Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

M.C.J. van Pul

# Asymptotic properties of statistical models in software reliability

Department of Operations Research, Statistics, and System Theory Report BS-R9011 May

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 Research (N.W.O.).

# Asymptotic Properties of Statistical Models in Software Reliability

Mark C. van Pul
*Centre for Mathematics and Computer Science*
*P.O. Box 4079, 1009 AB Amsterdam*
*The Netherlands*

In software reliability theory many different models have been proposed and investigated. A number of important models are completely characterized by their intensity function. In this paper a rather general class of intensity functions is considered. Sufficient conditions are given under which some important asymptotic properties of the model and of the maximum likelihood estimators for the model parameters can be proved. An application to software reliability theory is presented.

*1980 Mathematics Subject Classification:* 62N05, 62M99.
*Keywords and Phrases:* counting process, intensity function, maximum likelihood estimation, asymptotic behaviour, consistency, asymptotic normality and efficiency, LAN, software reliability models of Musa and Littlewood, martingale theory.

## 1. Introduction

Computer systems have become more and more important in modern society. The problem of estimating the reliability of computer software has therefore, over the last two decades, been receiving a great deal of attention. For this purpose a considerable number of models has been proposed [21]. There are now more than 40 in existence. Each of these statistical models, based on certain assumptions, is a simplification of reality which we want to describe or understand better. The development of so many different models, which are all supposed to describe the same thing — the evolution of the failure behaviour of a piece of software undergoing debugging — is largely due to a lack of agreement among modellers about how nature works.

When one wants to predict the reliability of computer software on the basis of past failure data, however, one needs more than just a software reliability model. The model parameter inference procedure and the incorporation of the results in prediction are also very important. In this paper we will study some important asymptotic properties of the maximum likelihood estimation procedure. We will derive, for a general class of counting processes, conditions on the intensity function, that are sufficient for these asymptotic properties to hold. We will further show that the intensity functions of some fairly well-known software reliability models, namely the models of Musa and Littlewood, satisfy these conditions.

Report BS-R9011
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Let be given a counting process $n(t)$, which counts up to an unknown value $n(\infty)=N$, $N<\infty$, which will be regarded as one of the model parameters. Only during a specific time interval $[0,\tau]$, jumps of the counting process $n(t)$ are observed. Its associated intensity function will be denoted by $\lambda(t;N,\psi)$, $t\in[0,\tau]$, $N\in\mathbb{N}$, $\psi\in\Psi$, $\Psi\subset\mathbb{R}^{p-1}$ for an integer $p$. Let $N_0$ and $\psi_0$ be the true parameter values. We want to obtain estimates for $N_0$ and $\psi_0$ and we are particularly interested in the behaviour of the parameter estimation when $N_0$ is large. Therefore we introduce a series of counting processes $n_\nu(t)$, $t\in[0,\tau], \nu=1,2,...$ by defining their associated intensity functions:

$$\lambda_\nu(t;\gamma,\psi) := \lambda(t;[\nu\gamma+½],\psi), \quad \nu=1,2,..., \tag{1.1}$$

where [.] stands for Ent(.) and $t\in[0,\tau]$, $\gamma\in\mathbb{R}^+$, $\psi\in\Psi$. If the real-life situation has $\nu=N_0$, then $\gamma=\gamma_0=1$ and $\psi=\psi_0$. We know from the theory of counting processes, that

$$m_\nu(t;\gamma,\psi) := n_\nu(t)-\int_0^t\lambda_\nu(s;\gamma,\psi)ds, \quad \nu=1,2,..., \tag{1.2}$$

are local square integrable martingales. We define for $\nu=1,2,...$ the stochastic process $x_\nu(t)$ by:

$$x_\nu(t) := \nu^{-1}n_\nu(t), \quad t\in[0,\tau]. \tag{1.3}$$

In most practical situations, as we soon will see, this stochastic process converges uniformly on $[0,\tau]$ in probability to a deterministic function $x_0(t)$ as $\nu\to\infty$.

We will assume that the maximum likelihood estimator $(\hat\gamma_\nu,\hat\psi_\nu)$ for $(\gamma_0,\psi_0)$ exists. Typically, $(\hat\gamma_\nu,\hat\psi_\nu)$ is a root of the likelihood equations

$$\frac{\partial}{\partial(\gamma,\psi)}\log L_\nu(\gamma,\psi;\tau) = 0, \quad \nu=1,2,..., \tag{1.4}$$

where the likelihood function $L_\nu(\gamma,\psi;t)$ is given by (see Aalen [1]):

$$L_\nu(\gamma,\psi;t) := \exp\left[\int_0^t\log\lambda_\nu(s;\gamma,\psi)\,dn_\nu(s)-\int_0^t\lambda_\nu(s;\gamma,\psi)\,ds\right]. \tag{1.5}$$

An important observation is the fact that information obtained for the asymptotic behaviour of $\hat\gamma_\nu$ can be transformed back directly to $\hat N$, the estimator of most interest. More precisely:

$$\begin{aligned}&\forall a>0:\quad P(|\hat\gamma_\nu-\gamma_0|\geqslant a)\to 0,\ \ \nu\to\infty\\ \Rightarrow\ &\forall b>0:\quad P(|\frac{\hat N}{N}-1|\geqslant b)\to 0,\ \ N\to\infty\end{aligned} \tag{1.6}$$

and

$$\begin{aligned}&\sqrt{\nu}(\hat\gamma_\nu-\gamma_0)\xrightarrow{D}\mathcal{N}(0,\sigma^2),\ \ \nu\to\infty\\ \Rightarrow\ &\sqrt{N}(\frac{\hat N}{N}-1)\xrightarrow{D}\mathcal{N}(0,\frac{\sigma^2}{\gamma_0^2}),\ \ N\to\infty.\end{aligned} \tag{1.7}$$

In the next section we will give for intensity functions $\lambda_\nu$ of a special form sufficient (and weak) conditions under which we have consistency, asymptotic normality and efficiency of the maximum likelihood estimators (MLE) and local asymptotic normality (LAN) of the model. In the third section of this paper we will discuss an application of the results in software reliability theory. Finally, in the fourth and last section a few concluding remarks are given concerning the possibility of weakening some of the conditions. We mention some results from recent simulations, as well as some plans for the future.

## 2. Asymptotic properties

We consider a sequence of models $(\lambda_\nu, m_\nu, x_\nu), \nu=1,2,...$ as defined by (1.1)-(1.3) in the previous section. For reasons of notational convenience we take $\theta:=(\gamma,\psi)^T\in\Theta$, $\Theta\subset\mathbb{R}^p$ for some integer $p$. In the sequel of this paper we will assume that the intensity function $\lambda_\nu$ is of the following form:

$$\lambda_\nu(t;\theta) = \nu\beta(t,\theta,x_\nu), \tag{2.1}$$

for an arbitrary non-negative and non-anticipating function $\beta$: $[0,\tau]\times\Theta\times K \to \mathbb{R}^+$. Non-anticipating means, that $\beta(t,\theta,x_\nu)=\beta(t,\theta,x_\nu|_{[0,t)})$. On $K:=D([0,\tau])$, the space of cadlag functions on $[0,\tau]$, we put the usual supnorm. The likelihood function $L_\nu(\theta,t)$ now becomes for $\theta\in\Theta$, $t\in[0,\tau]$ and $\nu=1,2,...$:

$$L_\nu(\theta,t) := \exp\left[\int_0^t \log \nu\beta(s,\theta_\nu,x_\nu)dn_\nu(s) - \nu\int_0^t \beta(s,\theta_\nu,x_\nu)ds\right]. \tag{2.2}$$

Furthermore, we define for $\theta\in\Theta$, $t\in[0,\tau]$, $i,j,k\in\{1,2,...,p\}$ and $\nu=1,2,...$:

$$C_\nu(\theta,t) := \log L_\nu(\theta,t), \tag{2.3}$$

$$U_{\nu i}(\theta,t) := \frac{\partial}{\partial\theta_i}C_\nu(\theta,t), \qquad \text{(score function)} \tag{2.4}$$

$$I_{\nu ij}(\theta,t) := \frac{\partial^2}{\partial\theta_i\partial\theta_j}C_\nu(\theta,t), \quad \text{(minus information matrix)} \tag{2.5}$$

$$R_{\nu ijk}(\theta,t) := \frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}C_\nu(\theta,t). \tag{2.6}$$

Wherever we use limits, convergence is always meant to be in probability as $\nu\to\infty$ under $\theta=\theta_0$. Consider the following global conditions:

(GC1) For all $x\in K$ and for all $\theta\in\Theta$ the intensity function $\beta$ satisfies:

$$\sup_{t\leq\tau}\beta(t,\theta,x) < \infty. \tag{2.7}$$

(GC2) (Lipschitz-continuity) There exists a constant $L$, not depending on $t$, such that for all $x,y\in K$ and all $t\in[0,\tau]$:

$$|\beta(t,\theta,x) - \beta(t,\theta,y)| \leq L\sup_{s\leq t}|x(s) - y(s)|. \tag{2.8}$$

Under the global conditions (GC1)-(GC2) the stochastic process $x_\nu(t)$, as defined in (1.3), converges uniformly on $[0,\tau]$ in probability to $x_0(t)$ as $\nu\to\infty$, where $x_0\in D([0,\tau])$ is the unique solution of

$$x(t) = \int_0^t \beta(s,\theta_0,x)ds. \tag{2.9}$$

This is proved by Kurtz [18]. Moreover, we consider the following local conditions:

(LC1) There exist neighbourhoods $\Theta_0$ and $K_0$ of $\theta_0, x_0$ respectively, such that the function $\beta(t,\theta,x)$ and its derivatives with respect to $\theta$ of the first, second and third order exist, are continuous functions of $\theta$ and $x$, bounded on $[0,\tau]\times\Theta_0\times K_0$.

(LC2) The function $\beta(t,\theta,x)$ is bounded away from zero on $[0,\tau]\times\Theta_0\times K_0$.

(LC3) The matrix $\Sigma=\{\sigma_{ij}(\theta_0)\}$ is positive definite, with for $i,j\in\{1,2,..,p\}$, $\theta\in\Theta_0$:

$$\sigma_{ij}(\theta) = \int_0^\tau \frac{\frac{\partial}{\partial\theta_i}\beta(s,\theta,x_0)\frac{\partial}{\partial\theta_j}\beta(s,\theta,x_0)}{\beta(s,\theta,x_0)}ds. \tag{2.10}$$

We are now able to formulate the main result of this paper. The proof of this theorem will be given in Appendix A.

**Theorem 1:**
*Given a counting process with intensity function $\lambda(t;N,\psi)$, $(N,\psi)$ an unknown p-dimensional parameter. As in section 1 ((1.1)-(1.3)), we can define an associated sequence of experiments by letting $\nu\rightarrow\infty$. Let $\theta_0=(\gamma_0,\psi_0)$ the true value of the parameter. Assume that for all $\nu$ the intensity function $\lambda_\nu(t;\theta)$ in the $\nu$-th experiment is of the form (2.1) for a certain function $\beta$ satisfying conditions (GC1)-(GC2) and (LC1)-(LC3). Then we have:*

(i) *Consistency of ML-estimators: With a probability tending to 1, the likelihood equations*

$$\frac{\partial}{\partial\theta}\log L_\nu(\theta,\tau) = 0, \quad \nu=1,2,... \tag{2.11}$$

*have exactly one consistent solution $\hat{\theta}_\nu$. Moreover this solution provides a local maximum of the likelihood function (2.2).*

(ii) *Asymptotic normality of the ML-estimators: Let $\hat{\theta}_\nu$ be the consistent solution of the maximum likelihood equations (2.11), then*

$$\sqrt{\nu}(\hat{\theta}_\nu-\theta_0) \overset{D}{\rightarrow} \mathcal{N}(0,\Sigma^{-1}), \quad \nu\rightarrow\infty, \tag{2.12}$$

*where $\Sigma$ is given by (2.10) and can be estimated consistently by $-I_\nu(\hat{\theta}_\nu,\tau)/\nu$.*

(iii) *Local asymptotic normality of the model: With $U_\nu$ given by (2.4), we have for all $h\in\mathbb{R}^p$:*

$$\log\frac{dP_{\theta_\nu}}{dP_{\theta_0}} - \nu^{-\frac{1}{2}}h^TU_\nu + \frac{1}{2}h^T\Sigma h \overset{P_{\theta_0}}{\rightarrow} 0, \quad \nu\rightarrow\infty, \tag{2.13}$$

*where $\theta_\nu=\theta_0+\nu^{-\frac{1}{2}}h$ and $\nu^{-\frac{1}{2}}U_\nu \overset{D}{\rightarrow} \mathcal{N}(0,\Sigma)$.*

(iv) *Asymptotic efficiency of the ML-estimators: $\hat{\theta}_\nu$ is asymptotically efficient in the sense that the limit distribution for any other regular estimator $\tilde{\theta}_\nu$ for $\theta_0$ satisfies:*

$$\lim_{\nu\rightarrow\infty}\mathcal{L}_{\theta_0}\left[\sqrt{\nu}(\tilde{\theta}_\nu-\theta_0)\right] = \mathcal{M}_{\theta_0} * \lim_{\nu\rightarrow\infty}\mathcal{L}_{\theta_0}\left[\sqrt{\nu}(\hat{\theta}_\nu-\theta_0)\right] \tag{2.14}$$

*for some distribution $\mathcal{M}_{\theta_0}$.*

An immediate consequence of this result on the asymptotic distribution of the ML-estimator $\hat{\theta}_\nu$ is the fact, that the Wald test statistic

$$-(\hat{\theta}_\nu-\theta_0)^TI_\nu(\hat{\theta}_\nu,\tau)(\hat{\theta}_\nu-\theta_0) \tag{2.15}$$

is asymptotically chi-squared distributed with $p$ degrees of freedom under the simple hypothesis $H_0:\theta=\theta_0$. The Rao test (or score) statistic

$$-U_\nu(\theta_0,\tau)^TI_\nu(\theta_0,\tau)^{-1}U_\nu(\theta_0,\tau) \tag{2.16}$$

and the Wilkes test (or likelihood ratio) statistic

$$2\left[C_\nu(\hat{\theta}_\nu,\tau)-C_\nu(\theta_0,\tau)\right] \tag{2.17}$$

have the same asymptotic distribution as the Wald test statistic. The equivalence of these tests is shown by Dzhaparidze [8]. In (2.15)-(2.17) $C_\nu, U_\nu$ and $I_\nu$ are given by (2.3)-(2.5).

## 3. An application to software reliability theory

Several statistical models are proposed in order to estimate the evolution in reliability of computer software during the debugging phase. In this section we will discuss two of these models; namely the Musa model [21] and the Littlewood model [19]. Both models concern the following experiment.

A computer program has been executed during a specified exposure period and the interfailure times are observed. The repairing of a fault takes place immediately after it produces a failure and no new faults are introduced with probability one. By using the information obtained from the experiment one can estimate the parameters of the underlying model, especially the total number of faults in the software. Maximum likelihood estimation will be used for this purpose.

We will need the following definitions. Let $N$ the unknown number of faults initially present in the software. Let the exposure period be $[0,\tau]$ and let $n(t)$, $t\in[0,\tau]$, denote the number of faults detected up to time $t$. Define $T_0 := 0$ and let $T_i$, $i=1,2,...,n(\tau)$, the failure time of the $i$-th occurring failure, while $t_i := T_i - T_{i-1}$, $i=1,2,...,n(\tau)$, denotes the interfailure time, that is the time between the $i$-th and the $(i-1)$-th occurring failure. Finally we define $t_{n(\tau)+1} := \tau - T_{n(\tau)}$.

**The Musa model:**
In the Musa model, Musa introduced in 1972, the failure rate of the program is at any time proportional to the number of remaining faults and each fault makes the same contribution to the failure rate. So if $(i-1)$ faults have already been detected, the failure rate for the $i$-th occurring failure, $\lambda_i$, becomes:

$$\lambda_i = \phi_0\left[N_0-(i-1)\right], \tag{3.1}$$

where $\phi_0$ is the true failure rate per fault (the occurrence rate) and $N_0$ is the true number of faults initially present in the software. In terms of counting processes we can write:

$$\lambda(t) = \phi_0\left[N_0-n(t-)\right], \tag{3.2}$$

where $\lambda(t)$, $t\in[0,\tau]$ denotes the failure rate at time $t$. The interfailure times $t_i$ are independent and exponentially distributed with parameter $\lambda_i$ given by (3.1) As in section 1 we let the true number of faults $N=\nu\gamma$ conceptually increase and we see that the corresponding sequence of intensity functions can be written in the standard form (2.1):

$$\lambda_\nu(t,\theta) = \nu\beta^{\mathrm{MU}}(t,\theta,x_\nu), \tag{3.3}$$

where $\theta=(\gamma,\phi)$, $x_\nu$ is given by (1.3) and

$$\beta^{\mathrm{MU}}(t;\gamma,\phi;x) = \phi\left[\gamma-x(t-)\right] \tag{3.4}$$

is defined on $[0,\tau]\times(\Theta\times K)$, where

$$\Theta\times K := \{(\gamma,\phi,x)\in\mathbf{R}^+\times\mathbf{R}^+\times D([0,\tau]) : 0\leqslant x(t)\leqslant\gamma\}. \tag{3.5}$$

**The Littlewood model:**
The main difference in the Littlewood model, introduced in 1979 by Littlewood, with respect to the Musa model is the fact that each fault does not make the same contribution to the failure rate $\lambda(t)$. Littlewood's argument for that is that larger faults will produce failures earlier than smaller ones. He treats $\phi_j$, the failure rate of fault $j$, as a stochastic variable and suggests a Gamma distribution:

$$\phi_j\sim\Gamma(a_0,b_0),\quad j=1,...,N. \tag{3.6}$$

Defining the expected occurrence rate of faults not occurred up to time $t$, as

$$\phi(t) := \mathrm{E}\{\phi_j \mid T_j > t\}, \tag{3.7}$$

with

$$\phi_j \sim \Gamma(a_0, b_0), \tag{3.8}$$

$$T_j \mid \phi_j = \phi \sim \exp(\phi), \tag{3.9}$$

a simple calculation yields:

$$\phi_j \mid T_j > t \sim \Gamma(a_0, b_0 + t) \tag{3.10}$$

and hence:

$$\phi(t) = \frac{a_0}{b_0 + t}. \tag{3.11}$$

An application of the so called *Innovation-theorem* (Aalen [1]) now shows, that the failure intensity of the software at time $t$ is given by:

$$\lambda(t) = \frac{a_0 \left[N_0 - n(t-)\right]}{b_0 + t}. \tag{3.12}$$

Letting $N = \nu\gamma$ conceptually increase as usual, we see that the corresponding sequence of intensity functions can again be written in the standard form (2.1):

$$\lambda_\nu(t;\theta) = \nu\beta^{\mathrm{LW}}(t;\theta,x_\nu) \tag{3.13}$$

where $\theta = (\gamma, a, b)$, $x_\nu$ is given by (1.3) and

$$\beta^{\mathrm{LW}}(t;\gamma,a,b;x) = \frac{a\left[\gamma - x(t-)\right]}{b+t} \tag{3.14}$$

is defined on $[0,\tau] \times (\Theta \times K)$, where

$$\Theta \times K := \{ (\gamma,a,b,x) \in \mathbf{R}^+ \times \mathbf{R}^+ \times \mathbf{R}^+ \times D([0,\tau]) : 0 \leqslant x(t) \leqslant \gamma \}. \tag{3.15}$$

By a simple reparametrization, namely:

$$\rho = \frac{1}{a}, \quad \mu = \frac{b}{a}, \tag{3.16}$$

we get from (3.14):

$$\beta(t;\gamma,\mu,\rho;x) = \frac{\gamma - x(t-)}{\mu + \rho t}, \tag{3.17}$$

defined on $[0,\tau] \times (\Theta \times K)$, where

$$\Theta \times K := \{ (\gamma,\mu,\rho,x) \in \mathbf{R}^+ \times \mathbf{R}^+ \times \mathbf{R} \times D([0,\tau]) : \mu + \rho\tau > 0,\ 0 \leqslant x(t) \leqslant \gamma \}. \tag{3.18}$$

Actually this is an extension of the Littlewood model, including also small values of $\rho \leqslant 0$. Note that when $\rho = 0$ we are dealing with the model we discussed earlier, namely the Musa model. We can therefore treat the Musa model as a special (limit-)case of the Littlewood model.

As an application we will show that the results of theorem 1 hold both for the Musa and the Littlewood model.

**Theorem 2:**
*Let $\tau>0$. We assume that the failure data are generated by the intensity function $\beta$ in (3.17) with true parameter value $\theta_0:=(\gamma_0,\mu_0,\rho_0)$ satisfying $\gamma_0>0$ and $\mu_0+\rho_0\tau>0$. If we define for $t\in[0,\tau]$*

$$x_0(t) := \begin{cases} \gamma_0\left[1-e^{-t/\mu_0}\right] & , \text{ if } \rho_0=0, \\ \gamma_0\left[1-\left[\dfrac{\mu_0}{\mu_0+\rho_0 t}\right]^{1/\rho_0}\right] & , \text{ if } \rho_0\neq 0, \end{cases} \tag{3.19}$$

*then $\beta$ satisfies conditions (GC1)-(GC2)&(LC1)-(LC3) and hence:*

(i) *With a probability tending to one, the likelihood equations have exactly one consistent solution $\hat{\theta}$.*
(ii) *A consistent solution $\hat{\theta}$ is asymptotically normal and efficient.*
(iii) *The model satisfies the LAN-property.*

The proof of theorem 2 will be given in appendix B. We will show there, that the following choice of $\Theta_0$ and $K_0$ will be appropriate:

$$\Theta_0 := [\epsilon_\gamma,M_\gamma]\times[\epsilon_\mu,M_\mu]\times[\epsilon_\rho,M_\rho], \tag{3.20}$$

$$K_0 := \{\, x\in K : ||x-x_0||_{\sup}\leqslant\epsilon_x \,\}, \tag{3.21}$$

where

$$0<\frac{1}{2}[\gamma_0+x_0(\tau)]<\epsilon_\gamma<\gamma_0<M_\gamma, \tag{3.22}$$

$$0<\epsilon_\mu<\mu_0<M_\mu, \tag{3.23}$$

$$-\frac{\mu_0}{\tau}<\epsilon_\rho<\rho_0<M_\rho \qquad \text{with } \epsilon_\rho<0<M_\rho \tag{3.24}$$

and

$$0<\epsilon_x<\frac{1}{2}[\gamma_0-x_0(\tau)]. \tag{3.25}$$

## 4. Concluding remarks, future investigations and open problems.

As stated in appendix A, theorems 1 and 2 remain valid if we replace (GC1)-(GC2)&(LC1)-(LC3) by the weaker set of conditions (C1)-(C4). Conditions comparable to these ones are also given by Cramer [6] and Kulldorff [17], using classical statistical techniques to prove consistency and asymptotic

normality of maximum likelihood estimators.
Nowadays modern methods have been developed by among others Ibragimov & Has'minskii [13], Jacod & Shiryaev [14] and Dzhaparidze & Valkeila [7] leading to the same results (and even more), without requiring the existence of higher derivatives of the intensity function and so weakening condition (LC1) (and (C2)) considerably. In [13] the parametric case is considered, but no theory for counting processes is developed, while in [14] and [7] only binary experiments for counting processes are studied. Also the work of Gill [11] and Van der Vaart [22] should be mentioned here. Therefore it seems very plausible that such methods can be applied also in our case. Indeed, the assumption of existence of the third derivative of $\beta$ with respect to $\theta$ can be abandoned (and for consistency even the existence of the second derivative!). Other conditions on $\beta$, - maybe weaker, but harder to verify - will replace them. In practical situations, however, the intensity functions tend to be very smooth and determining the existence of derivatives (w.r.t. $\theta$) is relatively easy.

Recently, we have made a beginning with the study of the behaviour of the ML-estimators in practice, computed from simulated data generated by the Musa model. The simulation results, which are to appear soon, confirm the asymptotic theory as derived in this paper. They also show on the other hand, that the convergence in distribution is rather slow and that for small values of $N_0$ the distributions of $\hat{N}$ and $\hat{\phi}$ can be very skew. With use of the Wilkes likelihood ratio test statistic (2.17), however, we were able to build confidence intervals for the model parameters that are very satisfactory.
Moreover, we think we can improve the construction of confidence intervals by making use of parametric bootstrap methods. The validity of the bootstrap method will follow by standard arguments on contiguity, regular estimators and the Skorohod-Dudley-Wichura almost sure representation theorem (see Gill [12]).
Furthermore, another topic of future investigation will be the study of goodness of fit tests. We intend to follow the martingale approach of Khmaladze [15]. See also Geurts et al. [10].
Of course, our ultimate goal will be to study more realistic models, incorporating imperfect repair and software growth. We have recently constructed such a model and are now investigating, whether this model fits in the theory developed so far.

Finally, we note that theorem 1 doesn't claim that the maximum likelihood equations have a unique solution. It only states that with a probability tending to one, among all these solutions, only one of them will be consistent. Moek [20] developed an easy criterion for the existence of a unique solution of the ML equations for the Musa model. The problem in case of the Littlewood model, however, is much harder and in fact still an open question. Finding such a criterion in the Littlewood case or developing an algorithm in order to determine the consistent one from a set of solutions from the ML equations — probably with use of compactification theory (see Bahadur [3]) — certainly will be an important topic in future investigations.

**Acknowledgement:**
I would like to thank Richard D. Gill and Kacha Dzhaparidze for their many helpful comments.

**Appendix A: Proof of theorem 1.**

One can easily check that (GC1)-(GC2)&(LC1)-(LC2) are sufficient for the alternative set of conditions (C1)-(C3) to hold (see Borgan [5]):

(C1) The function $\beta$ is continuous with respect to $\theta$, and strictly positive.

(C2) There exists a non-negative deterministic function $x_0 \in K$ and neighbourhoods $\Theta_0, K_0$ of $\theta_0$ and $x_0$ respectively, such that the derivatives of $\beta(t,\theta,x)$ with respect to $\theta$ of the first, second and third order exist and are continuous functions of $\theta$, on $[0,\tau] \times \Theta_0 \times K_0$. With $x_\nu, \nu=1,2,...$ the stochastic process given in (1.2), $x_0 \in K$ has to satisfy for all $i,j \in \{1,2,...,p\}$ as $\nu \rightarrow \infty$:

$$\int_0^\tau \frac{\frac{\partial}{\partial\theta_i}\beta(s;\theta_0,x_\nu)\frac{\partial}{\partial\theta_j}\beta(s;\theta_0,x_\nu)}{\beta(s;\theta_0,x_\nu)}ds \xrightarrow{P} \int_0^\tau \frac{\frac{\partial}{\partial\theta_i}\beta(s;\theta_0,x_0)\frac{\partial}{\partial\theta_j}\beta(s;\theta_0,x_0)}{\beta(s;\theta_0,x_0)}ds < \infty \tag{A.1}$$

$$\int_0^\tau \left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\beta(s;\theta_0,x_\nu)\right]^2 \beta(s;\theta_0,x_\nu)ds \xrightarrow{P} \int_0^\tau \left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\beta(s;\theta_0,x_0)\right]^2 \beta(s;\theta_0,x_0)ds < \infty. \tag{A.2}$$

(C3) There exist functions G and H and neighbourhoods $\Theta_0, K_0$ of $\theta_0$ and $x_0$ respectively, such that for all $t \in [0,\tau]$ and $x \in K_0$:

$$\sup_{\theta\in\Theta_0} \left| \frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\beta(t;\theta,x)\right| \leqslant G(t,x), \tag{A.3}$$

$$\sup_{\theta\in\Theta_0} \left| \frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\log\beta(t,\theta,x)\right| \leqslant H(t,x), \tag{A.4}$$

and moreover the functions G and H satisfy as $\nu \rightarrow \infty$:

$$\int_0^\tau G(s,x_\nu)ds \xrightarrow{P} \int_0^\tau G(s,x_0)ds < \infty, \tag{A.5}$$

$$\int_0^\tau H(s,x_\nu)\beta(s,\theta_0,x_\nu)ds \xrightarrow{P} \int_0^\tau H(s,x_0)\beta(s,\theta_0,x_0)ds < \infty, \tag{A.6}$$

$$\int_0^\tau H^2(s,x_\nu)\beta(s,\theta_0,x_\nu)ds \xrightarrow{P} \int_0^\tau H^2(s,x_0)\beta(s,\theta_0,x_0)ds < \infty. \tag{A.7}$$

In the proof of (i) and (ii) we will follow the lines of Borgan [5].

(i) **Consistency of ML-estimators:**

By a Taylor series expansion we get for any $\theta \in \Theta$:

$$U_{\nu i}(\theta,\tau) = U_{\nu i}(\theta_0,\tau)+\sum_{j=1}^{p}(\theta_j-\theta_{j0})I_{\nu ij}(\theta_0,\tau) + \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p}(\theta_j-\theta_{j0})(\theta_k-\theta_{k0})R_{\nu ijk}(\theta_\nu^*,\tau), \tag{A.8}$$

where $\theta_\nu^* = \theta_\nu^*(\theta)$ is on the line segment joining $\theta$ and $\theta_0$. We shall show that:

$$\nu^{-1}U_{\nu i}(\theta_0,\tau) \overset{P}{\to} 0, \tag{A.9}$$

$$\nu^{-1}I_{\nu ij}(\theta_0,\tau) \overset{P}{\to} -\sigma_{ij}(\theta_0), \tag{A.10}$$

$$\lim_{\nu\to\infty} P\left[\,|\nu^{-1}R_{\nu ijk}(\theta,\tau)|<M\right] = 1, \tag{A.11}$$

for all $i,j,k \in \{1,2,...,p\}$, all $\theta \in \Theta_0$ and a certain finite constant $M$, not depending on $\theta$. From (A.9)-(A.11) the statement (i) in the theorem will follow by a standard argument (see Billingsley [4], pp. 10-16).

Let us first prove (A.9). From (1.2) we get that (2.4) evaluated at the true parameter $\theta_0$, equals:

$$\begin{aligned} U_{\nu i}(\theta_0,t) &= \frac{\partial}{\partial\theta_i}\left[\int_0^t \log\nu\beta(s,\theta_0,x_\nu)dn_\nu(s) - \int_0^t \nu\beta(s,\theta_0,x_\nu)ds\right] \\ &= \int_0^t \frac{\partial}{\partial\theta_i}\left[\log\beta(s,\theta_0,x_\nu)\right] d\left[m_n(s,\theta_0)+\int_0^s \nu\beta(u,\theta_0,x_\nu)du\right] - \nu\int_0^t \frac{\partial}{\partial\theta_i}\beta(s,\theta_0,x_\nu)ds \\ &= \int_0^t \frac{\frac{\partial}{\partial\theta_i}\beta(s,\theta_0,x_\nu)}{\beta(s,\theta_0,x_\nu)} dm_\nu(s,\theta_0). \end{aligned} \tag{A.12}$$

Because $\beta$ is a non-anticipating function, it follows that $U_{\nu i}(\theta_0,t)$ is a stochastic integral of a predictable process w.r.t. a local martingale and hence a local square integrable martingale. Its variance process equals

$$<U_{\nu i}(\theta_0,\cdot), U_{\nu i}(\theta_0,\cdot)>(t) = \int_0^t \left[\frac{\frac{\partial}{\partial\theta_i}\beta(s,\theta_0,x_\nu)}{\beta(s,\theta_0,x_\nu)}\right]^2 \nu\beta(s,\theta_0,x_\nu)ds. \tag{A.13}$$

By condition (C2), $\nu^{-1}<U_{\nu i}(\theta_0,\cdot), U_{\nu i}(\theta_0,\cdot)>(\tau)$ converges in probability to some finite quantity as $\nu\to\infty$. Therefore, an application of Lenglart's inequality (see Andersen & Gill [2]) gives that for all $\delta,\eta>0$ we have

$$P\left[\sup_{t\in[0,\tau]}|\nu^{-1}U_{\nu i}(\theta_0,t)|>\eta\right] \leqslant \frac{\delta}{\eta^2}+P\left[\nu^{-2}<U_{\nu i}(\theta_0,\cdot), U_{\nu i}(\theta_0,\cdot)>(\tau)>\delta\right] \tag{A.14}$$

which proves (A.9).

To prove (A.10), note that by using (1.2):

$$\nu^{-1}I_{\nu ij}(\theta_0,\tau) = \nu^{-1}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left[\int_0^\tau \log\nu\beta(s,\theta_0,x_\nu)dn_\nu(s) - \int_0^\tau \nu\beta(s,\theta_0,x_\nu)ds\right]$$

$$= \nu^{-1}\int_0^\tau \frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\beta(s,\theta_0,x_\nu)d\left[m_\nu(s,\theta_0)+\int_0^s \nu\beta(u,\theta_0,x_\nu)du\right]$$

$$-\int_0^\tau \frac{\partial^2}{\partial\theta_i\partial\theta_j}\beta(s,\theta_0,x_\nu)ds$$

$$= \nu^{-1}\int_0^\tau \frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\beta(s,\theta_0,x_\nu)dm_\nu(s,\theta_0)$$

$$+\int_0^\tau \left[\beta(s,\theta_0,x_\nu)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\beta(s,\theta_0,x_\nu) - \frac{\partial^2}{\partial\theta_i\partial\theta_j}\beta(s,\theta_0,x_\nu)\right]ds$$

$$= \nu^{-1}\int_0^\tau \frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\beta(s,\theta_0,x_\nu)dm_\nu(s,\theta_0)$$

$$-\int_0^\tau \frac{\frac{\partial}{\partial\theta_i}\beta(s,\theta_0,x_\nu)\frac{\partial}{\partial\theta_j}\beta(s,\theta_0,x_\nu)}{\beta(s,\theta_0,x_\nu)}ds. \tag{A.15}$$

That the first of the last two terms in (A.15) converges in probability to zero follows by condition (C2) and an application of Lenglart's inequality similar to (A.14). By (2.8), (A.10) is an immediate consequence.

Finally to prove (A.11), we note that (A.3)&(A.4) in (C3) give for all $i,j,k \in \{1,2,...,p\}$ and all $\theta \in \Theta_0$:

$$\left|\nu^{-1}R_{\nu ijk}(\theta,\tau)\right| = \left|\nu^{-1}\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\left[\int_0^\tau \log\nu\beta(s,\theta,x_\nu)dn_\nu(s) - \int_0^\tau \nu\beta(s,\theta,x_\nu)ds\right]\right|$$

$$\leqslant \nu^{-1}\int_0^\tau \left|\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\log\beta(s,\theta,x_\nu)\right|dn_\nu(s) + \int_0^\tau \left|\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\beta(s,\theta,x_\nu)\right|ds$$

$$\leqslant \nu^{-1}\int_0^\tau H(t,x_\nu)dn_\nu(s) + \int_0^\tau G(t,x_\nu)ds. \tag{A.16}$$

By another application of Lenglart's inequality we get for all $\delta,\eta>0$:

$$P\left[\sup_{t\in[0,\tau]}\left|\nu^{-1}\int_0^t H(s,x_\nu)dn_\nu(s) - \nu^{-1}\int_0^t H(s,x_\nu)\nu\beta(s,\theta_0,x_\nu)ds\right| > \eta\right]$$

$$\leqslant \frac{\delta}{\eta^2} + P\left[\nu^{-2}\int_0^\tau H^2(s,x_\nu)\nu\beta(s,\theta_0,x_\nu)ds\right]. \tag{A.17}$$

So (A.6)&(A.7) in (C3) yield

$$\nu^{-1}\int_0^\tau H(s,x_\nu)dn_\nu(s) \rightarrow \int_0^\tau H(s,x)\beta(s,\theta_0,x)ds \tag{A.18}$$

in probability as $\nu\rightarrow\infty$. Combining this with (A.5) in (C3) and with (A.17) , we get (A.11) and hence the consistency of the ML-estimators.

**(ii) Asymptotic normality of the ML-estimators:**

Let $\hat{\theta}_\nu$ be the consistent solution of the likelihood equations (2.11). Taylor expanding $U_{\nu i}(\hat{\theta}_\nu,\tau)$ around $\theta_0$ gives:

$$\begin{aligned} 0 &= \nu^{-\frac{1}{2}} U_{\nu i}(\hat{\theta}_\nu,\tau) \\ &= \nu^{-\frac{1}{2}} U_{\nu i}(\theta_0,\tau)+\sum_{j=1}^{p}\nu^{\frac{1}{2}}(\hat{\theta}_{\nu j}-\theta_{0j})\nu^{-1}I_{\nu ij}(\theta_\nu^*,\tau), \end{aligned} \tag{A.19}$$

where $\theta_\nu^*=\theta_\nu^*(\hat{\theta}_\nu)$ is on the line segment between $\hat{\theta}_\nu$ and $\theta_0$. Therefore, the statement (ii) in theorem 1 will follow by an argument in Billingsley [4], if we can prove that for all $i,j\in\{1,2,...,p\}$ and for any random $\theta_\nu^*$, such that $\theta_\nu^*\to\theta_0$ in probability as $\nu\to\infty$:

$$\nu^{-\frac{1}{2}} U_\nu(\theta_0,\tau) \overset{D}{\to} \mathcal{N}(0,\Sigma) \tag{A.20}$$

$$\nu^{-1}I_{\nu ij}(\theta_\nu^*,\tau) \overset{P}{\to} -\sigma_{ij}(\theta_0). \tag{A.21}$$

Let us first prove (A.20). By (A.12) and condition (C2)

$$\begin{aligned} &<\nu^{-\frac{1}{2}} U_{\nu i}(\theta_0,\cdot),\nu^{-\frac{1}{2}} U_{\nu j}(\theta_0,\cdot)>(\tau) \\ &= \int_0^\tau \frac{\partial}{\partial\theta_i}\log\beta(s,\theta_0,x_\nu)\frac{\partial}{\partial\theta_j}\log\beta(s,\theta_0,x_\nu)\cdot\beta(s,\theta_0,x_\nu)ds \\ &= \int_0^\tau \frac{\frac{\partial}{\partial\theta_i}\beta(s,\theta_0,x_\nu)\frac{\partial}{\partial\theta_j}\beta(s,\theta_0,x_\nu)}{\beta(s,\theta_0,x_\nu)}ds \\ &\overset{P}{\to} \int_0^\tau \frac{\frac{\partial}{\partial\theta_i}\beta(s,\theta_0,x)\frac{\partial}{\partial\theta_j}\beta(s,\theta_0,x)}{\beta(s,\theta_0,x)}ds \end{aligned} \tag{A.22}$$

as $\nu\to\infty$ for all $i,j\in\{1,2,...,p\}$. Furthermore condition (C2) and dominated convergence give that for all $j\in\{1,2,...,p\}$ and all $\epsilon>0$ as $\nu\to\infty$:

$$\int_0^\tau \left[\frac{\frac{\partial}{\partial\theta_i}\beta(s,\theta_0,x_\nu)}{\nu^{\frac{1}{2}}\beta(s,\theta_0,x_\nu)}\right]^2 \nu\beta(s,\theta_0,x_\nu)I\left\{\left|\frac{\frac{\partial}{\partial\theta_i}\beta(s,\theta_0,x_\nu)}{\nu^{\frac{1}{2}}\beta(s,\theta_0,x_\nu)}\right|>\epsilon\right\}ds \overset{P}{\to} 0 \tag{A.23}$$

in probability as $\nu\to\infty$. From this (A.20) follows by an application of the martingale central limit theorem (see Andersen & Gill [2]).

Let us return to (A.21). By a Taylor series expansion we have when $\theta_\nu^*\in\Theta$:

$$\nu^{-1}I_{\nu ij}(\theta_\nu^*,\tau) = \nu^{-1}I_{\nu ij}(\theta_0,\tau)+\nu^{-1}\sum_{k=1}^{p}(\theta_{\nu k}^*-\theta_{0k})R_{\nu ijk}(\bar{\theta}_\nu,\tau), \tag{A.24}$$

where $R_{\nu ijk}(\theta,t)$ is defined in (A.5) and $\bar{\theta}_\nu=\bar{\theta}_\nu(\theta_\nu^*)$ is on the line segment joining $\theta_\nu^*$ and $\theta_0$. By (A.10) and (A.11), the first term in the right hand side of (A.24) converges in probability to $-\sigma_{ij}(\theta_0)$ as $\nu\to\infty$, while the second term is bounded in probability by $pM|\theta_\nu^*-\theta_0|_p$ for some finite constant $M$ not depending on $\theta_\nu^*$. This proves (A.21) and thus the asymptotic normality of the ML-estimators.

**(iii) Local asymptotic normality of the model:**

For sake of convenience, we introduce some notations. For a function $f: \mathbb{R}^p \to \mathbb{R}$, which is at least three times differentiable, we write:

$$\frac{\partial^3}{\partial x^3} f(x_0) := \left[ \frac{\partial^3}{\partial x_i \partial x_j \partial x_k} f|_{x=x_0} \right]_{1 \leq i,j,k \leq p}, \tag{A.25}$$

the (three-dimensional) $p \times p \times p$ matrix of third order partial derivatives, evaluated in $x_0$. Furthermore, for a (three-dimensional) $p \times p \times p$ matrix $Y = (y_{ijk})$, and a $p$-vector $g = (g_i)$, we define:

$$g^T Y g^{<2>} := \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} g_i g_j g_k y_{ijk}. \tag{A.26}$$

We define for $h \in \Theta$:

$$\theta_\nu(h) \quad := \quad \theta_0 + \nu^{-\frac{1}{2}} h, \quad \nu = 1, 2, .... \tag{A.27}$$

For fixed $h$ and $\nu$, using the fact that $\lambda_\nu = \nu\beta$, we have that the log likelihood ratio for $\theta_\nu(h)$ against $\theta_0$ is:

$$\begin{aligned} Q_\nu(h) &= \log \frac{dP_{\theta_\nu(h)}}{dP_{\theta_0}} \\ &= \log dP_{\theta_\nu(h)} - \log dP_{\theta_0} \\ &= [\int_0^\tau \log\lambda_\nu(s, \theta_\nu(h)) dn_\nu(s) - \int_0^\tau \lambda_\nu(s, \theta_\nu(h)) ds] - [\int_0^\tau \log\lambda_\nu(s, \theta_0) dn_\nu(s) - \int_0^\tau \lambda_\nu(s, \theta_0) ds] \\ &= C_\nu(\theta_\nu(h), \tau) - C_\nu(\theta_0, \tau), \end{aligned} \tag{A.28}$$

where $C_\nu$ is given by (2.3). Of course $Q_\nu(0) = 0$, and because

$$\frac{\partial}{\partial h} \theta_\nu(h) = \nu^{-\frac{1}{2}}, \tag{A.29}$$

the first, second and third order derivatives of $Q_\nu$ with respect to $h$ are:

$$\frac{\partial}{\partial h} Q_\nu(h) \quad = \quad \nu^{-\frac{1}{2}} \; U_\nu(\theta_\nu(h), \tau), \tag{A.30}$$

$$\frac{\partial^2}{\partial h^2} Q_\nu(h) \quad = \quad \nu^{-1} \; I_\nu(\theta_\nu(h), \tau), \tag{A.31}$$

$$\frac{\partial^3}{\partial h^3} Q_\nu(h) \quad = \quad \nu^{-\frac{3}{2}} \; R_\nu(\theta_\nu(h), \tau), \tag{A.32}$$

where $U_\nu$, $I_\nu$ and $R_\nu$ are given by (2.4)-(2.6). Hence we get the Taylor expansion:

$$\begin{aligned} Q_\nu(h) \quad = \quad & \nu^{-\frac{1}{2}} h^T \; U_\nu(\theta_0, \tau) \\ & + \frac{1}{2} \nu^{-1} h^T \; I_\nu(\theta_0, \tau) \; h \\ & + \frac{1}{6} \nu^{-\frac{3}{2}} h^T \; R_\nu(\theta_\nu^*, \tau) \; h^{<2>}, \end{aligned} \tag{A.33}$$

where $\theta_\nu^*$ is somewhere on the linesegment between $\theta_0$ and $\theta_\nu(h)$.

We have already proved (see (A.20)&(A.10)) that:

$$\nu^{-\frac{1}{2}} U_\nu(\theta_0,\tau) \xrightarrow{D} \mathcal{N}(0,\Sigma), \tag{A.34}$$

$$\nu^{-1} I_\nu(\theta_0) \xrightarrow{P} -\Sigma, \tag{A.35}$$

as $\nu \rightarrow \infty$. We see that we will have the local asymptotic normality (LAN) property (2.13), if only we can show that

$$\nu^{-\frac{3}{2}} R_\nu(\theta_\nu^*,\tau) \xrightarrow{P} 0, \tag{A.36}$$

as $\nu \rightarrow \infty$, for all sequences $(\theta_\nu^*)$ converging to $\theta_0$. But this follows directly from (A.11).

This proves part (iii) of theorem 1.

(iv) **Asymptotic efficiency of the ML-estimators:**

In (ii), the proof of the asymptotic normality of $\hat{\theta}_\nu$, we found that:

$$\nu^{-\frac{1}{2}} U_\nu \xrightarrow{D(\theta_\nu)} \mathcal{N}(0,\Sigma), \tag{A.37}$$

$$\sqrt{\nu}(\hat{\theta}_\nu - \theta_0) - \Sigma^{-1}\nu^{-\frac{1}{2}} U_\nu + o_P(1) \xrightarrow{D(\theta_0)} \mathcal{N}(0,\Sigma^{-1}). \tag{A.38}$$

Moreover, the LAN-property of the model, proved in (iii), gives us by using (A.37):

$$\log\frac{dP_{\theta_\nu}}{dP_{\theta_0}} = \nu^{-\frac{1}{2}} h^T U_\nu - \frac{1}{2} h^T \Sigma h + o_P(1) \xrightarrow{D(\theta_0)} \mathcal{N}(-\frac{1}{2}h^T\Sigma h, h^T \Sigma h). \tag{A.39}$$

Hence:

$$\begin{bmatrix} \sqrt{\nu}(\hat{\theta}_\nu - \theta_0) \\ \log\dfrac{dP_{\theta_\nu}}{dP_{\theta_0}} \end{bmatrix} \xrightarrow{D(\theta_0)} \mathcal{N}\left(\begin{bmatrix} 0 \\ -\frac{1}{2}h^T\Sigma h \end{bmatrix}, \begin{bmatrix} \Sigma^{-1} & h \\ h^T & h^T\Sigma h \end{bmatrix}\right). \tag{A.40}$$

From Le Cam's third lemma (see Van der Vaart [22], pp. 180-181), we can now conclude the contiguity of $P_{\theta_\nu}$ and $P_{\theta_0}$ and we have:

$$\sqrt{\nu}(\hat{\theta}_\nu - \theta_0) \xrightarrow{D(\theta_\nu)} \mathcal{N}(h, \Sigma^{-1}) \tag{A.41}$$

and thus

$$\sqrt{\nu}(\hat{\theta}_\nu - \theta_\nu) = \sqrt{\nu}(\hat{\theta}_\nu - (\theta_0 + \nu^{-\frac{1}{2}} h)) \xrightarrow{D(\theta_\nu)} \mathcal{N}(0,\Sigma^{-1}). \tag{A.42}$$

Combining (A.38)&(A.42), we see that for all $h \in \Theta$:

$$\lim_{\nu \to \infty} \mathscr{L}_{\theta_\nu}\left[\sqrt{\nu}(\hat{\theta}_\nu - \theta_\nu)\right] = \lim_{\nu \to \infty} \mathscr{L}_{\theta_0}\left[\sqrt{\nu}(\hat{\theta}_\nu - \theta_0)\right]; \tag{A.43}$$

this means by definition the regularity of the maximum likelihood estimator $\hat{\theta}_\nu$. Now we use an appropriate version of the well-known convolution theorem (see [16] or [22]), which states in our case that the limit-distribution of any regular estimator $\tilde{\theta}_\nu$ of $\theta_0$ satisfies:

$$\lim_{\nu \to \infty} \mathscr{L}_{\theta_0}\left[\sqrt{\nu}(\tilde{\theta}_\nu - \theta_0)\right] = \mathscr{N}(0, \Sigma^{-1}) * \mathscr{M}_{\theta_0}. \tag{A.44}$$

Because (A.38) implies that for $\tilde{\theta}_\nu := \hat{\theta}_\nu$ we get $\mathscr{M}_{\theta_0} \equiv 1$ in (A.44), we have proved that the maximum likelihood estimator $\hat{\theta}_\nu$ is asymptotically efficient.

Theorem 1 is now completely proved. □

**Appendix B: Proof of theorem 2.**

We will prove that the intensity function $\beta$ of the model, given by (3.17)-(3.19), satisfies the conditions (GC1)-(GC2)&(LC1)-(LC3).

(GC1) **Boundedness:**

For all $t \in [0,\tau]$, $\theta \in \Theta$ and $x \in K$, we have:

$$\sup_{s \leqslant t} \beta(s,\theta,x) = \sup_{s \leqslant t} \frac{(\gamma - x(s-))}{\mu + \rho s} \leqslant \frac{\gamma}{\min(\mu, \mu+\rho\tau)} < \infty, \tag{B.1}$$

because $\mu > 0$ and $\mu + \rho\tau > 0$. This proves (GC1).

(GC2) **Lipschitz-continuity:**

For fixed $\theta \in \Theta$ define a constant $L := [\min(\mu, \mu+\rho\tau)]^{-1}$. Then for all $x,y \in K$ and all $t \in [0,\tau]$ we have:

$$\begin{aligned} |\beta(t,\theta,x) - \beta(t,\theta,y)| &= \left| \frac{\gamma - x(t-)}{\mu+\rho t} - \frac{\gamma - y(t-)}{\mu + \rho t} \right| \\ &= \left| \frac{y(t-) - x(t-)}{\mu + \rho t} \right| \\ &\leqslant \frac{|x(t-) - y(t-)|}{\min(\mu, \mu+\rho\tau)} \\ &\leqslant L \sup_{s \leqslant t} |x(s) - y(s)|. \end{aligned} \tag{B.2}$$

This proves the Lipschitz-continuity.

Hence $x_\nu(t) \to x_0(t)$ uniformly on $[0,\tau]$ in probability. The deterministic function $x_0 \in K$ is the unique solution of the integral equation:

$$x(t) = \int_0^t \beta(s, \theta_0, x(s-)) ds. \tag{B.3}$$

Equivalently we can solve the Dirichlet differential equation:

$$\frac{d}{dt} x(t) = \frac{\gamma_0 - x(t-)}{\mu_0 + \rho_0 t}, \quad x(0) = 0. \tag{B.4}$$

One can easily find that the unique solution of (B.4) is given by (3.19).

(LC1) **Differentiability of $\beta$ w.r.t. $\theta$:**

Let us consider $\beta$ for fixed $t \in [0,\tau]$ as a function of $\theta$ and $x$ only:

$$\beta(\gamma, \mu, \rho, x; t) = \frac{\gamma - x(t-)}{\mu + \rho t} \tag{B.5}$$

The function $\beta$ itself and all its derivatives with respect to $\theta$ of the first, second and third order are of the form:

$$H(\gamma,\mu,\rho,x;t) = \frac{ct^k[\gamma-x(t-)]^l}{[\mu+\rho t]^m}, \tag{B.6}$$

where $c\in\mathbf{Z}$, $k\in\{0,1,2,3\}$, $l\in\{0,1\}$ and $m\in\{1,2,3,4\}$. This rational function is for fixed $t\in[0,\tau]$ obviously continuous in $\theta$ on $\Theta$ and even Lipschitz continuous in $x$ on $K$ with respect to the supnorm. To overcome boundedness-problems, however, we have to restrict ourselves to the neighbourhood $\Theta_0$ of $\theta_0$ given by (3.20). We have uniform in $x\in K$ and $t\in[0,\tau]$ for all $\theta\in\Theta_0$:

$$|H(a,b,\gamma,x;t)| \leqslant \frac{|c|\max(1,\tau^k)M_\gamma^l}{[\epsilon_\mu+\epsilon_\rho\tau)]^m} < \infty. \tag{B.7}$$

This proves (LC1).

(LC2) **$\beta$ bounded away from zero:**

With $\Theta_0$ and $K_0$ defined in (3.20)-(3.21) we have uniform in $t$ on $[0,\tau]\times\Theta_0\times K_0$:

$$\begin{aligned}\beta(t,\theta,x) &= \frac{\gamma-x(t-)}{\mu+\rho t}\\ &\geqslant \frac{\epsilon_\gamma-[x_0(t-)+\epsilon_x]}{M_\mu+M_\rho t}\\ &> \frac{\frac{1}{2}[\gamma_0+x_0(\tau)] - x_0(t) - \frac{1}{2}[\gamma_0-x_0(\tau)]}{M_\mu+M_\rho\tau}\\ &\geqslant 0.\end{aligned} \tag{B.8}$$

Hence $\beta$ is bounded away from zero on $[0,\tau]\times\Theta_0\times K_0$.

(LC3) **The matrix $\Sigma$ is positive definite:**

Using (LC1)-(LC2), we see that the coefficients of the matrix $\Sigma$ are well-defined and can easily be computed. Writing

$$\Sigma := \begin{bmatrix}\sigma_{\gamma\gamma} & \sigma_{\mu\gamma} & \sigma_{\rho\gamma}\\ \sigma_{\gamma\mu} & \sigma_{\mu\mu} & \sigma_{\rho\mu}\\ \sigma_{\gamma\rho} & \sigma_{\mu\rho} & \sigma_{\rho\rho}\end{bmatrix}, \tag{B.9}$$

$$Q := \frac{\mu_0}{\mu_0+\rho_0\tau}, \tag{B.10}$$

we find in case $\rho_0\neq0$:

$$\sigma_{\gamma\gamma} = \frac{1}{\gamma_0}[Q^{-1/\rho_0}-1], \tag{B.11}$$

$$\sigma_{\mu\mu} = \frac{\gamma_0}{\mu_0^2(2\rho_0+1)}[1-Q^{2+1/\rho_0}], \tag{B.12}$$

$$\sigma_{\rho\rho} = \frac{\gamma_0}{\rho_0^2}\left[[1-Q^{-1/\rho_0}]-\frac{1}{1+\rho_0}[1-Q^{1+1/\rho_0}]+\frac{1}{1+2\rho_0}[1-Q^{2+1/\rho_0}]\right], \tag{B.13}$$

$$\sigma_{\gamma\mu} = \sigma_{\mu\gamma} = \frac{-1}{\mu_0\rho_0}[1-Q], \tag{B.14}$$

$$\sigma_{\gamma\rho} = \sigma_{\rho\gamma} = \frac{1}{\rho_0^2}[1-Q+\log Q], \tag{B.15}$$

$$\sigma_{\mu\rho} = \sigma_{\rho\mu} = \frac{\gamma_0}{\mu_0\rho_0}\left[\frac{1}{1+\rho_0}[1-Q^{1+1/\rho_0}]-\frac{1}{1+2\rho_0}[1-Q^{2+1/\rho_0}]\right]. \tag{B.16}$$

It is easy to verify that the limits of these coefficients exist when $\rho_0 \to 0$; the resulting matrix $\Sigma$ in case $\rho_0 = 0$, is given by:

$$\Sigma = \begin{bmatrix} \frac{1}{\gamma_0}[e^{\frac{\tau}{\mu_0}}-1] & -\frac{\tau}{\mu_0^2} & -\frac{\tau^2}{\mu_0^2} \\ -\frac{\tau}{\mu_0^2} & \frac{\gamma_0}{\mu_0^2}[1-e^{-\frac{\tau}{\mu_0}}] & \frac{\gamma_0}{\mu_0}[1-(1+\frac{\tau}{\mu_0})e^{-\frac{\tau}{\mu_0}}] \\ -\frac{\tau^2}{\mu_0^2} & \frac{\gamma_0}{\mu_0}[1-(1+\frac{\tau}{\mu_0})e^{-\frac{\tau}{\mu_0}}] & 2\gamma_0[1-(\frac{\tau^2}{2\mu_0^2}+\frac{\tau}{\mu_0}+1)e^{-\frac{\tau}{\mu_0}}] \end{bmatrix}. \tag{B.17}$$

In both cases, it is possible to show that $\Sigma$ is positive definite for all $\tau > 0$, using a well-known theorem in linear algebra (see Gantmacher [9], pp. 304-308), which states that a real, symmetric $n \times n$-matrix $A$ is positive definite if and only if the successive principal minors $D_1, \ldots, D_n$ are all strictly positive. The verification of this requires some long and tedious computations and is therefore omitted in this appendix. We have found that $\Sigma$ is positive definite and hence non-singular for all $\tau > 0$.

We have shown that the conditions (GC1)-(GC2)&(LC1)-(LC3) are satisfied. Hence we can apply theorem 1. This proves theorem 2. □

**References**


[1] O.O. Aalen (1978), Non-parametric inference for a family of counting processes. *Annals of Statistics* **6** 701-726.

[2] P.K. Andersen & R.D. Gill (1982), Cox's Regression model for counting processes: A large sample study. *Annals of Statistics* **10** 1100-1120.

[3] R.R. Bahadur (1967), Rates of convergence of estimates and some test statistics. *Annals of Mathematical Statistics* **38** 303-324.

[4] P. Billingsley (1961), *Statistical Inference for Markov Processes.* University of Chicago Press, Chicago.

[5] Ø. Borgan (1984), Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scandinavian Journal of Statistics* **1** 1-16.

[6] H. Cramer (1946), *Mathematical Methods of Statistics.* Princeton University Press, Princeton.

[7] K. Dzhaparidze & E. Valkeila (1988), On the Hellinger type distances for filtered experiments. *Probability Theory and Related Fields* **85** 105-117.

[8] K. Dzhaparidze (1977), Tests of composite hypotheses for random variables and stochastic processes. *Theory of Probability and its Applications* **22** 104-118.

[9] F.R. Gantmacher (1954), *The Theory of Matrices.* Chelsea Publishing Company, New York.

[10] W.A.J. Geurts, M.M.A. Hasselaar & J.H. Verhagen (1988), *Large Sample Theory for Statistical Inference in Several Software Reliability Models.* Report MS-R8807, Centre for Mathematics and Computer Science, Amsterdam.

[11] R.D. Gill (1980), *Censoring and Stochastic Integrals.* Tract 124, Centre for Mathematics and Computer Science, Amsterdam.

[12] R.D. Gill (1989) Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part 1). *Scandinavian Journal of Statistics* **16** 97-128.

[13] I.A. Ibragimov & R.Z. Has'minskii (1979), *Statistical Estimation - Asymptotic Theory.* Springer-Verlag, New York.

[14] J. Jacod & A.N. Shiryaev (1988), *Limit Theorems for Stochastic Processes.* Springer-Verlag, New York.

[15] E.V. Khamaladze (1981), Martingale approach in the theory of goodness of fit tests. *Theory of Probability and Applications* **16** 240-257.

[16] P. Kragh, Ø. Borgan, R.D. Gill & N. Keiding (1991), *Statistical Methods Based on Counting Processes.* Spinger-Verlag, New York. To appear.

[17] G. Kulldorff (1957), On the conditions for consistency and asymptotic efficiency of maximum likelihood estimates. *Skandinavisk Aktuarietidskrift* **40** 129-144.

[18] T.G. Kurtz (1983), Gaussian approximations for Markov chains and counting processes. *Bulletin of the International Statistical Institute* **50** 361-375.

[19] B. Littlewood (1980), Theories of software reliability: How good are they and how can they be improved? *IEEE Transactions on Software Engineering* **6** 489-500.

[20] G. Moek (1983), *Software Reliability Models on Trial: Selection, Improved Estimation, and Practical Results.* Report MP 83059 U, National Aerospace Laboratory NLR, Amsterdam.

[21] J.D. Musa, A. Iannino & K. Okumoto (1987), *Software Reliability: Measurement, Prediction, Application.* McGraw-Hill Book Company, New York.

[22] A.W. van der Vaart (1987), *Statistical Estimation in Large Parameter Spaces.* Tract 44, Centre for Mathematics and Computer Science, Amsterdam.