hat{u}) = 140.32$			
	frequentist	$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
	bootstrap	[9,66]	[9,66]	[9,66]
	normal	[0,48]	[4,54]	[3,52]
	Bayesian			
	$k1 = -1, k2 = 0$		[4,64]	
	$k1 = k2 = 0$		[4,59]	
DS4				
$T1 = 2.50$	$u(\hat{\eta}) = 75, \hat{V}(\hat{u}) = 1140.17$			
	frequentist	$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
	bootstrap	[40,192]	[38,195]	[39,194]
	normal	[6,143]	[22,158]	[18,152]
	Bayesian			
	$k1 = -1, k2 = 0$		[23,184]	
	$k1 = k2 = 0$		[23,173]	
$T1 = 4.62$	$u(\hat{\eta}) = 50, \hat{V}(\hat{u}) = 263.45$			
	frequentist	$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
	bootstrap	[25,97]	[25,97]	[25,97]
	normal	[15,84]	[21,90]	[19,88]
	Bayesian			
	$k1 = -1, k2 = 0$		[21,94]	
	$k1 = k2 = 0$		[21,92]	

Table 4.7: One-sided 95% prediction intervals of N3 based on EOSN for DS3-4 and two values of $T1$. Time is measured in hours.

Data attributes			*prediction intervals*		
DS3					
$T1 = 15.0$	$u(\hat{\eta}) = 30, \hat{V}(\hat{u}) = 459.30$				
	frequentist		$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
		bootstrap	[0,106]	[0,109]	[0,108]
		normal	[0,66]	[0,77]	[0,73]
	Bayesian				
		$k1 = -1, k2 = 0$		[0,132]	
		$k1 = k2 = 0$		[0,111]	
$T1 = 28.70$	$u(\hat{\eta}) = 23, \hat{V}(\hat{u}) = 140.32$				
	frequentist		$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
		bootstrap	[0,56]	[0,55]	[0,56]
		normal	[0,44]	[0,48]	[0,47]
	Bayesian				
		$k1 = -1, k2 = 0$		[0,63]	
		$k1 = k2 = 0$		[0,58]	
DS4					
$T1 = 2.50$	$u(\hat{\eta}) = 75, \hat{V}(\hat{u}) = 1140.17$				
	frequentist		$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
		bootstrap	[0,167]	[0,167]	[0,167]
		normal	[0,132]	[0,143]	[0,138]
	Bayesian				
		$k1 = -1, k2 = 0$		[0,181]	
		$k1 = k2 = 0$		[0,170]	
$T1 = 4.62$	$u(\hat{\eta}) = 50, \hat{V}(\hat{u}) = 263.45$				
	frequentist		$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
		bootstrap	[0,86]	[0,86]	[0,86]
		normal	[0,78]	[0,83]	[0,81]
	Bayesian				
		$k1 = -1, k2 = 0$		[0,92]	
		$k1 = k = 0$		[0,89]	

they become less skewed and the median, mean and mode decrease. The calculated Bayesian and frequentist two- and one-sided 95% prediction intervals, $u(\hat{\eta})$ and $\hat{V}(\hat{u})$ are given in Tables 4.6 and 4.7, respectively. As expected, $\hat{V}(\hat{u})$ decreases when $T1$ is increased. For the most part, similar results hold for the one-sided as for the two-sided 95% prediction intervals. The exception is that, although they are slightly wider overall, the bootstrap-based one-sided prediction intervals are comparable to the Bayesian intervals based on the prior density with $k1 = 0$, while the two-sided bootstrap-based prediction intervals are comparable to the Bayesian intervals based on the prior with $k1 = -1$. As expected, the normal-based prediction intervals are narrower than the bootstrap-based prediction intervals. The Bayesian intervals based on the prior with $k1 = 0$ are slightly narrower than those based on the prior with $k1 = -1$.

Next, we discuss the results for the simulation study. Plots of the Bayesian predictive distributions for $N3$ based on the two prior densities are given in Figures 4.10 and 4.11. The plots are similar to those observed for DS3-4. Regardless of the prior used, the predictive distributions are skewed. As $T1$ is increased, they become less skewed and the median, mean and mode decrease a substantial amount. The calculated Bayesian and frequentist two- and one-sided 95% prediction intervals, $u(\hat{\eta})$, and $\hat{V}(\hat{u})$ are given in Tables 4.8 and 4.9, respectively. Similar results to those found for DS3-4 were obtained, with two noticeable exceptions. For one thing, regardless of the size of v, $\hat{V}(\hat{u})$ decreased substantially when $T1$ was increased to reflect a change from 50% to 90% of the expected total number of failures to have been observed by $T1$. Secondly, the upper limits for the bootstrap-based prediction intervals are infinite for $v = 50$ (for one- and two-sided intervals) and $v = 100$ (for two-sided intervals only). This occurs because we have *not* ignored those cases for which we did not obtain a finite, positive-valued m.l.e. for v. Recall that for

Figure 4.10: Plots of the predictive distribution with $k1 = -1$ for N3 and the three simulated samples from EOSN.

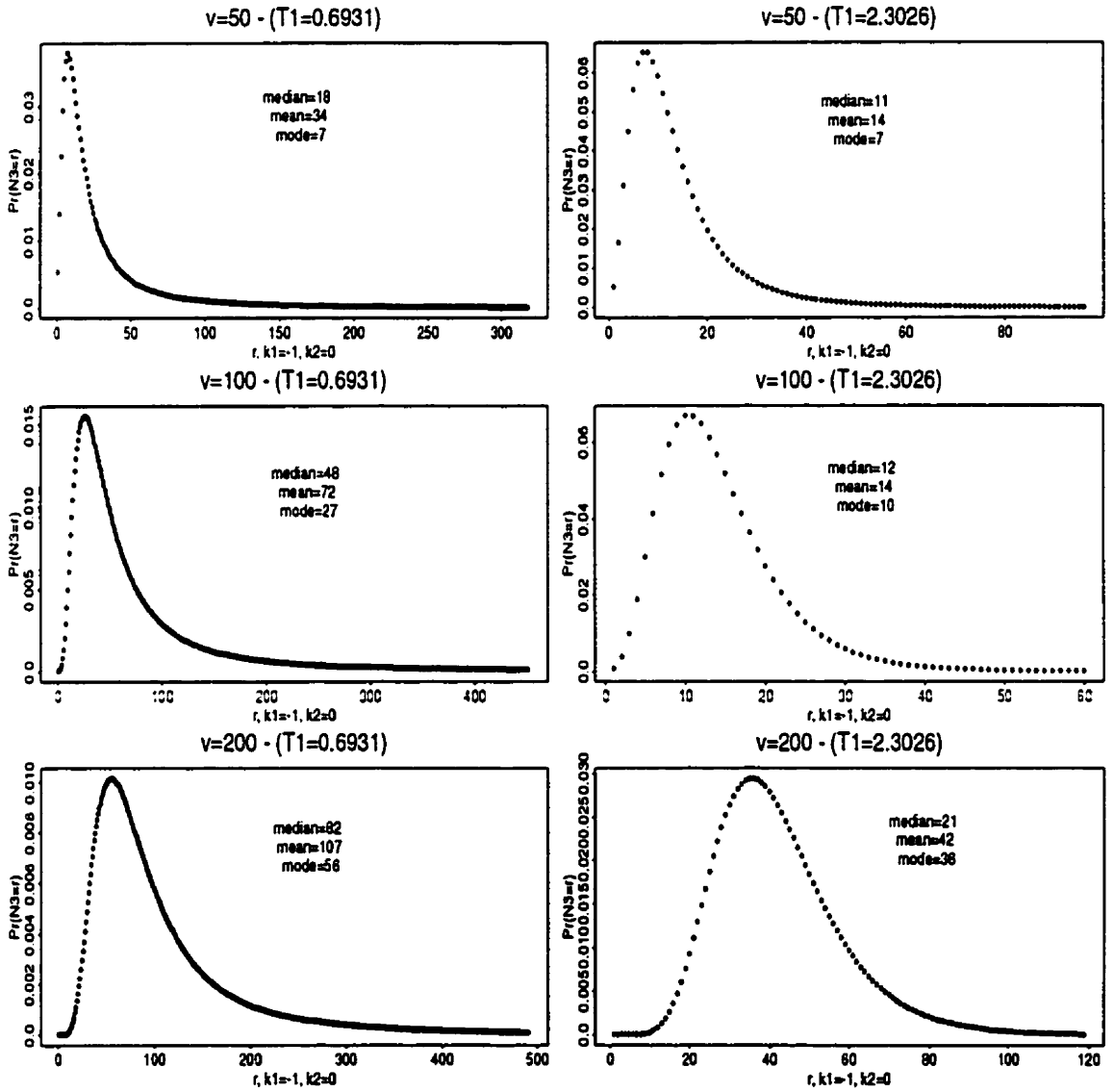


Figure 4.11: Plots of the predictive distribution with $k1 = 0$ for N3 and the three simulated samples from EOSN.

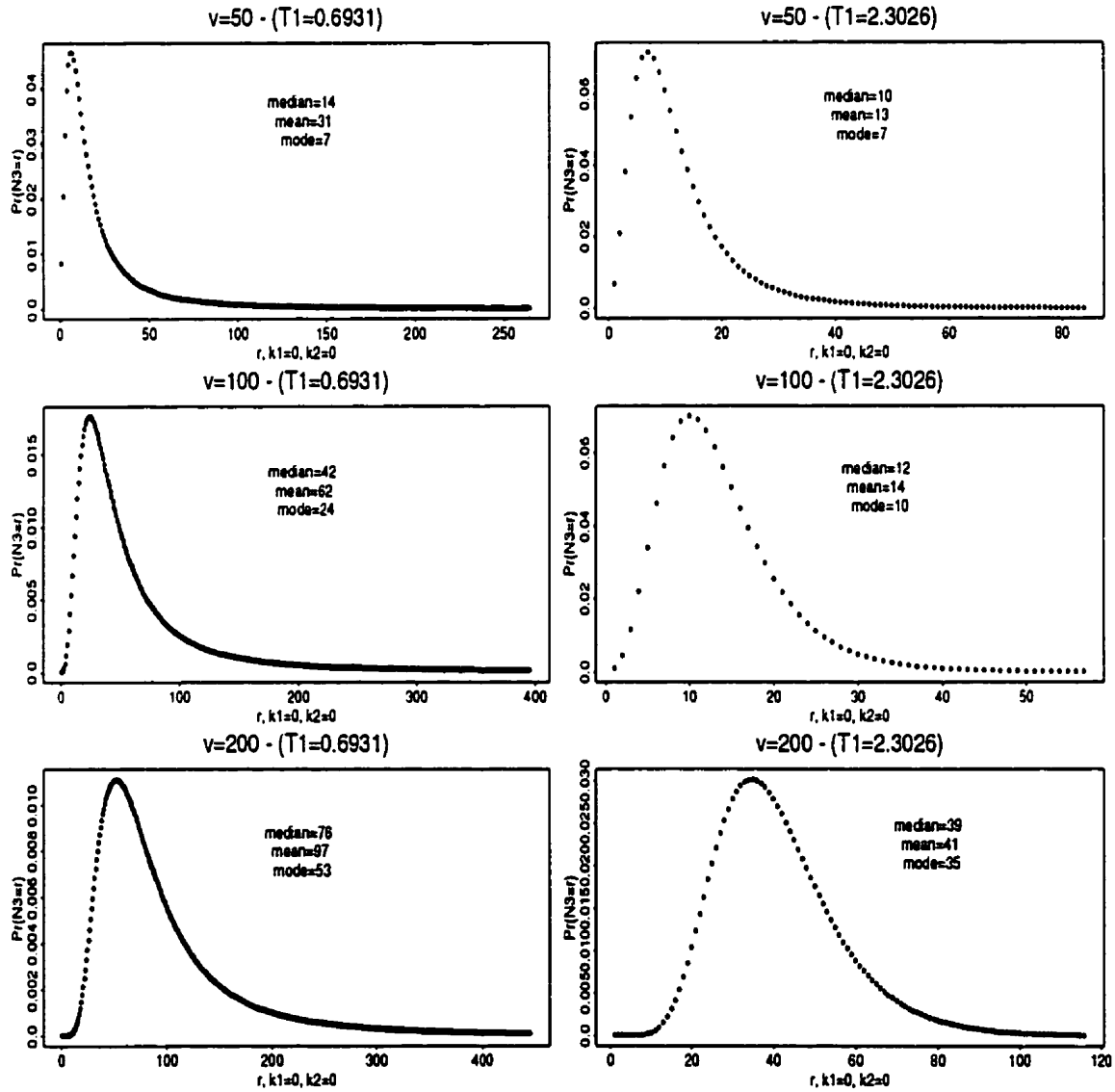


Table 4.8: Two-sided 95% prediction intervals of N3 - for three simulated samples from the EOSN model.

Data attributes			*prediction intervals*		
$v = 50$					
$T1 = 0.69$	$N3 = 30$, *frequentist*	$u(\hat{\eta}) = 16, \hat{V}(\hat{u}) = 351.76$			
			$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
		boot	[7,∞]	[6,∞]	[6,∞]
		normal	[0,53]	[0,75]	[0,65]
	Bayesian	$k1 = -1, k2 = 0$		[0,125]	
		$k1 = k2 = 0$		[0,80]	
$T1 = 2.30$	$N3 = 5$, *frequentist*	$u(\hat{\eta}) = 9, \hat{V}(\hat{u}) = 36.05$			
			$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
		boot	[2,84]	[2,81]	[2,82]
		normal	[0,22]	[0,27]	[0,25]
	Bayesian	$k1 = -1, k2 = 0$		[0,34]	
		$k1 = k2 = 0$		[0,30]	
$v = 100$					
$T1 = 0.69$	$N3 = 51$, *frequentist*	$u(\hat{\eta}) = 45, \hat{V}(\hat{u}) = 1459.55$			
			$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
		boot	[24,∞]	[23,∞]	[23,∞]
		normal	[0,121]	[1,153]	[0,139]
	Bayesian	$k1 = -1, k2 = 0$		[3,225]	
		$k1 = k2 = 0$		[0,188]	
$T1 = 2.30$	$N3 = 10$, *frequentist*	$u(\hat{\eta}) = 11, \hat{V}(\hat{u}) = 27.03$			
			$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
		boot	[3,29]	[3,28]	[3,29]
		normal	[0,23]	[2,26]	[1,25]
	Bayesian	$k1 = -1, k2 = 0$		[0.28]	
		$k1 = k2 = 0$		[0,26]	
$v = 200$					
$T1 = 0.69$	$N3 = 111$, *frequentist*	$u(\hat{\eta}) = 78, \hat{V}(\hat{u}) = 2429.06$			
			$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
		boot	[44,300]	[41,318]	[42,311]
		normal	[0,176]	[10,207]	[5,194]
	Bayesian	$k1 = -1, k2 = 0$		[15,274]	
		$k1 = k2 = 0$		[15,245]	
$T1 = 2.30$	$N3 = 17$, *frequentist*	$u(\hat{\eta}) = 38, \hat{V}(\hat{u}) = 157.13$			
			$\mathcal{D}_1$	$\mathcal{D}_{21}$	$\mathcal{D}_3$
		boot	[19,77]	[18,77]	[18,77]
		normal	[10,65]	[15,70]	[14,68]
	Bayesian	$k1 = -1, k2 = 0$		[16,73]	
		$k1 = k2 = 0$		[15,71]	

Table 4.9: One-sided 95% prediction intervals of N3 - for three simulated samples from the EOSN model.

Data attributes			*prediction intervals*
$v = 50$			
$T1 = 0.69$	$N3 = 30$, *frequentist*	$u(\hat{\eta}) = 16, \hat{V}(\hat{u}) = 351.76$	$\mathcal{D}_1$ $\mathcal{D}_{21}$ $\mathcal{D}_3$
		boot	[0,∞] [0,∞] [0,∞]
		normal	[0,47] [0,63] [0,56]
	Bayesian	$k1 = -1, k2 = 0$	[0,125]
		$k1 = k2 = 0$	[0,80]
$T1 = 2.30$	$N3 = 5$, *frequentist*	$u(\hat{\eta}) = 9$, $\hat{V}(\hat{u}) = 36.05$	$\mathcal{D}_1$ $\mathcal{D}_{21}$ $\mathcal{D}_3$
		boot	[0,56] [0,56] [0,56]
		normal	[0,20] [0,23] [0,22]
	Bayesian	$k1 = -1, k2 = 0$	[0,34]
		$k1 = k2 = 0$	[0,30]
$v = 100$			
$T1 = 0.69$	$N3 = 51$, *frequentist*	$u(\hat{\eta}) = 45, \hat{V}(\hat{u}) = 1459.55$	$\mathcal{D}_1$ $\mathcal{D}_{21}$ $\mathcal{D}_3$
		boot	[0,300] [0,338] [0,322]
		normal	[0,108] [0,131] [0,122]
	Bayesian	$k1 = -1, k2 = 0$	[0,224]
		$k1 = k2 = 0$	[0,188]
$T1 = 2.30$	$N3 = 10$, *frequentist*	$u(\hat{\eta}) = 11, \hat{V}(\hat{u}) = 27.03$	$\mathcal{D}_1$ $\mathcal{D}_{21}$ $\mathcal{D}_3$
		boot	[0,26] [0,25] [0,26]
		normal	[0,21] [0,23] [0,23]
	Bayesian	$k1 = -1, k2 = 0$	[0,28]
		$k1 = k2 = 0$	[0,26]
$v = 200$			
$T1 = 0.69$	$N3 = 111$, *frequentist*	$u(\hat{\eta}) = 78, \hat{V}(\hat{u}) = 2429.06$	$\mathcal{D}_1$ $\mathcal{D}_{21}$ $\mathcal{D}_3$
		boot	[0,242] [0,249] [0,247]
		normal	[0,160] [0,182] [0,173]
	Bayesian	$k1 = -1, k2 = 0$	[0,273]
		$k1 = k2 = 0$	[0,243]
$T1 = 2.30$	$N3 = 17$, *frequentist*	$u(\hat{\eta}) = 38, \hat{V}(\hat{u}) = 157.13$	$\mathcal{D}_1$ $\mathcal{D}_{21}$ $\mathcal{D}_3$
		boot	[0,69] [0,69] [0,69]
		normal	[0,61] [0,64] [0,63]
	Bayesian	$k1 = -1, k2 = 0$	[0,71]
		$k1 = k2 = 0$	[0,69]

the original parameterization of the model, $u3(\boldsymbol{\eta}) = v\beta exp(-\beta T1)$. From this expression, we see that when $\hat{v}$ is infinite, $u3(\hat{\boldsymbol{\eta}})$ is also infinite. If we refer back to Table 2.2, we note that when only 50% of $v = 50, 100$ have not occurred by time $T1$, approximately 16.70% and 7.90% of the estimated values for v will *not* be finite and positive-valued. With these results in mind, one can easily see why the upper limits of the bootstrap distribution for $N3$ are infinite for these cases. These calculated prediction intervals simply reflect the fact that insufficient information is available for making adequate inferences about $N3$.

4.4 Conclusions

The proposed frequentist-based interval predictors for $N2 = N(T1, T2]$, the number of events in the future time interval $(T1, T2]$ based on the observed data up to time $T1$, may be used for any NHPP model observed under the time-truncated sampling scheme. For the general NHPP model, the critical values of the distribution for any of the approximate pivotals $\mathcal{R}_i$'s used to obtain the prediction intervals will depend on the values for $\boldsymbol{\beta}$, $T1$ and $T2$. For those models which have a single β scale parameter, such as the GAMN and EOSN model, the critical values will depend on β, $T1$ and $T2$ only through the combination of $\beta T1$ and $\beta T2$.

For the EOSN model, both the normal- and the bootstrap-based prediction intervals for $N2$ have coverage probabilities close to the nominal values (either 0.95 or 0.90). The exception is when roughly 50% of the total number of expected failures v have been observed by $T1$. The normal-based prediction intervals for $N2$ are relatively shorter and have smaller coverage probabilities than those based on the bootstrap-t, regardless of which $\mathcal{R}_i^*$ is used. A conservative, but more computer-intensive, approach would be to use the bootstrap approximation all the time, or

at least when v is expected to be small. Nonetheless, for v known to be large, the quick and easy normal-based approximation seems appropriate for those prediction intervals based on any of the $\mathcal{R}_2$'s and possibly the $\mathcal{R}_3$. We do not recommend using the normal-based approximation for those based on $\mathcal{R}_1$, unless v is known to be very large.

Regardless of the $\mathcal{R}_i$ used and the value of v, there is a lot of uncertainty observed for predicting $N2$ for large $T2$ when roughly 50% of the total number of expected failures v have been observed by $T1$. This behaviour is reflected in the extreme average lengths and the less-than-nominal coverage probabilities of all the prediction intervals for $N2$ based on both approximations to the $\mathcal{R}_i$'s. Similar behaviour was observed for predicting $N3$ in the same context. This makes sense because we have inadequate data available to provide good predictions of v (and hence, $N2$ and $N3$) in these situations. As discussed in Chapter 2, more data is required in order to obtain good predictions.

Similar results hold for the one-sided as for the two-sided 95% prediction intervals, with one noticeable exception. It was observed for DS3-4 that, although they are slightly wider overall, the bootstrap-based one-sided prediction intervals for $N3$ are comparable to the Bayesian intervals based on the prior density with $k1 = 0$, while the two-sided bootstrap-based prediction intervals are comparable to the Bayesian intervals based on the prior with $k1 = -1$. Even though we did not obtain Bayesian prediction intervals for $N2$, we anticipate similar comparative results between the Bayesian and bootstrap-based prediction intervals.

Chapter 5

Further work

5.1 Continuation of current research

In Chapter 2, we determined conditions for a positive, finite $\hat{v} = \hat{\Lambda}(\infty)$ for specific models belonging to the general family of NHPP models with intensity $\lambda(t;\boldsymbol{\eta}) = v\,\lambda_0(t;\boldsymbol{\beta})$. For NHPP models (other than the EOS model) which have an equivalent GOS model, it is of interest to compare the conditions needed to obtain positive, finite-valued m. l. e. solutions. In particular, if the likelihood under the NHPP model is maximized at $v = \infty$, does this imply that the equivalent GOS likelihood is also maximized at $N = \infty$? Our results comparing the EOSN and EOS models indicate that there is a greater probability of obtaining $v = \infty$ using the EOSN model, than obtaining $N = \infty$ using the EOS model in certain situations. It would be nice to know if this holds for the other models.

In Chapter 3, we proposed a new approach for testing the goodness of fit of NHPP models having the above general form. We obtained conditional (on n) critical values of the $W^2_{n,\tilde{\boldsymbol{\beta}}}$ statistic, it is also of interest to examine the conditional and unconditional power of the test for various models. In addition, we also proposed

a four-parameter NEWN model that nests the EOSN, GAMN, WOSN and LEEN models. We want to investigate further the differences in the NEWN, WOSN and LEEN models by looking at their respective intensities and the extent to which we can discriminate among them with typical data.

In Chapter 4, we proposed a frequentist approach for providing approximate interval predictors of $N2 = N(T1, T2]$, the number of events in the future time interval $(T1, T2]$, based on the observed data up to time $T1$. We also used this method to assess the effect of data accumulation on prediction of $N3 = N(T1, \infty]$, the number of remaining events to be eventually observed given data has been observed up to time $T1$. We also discussed how to obtain Bayesian prediction intervals and compared them with frequentist-based prediction intervals in some examples with the EOSN model. In addition, it would be of interest to conduct similar simulation studies for the other models in Table 1.1. Some preliminary results (not given) on the prediction intervals constructed using the bootstrap-t and normal approximations to the $\mathcal{R}_i$'s for the real data sets D1-4 indicate that the results will be similar for the other models. Lastly, we are interested in assessing how the predictions of $N2$ and $N3$ based on various models will differ for specific data sets.

5.2 Additional research of interest

As mentioned in Section 1.2, both the GOS and NHPP families of models are not highly realistic models for modelling the occurrence of failures in the debugging phase of software program development. If more reported information about the underlying failure detection process were available, more ambitious models could be used to incorporate this information. In addition, a thorough understanding of

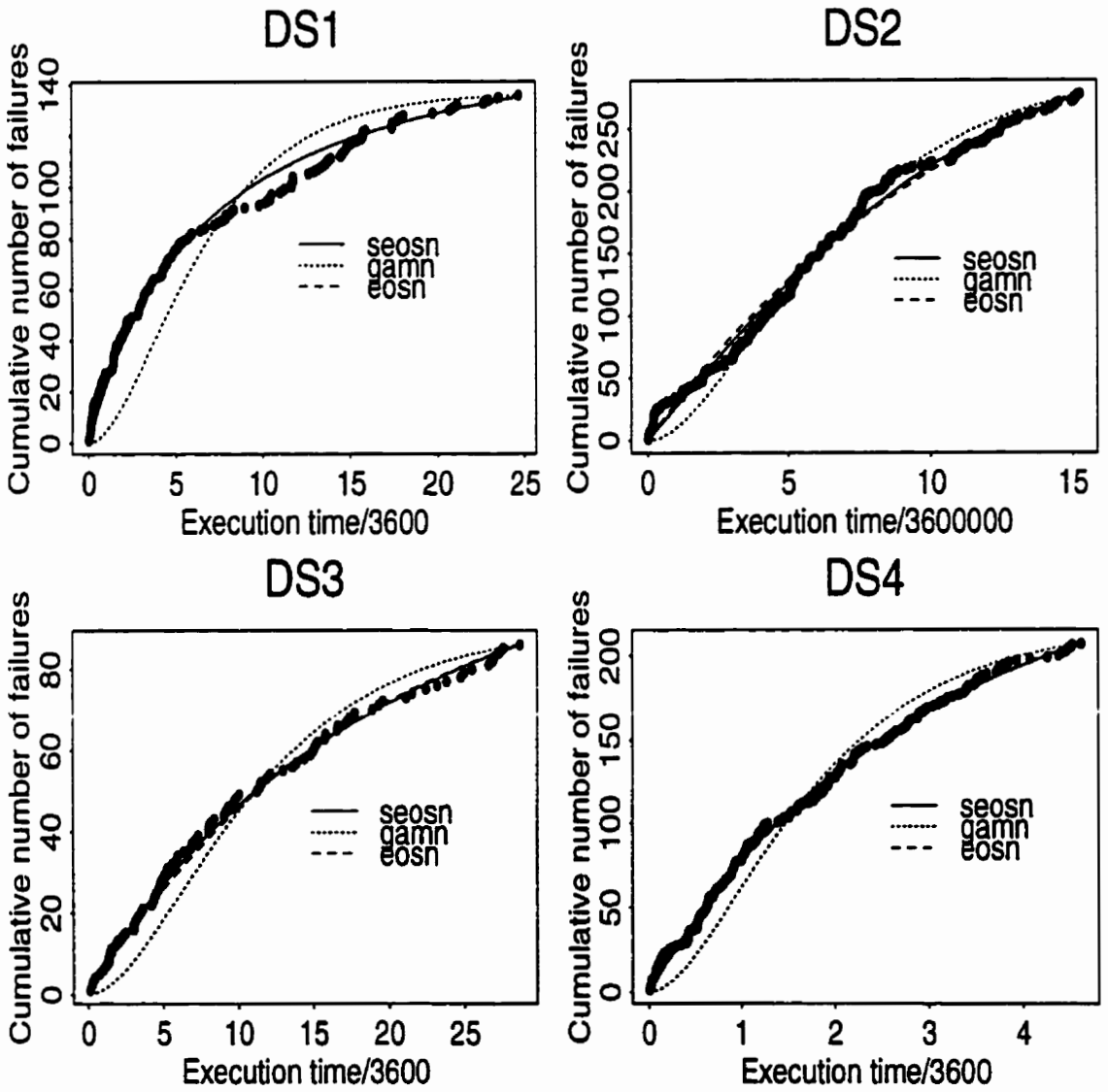
the underlying failure detection process could be used to evaluate the underlying assumptions of the GOS and NHPP models in this context.

As shown in Figure 3.2, there is an apparent lack of fit between the estimated mean value functions based on the models in Table 1.1 to the cumulative number of failures of DS1-2. This lack of fit is due mainly to the presence of change-points in the failure time behaviour over time. In order to reduce the model bias for predictive purposes, it is of interest to formulate tractable models that a) incorporate one or more change-point effects, and; b) can be used to test the statistical significance of these effects against an appropriate NHPP model alternative.

A change-point in the intensity $\lambda(t;\boldsymbol{\eta})$ of a NHPP can be generally defined as the point at which $\lambda(t;\boldsymbol{\eta})$ changes its trend or its functional form. Usually the change-point is unknown. Poisson models which take into account a change-point in the process have been investigated by some authors ([3], [76], [89], [102], [103]). It is of interest to develop and compare models that take into account change-point behaviour in the intensity function. For those models with an intensity function that is a function of the change-point τ, we will initially assume τ to be known, but we anticipate relaxing this assumption. When τ is considered unknown, it is of interest to test for the presence of the change-point. Complications arise in that under the null hypothesis, τ is meaningless and the m.l.e. $\hat{\tau}$ does not satisfy regularity conditions required for standard asymptotic theory (the parameter space under the null is on the boundary), as discussed in [3], [76], and [89].

Within the software reliability context, [102] and [103] attributed change-point behaviour to learning and skill improvements made by the programmers during the debugging phase of the failure process. They developed the "s-shaped" (GAMN) and "inflection s-shaped" (SEOSN) NHPP models to model change-point behavior. The intensity for the SEOSN model is given in Section 3. A particular feature shared

Figure 5.1: $\hat{\Lambda}(t;\eta)$ versus t for the EOSN, SEOSN, and GAMN models and DS1-4. Time is measured in hours.



by both models is that their intensity is not a function of τ. The SEOSN model is more flexible than the GAMN model in that it is not forced to be "s-shaped". The β_2 inflection parameter adds flexibility to the model. Plots of the estimated mean value functions under the EOSN, GAMN and SEOSN models and the cumulative number of failures versus time for DS1-4 are given in Figure 5.1. While GAMN does not fit the data, the SEOSN model fits well overall. In fact, the SEOSN and EOSN models give very similar fits. Since the EOSN model is a subset of the SEOSN model, a LR test may be performed to test the null hypothesis that the EOSN model is appropriate. At the $\alpha = 0.5$ level, we can not reject the null for DS2-4, but we can reject the null hypothesis for DS1. The observed AIC value was also obtained for the SEOSN model and given in Table 3.5. The SEOSN model is not the preferred model for any of the data sets, including DS1. In fact, the WOSN, LOGN, LEEN and NEWN models perform better than the EOSN model for DS1.

Non-Bayesian and Bayesian approaches to inference under a change-point HPP model that has an intensity that changes from λ_1 to λ_2 at τ is considered by [3] and [89], respectively. The change-point τ is considered unknown, as are λ_1 and λ_2. A log-linear NHPP change-point model with intensity

$$\lambda(t;\eta) = \begin{cases} \exp(\alpha_1 + \alpha_2 t) & \text{for } 0 \le t \le \tau \\ \exp(\alpha_1 + \delta_1 + \alpha_2 t) & \text{for } \tau < t \le T1, \end{cases}$$

for unknown parameters α_1, α_2, δ_1 and change-point τ, is considered by [76]. We are also interested in developing similar τ-dependent models in our context. Note that the above model used by [76] is a special case of the model given by

$$\lambda(t;\eta) = \begin{cases} \lambda(t;\eta) & \text{for } 0 \le t \le \tau \\ \delta\,\lambda(t;\eta) & \text{for } \tau < t \le T1, \end{cases} \tag{5.1}$$

where $\lambda(t;\eta) = \exp(\alpha_1 + \alpha_2 t)$, $\delta_1 = \exp(\delta)$ and τ unknown. This intensity is simply that of the log-linear reparameterized EOSN model.

It is of interest to compare the models given by equation (5.1) to their appropriate counterparts given by

$$\lambda(t;\eta) = \begin{cases} \lambda(t;\eta) & \text{for } 0 \leq t \leq \tau \\ \delta\,\lambda(\delta\,t;\eta) & \text{for } \tau < t \leq T1 \end{cases} \tag{5.2}$$

Our motivation for considering such models arises from a possible change-point effect due to a change in the testing compression factor, as discussed in detail in [82] and [83]. It is mentioned in both papers that stress (inputs) placed upon the software system during testing is often accelerated, as is done in hardware testing. The resultant effect is that the time-scale is multiplied by the testing compression factor. Although it seems more reasonable to anticipate the stress load placed on the software system to change often, we will only consider the case when it is assumed to change once; hence, the model given in equation (5.2). In addition to that of the EOSN model, it would be of interest to use the LOGN and POWN intensities in equations (5.1) and (5.2). Since the GAMN model already accounts for change-point behaviour, it does not seem reasonable to consider it. The WOSN and POWN models already have three parameters, so a fourth could make them more intractable. As discussed earlier, it seems reasonable to first assume that the change-point τ is known for models given by equations (5.1) and (5.2) in order to check the possibility of using these models under this situation. One may then relax this assumption and then check on their feasibility again.

Bibliography

[1] Aalen, O. O. and Hoem, J. M. (1978) "Random time changes for multivariate counting processes", *Scandinavian Actuarial Journal*, 81-101.

[2] Abdalla, A. A., Chan, P. Y. and Littlewood, B. (1986) "Evaluation of competing software reliability predictions", *IEEE Transactions on Software Engineering*, 12, 950-967.

[3] Akman, V. E. and Raftery, A. E. (1986) "Asymptotic inference for a change-point Poisson process", *The Annals of Statistics*, 14, 1583-1590.

[4] Akman, V. E. and Raftery, A. E. (1986) "Bayes factors for non-homogeneous Poisson processes with vague prior information", *Journal of the Royal Statistical Society*, B, 48, 322-329.

[5] Andersen, P. K., Borgan, O., Gill, R. G., and Keiding, N. (1993) *Statistical Methods Based on Counting Processes*, New York: Springer-Verlag.

[6] Anscombe, F. J. (1948) "The transformation of Poisson, binomial, and negative binomial data", *Biometrika*, 35, 246-254.

[7] Anscombe, F. J. (1953) Contribution to the discussion of H. Hotelling's paper. *Journal of the Royal Statistical Society*, B, 15, 229-230.

[8] Ascher, H. and Feingold, H. (1984) *Repairable Systems Reliability*, New York: Marcel Dekker.

[9] Bain, L. J., Engelhardt, M. (1980) "Inferences on the parameters and current system reliability for a time truncated Weibull process", *Technometrics*, 22, 421-426.

[10] Bain, L. J., Engelhardt, M. and Wright, F. T. (1985) "Tests for an increasing trend in the intensity of a Poisson process: a power study", *Journal of the American Statistical Association*, 80, 419-422.

[11] Baker, R. D. (1996) "Some new tests of the power law process", *Technometrics*, 38, 256-265.

[12] Barendregt, L. G. and van Pul, M. C. (1995) "On the estimation of the parameters for the Littlewood model in software reliability", *Statistica Neerlandica*, 49, 165-184.

[13] Bar-Lev, S. K., Lavi, I. and Reiser, B. (1992) "Bayesian inference for the power law process", *Annals of Institute of Statistical Mathematics*, 44, 623-639.

[14] Barlow, R. E. and Campo, R. (1975) "Total time on test processes and application to failure data analysis", in *Reliability and Fault Tree Analysis*, edited by Barlow, R. E., Fussell, J., and Singpurwalla, N. D., 451-481, Philadelphia: SIAM.

[15] Bartlett, M. S. (1936) "The square root transformation in the analysis of variance", *Journal of the Royal Statistical Society Supplement*, 3, 68.

[16] Bastani, F. B. (1993) "Software Reliability", special issue of *IEEE Transactions on Software Engineering*, IEEE Computer Society, Washington, DC.

[17] Bell, C. B. (1984) "Inference for goodness-of-fit problems with nuisance parameters (application to signal detection) ", *Journal of Statistical Planning and Inference*, 9, 273-284.

[18] Beran, R. (1990) "Calibrating prediction regions", *Journal of the American Statistical Association*, 85, 715-723.

[19] Berger, J. O. (1980) *Statistical Decision Theory*, New York: Springer-Verlag.

[20] Breidt, F. J. Davis, R. A. and Dunsmuir, W. T. M. (1995) "Improved bootstrap prediction intervals for autoregressions", *Journal of Time Series Analysis*, 16, 177-200.

[21] Brocklehurst, S. Chan, P. Y. Littlewood, B. and Snell, J. (1990), "Recalibrating software reliability models", *IEEE Transactions on Software Engineering*, 16, 458-470.

[22] Brocklehurst, S. and Littlewood, B. (1992) "New ways to get accurate reliability measures", *IEEE Software*, 9, 34-42.

[23] Calabria, R., Guida, M., and Pulcini, G. (1989) "Bayes estimation of prediction intervals for a power law process", *Communications in Statistics - Theory and Methods*, 19, 3023-3035.

[24] Campodónico, S. and Singpurwalla, N. D. (1994) "A Bayesian analysis of the logarithmic-Poisson execution time model based on expert opinion and failure data", *IEEE Transactions on Software Engineering*, 20, 677-681.

[25] Campodónico, S. and Singpurwalla, N. D. (1995) "Inference and predictions from Poisson point processes incorporating expert knowledge", *Journal of the American Statistical Association*, 90, 220-226.

[26] Cox, D. R. and Isham, V. (1980) *Point Processes*, London: Chapman and Hall.

[27] Cox, D. R. and Lewis, P. A. W. (1966) *The Statistical Analysis of Series of Events*, London: Methuen.

[28] Crow, L. H. (1974) "Reliability analysis for complex, repairable systems", *Reliability and Biometry*, 379-410.

[29] Crow, L. H. (1982) "Confidence interval procedures for the Weibull process with applications to reliability growth", *Technometrics*, 24, 67-72.

[30] Crowder, M. J., Kimber, A. L., Smith, R. L. and Sweeting, T. J. (1991). *Statistical Analysis of Reliability Data*, London: Chapman and Hall.

[31] D'Agostino, R. B. and Stephens, M. A. (1986) *Goodness-of-fit Techniques*, New York: Marcel Dekker.

[32] Dalal, S. R. and Mallows, C. L. (1988) "When should one stop testing software?", *Journal of the American Statistical Association*, 83, 872-879.

[33] DiCiccio, T. J. and Efron, B. (1996) "Bootstrap confidence intervals", *Statistical Science*, 14, 189-228.

[34] Efron, B. and Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*, New York: Chapman and Hall.

[35] Engelhardt, M. (1995) "Models and analyses for the reliability of a single repairable system", *Recent advances in life-testing and reliability*, Boca Raton: CRC, 79-106.

[36] Engelhardt, M. (1988) "Weibull processes", *Encyclopedia of Statistical Sciences*, 9, 557- 561.

[37] Engelhardt, M. (1986) "On the asymptotic behavior of the mean time between failures for repairable systems", *Reliability and quality control*, Amsterdam: North-Holland, 1-7.

[38] Engelhardt, M. and Bain, L. J. (1978) "Prediction intervals for the Weibull process", *Technometrics*, 20, 167-169.

[39] Engelhardt, M. Williams, D. H. and Bain, L. J. (1993) "Statistical analysis of a power-law process with left-truncated data", *Advances in Reliability*, 105-121.

[40] Evans, M. and Swartz, T. (1995) "Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems",*Statistical Science*, 10, 254-272.

[41] Evans, M. and Swartz, T. (1995) "Comments on 'Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems' ", *Statistical Science*, 11, 54-64.

[42] Finkelstein, J. M. (1976), "Confidence bounds on the parameters of the Weibull process", *Technometrics*, 18, 115-117.

[43] Forman, E. H. and Singpurwalla, N. D. (1977) "An empirical stopping rule for debugging and testing computer software", *Journal of the American Statistical Association*, 72, 750-757.

[44] Geurts, W. A. J., Hasselaar, M. M. A. and Verhagen, J. H. (1988), "Large sample theory for statistical inference in several software reliability models", *Technical Report MS-8807*, Amsterdam: Centre for Mathematics and Computer Science.

[45] Gill, R. D. (1986) "The total time on test plot and the cumulative total time on test statistic for a counting process", *The Annals of Statistics*, 14, 1234-1239.

[46] Goel, A. L. and Okumoto, K. (1979) "Time-dependent error-detection rate model for software reliability and other performance measures," *IEEE Transactions on Reliability*, 28, 206-211.

[47] Goel, A. L. (1985) "Software reliability models", special issue of *IEEE Transactions on Software Engineering*, IEEE Computer Society, Washington, DC.

[48] Guida, M., Calabria, R., and Pulcini, G. (1989) "Bayes Inference for a non-homogeneous Poisson process with power intensity law", *IEEE Transactions on Reliability*, 38, 603-609.

[49] Haber, S. (1970) "Numerical techniques of multiple integrals", *SIAM Review*, 12, 481-526.

[50] Hahn, G. J. and Nelson, W. (1973) "A survey of prediction intervals and their applications", *Journal of Quality Technology*, 5, 178-188.

[51] Higgins, J. J. and Tsokos, C. P. (1981) "A quasi-Bayes estimate of the failure intensity of a reliability growth model", *IEEE Transactions on Reliability*, 30, 471-475.

[52] Hjort, N. L. (1990) "Goodness of fit tests in models for life history data based on cumulative hazard rates", *The Annals of Statistics*, 18, 1221-1258.

[53] Hossain, S. A. and Dahiya, R. C. (1992) "Estimation of parameters of Yamada et. al. model and modification of the model", *Communications in Statistics - Theory and Methods*, 21 (10), 2881-2898.

[54] Jelinski, Z. and Moranda, P. B. (1972) "Software reliability research", in *Statistical Computer Performance Evaluation*, ed. W. Freiberger, New York:Academic Press, 465-497.

[55] Joe, H. (1989). "Statistical inference for general order statistics and nonhomogeneous Poisson process software reliability models", *IEEE Transactions on Software Engineering*, 15, 1485-1489.

[56] Joe, H. and Reid, N. (1985) "Estimating the number of faults in a system", Journal of the American Statistical Association, 80, 222-226.

[57] Johnson, N. L., Kotz, S. and Kemp, A. W. (1992) *Univariate Discrete Distributions*, 2nd ed. New York: Wiley.

[58] Kalbfleisch, J. G. (1985) *Probability and Statistical Inference, Volume 2: Statistical Inference*. New York: Springer-Verlag.

[59] Khmaladze, E. V. (1981) "Martingale approach to the goodness of fit tests", *Theory of Probability and its Applications*, 26, 240-265.

[60] Khmaladze, E. V. (1988) "An innovation approach to goodness of fit tests in R^m", *The Annals of Statistics*, 16, 1503-1516.

[61] Khoshgoftaar, T. (1988) "Nonhomogeneous Poisson processes for software reliability growth", *8th Symposium in Computational Statistics*, August 1988, (Compstat '88), 11-12.

[62] Khoshgoftaar, T. and Woodcock, T. (1992) "Software reliability model selection", *Quality and Reliability Engineering International*, 8, 457-469.

[63] Klefsjö. B. and Kumar, U. (1992) "Goodness-of-fit tests for the power-law process based on the TTT-Plot", *IEEE Transactions on Reliability*, 41, 593-598.

[64] Kuo, L. and Yang, T. Y. (1995) "Bayesian computation of software reliability", *Journal of Computational and Graphical Statistics*, 4, 65-82.

[65] Kuo, L. and Yang, T. Y. (1996) "Bayesian computation for nonhomogeneous Poisson processes in software reliability", *Journal of the American Statistical Association*, 91, 763-773.

[66] Kyparisis, J. and Singpurwalla, N. D. (1985) "Bayesian inference for the Weibull process with applications to assessing software reliability growth and predicting software errors", in *Computer Science and Statistics*, ed. L. Billard, London: Elsevier Applied Science, 57-64.

[67] Lawless, J. F. (1982) *Statistical models and methods for lifetime data*, New York: Wiley.

[68] Lawless, J. F. and Thiagarajah, K. (1996) "A point-process model incorporating renewals and time trends, with application to repairable systems", *Technometrics*, 38, 131-138.

[69] Lee, L. (1980) "Testing adequacy of the Weibull and the log linear rate models for a Poisson process", *Technometrics*, 22, 195-200.

[70] Lee, L. and Lee, S. K. (1978) "Some results in inference for the Weibull process", *Technometrics*, 20, 41-45.

[71] Lee, S. M., Bell, C. B. and Mason, A. W. I. (1988) "Discrimination, signal detection, and estimation for Weibull-type Poisson processes", *IEEE Transactions on information theory*, 34, 576-580.

[72] Lindsay, B. G. and Roeder, K. (1987) "A unified treatment of integer parameter models", *Journal of the American Statistical Association*, 82, 758-764.

[73] Littlewood, B. and Verrall, J. L. (1973 a) "A Bayesian reliability growth model for computer software", *Applied Statistics*, 22, 332-346.

[74] Littlewood, B. and Verrall, J. L. (1973 b) "A Bayesian reliability growth model for computer software", *Record of the 1973 IEEE Symposium on Computer Software Reliability*, IEEE, New York, 70-77.

[75] Littlewood, B. and Verrall, J. L. (1981) "Likelihood function of a debugging model for computer software reliability", *IEEE Transactions on Reliability*, 30, 145-148.

[76] Loader, C. R. (1992) "A log-linear model for a Poisson process change point", *The Annals of Statistics*, 20, 1391-1411.

[77] Miller, D. R. (1986) "Exponential order statistic models of software reliability growth," *IEEE Transactions on Software Engineering*, 12, 12-24.

[78] Moek, G. (1984) "Comparison of some software reliability models for simulated and real failure data", *International Journal of Modelling and Simulation*, 4, 29-41.

[79] Møller, S. K. (1976) "The Rasch-Weibull process", *Scandinavian Journal of Statistics*, 3, 107-115.

[80] Musa, J. D. (1975) "A theory of software reliability and its applications", *IEEE Transactions on Software Engineering*, 1, 312-327.

[81] Musa, J. D. (1979) "Software reliability data". Data Analysis Centre for Software, Rome Air Development Centre, Rome, N. Y. Technical Report.

[82] Musa, J. D. and Ackerman, A. F. (1989) "Quantifying software validation: when to stop testing?", *IEEE Software*, 5, 19-27.

[83] Musa, J. D., Iannino, A. and Okumoto, K. (1987) *Software reliability measurement, prediction, application*, McGraw-Hill.

[84] Musa, J. D. and Okumoto, K. (1984) "A logarithmic-Poisson execution time model for software reliability measurement" in *Proceedings of the 7th International Conference on Software Engineering*, 230-238.

[85] National Research Council (NRC) (1995) *Statistical software engineering*, Committee on Applied and Theoretical Statistics, Board on Mathematical Sciences. Washington, DC: National Academy Press.

[86] O'Hagan, A. (1994) *Kendall's advanced theory of statistics Volume 2B: Bayesian inference*, Great Britain: Wiley.

[87] Park, W. J. and Kim, Y. G. (1992) "Goodness-of-fit tests for the power-law process", *IEEE Transactions on Reliability*, 41, 107-110.

[88] Patel, J. K. (1989) "Prediction intervals - a review", *Communications in Statistics - Theory and Methods*, 18 (7), 2393-2465.

[89] Raftery, A. E. and Akman, V. E. (1986) "Bayesian analysis of a Poisson process with a change-point", *Biometrika*, 73, 85-89.

[90] Raftery, A. E. (1987) "Inference and prediction for a General Order Statistic model with unknown population size", *Journal of the American Statistical Association*, 82, 1163-1168.

[91] Raftery, A. E. (1988) "Analysis of a Simple Debugging Model", *Applied Statistics*, 37, 12-22.

[92] Rigdon, S. E. (1989) "Testing goodness-of-fit for the power law process", *Communications in Statistics - Theory and Methods*, 18, 4665-4676.

[93] Rigdon, S. E. and Basu, A. P. (1989) "The power law process: a model for the reliability of repairable systems", *Journal of Quality Technology*, 21, 251-260.

[94] Schneidewind, N. F. (1993) "Software reliability model with optimal selection of failure data", *IEEE Transactions on Software Engineering*, 19, 1095-1104.

[95] Seber, G. A. F. and Wild, C. J. (1989) *Nonlinear Regression*. New York: Wiley.

[96] Singpurwalla, N. D. and Wilson, S. P. (1994) "Software Reliability Modeling", *International Statistical Review*, 62, 289-317.

[97] van Pul, M. C. (1992) "Asymptotic properties of a class of statistical models in software reliability", *Scandinavian Journal of Statistics*, 19, 235-253.

[98] van Pul, M. C. (1993) "Statistical analysis of software reliability models", *CWI Tract*, 95, Stichting Mathematisch Centrum, Centrum voor Wiskunde en Informatica, Amsterdam.

[99] van Pul, M. C. (1994) "A general introduction to software reliability", *CWI Quarterly*, 7, 203-244.

[100] Xie, M. (1993) "Software reliability models - a selected annotated bibliography", *Software Testing, Verification and Reliability*, 3, 3-28.

[101] Yamada, S., Narihisa, H. and Osaki, S. (1984) "Optimum release policies for a software system with a scheduled software delivery time", *International Journal of Systems Science*, 15, 905-914.

[102] Yamada, S., Ohba, M. and Osaki, S. (1983) "S-shaped reliability growth modeling for software error detection", *IEEE Transactions on Reliability*, 32, 475-478.

[103] Yamada, S., Ohba, M. and Osaki, S. (1984) "S-shaped software reliability growth models and their applications", *IEEE Transactions on Reliability*, 33, 289-291.

[104] Zeephongsekul, P., Xia, G. and Kumar, S. (1994) "Software reliability growth models based on cluster point processes", *International Journal of Systems Science*, 25, 737-751.

[105] Zhao, M. and Xie, M. (1996) "On maximum likelihood estimation for a general non-homogeneous Poisson process", *Scandinavian Journal of Statistics*, 23, 597-607.