CWI

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

W.A.J. Geurts, M.M.A. Hasselaar, J.H. Verhagen

Large sample theory for statistical inference in several software reliability models

Department of Mathemàtical Statistics Report MS-R8807 June

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

Large Sample Theory for Statistical Inference in Several Software Reliability Models

W.A.J. Geurts

M.M.A. Hasselaar

J.H. Verhagen
Mathematical Institute,
University of Utrecht.

The theory of counting processes is applied to several error-counting, debugging models from software reliability, the aim being to estimate the total number of bugs in the system and other parameters. Large-sample results on the maximum likelihood estimators are obtained as the number of software bugs increases while the observation period remains fixed. Testing goodness-of-fit of the (parametric) models is also considered, using an "innovations" approach.

1980 Mathematics Subject Classification: 62N05, 62M99
Keywords and Phrases: counting process, intensity process, estimating population size
Note: This report was written as part of the master's degree programme of the authors, under the supervision of Prof. R.D. Gill, Department of Mathematical Statistics, Centre for Mathematics and Computer Science, Amsterdam.

PREFACE

This paper is the report of a project to complete our studies in mathematics at the Mathematical Institute of the University of Utrecht. This project was carried out in a team. We worked from January 1987 on this project under the supervision of Prof. R.D. Gill. We want to thank Richard Gill for the inspiring supervision and the extremely useful contributions to the project. We also thank Johan Grasman for his useful comments and for his willingness to help us as much as possible. Finally we want to express our gratitude and appreciation to Geert Moek for his presence at our regular meetings and for giving us the opportunity to make use of his practical experience.

Report MS-R8807
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

CONTENTS

INTRODUCTION

Computer systems have become more and more important in modern society. Therefore it is important to give an estimation of the reliability of those systems, and in particular of their software. The discipline that deals with the reliability of computer-programming is called Software Reliability. Within that discipline, software reliability has been defined as the probability that a failure does not occur during a specified exposure time. A failure is interpreted as the result of a software fault which causes devation from the required output by more then specified tolerances. Several investigaters have built statistical models in order to estimate the reliability of software. They all considered the following experiment. A computer program has been executed during a specified exposure period and the failure-times are observed. If the repairing of a fault takes place immediately after it produces a failure, we are dealing with a so called debugging model. Many software reliability models consider also the number of remaining errors or faults in the program. These models are called error-counting models. By using the information obtained from the experiment one can estimate the parameters of the underlying model, especially the total number of remaining faults in the software. Mostly maximum likelihood estimation is used for this purpose. The debugging and error-counting models we consider are the model J.D. MUSA introduced in 1975 (also known as the model of JELINSKI & MORANDA (1972)), the model of A.L. GOEL & K. OKUMOTO from 1980 and the model B. LITTLEWOOD introduced in 1980. For a summary of the most common software reliability models we refer to RAMAMOORTHY & BASTANI (1982).

As we have already implied, we stopped the experiment using a stop-criterion, for example "stop after a specified exposure time". As a consequence of this stopping rule, the number of failures in the exposure period will be stochastic. This is the reason why it is not clear that classical theory about asymptotic normality of the maximum likelihood estimators will hold. The aim of this paper is therefore among other things to prove the consistency of the model parameter-estimators and moreover to derive their asymptotic distribution.

In the first section we will sketch our approach to the problem. We will introduce the theory of counting processes and martingales and we will make clear in what way we will apply asymptotic theory. A novel aspect of our approach is the fact that instead of increasing the exposure period of the program we will increase the total number of faults in the software. A general introduction of the software reliability models to be considered is given in the second section. In the third, fourth and fifth section we will consider respectively the Musa, Goel & Okumoto and Littlewood model. In the sixth section we will consider goodness of fit tests, especially for the model of Musa. Along the way, we indicate how results can be obtained for a general class of models including the three we study in detail.

1. Approach to the problem and mathematical framework

For the time being, let τ be that specified (non-random) stopping time of the experiment (for a stochastic stopping time τ, see the Remark on page 22). The fact that the number of faults detected in the time-interval $[0,\tau]$ will be stochastic is the reason why we cannot use classical maximum likelihood theory in deriving asymptotic theory. In particular, it is not clear that the approximate variance of the estimators can be found using the Fisher-information matrix or even that the estimators will be approximately normally distributed.

We now introduce a new instrument we will use: the theory of counting processes and martingales. For a complete summary we refer to Andersen & Borgan (1985). Here we give a brief, mainly heuristic, introduction. In the following random variables are printed in bold type.

A *counting process* $\mathbf{N}(t)$ or $\mathbf{n}_t$ is a stochastic process that can be thought of as the counting of events, in this case software failures, up to time t. Thus $\mathbf{n}_t$ is a non-decreasing integer valued function of time with jumps of size one only; it is right continuous and $\mathbf{n}_0=\mathbf{N}(0)=0$. In regular cases, accompanying a counting process $\mathbf{n}_t$ is an intensity process $\boldsymbol{\lambda}_t$. It is interpreted heuristically as the conditional probability rate that $\mathbf{n}_t$ jumps in a small time interval $[t,t+\mathrm{d}t]$ around t given all that has happened up to but not including t

$$\boldsymbol{\lambda}_t \mathrm{d}t=\Pr\{\mathbf{n}_{t+\mathrm{d}t}-\mathbf{n}_{t-}=1 \mid \mathcal{F}_{t-}\}\ , \tag{1.1}$$

where $\mathcal{F}_{t-}=\sigma\{\mathbf{n}_s,\ s<t\}$, the σ-algebra generated by the paths of $\mathbf{n}_s$ on the interval $[0,t)$. A formal definition of an intensity process will be given later.

A process $\mathbf{x}_t$, measurable with respect to $\mathcal{F}_t$ for each t is called *adapted*.

Notice that $\boldsymbol{\lambda}_t$ given the strict past is non-stochastic and therefore called *predictable*. (As a function of (ω,t) it is measurable with respect to the σ-algebra on $\Omega\times\mathbb{R}^+$ generated by the left continuous adapted processes.)

A *martingale* $\mathbf{m}_t$ or $\mathbf{M}(t)$ is a stochastic process with the property that the increment over a time-interval $(t,t+h]$ given the past has zero expectation. So formally

$$\mathrm{E}[\mathbf{m}_t \mid \mathcal{F}_s]=\mathbf{m}_s\quad \forall\ s<t\ . \tag{1.2}$$

Informally, taking $h=t-s$ small

$$\mathrm{E}[\mathrm{d}\mathbf{m}_t \mid \mathcal{F}_{t-}]=0\ .$$

Let

$$\mathbf{m}_t=\mathbf{n}_t-\int_0^t \boldsymbol{\lambda}_s \mathrm{d}s\ , \tag{1.3}$$

then $\mathrm{E}[\mathrm{d}\mathbf{m}_t \mid \mathcal{F}_{t-}]=\mathrm{E}[\mathrm{d}\mathbf{n}_t-\boldsymbol{\lambda}_t\mathrm{d}t \mid \mathcal{F}_{t-}]=\mathrm{E}[\mathrm{d}\mathbf{n}_t \mid \mathcal{F}_{t-}]-\mathrm{E}[\boldsymbol{\lambda}_t\mathrm{d}t \mid \mathcal{F}_{t-}]$. This equals $\mathrm{E}[\mathrm{d}\mathbf{n}_t \mid \mathcal{F}_{t-}]-\boldsymbol{\lambda}_t\mathrm{d}t$, since $\boldsymbol{\lambda}_t$ is predictable. Furthermore $\mathrm{E}[\mathrm{d}\mathbf{n}_t \mid \mathcal{F}_{t-}]=1\cdot\boldsymbol{\lambda}_t\mathrm{d}t+0\cdot(1-\boldsymbol{\lambda}_t\mathrm{d}t)=\boldsymbol{\lambda}_t\mathrm{d}t$. Hence, $\mathrm{E}[\mathrm{d}\mathbf{m}_t \mid \mathcal{F}_{t-}]=0$, so indeed $\mathbf{n}_t-\int_0^t \boldsymbol{\lambda}_s\mathrm{d}s$ is a martingale. $\int_0^t \boldsymbol{\lambda}_s\mathrm{d}s \overset{\mathrm{def}}{=} \mathbf{A}_t$ is called the compensator of the counting process $\mathbf{n}_t$, it is also a predictable process. Formally, we say that $\boldsymbol{\lambda}_t$ is the intensity of $\mathbf{n}_t$ if it is predictable and $\mathbf{n}_t-\int_0^t \boldsymbol{\lambda}_s\mathrm{d}s$ is a martingale.

Some important facts on martingales and counting processes we will use are (see ANDERSEN & BORGAN (1985)):

The predictable variation process:

$$d\langle \mathbf{m}\rangle(t) \stackrel{\text{def}}{=} \text{var}\{d\mathbf{m}_t \mid \mathfrak{F}_{t-}\} = \text{var}\{d\mathbf{n}_t - \boldsymbol{\lambda}_t dt \mid \mathfrak{F}_{t-}\} = \boldsymbol{\lambda}_t dt\,. \tag{1.4}$$

That this is true follows from the fact that $\boldsymbol{\lambda}_t$ is predictable and from the fact that $d\mathbf{n}_t$ is either 0 or 1. So $\text{var}\{d\mathbf{n}_t \mid \mathfrak{F}_{t-}\} = \boldsymbol{\lambda}_t dt \cdot (1-\boldsymbol{\lambda}_t dt) = \boldsymbol{\lambda}_t dt$. Formally for any (locally) square integrable martingale $\mathbf{m}$, $\langle \mathbf{m}\rangle$, the predictable variation process of $\mathbf{m}$, is the non-decreasing predictable process such that $\mathbf{m}^2 - \langle \mathbf{m}\rangle$ is a (local) martingale. In our case $\langle \mathbf{m}\rangle = \mathbf{A}$. (A process has a property *locally* if there exists an increasing sequence of stopping times converging to infinity, such that the process stopped at each stopping time has the property.)

Stochastic integration and predictable (co)variation:
If $\mathbf{h}_t$ is a predictable process then

$$\text{var}\{\mathbf{h}_t d\mathbf{n}_t \mid \mathfrak{F}_{t-}\} = \mathbf{h}_t^2 \text{ var}\{d\mathbf{m}_t \mid \mathfrak{F}_{t-}\} = \mathbf{h}_t^2 d\langle \mathbf{m}\rangle(t)\,. \tag{1.5}$$

In fact $\int_0^t \mathbf{h}_s d\mathbf{m}_s$ is a (local) square integrable martingale and $\langle \int \mathbf{h} d\mathbf{m}\rangle = \int \mathbf{h}^2 d\langle \mathbf{m}\rangle$. More generally

$$d\langle \mathbf{m}_1, \mathbf{m}_2\rangle(t) \stackrel{\text{def}}{=} \text{cov}\{d\mathbf{m}_1(t), d\mathbf{m}_2(t) \mid \mathfrak{F}_{t-}\} \tag{1.6}$$

where $\mathbf{m}_1(t)$ and $\mathbf{m}_2(t)$ are (local square integrable) martingales. Again $\mathbf{m}_1 \cdot \mathbf{m}_2 - \langle \mathbf{m}_1, \mathbf{m}_2\rangle$ is a local martingale. We find

$$\langle \int \mathbf{h}_1 d\mathbf{m}_1, \int \mathbf{h}_2 d\mathbf{m}_2\rangle(t) = \int_0^t \mathbf{h}_1(s)\mathbf{h}_2(s) d\langle \mathbf{m}_1, \mathbf{m}_2\rangle(s)\,, \tag{1.7}$$

where $\mathbf{h}_1$ and $\mathbf{h}_2$ are predictable.

Lenglart's inequality:
Let $\mathbf{m}_t$ be a local square integrable martingale.
Then for all δ, $\mu > 0$

$$\Pr\{\sup_{t\in[0,\tau]} |\mathbf{m}_t| > \mu\} \leqslant \frac{\delta}{\mu^2} + \Pr\{\langle \mathbf{m}, \mathbf{m}\rangle(\tau) > \delta\}\,. \tag{1.8}$$

This is an application of Lenglart's inequality.

Weak convergence:
Consider a sequence $\mathbf{n}_t^{(\nu)}$ of counting processes, with intensity process $\boldsymbol{\lambda}_t^{(\nu)}$ and a sequence $\mathbf{H}^{(\nu)}$ of $(p \times 1)$-vectors of predictable processes, and define

$$\mathbf{Z}_j^{(\nu)}(t) = \int_0^t \mathbf{H}_j^{(\nu)}\{d\mathbf{n}_s^{(\nu)} - \boldsymbol{\lambda}_s^{(\nu)} ds\} \quad j = 1, 2, \ldots, p\,. \tag{1.9}$$

If as $\nu \to \infty$

$$\langle \mathbf{Z}^{(\nu)}\rangle(t) \xrightarrow{P} G(t) \quad \forall \quad 0 \leqslant t \leqslant \tau\,, \tag{1.10}$$

where G is a $p \times p$ continuous matrix function, $\langle \mathbf{Z}^{(\nu)}\rangle$ is the matrix of elements $\langle \mathbf{Z}_i^{(\nu)}, \mathbf{Z}_j^{(\nu)}\rangle$

and if for all $\epsilon>0$ as $\nu\rightarrow\infty$

$$\int_0^\tau [\mathbf{H}_j^{(\nu)}(t)]^2 \boldsymbol{\lambda}^{(\nu)}(t) \cdot I\{|\mathbf{H}_j^{(\nu)}(t)|>\epsilon\}\mathrm{d}t \xrightarrow{\mathrm{P}} 0 \quad \text{for } j=1,\ldots,p \tag{1.11}$$

then $\mathbf{Z}^{(\nu)}$ converges in distribution to $\mathbf{Z}^{(\infty)}$ in the space $(D[0,\tau])^p$, where $\mathbf{Z}^{(\infty)}$ is a p-variate Gaussian martingale with covariance function G and $\mathbf{Z}^{(\infty)}(0)=0$. This is a special case of a martingale central limit theorem, see for example HELLAND (1982). (For a description of the space $(D[0,\tau])^p$ we refer to BILLINGSLEY (1968), Chapter 3.)

Now let us look at the way we will treat asymptotic behaviour. It does not make sense to let τ, the stopping time, grow to infinity. In many error-counting models (those in which each fault is completely removed when it produces a failure) in the long run the estimate of the total number of faults will trivially be equal to the true total number of faults. It makes more sense to (conceptually) increase the number of faults in the program. Let N be the true number of faults; we introduce a dummy variable β_0, where $N=\nu\beta_0$, ν some constant. The suggestion is then to increase ν $(\nu\rightarrow\infty)$. The idea is that then asymptotics should be relevant to the practical situation in which N is large and $\mathbf{n}_\tau/N$ not close to zero or one. Define

$\mathbf{n}_t^{(\nu)}$ the counting process $\mathbf{n}_t$ in the ν'th experiment, $\beta_0=N/\nu$ is some constant unknown fraction to be estimated, together with possibly several other parameters;
$\boldsymbol{\lambda}_t^{(\nu)}$ the intensity process of $\mathbf{n}_t^{(\nu)}$.

We will prove for the three models considered in detail later that the counting process $\mathbf{n}_t^{(\nu)}$ divided by ν will converge in probability uniformly on $[0,\tau]$ to a deterministic function of time depending on the true model parameters. Notice that it is very unusual to increase a model parameter itself, in this case N. This complication is solved by estimating β_0. In final results, e.g. asymptotic confidence regions, test-statistics, one can substitute N/ν for β_0. Besides β_0, also ν will then vanish, so we will have expressions containing N only.
We will illustrate our idea by Figures 1.1 and 1.2.

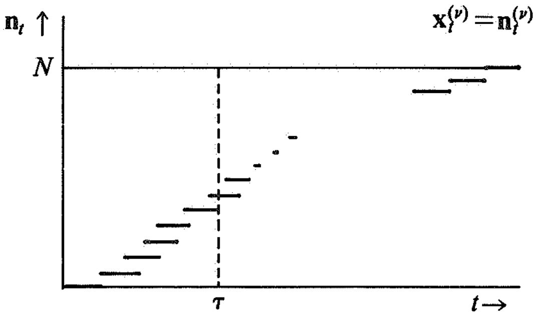

FIGURE 1.1.

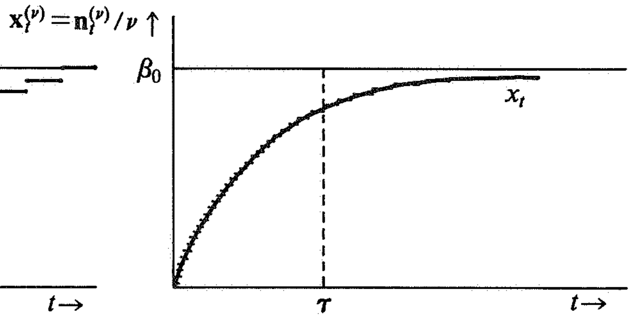

FIGURE 1.2.

In Figure 1.1 we see indeed that for $\tau\rightarrow\infty$, $\mathbf{n}_\tau$ eventually equals N, the true number of faults. In Figure 1.2 the vertical axis is rescaled. While increasing ν, $\mathbf{x}_t^{(\nu)} \overset{\text{def}}{=} \mathbf{n}_t^{(\nu)}/\nu$ will jump more often; more failures will occur in the time-interval $[0,\tau]$. We have also illustrated $\mathbf{x}_t^{(\nu)}$ approaching a deterministic function x_t with increasing ν.
To prove this convergence in probability of $\mathbf{x}_t^{(\nu)}$ to the deterministic function x_t in Figure 1.2., we will apply Theorem 2.1 of T.G. KURTZ (1983) (cf. Theorem 8.1, p. 52 of KURTZ (1981)). Here we

state a simplified version of the theorem which is sufficient for our applications.

THEOREM 1.1. *Suppose we can write* $\boldsymbol{\lambda}_t^{(\nu)}/\nu$ *as a non-anticipating function* $\beta(t, \mathbf{x}_s^{(\nu)}; 0 \leq s < \infty) = \beta(t, \mathbf{x}_s^{(\nu)}\lambda_t : 0 \leq s < \infty)$, *with the following properties:*

$$\beta : [0,\infty) \times D[0,\infty) \to [0,\infty) \text{ is non-negative and Borel-measurable} \tag{1.12}$$

$$\forall x \in D[0,\infty),\ \forall t \geq 0 \ \sup_{s \leq t} \beta(s,x) < \infty \tag{1.13}$$

$$\sup_{s \leq t} \beta(s,x) \leq \alpha(M_1 + M_2 \sup_{s \leq t} |x(s)|), \text{ for some constants } \alpha,\ M_1 \text{ and } M_2 \tag{1.14}$$

$$\exists\, M < \infty : |\beta(t,x) - \beta(t,y)| \leq M \cdot \sup_{s \leq t} |x(s) - y(s)| \quad \forall x,y \in D[0,\infty). \tag{1.15}$$

Then

$$\sup_{t \in [0,T]} |\mathbf{x}_t^{(\nu)} - x_t| \xrightarrow{P} 0 \ \ as \ \ \nu \to \infty, \tag{1.16}$$

where x_t *is the solution of*

$$x_t = \int_0^t \beta(s,x)\mathrm{d}s. \tag{1.17}$$

PROOF. See KURTZ (1981, 1983). □

Hence we may conclude from Theorem 1.1 that

$$\forall \epsilon > 0 \ \lim_{\nu \to \infty} \Pr\left[\sup_{t \in [0,T]} \left|\frac{\mathbf{n}_t^{(\nu)}}{\nu} - x_t\right| > \epsilon\right] = 0. \tag{1.18}$$

In particular,

$$\mathbf{x}_t^{(\nu)} = \frac{\mathbf{n}_t^{(\nu)}}{\nu} \xrightarrow{P} x_t \ \ (\nu \to \infty). \tag{1.19}$$

To conclude this section we will give the correct expression for the likelihood function. AALEN (1978) showed using Radon-Nikodym derivatives that the likelihood function for parameters of the intensity of a counting process $\mathbf{n}$ observed on $[0,t]$ is given by

$$\mathbf{L}_t = \left[\prod_{i=1}^{\mathbf{n}_t} \boldsymbol{\lambda}_{\mathbf{T}_i}\right] \exp\left[-\int_0^t \boldsymbol{\lambda}_s \mathrm{d}s\right], \tag{1.20}$$

where $\boldsymbol{\lambda}_t$ is the intensity process of $\mathbf{n}_t$ with respect to $\mathcal{N}_t = \sigma\{\mathbf{n}_s ; s \leq t\}$ and where $\mathbf{T}_i$ is the i'th failure time. From (1.20) we can derive another expression for the likelihood function:

$$\mathbf{L}_t = \exp\left[-\int_0^t \boldsymbol{\lambda}_s ds + \int_0^t \log \boldsymbol{\lambda}_s \mathrm{d}\mathbf{n}_s\right]. \tag{1.21}$$

Suppose $\boldsymbol{\lambda}_t = \boldsymbol{\lambda}_{t:\theta}$ for some parameter $\theta \in \Theta \subset \mathbb{R}^p$ and $\mathrm{d}\mathbf{m}_{t:\theta} = \mathrm{d}\mathbf{n}_t - \boldsymbol{\lambda}_{t:\theta}\mathrm{d}t$. Then a simple calculation shows that under regularity conditions

$$\frac{\partial}{\partial\theta} \log \mathbf{L}_t = \int_0^t \left[\frac{\partial}{\partial\theta} \log \boldsymbol{\lambda}_{s:\theta}\right] \mathrm{d}\mathbf{m}_{s:\theta} \tag{1.22}$$

$$\frac{\partial^2}{\partial\theta^2} \log \mathbf{L}_t = \int_0^t \left[\frac{\partial}{\partial\theta} \log \boldsymbol{\lambda}_{s:\theta}\right]^2 \boldsymbol{\lambda}_{s:\theta}\mathrm{d}s + \int_0^t \left[\frac{\partial^2}{\partial\theta^2} \log \boldsymbol{\lambda}_{s:\theta}\right] \mathrm{d}\mathbf{m}_{s:\theta} \tag{1.23}$$

Since $\frac{\partial}{\partial\theta}\log \lambda_{s;\theta}$ is predictable, this shows that $\frac{\partial}{\partial\theta}\log \mathbf{L}_t$ is a martingale, with

$$<\frac{\partial}{\partial\theta}\log \mathbf{L}>_t = \int_0^t \left[\frac{\partial}{\partial\theta}\log \lambda_{s;\theta}\right]^2 \lambda_{s;\theta}\mathrm{d}s \,. \tag{1.24}$$

In Sections 3, 4 and 5 we will use this property to prove asymptotic normality of the parameter-estimators.

Borgan (1984) has formulated two theorems (p. 6 Theorem 1, p. 8 Theorem 2) to prove consistency and asymptotic normality of ML-estimators in counting process-models with multiplicative intensities; that is, $\lambda_{t;\theta}^{(\nu)} = \alpha_{t;\theta}\cdot \mathbf{Y}_t^{(\nu)}$, where $\mathbf{Y}_t$ is a predictable process and $\alpha_{t;\theta}$ is a deterministic function. Unfortunately, in the three models we consider, the intensities are not of this type. Nevertheless, it seems possible to formulate a general theorem giving asymptotic normality (when we assume consistency) of the ML-estimators in counting process models with intensities such as we are dealing with: $\lambda_{t;\theta}^{(\nu)} = \nu\beta(t,\mathbf{x}^{(\nu)};\theta)$. The result to be expected is

$$\sqrt{\nu}\,(\hat{\theta}^{(\nu)} - \theta_0) \xrightarrow{\mathfrak{D}} \mathfrak{N}(0,\Sigma^{-1})\,, \tag{1.25}$$

where

$$\Sigma = \int_0^T (\frac{\partial}{\partial\theta}\log \beta(t,x;\theta_0))^{\otimes 2}\beta(t,x;\theta_0)\mathrm{d}t \tag{1.26}$$

and x is the solution of (1.17). (For a column vector a, $a^{\otimes 2}$ is the matrix $a\cdot a^{\mathrm{T}}$). θ_0 is the vector of true model parameters.

Remark. Obtaining a convenient set of regularity conditions could be a useful future research project.

2. General introduction to the three software reliability models

We have already mentioned in the introduction that the three models to be considered are so called error-counting and debugging models. In the Musa-model and the Littlewood-model the faults are corrected with probability one immediately when they produce a failure. This in contrast to the Goel and Okumoto-model; here a fault detected on account of a failure having occurred may not be removed and as a result may cause additional failures at a later stage. The three models are supposed to be applicable under the following general assumptions:

- faults produce failures; each failure is observed;
- no *new* faults are introduced during the repairing of detected faults;
- the test-inputs are selected randomly from the input-set and the test-inputs are representative for the operational inputs;
- software execution is stopped when a failure occurs.

For a discussion on these assumptions we refer to Moek (1984), p. 1-2.

In the following, we mean by $\mathbf{T}_i$ the failure time of the i'th occuring failure, while $\mathbf{t}_i = \mathbf{T}_i - \mathbf{T}_{i-1}$ denotes the interfailure time; that is the time between the i'th and $(i-1)$'th failure. Recall that at time τ, $\mathbf{n}_\tau$ failures have been observed. When we define $\tilde{\mathbf{t}}_i$ as $(\mathbf{T}_i\wedge\tau) - (\mathbf{T}_{i-1}\wedge\tau)$, then $\tilde{\mathbf{t}}_i$ equals $\mathbf{t}_i$ for $i = 1, \ldots, \mathbf{n}_\tau$ and $\tilde{\mathbf{t}}_{\mathbf{n}_\tau+1}$ denotes the time between the last failure time $\mathbf{T}_{\mathbf{n}_\tau}$ and the stopping time τ. We will now give a brief specification of the models.

In the *Musa-model*, when a failure occurs the corresponding fault is completely removed, the failure rate of the program is proportional to the number of remaining faults and each fault makes the same contribution to the failure rate of the program. So if $(i-1)$ faults have already been detected, the failure rate for the i'th failure λ_i becomes

$$\lambda_i = \phi_0(N - (i-1))\,, \tag{2.1}$$

where ϕ_0 is the true failure rate per fault or the occurrence rate and N is the true number of faults in the program. In statistical terms: we observe the order statistics (smaller than or equal to τ) of a sample of N independent exponentially distributed random variables with parameter ϕ_0, where N and ϕ_0 are unknown. In terms of counting processes we can write

$$\boldsymbol{\lambda}_t = \phi_0(N - \mathbf{n}_{t-}) , \tag{2.2}$$

where $\boldsymbol{\lambda}_t$ denotes the failure rate at time t and where $\mathbf{n}_{t-}$ denotes the number of detected faults up to t. The interfailure times are independent and exponentially distributed with parameter λ_i,

$$f(t_i \mid \phi_0, N) = \phi_0(N-i+1)\exp\{-\phi_0(N-i+1)t_i\} .$$

In the *Goel and Okumoto-model* the failures occur according to a non-homogeneous Poisson-process with failure rate $\lambda(t) = N \cdot \lambda_0 \cdot e^{\phi_0 t}$. Notice that $\lambda(t)$ does not depend on the process $\mathbf{n}_t$; it is a simple deterministic function of time. One can check that the expected number of failures in $[0,\infty)$ equals $E(\int_0^\infty \lambda(s)ds) = N$. Thus we have N faults or sources of failures, each producing failures at an exponentially decreasing rate.

The main difference in the *Littlewood-model* with respect to the previous two models is the fact that each fault does not make the same contribution to the failure rate $\boldsymbol{\lambda}_t$. Littlewood's argument for that is that larger faults will produce failures earlier then smaller ones. He treats ϕ_j, the failure rate for fault number j as a stochastic variable $\boldsymbol{\phi}_j$. Notice that we have numbered the N faults arbitrarily, so do not confuse fault number j with the i'th occurring failure.

Littlewood's model is an empirical Baysian model and he himself suggests a gamma distribution $\Gamma(a,b)$ for the a-priori probability distribution of $\boldsymbol{\phi}_j$ for all j. In section 5 we will show that the failure rate of the program at time t is then given by

$$\boldsymbol{\lambda}_t = \frac{a(N-\mathbf{n}_{t-})}{b+t} . \tag{2.4}$$

So as in the Musa-model, $\boldsymbol{\lambda}_t$ depends on the past of the counting process $\mathbf{n}_t$.

A more extended model we will not consider in this paper, is the *Poisson-Gamma model.* Here, besides the occurrence rates $\boldsymbol{\phi}_j$, the total number of faults is also treated as a random variable: $\mathbf{N}$. In this model $\mathbf{N}$ and $\boldsymbol{\phi}_j$, $j=1,\ldots,\mathbf{N}$ are assumed to be independent at time zero and often a Poisson distribution is used for the a-priori distribution of $\mathbf{N}$. For a complete description of this model see MOEK (1983), SPREIJ and KOCH (1983).

In Figures 2.1, 2.2 and 2.3 we have illustrated the failure rates of the three models as a function of time. Again $\mathbf{T}_i$ denotes the time of the i'th occurring failure.

$$\boldsymbol{\lambda}_t = \phi_0(N - \mathbf{n}_{t-}) I\{\mathbf{n}_{t-} \leqslant N\}$$

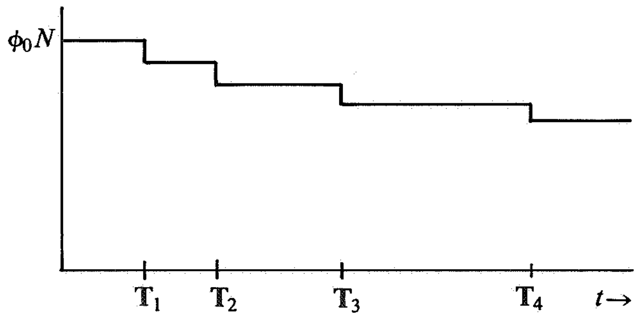

FIGURE 2.1. MUSA

$$\lambda_t = \phi_0 N e^{-\phi_0 t}$$

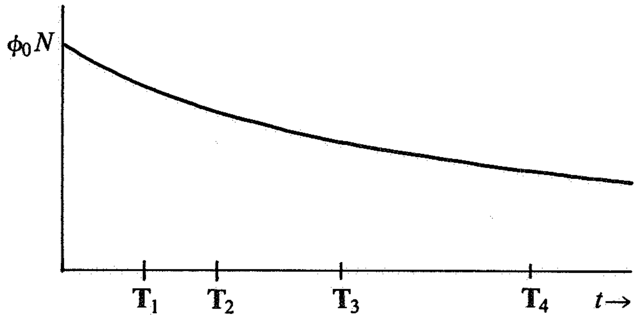

FIGURE 2.2. GOEL AND OKUMOTO

$$\lambda_t = \frac{a(N - \mathbf{n}_{t-})}{b+t} I\{\mathbf{n}_{t-} \leqslant N\}$$

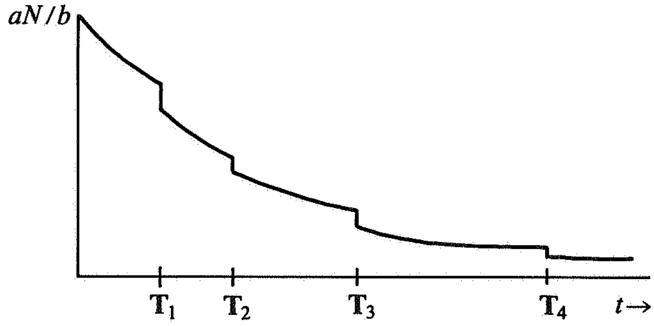

FIGURE 2.3. LITTLEWOOD

To ensure the failure rates are non-negative we have added an indicator-function $I\{\mathbf{n}_{t-} \leqslant N\}$ to the failure rates of the Musa-model and the Littlewood-model. So when all faults have been detected ($t \geqslant \mathbf{T}_N$) the intensities in the Musa and Littlewood model become zero. This is exactly how we want the intensity to act.

In the Goel and Okumoto-model the failure rate never becomes zero! This is supposed to reflect the fact that a detected error may not be removed and may cause additional errors. However, the exponential hazard rate is completely arbitrary and we feel that the Musa and Littlewood models are much more realistic then the Goel and Okumoto model. Perhaps Goel and Okumoto's model should be considered as an easily analysable approximation to Musa's.

In the Musa-model and the model of Goel and Okumoto we have to estimate two parameters: ϕ_0 and N. This in contrast with the Littlewood model where we have to estimate three parameters; the shape and scale parameters of the a priori gamma distribution (a, shape, and b, 1/scale) and N, the true number of faults. From these, one can determine in the Musa and Littlewood models estimates of the remaining number of faults. In the Goel and Okumoto-model one can determine an estimate of the expected number of failures still to occur ($\hat{N} e^{-\hat{\phi}\tau}$, where $\hat{N}$ and $\hat{\phi}$ are the MLE's of N and ϕ).

All models assume that $\lambda_t^{(\nu)}$ is decreasing in t. If the observed rate of failures does not actually decrease over $[0,\tau]$ in some sense, maximum likelihood estimates will fail to exist. Under each model, we will show that the estimates do exist with probability converging to 1 as $\nu \rightarrow \infty$.

3. THE MUSA-MODEL

In the Musa-model the failure rate of the program at time t is given by

$$\lambda_t = \phi_0(N - \mathbf{n}_{t-}) . \tag{3.1}$$

So we can write down the likelihood function using (1.21) from Section 1:

$$L(N, \phi_0 \mid \tilde{\mathbf{t}}_i, i=1,\ldots,\mathbf{n}_{\tau+1}) = \exp\{-\int_0^\tau \phi_0(N-\mathbf{n}_{s-})ds + \int_0^\tau \log \phi_0(N-\mathbf{n}_{s-})d\mathbf{n}_s\} \tag{3.2}$$

or equivalently

$$L(N, \phi_0 \mid \tilde{\mathbf{t}}_i, i=1,\ldots,\mathbf{n}_{\tau+1}) = \exp\{-\sum_{i=1}^{\mathbf{n}_\tau+1} \phi_0(N-i+1)\mathbf{t}_i + \sum_{i=1}^{\mathbf{n}_\tau} \log \phi_0(N-i+1)\} . \tag{3.3}$$

Note that $\mathbf{n}_s : s \leq \tau$ is equivalent to $\mathbf{t}_1, \ldots, \mathbf{t}_{\mathbf{n}_\tau}, \tilde{\mathbf{t}}_{\mathbf{n}_\tau+1}$, where $\mathbf{t}_i$, $i=1,\ldots,\mathbf{n}_\tau$ is the time between the i'th and $(i-1)$'th failure time and $\tilde{\mathbf{t}}_{\mathbf{n}_\tau+1}$ is the time between the stopping time τ and the $\mathbf{n}_\tau$'th failure time. The likelihood function in (3.3) is the same likelihood function as what one would intuitively have expected:

$$L(N, \phi_0 \mid \mathbf{t}_1, \ldots, \mathbf{t}_{\mathbf{n}_\tau}, \tilde{\mathbf{t}}_{\mathbf{n}_\tau+1}) = (1 - F_{\mathbf{n}_\tau+1}(\tilde{\mathbf{t}}_{\mathbf{n}_\tau+1} \mid \phi_0, N)) \prod_{i=1}^{\mathbf{n}_\tau} f_i(\mathbf{t}_i \mid \phi_0, N) \tag{3.4}$$

where in the Musa-model

$$f_i(\mathbf{t}_i \mid \phi_0, N) = \phi_0(N-i+1)\exp\{-\phi_0(N-i+1)\mathbf{t}_i\} , \tag{3.5}$$

and F_i is the corresponding distribution function. Thus $(1-F_{\mathbf{n}_\tau+1}(\tilde{\mathbf{t}}_{\mathbf{n}_\tau+1} \mid \phi_0, N))$ equals the probability that the $(\mathbf{n}_\tau+1)$'th failure has not occurred yet (i.e. at elapsed time $\mathbf{t}_{\mathbf{n}_\tau+1}$ since the previous failure). The logarithm of (3.3) becomes

$$\log L(N, \phi_0 \mid \tilde{\mathbf{t}}_i, i=1,\ldots,\mathbf{n}_\tau+1) = -\sum_{i=1}^{\mathbf{n}_\tau+1} \phi_0(N-i+1)\,\tilde{\mathbf{t}}_i + \sum_{i=1}^{\mathbf{n}_\tau} \log \phi_0(N-i+1) . \tag{3.6}$$

Hence, the likelihood equations become

$$\frac{\partial}{\partial N} \log \mathbf{L} \mid_{\hat{\phi},\hat{\mathbf{N}}} = -\sum_{i=1}^{\mathbf{n}_\tau+1} \hat{\phi}\tilde{\mathbf{t}}_i + \sum_{i=1}^{\mathbf{n}_\tau} \frac{1}{\hat{\mathbf{N}}-i+1} = 0 \tag{3.7}$$

$$\frac{\partial}{\partial N} \log \mathbf{L} \mid_{\hat{\phi},\hat{\mathbf{N}}} = -\sum_{i=1}^{\mathbf{n}_\tau+1} (\hat{\mathbf{N}}-i+1)\tilde{\mathbf{t}}_i + \frac{\mathbf{n}_\tau}{\hat{\phi}} = 0 . \tag{3.8}$$

where $\hat{\phi}$ and $\hat{\mathbf{N}}$ are the maximum likelihood estimators of respectively ϕ_0 and N. Note that we let N vary continuously, where in fact it is a discrete variable. From (3.8) we get

$$\hat{\phi} = \frac{\mathbf{n}_\tau}{\sum_{i=1}^{\mathbf{n}_\tau+1} (\hat{\mathbf{N}}-i+1)\tilde{\mathbf{t}}_i} . \tag{3.9}$$

Using (3.9), (3.7) now becomes

$$\mathbf{f}(\hat{\mathbf{N}}) \stackrel{\text{def}}{=} \frac{-\mathbf{n}_\tau}{\hat{\mathbf{N}} - \frac{\sum_{i=1}^{\mathbf{n}_\tau+1} (i-1)\tilde{\mathbf{t}}_i}{\tau}} + \sum_{i=1}^{\mathbf{n}_\tau} \frac{1}{\hat{\mathbf{N}}-i+1} = 0 , \tag{3.10}$$

since

$$\sum_{i=1}^{\mathbf{n}_\tau+1} \tilde{\mathbf{t}}_i = \tau \, .$$

So $\hat{\mathbf{N}}$ can be calculated numerically by (3.10) and afterwards $\hat{\phi}$ can be found using (3.9). We will now prove consistency and asymptotic normality of the estimators $\hat{\phi}$ and $\hat{\mathbf{N}}$ in this model. In both proofs we will apply Theorem 1 of Section 1. Let us therefore write $N=\nu\beta_0$, where β_0 is the true dummy parameter, $\hat{\mathbf{N}}^{(\nu)}=\nu\hat{\boldsymbol{\beta}}^{(\nu)}$, $\boldsymbol{\lambda}_t^{(\nu)}=\phi_0(N-\mathbf{n}_t^{(\nu)})$, where $\hat{\boldsymbol{\beta}}^{(\nu)}$ and $\hat{\mathbf{N}}^{(\nu)}$ are the ML-estimators as defined by (3.9) and (3.10) of β_0 and N in the ν'th experiment, if these exist.

THEOREM 3.1. *In the Musa-model*

$$\sup_{t\in[0,\tau]} \left| \frac{\mathbf{n}_t^{(\nu)}}{\nu} - \beta_0(1-e^{-\phi_0 t}) \right| \xrightarrow{P} 0 \text{ as } \nu\to\infty \, , \tag{3.11}$$

where β_0 and ϕ_0 are the true model parameters.

PROOF. Let in Theorem 1.1

$$\beta(t,x)=\phi_0(\beta_0-x(t-))\cdot I\{0\leqslant x(t-)\leqslant\beta_0\} \, . \tag{3.12}$$

(Recall $\mathbf{x}_t^{(\nu)}=\dfrac{\mathbf{n}_t^{(\nu)}}{\nu}$, so $\mathbf{x}_0^{(\nu)}=0$). Notice that in our experiment $x(t-)$ is never outside $[0,\beta_0]$, so in fact the indicator function in (3.12) is irrelevant. We will now verify conditions (1.14) and (1.15) of Theorem 1.1. We have

$$\sup_{s\leqslant t} \beta(s,x)\leqslant \sup_{s\leqslant t} | \beta(s,x) | \leqslant \sup_{s\leqslant t} | \phi_0 | \cdot | \beta_0-x(s-) | \leqslant | \phi_0 | \cdot | \beta_0 | <\infty \, .$$

$| \phi_0 |$ and $| \beta_0 |$ are finite so (1.14) holds. Secondly, let x, $y\in D[0,\infty)$, then

$$\begin{aligned} | \beta(t,x)-\beta(t,y) | &= | \phi_0(\beta_0-x(t-))-\phi_0(\beta_0-y(t-)) | = \\ &= | \phi_0(y(t-)-x(t-)) | \leqslant | \phi_0 | \cdot \sup_{s\leqslant t} | y(s-)-x(s-) | \, . \end{aligned}$$

Therefore (1.15) holds. So $\sup_{t\in[0,\tau]} \left| \dfrac{\mathbf{n}_t^{(\nu)}}{\nu} - x_t \right| \xrightarrow{P} 0$ $(\nu\to\infty)$, where x_t is the solution of $x_t=\int_0^t \phi_0(\beta_0-x_{s-})ds$, that is,

$$x_t=\beta_0(1-e^{-\phi_0 t}) \, . \tag{3.13}$$

□

We see that x_t is an increasing function of time and that $\lim_{t\to\infty} x_t = \lim_{t\to\infty}\beta_0(1-e^{-\phi_0 t})=\beta_0$, as was to be expected. An immediate consequence of Theorem 3.1 is the following corollary.

COROLLARY 3.1. *In the Musa-model*

$$\frac{\mathbf{n}_t^{(\nu)}}{\nu} \xrightarrow{P} \beta_0(1-e^{-\phi_0 t}) \text{ as } \nu\to\infty \, . \tag{3.14}$$

THEOREM 3.2. *In the Musa-model, let $\theta_0=\begin{bmatrix}\beta_0\\ \phi_0\end{bmatrix}$ be the 2×1 vector of true model-parameters and let $\hat{\boldsymbol{\theta}}^{(\nu)}=\begin{bmatrix}\hat{\boldsymbol{\beta}}^{(\nu)}\\ \hat{\boldsymbol{\phi}}^{(\nu)}\end{bmatrix}$ the maximum likelihood estimator of θ_0 in the ν'th experiment. Then*

$$\hat{\boldsymbol{\theta}}^{(\nu)} \xrightarrow{\mathrm{P}} \theta_0 \quad (\nu \to \infty)\,.$$

That is, $\hat{\boldsymbol{\theta}}^{(\nu)}$ is a consistent estimator of θ_0.

An equivalent to (3.10) is, after some calculations,

$$\mathrm{U}^{(\nu)}(\hat{\boldsymbol{\beta}}^{(\nu)}) \stackrel{\text{def}}{=} \int_0^T \frac{1}{\hat{\boldsymbol{\beta}}^{(\nu)} - \frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}}\, \mathrm{d}\frac{\mathbf{n}_t^{(\nu)}}{\nu} - \frac{\mathbf{n}_t^{(\nu)}/\nu}{\hat{\boldsymbol{\beta}}^{(\nu)} - \frac{1}{T}\int_0^T \frac{\mathbf{n}_t^{(\nu)}}{\nu}\mathrm{d}t} = 0\,. \tag{3.15}$$

Here is the idea of the proof of Theorem 3.2: we will show that $\mathrm{U}^{(\nu)}(\beta)$ defined by (3.15) has a unique root $\hat{\boldsymbol{\beta}}^{(\nu)}$, at least with probability approaching to 1 as $\nu \to \infty$. Then we will define another continuous function $U(\beta)$, whose only zero is the true model parameter β_0. Finally we will show that for each fixed β, $\mathrm{U}^{(\nu)}(\beta)$ will converge in probability to $U(\beta)$. Recall (3.15); let

$$\mathrm{U}^{(\nu)}(\beta) \stackrel{\text{def}}{=} \int_0^T \frac{1}{\beta - \frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}}\, \mathrm{d}\frac{\mathbf{n}_t^{(\nu)}}{\nu} - \frac{\mathbf{n}_T^{(\nu)}/\nu}{\beta - \frac{1}{T}\int_0^T \frac{\mathbf{n}_t^{(\nu)}}{\nu}\mathrm{d}t} \tag{3.16}$$

$$= \frac{1}{\nu}\left[\sum_{i=1}^{\mathbf{n}_T^{(\nu)}} \frac{1}{\beta - \frac{(i-1)}{\nu}} - \frac{\mathbf{n}_T^{(\nu)}}{\beta - \frac{c(\mathbf{n})}{\nu}}\right], \tag{3.17}$$

where $c(\mathbf{n}) \stackrel{\text{def}}{=} \frac{1}{T}\sum_{i=1}^{\mathbf{n}_\tau+1} (i-1)\tilde{\mathbf{t}}_i = \frac{1}{T}\int_0^T \mathbf{n}_t^{(\nu)}\mathrm{d}t$. Notice that $\mathrm{U}^{(\nu)}(\beta)$ has only been defined for $\beta > \frac{\mathbf{n}_T^{(\nu)} - 1}{\nu}$.

LEMMA 3.1. *If $\mathrm{U}^{(\nu)}(\beta) = 0$ has a solution, it is unique.*

PROOF. We will show that at each zero, $\mathrm{U}^{(\nu)}(\beta)$ has a negative derivative

$$\frac{\partial}{\partial\beta}\mathrm{U}^{(\nu)}(\beta) = \frac{1}{\nu}\left[\frac{\mathbf{n}_T^{(\nu)}}{\left[\beta - \frac{c(\mathbf{n})}{\nu}\right]^2} - \sum_{i=1}^{\mathbf{n}_T^{(\nu)}} \frac{1}{\left[\beta - \frac{(i-1)}{\nu}\right]^2}\right]. \tag{3.18}$$

Using $\mathrm{U}^{(\nu)}(\beta) = 0$, (3.18) becomes

$$\frac{\partial}{\partial\beta}\mathrm{U}^{(\nu)}(\beta) = \frac{1}{\nu}\left[\frac{1}{\mathbf{n}_T^{(\nu)}}\left[\sum_{i=1}^{\mathbf{n}_T^{(\nu)}} \frac{1}{\beta - \frac{(i-1)}{\nu}}\right]^2 - \sum_{i=1}^{\mathbf{n}_T^{(\nu)}} \left[\frac{1}{\beta - \frac{(i-1)}{\nu}}\right]^2\right]. \tag{3.19}$$

The right-hand side of (3.19) is non-positive, since

$$n \cdot \sum_{i=1}^{n} x_i^2 \geqslant \left[\sum_{i=1}^{n} x_i\right]^2 \tag{3.20}$$

for an arbitrary sequence x_i, $i = 1, \ldots, n$. The $\geqslant$-sign can be replaced by the $>$-sign if not all x_i are equal. So

$$\frac{\partial}{\partial\beta}\mathrm{U}^{(\nu)}(\beta) < 0 \quad \text{at zeros of } \mathrm{U}^{(\nu)}(\beta) \text{ if } \mathbf{n}_\tau > 1\,. \tag{3.21}$$

The same argument shows that $\frac{\partial}{\partial\beta}U^{(\nu)}(\beta)\leqslant 0$, whenever $U^{(\nu)}(\beta)>0$:

$$U^{(\nu)}(\beta)>0 \Rightarrow 0\leqslant \frac{\mathbf{n}_T^{(\nu)}}{\beta-\frac{c(\mathbf{n})}{\nu}} < \sum_{i=1}^{\mathbf{n}_T^{(\nu)}} \frac{1}{\beta-\frac{(i-1)}{\nu}} . \tag{3.22}$$

Therefore

$$\frac{\mathbf{n}_T^{(\nu)}}{\left[\beta-\frac{c(\mathbf{n})}{\nu}\right]^2} < \frac{1}{\mathbf{n}_T^{(\nu)}} \left[\sum_{i=1}^{\mathbf{n}_T^{(\nu)}} \frac{1}{\beta-\frac{(i-1)}{\nu}}\right]^2 . \tag{3.23}$$

Hence

$$\frac{\partial}{\partial\beta}U^{(\nu)}(\beta)=\frac{1}{\nu}\left[\frac{\mathbf{n}_T^{(\nu)}}{\left[\beta-\frac{c(\mathbf{n})}{\nu}\right]^2} - \sum_{i=1}^{\mathbf{n}_T^{(\nu)}} \frac{1}{\left[\beta-\frac{(i-1)}{\nu}\right]^2}\right] \tag{3.24}$$

$$<\frac{1}{\nu}\left[\frac{1}{\mathbf{n}_T^{(\nu)}}\left[\sum_{i=1}^{\mathbf{n}_T^{(\nu)}} \frac{1}{\beta-\frac{(i-1)}{\nu}}\right]^2 - \sum_{i=1}^{\mathbf{n}_T^{(\nu)}} \left[\frac{1}{\beta-\frac{(i-1)}{\nu}}\right]^2\right] \leqslant 0 .$$

As a consequence of the fact that $U^{(\nu)}(\beta)$ is continuous in β, we can conclude that $U^{(\nu)}(\beta)$ has at most one solution. □

Moek (1984) developed a criterion for the existence of a solution of $U^{(\nu)}(\beta)=0$:

$$\text{There is a unique solution if and only if } c(\mathbf{n})>\frac{\mathbf{n}_T^{(\nu)}-1}{2} . \tag{3.25}$$

To make this intuitively likely, note that for large values of ν, $U^{(\nu)}(\beta)$ can be approximated by

$$U^{(\nu)}(\beta) \approx \frac{\mathbf{n}_T^{(\nu)}}{\nu^2\beta^2}\left[\frac{\mathbf{n}_T^{(\nu)}-1}{2}-c(\mathbf{n})\right] . \tag{3.26}$$

In Figures 3.1 and 3.2 we have illustrated this criterion.

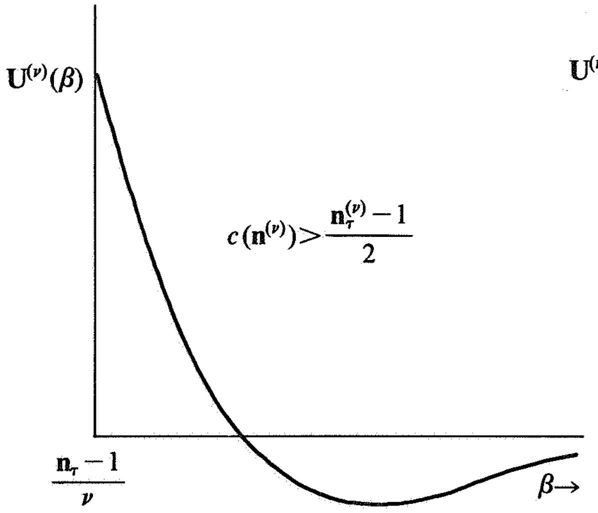

FIGURE 3.1.

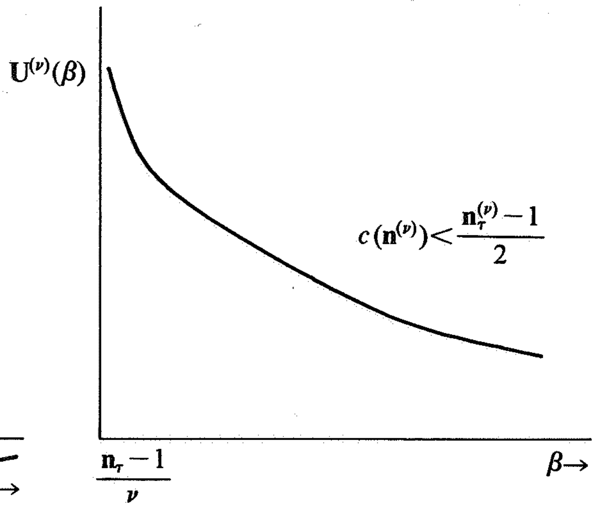

FIGURE 3.2.

Now define $U(\beta)$ as

$$U(\beta)=\int_0^T \frac{1}{\beta-x_t}\mathrm{d}x_t - \frac{x_T}{\beta-\frac{1}{T}\int_0^T x_t\,\mathrm{d}t}, \quad \beta>x_T>\frac{1}{T}\int_0^T x_t\,\mathrm{d}t, \tag{3.27}$$

where $x_t=\beta_0(1-e^{-\phi_0 t})$, β_0 and ϕ_0 the true model parameters.

LEMMA 3.2. *β_0 is the unique root of $U(\beta)=0$.*

PROOF. For simplicity let $\bar{x}_T \stackrel{\text{def}}{=} \frac{1}{T}\int_0^T x_t\,\mathrm{d}t = c(x)$

$$\int_t^T \frac{1}{\beta_0-x_t}\mathrm{d}x_t = -[\ln(\beta_0-x_t)]_0^T = \phi_0\cdot T$$

and

$$\frac{x_T}{\beta_0-\bar{x}_T} = \frac{x_T}{\beta_0-\beta_0+\frac{\beta_0}{\phi_0 T}(1-e^{-\phi_0 T})} = \phi_0\cdot T\,.$$

So indeed $U(\beta_0)=0$. Consider

$$\frac{\partial}{\partial\beta}U(\beta) = -\int_0^T \frac{1}{(\beta-x_t)^2}\mathrm{d}x_t + \frac{x_t}{(\beta-\bar{x}_T)^2}\,. \tag{3.28}$$

So we have for a zero of $U(\beta)$ (use $U(\beta)=0$)

$$\frac{\partial}{\partial\beta}U(\beta) = -\int_0^T \frac{1}{(\beta-x_t)^2}\mathrm{d}x_t + \frac{1}{x_T}\left(\int_0^T \frac{1}{\beta-x_t}\mathrm{d}x_t\right)^2\,. \tag{3.29}$$

We can apply Jensen's inequality for a stochastic variable $\mathbf{x}$:

$$\mathrm{E}(\mathbf{x}^2) \leqslant \mathrm{E}^2(\mathbf{x})\,. \tag{3.30}$$

In our case

$$\int_0^T \frac{1}{(\beta-x_t)^2}\,\mathrm{d}\left(\frac{x_t}{x_T}\right) \leqslant \left(\int_0^T \frac{1}{\beta-x_t}\,\mathrm{d}\left(\frac{x_t}{x_T}\right)\right)^2 . \tag{3.31}$$

So indeed the righthand side of (3.29) is non-positive and so $\frac{\partial}{\partial\beta}U(\beta)\leqslant 0$ whenever $U(\beta)=0$. In the same way we can show that $\frac{\partial}{\partial\beta}U(\beta)\leqslant 0$, whenever $U(\beta)>0$. We know $\lim_{\beta\to\infty} U(\beta)=0$ and $\lim_{\beta\downarrow x_T} U(\beta)=+\infty$. So using the continuity of $U(\beta)$, we have proved the desired result. □

In Figure 3.3 we have illustrated $U(\beta)$.

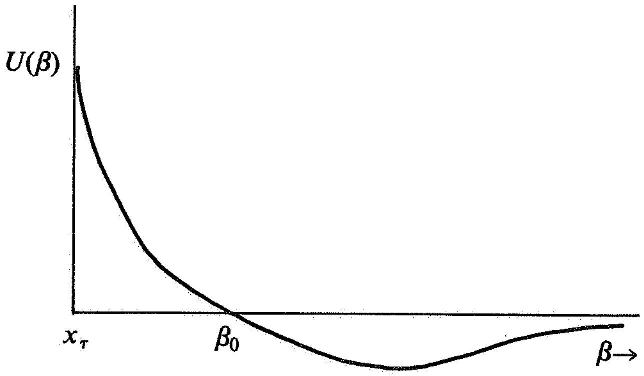

FIGURE 3.3.

LEMMA 3.3. *Let* $\mathbf{U}^{(\nu)}(\beta)$ *and* $U(\beta)$ *be as defined in (3.16) and (3.27). Then*

$$\mathbf{U}^{(\nu)}(\beta) \xrightarrow{\mathrm{P}} U(\beta) \quad (\nu\to\infty)$$

for each fixed $\beta>x_T$. PROOF.

$$\Pr\left[\,|\mathbf{U}^{(\nu)}(\beta)-U(\beta)\,|>\epsilon\right]=\Pr\left[\left|\int_0^T \frac{1}{\beta-\mathbf{x}_{t-}^{(\nu)}}\mathrm{d}\mathbf{x}_t^{(\nu)} - \int_0^T \frac{1}{\beta-x_t}\mathrm{d}x_t + \frac{\mathbf{x}_T^{(\nu)}}{\beta-\bar{\mathbf{x}}_T^{(\nu)}} - \frac{x_T}{\beta-\bar{x}_T}\right|>\epsilon\right], \tag{3.32}$$

where $\mathbf{x}_t^{(\nu)}=\frac{\mathbf{n}_t^{(\nu)}}{\nu}$ in (3.17). We know that

$$\begin{aligned}\{\omega:\ |\,\mathbf{U}^{(\nu)}(\beta)-U(\beta)|>\epsilon\}\subseteq\{\omega:\ &\left|\int_0^T \frac{1}{\beta-\mathbf{x}_{t-}^{(\nu)}}\mathrm{d}\mathbf{x}_t^{(\nu)}-\int_0^T\frac{1}{\beta-x_{t-}}\mathrm{d}\mathbf{x}_t^{(\nu)}\right|>\frac{\epsilon}{3}\}\\ \cup\{\omega:\ &\left|\int_0^T \frac{1}{\beta-x_{t-}}\mathrm{d}\mathbf{x}_t^{(\nu)}-\int_0^T\frac{1}{\beta-x_{t-}}\mathrm{d}x_t\right|>\frac{\epsilon}{3}\}\\ \cup\{\omega:\ &\left|\frac{\mathbf{x}_T^{(\nu)}}{\beta-\bar{\mathbf{x}}_T^{(\nu)}}-\frac{x_T}{\beta-\bar{x}_T}\right|>\frac{\epsilon}{3}\}.\end{aligned} \tag{3.33}$$

So

$$\Pr\left[\,|\,\mathbf{U}^{(\nu)}(\beta)-U(\beta)|>\epsilon\right]\leqslant\Pr\left[\,\left|\int_0^T \frac{1}{\beta-\mathbf{x}_{t-}^{(\nu)}}\mathrm{d}\mathbf{x}_t^{(\nu)}-\int_0^T\frac{1}{\beta-x_{t-}}\mathrm{d}\mathbf{x}_t^{(\nu)}\right|>\frac{\epsilon}{3}\right]$$

$$+\Pr\left[\ \left|\int_0^\tau \frac{1}{\beta-x_{t-}}\mathrm{d}\mathbf{x}_t^{(\nu)}-\int_0^\tau \frac{1}{\beta-x_{t-}}\mathrm{d}x_t\ \right|>\frac{\epsilon}{3}\right] \tag{3.34}$$

$$+\Pr\left[\ \left|\frac{\mathbf{x}_\tau^{(\nu)}}{\beta-\bar{\mathbf{x}}_\tau^{(\nu)}}-\frac{x_\tau}{\beta-\bar{x}_\tau}\ \right|>\frac{\epsilon}{3}\right].$$

Look at the first term on the righthand side of (3.34)

$$\left|\int_0^\tau \frac{1}{\beta-\mathbf{x}_{t-}^{(\nu)}}\mathrm{d}\mathbf{x}_t^{(\nu)}-\int_0^\tau \frac{1}{\beta-x_{t-}}\mathrm{d}\mathbf{x}_t^{(\nu)}\right| \leqslant \sup_{t\in[0,\tau]}\left|\frac{1}{\beta-\mathbf{x}_{t-}^{(\nu)}}-\frac{1}{\beta-x_{t-}}\right|\cdot \mathbf{x}_\tau^{(\nu)}\xrightarrow{P} 0,$$

because $\mathbf{x}_\tau^{(\nu)}\xrightarrow{P} x_\tau=\beta_0(1-e^{-\phi_0\tau})<\infty$ as a consequence of Corollary 3.1 and

$$0\leqslant \sup_{t\in[0,\tau]}\left|\frac{1}{\beta-\mathbf{x}_{t-}^{(\nu)}}-\frac{1}{\beta-x_{t-}}\right| = \sup_{t\in[0,\tau]}\left|\frac{\mathbf{x}_{t-}^{(\nu)}-x_{t-}}{(\beta-\mathbf{x}_{t-}^{(\nu)})(\beta-x_{t-})}\right|$$

$$\leqslant \frac{\sup_{t\in[0,\tau]}\left|\mathbf{x}_{t-}^{(\nu)}-x_{t-}\right|}{(\beta-\mathbf{x}_\tau^{(\nu)})(\beta-x_\tau)}\xrightarrow{P} 0$$

as a consequence of Theorem 3.1 and Corollary 3.1. So

$$\lim_{\nu\to\infty}\Pr\left[\ \left|\int_0^\tau \frac{1}{\beta-\mathbf{x}_{t-}^{(\nu)}}\mathrm{d}\mathbf{x}_t^{(\nu)}-\int_0^\tau \frac{1}{\beta-x_{t-}}\mathrm{d}\mathbf{x}_t^{(\nu)}\ \right|>\frac{\epsilon}{3}\right]=0. \tag{3.35}$$

Similarly

$$\lim_{\nu\to\infty}\Pr\left[\ \left|\int_0^\tau \frac{1}{\beta-x_{t-}}\mathrm{d}\mathbf{x}_t^{(\nu)}-\int_0^\tau \frac{1}{\beta-x_{t-}}\mathrm{d}x_t\ \right|>\frac{\epsilon}{3}\right]=0, \tag{3.36}$$

because of the fact that

$$\int_0^\tau \frac{1}{\beta-x_{t-}}\mathrm{d}(\mathbf{x}_t^{(\nu)}-x_t)=(\mathbf{x}_\tau^{(\nu)}-x_\tau)\cdot\frac{1}{\beta-x_\tau}-\int_0^\tau(\mathbf{x}_t^{(\nu)}-x_t)\,\mathrm{d}\left(\frac{1}{\beta-x_t}\right). \tag{3.37}$$

The first term converges in probability to zero because $\frac{1}{\beta-x_\tau}$ is finite and $\mathbf{x}_\tau^{(\nu)}\xrightarrow{P} x_\tau$ for $\nu\to\infty$ (Corollary 3.1). The second term converges in probability to zero, since

$$\left|\int_0^\tau\left(\mathbf{x}_t^{(\nu)}-x_t\right)\mathrm{d}\left(\frac{1}{\beta-x_t}\right)\right|\leqslant\sup_{t\in[0,\tau]}\left|\mathbf{x}_t^{(\nu)}-x_t\right|\cdot\left(\frac{1}{\beta-x_\tau}-\frac{1}{\beta}\right).$$

$\left(\frac{1}{\beta-x_\tau}-\frac{1}{\beta}\right)$ is finite and $\sup_{t\leqslant\tau}\left|\mathbf{x}_t^{(\nu)}-x_t\right|\xrightarrow{P} 0$ according to Theorem 3.1. Finally

$$\lim_{\nu\to\infty}\Pr\left[\left|\frac{\mathbf{x}_\tau^{(\nu)}}{\beta-\bar{\mathbf{x}}_\tau^{(\nu)}}-\frac{x_\tau}{\beta-\bar{x}_\tau}\right|>\frac{\epsilon}{3}\right]=0, \tag{3.38}$$

as a consequence of Slutsky's theorem: define

$$g(\alpha,\bar{\alpha})=\frac{\alpha}{\beta-\bar{\alpha}} \tag{3.39}$$

then $g(\alpha,\bar{\alpha})$ is a continuous function at a point with $\bar{\alpha}<\beta$. Since

$$\mathbf{x}_\tau^{(\nu)} \xrightarrow{P} x_\tau<\beta \text{ and } \bar{\mathbf{x}}_\tau^{(\nu)} \xrightarrow{P} \bar{x}_\tau<x_\tau<\beta\,,$$

$$g(\mathbf{x}_\tau^{(\nu)},\bar{\mathbf{x}}_\tau^{(\nu)}) \xrightarrow{P} g(x_\tau,\bar{x}_\tau) \quad (\nu\to\infty) \tag{3.40}$$

and so (3.38) holds. So we can conclude $\lim_{\nu\to\infty} \Pr\left\{\left|\mathbf{U}^{(\nu)}(\beta)-U(\beta)\right|>\epsilon\right\}=0$, $\forall\epsilon>0$, for all fixed $\beta>x_\tau$ and so $\mathbf{U}^{(\nu)}(\beta) \xrightarrow{P} U(\beta)$ ($\nu\to\infty$, β fixed). □

PROOF of Theorem 3.2. We showed that $\mathbf{U}^{(\nu)}(\beta)$ has at most one solution $\hat{\boldsymbol{\beta}}^{(\nu)}$ (Lemma 3.1). Further, β_0, the true model parameter is the unique solution of $U(\beta)=0$ (Lemma 3.2). In Lemma 3.3 we showed that $\mathbf{U}^{(\nu)}(\beta)\to U(\beta)$ ($\nu\to\infty$) for each fixed $\beta>x_\tau$. Recall that $\mathbf{U}^{(\nu)}(\beta)$ was only defined for $\beta>\mathbf{x}_\tau^{(\nu)}-\frac{1}{\nu}$ and $U(\beta)$ for $\beta>x_\tau$. Now look at an interval $[\beta',\beta'']$ around β_0, where $x_\tau<\beta'<\beta_0<\beta''$. Let β' satisfy $\beta'=x_\tau+\epsilon$ where ϵ can be arbitrary small, $\epsilon>0$, then one can easily see that $\lim \Pr(\mathbf{x}_\tau^{(\nu)}<\beta')=1$. We now have as a consequence of the continuity of $\mathbf{U}^{(\nu)}(\beta)$ in β that $\lim_{\nu\to\infty}\Pr$ (there is a solution of $\mathbf{U}^{(\nu)}(\beta)=0$ in $[\beta',\beta'']=1$). Moreover this solution is to the right of $\mathbf{x}_\tau^{(\nu)}$. This solution is unique by Lemma 3.1. This leads to the desired result $\hat{\boldsymbol{\beta}}^{(\nu)} \xrightarrow{P} \beta_0$ ($\nu\to\infty$), since the interval $[\beta',\beta'']$ round β_0 can be taken arbitrary small. Now it is easy to show that $\hat{\boldsymbol{\phi}}^{(\nu)} \xrightarrow{P} \phi_0$ ($\nu\to\infty$):

$$\hat{\boldsymbol{\phi}}^{(\nu)}=\frac{\mathbf{n}_\tau^{(\nu)}}{\sum_{i=1}^{\mathbf{n}_\tau^{(\nu)}+1}\left[\nu\hat{\boldsymbol{\beta}}^{(\nu)}-\frac{(i-1)}{\nu}\right]\tilde{\mathbf{t}}_i}=\frac{\mathbf{x}_\tau^{(\nu)}}{\int_0^\tau(\hat{\boldsymbol{\beta}}^{(\nu)}-\mathbf{x}_t^{(\nu)})dt} \xrightarrow{P} \frac{x_\tau}{\int_0^\tau(\beta_0-x_t)dt}=\phi_0 \tag{3.41}$$

as a consequence of Theorem 3.1 and $\hat{\boldsymbol{\beta}}^{(\nu)} \xrightarrow{P} \beta_0$ ($\nu\to\infty$). Conclusion:

$$\hat{\boldsymbol{\theta}}^{(\nu)}=\begin{bmatrix}\hat{\boldsymbol{\beta}}^{(\nu)}\\ \hat{\boldsymbol{\phi}}^{(\nu)}\end{bmatrix} \xrightarrow{P} \begin{bmatrix}\beta_0\\ \phi_0\end{bmatrix}=\theta_0 \quad (\nu\to\infty)\,.$$ □

THEOREM 3.3. *Let again* $\hat{\boldsymbol{\theta}}^{(\nu)}=\begin{bmatrix}\hat{\boldsymbol{\beta}}^{(\nu)}\\ \hat{\boldsymbol{\phi}}^{(\nu)}\end{bmatrix}$ *be the maximum likelihood estimators in the ν'th experiment of* $\theta_0=\begin{bmatrix}\beta_0\\ \phi_0\end{bmatrix}$. *Then*

$$\sqrt{\nu}(\hat{\boldsymbol{\theta}}^{(\nu)}-\theta_0) \xrightarrow{\mathcal{D}} \mathfrak{N}(0,\Sigma_M(\theta_0)^{-1}),\quad (\nu\to\infty)$$

where (the subscript M denotes Musa)

$$\Sigma_M(\theta_0)=\begin{bmatrix}\frac{1}{\beta_0}(e^{\phi_0\tau}-1) & \tau\\ \tau & -\frac{\beta_0}{\phi_0^2}(e^{-\phi_0\tau}-1)\end{bmatrix}. \tag{3.42}$$

Proof. Consider the following Taylor-expansions

$$0=\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\hat{\theta}^{(\nu)})_i=\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0)_i+\sqrt{\nu}(\hat{\theta}^{(\nu)}-\theta_0)^{\mathrm{T}}\frac{1}{\nu}\frac{\partial^2}{\partial\theta\partial\theta_i}\log \mathbf{L}(\overline{\theta}^{(i)}),$$

where

$$|(\overline{\theta}^{(i)}-\theta_0)_j| \leq |(\hat{\theta}^{(\nu)}-\theta_0)_j| \quad i=1,2;\ j=1,2. \tag{3.43}$$

Following Cramér (1946), pp. 500-504 (cf. Billingsley (1961) Theorems 2.2 and 10.1 and Borgan (1984) Theorem 2) it is sufficient to show that

$$\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0)\xrightarrow{\mathcal{D}}\mathfrak{N}(0,\Sigma_M(\theta_0)) \quad (\nu\to\infty) \tag{3.44}$$

and

$$\frac{1}{\nu}\frac{\partial^2}{\partial\theta^2}\log \mathbf{L}(\overline{\theta})\xrightarrow{\mathrm{P}} -\Sigma_M(\theta_0) \quad (\nu\to\infty). \tag{3.45}$$

We will use martingale theory to prove (3.44) and (3.45). Let us first look at (3.44). Recall (3.6), where $N=\nu\beta_0$

$$\log \mathbf{L}(\theta_0)=-\sum_{i=1}^{\mathbf{n}_\tau^{(\nu)}+1}\phi_0(\nu\beta_0-i+1). \tag{3.46}$$

Then we get

$$\frac{\partial}{\partial\beta}\log \mathbf{L}(\theta_0) = -\sum_{i=1}^{\mathbf{n}_\tau^{(\nu)}+1}\nu\phi_0\tilde{\mathbf{t}}_i+\sum_{i=1}^{\mathbf{n}_\tau^{(\nu)}}\frac{\nu}{\nu\beta_0-i+1} = \int_0^\tau\frac{\nu}{\nu\beta_0-\mathbf{n}_{t-}^{(\nu)}}d\mathbf{m}_t^{(\nu)}\stackrel{\text{def}}{=}\int_0^\tau\mathbf{h}_1^{(\nu)}d\mathbf{m}_t^{(\nu)}, \tag{3.47}$$

where $\mathbf{m}_t^{(\nu)}=\mathbf{n}_t^{(\nu)}-\int_0^t\boldsymbol{\lambda}_s^{(\nu)}ds$ is a martingale. Similarly

$$\frac{\partial}{\partial\phi}\log \mathbf{L}(\theta_0) = -\sum_{i=1}^{\mathbf{n}_\tau^{(\nu)}+1}(\nu\beta_0-i+1)\tilde{\mathbf{t}}_i+\frac{\mathbf{n}_\tau^{(\nu)}}{\phi_0}=\int_0^\tau\frac{1}{\phi_0}d\mathbf{m}_t^{(\nu)}\stackrel{\text{def}}{=}\int_0^\tau\mathbf{h}_2^{(\nu)}d\mathbf{m}_t^{(\nu)}. \tag{3.48}$$

Notice that $\mathbf{h}_1^{(\nu)}$ and $\mathbf{h}_2^{(\nu)}$ equal $\frac{\partial}{\partial\beta}\log\boldsymbol{\lambda}_t^{(\nu)}$ and $\frac{\partial}{\partial\phi}\log\boldsymbol{\lambda}_t^{(\nu)}$, since $\boldsymbol{\lambda}_t^{(\nu)}=\nu\phi_0(\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu})$. So

$$\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\beta}\log \mathbf{L}(\theta_0)=\frac{1}{\sqrt{\nu}}\int_0^\tau\frac{1}{\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}}d\mathbf{m}_t^{(\nu)} \tag{3.49}$$

and

$$\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\phi}\log \mathbf{L}(\theta_0)=\frac{1}{\sqrt{\nu}}\int_0^\tau\frac{1}{\phi_0}d\mathbf{m}_t^{(\nu)}. \tag{3.50}$$

We will now use the martingale central limit theorem mentioned in Section 1. Therefore we have to calculate the matrix predictable covariation process

$$\left\langle\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0)\right\rangle(t)=\frac{1}{\nu}\begin{pmatrix}\int_0^t\mathbf{h}_1^{(\nu)}\mathbf{h}_1^{(\nu)}d\langle\mathbf{m}^{(\nu)}\rangle(s) & \int_0^t\mathbf{h}_1^{(\nu)}\mathbf{h}_2^{(\nu)}d\langle\mathbf{m}^{(\nu)}\rangle(s)\\ \int_0^t\mathbf{h}_1^{(\nu)}\mathbf{h}_2^{(\nu)}d\langle\mathbf{m}^{(\nu)}\rangle(s) & \int_0^t\mathbf{h}_2^{(\nu)}\mathbf{h}_2^{(\nu)}d\langle\mathbf{m}^{(\nu)}\rangle(s)\end{pmatrix}. \tag{3.51}$$

Here $\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0)$ is considered as a process, replacing τ by t in (3.49) and (3.50). Using $\mathrm{d}<\mathbf{m}^{(\nu)}>(s)=\boldsymbol{\lambda}_s^{(\nu)}\mathrm{d}s$, (3.47) and (3.48) we get

$$<\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0)>(t)=\frac{1}{\nu}\begin{bmatrix} \int_0^t \frac{\boldsymbol{\lambda}_s^{(\nu)}}{(\beta_0-\frac{\mathbf{n}_{s-}^{(\nu)}}{\nu})^2}\mathrm{d}s & \int_0^t \nu\mathrm{d}s \\ \int_0^t \nu\mathrm{d}s & \int_0^t \frac{1}{\phi_0^2}\boldsymbol{\lambda}_s^{(\nu)}\mathrm{d}s \end{bmatrix}$$

$$=\begin{bmatrix} \phi_0\int_0^t \frac{\mathrm{d}s}{\beta_0-\frac{\mathbf{n}_{s-}^{(\nu)}}{\nu}} & t \\ t & \frac{1}{\phi_0}\int_0^t(\beta_0-\frac{\mathbf{n}_{s-}^{(\nu)}}{\nu})\mathrm{d}s \end{bmatrix}$$

$$\overset{\mathrm{P}}{\to}\begin{bmatrix} \frac{1}{\beta_0}(e^{\phi_0 t}-1) & t \\ t & -\frac{\beta_0}{\phi_0^2}(e^{-\phi_0 t}-1) \end{bmatrix} \quad (\nu\to\infty)$$

$$=\Sigma_M(\theta_0) \quad \text{when } t=\tau .$$

This convergence in probability is a consequence of Theorem 3.1:

$$\sup_{t\in[0,\tau]} \mid \mathbf{x}_t^{(\nu)}-x_t \mid \overset{\mathrm{P}}{\to} 0 \quad \text{as } \nu\to\infty .$$

In order to apply the martingale central limit theorem, we hereby show that $\forall\epsilon>0$

$$\int_0^\tau \frac{\nu}{(\nu\beta_0-\mathbf{n}_{t-}^{(\nu)})^2}\,\phi_0(\nu\beta_0-\mathbf{n}_{t-}^{(\nu)})\cdot I\left\{\left|\frac{1}{\nu\beta_0-\mathbf{n}_{t-}^{(\nu)}}\right|>\epsilon\right\}\mathrm{d}t \overset{\mathrm{P}}{\to} 0 \quad (\nu\to\infty) \tag{3.53}$$

and

$$\int_0^\tau \frac{1}{\phi_0^2\cdot y}\,\phi_0(\nu\beta_0-\mathbf{n}_{t-}^{(\nu)})\, I\left\{\left|\frac{1}{\sqrt{\nu}\,\phi_0}\right|>\epsilon\right\}\mathrm{d}t \overset{\mathrm{P}}{\to} 0 \quad (\nu\to\infty). \tag{3.54}$$

The left-hand side of (3.53) equals $\int_0^\tau \frac{\phi_0}{\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}}\cdot I\left\{\left|\frac{1}{\sqrt{\nu}}\cdot\frac{1}{(\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu})}\right|>\epsilon\right\}\mathrm{d}t$. This equals 0 when $\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}>\frac{1}{\epsilon\cdot\sqrt{\nu}}$. So

$$\Pr\left[\int_0^\tau \frac{\phi_0}{\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}}\cdot I\left\{\left|\frac{1}{\sqrt{\nu}}\cdot\frac{1}{(\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu})}\right|>\epsilon\right\}\mathrm{d}t\neq 0\right]\to 0 \text{ as } \nu\to\infty .$$

In the same way, the left-hand side of (3.53) equals 0 when $\sqrt{\nu} > \frac{1}{\epsilon \cdot \phi_0}$. So indeed (3.53) and (3.54) hold. Now we can apply the martingale central limit theorem and conclude that at $t=\tau$

$$\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma_M(\theta_0)) \quad (\nu\to\infty),$$

where $\Sigma_M(\theta_0)$ is given by (3.42).

To prove (3.45) let us first look at $\frac{1}{\nu}\frac{\partial^2}{\partial\theta^2}\log \mathbf{L}(\theta_0)$.

$$\begin{aligned}
\frac{\partial^2}{\partial\beta^2}\log \mathbf{L}(\theta_0) &= -\nu^2\cdot\sum_{i=1}^{\mathbf{n}_\tau^{(\nu)}}\frac{1}{(\nu\beta_0-i+1)^2} \\
&= -\nu^2\left[\int_0^\tau \frac{1}{(\nu\beta_0-\mathbf{n}_{t-}^{(\nu)})^2}\mathrm{d}\mathbf{m}_t^{(\nu)} + \int_0^\tau \frac{\phi_0}{\nu\beta_0-\mathbf{n}_{t-}^{(\nu)}}\mathrm{d}t\right]. \qquad (3.55)
\end{aligned}$$

$$\frac{\partial^2}{\partial\phi^2}\log \mathbf{L}(\theta_0) = \frac{-\mathbf{n}_\tau^{(\nu)}}{\phi_0^2} = -\int_0^\tau \frac{\nu}{\phi_0}\left(\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}\right)\mathrm{d}t - \int_0^\tau \frac{1}{\phi_0^2}\mathrm{d}\mathbf{m}_t^{(\nu)}. \qquad (3.56)$$

$$\frac{\partial^2}{\partial\beta\partial\phi}\log \mathbf{L}(\phi_0) = \frac{\partial^2}{\partial\phi\partial\beta}\log \mathbf{L}(\theta_0) = -\nu\cdot\tau. \qquad (3.57)$$

So

$$\frac{1}{\nu}\begin{bmatrix} \frac{\partial^2}{\partial\beta^2}\log \mathbf{L}(\theta_0) & \frac{\partial^2}{\partial\beta\partial\phi}\log \mathbf{L}(\phi_0) \\ \frac{\partial^2}{\partial\phi\partial\beta}\log \mathbf{L}(\theta_0) & \frac{\partial^2}{\partial\phi^2}\log \mathbf{L}(\theta_0) \end{bmatrix} =$$

$$\begin{bmatrix} -\int_0^\tau \frac{\phi_0}{\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}}\mathrm{d}t - \frac{1}{\nu}\int_0^\tau \frac{1}{(\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu})^2}\mathrm{d}\mathbf{m}_t^{(\nu)} & -\tau \\ -\tau & -\frac{1}{\phi_0}\int_0^\tau\left(\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}\right)\mathrm{d}t - \frac{1}{\phi_0^2\cdot\nu}\cdot\int_0^\tau \mathrm{d}\mathbf{m}_t^{(\nu)} \end{bmatrix}. \qquad (3.58)$$

We will show that $\frac{1}{\nu}\int_0^\tau \frac{1}{(\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu})}\mathrm{d}\mathbf{m}_t^{(\nu)}$ and $\frac{1}{\phi_0^2\cdot\nu}\cdot\int_0^\tau \mathrm{d}\mathbf{m}_t^{(\nu)}$ converge in probability to zero. Therefore we apply Lenglart's inequality (see Section 1):

$$\Pr\left(\sup_{t\in[0,\tau]}\left|\frac{1}{\nu}\int_0^t \frac{1}{\left[\beta_0-\frac{\mathbf{n}_{s-}^{(\nu)}}{\nu}\right]^2}\mathrm{d}\mathbf{m}_s^{(\nu)}\right|>\mu\right) \leqslant \frac{\delta}{\mu^2} + \Pr\left(\left\langle\frac{1}{\nu}\int_0^\cdot \frac{1}{\left[\beta_0-\frac{\mathbf{n}_{s-}^{(\nu)}}{\nu}\right]^2}\mathrm{d}\mathbf{m}_s^{(\nu)}\right\rangle_\tau > \delta\right) \qquad (3.59)$$

and

$$\Pr\left[\sup_{t\in[0,\tau]}\left|\frac{1}{\nu}\cdot\frac{1}{\phi_0^2}\int_0^t \mathrm{d}\mathbf{m}_s^{(\nu)}\right|>\mu\right] \leqslant \frac{\delta}{\mu^2} + \Pr\left[\left\langle\frac{1}{\nu}\cdot\frac{1}{\phi_0^2}\int_0^\cdot \mathrm{d}\mathbf{m}_s^{(\nu)}\right\rangle_\tau > \delta\right] \qquad (3.60)$$

for all δ, $\mu>0$.

We have

$$<\frac{1}{\nu}\int_0 \frac{1}{\left[\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}\right]^2}\mathrm{d}\mathbf{m}_t^{(\nu)}>_\tau = \frac{1}{\nu}\int_0^\tau \frac{\phi_0}{\left[\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}\right]^3}\mathrm{d}t \xrightarrow{P} 0 \quad (\nu\to\infty) \tag{3.61}$$

$$<\frac{1}{\nu}\cdot\frac{1}{\phi_0^2}\int_0 \mathrm{d}\mathbf{m}_t^{(\nu)}>_\tau = \frac{1}{\phi_0^3}\int_0^\tau \left[\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}\right]\mathrm{d}t \xrightarrow{P} 0 \quad (\nu\to\infty) \tag{3.62}$$

both as a consequence of Theorem 3.1. Now from (3.59) and (3.60) we see easily that $\frac{1}{\nu}\int_0^\tau \frac{\phi_0}{(\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu})^2}\mathrm{d}\mathbf{m}_t^{(\nu)}$ and $\frac{1}{\phi_0^2\cdot\nu}\int_0^\tau \mathrm{d}\mathbf{m}_t^{(\nu)}$ converge to zero for $\nu\to\infty$. Again using Theorem 3.1, we now get from (3.58)

$$\frac{1}{\nu}\begin{bmatrix} \frac{\partial^2}{\partial\beta^2}\log \mathbf{L}(\theta_0) & \frac{\partial^2}{\partial\beta\partial\phi}\log \mathbf{L}(\theta_0) \\ \frac{\partial^2}{\partial\phi\partial\beta}\log \mathbf{L}(\theta_0) & \frac{\partial^2}{\partial\phi^2}\log \mathbf{L}(\theta_0)\end{bmatrix} \xrightarrow{P} \begin{bmatrix} -\frac{1}{\beta_0}(e^{\phi_0\tau}-1) & -\tau \\ -\tau & \frac{\beta_0}{\phi_0^2}(e^{-\phi_0\tau}-1)\end{bmatrix} = -\Sigma_M(\theta_0)\,. \tag{3.63}$$

Now look at the following Taylor-expansion ($\overline{\boldsymbol{\theta}}=\overline{\boldsymbol{\theta}}^{(i)}$)

$$\begin{aligned}\frac{1}{\nu}\cdot\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log \mathbf{L}(\overline{\boldsymbol{\theta}}) &= \frac{1}{\nu}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log \mathbf{L}(\theta_0) \\ &\quad + \frac{1}{\nu}\sum_{k=1}^{2}(\overline{\boldsymbol{\theta}}-\theta_0)_k\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\log \mathbf{L}(\boldsymbol{\theta}^*)\,,\quad i=1,2\,;\ j=1,2\end{aligned} \tag{3.64}$$

where $|\ (\boldsymbol{\theta}^*-\theta_0)_k\ | \leqslant |\ (\overline{\boldsymbol{\theta}}-\theta_0)_k\ | \leqslant (\hat{\boldsymbol{\theta}}-\theta_0)_k\ |$ and $\boldsymbol{\theta}^*$ actually depends on i and j. We have

$$\frac{1}{\nu}\frac{\partial^3}{\partial\beta^3}\log(\boldsymbol{\theta}^*) = \frac{1}{\nu}\sum_{i=1}^{\mathbf{n}_\tau^{(\nu)}}\frac{2}{(\beta^*-\frac{i}{\nu})^3} = 2\int_0^\tau \frac{\mathrm{d}\mathbf{x}_t^{(\nu)}}{(\beta^*-\mathbf{x}_{t-}^{(\nu)})^3} \tag{3.65}$$

$$\frac{1}{\nu}\frac{\partial^3}{\partial\phi^3}\log(\boldsymbol{\theta}^*) = \frac{1}{\nu}\cdot\frac{2\mathbf{n}_\tau^{(\nu)}}{(\phi^*)^3} = \frac{2\mathbf{x}_\tau^{(\nu)}}{(\phi^*)^3}\,. \tag{3.66}$$

The other derivatives of the third order are equal to zero. Notice that $\boldsymbol{\theta}^*$ and $\overline{\boldsymbol{\theta}}^{(i)}$ may not actually be measurable; however the usual outer probability arguments take care of this. We have already proved $\hat{\boldsymbol{\theta}}^{(\nu)} \xrightarrow{P} \theta_0$ $(\nu\to\infty)$, so for $\nu\to\infty$ we get from (3.64) that also $\boldsymbol{\theta}^* \xrightarrow{P} \theta_0$ and therefore

$$\Pr\left[\frac{1}{\nu}\left|\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\log \mathbf{L}(\boldsymbol{\theta}^*)\right|<M,\ \text{some finite constant}\right]\to 1 \quad (\nu\to\infty)\,. \tag{3.67}$$

So the whole second term on the right-hand side of (3.64) will converge in probability to zero for $\nu\to\infty$. Hence

$$\frac{1}{\nu}\frac{\partial^2}{\partial\theta^2}\log \mathbf{L}\,(\overline{\boldsymbol{\theta}}) \xrightarrow{P} -\Sigma_M\,(\theta_0)\,. \tag{3.68}$$

So indeed (3.44) and (3.45) hold. Using CRAMÉR (1946) or BILLINGSLEY (1961) (Theorems 2.2 and 10.1) we may conclude

$$\sqrt{\nu}\,(\hat{\boldsymbol{\theta}}^{(\nu)}-\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\ \Sigma_M\,(\theta_0)^{-1}) \quad (\nu\to\infty)\,. \tag{3.69}$$

□

The conclusion of this section is that the ML-estimators $\hat{\beta}^{(\nu)}$ and $\hat{\phi}^{(\nu)}$ are consistent (Theorem 3.2) and that

$$\sqrt{\nu}\,(\hat{\theta}^{(\nu)} - \theta_0) \xrightarrow{\mathcal{D}} \mathfrak{N}(0, \Sigma_M(\theta_0)^{-1}),$$

where

$$\Sigma_M(\theta_0) = \begin{bmatrix} -\frac{1}{\beta_0}(1-e^{\phi_0\tau}) & \tau \\ \tau & \frac{\beta_0}{\phi_0^2}(1-e^{-\phi_0\tau}) \end{bmatrix}$$

(Theorem 3.3). Note: the Fisher observed-information matrix converges also in probability to $-\Sigma_M(\theta_0)$:

$$\frac{1}{\nu}\frac{\partial^2}{\partial\theta^2}\log \mathbf{L}(\hat{\theta}^{(\nu)}) \xrightarrow{P} -\Sigma_M(\theta_0) \quad (\nu\to\infty) \tag{3.70}$$

(use (3.63) and Theorem (3.2)).

REMARK. Both Theorem 3.2 and 3.3 have been proved for stopping at time τ (fixed). These theorems can also be proved when we stop at any stopping time $\tau^{(\nu)}$ such that $\tau^{(\nu)} \xrightarrow{P} \tau$, as $\nu\to\infty$, for instance at time $\tau^{(\nu)} = \mathbf{T}^{(\nu)}_{[p\nu]}$ (p fixed), i.e., we would decide to stop after a fixed number of n failures, which we a priori expect to be a significant proportion of all.

The above model has been applied to data from the following project: Project A (see MOEK, 1983) concerning an information system for registering aircraft movements. It can be classified as a transaction oriented, on-line system where relatively small jobs are submitted by several users at arbitrary points of time. The jobs concern information retrieval as well as update of existing information. The information system consists of three identical subsystems, each comprising a PDP 11-34 computer with 256 Kbyte direct access memory, working together in such a way that each user has continually at his disposal the most recent information. A distributed database system is used for this purpose. The software consists of about 600 subroutines with an average length of 70 lines of executable code written in BASIC-PLUS-2 (a compiler oriented language) and about 40 subroutines with a average length of 200 lines of executable code written in assembler. Each subsystem has 20 Mbyte disc capacity. The number of user terminals connected is 35 and the number of line printers is 4. The entire system was developed in-house and, after some testing, was carried over to the users who continued testing in the operational environment. Failure data collected during the latter stage are given in Table 3.1. Some faults caused the same failure more than once, because immediate correction of the faults was not always possible. Only the first occurrence of such failures was counted for Table 3.1. It is not likely that we actually stopped after exactly 0.57657 Msec. A more likely used stop-criterion was to stop after 0.6 Msec, or perhaps at the first failure after 0.5 Msec. We pretend below that the stopping time was $\mathbf{T}_n$ for various values of n.

failure number i	inter-failure time $\mathbf{t}_i$	failure time $\mathbf{T}_i$	failure number i	inter-failure time $\mathbf{t}_i$	failure time $\mathbf{T}_i$
1	880	880	23	4450	133210
2	3430	4310	24	4860	138070
3	2860	7170	25	640	138710
4	11760	18930	26	3990	142700
5	4750	23680	27	26840	169540
6	240	23920	28	2270	171810
7	2300	26220	29	200	172010
8	8570	34790	30	39180	211190
9	4620	39410	31	14910	226100
10	1060	40470	32	14670	240770
11	3820	44290	33	16310	257080
12	14800	59090	34	38410	295490
13	1770	60860	35	1120	296610
14	24270	85130	36	30560	327170
15	4800	89930	37	6210	333380
16	470	90400	38	120	333500
17	40	90440	39	20210	353710
18	10170	100610	40	26400	380110
19	1120	101730	41	37800	417910
20	980	102710	42	74220	492130
21	24300	127010	43	84440	576570
22	1750	128760			

TABLE 3.1. Failure times for project A (CPU time in Msec $\times 10^6$)

We constructed the upperbounds of one-sided $(1-\alpha)$ confidence intervals for N after detection of n failures, by several methods, and at different values of n. For $\theta = \begin{bmatrix} N \\ \phi \end{bmatrix}$, $\hat{\boldsymbol{\theta}}$ denotes the maximum likelihood estimator of $\boldsymbol{\theta}$. First of all, we constructed upperbounds for N by using (asymptotic) expected Fisher information and observed information respectively. By using Theorem 3.3, we have

$$\sqrt{\nu}\,(\hat{\boldsymbol{\theta}}^{(\nu)} - \theta_0) \xrightarrow{\mathcal{D}} \mathfrak{N}(0,\ \Sigma_M\,(\theta_0)^{-1}) \quad (\nu \to \infty)$$

$$\Sigma_M\,(\theta_0)^{-1} = \begin{bmatrix} \frac{1}{\beta_0}\,(e^{\phi_0\tau}-1) & \tau \\ \tau & \frac{-\beta_0}{\phi_0^2}\,(e^{-\phi_0\tau}-1) \end{bmatrix}^{-1}$$

$$= D^{-1} \begin{bmatrix} \frac{-\beta_0}{\phi_0^2}\,(e^{-\phi_0\tau}-1) & -\tau \\ -\tau & \frac{1}{\beta_0}\,(e^{\phi_0\tau}-1) \end{bmatrix}$$

(asymptotic expected information), in which

$$D = -\frac{1}{\phi_0^2}\,(2 - e^{\phi_0\tau} - e^{-\phi_0\tau}) - \tau^2\,.$$

So we get

$$\hat{\sigma}_{\beta}^{2} = \frac{\frac{\hat{\beta}}{\hat{\phi}^{2}}\,(1-e^{\hat{\phi}\tau})}{\frac{1}{\hat{\phi}^{2}}\,e^{\hat{\phi}\tau}(1-e^{\hat{\phi}\tau})^{2}} - \tau^{2}\,.$$

Notice $\hat{\mathrm{N}} = \nu\hat{\beta}$, so var $\hat{\mathrm{N}} = \nu^{2}$ var $\hat{\beta}$. By taking $\nu = n$, we get

$$\hat{\sigma}^{2}_{\mathrm{N},exp} = \frac{n\hat{\mathrm{N}}\,(e^{-\hat{\phi}\tau}-1)}{2-e^{\hat{\phi}\tau}-e^{-\hat{\phi}\tau}+\tau^{2}\hat{\phi}^{2}} \tag{3.71}$$

(for values of $\hat{\sigma}_{\mathrm{N},exp}$, see Table 3.2).

Another estimator of Σ is $\frac{1}{\nu}\,\mathbf{I}^{(\nu)}$, with

$$\mathbf{I} = -\begin{pmatrix} \frac{\partial^{2}}{\partial N^{2}} \log \mathbf{L}\,|_{\hat{\mathrm{N}},\hat{\phi}} & \frac{\partial^{2}}{\partial N \partial\phi} \log \mathbf{L}\,|_{\hat{\mathrm{N}},\hat{\phi}} \\ \frac{\partial^{2}}{\partial\phi\partial N} \log \mathbf{L}\,|_{\hat{\mathrm{N}},\hat{\phi}} & \frac{\partial^{2}}{\partial\phi^{2}} \log \mathbf{L}\,|_{\hat{\mathrm{N}},\hat{\phi}} \end{pmatrix}$$

Here $\mathbf{I}$ is the observed information for θ. In the case when we stop after detection of n (fixed $=[p\nu]$) failures, the log-likelihood function becomes

$$\begin{aligned}\log L(N,\phi|\tilde{\mathrm{t}}_{i},\ i=1,\ldots,n) &= -\sum_{i=1}^{n} \phi(N-i+1)\,\tilde{\mathrm{t}}_{i} + \sum_{i=1}^{n} \log \phi(N-i+1) \\ &= -\sum_{i=1}^{n} \phi(N-i+1)\,\mathrm{t}_{i} + \sum_{i=1}^{n} \log \phi(N-i+1)\,.\end{aligned}$$

So,

$$\mathbf{I} = \begin{pmatrix} \sum_{i=1}^{n} \frac{1}{(\hat{\mathrm{N}}-i+1)^{2}} & \tau \\ \tau & \frac{n}{\hat{\phi}^{2}} \end{pmatrix}.$$

Taking $\nu=n$, $(\frac{1}{\nu}\,\mathbf{I})^{-1} = (\frac{1}{n}\,\mathbf{I})^{-1}$ is an estimator of Σ^{-1}, with

$$\left(\frac{1}{n}\mathbf{I}\right)^{-1} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{(\hat{\mathrm{N}}-i+1)^{2}} & \tau/n \\ \tau/n & \frac{n}{\hat{\phi}^{2}} \end{pmatrix} = (D')^{-1} \begin{pmatrix} \frac{1}{\hat{\phi}^{2}} & \frac{-\tau}{n} \\ \frac{-\tau}{n} & \frac{1}{n}\sum_{i=1}^{n} \frac{1}{(\hat{\mathrm{N}}-i+1)^{2}} \end{pmatrix},$$

where

$$\mathbf{D}' = \frac{1}{n\hat{\phi}^{2}} \sum_{i=1}^{n} \frac{1}{(\hat{\mathrm{N}}-i+1)^{2}} - \frac{\tau^{2}}{n^{2}}\,.$$

So, we get

$$\hat{\sigma}^{2}_{\mathrm{N},obs} = (D')^{-1}\,\frac{1}{\hat{\phi}^{2}} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{(\hat{\mathrm{N}}-i+1)^{2}} - \frac{\tau^{2}\hat{\phi}^{2}}{n^{2}}} \tag{3.72}$$

(for realized values of $\hat{\sigma}_{\hat{N},obs}$, see Table 3.2). We are now able to construct approximate one-sided $(1-\alpha)$ confidence intervals for N, for which the upperbound for N is to be calculated from

$$\Pr(|\mathbf{z}| \leqslant k_\alpha) = 1-2\alpha,$$

where

$$\frac{\sqrt{n}\,(\hat{N}-N)}{\hat{\sigma}_{\hat{N}}} \overset{\mathcal{D}}{\simeq} \mathbf{z} \overset{\mathcal{D}}{=} \mathcal{N}(0,1)$$

($(1-2\alpha)$ two-sided confidence interval). In this case the upperbound for N will equal

$$\hat{N} + k_\alpha \frac{\hat{\sigma}_{\hat{N}}}{\sqrt{n}}; \quad \Pr\Big(N \leqslant \hat{N} + k_\alpha \frac{\hat{\sigma}_{\hat{N}}}{\sqrt{n}}\Big) \simeq 1-\alpha .$$

We calculated the upperbounds for N for both $\alpha=.05$ and $\alpha=.30$ ($k_{.05}=1.645$ and $k_{.3}=0.525$). The values are given in Table 3.3a and 3.3b respectively. SPREY (1985) constructed exact one-sided $(1-\alpha)$ confidence intervals for N, by using the exact distribution of $\zeta = \dfrac{\sum_{i=1}^{n}(i-1)\mathbf{t}_i}{\sum_{i=1}^{n}\mathbf{t}_i}$, for n fixed. Those values are also tabled in 3.3a and 3.3b. Furthermore we determined upperbounds for N by making use of the Wilkes' Likelihood Ratio Test. Let $\mathbf{N}^*$ be the larger of the two values for which

$$2\,[\log \mathbf{L}(\hat{N}) - \log \mathbf{L}(N^*)] = C_\alpha = c\,,$$

where

$$\mathbf{L}(N) = \max_{\phi} \mathbf{L}(N,\phi)$$

$$\Pr(|\mathbf{w}| \leqslant C_\alpha) = 1-2\alpha$$

and

$$2\,[\log \mathbf{L}(\hat{N}) - \log \mathbf{L}(N)] \overset{\mathcal{D}}{\simeq} \mathbf{w} \overset{\mathcal{D}}{=} \chi_1^2\,,$$

i.e., $\mathbf{N}^*$ is the larger solution of

$$(\phi^*(N^*))^n \prod_{i=1}^{n} (N^*-i+1) = (\hat{\phi})^n \prod_{i=1}^{n} (\hat{N}-i+1)\exp\Big(-\frac{c}{2}\Big),$$

where

$$\phi^*(N^*) = \frac{n}{\sum_{i=1}^{n}(N^*-i+1)\mathbf{t}_i}\,.$$

For the LRT-upperbounds for N, we also refer to Table 3.3a and 3.3b. Finally, we determined 'modified expected information' upperbounds for N, by taking better account of the presence of the unknown N in the expected information. We computed N_u from the equality:

$$N_u = \hat{N} + k_\alpha \frac{\sigma(N_u)}{\sqrt{n}}\,,$$

with k_α as mentioned before. Here

$$\sigma^2(N_u) = \frac{1}{\dfrac{1}{n}\sum_{i=1}^{n}\dfrac{1}{(N_u-i+1)^2} - \left[\dfrac{1}{n}\sum_{i=1}^{n}\dfrac{1}{(N_u-i+1)}\right]^2} \qquad (3.73)$$

In Table 3.3a (α=.05) it can be seen that we couldn't find any value for N_u, because $N_u - k_{.05} \frac{\sigma(N_u)}{\sqrt{n}} < \hat{N}$, $\forall N_u > \hat{N}$. In such cases one should take $N_u = \infty$.

Working out the cases when $n=$ 10, 14, 20, 30, 32, 34, 36, 38 and 40 failures are detected, we are able to construct Tables 3.2, 3.3a and 3.3b.

n	$\hat{N}$	$\hat{\phi} = \frac{1}{T} \sum_{i=1}^{n} (\hat{N}-i+1) = \frac{n}{\sum_{i=1}^{n} (\hat{N}-i+1)t_i}$	$\hat{\sigma}_{\hat{N}, exp}$	$\hat{\sigma}_{\hat{N}, obs}$
10	123.6	2.08	23986	24328
14	16.2	21.34	187	198
20	111.8	1.91	3225	3248
30	39.9	6.42	2961	3007
32	40.7	6.23	2231	2266
34	38.8	6.78	745	759
36	41.2	6.08	860	875
38	47.2	4.78	2450	2482
40	47.4	4.74	1618	1640
		(per Msec)		

TABLE 3.2.

n	$\hat{N} + 1.645 \frac{\hat{\sigma}_{\hat{N}, exp}}{\sqrt{n}}$	$\hat{N} + 1.645 \frac{\hat{\sigma}_{\hat{N}, obs}}{\sqrt{n}}$	Sprey	LRT	N_u
10	2671.3	2689.4	∞	∞	∞
14	22.22	22.39	60	72	∞
20	772.41	774.73	∞	∞	∞
30	56.24	56.37	113	121	∞
32	54.44	54.54	84	89	∞
34	46.50	46.57	54	57	∞
36	49.24	49.31	55	59	∞
38	60.41	60.50	80	83	∞
40	57.86	57.93	69	72	∞

TABLE 3.3a : Upperbounds for N, based on approximate 95 percent one-sided confidence intervals for N

n	$\hat{N}+0.525\,\frac{\hat{\sigma}_{\hat{N},\,exp}}{\sqrt{n}}$	$\hat{N}+0.525\,\frac{\hat{\sigma}_{\hat{N},\,obs}}{\sqrt{n}}$	Sprey	LRT	N_u
10	936.7	942.5	∞	∞	∞
14	18.12	18.17	18	19	∞
20	322.63	323.38	∞	∞	∞
30	45.12	45.16	46	48	∞
32	45.08	45.12	45	47	53
34	41.26	41.28	41	43	44
36	43.77	43.79	43	45	48
38	51.42	51.44	51	53	57
40	50.74	50.76	50	52	54

TABLE 3.3b : Upperbounds for N, based on approximate 70 percent one-sided confidence intervals for N.

We expected that the upperbounds for N using observed information would be closer to the Sprey-upperbounds, than the upperbounds based on (asymptotic) expected information. Table 3.3a and 3.3b do not confirm this conjecture. As we expected the bounds based on the likelihood Ratio Test do not differ very much from the upperbounds of Sprey. The differences in Table 3.3b are smaller that in 3.3a, possibly because it is easier to estimate a central than an extreme quantile of the distribution of $\hat{N}$. The great differences in Table 3.3a. are also caused by using small sample sizes. The results in Table 3.3b are very satisfactory.

4. The Goel and Okumoto-model

The failure rate at time t in this model is given by

$$\lambda(t) = N\phi_0 e^{-\phi_0 t}, \quad N, \phi_0 > 0. \tag{4.1}$$

Using (1.20), the likelihood function becomes

$$L(N, \phi_0|\tilde{\mathbf{t}}_i : i=1, \ldots, \mathbf{n}_\tau+1) = (N\phi_0)^{\mathbf{n}_\tau} e^{-\phi_0 \sum_{i=1}^{\mathbf{n}_\tau} \mathbf{T}_i} \cdot e^{-N(1-e^{-\phi_0\tau})}, \tag{4.2}$$

where $\mathbf{T}_i = \sum_{j=1}^{i} \tilde{\mathbf{t}}_j$ is the time measured from the beginning of the experiment up to the i'th failure time and again τ is the specified non-random stopping time. The logarithm of (4.2) becomes

$$\log L(N, \phi_0|\mathbf{t}_i, i=1, \ldots, \mathbf{n}_\tau+1) = \mathbf{n}_\tau \log N\phi_0 - \phi_0 \sum_{i=1}^{\mathbf{n}_\tau} \mathbf{T}_i - N(1-e^{-\phi_0\tau}). \tag{4.3}$$

So the likelihood equations become

$$\frac{\partial}{\partial N} \log \mathbf{L}|_{\hat{\phi},\hat{N}} = \frac{\mathbf{n}_\tau}{\hat{N}} - (1-e^{-\hat{\phi}\tau}) = 0 \tag{4.4}$$

and

$$\frac{\partial}{\partial \phi} \log \mathbf{L}|_{\hat{\phi},\hat{N}} = \frac{\mathbf{n}_\tau}{\hat{\phi}} - \hat{N}\tau e^{-\hat{\phi}\tau} - \sum_{i=1}^{\mathbf{n}_\tau} \mathbf{T}_i = 0 \tag{4.5}$$

From (4.4) we get $\hat{N} > \mathbf{n}_\tau$ and

$$\hat{\phi} = -\frac{1}{\tau}\log(1-\frac{\mathbf{n}_\tau}{\hat{\mathbf{N}}}) \tag{4.6}$$

and $\hat{\mathbf{N}}$ can be got numerically from

$$\mathbf{f}(\hat{\mathbf{N}}) \stackrel{\text{def}}{=} \left|1-\frac{\mathbf{n}_\tau}{\hat{\mathbf{N}}}\right|^{c_{\mathbf{n}:\tau}-\hat{\mathbf{N}}(1-\frac{\mathbf{n}_\tau}{\hat{\mathbf{N}}})} - e^{-\mathbf{n}_\tau} = 0 \tag{4.7}$$

or equivalently

$$\mathbf{g}(\hat{\mathbf{N}}) \stackrel{\text{def}}{=} \log(1-\frac{\mathbf{n}_\tau}{\hat{\mathbf{N}}}) + \frac{\mathbf{n}_\tau}{\hat{\mathbf{N}}-\mathbf{n}_\tau+c_{\mathbf{n}:\tau}} = 0. \tag{4.8}$$

where

$$c_{\mathbf{n}:\tau} \stackrel{\text{def}}{=} \frac{1}{\tau}\sum_{i=1}^{\mathbf{n}_\tau}\mathbf{T}_i = \frac{1}{\tau}\int_0^\tau t\,d\mathbf{n}_t\,. \tag{4.9}$$

Our aim is to prove consistency and asymptotic normality of the estimators. As we did in the previous section, let us write $N=\nu\beta_0$ and let ν grow to infinity. By $\mathbf{n}_t^{(\nu)}$ we mean again the counting process $\mathbf{n}_t$ in the ν'th experiment. An equivalent of (4.8) is then

$$\mathbf{U}^{(\nu)}(\hat{\boldsymbol{\beta}}^{(\nu)}) = \log(1-\frac{\mathbf{n}_\tau^{(\nu)}}{\nu\hat{\boldsymbol{\beta}}^{(\nu)}}) + \frac{\mathbf{n}_t^{(\nu)}/\nu}{c_{\mathbf{n}:\tau}^{(\nu)}/\nu+\hat{\boldsymbol{\beta}}^{(\nu)}-\frac{\mathbf{n}_\tau^{(\nu)}}{\nu}} = 0. \tag{4.10}$$

where $c_{\mathbf{n}^{(\nu)}:\tau} \stackrel{\text{def}}{=} \frac{1}{\tau}\int_0^\tau t\,d\mathbf{n}_t^{(\nu)}$ and where $\hat{\boldsymbol{\beta}}^{(\nu)}$ is ML-estimator of β_0 in the ν'th experiment.

THEOREM 4.1. *In the model of Goel and Okumoto*

$$\sup_{t\in[0,\tau]}\left|\frac{\mathbf{n}_t^{(\nu)}}{\nu}-\beta_0(1-e^{-\phi_0 t})\right| \xrightarrow{P} 0 \text{ as } \nu\to\infty \tag{4.11}$$

and

$$\frac{c_{\mathbf{n}^{(\nu)}:\tau}}{\nu} \xrightarrow{P} -\beta_0 e^{-\phi_0\tau} + \frac{\beta_0}{\phi_0\tau}(1-e^{-\phi_0\tau}) \quad (\nu\to\infty)\,. \tag{4.12}$$

PROOF. Let in Theorem 1.1 $\beta(s,x)$ be defined by

$$\beta(s,x) = \phi_0\beta_0 e^{-\phi_0 s}\,. \tag{4.13}$$

Since $\beta(s,x)$ does not depend on x in this case, the conditions (1.14) and (1.15) of Theorem 1.1 trivially hold. Therefore

$$\sup_{t\in[0,\tau]}\left|\frac{\mathbf{n}_t^{(\nu)}}{\nu}-x_t\right| \xrightarrow{P} 0 \quad (\nu\to\infty), \tag{4.14}$$

where

$$x_t = \int_0^t \phi_0\beta_0 e^{-\phi_0 s}ds = \beta_0(1-e^{-\phi_0 t})\,. \tag{4.15}$$

Using partial integration, we get

$$\frac{c_{\mathbf{n}^{(\nu)};T}}{\nu} = \frac{1}{\nu}\cdot\frac{1}{T}\int_0^T t\,\mathrm{d}\mathbf{n}_t^{(\nu)} = \frac{1}{T}\int_0^T t\,\mathrm{d}\frac{\mathbf{n}_T^{(\nu)}}{\nu} - \frac{1}{T}\int_0^T \frac{\mathbf{n}_t^{(\nu)}}{\nu}\,\mathrm{d}t\,. \tag{4.16}$$

So using (4.14) we get

$$\frac{c_{\mathbf{n}^{(\nu)};T}}{\nu} \xrightarrow{\mathrm{P}} -\beta_0 e^{-\phi_0 T} + \frac{\beta_0}{\phi_0 T}\,(1-e^{-\phi_0 T})\ \ (\nu\to\infty)\,. \tag{4.17}$$

□

From Theorem 4.1 we get immediately :

COROLLARY 4.1. *In the Goel and Okumoto model*

$$\frac{\mathbf{n}_t^{(\nu)}}{\nu} \xrightarrow{\mathrm{P}} \beta_0(1-e^{-\phi_0 t})\ \ (\nu\to\infty)\,. \tag{4.18}$$

MOEK (1984) proved the following lemma:

LEMMA 4.1. *Equation (4.7) has a unique solution* $\hat{\mathbf{N}}$ *if and only if* $c_{\mathbf{n};T} < \frac{\mathbf{n}_T}{2}$. *This unique solution will lie in the interval*

$$(\mathbf{n}_T,\ \mathbf{n}_T + \frac{c^2_{\mathbf{n};T}}{\mathbf{n}_T - 2c_{\mathbf{n};T}}]\,.$$

PROOF. $\mathbf{f}(N)$ from (4.7) can be approximated by

$$\mathbf{f}(N) \approx e^{-\mathbf{n}_T}\left[e^{-(c_{\mathbf{n};T}-\frac{\mathbf{n}_T}{2})\cdot\frac{\mathbf{n}_T}{N}} - 1\right] \tag{4.19}$$

therefore $\mathbf{f}(N)\downarrow 0$ for $N\to\infty$, whenever $c_{\mathbf{n};T} < \frac{\mathbf{n}_T}{2}$. From (4.7) we see that $\mathbf{f}(\mathbf{n}_T) < 0$. From the sign of the derivative of a possible root and from the continuity of $\mathbf{f}(N)$, we may conclude the uniqueness of the solution of (4.7). □

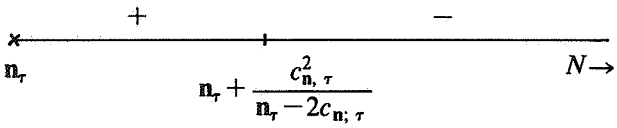

FIGURE 4.1. Sign of the derivative at a root of $\mathbf{f}(N)=0$ when $c_{\mathbf{n};T}<\mathbf{n}_T/2$.

Of course this lemma implies the following corollary:

COROLLARY 4.2. $\mathbf{U}^{(\nu)}(\beta)$ *defined by (4.10) for* $\beta>\mathbf{x}_T^{(\nu)}$ *has a unique solution* $\hat{\boldsymbol{\beta}}^{(\nu)}$ *if and only if* $c^{(\nu)}_{\mathbf{n};T} < \frac{\mathbf{n}_T^{(\nu)}}{2}$.

Now define

$$U(\beta) = \log\,(1-\frac{\alpha\beta_0}{\beta}) + \frac{\alpha}{\dfrac{\alpha}{\phi_0\tau} + \dfrac{\beta}{\beta_0} - 1} \quad \text{for } \beta > x_\tau = \alpha\beta_0\,. \tag{4.20}$$

where $\alpha = (1-e^{-\phi_0\tau})$, β_0 and ϕ_0 the true model parameters. Notice that the value of β, for which $\frac{\alpha}{\phi_0\tau} + \frac{\beta}{\beta_0} - 1 = 0$, that is $\bar{\beta} \overset{\text{def}}{=} \beta_0(1-\frac{\alpha}{\phi_0 t})$ is irrelevant since it will be smaller then x_τ as a consequence of $\log(1+x) \leqslant x \;\; \forall x > -1$.

LEMMA 4.2. *Let $U(\beta)$ be defined by (4.20). Then β_0 is the unique solution of $U(\beta) = 0$.*

PROOF.

$$U(\beta_0) = \log\,(1-\frac{\alpha\beta_0}{\beta_0}) + \frac{\alpha}{\dfrac{\alpha}{\phi_0\tau} + \dfrac{\beta}{\beta_0} - 1} = -\phi_0\tau + \phi_0\tau = 0\,, \tag{4.21}$$

$$\frac{\partial}{\partial\beta}\,U(\beta) = \alpha\left[\frac{1}{\dfrac{\beta^2}{\beta_0} - \alpha\beta} - \frac{1}{\beta_0(\dfrac{\alpha}{\phi_0\tau} + \dfrac{\beta}{\beta_0} - 1)^2}\right]. \tag{4.22}$$

Using $U(\beta)=0$, the derivative at a root becomes

$$\frac{\partial}{\partial\beta}U(\beta) = \alpha\left[\frac{1}{\dfrac{\beta^2}{\beta_0} - \alpha\beta} - \frac{1}{\alpha^2\beta_0}[\log\,(1-\frac{\alpha\beta_0}{\beta})]^2\right]$$

$$= \frac{\beta_0}{\beta^2}\,\alpha\left[\frac{1}{1-\alpha\dfrac{\beta_0}{\beta}} - \left(\frac{\beta}{\alpha\beta_0}\right)^2[\log\,(1-\frac{\alpha\beta_0}{\beta})]^2\right]. \tag{4.23}$$

Let $\alpha = \dfrac{\alpha\beta_0}{\beta}$, then we want to prove

$$\frac{1}{1-x} - \frac{\log^2(1-x)}{x^2} \geqslant 0\,,\; 0 < x = \frac{\alpha\beta_0}{\beta} < 1\,. \tag{4.24}$$

From $-\log(1-x) \geqslant x$, $0<x<1$ we get

$$\frac{1}{1-x} - \frac{\log^2(1-x)}{x^2} \geqslant \frac{1}{1-x} + \frac{\log(1-x)}{x} \overset{\text{def}}{=} h(x),\; 0<x<1\,. \tag{4.25}$$

One can easily verify

$$\frac{\partial}{\partial x}\,h(x) > 0,\;\; 0<x<1 \tag{4.26}$$

$$\lim_{x\uparrow 1} h(x) = +\infty \tag{4.27}$$

and

$$\lim_{x\downarrow 0} h(x) = 0\,. \tag{4.28}$$

So $h(x)$ is stricly increasing and non-negative. Hence from (4.23) and (4.25) we may conclude

$$\frac{\partial}{\partial\beta}\,U(\beta) \geqslant 0,\ \text{whenever}\ U(\beta) = 0\,.$$

Furthermore, we know

$$\lim_{\beta\downarrow\alpha\beta_0 = x_\tau} U(\beta) = -\infty \tag{4.29}$$

and

$$\lim_{\beta\to\infty} U(\beta) = 0\,. \tag{4.30}$$

Using the continuity of $U(\beta)$ we may now conclude that β_0 is the unique solution of $U(\beta) = 0$. □

In Figure 4.2 we have illustrated the shape of $U(\beta)$.

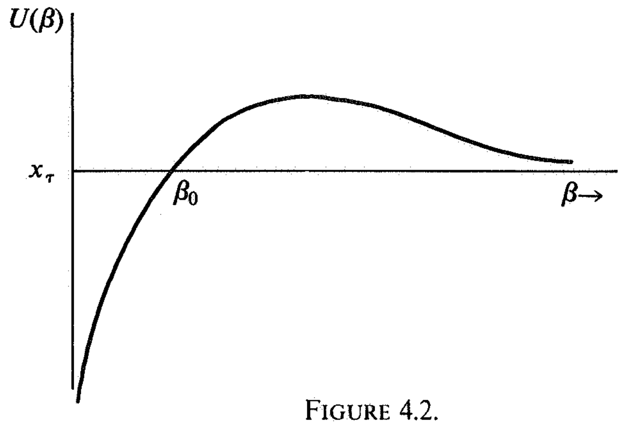

FIGURE 4.2.

LEMMA 4.3. *Let* $\mathrm{U}^{(\nu)}(\beta)$ *be defined by (4.10) for* $\beta > \mathrm{x}_\tau^{(\nu)}$ *and* $U(\beta)$ *by (4.20) for* $\beta > x_\tau$, *then*

$$\mathrm{U}^{(\nu)}(\beta) \xrightarrow{\mathrm{P}} U(\beta)\ \ (\nu\to\infty),$$

for all fixed $\beta > x_\tau$.

PROOF: follows immediately from Theorem 4.1 and Corollary 4.1. □

THEOREM 4.2. *Let*

$$\hat{\boldsymbol{\theta}}^{(\nu)} = \begin{bmatrix} \hat{\boldsymbol{\beta}}^{(\nu)} \\ \hat{\boldsymbol{\phi}}^{(\nu)} \end{bmatrix} \ \textit{and}\ \ \theta_0 = \begin{bmatrix} \beta_0 \\ \phi_0 \end{bmatrix},$$

then

$$\hat{\boldsymbol{\theta}}^{(\nu)} \xrightarrow{\mathrm{P}} \theta_0\ \ (\nu\to\infty)\,.$$

PROOF:
— If $\mathrm{U}^{(\nu)}(\beta) = 0$ has a solution for $\beta > \mathrm{x}_\tau^{(\nu)}$, it will be unique: $\hat{\boldsymbol{\beta}}^{(\nu)}$ (Corollary 4.2).
— β_0 is the unique solution of $U(\beta) = 0$ (Lemma 4.2)
— $\mathrm{U}^{(\nu)}(\beta) \xrightarrow{\mathrm{P}} U(\beta)$ $(\nu\to\infty)$ for all fixed $\beta > x_\tau$ (Lemma 4.3)

— $\mathbf{U}^{(\nu)}(\beta)$ and $U(\beta)$ are continuous in β, so let $x_\tau<\beta'<\beta_0<\beta''$:

$$\lim_{\nu\to\infty} \Pr \text{ (there is a solution of } \mathbf{U}^{(\nu)}(\beta)=0 \text{ in } [\beta',\beta'']) = 1 .$$

Since this solution will be unique,

$$\hat{\boldsymbol{\beta}}^{(\nu)} \xrightarrow{P} \beta_0 \quad (\nu\to\infty) . \tag{4.31}$$

Now look at $\hat{\boldsymbol{\phi}}^{(\nu)}$ from (4.6):

$$\hat{\boldsymbol{\phi}}^{(\nu)} = -\frac{1}{\tau}\log(1-\frac{\mathbf{x}_\tau^{(\nu)}}{\hat{\boldsymbol{\beta}}^{(\nu)}}) \xrightarrow{P} -\frac{1}{\tau}\log(1-\frac{x_\tau}{\beta_0}) = \phi_0 \tag{4.32}$$

as a consequence of (4.31) and Corollary 4.1. □

THEOREM 4.3.

$$\sqrt{\nu}\,(\hat{\boldsymbol{\theta}}^{(\nu)} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{GO}(\theta_0)^{-1}),$$

where (GO denotes Goel and Okumoto)

$$\Sigma_{GO}(\theta_0) = \begin{bmatrix} \frac{1}{\beta_0}(1-e^{\phi_0\tau}) & \tau e^{-\phi_0\tau} \\ \tau e^{-\phi_0\tau} & \frac{\beta_0}{\phi_0^2} - \frac{\beta_0}{\phi_0^2}e^{\phi_0\tau} - \beta_0\tau^2 e^{-\phi_0\tau} \end{bmatrix}. \tag{4.33}$$

The proof of this theorem is a complete analogon of the proof of Theorem 3.3. Therefore the proof of Theorem 4.3 can be found in Appendix A.

We end this section, comparing the asymptotic variances of the ML-estimators in the Musa-model and in the model of Goel and Okumoto. Recall (3.42)

$$\Sigma_M(\theta_0) = \begin{bmatrix} \frac{1}{\beta_0}(e^{\phi_0\tau}-1) & \tau \\ \tau & \frac{-\beta_0}{\phi_0^2}(e^{-\phi_0\tau}-1) \end{bmatrix}$$

Hence

$$\Sigma_M(\theta_0)^{-1} = \frac{1}{\det_M}\begin{bmatrix} \frac{\beta_0}{\phi_0^2}(1-e^{-\phi_0\tau}) & -\tau \\ -\tau & -\frac{1}{\beta_0}(1-e^{\phi_0\tau}) \end{bmatrix}, \tag{4.34}$$

where

$$\det_M = \frac{1}{\phi_0^2}e^{\phi_0\tau} - \frac{2}{\phi_0^2} + \frac{1}{\phi_0^2}e^{-\phi_0\tau} - \tau^2 \geqslant 0 .$$

From (4.33) we get

$$\Sigma_{GO}(\theta_0)^{-1} = \frac{1}{\det_{GO}}\begin{bmatrix} \frac{\beta_0}{\phi_0^2} - \frac{\beta_0}{\phi_0^2}e^{-\phi_0\tau} - \beta_0\tau^2 e^{-\phi_0\tau} & -\tau e^{-\phi_0\tau} \\ -\tau e^{-\phi_0\tau} & \frac{1}{\beta_0}(1-e^{-\phi_0\tau}) \end{bmatrix}. \tag{4.35}$$

Since $\det_{GO} = e^{-\phi_0\tau}.\ \det_M$, (4.35) becomes

$$\Sigma_{GO}(\theta_0)^{-1} = \frac{1}{\det_M}\begin{bmatrix} -\frac{\beta_0}{\phi_0^2}(e^{-\phi_0\tau}-1)e^{\phi_0\tau} - \beta_0\tau^2 & -\tau \\ -\tau & \frac{1}{\beta_0}(e^{\phi_0\tau}-1) \end{bmatrix}. \tag{4.36}$$

Comparing (4.34) and (4.36), we see that only the top-left elements $\Sigma_M^{11}(\theta_0)$ and $\Sigma_{GO}^{11}(\theta_0)$ differ, that is, the asymptotic variance of $\sqrt{\nu}(\hat{\beta}^{(\nu)} - \beta_0)$:

$$\Sigma_{GO}^{11} - \Sigma_M^{11} = (e^{\phi_0\tau}-1)\,\Sigma_M^{11} - \frac{\beta_0\tau^2}{\det_M}$$

$$= \frac{1}{\det_M}\left[\frac{\beta_0}{\phi_0^2}(e^{\phi_0\tau}-1)(1-e^{-\phi_0\tau}) - \beta_0\tau^2\right]$$

$$= \frac{\beta_0}{\phi_0^2 \det_M}\left[(e^{\phi_0\tau} - e^{-\phi_0\tau} - 2) - (\phi_0\tau)^2\right] \geqslant 0\,. \tag{4.38}$$

One may conclude from this that the Goel and Okumoto model gives less information about the β_0-parameter then the Musa-model. Intuitively speaking, the Goel and Okumoto model is more random than the Musa model although the intensities are asymptotically equivalent. Under the model of Goel and Okumoto we only *expect* to finally have N failures; in the Musa model we finally certainly do.

5. The Littlewood-model

Recall that in the Littlewood-model $\boldsymbol{\phi}_j$, $j = 1, \ldots, N$ are independently gamma distributed $\Gamma(a,b)$, where $\boldsymbol{\phi}_j$ denotes the failure rate of fault number j. Assume that we have already detected $i-1$ faults up to time t (see figure 5.1), so $N-i+1$ faults are still remaining in the program

Figure 5.1

Define the counting process $\mathbf{n}_t^{(j)}$ as

$$\mathbf{n}_t^{(j)} = I\ \{\text{failure } j \text{ is detected in the time interval } [0,t]\}\,. \tag{5.1}$$

Notice that $\mathbf{n}_t^{(j)}$ jumps only once for each j. Define

$$\mathbf{n}_t = \sum_{j=1}^{N} \mathbf{n}_t^{(j)}, \tag{5.2}$$

then the counting process $\mathbf{n}_t$ is observable and denotes the number of detected faults up to t. So $\mathbf{n}_t$ takes the value $i-1$ in our case.

In the first instance our goal is to find the intensity $\boldsymbol{\lambda}_t$, with respect to the correct filtration, i.e. generated by the aggregate process $\mathbf{n}_t$. Let

$$\mathcal{F}_t = \sigma\{\boldsymbol{\phi}_j,\ \mathbf{n}_s^{(j)},\ s \leqslant t,\ j = 1, \ldots, N\} \tag{5.3}$$

and

$$\mathcal{N}_t = \sigma\{\mathbf{n}_s,\ s \leqslant t\}\,. \tag{5.4}$$

We want to find $\boldsymbol{\lambda}_t^{\mathcal{N}_t}$, that is, the intensity with respect to $\mathcal{N}_t$. We know from (5.1) that the intensity of a failure from fault number j is given by

$$\lim_{h\downarrow 0} \frac{1}{h}\Pr(\mathbf{n}_{t+h}^{(j)} - \mathbf{n}_t^{(j)} = 1|\mathcal{F}_{t-}) = \phi_j(1-\mathbf{n}_{t-}^{(j)}) = \phi_j I\{\mathbf{n}_{t-}^{(j)} = 0\} \tag{5.5}$$

(if fault j has already occured the intensity equals zero, otherwise ϕ_j). So

$$\lambda_t^{\mathcal{F}_t} = \sum_{j=1}^{N} \phi_j \; I\{\mathbf{n}_{t-}^{(j)} = 0\}. \tag{5.6}$$

We will now apply a theorem, which is known as the "Innovation-theorem": *if* $\mathbf{n}_t$ *is a counting process and* $\mathcal{N}_t$ *and* $\mathcal{F}_t$ *are two families of σ-algebra's, where* $\mathcal{N}_t \subset \mathcal{F}_t$ *for all* t *and* $(\mathbf{n}_t)_{t\geqslant 0}$ *is adapted to* $(\mathcal{N}_t)_{t\geqslant 0}$, *then*

$$\lambda_t^{\mathcal{N}_t} = \mathrm{E}\,[\lambda_t^{\mathcal{F}_t}|\mathcal{N}_{t-}], \tag{5.7}$$

see for example BREMAUD (1977) p. 372 example or AALEN (1978) p. 706 Theorem 3.4. Hence, for any given i

$$\begin{aligned}\lambda_t^{\mathcal{N}_t} = \mathrm{E}\,[\lambda_t^{\mathcal{F}_t}|\mathcal{N}_{t-}] &= N\; \mathrm{E}\{\phi_j I\{\mathbf{n}_{t-}^{(j)} = 0|\mathcal{N}_{t-}\} \\ &= N.\; \mathrm{E}\,[\phi_j I\{\mathbf{n}_{t-}^{(j)} = 0\}|\mathcal{N}_{t-}, \text{ fault } j \text{ is not detected in } [0,t)]\times \\ &\qquad \Pr(\text{fault } j \text{ is not detected in } [0,t)|\mathcal{N}_{t-}) \\ &\quad + N.\; \mathrm{E}\,[\phi_j I\{\mathbf{n}_{t-}^{(j)} = 0\}|\mathcal{N}_{t-}, \text{ fault } j \text{ is detected in } [0,t)]\times \\ &\qquad \Pr(\text{fault } j \text{ is detected in } [0,t)|\mathcal{N}_{t-})\end{aligned} \tag{5.8}$$

The second term on the right hand side of (5.8) is zero, and

$$\Pr(\text{fault } j \text{ is not detected in } [0,t)\,|\,\mathcal{N}_{t-}) = \frac{N-\mathbf{n}_{t-}}{N} = \frac{N-i+1}{N}. \tag{5.9}$$

So (5.8) becomes

$$\begin{aligned}\lambda_t^{\mathcal{N}_t} &= N.\mathrm{E}\,[\phi_j I\{\mathbf{n}_{t-}^{(j)} = 0\}|\mathcal{N}_{t-}, \text{ fault } j \text{ is not detected in } [0,t)]\times\frac{N-\mathbf{n}_{t-}}{N} \\ &= (N-\mathbf{n}_{t-}).\; \mathrm{E}[\phi_j|\mathbf{T}_j \geqslant t],\end{aligned} \tag{5.10}$$

where $\mathbf{T}_j$ denotes the failure time of fault j.

$\boldsymbol{\phi}_j$ is gamma-distributed and given $\boldsymbol{\phi}_j = \phi_j$ the failure times are independent exponentially distributed with parameter ϕ_j, so

$$\boldsymbol{\phi}_j \sim \Gamma(a,b)\,,$$

$$\mathbf{T}_j|\boldsymbol{\phi}_j = \phi_j \sim \exp(\phi_j)\,.$$

The joint probability density function of $\boldsymbol{\phi}_j$ and $\mathbf{T}_j$ is therefore given by

$$f(\phi,T) = \phi\cdot e^{-T\phi}.\frac{b^a e^{-b\phi}\phi^{a-1}}{\Gamma(a)}, \quad T\geqslant 0,\; \phi\geqslant 0. \tag{5.11}$$

So the conditional pdf of $\boldsymbol{\phi}_j$ given $\mathbf{T}_j > t$ becomes

$$\begin{aligned}f(\phi|\mathbf{T}>t) &\propto \int_t^{\infty} \phi e^{-T\phi}\,.\,\frac{b^a e^{-b\phi}\phi^{a-1}}{\Gamma(a)}\mathrm{d}T \\ &\propto \phi^{a-1}e^{-(b+t)\phi},\end{aligned} \tag{5.12}$$

i.e. a gamma pdf: $\Gamma(a,b+t)$. Hence

$$\mathrm{E}\,[\phi_j|\mathbf{T}_j>t] = \frac{a}{b+t} \tag{5.13}$$

and therefore

$$\lambda_t^{\mathfrak{N}_t} = \frac{(N-\mathbf{n}_{t-})a}{b+t}. \tag{5.14}$$

LITTLEWOOD (1980) showed informally that, conditional on $\mathfrak{N}_t$, $\lambda_t^{\mathfrak{N}_t} \sim \Gamma((N-\mathbf{n}_{t-})a,\, b+t)$. This leads of course to the same result (5.14). The likelihood function now follows from (1.20):

$$\mathbf{L}_\tau = \left[\prod_{i=1}^{\mathbf{n}_\tau} \lambda_{\mathbf{T}_i}\right] \exp\left[-\int_0^\tau \lambda_s \mathrm{d}s\right]$$

$$= \left[\prod_{i=1}^{\mathbf{n}_\tau} \lambda_{\mathbf{T}_i}\right] \cdot \exp\left[-\sum_{i=1}^{\mathbf{n}_\tau+1} \int_{\mathbf{T}_{i-1}}^{\mathbf{T}_i\wedge\tau} \lambda_s \mathrm{d}s\right]. \tag{5.15}$$

In our case

$$\mathbf{L}_\tau = \prod_{i=1}^{\mathbf{n}_\tau} \frac{(N-i+1)a}{b+\mathbf{T}_i} \exp\left[-\sum_{i=1}^{\mathbf{n}_\tau+1} \int_{\mathbf{T}_{i-1}}^{\mathbf{T}_i\wedge\tau} \frac{(N-i+1)a}{b+t}\mathrm{d}t\right]$$

$$= \prod_{i=1}^{\mathbf{n}_\tau} \frac{(N-i+1)a}{b+\mathbf{T}_i}\cdot \exp\left[-\sum_{i=1}^{\mathbf{n}_\tau+1} (N-i+1)a \log\frac{b+\mathbf{T}_i\wedge\tau}{b+\mathbf{T}_{i-1}}\right]$$

$$= \prod_{i=1}^{\mathbf{n}_\tau} \frac{(N-i+1)a}{b+\mathbf{T}_i}\cdot \prod_{i=1}^{\mathbf{n}_\tau+1} \left[\frac{b+\mathbf{T}_{i-1}}{b+\mathbf{T}_i\wedge\tau}\right]^{(N-i+1)a}$$

$$= \left[\frac{b+\mathbf{T}_{\mathbf{n}_\tau}}{b+\tau}\right]^{(N-\mathbf{n}_\tau)a} \cdot \prod_{i=1}^{\mathbf{n}_\tau} \frac{(N-i+1)a(b+\mathbf{T}_{i-1})^{(N-i+1)a}}{(b+\mathbf{T}_i)^{(N-i+1)a+1}}. \tag{5.16}$$

Again this likelihood function equals the likelihood function that was intuitively to be expected:

$$\mathbf{L}_\tau = (1-F_{\mathbf{n}_\tau+1}(\tilde{\mathbf{t}}_{\mathbf{n}_\tau+1}|\mathbf{T}_{\mathbf{n}_\tau}))\cdot \prod_{i=1}^{\mathbf{n}_\tau} f_i(\mathbf{t}_i|\mathbf{T}_{i-1}), \tag{5.17}$$

since it can be shown that given $\mathbf{t}_1,\ldots\mathbf{t}_{i-1}$, the i'th interfailure time $\mathbf{t}_i$ is Pareto-distributed, with parameters only depending on $\mathbf{T}_{i-1}$, in particular

$$f_i(\mathbf{t}_i|\mathbf{T}_{i-1},a,b) = \frac{(N-i-1)a(b+\mathbf{T}_{i-1})^{(N-i-1)a}}{(b+\mathbf{T}_{i-1}+\mathbf{t}_i)^{(N-i+1)a+1}} \tag{5.18}$$

and

$$F_i(\mathbf{t}_i|\mathbf{T}_{i-1},a,b) = 1-\left[\frac{b+\mathbf{T}_{i-1}}{b+\mathbf{T}_{i-1}+\mathbf{t}_i}\right]^{(N-1+1)a}. \tag{5.19}$$

It is now easily seen that (5.17) equals (5.16). The log-likelihood function now becomes, where $\mathbf{L}_\tau = \mathbf{L}(N,a,b)$

$$\log \mathbf{L}(N,a,b) = \mathbf{n}_\tau \log a + (N-\mathbf{n}_\tau)a \log\frac{b+\mathbf{T}_{\mathbf{n}_\tau}}{b+\tau} +$$

$$\sum_{i=1}^{\mathbf{n}_\tau} [\log(N-i+1) + (N-i+1)a \log(b+\mathbf{T}_{i-1}) - \tag{5.20}$$

$$((N-i+1)a+1)\log(b+\mathbf{T}_i)].$$

The likelihood equations now become

$$\frac{\partial}{\partial N}\log \mathbf{L}|_{\hat{\mathbf{N}},\hat{\mathbf{a}},\hat{\mathbf{b}}} = -\hat{\mathbf{a}}\log\frac{\hat{\mathbf{b}}+\tau}{\hat{\mathbf{b}}} + \sum_{i=1}^{\mathbf{n}_\tau}\frac{1}{\hat{\mathbf{N}}-i+1} = 0 \tag{5.21}$$

$$\frac{\partial}{\partial a}\log \mathbf{L}|_{\hat{\mathbf{N}},\hat{\mathbf{a}},\hat{\mathbf{b}}} = \frac{\mathbf{n}_\tau}{\hat{\mathbf{a}}} + \hat{\mathbf{N}}\log\hat{\mathbf{b}} - (\hat{\mathbf{N}}-\mathbf{n}_\tau)\log(\hat{\mathbf{b}}+\tau) - \sum_{i=1}^{\mathbf{n}_\tau}\log(\hat{\mathbf{b}}+\mathbf{T}_i) = 0 \tag{5.22}$$

$$\frac{\partial}{\partial b}\log L|_{\hat{\mathbf{N}},\hat{\mathbf{a}},\hat{\mathbf{b}}} = \frac{\hat{\mathbf{a}}\hat{\mathbf{N}}}{\hat{\mathbf{b}}} - (1+\hat{\mathbf{a}})\sum_{i=1}^{\mathbf{n}_\tau}\frac{1}{\hat{\mathbf{b}}+\mathbf{T}_i} - \frac{\hat{\mathbf{a}}(\hat{\mathbf{N}}-\mathbf{n}_\tau)}{\hat{\mathbf{b}}+\mathbf{T}_{\mathbf{n}_\tau}} + \frac{\hat{\mathbf{a}}(\hat{\mathbf{N}}-\mathbf{n}_\tau)(\tau-\mathbf{T}_{\mathbf{n}_\tau})}{(\hat{\mathbf{b}}+\mathbf{T}_{\mathbf{n}_\tau})(\hat{\mathbf{b}}+\tau)} = 0, \tag{5.23}$$

when $\hat{\mathbf{a}},\hat{\mathbf{b}}$ and $\hat{\mathbf{N}}$ are the ML-estimators. As we did in the previous sections, let us write $N=\nu\beta_0$. In the following we denote by a_0,b_0 and β_0 the true model-parameters. $\hat{\mathbf{a}}^{(\nu)}$, $\hat{\mathbf{b}}_{(\nu)}$ and $\hat{\boldsymbol{\beta}}_{(\nu)}$ are the MLE's of a_0,b_0 and β_0 in the ν'th experiment. Then from (5.21)—(5.23) we get

$$0 = -\hat{\mathbf{a}}^{(\nu)}\log\frac{\hat{\mathbf{b}}^{(\nu)}+\tau}{\hat{\mathbf{b}}^{(\nu)}} + \sum_{i=1}^{\mathbf{n}_\tau^{(\nu)}}\frac{1}{\nu\hat{\boldsymbol{\beta}}^{(\nu)}-i+1} = \int_0^\tau \frac{1}{\nu\hat{\boldsymbol{\beta}}^{(\nu)}-\mathbf{n}_{t-}^{(\nu)}}\mathrm{d}\mathbf{m}_t^{(\nu)} \tag{5.24}$$

$$0 = \frac{\mathbf{n}_\tau^{(\nu)}}{\hat{\mathbf{a}}^{(\nu)}} + \nu\hat{\boldsymbol{\beta}}^{(\nu)}\log\hat{\mathbf{b}}^{(\nu)} - (\nu\hat{\boldsymbol{\beta}}^{(\nu)}-\mathbf{n}_\tau^{(\nu)})\log(\hat{\mathbf{b}}^{(\nu)}+\tau) - \sum_{i=1}^{\mathbf{n}_\tau^{(\nu)}}\log(\hat{\mathbf{b}}^{(\nu)}+\mathbf{T}_i)$$

$$= \int_0^\tau \frac{1}{\hat{\mathbf{a}}^{(\nu)}}\mathrm{d}\mathbf{m}_t^{(\nu)} \tag{5.25}$$

$$0 = \frac{\hat{\mathbf{a}}^{(\nu)}.\nu.\hat{\boldsymbol{\beta}}^{(\nu)}}{\hat{\mathbf{b}}^{(\nu)}} - (1+\hat{\mathbf{a}}^{(\nu)})\sum_{i=1}^{\mathbf{n}_\tau^{(\nu)}}\frac{1}{\hat{\mathbf{b}}^{(\nu)}+\mathbf{T}_i} - \frac{\hat{\mathbf{a}}^{(\nu)}(\nu\hat{\boldsymbol{\beta}}^{(\nu)}-\mathbf{n}_\tau^{(\nu)})}{\hat{\mathbf{b}}^{(\nu)}+\mathbf{T}_{\mathbf{n}_\tau^{(\nu)}}} +$$

$$\frac{\hat{\mathbf{a}}^{(\nu)}(\nu\hat{\boldsymbol{\beta}}^{(\nu)}-\mathbf{n}_\tau^{(\nu)})(\tau-\mathbf{T}_{\mathbf{n}_\tau^{(\nu)}})}{(\hat{\mathbf{b}}^{(\nu)}+\mathbf{T}_{\mathbf{n}_\tau}^{(\nu)})(\hat{\mathbf{b}}^{(\nu)}+\tau)} = \int_0^\tau \frac{-1}{\hat{\mathbf{b}}^{(\nu)}+t}\mathrm{d}\mathbf{m}_t^{(\nu)}. \tag{5.26}$$

where

$$\mathbf{m}_t^{(\nu)} = \mathbf{n}_t^{(\nu)} - \int_0^t \boldsymbol{\lambda}_s^{(\nu)}\mathrm{d}s.$$

From (5.25) we get

$$0 = \int_0^\tau \frac{1}{\hat{\mathbf{a}}^{(\nu)}}\mathrm{d}\mathbf{m}_t^{(\nu)} = \frac{\mathbf{n}_\tau^{(\nu)}}{\hat{\mathbf{a}}^{(\nu)}} - \int_0^\tau \frac{\nu\hat{\mathbf{b}}^{(\nu)}-\mathbf{n}_{t-}^{(\nu)}}{\hat{\mathbf{b}}^{(\nu)}+t}\mathrm{d}t.$$

So

$$\hat{\mathbf{a}}^{(\nu)} = \frac{\mathbf{n}_\tau^{(\nu)}/\nu}{\displaystyle\int_0^\tau\Big(\frac{\hat{\boldsymbol{\beta}}^{(\nu)}-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}}{\hat{\mathbf{b}}^{(\nu)}+t}\Big)\mathrm{d}t} = \frac{\mathbf{x}_\tau^{(\nu)}}{\hat{\boldsymbol{\beta}}^{(\nu)}\log\Big(\frac{\hat{\mathbf{b}}^{(\nu)}+\tau}{\mathbf{b}^{(\nu)}}\Big) - \displaystyle\int_0^\tau\frac{\mathbf{x}_{t-}^{(\nu)}}{\hat{\mathbf{b}}^{(\nu)}+t}\mathrm{d}t}. \tag{5.27}$$

From (5.26) we get

$$0 = \int_0^\tau \frac{-1}{\hat{\mathbf{b}}^{(\nu)}+t}\mathrm{d}\mathbf{m}_t^{(\nu)} = \int_0^\tau \frac{-1}{\hat{\mathbf{b}}^{(\nu)}+t}\mathrm{d}\mathbf{n}_t^{(\nu)} + \int_0^\tau \frac{\mathbf{a}^{(\nu)}(\nu\hat{\boldsymbol{\beta}}^{(\nu)}-\mathbf{n}_{t-}^{(\nu)})}{(\hat{\mathbf{b}}^{(\nu)}+t)^2}\mathrm{d}t.$$

This leads to

$$\hat{\beta}^{(\nu)} = \frac{\hat{\mathbf{b}}^{(\nu)}(\hat{\mathbf{b}}^{(\nu)}+\tau)}{\tau}\left[\int_0^\tau \frac{\mathbf{n}_{t-}^{(\nu)}/\nu}{(\hat{\mathbf{b}}^{(\nu)}+t)^2}\mathrm{d}t + \frac{1}{\hat{\mathbf{a}}^{(\nu)}}\int_0^\tau \frac{1}{\hat{\mathbf{b}}^{(\nu)}+t}\,\mathrm{d}\,\frac{\mathbf{n}_t^{(\nu)}}{\nu}\right]. \tag{5.28}$$

Now we can combine (5.27) and (5.28). Then (5.24)—(5.26) is equivalent to

$$\mathbf{f}^{(\nu)}(\mathbf{b}^{(\nu)}) \stackrel{\text{def}}{=} -\mathbf{a}^{(\nu)}(\hat{\mathbf{b}}^{(\nu)})\log\frac{\hat{\mathbf{b}}^{(\nu)}+\tau}{\hat{\mathbf{b}}^{(\nu)}} + \int_0^\tau \frac{1}{\beta^{(\nu)}(\hat{\mathbf{b}}^{(\nu)})-\frac{\mathbf{n}_t^{(\nu)}}{\nu}}\,\mathrm{d}\,\frac{\mathbf{n}_t^{(\nu)}}{\nu} = 0. \tag{5.29}$$

where

$$\mathbf{a}^{(\nu)}(b) \stackrel{\text{def}}{=} \frac{\mathbf{n}_\tau^{(\nu)}/\nu - (\log\frac{b+\tau}{b})\,\frac{b(b+\tau)}{\tau}\int_0^\tau \frac{1}{b+b}\mathrm{d}\frac{\mathbf{n}_t^{(\nu)}}{\nu}}{(\log\frac{b+\tau}{b})\,\frac{b(b+\tau)}{b}\int_0^\tau \frac{\mathbf{n}_{t-}^{(\nu)}/\nu}{(b+t)^2}\mathrm{d}t - \int_0^\tau \frac{\mathbf{n}_{t-}^{(\nu)}/\nu}{b+t}\mathrm{d}t} \tag{5.30}$$

and

$$\beta^{(\nu)}(b) \stackrel{\text{def}}{=} \frac{b(b+\tau)}{\tau}\left[\int_0^\tau \frac{\mathbf{n}_{t-}^{(\nu)}/\nu}{(b+t)^2}\mathrm{d}t + \frac{1}{\mathbf{a}^{(\nu)}(b)}\int_0^\tau \frac{1}{b+t}\mathrm{d}\frac{\mathbf{n}_t^{(\nu)}}{\nu}\right]. \tag{5.31}$$

So $\hat{\mathbf{b}}^{(\nu)}$ is a solution of $\mathbf{f}^{(\nu)}(b)=0$, where $\mathbf{f}^{(\nu)}(b)$ is given by (5.29)—(5.31)

Notice that it is very hard to find the maximum likelihood estimators. In practice the problem is often reduced by stating $a=1$. Thus one assumes an exponential a-priori distribution of ϕ_j.

Now we will look at the consistency of the estimators.

Theorem 5.1. *In the Littlewood-model*

$$\sup_{t\in[0,\tau]}\left|\frac{\mathbf{n}_t^{(\nu)}}{\nu} - \beta_0\left(1-\left(\frac{b_0}{b_0+t}\right)^{a_0}\right)\right| \xrightarrow{\mathrm{P}} 0 \ (\nu\to\infty), \tag{5.32}$$

where β_0, b_0 and a_0 are the true model-parameters.

Proof. Let in Theorem 1.1

$$\beta(s,x) \stackrel{\text{def}}{=} \frac{(\beta_0-x(s-))a_0}{b_0+s}\,.\ I\{0 \leqslant x(s-) \leqslant \beta_0\} \tag{5.33}$$

As in section 3 and 4, the indicator function is irrelevant, since $x(s-)$ will never exceed β_0.

We will verify conditions (1.14) and (1.15):

$$\begin{aligned}\sup_{s\leqslant t}\beta(s,x) &\leqslant \sup_{0\leqslant s\leqslant t}\left|\frac{(\beta_0-x(s-))a_0}{b_0+s}\right| \leqslant \sup_{0\leqslant s\leqslant t}\left|\frac{a_0}{b_0+s}\right|\cdot|\beta_0-x(s-)|\\ &\leqslant \left|\frac{a_0}{b_0}\right|\cdot|\beta_0|\end{aligned} \tag{5.34}$$

Let y denote another element of $D[0,\infty)$, then

$$\begin{aligned}|\beta(t,x)-\beta(t,y)| &= \left|\frac{(\beta_0-x(t-))a_0}{b_0+t} - \frac{(\beta_0-y(t-))a_0}{b_0+t}\right| \leqslant\\ &\left|\frac{a_0}{b_0+t}\right|\cdot\big|x(t-)-y(t-)\big|\ \big| \leqslant \left|\frac{a_0}{b_0}\right|\sup_{s\leqslant t}|\,x(s-)-y(s-)|\end{aligned} \tag{5.35}$$

So according to Theorem 1.1

$$\sup_{t \in [0,\tau]} \left| \frac{\mathbf{n}_t^{(\nu)}}{\nu} - x_t \right| \xrightarrow{\mathrm{P}} 0 \quad (\nu \to \infty),$$

where x_t is the solution of

$$\mathrm{d}x_t = \frac{(\beta_0 - x_s)a_0}{b_0 + s} \mathrm{d}s \,.$$

One can easily see that this solution is given by

$$x_t = \beta_0 \left(1 - \left(\frac{b_0}{b_0 + t}\right)^{a_0}\right). \qquad \square$$

COROLLARY 5.1. *In the Littlewood-model*

$$\frac{\mathbf{n}_t^{(\nu)}}{\nu} \xrightarrow{\mathrm{P}} \beta_0 \left(1 - \left(\frac{b_0}{b_0 + t}\right)^{a_0}\right) \quad (\nu \to \infty). \tag{5.36}$$

Define $a(b)$ as

$$a(b) = \frac{x_\tau - \left(\log\frac{b+\tau}{b}\right) \frac{b(b+\tau)}{\tau} \int_0^\tau \frac{1}{b+t} \mathrm{d}x_t}{\left(\log\frac{b+\tau}{b}\right) \frac{b(b+\tau)}{\tau} \int_0^\tau \frac{x_t}{(b+t)^2} \mathrm{d}t - \int_0^\tau \frac{x_t}{b+t} \mathrm{d}t}, \tag{5.37}$$

where

$$x_t = \beta_0 \left(1 - \left(\frac{b_0}{b_0 + t}\right)^{a_0}\right).$$

Since

$$\int_0^\tau \frac{1}{b_0 + t} \mathrm{d}x_t = \frac{\beta_0 a_0 (b_0 + \tau)^{a_0 + 1} - \beta_0 a_0 b_0^{a_0 + 1}}{(a_0 + 1) b_0 (b_0 + \tau)^{a_0 + 1}} \tag{5.38}$$

$$\int_0^\tau \frac{x_t}{(b_0 + t)^2} \mathrm{d}t = \frac{-\beta_0}{b_0 + \tau} + \frac{\beta_0}{b_0} + \frac{\beta_0 b_0^{a_0 + 1} - \beta_0 (b_0 + \tau)^{a_0 + 1}}{(a_0 + 1) b_0 (b_0 + \tau)^{a_0 + 1}} \tag{5.39}$$

and

$$\int_0^\tau \frac{x_t}{b_0 + t} \mathrm{d}t = \beta_0 \log\frac{b_0 + \tau}{b_0} + \frac{\beta_0 b_0^{a_0}}{a_0} (b_0 + \tau)^{-a_0} - \frac{\beta_0 b_0^{a_0}}{a_0} b_0^{-a_0}, \tag{5.40}$$

it follows that

$$a(b_0) = a_0 \,. \tag{5.41}$$

We may conclude from Theorem 5.1, that

$$\mathbf{a}^{(\nu)}(b) \xrightarrow{\mathrm{P}} a(b) \quad (\nu \to \infty) \tag{5.42}$$

for each fixed b. It follows also from Theorem 5.1, that

$$\boldsymbol{\beta}^{(\nu)}(b) \xrightarrow{\mathrm{P}} \beta(b) \quad (\nu \to \infty) \tag{5.43}$$

for each fixed b, where

$$\beta(b) \stackrel{\text{def}}{=} \frac{b(b+\tau)}{\tau}\left[\int_0^\tau \frac{x_t}{(b+t)^2}\mathrm{d}t + \frac{1}{a(b)}\int_0^\tau \frac{1}{b+t}\mathrm{d}x_t\right]. \tag{5.44}$$

One can easily see that $\beta(b_0)=\beta_0$. Define $f(b)$ as

$$f(b) = -a(b)\log\frac{b+\tau}{b} + \int_0^\tau \frac{1}{\beta(b)-x_t}\mathrm{d}x_t. \tag{5.45}$$

Then

$$\mathbf{f}^{(\nu)}(b) \xrightarrow{\mathbf{P}} f(b) \quad (\nu\to\infty),$$

for each fixed b as a consequence of Theorem 5.1. We have

$$f(b_0) = -a_0\log\frac{b_0+\tau}{b_0} + \int_0^\tau \frac{1}{\beta_0-x_t}\mathrm{d}x_t = -a_0\log\frac{b_0+\tau}{b_0} + \log\left(\frac{b_0+\tau}{b_0}\right)^{a_0} = 0. \tag{5.46}$$

We know that $\mathbf{f}^{(\nu)}(\hat{\mathbf{b}}^{(\nu)})=0$, if the MLE $\hat{\mathbf{b}}^{(\nu)}$ of b_0 exists. So for each $b' < b_0 < b''$

$$\lim_{\nu\to\infty} \Pr\,(\text{there is a solution of } \mathbf{f}^{(\nu)}(b)=0 \text{ in } [b',b'']) = 1. \tag{5.47}$$

Notice that we have not proved the uniqueness of the estimators. We showed that if the likelihood equations have one or more solutions at least one of them will be consistent, if we can make this choice using the data only.

For that consistent $\hat{\mathbf{b}}^{(\nu)}$ it can be verified that we have

$$\mathbf{a}^{(\nu)}(\hat{\mathbf{b}}^{(\nu)}) \xrightarrow{\mathbf{P}} a(b_0) = a_0 \tag{5.48}$$

$$\boldsymbol{\beta}^{(\nu)}(\hat{\mathbf{b}}^{(\nu)}) \xrightarrow{\mathbf{P}} \beta(b_0) = \beta_0\,. \tag{5.49}$$

In conclusion, if there is more then one set of solutions $\begin{bmatrix}\hat{\boldsymbol{\beta}}^{(\nu)}\\ \hat{\mathbf{a}}^{(\nu)}\\ \hat{\mathbf{b}}^{(\nu)}\end{bmatrix}$ of (5.29)—(5.31) at least one set $\begin{bmatrix}\hat{\boldsymbol{\beta}}^{(\nu)}\\ \hat{\mathbf{a}}^{(\nu)}\\ \hat{\mathbf{b}}^{(\nu)}\end{bmatrix}$ will be consistent.

At the end of this section, we look at the asymptotic distribution of the MLE $\hat{\boldsymbol{\theta}}^{(\nu)}$. If we assume consistency of $\hat{\boldsymbol{\theta}}^{(\nu)}$, asymptotic normality will follow easily.

THEOREM 5.2. *If $\hat{\boldsymbol{\theta}}^{(\nu)}$ is a consistent estimator of $\theta_0 = \begin{bmatrix}\beta_0\\ a_0\\ b_0\end{bmatrix}$, then for $\nu\to\infty$:*

$$\sqrt{\nu}\,(\hat{\boldsymbol{\theta}}^{(\nu)}-\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma_L(\theta_0)^{-1}), \tag{5.50}$$

where $\Sigma_L(\theta_0)$ (here L denotes Littlewood) is given by $\Sigma_L(\theta_0) =$

$$\begin{bmatrix} ((b_0+\tau)^{a_0}-b_0^{a_0})\,\dfrac{1}{\beta_0 b_0^{a_0}} & \log(\dfrac{b_0+\tau}{b_0}) & \dfrac{-\alpha_0\tau}{b_0(b_0+\tau)} \\ \log(\dfrac{b_0+\tau}{b^0}) & \dfrac{\beta_0}{a_0^2}(1-(\dfrac{b_0}{b_0+\tau})^{a_0}) & \dfrac{\beta_0}{b_0(a_0+1)}((\dfrac{b_0}{b_0+\tau})^{a_0+1}-1) \\ \dfrac{-\alpha_0\tau}{b_0(b_0+\tau)} & \dfrac{\beta_0}{b_0(a_0+1)}((\dfrac{b_0}{b_0+\tau})^{a_0+1}-1) & \dfrac{a_0\beta_0}{b_0^2(a_0+2)}(1-(\dfrac{b_0}{b_0+\tau})^{a_0+2}) \end{bmatrix}. \tag{5.51}$$

PROOF. See Appendix B. □

We will try to find estimators for the parameters in the model of Littlewood, with the data from Table 3.1. Because it was very difficult to find maximum likelihood estimators in this case we also made use of two other different methods.

The first alternative method comes from an idea of K.O. Dzhaparidze. $\mathbf{n}_t^{(\nu)}$ is a counting process, with intensity

$$\boldsymbol{\lambda}^{(\nu)}(t,\theta) = \frac{\nu\beta - \mathbf{n}_{t-}^{(\nu)}}{b+t} \cdot a \cdot I\{\mathbf{n}_{t-}^{(\nu)} \leqslant \nu\beta\}\ ,\ \theta = \begin{bmatrix} a \\ b \\ \beta \end{bmatrix}.$$

$\mathbf{n}_t^{(\nu)}$ has compensator

$$\mathbf{a}_{t:\theta}^{(\nu)} = \int_0^t \boldsymbol{\lambda}^{(\nu)}(s;\theta)\mathrm{d}s\,.$$

Define

$$\mathbf{m}_{t:\theta}^{(\nu)} = \mathbf{n}_t^{(\nu)} - \mathbf{a}_{t:\theta}^{(\nu)},$$

also define $\mathbf{x}_t^{(\nu)} = \nu^{-1}\mathbf{n}_t^{(\nu)}$ and $\mathbf{w}_{t:\theta}^{(\nu)} = \nu^{-1}\mathbf{m}_{t:\theta}^{(\nu)}$. Consider $\int_0^T \boldsymbol{\psi}_{t:\theta}^{(\nu)}\,\mathrm{d}\mathbf{w}_{t:\theta}^{(\nu)}$ for some predictable process $\boldsymbol{\psi}^{(\nu)}$, depending on the parameters. Suppose:

$$\boldsymbol{\psi}_{t:\theta}^{(\nu)} \xrightarrow{\mathrm{P}} \psi_{t:\theta} \quad \text{as} \quad \nu \to \infty\,.$$

Therefore we expect

$$\int_0^T (\boldsymbol{\psi}_{t:\theta}^{(\nu)})^2\ \nu^{-1}\ \boldsymbol{\lambda}_{t:\theta}^{(\nu)}\ \mathrm{d}t \xrightarrow{\mathrm{P}} M,$$

with M finite. By Lenglart's inequality (1.8) we therefore have

$$\int_0^T \boldsymbol{\psi}_{t:\theta}^{(\nu)}\ \mathrm{d}\mathbf{w}_{t:\theta}^{(\nu)} \xrightarrow{\mathrm{P}} 0 \quad (\nu \to \infty)\,.$$

We also expect that both

$$\int_0^T \boldsymbol{\psi}_{t:\theta}^{(\nu)}\ \mathrm{d}\mathbf{x}_t^{(\nu)} \xrightarrow{\mathrm{P}} \int_0^T \psi_t \mathrm{d}x_t$$

and

$$\int_0^T \boldsymbol{\psi}_{t:\theta}^{(\nu)}\ \nu^{-1}\ \boldsymbol{\lambda}_{t:\theta}^{(\nu)} \xrightarrow{\mathrm{P}} \int_0^T \psi_t \mathrm{d}x_t\,.$$

This suggests estimating θ by choosing three different processes $\boldsymbol{\psi}^{(\nu)}$ and solving the equations:

$$\int_0^T \boldsymbol{\psi}_{t:\theta}^{(\nu)}\ \mathrm{d}\mathbf{w}_{t:\theta}^{(\nu)} = 0.$$

Taking $\boldsymbol{\psi}_{t:\theta}^{(\nu)} = \frac{\partial}{\partial\theta} \log \boldsymbol{\lambda}_{t:\theta}^{(\nu)}$ gives the likelihood equations which are hard to solve. However, if we take $\boldsymbol{\psi}_{t:\theta}^{(\nu)} = (b+t)t^r$, for $r=0,1$ and 2, we get an easy problem. In this case the estimating equations become (on the event $\mathbf{x}_T^{(\nu)} < \beta$):

$$b(\int_0^T t^r\ \mathrm{d}\mathbf{x}_t^{(\nu)}) + (\int_0^T t^{r+1}\ \mathrm{d}\mathbf{x}_t^{(\nu)}) - \beta a(\int_0^T t^r\ \mathrm{d}t) + a\int_0^T t^r\ \mathbf{x}_t^{(\nu)}\ \mathrm{d}t = 0 \qquad r=0,1,2. \tag{5.52}$$

The inequality of Lenglart can be applied, showing that the left hand side of (5.52) converges in probability to zero, as $\nu\to\infty$, at the true parameter values. By KURTZ (1981, 1983) each term in (5.52) involving $\mathbf{x}_t^{(\nu)}$ converges in probability to the corresponding term in x_t. We rewrite (5.52) as:

$$\begin{bmatrix} \mathbf{x}_t^{(\nu)} & -\tau & \tau\mathbf{x}_\tau^{(\nu)} - \int_0^\tau t\,d\mathbf{x}_t^{(\nu)} \\ \int_0^\tau t\,d\mathbf{x}_t^{(\nu)} & -\frac{1}{2}\tau^2 & \frac{1}{2}\tau^2\mathbf{x}_\tau^{(\nu)} - \frac{1}{2}\int_0^\tau t^2 d\mathbf{x}_t^{(\nu)} \\ \int_0^\tau t^2\,d\mathbf{x}_t^{(\nu)} & -\frac{1}{3}\tau^3 & \frac{1}{3}\tau^3\mathbf{x}_\tau^{(\nu)} - \frac{1}{3}\int_0^\tau t^3 d\mathbf{x}_t^{(\nu)} \end{bmatrix} \begin{bmatrix} b \\ \beta a \\ a \end{bmatrix} = \begin{bmatrix} -\int_0^\tau t\,d\mathbf{x}_t^{(\nu)} \\ -\int_0^\tau t^2 d\mathbf{x}_t^{(\nu)} \\ -\int_0^\tau t^3 d\mathbf{x}_t^{(\nu)} \end{bmatrix} \tag{5.53}$$

then, provided the probability limit of the matrix on the left hand side of (5.53) is nonsingular, the solution $(\hat{\mathbf{b}}^{(\nu)}\ (\hat{\beta\mathbf{a}})^{(\nu)}\ \hat{\mathbf{a}}^{(\nu)})^{\mathrm{T}}$ is a consistent estimator of $(b\ \beta a\ a)^{\mathrm{T}}$, and hence also yields a consistent estimator of $\theta = (a\ b\ \beta)^{\mathrm{T}}$. The mentioned nonsingularity has unfortunately not been proved yet.

For practical use, take $\nu = \mathbf{n}_\tau$, let $\overline{\mathbf{T}}^{(r)}, r = 1,2,3$, be the means of $\mathbf{T}_i^r$, $i = 1,\ldots,\mathbf{n}_\tau$, where $\mathbf{n}_{\mathrm{T}_i-} = i-1$ and $\mathbf{n}_{\mathrm{T}_i} = i$. In this case the estimating equations become

$$\begin{bmatrix} 1 & -\tau & \tau-\overline{\mathbf{T}}^{(1)} \\ \overline{\mathbf{T}}^{(1)} & -\frac{1}{2}\tau^2 & \frac{1}{2}\tau^2-\frac{1}{2}\overline{\mathbf{T}}^{(2)} \\ \overline{\mathbf{T}}^{(2)} & -\frac{1}{3}\tau^3 & \frac{1}{3}\tau^3-\frac{1}{3}\overline{\mathbf{T}}^{(3)} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\beta\mathbf{a}} \\ \hat{\mathbf{a}} \end{bmatrix} = - \begin{bmatrix} \overline{\mathbf{T}}^{(1)} \\ \overline{\mathbf{T}}^{(2)} \\ \overline{\mathbf{T}}^{(3)} \end{bmatrix} \tag{5.54}$$

with $\hat{\mathbf{N}} = \hat{\boldsymbol{\beta}}\,\mathbf{n}_\tau$. Later in this section we will give the practical results.

Our second method is based on the least squares principle. First a remark: in the model of Littlewood

$$\boldsymbol{\lambda}_t = \frac{(N-\mathbf{n}_{t-})a}{b+t}\,1\{\mathbf{n}_{t-} \leqslant N\} = \frac{(N-\mathbf{n}_{t-})}{\frac{b}{a}+\frac{1}{a}t}\,1\{\mathbf{n}_{t-} \leqslant N\}$$

we use a reparametrization, namely $\rho = \frac{1}{a}$ and $\mu = \frac{b}{a}$, so we get

$$\boldsymbol{\lambda}_t = \frac{(N-\mathbf{n}_{t-})}{\mu+\rho t}\,1\{\mathbf{n}_{t-} \leqslant N\}. \tag{5.55}$$

Actually this is an extension of our previous model if we now also include values of $\rho\leqslant 0$. When $\rho=0$ we deal with the Musa-model. Therefore we conclude that the model of Musa is a special case of the Littlewood-model.

In the model of Littlewood $x_t = \beta(1-(\frac{b}{b+t})^a)$ (5.36). By the same reparametrization we can rewrite

$$x_t = \beta(1-(\frac{\mu}{\mu+\rho t})^{\frac{1}{\rho}}). \tag{5.56}$$

Notice that

$$\lim_{\rho\to 0} x_t = \beta(1-e^{-t/\mu}).$$

The limit is the solution x_t in the model of Musa (see (3.13)).

Define

$$V(\mu,\rho,\beta) = \sum_{i=1}^{\mathbf{n}_\tau} (\mathbf{x}_{\mathrm{T}_i} - \beta\,(1-(\frac{\mu}{\mu+\rho t})^{\frac{1}{\rho}}))^2, \quad \mathbf{x}_{\mathrm{T}_i} = \mathbf{n}_{\mathrm{T}_i}\,.$$

Following the least squares method we try to calculate

$$\min_{\mu,\rho,\beta} V(\mu,\rho,\beta)\,.$$

We didn't prove the consistency of the estimators based on the least-squares method.

RESULTS. Using the method of *Dzhaparidze* we get:

$$\hat{\mathbf{N}} = \nu\hat{\boldsymbol\beta} = 46.4$$
$$\hat{\mathbf{b}} = 2.2$$
$$\hat{\mathbf{a}} = 11.3$$

Using the *least-squares* method we get:

$$\hat{\mathbf{N}} = \nu\hat{\boldsymbol\beta} = 47.2 \qquad \hat{\mathbf{b}} = 1.4$$
$$\hat{\mu} = 0.19 \qquad \hat{\mathbf{a}} = 7.3$$
$$\hat{\rho} = 0.14.$$

Both methods give realistic estimators for N. The estimators for μ and ρ can be interpreted as follows: any failure has expected failure rate $\frac{1}{\hat{\mu}} = 5.2$ per Msec with standard deriation $\frac{1}{\hat{\mu}}\sqrt{\hat{\rho}} = \sqrt{3.8}$ per Msec (results least squares method).

Maximum likelihood estimation: In the Littlewood-model

$$\boldsymbol\lambda_t = \frac{(N-\mathbf{n}_{t-})}{\mu+\rho t}\,1\{\mathbf{n}_{t-} \leqslant N\}. \tag{5.55}$$

The log-likelihood function equals

$$\log\,(\prod_i \boldsymbol\lambda_{\mathbf{T}_i}\cdot\exp(-\int_0^T \boldsymbol\lambda_s \mathrm{d}s)). \tag{5.57}$$

Notice

$$\int_{\mathbf{T}_{i-1}}^{\mathbf{T}_i} \boldsymbol\lambda_t \mathrm{d}t = \frac{(N-i+1)}{\rho}\log(\frac{\mu+\rho\mathbf{T}_i}{\mu+\rho\mathbf{T}_{i-1}})$$

$$= \frac{-(N-i+1)}{\rho}\log(1-\frac{\rho\mathbf{t}_i}{\mu+\rho\mathbf{T}_i}),\quad \mathbf{t}_i = \mathbf{T}_i - \mathbf{T}_{i-1}\,. \tag{5.58}$$

In case we stop at a certain (43rd) failure time (observed $n = 43$) the log-likelihood equation becomes:

$$\sum_{i=1}^{n}\left\{\log(\frac{N-i+1}{\mu+\rho\mathbf{T}_i}) + \frac{(N-i+1)}{\rho}\log(1-\frac{\rho\mathbf{t}_i}{\mu+\rho\mathbf{T}_i})\right\}. \tag{5.59}$$

For $\frac{1}{\rho}\log(1-\frac{\rho\mathbf{t}_i}{\mu+\rho\mathbf{T}_i})$ we make use of a Taylor-expension (ρ close to zero): for $|\rho| < 0.01$ we use

$$\frac{1}{\rho}\log\,(1-\frac{\rho\mathbf{t}_i}{\mu+\rho\mathbf{T}_i}) \approx -\frac{\mathbf{t}_i}{\mu+\rho\mathbf{T}_i} - \tfrac{1}{2}\rho\,(\frac{\mathbf{t}_i}{\mu+\rho\mathbf{T}_i})^2\,.$$

To determine the maximum of the log-likelihood we use the following method: We search for the maximum of the log-likelihood in a $5 \times 5 \times 5$ grid. For example: $N = 43, 44, 45, 46$ and 47; $\mu=0.10$, $1.15, 0.20, 0.25$ and 0.30; $\rho = 0, 0.01, 0.1, 1.10$. We then compute the maximum of the log-likelihood in a new $5 \times 5 \times 5$ grid, in the neighboorhood of the values of N, μ and ρ for which the maximum was achieved in the preceding step, and so on.

REMARK 1: For negative values of ρ close to zero, this still yields a meaningful model.

REMARK 2: Another useful method is one-step Newton-Raphson iteration starting with consistent estimators, for example, using the method of Dzhaparidze, when—in that case—consistency has been proved. Denote $\hat{\boldsymbol{\theta}}_D$ as the estimators using the Dzhaparidze-method. We can write the one-step Newton-Raphson method as:

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_D - \left(\frac{\partial^2}{\partial\theta\partial\theta^{\mathrm{T}}} \log \mathbf{L}(\hat{\boldsymbol{\theta}}_D)\right)^{-1} \frac{\partial}{\partial\theta} \log \mathbf{L}\,(\hat{\boldsymbol{\theta}}_D)$$

RESULTS. Using the grid method and putting $\rho \geqslant 0$, the maximum of the log-likelihood turns out to be 156.4. This value is achieved for $N=44.5$; $\mu=0.18$ and $\rho=0$. Notice that $\rho=0$ means that we are in the Musa-model. This confirms our conjecture that the maximum of (5.20) is achieved for infinite a.

We also computed the log-likelihood, using the results from the method of Dzhaparidze and the least squares method respectively (see above). The results are given in Table 5.1. When we also allow negative values of ρ (notice that $\mu+\rho\ \mathrm{T}_i>0\ \forall_i$, and also $N \geqslant 43$) the maximum of the log-likelihood equals 156.9.

N	μ	ρ	log-likelihood	
44.5	0.18	0	156.4	Musa
46.34	0.21	0.09	156.1	Dzhaparidze
46.92	0.21	0.14	156.1	Least Squares
43	0.21	-0.26	156.9	M.L.E.

TABLE 5.1. Comparison of results.

Notice that the difference between the maximum log-likelihood (156.9) and the log-likelihood at $\rho=0$ (Musa: 156.4) is smaller than $\frac{1}{2}\chi_1^2$ (.95) = 1.97, so we cannot reject the Musa model. By computing $\mathrm{E}(\mathbf{n}_t)$ we can draw the theoretical curve which we are able to compare with the given failure data.

$$\mathrm{E}(\mathbf{n}_t) = n_t \ = N\left(1 - \frac{\mu^{1/\rho}}{(\mu+\rho t)^{1/\rho}}\right) \qquad \text{(see (5.56))}$$

$$= N\left(1-\left(1+\frac{\rho t}{\mu}\right)^{-1/\rho}\right) \qquad (5.59)$$

Notice that $n_t \approx N(1- e^{t/\mu})$ in the case that $\rho \approx 0$. Both theoretical curve and failure data are drawn in Figure 5.2.

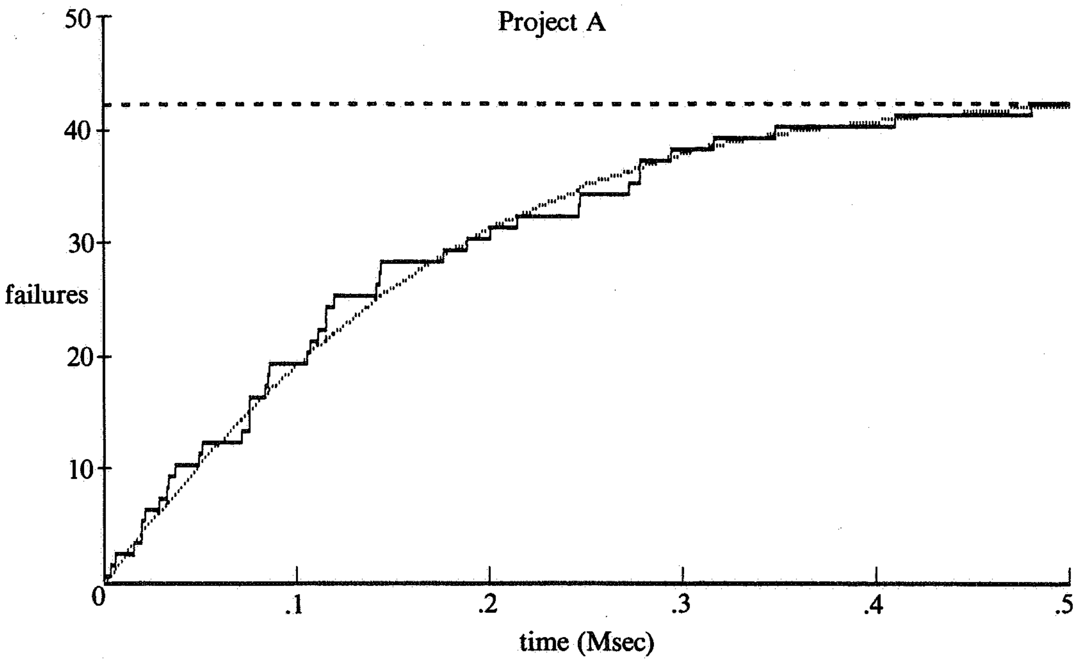

FIGURE 5.2. Project A.

6. GOODNESS OF FIT TESTS

A classical problem in statistics is the construction of the so-called goodness of fit-tests, to check whether a model provides a good fit to a given set of (failure) data. Let us consider the classical situation. Let $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be a sequence of independent identically distributed random variables with an unknown distribution function $F(x)$.We consider the goodness of fit test:

$$H_0 : \mathbf{x}_1, \ldots, \mathbf{x}_n \sim F_\theta, \text{ for some } \theta \in \Theta \subseteq \mathbb{R}^p$$

versus

$$H_1 : \mathbf{x}_1, \ldots, \mathbf{x}_n \sim F \neq F_\theta, \text{ for any } \theta .$$

Thus we test the null hypothesis that the $\mathbf{x}_i$ come from a specified parametric model, $F = F_\theta$, for some $\theta \in \Theta$.

A way to attack this problem is as follows. We can compute both $\mathbf{F}_n$, the empirical distribution function of $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and $F_{\hat{\theta}_n}$. Consider the difference $\mathbf{F}_n - F_{\hat{\theta}_n}$; under H_0 this should be small in some sense, but larger under H_1. A sensible goodness of fit test is the Kolmogorov-Smirnov test. The K-S-test-statistic equals $\sup_t |\, \mathbf{F}_n(t) - F_{\hat{\theta}_n}(t)|$. We recall that for the set Θ consisting of a simple point θ_0, that is when a simple hypothesis is tested, $n^{\frac{1}{2}}(\mathbf{F}_n - F_0) \xrightarrow{\mathcal{D}} B^0 \circ (F_0)$, $F_0 = F_{\theta_0} = F_{\hat{\theta}_n}$. Here B^0 denotes the *Brownian bridge.* Otherwise the large sample distribution of $\mathbf{F}_n - F_{\hat{\theta}_n}$ and hence also of the Kolmogorov-Smirnov test statistic is rather complicated. A large literature exists on how to attack this problem. We come back to various proposals later.

More focussed on our situation we test:

$$H_0 : \mathbf{n} \text{ has intensity } \boldsymbol{\lambda}_{t,\theta}, \;\; \theta \in \Theta$$

versus

$$H_1 : \mathbf{n} \text{ does not have intensity } \boldsymbol{\lambda}_{t,\theta}, \; \theta \in \Theta .$$

Instead of $\mathbf{F}_n - F_{\hat{\theta}_n}$, we now look at $\mathbf{n} - \mathbf{a}_{\hat{\theta}}$, with $\mathbf{a}_{t;\hat{\theta}} = \int_0^t \boldsymbol{\lambda}(s, \hat{\theta}) ds$.

REMARK 1. In the case that $\Theta = \{\theta_0\}$, that is H_0:$\mathbf{a} = \mathbf{a}_0$ ('known'), $\mathbf{n} \circ \mathbf{a}_0^{-1}$ is distributed as a Poisson process, so that a goodness of fit test is quite easy to construct. Another special case is H_0:$\mathbf{a} = c\mathbf{a}_0$ ($\mathbf{a}_0$ known, c unknown) which leads to the *total time on test plot* (of $\mathbf{n}$ versus $\mathbf{a}_0$), see GILL (1986).

$\mathbf{n}_t^{(\nu)}$ is a counting process with intensity $\boldsymbol{\lambda}^{(\nu)}(t;\theta) = \nu\beta_t\,(\mathbf{x}^{(\nu)};\theta)$ and compensator $\mathbf{a}_{t;\theta}^{(\nu)} = \int_0^t \boldsymbol{\lambda}^{(\nu)}(s,\theta)\mathrm{d}s$. Here $\mathbf{x}_t^{(\nu)} = \mathbf{n}_t^{(\nu)}/\nu$, let x_t be the solution of $x_0 = 0$,

$$\mathrm{d}x_t = \beta_t(x;\theta)\mathrm{d}t. \tag{6.1}$$

Furthermore we define

$$\mathbf{m}_\theta^{(\nu)} = \mathbf{n}^{(\nu)} - \mathbf{a}_\theta^{(\nu)}, \text{ and } \mathbf{w}_\theta^{(\nu)} = \nu^{-\frac{1}{2}}\,\mathbf{m}_\theta^{(\nu)}. \tag{6.2}$$

Let

$$\mathbf{h}_{t;\theta}^{(\nu)} = \frac{\partial}{\partial\theta}\log\beta_t\,(\mathbf{x}^{(\nu)};\theta),\; h_{t;\theta} = \frac{\partial}{\partial\theta}\log\beta_t(x;\theta). \tag{6.3}$$

Suppose $\hat{\boldsymbol{\theta}}^{(\nu)}$ is a consistent maximum likelihood estimator of θ (under the null hypothesis). Using the martingale central limit theorem one could prove under H_0 (under regularity assumptions) that

$$\nu^{\frac{1}{2}}\left[\hat{\boldsymbol{\theta}}^{(\nu)} - \theta\right] \xrightarrow{\mathcal{D}} \left[\int_0^T h^{\otimes 2}\mathrm{d}x\right]^{-1}\left[\int_0^T h\mathrm{d}w\right], \tag{6.4}$$

where $h^{\otimes 2} = hh^{\mathrm{T}}$ (i.e. a square matrix with elements h_ih_j).

This result would be obtained as follows. Consider the following Taylor-expansions

$$0 = \left[\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0)|_{\hat{\boldsymbol{\theta}}}\right]_i = \left[\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log\mathbf{L}(\theta_0)\right]_i + \frac{1}{\nu}\sum_{j=1}^{2}\sqrt{\nu}\,(\hat{\boldsymbol{\theta}}-\theta_0)_j\,\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\mathbf{L}(\theta_0)$$

$$+\frac{1}{2}\frac{1}{\nu}\sum_{j=1}^{2}\sum_{k=1}^{2}\sqrt{\nu}\,(\hat{\boldsymbol{\theta}}-\theta_0)_j\,(\hat{\boldsymbol{\theta}}-\theta_0)_k\,\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\log\mathbf{L}\,(\hat{\boldsymbol{\theta}}^{(ij)}),\; i = 1,2$$

with

$$|(\hat{\boldsymbol{\theta}}^{(ij)} - \theta_0)_l| < |(\hat{\boldsymbol{\theta}} - \theta_0)_l|\;\forall l \tag{6.5}$$

(using the mean value theorem for every i and j separately). Under mild conditions it can be proved that

$$\Pr\left[\frac{1}{\nu}\left|\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\log\mathbf{L}\,(\hat{\boldsymbol{\theta}}^{(ij)})\right| < M,\text{ some finite constant}\right] \to 1 \quad (\nu\to\infty),$$

provided that $\hat{\boldsymbol{\theta}}$ is a consistent maximum likelihood estimator of θ, hence:

$$0 \approx \left[\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log\mathbf{L}\,(\theta_0)\right]_i + \frac{1}{\nu}\sum_{j=1}^{2}\sqrt{\nu}(\hat{\boldsymbol{\theta}}-\theta_0)_j\,\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\mathbf{L}(\theta_0).$$

so,

$$0 \approx \frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log\mathbf{L}(\theta_0) + \sqrt{\nu}\,(\hat{\boldsymbol{\theta}}^{(\nu)} - \theta_0)\cdot\frac{1}{\nu}\frac{\partial^2}{\partial\theta\partial\theta^{\mathrm{T}}}\log\mathbf{L}(\theta_0). \tag{6.6}$$

Here,

$$\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log\mathbf{L}(\theta_0) = \frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\left[\int_0^T\log\boldsymbol{\lambda}^{(\nu)}(t;\theta)\mathrm{d}\mathbf{n}_t^{(\nu)} - \int_0^T\boldsymbol{\lambda}^{(\nu)}(t;\theta)\mathrm{d}t\right]$$

$$= \frac{1}{\sqrt{\nu}} \int_0^\tau \frac{\partial}{\partial\theta} \log \lambda^{(\nu)}(t;\theta)\mathrm{d}n_t^{(\nu)} - \int_0^\tau \frac{\partial}{\partial\theta} \lambda^{(\nu)}(t;\theta)\mathrm{d}t$$

$$= \frac{1}{\sqrt{\nu}} \left[\int_0^\tau \frac{\partial}{\partial\theta} \log \beta_t(\mathbf{x}^{(\nu)};\theta)\mathrm{d}n_t^{(\nu)} - \int_0^\tau \frac{\partial}{\partial\theta} \log \beta_t(\mathbf{x}^{(\nu)};\theta) \cdot \lambda^{(\nu)}(t;\theta)\mathrm{d}t\right]$$

$$= \frac{1}{\sqrt{\nu}} \int_0^\tau \mathbf{h}^{(\nu)}\mathrm{d}\mathbf{m}_t^{(\nu)} = \int_0^\tau \mathbf{h}^{(\nu)}\mathrm{d}\mathbf{w}_t^{(\nu)}. \tag{6.7}$$

In the same way it can be proved that

$$\frac{1}{\nu} \frac{\partial^2}{\partial\theta\partial\theta^{\mathrm{T}}} \log \mathbf{L}(\theta_0) = -\int_0^\tau \mathbf{h}^{(\nu)\otimes 2}\mathrm{d}\mathbf{x}^{(\nu)} \xrightarrow{\mathrm{P}} -\int_0^\tau h^{\otimes 2}\mathrm{d}x. \tag{6.8}$$

Result (6.4) follows from (6.6), (6.7) and (6.8).

Summarizing, using KURTZ (1981, 1983) and the martingale central limit theorem we can obtain in the usual way, under reasonable smoothness assumptions and consistency of the MLE

$$\mathbf{x}^{(\nu)} \xrightarrow{\mathrm{P}} x \ (\nu\to\infty) \text{ in } D[0,\tau],\ \|\cdot\|$$

$$\begin{bmatrix} \nu^{-\frac{1}{2}}\ \mathbf{m}^{(\nu)} \\ \nu^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^{(\nu)} - \theta_0) \end{bmatrix} \xrightarrow{\mathcal{D}} \begin{bmatrix} \mathbf{w} \\ (\int_0^\tau h^{\otimes 2}\mathrm{d}x)^{-1}\ (\int_0^\tau h\mathrm{d}\mathbf{w}) \end{bmatrix} \tag{6.9}$$

as $\nu\to\infty$ in $(D[0,\tau]\times\mathbb{R})$, $J\times d$. Here $\mathbf{w}$ is a zero mean Gaussian martingale with

$$<\mathbf{w}>_t = \mathrm{var}\,(\mathbf{w}_t) = x_t$$

($\|\cdot\|$ is the supremum norm in $D[0,\tau]$; J is the Skorohod-metric on $D[0,\tau]$, and d is Euclidean distance in $\mathbb{R}$).

It seems reasonable to base a goodness of fit test on the process $\mathbf{m}^{(\nu)}{}_{\hat\theta^{(\nu)}} = \mathbf{n}^{(\nu)} - \mathbf{a}^{(\nu)}{}_{\hat\theta^{(\nu)}}$. Notice that $\mathbf{m}^{(\nu)}{}_{\hat\theta^{(\nu)}} \circ (\mathbf{a}^{(\nu)}{}_{\hat\theta^{(\nu)}})^{-1}$ is the total time on test plot (see e.g. GILL (1986)) before rescaling to $[0,1] \times [0,1]$. Thus, in particular, the Kolmogorov-Smirnov test-statistic based on $\mathbf{m}^{(\nu)}{}_{\hat\theta^{(\nu)}}$ — i.e. the statistic $\|\mathbf{m}^{(\nu)}{}_{\hat\theta^{(\nu)}}\|$ — is the same as that based on the total time on test plot, up to rescaling by the amount $\mathbf{n}_\tau^{(\nu)}$ (and $\nu^{-1}\ \mathbf{n}_\tau^{(\nu)} \xrightarrow{\mathrm{P}}$ a constant). By a Taylor-expansion, it can be proved

$$\nu^{-\frac{1}{2}}\ (\mathbf{a}^{(\nu)}{}_{\hat\theta^{(\nu)}} - \mathbf{a}_\theta^{(\nu)}) \approx \nu^{\frac{1}{2}}\ (\hat{\boldsymbol{\theta}}^{(\nu)} - \theta)^{\mathrm{T}} \left[\int_0^{(\cdot)} h\mathrm{d}x\right]. \tag{6.10}$$

We give a heuristic proof of (6.10). We know

$$\nu^{\frac{1}{2}} \left[\frac{\mathbf{a}^{(\nu)}{}_{\hat\theta^{(\nu)}}}{\nu} - \frac{\mathbf{a}_\theta^{(\nu)}}{\nu}\right] = \nu^{\frac{1}{2}}\ (\hat{\boldsymbol{\theta}}^{(\nu)} - \theta)\ \frac{\partial}{\partial\theta}\ \frac{\mathbf{a}_\theta^{(\nu)}}{\nu} + \nu^{-\frac{1}{2}} \cdot \frac{1}{2}(\nu^{\frac{1}{2}}\ (\hat{\boldsymbol{\theta}}^{(\nu)} - \theta))^2\ \frac{\partial^2}{\partial\theta^2}\ \frac{\mathbf{a}^{(\nu)}{}_{\bar\theta}}{\nu}. \tag{6.11}$$

Using the fact that $\hat{\boldsymbol{\theta}}^{(\nu)}$ is a $\sqrt{\nu}$-consistent estimator for θ, it will be possible to prove that the last term in (6.11) converges in probability to zero, when $\nu\to\infty$. Furthermore

$$\frac{\partial}{\partial\theta}\ \frac{\mathbf{a}_\theta^{(\nu)}}{\nu} = \int_0^{(\cdot)} \mathbf{h}^{(\nu)}\boldsymbol{\beta}^{(\nu)} \xrightarrow{\mathrm{P}} \int_0^{(\cdot)} h\mathrm{d}x\,.$$

We can write

$$\nu^{-\frac{1}{2}}\ \mathbf{m}^{(\nu)}_{\hat{\theta}^{(\nu)}} = \nu^{-\frac{1}{2}}\ (\mathbf{n}^{(\nu)} - \mathbf{a}^{(\nu)}_{\hat{\theta}^{(\nu)}}) = \nu^{-\frac{1}{2}}\ (\mathbf{n}^{(\nu)} - \mathbf{a}^{(\nu)}_{\theta}) - \nu^{-\frac{1}{2}}\ (\mathbf{a}^{(\nu)}_{\hat{\theta}^{(\nu)}} - \mathbf{a}^{(\nu)}_{\theta}).$$

Recall

$$\nu^{-\frac{1}{2}}\ (\mathbf{n}^{(\nu)} - \mathbf{a}^{(\nu)}_{\theta}) \xrightarrow{\mathcal{D}} \mathbf{w},$$

so using (6.4) and (6.10), we have

$$\nu^{-\frac{1}{2}}\ \mathbf{m}^{(\nu)}_{\hat{\theta}^{(\nu)}} \xrightarrow{\mathcal{D}} \mathbf{z} = \mathbf{w} - \left[\int_0^{(\cdot)} h^{\mathrm{T}} \mathrm{d}x\right] \left[\int_0^{T} h^{\otimes 2} \mathrm{d}x\right]^{-1} \left[\int_0^{T} h \,\mathrm{d}\mathbf{w}\right]. \tag{6.12}$$

Notice that $\int_0^T h\,\mathrm{d}\mathbf{w}$ is independent of the previous process, everything is Gaussian, and we have

$$\operatorname{cov}\left[\mathbf{z}_t, \int_0^T h\,\mathrm{d}\mathbf{w}\right] = \int_0^t h\,\mathrm{d}x - \left[\int_0^T h^{\otimes 2}\mathrm{d}x\right]\left[\int_0^T h^{\otimes 2}\mathrm{d}x\right]^{-1} \int_0^t h\,\mathrm{d}x = 0$$

Thus

$$\mathbf{w} = \mathbf{z} + \left[\int_0^{(\cdot)} h^{\mathrm{T}}\mathrm{d}x\right]\left[\int_0^T h^{\otimes 2}\mathrm{d}x\right]^{-1}\left[\int_0^{(\cdot)} h\,\mathrm{d}\mathbf{w}\right]$$

is a decomposition of $\mathbf{w}$ as the sum of two independent zero mean Gaussian processes. By Anderson's lemma (ANDERSON, 1955), we therefore obtain:

$$\mathrm{P}(\|\mathbf{w}\| \geqslant a) \geqslant \mathrm{P}\,(\|\mathbf{z}\| \geqslant a);$$

i.e. the Kolmogorov-Smirnov test, based on $\mathbf{m}_{\hat{\theta}}$, but ignoring the estimation of θ, is conservative.
This leads to two usual approaches:

(a) Just use the Kolmogorov-Smirnov test (i.e. be conservative).
(b) When testing at the 5%-level, use the 20% critical value as an ad hoc correction to conservatism (an idea of A.O. ALLEN, 1978).

6.1. THE MUSA-MODEL

We want to check whether the model of Musa provides a good fit to the failure data of Moek (Table 3.1). In the model of Musa the intensity equals $\lambda^{(\nu)}(t;\theta) = \phi(\nu\beta - \mathbf{n}_{t-})$. We put $N = \nu\beta$, let $\theta = \begin{bmatrix}\beta\\ \phi\end{bmatrix}$ the parameter-vector. The solution of (6.1) in this case, turns out to be

$$x_t = \beta_0\,(1 - e^{-\phi_0 t}). \tag{6.13}$$

Furthermore,

$$h_t = \begin{bmatrix} \dfrac{1}{N - \mathbf{n}_{t-}} \\[2ex] \dfrac{1}{\phi} \end{bmatrix} \tag{6.14}$$

We already proved that $\hat{\theta}^{(\nu)}$ is a consistent maximum likelihood estimator of θ (Theorem 3.2). We also proved (6.4) in the case that we deal with the Musa-model. In this case we are able to prove (6.10) because now

$$\frac{\partial^2}{\partial\theta^2}\,\mathbf{a}^{(\nu)}_{t;\hat{\theta}^{(\nu)}} = \begin{bmatrix} 0 & t \\ t & 0 \end{bmatrix}.$$

and we know $0 \leqslant t \leqslant \tau < \infty$. Together with the proved consistency of $\hat{\theta}^{(\nu)}$, the last term in (6.11) converges in probability to zero. Now using (6.11) together with $\frac{\partial}{\partial\theta}\frac{\mathbf{a}_\theta^{(\nu)}}{\nu} \xrightarrow{P} \int_0^{(\cdot)} h\mathrm{d}x$, this completes the proof of (6.10) in the Musa-case.

We now perform the Kolmogorov-Smirnov goodness-of-fit test to check the adequacy of the fitted model, by using approach (b). For $\mathbf{n}_\tau = 43$, the n failure times $0 \leqslant \mathbf{T}_1 \leqslant \mathbf{T}_2 \leqslant \cdots \leqslant \mathbf{T}_{43} = \tau$ are random variables. In fact τ should be fixed (for example 0.6 Msec), but we don't know what happened after detection of the 43rd bug. We put $\tau = \mathbf{T}_{43}$. Let

$$F_n(\mathbf{T}_i) = \frac{\mathbf{n}(\mathbf{T}_i)}{\mathbf{n}(\tau)} = \frac{i}{n}, \text{ and } F(\mathbf{T}_i) = \frac{1}{n}\sum_{j=1}^{i} \hat{\phi}\,(\hat{\mathbf{N}} - j + 1)\,\mathbf{t}_j,\ i = 1, \ldots, n.$$

Then we calculate the Kolmogorov-Smirnov test statistic D by

$$D_i = \max\,\{|\,F(\mathbf{T}_i) - \frac{i}{n-1}|,\ |F(\mathbf{T}_i) - \frac{i-1}{n-1}|\},\quad D = \max_i D_i\,.$$

Using the data of Moek (Table 3.1.), we found $D = 0.075$ and when we compare this with the reference value $D_{42;.20} = 0.160$ (remember we used approach (b)), we accept the null-hypothesis. So, by using this approach, the model of Musa provides a very good fit to the data we used. The same also turns out to be true for the model of Goel and Okumoto.

We can alternatively make use of two other approaches leading to (asymptotically) exact tests, avoiding both conservatism (a), and ad hoc approximation (b):

(c) Compute the distribution of $\|\mathbf{z}\|$ by simulation or by numerical intergration, cf. KARDAUN, 1986.

(d) Transform $\mathbf{z}$, and correspondingly $\mathbf{z}^{(\nu)}$ into another process for which the distribution of its $\|\cdot\|$ is easy to find or well known. We are going to use approach (d), following an idea of E. KHMALADZE (1981) in the context of empirical processes. Consider the asymptotic situation, and let $\mathbf{z}^* = \mathbf{z} - \tilde{\mathbf{z}}$, where $\tilde{\mathbf{z}}$ is the compensator of $\mathbf{z}$ with respect to its natural filtration. In fact, we consider

$$\mathcal{G}_t = \sigma\,\{\mathbf{w}_s,\ s \leqslant t;\ \int_0^\tau h\mathrm{d}\mathbf{w}\}.$$

$\mathbf{z}^*$ will be a Gaussian martingale, with easy properties. KHMALADZE (1981) shows that the transformation from $\mathbf{z}$ to $\mathbf{z}^*$ loses no statistical information, in a certain sense. Clearly $\sigma\{\mathbf{z}_s : s \leqslant t\} \subseteq \mathcal{G}_t$ and equality can also be shown, modulo completion by null sets.

For $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2) \sim \mathcal{N}(\mu, \Sigma)$:

$$\mathrm{E}(\mathbf{x}_1|\mathbf{x}_2) = \mu_1 + \Sigma_{12}\,\Sigma_{22}^{-1}\,(\mathbf{x}_2 - \mu_2), \text{ with } \mathrm{E}(\mathbf{x}_i) = \mu_i \text{ and } \mathrm{cov}\,(\mathbf{x}_i, \mathbf{x}_j) = \Sigma_{ij}\,.$$

Because both $\mathrm{d}\mathbf{w}_t$ and $\int_t^\tau h_s \mathrm{d}\mathbf{w}_s$ are Gaussian, heuristically,

$$\mathrm{E}(\mathrm{d}\mathbf{w}_t|\mathcal{G}_t) = \mathrm{E}(\mathrm{d}\mathbf{w}_t|\int_t^\tau h\mathrm{d}\mathbf{w}) = h_t^{\mathrm{T}}\mathrm{d}x_t\,(\int_t^\tau h^{\otimes 2}\mathrm{d}x)^{-1}\,(\int_t^\tau h\mathrm{d}\mathbf{w})$$

and hence we expect

$$\begin{aligned}\tilde{\mathbf{z}}_t = \int_0^t &\left\{h_s^{\mathrm{T}}\mathrm{d}x_s\,(\int_s^\tau h^{\otimes 2}\mathrm{d}x)^{-1}\,(\int_s^\tau h\mathrm{d}\mathbf{w})\right\} \\ &- \left[\int_0^t h_s^{\mathrm{T}}\mathrm{d}x_s\right]\left\{\left[\int_0^\tau h^{\otimes 2}\mathrm{d}x\right]^{-1}\left[\int_0^\tau h\mathrm{d}\mathbf{w}\right]\right\} \qquad (6.15) \\ \mathbf{z}_t^* = \mathbf{z}_t - \tilde{\mathbf{z}}_t = \mathbf{w}_t - &\int_{s=0}^t \left\{h_s^{\mathrm{T}}\mathrm{d}x_s\left[\int_{u=s}^\tau h_u^{\otimes 2}\mathrm{d}x_u\right]^{-1}\left[\int_{u=s}^\tau h_u\mathrm{d}\mathbf{w}_u\right]\right\}\end{aligned}$$

$$= \mathbf{w}_t - \tilde{\mathbf{w}}_t = \mathbf{w}^*_t ,$$

where $\tilde{\mathbf{w}}$ is the compensator of $\mathbf{w}$ with respect to $\{\mathcal{F}_t\}$.

We can write $\mathbf{z}^*$ by reversing the order of integration as

$$\mathbf{z}^*_t = \mathbf{w}_t - \int_{u=0}^{\tau} \left\{ \int_{s=0}^{u \wedge t} h_s^{\mathrm{T}} \mathrm{d}x_s \left[\int_s^{\tau} h^{\otimes 2} \mathrm{d}x \right]^{-1} \right\} h_u \mathrm{d}\mathbf{w}_u .$$

From this we can compute $\mathrm{cov}(\mathbf{z}^*_s, \mathbf{z}^*_t) =$

$$x_{s \wedge t} - \int_{u=0}^{s} \left\{ \int_{v=0}^{u \wedge t} h_v^{\mathrm{T}} \mathrm{d}x_v \left[\int_v^{\tau} h^{\otimes 2} \mathrm{d}x \right]^{-1} \right\} h_u \mathrm{d}x_u$$

$$- \int_{u=0}^{t} \left\{ \int_{v=0}^{u \wedge s} h_v^{\mathrm{T}} \mathrm{d}x_v \left[\int_v^{\tau} h^{\otimes 2} \mathrm{d}x \right]^{-1} \right\} h_u \mathrm{d}x_u$$

$$+ \int_{u=0}^{\tau} \left[\left\{ \int_{v=0}^{u \wedge t} h_v^{\mathrm{T}} \mathrm{d}x_v \left[\int_v^{\tau} h^{\otimes 2} \mathrm{d}x \right]^{-1} \right\} h_u^{\otimes 2} \mathrm{d}x_u \left\{ \int_{v=0}^{u \wedge s} \left[\int_v^{\tau} h^{\otimes 2} \mathrm{d}x \right]^{-1} h_v \mathrm{d}x_v \right\} \right]$$

$$= x_{s \wedge t} - \int_{v=0}^{s \wedge t} \int_{u=v}^{s} h_v^{\mathrm{T}} \mathrm{d}x_v \left[\int_v^{\tau} h^{\otimes 2} \mathrm{d}x \right]^{-1} h_u \mathrm{d}x_u$$

$$- \int_{v=0}^{s \wedge t} \int_{u=v}^{t} h_v^{\mathrm{T}} \mathrm{d}x_v \left[\int_v^{\tau} h^{\otimes 2} \mathrm{d}x \right]^{-1} h_u \mathrm{d}x_u$$

$$+ \int_{v=0}^{t} \int_{v'=0}^{s} \int_{u=v \vee v'}^{\tau} h_v^{\mathrm{T}} \mathrm{d}x_v \left[\int_v^{\tau} h^{\otimes 2} \mathrm{d}x \right]^{-1} h_u^{\otimes 2} \mathrm{d}x_u \left[\int_{v'}^{\tau} h^{\otimes 2} \mathrm{d}x \right]^{-1} h_{v'} \mathrm{d}x_{v'} . \quad (6.16)$$

We split the final term into two terms corresponding to $v < v'$ and $v > v'$. The first of these is

$$\int_{v=0}^{t} \int_{v'=0}^{s} h_v^{\mathrm{T}} \mathrm{d}x_v \left[\int_v^{\tau} h^{\otimes 2} \mathrm{d}x \right]^{-1} h_{v'} \mathrm{d}x_{v'}$$

$$= \int_{v=0}^{t \wedge s} \int_{u=v}^{s} h_v^{\mathrm{T}} \mathrm{d}x_v \left[\int_v^{\tau} h^{\otimes 2} \mathrm{d}x \right]^{-1} h_u \mathrm{d}x_u \quad (6.17a)$$

and similarly the second is

$$\int_{v=0}^{s \wedge t} \int_{u=v}^{t} h_v^{\mathrm{T}} \mathrm{d}x_v \left[\int_v^{\tau} h^{\otimes 2} \mathrm{d}x \right]^{-1} h_u \mathrm{d}x_u . \quad (6.17b)$$

Substituting (6.17a) and (6.17b) in (6.16) gives

$$\mathrm{cov}(\mathbf{z}^*_s, \mathbf{z}^*_t) = x_{s \wedge t} = \mathrm{cov}(\mathbf{w}_s, \mathbf{w}_t) \text{ or } \langle \mathbf{z}^* \rangle = \langle \mathbf{w} \rangle .$$

This shows that $\mathbf{z}^*$ is indeed a martingale. So if we can construct a transform of $\nu^{-\frac{1}{2}} \mathbf{m}^{(\nu)}_{\hat{\theta}^{(\nu)}}$ which is asymptotically distributed as $\mathbf{z}^*$, this is in fact the asymptotic distribution of $\nu^{-\frac{1}{2}} \mathbf{m}_{\theta}$; thus after transformation we may entirely forget the estimation of θ.

The most obvious transformation seems to be the following. We simply substitute $\hat{\mathbf{h}}^{(\nu)}$, $\mathbf{x}^{(\nu)}$ and $\nu^{-\frac{1}{2}}\,\mathbf{m}^{(\nu)}{}_{\hat{\theta}^{(\nu)}}$ for h, x, and $\mathbf{w}$ in (6.15). Here $\hat{\mathbf{h}}_t^{(\nu)} = \dfrac{\partial}{\partial\theta}\log(\hat{\phi}(\hat{\mathbf{N}}-\mathbf{n}_{t-}))$. The substitution of the first two estimates seems harmless, the last one is more subtle. We should really want to substitute $\nu^{-\frac{1}{2}}\,\mathbf{m}_\theta^{(\nu)}$ for $\mathbf{w}$, because $\nu^{-\frac{1}{2}}\,\mathbf{m}_\theta^{(\nu)} \to \mathbf{w}$. The difference, however, between $\nu^{-\frac{1}{2}}\,\mathbf{m}^{(\nu)}{}_{\hat{\theta}^{(\nu)}}$ and $\nu^{-\frac{1}{2}}\,\mathbf{m}_\theta^{(\nu)}$ is asymptotically

$$\left[\int_0^{(\cdot)} h^{\mathrm{T}}\mathrm{d}x\right]\left[\int_0^{\tau} h^{\otimes 2}\mathrm{d}x\right]^{-1}\left[\int_0^{\tau} h\,\mathrm{d}\mathbf{w}\right]. \tag{6.18}$$

Replacing $\mathbf{w}$ in (6.15) by (6.18) gives zero. In other words, the mapping $\mathbf{w} \to \mathbf{w} - \tilde{\mathbf{w}}$ is a *projection.* Notice also that $\int_0^{\tau}\hat{\mathbf{h}}^{(\nu)}\mathrm{d}\mathbf{m}^{(\nu)}{}_{\hat{\theta}^{(\nu)}} \equiv 0$; this is precisely the likelihood equation.

So, we conclude that we can probably use the goodness of fit test statistic

$$\nu^{-\frac{1}{2}}\left\|\left|\hat{\mathbf{m}} - \int_0^{t}\left\{\hat{\mathbf{h}}_s^{\mathrm{T}}\mathrm{d}\hat{\mathbf{x}}_s\left[\int_s^{\tau}\hat{\mathbf{h}}^{\otimes 2}\mathrm{d}\hat{\mathbf{x}}\right]^{-1}\int_s^{\tau}\hat{\mathbf{h}}\mathrm{d}\hat{\mathbf{m}}\right\}\right|\right\| / \sqrt{\hat{\mathbf{x}}_\tau} \tag{6.19}$$

which is expected to be asymptotically distributed as $\sup_{t\in[0,1]}|\mathbf{B}_t|$ where $\mathbf{B}$ is a standard Brownian motion. Notice (6.19) only exists if $t<\tau$ (KHMALADZE, 1981, page 251). In (6.19)

$$\hat{\mathbf{x}} = \mathbf{x}^{(\nu)} = \mathbf{n}^{(\nu)}/\nu, \quad \hat{\mathbf{h}} = h(\hat{\mathbf{x}};\,\hat{\boldsymbol{\theta}}^{(\nu)}), \quad \hat{\mathbf{m}} = \mathbf{n}^{(\nu)} - \mathbf{a}^{(\nu)}{}_{\hat{\theta}^{(\nu)}}$$

and taking e.g. $\nu = \mathbf{n}_\tau^{(\nu)}$ (that doesn't make any difference) the test-statistic becomes:

$$\mathbf{n}_\tau^{-\frac{1}{2}}\;\|\hat{\mathbf{m}} - \int_0^{(\cdot)}(\hat{\mathbf{h}}_s^{\mathrm{T}}\mathrm{d}\mathbf{n}_s\,(\int_s^{\tau}\hat{\mathbf{h}}^{\otimes 2}\mathrm{d}\mathbf{n})^{-1}\int_s^{\tau}\hat{\mathbf{h}}\mathrm{d}\hat{\mathbf{m}})\| \tag{6.20}$$

with

$$\mathbf{h} = h\,(\frac{\mathbf{n}}{\mathbf{n}_\tau};\hat{\boldsymbol{\theta}}).$$

In the model of Musa, with $\boldsymbol{\lambda}_t = \phi(N-\mathbf{n}_{t-})$, $\theta = \begin{bmatrix} N \\ \phi \end{bmatrix}$ the parameter vector, we have

$$\hat{\mathbf{h}} = \begin{bmatrix} \hat{\mathbf{h}}_{1:t} \\ \hat{\mathbf{h}}_{2:t} \end{bmatrix} = \begin{bmatrix} \dfrac{\mathbf{n}_\tau}{\hat{\mathbf{N}}-\mathbf{n}_{t-}} \\ \dfrac{1}{\hat{\phi}} \end{bmatrix}. \tag{6.21}$$

It can be seen in (6.20) that the statistic is invariant under a rescaling of the components of $\hat{\mathbf{h}}$; i.e. premultiplying by a diagonal matrix. Thus the $\mathbf{n}_\tau$ in the numerator of $\hat{\mathbf{h}}_{1:t}$ can be replaced by 1.

In general, therefore, we may use the statistic

$$\mathbf{n}_\tau^{-\frac{1}{2}}\;\|\hat{\mathbf{m}}^*\| = \mathbf{n}_\tau^{-\frac{1}{2}}\;\|\hat{\mathbf{m}} - \int_0^{(\cdot)}(\hat{\mathbf{h}}_s^{\mathrm{T}}\mathrm{d}\mathbf{n}_s\,(\int_s^{\tau}\hat{\mathbf{h}}^{\otimes 2}\mathrm{d}\mathbf{n})^{-1}\int_s^{\tau}\hat{\mathbf{h}}\mathrm{d}\hat{\mathbf{m}})\|$$

where

$$\hat{\mathbf{m}} = \mathbf{n} - \int_0^{(\cdot)}\boldsymbol{\lambda}\,(t;\hat{\phi})\mathrm{d}t.$$

In our case the test-statistic can be rewriten as:

$$\frac{1}{\sqrt{\mathbf{n}_T}} \sup_{1 \leqslant n \leqslant \mathbf{n}_T - 1} \left| \left[n - \sum_{i=1}^{n} \hat{\phi}\, (\hat{\mathbf{N}} - i + 1)\mathbf{t}_i \right] - \sum_{s=1}^{n} \left[\frac{1}{\hat{\mathbf{N}} - s + 1} \cdot \frac{1}{\det} \cdot \right.\right.$$

$$\left\{ \left[\frac{\mathbf{n}_T - s + 1}{\hat{\phi}^2} \right] \cdot \left[-\sum_{i=s}^{\mathbf{n}_T} \hat{\phi} \mathbf{t}_i + \sum_{i=s}^{\mathbf{n}_T} \frac{1}{\hat{\mathbf{N}} - i + 1} \right] \right.$$

$$\left. - \left[\sum_{i=s}^{\mathbf{n}_T} \frac{1}{\hat{\phi}(\hat{\mathbf{N}} - i + 1)} \right] \cdot \left[-\sum_{i=s}^{\mathbf{n}_T} (\hat{\mathbf{N}} - i + 1)\mathbf{t}_i + \frac{\mathbf{n}_T - s + 1}{\hat{\phi}} \right] \right\}$$

$$+ \frac{1}{\hat{\phi}} \cdot \frac{1}{\det} \cdot \left\{ \left[\sum_{i=s}^{\mathbf{n}_T} \frac{1}{\hat{\phi}(\hat{\mathbf{N}} - i + 1)} \right] \cdot \left[-\sum_{i=s}^{\mathbf{n}_T} \hat{\phi} \mathbf{t}_i + \sum_{i=s}^{\mathbf{n}_T} \frac{1}{\hat{\mathbf{N}} - i + 1} \right] + \right.$$

$$\left.\left.\left. + \left[\sum_{i=s}^{\mathbf{n}_T} \frac{1}{(\hat{\mathbf{N}} - i + 1)^2} \right] \cdot \left[-\sum_{i=s}^{\mathbf{n}_T} (\hat{\mathbf{N}} - i + 1)\mathbf{t}_i + \frac{\mathbf{n}_T - s + 1}{\hat{\phi}} \right] \right\} \right] \right| \tag{6.22}$$

where det (determinant) equals

$$\frac{\mathbf{n}_T - s + 1}{\hat{\phi}^2} \sum_{i=s}^{\mathbf{n}_T} \frac{1}{(\hat{\mathbf{N}} - i + 1)^2} - \left[\sum_{i=s}^{\mathbf{n}_T} \frac{1}{\hat{\phi}(\hat{\mathbf{N}} - i + 1)} \right]^2 . \tag{6.23}$$

Notice that (6.23) equals zero when $s = \mathbf{n}_T$, this could cause problems, but we avoided this by computing the supremum over values of n for which $1 \leqslant n \leqslant \mathbf{n}_T - 1$.

The critical value, at the 95 percent level, equals 2.2454. The value of the last test-statistic we found with $n = 43$ equals 0.3948, which is far below the critical value. So we have to conclude that the Musa-model fits adequately.

APPENDIX A: PROOF OF THEOREM 4.3

Look at the following Taylor-expansions

$$0 = \frac{1}{\sqrt{\nu}} \frac{\partial}{\partial \theta} \log \mathbf{L}\, (\hat{\theta}^{(\nu)})_i = \frac{1}{\sqrt{\nu}} \frac{\partial}{\partial \theta} \log \mathbf{L}\, (\theta_0)_i + \sqrt{\nu}(\hat{\theta}^{(\nu)} - \theta_0)^{\mathrm{T}} \cdot \frac{1}{\nu} \frac{\partial^2}{\partial \theta \partial \theta_i} \log \mathbf{L}(\overline{\theta}^{(i)}),$$

where

$$|(\hat{\theta}^{(i)} - \theta_0)_j| \leqslant |(\hat{\theta}^{(\nu)} - \theta_0)_j| \quad i = 1,2;\ j = 1,2. \tag{A.1}$$

It is sufficient to show that

$$\frac{1}{\sqrt{\nu}} \frac{\partial}{\partial \theta} \log \mathbf{L}\, (\theta_0) \xrightarrow{\mathcal{D}} \mathfrak{N}(0, \Sigma_{GO}\, (\theta_0)\ \ (\nu \to \infty) \tag{A.2}$$

and that

$$\frac{1}{\nu} \frac{\partial}{\partial \theta^2} \log \mathbf{L}\, (\overline{\theta}) \xrightarrow{\mathrm{P}} -\Sigma_{GO}\, (\theta_0)\ \ (\nu \to \infty). \tag{A.3}$$

Let us first look at (A.2). From (4.3) we get ($N = \nu\beta_0$)

$$\log \mathbf{L}\, (\theta_0) = \mathbf{n}_T^{(\nu)} \log \nu\beta_0\phi_0 - \phi_0 \sum_{i=1}^{\mathbf{n}_T^{(\nu)}} \mathbf{T}_i - \nu\beta_0\, (1 - e^{-\phi_0 T}). \tag{A.4}$$

So

$$\frac{\partial}{\partial \beta} \log \mathbf{L}\, (\theta_0) = \frac{\mathbf{n}_T^{(\nu)}}{\beta_0} - \nu(1 - e^{-\phi_0 T}) = \int_0^T \frac{1}{\beta_0}\, \mathrm{d}\mathbf{m}_t^{(\nu)} \tag{A.5}$$

$$\frac{\partial}{\partial\phi}\log \mathbf{L}(\theta_0)=\frac{\mathbf{n}_\tau^{(\nu)}}{\phi_0}-\sum_{i=1}^{\mathbf{n}_\tau^{(\nu)}}\mathbf{T}_i-\nu\beta_0\,\tau e^{-\phi_0\tau}=\int_0^\tau\left(\frac{1}{\phi_0}-t\right)\mathrm{d}\mathbf{m}_t^{(\nu)}. \tag{A.6}$$

where

$$\mathbf{m}_t^{(\nu)}=\mathbf{n}_t^{(\nu)}-\int_0^t \boldsymbol{\lambda}_s^{(\nu)}\,\mathrm{d}s.$$

Notice that $\frac{1}{\beta_0}$ equals $\frac{\partial}{\partial\beta}\log\boldsymbol{\lambda}_t$ and $(\frac{1}{\phi_0}-t)$ equals $\frac{\partial}{\partial\phi}\log\boldsymbol{\lambda}_t$, since $\boldsymbol{\lambda}_t=\nu\beta_0\phi_0\exp[-\phi_0 t]$.

$$<\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0)>(t)=\frac{1}{\nu}\begin{bmatrix}\int_0^t \frac{1}{\beta_0^2}\,\mathrm{d}<\mathbf{m}^{(\nu)}>(s) & \int_0^t \frac{1}{\beta_0}\left(\frac{1}{\phi_0}-s\right)\mathrm{d}<\mathbf{m}^{(\nu)}>(s)\\ \int_0^t \frac{1}{\beta_0}\left(\frac{1}{\beta_0}-s\right)\mathrm{d}<\mathbf{m}^{(\nu)}>(s) & \int_0^t\left(\frac{1}{\phi_0}-s\right)^2\mathrm{d}<\mathbf{m}^{(\nu)}>(s)\end{bmatrix} \tag{A.7}$$

Using $\mathrm{d}<\mathbf{m}^{(\nu)}>(t)=\boldsymbol{\lambda}_t^{(\nu)}\mathrm{d}t=\nu\beta_0\phi_0 e^{-\phi_0 t}$, (A.7) becomes

$$<\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0)>(t)=\begin{bmatrix}\int_0^t \frac{\phi_0}{\beta_0}e^{-\phi_0 s}\mathrm{d}s & \int_0^t \phi_0\left(\frac{1}{\phi_0}-s\right)e^{-\phi_0 s}\mathrm{d}s\\ \int_0^t \theta_0\left(\frac{1}{\phi_0}-s\right)e^{-\phi_0 s}\mathrm{d}s & \int_0^t \phi_0\,\beta_0\left(\frac{1}{\phi_0}-s\right)^2 e^{-\phi_0 s}\mathrm{d}s\end{bmatrix}$$

$$=\begin{bmatrix}\frac{1}{\beta_0}(1-e^{-\phi_0 t}) & te^{-\phi_0 t}\\ te^{-\phi_0 t} & \frac{\beta_0}{\phi_0^2}-\frac{\beta_0}{\phi_0^2}e^{-\phi_0 t}-\beta_0 t^2 e^{-\phi_0 t}\end{bmatrix} \tag{A.8}$$

$$\stackrel{\text{def}}{=}\Sigma_{GO}(\theta_0)\ \text{when}\ t=\tau\,.$$

To apply a martingale central limit theorem, we further need $\forall\epsilon>0$

$$\int_0^\tau \frac{1}{\nu}\frac{1}{\beta_0^2}\boldsymbol{\lambda}_t\,I\left\{\left|\frac{1}{\sqrt{\nu}}\cdot\frac{1}{\beta_0}\right|>\epsilon\right\}\mathrm{d}t \xrightarrow{\mathrm{P}} 0\quad(\nu\to\infty) \tag{A.9}$$

and

$$\int_0^\tau \frac{1}{\nu}\left(\frac{1}{\phi_0}-t\right)^2\boldsymbol{\lambda}_t\,I\left\{\left|\frac{1}{\sqrt{\nu}}\left(\frac{1}{\phi_0}-t\right)\right|>\epsilon\right\}\mathrm{d}t \xrightarrow{\mathrm{P}} 0\quad(\nu\to\infty). \tag{A.10}$$

In (A.9) $\frac{1}{\beta_0^2}\cdot\frac{\boldsymbol{\lambda}_t}{\nu}\,I\{|\frac{1}{\sqrt{\nu}}\cdot\frac{1}{\beta_0}|>\epsilon\}$ equals 0 when $\sqrt{\nu}>\frac{1}{\epsilon.\beta_0}$, so indeed (A.9) holds. To show (A.10) notice that $\int_0^\tau(\frac{1}{\phi_0}-t)^2\cdot\frac{\boldsymbol{\lambda}_t^{(\nu)}}{\nu}\,\mathrm{d}t<\infty$ and is independent of ν. Furthermore,

$I\{|\frac{1}{\sqrt{\nu}}(\frac{1}{\phi_0}-t)|>\epsilon\}$ converges pointwise to zero for $t\neq\frac{1}{\phi_0}$. The use of dominated convergence now leads to (A.10). We may now apply the martingale central limit theorem mentioned in Section 1 and conclude that

$$\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{GO}(\theta_0)),$$

where $\Sigma_{GO}(\theta_0)$ is given by (A.8). Now look at (A.3), we will first prove that $\frac{1}{\nu}\frac{\partial^2}{\partial\theta^2}\log \mathbf{L}(\theta_0) \xrightarrow{P} -\Sigma_{GO}(\theta_0)$ $(\nu\to\infty)$. We have

$$\frac{1}{\nu}\frac{\partial^2}{\partial\beta^2}\log \mathbf{L}(\theta_0) = -\frac{\mathbf{n}_\tau^{(\nu)}}{\nu\beta_0^2} \xrightarrow{P} -\frac{1}{\beta_0}(1-e^{-\theta_0\tau}) \quad (\nu\to\infty), \tag{A.11}$$

$$\frac{1}{\nu}\frac{\partial^2}{\partial\phi^2}\log \mathbf{L}(\theta_0) = \frac{\mathbf{n}_\tau^{(\nu)}}{\nu\theta_0^2} + \beta_0\,\tau^2\,e^{-\theta_0\tau} \xrightarrow{P} -\frac{\beta_0}{\phi_0^2}(1-e^{-\phi_0\tau}) + \beta_0\,\tau^2\,e^{-\phi_0\tau} \quad (\nu\to\infty), \tag{A.12}$$

$$\frac{1}{\nu}\frac{\partial^2}{\partial\beta\partial\phi}\log \mathbf{L}(\theta_0) = \frac{1}{\nu}\frac{\partial^2}{\partial\phi\partial\beta}\log \mathbf{L}(\theta_0) = -\tau e^{-\phi_0\tau}. \tag{A.13}$$

The convergence in probability in (A.11) and (A.12) is a consequence of Corollary 4.1 :

$$\frac{\mathbf{n}_t^{(\nu)}}{\nu} \xrightarrow{P} \beta_0(1-e^{\phi_0\tau}) \quad (\nu\to\infty).$$

So indeed

$$\frac{1}{\nu}\frac{\partial^2}{\partial\theta^2}\log \mathbf{L}(\theta_0) \xrightarrow{P} -\Sigma_{GO}(\theta_0) \quad (\nu\to\infty). \tag{A.14}$$

Now look at $(\bar{\boldsymbol\theta}=\bar{\boldsymbol\theta}^{(i)})$:

$$\frac{1}{\nu}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log \mathbf{L}(\bar{\boldsymbol\theta}) = \frac{1}{\nu}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log \mathbf{L}(\theta_0) + \sum_{k=1}^{2}(\bar{\boldsymbol\theta}-\theta_0)_k\cdot\frac{1}{\nu}\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\log L(\boldsymbol\theta^*),$$

$$i=1,2;\ j=1,2, \text{ where } |(\boldsymbol\theta^*-\theta_0)_k| \leqslant |(\bar{\boldsymbol\theta}-\theta_0)_k| \leqslant |(\hat{\boldsymbol\theta}^{(\nu)}-\theta_0)_k|. \tag{A.15}$$

In Theorem 4.2 we have proved $\hat{\boldsymbol\theta}^{(\nu)} \xrightarrow{P} \theta_0$ $(\nu\to\infty)$. So from (A.15) we get $\boldsymbol\theta^* \xrightarrow{P} \theta_0$ $(\nu\to\infty)$ and $\bar{\boldsymbol\theta} \xrightarrow{P} \theta_0$ $(\nu\to\infty)$. Note that $\boldsymbol\theta^*$ and $\bar{\boldsymbol\theta}$ may not actually be measurable; however the usual outer probability argument takes care of this. In our case we have for $\nu\to\infty$:

$$\frac{1}{\nu}\frac{\partial^3}{\partial\beta^3}\log \mathbf{L}(\boldsymbol\theta^*) = \frac{2}{(\beta^*)^3}\cdot\frac{\mathbf{n}_\tau^{(\nu)}}{\nu} \xrightarrow{P} \frac{2}{\beta_0^2}(1-e^{-\phi_0\tau}) \quad (\nu\to\infty)$$

$$\frac{1}{\nu}\frac{\partial^3}{\partial\phi\partial\beta^2}\log \mathbf{L}(\boldsymbol\theta^*) = 0$$

$$\frac{1}{\nu}\frac{\partial^3}{\partial\phi^3}\log \mathbf{L}(\boldsymbol\theta^*) = \frac{2}{(\phi^*)^3}\cdot\frac{\mathbf{n}_t^{(\nu)}}{\nu} - \beta^*\,\tau^3\,e^{-\phi^*\tau} \xrightarrow{P} 2\frac{\beta_0}{\phi_0^3} - \beta_0\,\tau^2\,e^{-\phi_0\tau} \quad (\nu\to\infty)$$

$$\frac{1}{\nu}\frac{\partial^3}{\partial\beta\partial\phi^2}\log \mathbf{L}(\boldsymbol\theta^*) = \tau^2 e^{-\phi^*\tau} \xrightarrow{P} \tau^2 e^{-\phi_0\tau} \quad (\nu\to\infty).$$

All right hand sides are finite. So we have

$$\sum_{k=1}^{2}(\bar{\boldsymbol\theta}-\theta_0)_k\cdot\frac{1}{\nu}\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\log \mathbf{L}(\theta^*) \xrightarrow{P} 0 \quad (\nu\to\infty). \tag{A.16}$$

Hence we may conclude from (A.14) and (A.16) that

$$\frac{1}{\nu}\frac{\partial^2}{\partial\theta^2}\log \mathbf{L}(\bar{\boldsymbol{\theta}}) \xrightarrow{P} -\Sigma_{GO}(\theta_0) \quad (\nu\to\infty). \tag{A.17}$$

Using CRAMÉR (1946) or BILLINGSLEY (1961) (Theorems 2.2 and 10.1) we may conclude that

$$\sqrt{\nu}\,(\hat{\boldsymbol{\theta}}^{(\nu)} - \theta_0) \xrightarrow{\mathfrak{D}} \mathfrak{N}(0, \Sigma_{GO}(\theta_0)^{-1}).$$

□

APPENDIX B: PROOF OF THEOREM 5.2

Consider the following usual Taylor-expansions

$$0 = \frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\hat{\boldsymbol{\theta}}^{(\nu)})\,i = \frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0)_i + \sqrt{\nu}\,(\hat{\boldsymbol{\theta}}^{(\nu)} - \theta_0)^{\mathrm{T}} \cdot \frac{1}{\nu}\frac{\partial^2}{\partial\theta\cdot\partial\theta_i}\log \mathbf{L}(\bar{\boldsymbol{\theta}}^{(i)}),$$

where

$$|(\bar{\boldsymbol{\theta}}^{(i)} - \theta_0)_j| \leqslant |(\hat{\boldsymbol{\theta}}^{(\nu)} - \theta_0)_j| \quad i=1,2,3;\ j=1,2,3. \tag{B.1}$$

It is sufficient to shown that

$$\frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0) \xrightarrow{\mathfrak{D}} \mathfrak{N}(0, \Sigma_L(\theta_0)) \quad (\nu\to\infty) \tag{B.2}$$

and

$$\frac{1}{\nu}\frac{\partial^2}{\partial\theta^2}\log \mathbf{L}(\bar{\boldsymbol{\theta}}) \xrightarrow{P} -\Sigma_L(\theta_0) \quad (\nu\to\infty) \tag{B.3}$$

where $\Sigma_L(\theta_0)$ is given by (5.53). First look at (B.2)

$$\frac{\partial}{\partial\beta}\log \mathbf{L}(\theta_0) = \int_0^T \frac{1}{\beta_0 - \dfrac{\mathbf{n}_t^{(\nu)}}{\nu}}\,d\mathbf{m}_t^{(\nu)}, \tag{B.4}$$

$$\frac{\partial}{\partial a}\log \mathbf{L}(\theta_0) = \int_0^T \frac{1}{a_0}\,d\mathbf{m}_t^{(\nu)}, \tag{B.5}$$

$$\frac{\partial}{\partial b}\log \mathbf{L}(\theta_0) = \int_0^T \frac{-1}{b_0+t}\,d\mathbf{m}_t^{(\nu)}. \tag{B.6}$$

Notice again that $\dfrac{1}{\beta_0 - \dfrac{\mathbf{n}_t^{(\nu)}}{\nu}}$ equals $\dfrac{\partial}{\partial\beta}\log \boldsymbol{\lambda}_t$, $\dfrac{1}{a_0}$ equals $\dfrac{\partial}{\partial_a}\log \boldsymbol{\lambda}_t$ and $\dfrac{-1}{b_0+t}$ equals $\dfrac{\partial}{\partial b}\log \boldsymbol{\lambda}_t$.

Hence

$$< \frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0) >_t =$$

$$\frac{1}{\nu}\begin{bmatrix} \int_0^t \frac{1}{(\beta_0-\frac{\mathbf{n}_{s-}^{(\nu)}}{\nu})^2}\boldsymbol{\lambda}_s^{(\nu)}\mathrm{d}s & \int_0^t \frac{1}{(\beta_0-\frac{\mathbf{n}_{s-}^{(\nu)}}{\nu})a_0}\cdot\boldsymbol{\lambda}_s^{(\nu)}\mathrm{d}s & \int_0^t \frac{\boldsymbol{\lambda}_s^{(\nu)}}{(\beta_0-\frac{\mathbf{n}_{s-}^{(\nu)}}{\nu})(b_0+s)}\cdot\mathrm{d}s \\ * & \int_0^t \frac{\boldsymbol{\lambda}_s^{(\nu)}}{a_0^2}\mathrm{d}s & \int_0^t \frac{-\boldsymbol{\lambda}_s^{(\nu)}}{a_0(b_0+s)}\mathrm{d}s \\ * & * & \int_0^t \frac{\boldsymbol{\lambda}_s^{(\nu)}}{(b_0+s)^2}\mathrm{d}s \end{bmatrix}$$

$$= \begin{bmatrix} \int_0^t \frac{a_0}{(\beta_0-\frac{\mathbf{n}_{s-}^{(\nu)}}{\nu})(b_0+s)}\mathrm{d}s & \int_0^t \frac{1}{b_0+s}\mathrm{d}s & \int_0^t \frac{-a_0}{(b_0+s)^2}\mathrm{d}s \\ * & \int_0^t \frac{\beta_0-\frac{\mathbf{n}_{s-}^{(\nu)}}{\nu}}{a_0(b_0+s)}\mathrm{d}s & \int_0^t \frac{-(\beta_0-\frac{\mathbf{n}_{s-}^{(\nu)}}{\nu})}{(b_0+s)^2}\mathrm{d}s \\ * & * & \int_0^t \frac{a_0(\beta_0-\mathbf{n}_{s-}^{(\nu)}/\nu)}{(b_0+s)^3}\mathrm{d}s \end{bmatrix} \tag{B.7}$$

Using Theorem 5.1, this leads to

$$< \frac{1}{\sqrt{\nu}}\frac{\partial}{\partial\theta}\log \mathbf{L}(\theta_0)>_t$$

$$\xrightarrow{P} \begin{bmatrix} ((b_0+t)^{a_0}-b_0^{a_0})\frac{1}{\beta_0 b_0^{a_0}} & \log\left[\frac{b_0+t}{b_0}\right] & \frac{-a_0 t}{b_0(b_0+t)} \\ * & \frac{\beta}{a_0^2}\left[1-\left[\frac{b_0}{b_0+t}\right]^{a_0}\right] & \frac{\beta_0}{b_0(a_0+1)}\left[\left[\frac{b_0}{b_0+t}\right]^{a_0+1}-1\right] \\ * & * & \frac{a_0\beta_0}{b_0^2(a_0+2)}\left[1-\left[\frac{b_0}{b_0+t}\right]^{a_0+2}\right] \end{bmatrix} \tag{B.8}$$

$$\stackrel{\text{def}}{=}\Sigma_L(\theta_0) \text{ with } t=\tau\,.$$

Furthermore we need $\forall\epsilon>0$

$$\int_0^\tau \frac{1}{\nu}\cdot\frac{1}{(\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu})^2}\cdot\boldsymbol{\lambda}_t^{(\nu)}\, I\left\{\left|\frac{1}{\sqrt{\nu}}\,\frac{1}{(\beta_0-\frac{\mathbf{n}_{t-}^{(\nu)}}{\nu})}\right|>\epsilon\right\}\mathrm{d}t \xrightarrow{P} 0 \quad (\nu\to\infty)R\,, \tag{B.9}$$

$$\int_0^\tau \frac{1}{\nu}\cdot\frac{1}{a_0^2}\cdot\boldsymbol{\lambda}_t^{(\nu)}\cdot I\left\{\left|\frac{1}{\sqrt{\nu}}\cdot\frac{1}{a_0}\right|>\epsilon\right\}\mathrm{d}t \xrightarrow{P} 0 \quad (\nu\to\infty)\,, \tag{B.10}$$

$$\int_0^\tau \frac{1}{\nu}\cdot\frac{1}{(b_0+t)^2}\boldsymbol{\lambda}_t^{(\nu)}\cdot I\left\{\left|\frac{1}{\sqrt{\nu}}\cdot\frac{1}{(b_0+t)}\right|>\epsilon\right\}\mathrm{d}t \xrightarrow{P} 0 \quad (\nu\to\infty)\,. \tag{B.11}$$

The integrand of (B.9) equals 0 when $\sqrt{\nu}(\beta_0-\mathbf{x}_\tau^{(\nu)})\geqslant 1/\epsilon$, the integrand of (B.10) equals 0 when $\sqrt{\nu}a_0\geqslant 1/\epsilon$, and finally the integrand of (B.11) equals 0 when $\sqrt{\nu}\cdot b_0\geqslant 1/\epsilon$. We may now apply the martingale central limit theorem mentioned in Section 1. Conclusion:

$$\frac{1}{\sqrt{\nu}} \frac{\partial}{\partial \theta} \log \mathbf{L}(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_L(\theta_0)) \quad (\nu \to \infty).$$

Now look at (B.3). We will first show that

$$\frac{1}{\nu} \frac{\partial^2}{\partial \theta^2} \log \mathbf{L}(\theta_0) \xrightarrow{P} -\Sigma_L(\theta_0) \quad (\nu \to \infty) \tag{B.12}$$

$$\frac{1}{\nu} \cdot \frac{\partial^2}{\partial \beta^2} \log \mathbf{L}(\theta_0) = -\int_0^\tau \frac{1}{(\beta_0 - \frac{\mathbf{n}_{t-}^{(\nu)}}{\nu})^2} \, \mathrm{d} \frac{\mathbf{n}_t^{(\nu)}}{\nu} \xrightarrow{P} \frac{-a_0}{\beta_0 b_0^{a_0}} \int_0^\tau (b_0+t)^{a_0-1} \, \mathrm{d}t$$

$$= -\frac{1}{\beta_0 b_0^{a_0}} ((b_0+\tau)^{a_0} - b_0^{a_0}) \quad (\nu \to \infty). \tag{B.13}$$

$$\frac{1}{\nu} \frac{\partial^2}{\partial a^2} \log \mathbf{L}(\theta_0) = -\int_0^\tau \frac{1}{a_0^2} \, \mathrm{d} \frac{\mathbf{n}_t^{(\nu)}}{\nu} \xrightarrow{P} -\frac{\beta_0 b_0^{a_0}}{a_0} \int_0^\tau \frac{1}{(b_0+t)^{a_0+1}} \, \mathrm{d}t$$

$$= -\frac{\beta_0}{a_0^2} (1 - (\frac{b_0}{b_0+\tau})^{a_0}) \quad (\nu \to \infty). \tag{B.14}$$

$$\frac{1}{\nu} \frac{\partial^2}{\partial \beta^2} \log \mathbf{L}(\theta_0) = \int_0^\tau \frac{1}{(b_0+t)^2} \, \mathrm{d} \frac{\mathbf{n}_\tau^{(\nu)}}{\nu} - 2 \int_0^\tau \frac{a_0(\beta_0 - \frac{\mathbf{n}_\tau^{(\nu)}}{\nu})}{(b_0+t)^3} \, \mathrm{d}t \xrightarrow{P}$$

$$\frac{-a_0 \beta_0}{b_0^2 (a_0+2)} \left[1 - \left[\frac{b_0}{b_0+\tau} \right]^{a_0+2} \right] \quad (\nu \to \infty). \tag{B.15}$$

$$\frac{1}{\nu} \frac{\partial^2}{\partial a \partial \beta} \log \mathbf{L}(\theta_0) = \frac{1}{\nu} \frac{\partial^2}{\partial \beta \partial a} \log \mathbf{L}(\theta_0) = -\int_0^\tau \frac{1}{b_0+t} \, \mathrm{d}t = -\log \left[\frac{b_0+\tau}{b_0} \right]. \tag{B.16}$$

$$\frac{1}{\nu} \frac{\partial^2}{\partial a \partial b} \log \mathbf{L}(\theta_0) = \frac{1}{\nu} \frac{\partial^2}{\partial b \partial a} \log \mathbf{L}(\theta_0) = \int_0^\tau \frac{\beta_0 - \frac{\mathbf{n}_{t-}^{(\nu)}}{\nu}}{(b_0+t)^2} \, \mathrm{d}t \xrightarrow{P}$$

$$\frac{-\beta_0}{b_0(a_0+1)} \left[\left[\frac{b_0}{b_0+\tau} \right]^{a_0+1} - 1 \right]. \tag{B.17}$$

$$\frac{1}{\nu} \frac{\partial^2}{\partial b \partial \beta} \log \mathbf{L}(\theta_0) = \frac{1}{\nu} \frac{\partial^2}{\partial \beta \partial b} \log \mathbf{L}(\theta_0) = \int_0^\tau \frac{a_0}{(\beta_0+t)^2} \, \mathrm{d}t = \frac{a_0 \tau}{b_0(b_0+\tau)}. \tag{B.18}$$

All convergence in probability in (B.13)—(B.18) follows from Theorem 5.1. So we have indeed

$$\frac{1}{\nu} \frac{\partial^2}{\partial \theta^2} \log \mathbf{L}(\theta_0) = -\Sigma_L(\theta_0).$$

Now look at ($\bar{\boldsymbol{\theta}} = \bar{\boldsymbol{\theta}}^{(i)}$)

$$\frac{1}{\nu} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \mathbf{L}(\bar{\boldsymbol{\theta}}) = \frac{1}{\nu} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \mathbf{L}(\theta_0) + \sum_{k=1}^{3} (\bar{\boldsymbol{\theta}} - \theta_0)_k \cdot \frac{1}{\nu} \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \log \mathbf{L}(\boldsymbol{\theta}^*),$$

where

$$|(\theta^* - \theta_0)_k| \leq |(\bar{\theta} - \theta_0)_k| \leq |(\hat{\theta}^{(\nu)} - \theta_0)_k| \quad i,j,k = 1,2,3. \tag{B.19}$$

Again outer probability arguments take for the fact that θ^* may not be measurable. We assumed $\hat{\theta} \xrightarrow{P} \theta_0$ $(\nu \to \infty)$, so $\theta^* \xrightarrow{P} \theta_0$ $(\nu \to \infty)$ from (B.19). Therefore $\frac{\partial^3}{\partial\theta_i \partial\theta_j \partial\theta_k} \log \mathbf{L}(\theta^*)$ is finite for $\nu \to \infty$ for all i,j,k:

$$\frac{1}{\nu} \frac{\partial^2}{\partial\theta^2} \log(\hat{\theta}) \xrightarrow{P} -\Sigma_L(\theta_0).$$

We have proved (B.2) and (B.3), so according to CRAMÉR (1946) or BILLINGSLEY (1961) (Theorems 2.2 and 10.1)

$$\sqrt{\nu}\,(\hat{\theta}^{(\nu)} - \theta_0) \xrightarrow{\mathcal{D}} \mathfrak{N}(0, \Sigma_L(\theta_0)^{-1}),$$

where $\Sigma_L(\theta_0)$ is given by (5.53) or (B.8). □

REFERENCES

[1] O.O. AALEN (1978). Non-parametric inference for a family of counting processes. *Ann. Statist.* **6**, 701-726.
[2] A.O. ALLEN (1978). *Probability, Statistics and Queueing Theory*, Computer Science and Applied Mathematics, Academic Press, New York
[3] P.K. ANDERSEN & Ø. BORGAN (1985). Counting process models for life history data: a review. *Scand. J. Statist.* **12**, 97-158.
[4] T.W. ANDERSON (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequelities, *Proc. Amer. Math. Soc.* **6**, 170-176.
[5] P.K. ANDERSEN & R.D. GILL (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 110-1120.
[6] P. BILLINGSLEY (1961). *Statistical Inference for Markov-processes.* The University of Chicago Press, Chicago.
[7] P. BILLINGSLEY (1968). *Convergence of probability measures.* Wiley, New York.
[8] Ø. BORGAN (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scand. J. Statist.* **1**, 1-16. correction *ibid.* **11**, 275.
[9] P. BREMAUD (1977). Processes ponctuals et martingales: Résultats récents sur la modélisation et filtrage. *Adv. Appl. Prob.* **9**, 362-416.
[10] H. CRAMÉR (1946). *Mathematical methods of Statistics.* Princeton University Press, Princeton.
[11] K.O. DZHAPARIDZE (1983). On iterative procedures of asymptotic inference. *Statistical Neerlandica* **37**, 181-189.
[12] R.D. GILL (1984). Understanding Cox's regression model: a martingale approach. *J. Am. Statist. Ass.* **79**, 441-447.
[13] R.D. GILL (1986). The Total Time on Test Plot and the Cumulative Total Time on Test Statistic for a counting process. *Ann. Statist.* **14**, 1234-1239.
[14] A.L. GOEL & K. OKUMOTO (1980). *A time dependant error-detection rate model for software performance assesment with applications.* Rome Air Development Center, Griffiths Air Force Base, New York.
[15] I.S. HELLAND (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* **9**, 79-94.
[16] I.S. HELLAND (1983). Applications of central limit theorems for martingales with continuous time. *Bull. Inst. Int. Statist.* **50** (1), 346-360.

[17] Z. JELINSKI & P. MORANDA (1972). Software Reliability Research. *Statistical Computer Performance Evaluation*, 466-484, Academic Press, New York.

[18] J.D. KALBFLEISCH & R.L. PRENTICE (1980). *The statistical analysis of failure time data.* Wiley, New York.

[19] E.V. KHMALADZE (1981). Martingale approach in the theory of goodness of fit tests. *Th. Prob. Ap.* **16**, 240-257.

[20] E.V. KHMALADZE (1982). Some applications of the theory of martingales to statistics. *Russ. Math. Surv.* **37** (6), 215-237.

[21] E.V. KHMALADZE (1987). An innovation approach to goodness of fit tests in $\mathbb{R}^m$

[22] T.G. KURTZ (1981). *Approximation of population processes.* Regional Conference Series in Applied Mathematics **36**, SIAM Philadelphia, Pennsylvania.

[23] T.G. KURTZ (1983). Gaussian approximations for Markov chains and counting processes. *Bull. Inst. Int. Statist.* **50** (1), 361-375.

[24] B. LITTLEWOOD (1980). Theories of Software Reliability: How good are they and how can they be improved? *IEEE Trans. Software Eng.* **6**, 489-500.

[25] G. MOEK (1983). *Software Reliability models on trial: selection, improved estimation and practical results*, NLR Report MP 83059 U, Amsterdam.

[26] G. MOEK (1984). Comparison of some software reliability models for simulated and real failure data, *Int. J. of Modeling and Simulation*, **4**, 29-41.

[27] G. MOEK (1985). *Some theoretical and practical results on Littlewood's Baysian Differential Debugging model for software reliability*, NLR Report MP 85017 U, Amsterdam.

[28] J.D. MUSA (1975). A theory of software reliability and its applications. *Trans. Software Eng.* **1**, 312-327.

[29] C.V. RAMAMOORTHY & F.B. BASTANI (1982). Software Reliability—Status and Perspectives. *IEEE Trans. Software Eng.* **8**, 354-371.

[30] F.W. SCHOLZ (1986). Software Reliability Modeling and Analysis. *IEEE Trans. Software Eng.* **12**, 25-31.

[31] F.W. SCHOLZ (1987a). *Software Reliability Confidence Bounds.* Unpublished manuscript.

[32] F.W. SCHOLZ (1987b). *Asymptotics for the Jelinski Moranda model.* Unpublished manuscript.

[33] D.L. SNIJDER (1975). *Random Point Processes.* Wiley, New York.

[34] P.J.C. SPREY & G. KOCH (1983). A martingale approach to software reliability, *IEEE Trans. Reliability* **32**, 342-345.

[35] P.J.C. SPREY (1985). Parametric estimation for a specific software reliability model, *IEEE Trans. Reliability* **34**, 223-228.