Scand J Statist 19: 235–253, 1992

# Asymptotic Properties of a Class of Statistical Models in Software Reliability

MARK C. VAN PUL

*CWI, Amsterdam*

**ABSTRACT. In software reliability theory many different models have been proposed and investigated. A number of important software reliability models can be formulated in terms of counting processes and 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. A novel aspect of our approach is the way in which we treat asymptotic theory. An application to software reliability theory is presented.**

*Key words:* counting process, intensity function, martingale theory, maximum likelihood estimation (MLE), consistency, asymptotic normality and efficiency, local asymptotic normality (LAN), software reliability models

## 1. Introduction

Computer systems have become more and more important in modern society. The problem of estimating the reliability of computer software has therefore received a great deal of attention over the last two decades. For this purpose a considerable number of models has been proposed. We refer to Musa *et al.* (1987) for a complete overview of the most common software reliability models. 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 the human mind creates imperfect computer programs. 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.

Many important software reliability models can be formulated in terms of counting processes, counting the number of failure occurrences. In this paper we study some asymptotic properties of the maximum likelihood estimation procedure for parametric counting process models. For a general class of counting processes, we derive conditions on the intensity function that are sufficient for these asymptotic properties to hold. We show that the intensity functions of some fairly well-known software reliability models, the models of Jelinski–Moranda and Littlewood, satisfy these conditions. The software models analyzed are perhaps not the most realistic models for computer systems; however, they represent a natural starting point for such a study. A novel aspect of our approach is the fact that—in order to treat asymptotic theory—instead of increasing the time variable or the number of data as is usually the case, we (conceptually) increase one of the model parameters itself. To illustrate the problem and to motivate our concepts, we present here one of the oldest and most elementary software reliability models, that of Jelinski–Moranda (1972), as an example.

*Example 1. The Jelinski–Moranda model.* 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 a failure is produced and no new faults are introduced with probability one.

Let $N$ be 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 $T_i$, $i = 1, 2, \ldots, n(\tau)$, be the failure time of the $i$th occurring failure, while $t_i := T_i - T_{i-1}$, $i = 1, 2, \ldots, n(\tau)$, denotes the interfailure time, that is the time between the $i$th and the $(i-1)$th occurring failure. Finally define $t_{n(\tau)+1} := \tau - T_{n(\tau)}$.

In the Jelinski–Moranda model, introduced in 1972 and generalized a few years later by Musa (1975), the failure rate of the program at any given time is proportional to the number of remaining faults and each fault still present 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:

$$\lambda_i = \phi_0[N_0 - (i-1)], \tag{1.1}$$

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:

$$\lambda^{\mathrm{JM}}(t) = \phi_0[N_0 - n(t-)], \quad t \in [0, \tau], \tag{1.2}$$

where $\lambda(t)$, $t \in [0, \tau]$ denotes the failure rate at time $t$. The interfailure times $t_i$, $i = 1, \ldots, n(\tau)$, are independent and exponentially distributed with parameter $\lambda_i$ given by (1.1).

By using the information obtained from the test experiment, one can estimate the parameters of the underlying model. Maximum likelihood estimation is mostly used for this purpose. Now let us consider how we treat asymptotic behaviour. It does not make sense to let $\tau$, the stopping time, grow to infinity. In the long run the estimate of the total number of faults will trivially be equal to the true number of faults. It makes more sense to (conceptually) increase the number of faults in the program. The idea is that then asymptotics should be relevant to the practical situation in which $N_0$ is large and $n(\tau)/N_0$ not close to zero or one. □

In the next section we give some definitions, notation and background. Here we also state more precisely how asymptotic theory is applied. In section 3 (weak) sufficient conditions are given 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 fourth section we discuss an application of the results in software reliability. Two numerical examples based on both real and simulated data are presented in section 5. Finally, in the sixth section a few remarks are given concerning the possibility of weakening some of the conditions. We mention some results from recent investigations, as well as some plans for the future.

## 2. Some definitions, notation and background

A *counting process* $n(t)$ is a stochastic process that counts the number of certain events (for instance software failures) up to time $t$. Thus $n(t)$ is a non-decreasing integer-valued function of time with jumps of size one only; it is right-continuous and $n(0) = 0$. A *martingale* $m(t)$ is a stochastic process with the property that the increment over a time-interval $(t, t+h]$ given the past has zero expectation. The past here consists solely of the minimal (or self-exciting) history of the counting process $n(t)$. In regular cases, a counting process $n(t)$ is accompanied by an *intensity process* $\lambda(t)$. It is interpreted heuristically as the probability rate that $n(t)$ jumps in a small time interval $[t, t+dt]$ at $t$, conditioned on the past. A more formal definition states that $\lambda(t)$ is the intensity of $n(t)$ if it is predictable (that is non-stochastic given the past) and

$$m(t) := n(t) - \int_0^t \lambda(s)\, ds$$

is a martingale.

Let a counting process $n(t)$ be given. Jumps of the counting process $n(t)$ are observed only during a specific time interval $[0, \tau]$. In this paper we assume that the intensity function associated with the counting process exists and is a member of some specified parametric family, that is:

$$\lambda(t) := \lambda(t; N, \psi), \qquad t \in [0, \tau], \quad N \in \mathbb{N}, \quad \psi \in \Psi, \quad \Psi \subset \mathbb{R}^{p-1}$$

for an integer $p$. Let $N_0$ and $\psi_0$ be the true parameter values. Typically the parameter $N_0$ represents the scale or size of the problem (sometimes $N_0 = n(\infty)$), while $\psi_0$ is a nuisance vector parameter. We are interested in the estimation of $N_0$ and $\psi_0$ as $N_0 \to \infty$. We assume that the model is also meaningful for non-integer $N$. For instance the intensity function (1.2) of the Jelinski–Moranda model can be generalized to

$$\lambda^{\mathrm{JM}}(t) = \phi[N - n(t-)]I\{n(t) < N\}, \quad t \in [0, \tau],$$

where $I\{\cdot\}$ denotes the indicator function. As we are particularly interested in the parameter estimation when $N_0$ is large, we introduce a series of counting processes $n_\nu(t)$, $t \in [0, \tau]$, $\nu = 1, 2, \ldots$ and let $N_0$ conceptually increase. Let $N = N_\nu \to \infty$ for $\nu \to \infty$. By the reparametrisation

$$N_\nu = \nu\gamma_\nu$$

with a dummy variable $\gamma_\nu$, we can denote the associated intensity functions by

$$\lambda_\nu(t; \gamma, \psi) := \lambda(t; \nu\gamma, \psi), \qquad t \in [0, \tau], \quad \gamma \in \mathbb{R}^+, \quad \psi \in \Psi, \quad \nu = 1, 2, \ldots$$

Now we consider the estimation of $\gamma$ and $\psi$ as $\nu \to \infty$. If the real-life situation has $\nu = N_0$, then $\gamma = \gamma_0 = 1$ and $\psi = \psi_0$. It is rather unorthodox to increase a model parameter itself, in this case $N$. This complication is solved by estimating $\gamma$. We assume that the maximum likelihood estimators $(\hat{\gamma}_\nu, \hat{\psi}_\nu)$ for $(\gamma_0, \psi_0)$ exist. Typically, $(\hat{\gamma}_\nu, \hat{\psi}_\nu)$ is a root of the likelihood equations

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

where the likelihood function at time $t$ $L_\nu(\gamma, \psi; t)$ is given by (see Aalen, 1978):

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

We will take care later to rephrase results on the behaviour of $\hat{\gamma}_\nu$ as $\nu \to \infty$ in terms of $\hat{N}$ as $N_0 \to \infty$, where $\hat{N} = \nu\hat{\gamma}_\nu$. By invariance of the maximum likelihood estimation method the value of $\nu$ chosen in actual computations does not influence the value of the result $\hat{N}$. Also, estimated asymptotic variances etc. for $\hat{N}$ depend on $\nu$ and $\hat{\gamma}_\nu$ only through $\nu\hat{\gamma}_\nu$. As we said information obtained for the asymptotic behaviour of $\hat{\gamma}_\nu$ can be transformed back directly to $\hat{N}$, the estimator of main interest. More precisely, consistency of $\hat{\gamma}$ (or $\hat{\gamma}_\nu \to p\ \gamma_0$ as $\nu \to \infty$) implies $\hat{N}/N_0 \to_p 1$ as $N_0 \to \infty$. Similarly concerning the asymptotic normality:

$$\sqrt{\nu}[\hat{\gamma}_\nu - \gamma_0] \xrightarrow{D} \mathcal{N}[0, \sigma^2(\gamma_0, \psi_0)], \nu \to \infty \Rightarrow \sqrt{N_0}\left[\frac{\hat{N}}{N_0} - 1\right] \xrightarrow{D} \mathcal{N}\left[0, \frac{\sigma^2(\gamma_0, \psi_0)}{\gamma_0}\right], \quad N_0 \to \infty. \tag{2.1}$$

Result (2.1) states that $\hat{N}$ is asymptotically normally distributed with mean $N_0$ and variance $N_0\sigma^2(\gamma_0, \psi_0)/\gamma$. One will use this result in practice by estimating the variance as $\hat{N}\sigma^2(\hat{\gamma}, \hat{\psi})/\hat{\gamma}$. This quantity turns out not to depend on $\hat{\gamma}$ (see also remark 3 and (5.1) below). Alternatively, if one uses observed Fisher information, one also gets parametrization-free conclusions, immediately.

Finally, we know from the theory of counting processes that

$$m_\nu(t; \gamma, \psi) := n_\nu(t) - \int_0^t \lambda_\nu(s; \gamma, \psi)\, ds, \quad \nu = 1, 2, \ldots,$$

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

$$x_\nu(t) := \nu^{-1} n_\nu(t), \quad t \in [0, \tau]. \tag{2.2}$$

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

In the next section we give (weak) sufficient conditions for intensity functions $\lambda_\nu$ of a special form under which we have consistency, asymptotic normality and efficiency of the maximum likelihood estimators (MLE) and local asymptotic normality (LAN) of the model.

### 3. Asymptotic properties

We consider a sequence of models $(\lambda_\nu, m_\nu, x_\nu)$, $\nu = 1, 2, \ldots$ as defined 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 we assume that the intensity function $\lambda_\nu$ is of the form:

$$\lambda_\nu(t; \theta) = \nu\beta(t, \theta, x_\nu), \tag{3.1}$$

where $\beta := [0, \tau] \times \Theta \times K \to \mathbb{R}^+$ is an arbitrary non-negative and non-anticipating function. Non-anticipating means that $\beta(t, \theta, x_\nu)$ only depends on $x_\nu|_{[0,t)}$, the past of the stochastic process $x_\nu$ up to but not including time $t$. In fact, in all practical cases $\beta(t, \theta, x_\nu)$ will depend only on $x_\nu(t-)$. On $K := D([0, \tau])$, the space of right-continuous functions on $[0, \tau]$ with left limits (so-called cadlag functions), we put the usual supremum norm. The likelihood function $L_\nu(\theta, t)$ now becomes for $\theta \in \Theta$, $t \in [0, \tau]$ and $\nu = 1, 2, \ldots$:

$$L_\nu(\theta, t) := \exp\left[\int_0^t \log \nu\beta(s, \theta, x_\nu)\, dn_\nu(s) - \nu \int_0^t \beta(s, \theta, x_\nu)\, ds\right]. \tag{3.2}$$

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

$$C_\nu(\theta, t) := \log L_\nu(\theta, t), \tag{3.3}$$

$$U_{\nu i}(\theta, t) := \frac{\partial}{\partial\theta_i} C_\nu(\theta, t), \quad \text{(score function)} \tag{3.4}$$

$$I_{\nu ij}(\theta, t) := \frac{\partial^2}{\partial\theta_i\, \partial\theta_j} C_\nu(\theta, t), \quad \text{(minus observed information matrix)} \tag{3.5}$$

$$R_{\nu ijk}(\theta, t) := \frac{\partial^3}{\partial\theta_i\, \partial\theta_j\, \partial\theta_k} C_\nu(\theta, t). \tag{3.6}$$

Consider the following global conditions:

(G1) For all $x \in K$ and for all $\theta \in \Theta$ the intensity function $\beta$ satisfies:

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

(G2) (Lipschitz continuity) There exists a constant $L$, not depending on $t$, such that for all $x, y \in K$ and all $t \in [0, \tau]$:

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

Under the global conditions (G1)–(G2) the stochastic process $x_\nu(t)$, as defined in (2.10), 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

$$x(t) = \int_0^t \beta(s, \theta_0, x)\, ds.$$

This was proved by Kurtz (1983).

Moreover, we consider the following local conditions:

(L1) 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 $[0, \tau] \times \Theta_0 \times K_0$.

(L2) The function $\beta(t, \theta, x)$ is bounded away from zero on $[0, \tau] \times \Theta_0 \times K_0$.

(L3) The matrix $\Sigma = \{\sigma_{ij}(\theta_0)\}$ is positive definite, with for $i, j \in \{1, 2, \ldots, p\}, \theta \in \Theta_0$:

$$\sigma_{ij}(\theta) = \int_0^\tau \frac{\frac{\partial}{\partial \theta_i} \beta(s, \theta, x_0) \frac{\partial}{\partial \theta_j} \beta(s, \theta, x_0)}{\beta(s, \theta, x_0)}\, ds. \tag{3.7}$$

We are now able to formulate the main result of this paper.

**Theorem 1**

*Consider a counting process with intensity function $\lambda(t; N, \psi)$, where $(N, \psi)$ denotes an unknown $p$-dimensional parameter. As in section 2 we can define an associated sequence of experiments by letting $\nu \to \infty$. Let $\theta_0 = (\gamma_0, \psi_0)$ be 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 (3.1) for a certain function $\beta$ satisfying conditions (G1)–(G2) and (L1)–(L3). Then we have:*

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

$$\frac{\partial}{\partial \theta} \log L_\nu(\theta, \tau) = 0, \quad \nu = 1, 2, \ldots \tag{3.8}$$

*have exactly one consistent solution $\hat{\theta}_\nu$. Moreover this solution provides a local maximum of the likelihood function (3.2).*

(*ii*) *Asymptotic normality of the ML-estimators: let $\hat{\theta}_\nu$ be the consistent solution of the maximum likelihood equations (3.8), then*

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

*where $\Sigma$ is given by (3.7) and can be estimated consistently from the observed information matrix $I_\nu$, given in (3.5).*

(*iii*) *Local asymptotic normality of the model: with $U_\nu$ given by (3.4), we have for all $h \in \mathbb{R}^p$:*

$$\log \frac{dP_{\theta_\nu}}{dP_{\theta_0}} - \nu^{-1/2} h^T U_\nu + \tfrac{1}{2} h^T \Sigma h \xrightarrow{P_{\theta_0}} 0, \quad \nu \to \infty, \tag{3.9}$$

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

(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:*

$$\sqrt{\nu}(\tilde{\theta}_\nu - \theta_0) \xrightarrow{D(\theta_0)} Z + Y,$$

*where* $Z \sim_d \mathcal{N}(0, \Sigma^{-1})$, *Z and Y independent. (For a definition of the regularity of an estimator we refer to van der Vaart (1988) or to (3.14) below.)*

*Remark 1.* An immediate consequence of this result about the asymptotic distribution of the ML-estimator $\hat{\theta}_\nu$ is the fact that the Wald test statistic

$$-(\hat{\theta}_\nu - \theta_0)^T I_\nu(\hat{\theta}_\nu, \tau)(\hat{\theta}_\nu - \theta_0),$$

where $I_\nu$ is given by (3.5), is asymptotically chi-squared distributed with $p$ degrees of freedom under the simple hypothesis $H_0: \theta = \theta_0$. With $C_\nu$, $U_\nu$ and $I_\nu$ given by (3.3)–(3.5) the Rao test (or score) statistic

$$-U_\nu(\theta_0, \tau)^T I_\nu(\theta_0, \tau)^{-1} U_\nu(\theta_0, \tau)$$

and the Wilks test (or likelihood ratio) statistic

$$2[C_\nu(\hat{\theta}_\nu, \tau) - C_\nu(\theta_0, \tau)] \tag{3.10}$$

have the same asymptotic distribution as the Wald test statistic. Equivalence of these tests can be shown by the arguments of Rao (1973). □

*Proof of theorem 1.* One can easily check that (G1)–(G2) and (L1)–(L3) imply the following more classical looking set of conditions (C1)–(C3) (see Borgan, 1984):

(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 $x_\nu$, $\nu = 1, 2, \ldots$ the stochastic process given in (2.2), $x_0 \in K$ has to satisfy for all $i, j \in \{1, 2, \ldots, p\}$ as $\nu \to \infty$:

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

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

(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 $x \in K_0$:

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

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

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

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

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

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

Although our model (3.1) is not a special case of the multiplicative intensity model considered in Borgan (1984), the rest of the proof of (i) and (ii) of theorem 1, which is given in van Pul (1990), now follows exactly the lines of Borgan (1984) and is omitted here. Borgan starts with conditions of the type (C1)–(C3) and uses the same standard argumentation as given by Cramér (1946), who derived similar results for the classical case of i.i.d. random variables. Compared with the i.i.d. case the difference is that in the present context Lenglart's inequality is used to establish the convergence in probability results (instead of the law of large numbers in the classical case), while we have to use the martingale central limit theorem to establish the weak convergence result, which in the classical case is proved by the central limit theorem for i.i.d. random variables.

We will now give proofs of part (iii) and (iv) of theorem 1. It should be noted that Hjort in the discussion of the lecture of Andersen & Borgan (1985) already pointed out that Local Asymptotic Normality in Borgan's model could easily have been shown by him. In Hjort (1986) LAN is proved for the multiplicative model.

*(iii) Local asymptotic normality of the model.* For sake of convenience, we introduce some more notation. For a function $f$: $\mathbb{R}^p \to \mathbb{R}$, which is at least three times differentiable, we write:

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

the (three-dimensional) $p \times p \times p$ matrix of third order 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:

$$g^T Y g^{\langle 2 \rangle} := \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p g_i g_j g_k y_{ijk}.$$

We define for $h \in \Theta$:

$$\theta_\nu(h) := \theta_0 + \nu^{-1/2} h, \quad \nu = 1, 2, \ldots .$$

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:

$$\begin{aligned} Q_\nu(h) &= \log \frac{dP_{\theta,(h)}}{dP_{\theta_0}} \\ &= \log dP_{\theta,(h)} - \log dP_{\theta_0} \\ &= \left[\int_0^\tau \log \lambda_\nu(s, \theta_\nu(h))\, dn_\nu(s) - \int_0^\tau \lambda_\nu(s, \theta_\nu(h))\, ds\right] \\ &\quad - \left[\int_0^\tau \log \lambda_\nu(s, \theta_0)\, dn_\nu(s) - \int_0^\tau \lambda_\nu(s, \theta_0)\, ds\right] \\ &= C_\nu(\theta_\nu(h), \tau) - C_\nu(\theta_0, \tau), \end{aligned}$$

where $C_v$ is given by (3.3). Of course $Q_v(0) = 0$, because

$$\frac{\partial}{\partial h}\theta_v(h) = v^{-1/2},$$

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

$$\frac{\partial}{\partial h}Q_v(h) = v^{-1/2}U_v(\theta_v(h), \tau),$$

$$\frac{\partial^2}{\partial h^2}Q_v(h) = v^{-1}I_v(\theta_v(h), \tau),$$

$$\frac{\partial^3}{\partial h^3}Q_v(h) = v^{-3/2}R_v(\theta_v(h), \tau),$$

where $U_v$, $I_v$ and $R_v$ are given by (3.4)–(3.6). Hence we get the Taylor expansion:

$$Q_v(h) = v^{-1/2}h^TU_v(\theta_0, \tau) + \tfrac{1}{2}v^{-1}h^TI_v(\theta_0, \tau)h + \tfrac{1}{6}v^{-3/2}h^TR_v(\theta_v^*, \tau)h^{\langle 2\rangle},$$

where $\theta_v^*$ is somewhere on the linesegment between $\theta_0$ and $\theta_v(h)$. In the proofs of consistency and asymptotic normality (see van Pul, 1990) it is deduced that:

$$v^{-1/2}U_v(\theta_0, \tau) \to_d \mathcal{N}(0, \Sigma),$$

$$v^{-1}I_v(\theta_0) \to_p -\Sigma,$$

$$v^{-3/2}R_v(\theta_v^*, \tau) \to_p 0,$$

as $v \to \infty$, for all sequences $(\theta_v^*)$ converging to $\theta_0$. Hence, this yields us exactly the local asymptotic normality (LAN) property (3.9). This proves part (iii) of theorem 1.

(*iv*) *Asymptotic efficiency of the ML-estimators.* In the proof of asymptotic normality of $\hat{\theta}_v$ (van Pul, 1990), it is derived that:

$$v^{-1/2}U_v \to_{D(\theta_v)} \mathcal{N}(0, \Sigma), \tag{3.11}$$

$$\sqrt{v}(\hat{\theta}_v - \theta_0) - \Sigma^{-1}v^{-1/2}U_v + O_P(1) \to_{D(\theta_0)} \mathcal{N}(0, \Sigma^{-1}). \tag{3.12}$$

Moreover, the LAN-property of the model, proved in (iii), gives us by using (3.11):

$$\log\frac{dP_{\theta_v}}{dP_{\theta_0}} = v^{-1/2}h^TU_v - \tfrac{1}{2}h^T\Sigma h + O_P(1) \to_{D(\theta_0)} \mathcal{N}(-\tfrac{1}{2}h^T\Sigma h, h^T\Sigma h).$$

Hence:

$$\begin{bmatrix} \sqrt{v}(\hat{\theta}_v - \theta_0) \\ \log\dfrac{dP_{\theta_v}}{dP_{\theta_0}} \end{bmatrix} \to_{D(\theta_0)} \mathcal{N}\left(\begin{bmatrix} 0 \\ -\tfrac{1}{2}h^T\Sigma h \end{bmatrix}, \begin{bmatrix} \Sigma^{-1} & h \\ h^T & h^T\Sigma h \end{bmatrix}\right).$$

From Le Cam's third lemma (see van der Vaart, 1987, pp. 180–181), we can now conclude the contiguity of $P_{\theta_v}$ and $P_{\theta_0}$ and we have:

$$\sqrt{v}(\hat{\theta}_v - \theta_0) \to_{D(\theta_v)} \mathcal{N}(h, \Sigma^{-1})$$

and thus

$$\sqrt{v}(\hat{\theta}_v - \theta_v) = \sqrt{v}(\hat{\theta}_v - (\theta_0 + v^{-1/2}h)) \to_{D(\theta_v)} \mathcal{N}(0, \Sigma^{-1}). \tag{3.13}$$

Combining (3.12) and (3.13), we see that for all $h \in \Theta$:

$$\lim_{\nu \to \infty} \mathscr{L}_{\theta_\nu}[\sqrt{\nu}(\hat{\theta}_\nu - \theta_\nu)] = \lim_{\nu \to \infty} \mathscr{L}_{\theta_0}[\sqrt{\nu}(\hat{\theta}_\nu - \theta_0)];$$

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 van der Vaart, 1987), which states in our case that the limit-distribution of any regular estimator $\tilde{\theta}_\nu$ of $\theta_0$ satisfies:

$$\lim_{\nu \to \infty} \mathscr{L}_{\theta_0}[\sqrt{\nu}(\tilde{\theta}_\nu - \theta_0)] = \mathscr{N}(0, \Sigma^{-1}) * \mathscr{M}_{\theta_0}. \tag{3.14}$$

Because (3.12) implies that for $\tilde{\theta}_\nu := \hat{\theta}_\nu$ we get $\mathscr{M}_{\theta_0} \equiv 0$ in (3.14), we have proved that the maximum likelihood estimator $\hat{\theta}_\nu$ is asymptotically efficient. This proves part (iv) and hence completes the proof of theorem 1. □

## 4. An application to software reliability theory

Several statistical models have been proposed in order to estimate the evolution in reliability of computer software during the debugging phase. In the introduction we introduced the Jelinski–Moranda model. In this section we will present another well-known model in the theory of software reliability, namely the Littlewood model (1980), and we will discuss a generalisation of this model. Other well-known software reliability models that fit in our framework are the model of Goel & Okumoto (1980) and the Poisson–Gamma model discussed in Koch & Sprey (1983). For backgrounds and notation we refer to example 1 in the introduction of this paper.

*Example 2. The Littlewood model.* Recall that the failure intensity in the Jelinski–Moranda model is given by:

$$\lambda^{JM}(t) = \phi_0[N_0 - n(t-)], \quad t \in [0, \tau]. \tag{4.1}$$

Also in the model, introduced by Littlewood (1980), it is assumed that at any time the failure rate is proportional to the number of remaining errors. The main difference in the Littlewood model with respect to the Jelinski–Moranda 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:

$$\phi_j \sim \Gamma(a_0, b_0), \quad j = 1, \ldots, N.$$

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

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

with

$$\phi_j \sim \Gamma(a_0, b_0),$$

$$T_j \mid \phi_j = \phi \sim \exp(\phi),$$

a simple calculation yields:

$$\phi_j \mid T_j > t \sim \Gamma(a_0, b_0 + t)$$

and hence:

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

An application of the so called innovation theorem (Aalen, 1978) now shows that the failure intensity of the software at time $t$ is given by:

$$\lambda^{L}(t) = \frac{a_0[N_0 - n(t-)]}{b_0 + t} \tag{4.2}$$

By a simple reparametrization, namely:

$$\rho_0 = \frac{1}{a_0}, \qquad \mu_0 = \frac{b_0}{a_0},$$

we get from (4.2):

$$\lambda^{GL}(t) = \frac{N_0 - n(t-)}{\mu_0 + \rho_0 t}, \quad t \in [0, \tau]. \tag{4.3}$$

Actually this provides an extension of the Littlewood model, allowing also small values of $\rho_0 \leqslant 0$. Note that when $\rho_0 = 0$ we are dealing with the model we discussed earlier, namely the Jelinski–Moranda model. We can therefore treat the Jelinski–Moranda model as another special (limit) case of the Littlewood model. Note that both the Littlewood and the Jelinski–Moranda model have as a special case the Poisson model (with constant failure intensity); this is the limit case letting $N_0 \to \infty$ and $\mu_0 \to \infty$ in (4.3) such that $N_0/\mu_0$ is a constant. □

*Remark 2.* The models studied by Aalen (1980) are of the form

$$\lambda_j(t) = \sum_{i=1}^{p} \alpha_i Y_{ij}(t), \quad j = 1 \dots n,$$

where the $\alpha_i(t)$ are non-parametric functions of $t$ and the $Y_{ij}(t)$ are arbitrary observable processes. It should be noticed that (4.3) can be written as:

$$\lambda^{GL}(t) = \frac{N_0}{\mu_0 + \rho_0 t} 1 + \frac{-1}{\mu_0 + \rho_0 t} n(t-), \quad t \in [0, \tau],$$

where 1 and $n(t-)$ are indeed observable processes. In contrast to the models of Aalen, in software reliability models the coefficients $\alpha_i(t)$ are parametric functions of time (and sometimes even constant in time). Additionally, in software reliability models typically $n = 1$, while Aalen's $n$ is large. This calls for models and methods different from the typical Aalen type ones. □

Let us now consider how we apply asymptotic theory to the generalized model given by (4.3). Letting $N = \nu\gamma$ conceptually increase as described in section 2, we see that the corresponding sequence of intensity functions can be written in the standard form (3.1):

$$\lambda_{\nu}^{GL}(t, \theta) = \nu\beta^{GL}(t, \theta, x_{\nu}),$$

where $\theta = (\gamma, \mu, \rho)$, $x_{\nu}$ is given by (2.2) and

$$\beta^{GL}(t; \gamma, \mu, \rho; x) = \frac{\gamma - x(t-)}{\mu + \rho t}, \tag{4.4}$$

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

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

As an application we will show that the results of theorem 1 hold for the Jelinski–Moranda and the Littlewood model.

**Theorem 2**

*Let $\tau > 0$. We assume that the failure data are generated by the intensity function $\beta$ in (4.4) with true parameter value $\theta_0 := (\gamma_0, \mu_0, \rho_0)$ satisfying $\gamma_0 > 0$ and $\mu_0 + \rho_0 t > 0$, $t \in [0, \tau]$. If we define for $t \in [0, \tau]$*

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

*then $\beta$ satisfies conditions $(G1)$–$(G2)$ & $(L1)$–$(L3)$ and hence:*

(*i*) *With 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.*

Full details of the verification of (G1)–(G2) and (L1)–(L3), which is technical but routine, except perhaps for (L3), is given in van Pul (1990) and therefore omitted here. It is shown there that the following choice of $\Theta_0$ and $K_0$ will be appropriate:

$$\Theta_0 := [\varepsilon_\gamma, M_\gamma] \times [\varepsilon_\mu, M_\mu] \times [\varepsilon_\rho, M_\rho],$$

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

where

$$0 < \tfrac{1}{2}[\gamma_0 + x_0(\tau)] < \varepsilon_\gamma < \gamma_0 < M_\gamma,$$

$$0 < \varepsilon_\mu < \mu_0 < M_\mu,$$

$$-\frac{\mu_0}{\tau} < \varepsilon_\rho < \rho_0 < M_\rho \quad \text{with} \quad \varepsilon_\rho < 0 < M_\rho$$

and

$$0 < \varepsilon_x < \tfrac{1}{2}[\gamma_0 - x_0(\tau)].$$

*Verification of (L3).* To verify (L3) one may check for instance that $\det \Sigma \neq 0$; this is, however, extremely tedious. A much simpler approach is to note that $\Sigma$ is the covariance matrix of

$$d_1 := \int_0^\tau \frac{1}{N - n(t-)}\, dM(t)$$

$$d_2 := \int_0^\tau \frac{1}{\mu + \rho t}\, dM(t)$$

and

$$d_3 := \int_0^\tau \frac{t}{\mu + \rho t}\, dM(t).$$

Therefore, we have $\det \Sigma = 0$ if and only if there exist coefficients $a$, $b$ and $c$ (not all equal to zero), such that

$$D := ad_1 + bd_2 + cd_3 \tag{4.5}$$

is constant. Now we consider the following two cases. Firstly, with positive probability $n(t)$ makes no jump at all in $[0, \tau]$. Secondly, also with positive probability, the process $n(t)$ makes exactly one jump in $[0, \tau]$. Suppose this jump is at time $T$. One can easily check that in the case of no jump (4.5) is given by

$$D_0 = N \int_0^\tau \frac{b + ct}{(\mu + \rho t)^2} dt - \frac{a}{\rho} \log\left(1 + \frac{\rho}{\mu}\tau\right)$$

and in the case of exactly one jump at $T$

$$D_1(T) = \left[\frac{a}{N} + N \int_0^\tau \frac{b + ct}{(\mu + \rho t)^2} dt - \frac{a}{\rho} \log\left(1 + \frac{\rho}{\mu}\tau\right)\right] + \frac{b + cT}{\mu + \rho T} - \int_T^\tau \frac{b + ct}{(\mu + \rho t)^2} dt$$

Obviously, $D_1$ is a non-constant function of $T$, when $b$ and $c$ are not both equal to zero. But given $b = 0$, $c = 0$ we see that the constants $D_0$ and $D_1$ are different, except for the degenerate case that also $a = 0$. We have hence proved that there do not exist coefficients $a$, $b$ and $c$ (not all three equal to zero), such that $D$ is constant. This yields the non-singularity of $\Sigma$. □

*Remark 3.* The software reliability models, where $\theta = (\gamma, \psi)$ and $N = v\gamma$ is a parameter of interest, typically satisfy the rescaling condition

$$\beta(t; c\gamma, c\psi; cx) = c\beta(t; \gamma, \psi; x), \tag{4.29}$$

for all $c > 0$. It is easy to check that the asymptotic variance of $\hat{\psi}$ now does not depend on $\gamma$, while that of $\hat{\gamma}$ is proportional to $\gamma$. This guarantees that asymptotic confidence intervals for $N$ and $\psi$ do not depend on the (arbitrary) choice of $v$. □

## 5. Some numerical results

We have recently begun a study of the behaviour of the ML-estimators in practice, computed from both real data and from simulated data generated by the Jelinski–Moranda model. The simulation results (van Pul, 1991a), 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 Wilks likelihood ratio test statistic (3.10), however, we were able to build confidence intervals for the model parameters that are much more satisfactory than intervals based on the approximate normal test statistic.

In this section we will discuss two numerical examples. In example 3, which deals with some real data collected by Moek (1983), the Jelinski–Moranda model and the Littlewood model are compared. In example 4, data are simulated according to the Jelinski–Moranda model. The theoretical asymptotic normality is studied and coverage percentages of confidence intervals based on the asymptotic normal and on the WLRT statistic are compared. More background, calculations and detailed results, both on Moek's data and on the simulated data, can be found in van Pul (1991a).

*Example 3. A case study.* The models of Jelinski–Moranda and Littlewood have been applied to real data from Project A, concerning an information system for registering aircraft movements. For more details see Moek (1983, 1984). Failure data collected during the testing stage (in the operational environment) are given in Table 1. Note that there is a misprint in $T_{37}$ of the original data in Moek (1983). Furthermore, $T_{44}$ does not represent a failure time, but is assumed to be the stopping time of the testing process. Figure 1 gives the counting process associated with the data of Table 1.

We calculated maximum likelihood estimators for the model parameters of the models of Jelinski–Moranda (JM), Littlewood (L) and Generalized Littlewood (GL). Their intensity

Table 1. *Failure times for Moeks project A (CPU time in msec)*

| $i$ | $t_i$ | $T_i$ | $i$ | $t_i$ | $T_i$ |
|---|---|---|---|---|---|
| 1 | 0.00088 | 0.00088 | 23 | 0.00445 | 0.13321 |
| 2 | 0.00343 | 0.00431 | 24 | 0.00486 | 0.13807 |
| 3 | 0.00286 | 0.00717 | 25 | 0.00064 | 0.13871 |
| 4 | 0.01176 | 0.01893 | 26 | 0.00399 | 0.14270 |
| 5 | 0.00475 | 0.02368 | 27 | 0.02684 | 0.16954 |
| 6 | 0.00024 | 0.02392 | 28 | 0.00227 | 0.17181 |
| 7 | 0.00230 | 0.02622 | 29 | 0.00020 | 0.17201 |
| 8 | 0.00857 | 0.03479 | 30 | 0.03918 | 0.21119 |
| 9 | 0.00462 | 0.03941 | 31 | 0.01491 | 0.22610 |
| 10 | 0.00106 | 0.04047 | 32 | 0.01467 | 0.24077 |
| 11 | 0.00382 | 0.04429 | 33 | 0.01631 | 0.25708 |
| 12 | 0.01480 | 0.05909 | 34 | 0.03841 | 0.29549 |
| 13 | 0.00177 | 0.06086 | 35 | 0.00112 | 0.29661 |
| 14 | 0.02427 | 0.08513 | 36 | 0.03056 | 0.32717 |
| 15 | 0.00480 | 0.08993 | 37 | 0.00621 | 0.33338 |
| 16 | 0.00047 | 0.09040 | 38 | 0.00012 | 0.33350 |
| 17 | 0.00004 | 0.09044 | 39 | 0.02021 | 0.35371 |
| 18 | 0.01017 | 0.10061 | 40 | 0.02640 | 0.38011 |
| 19 | 0.00112 | 0.10173 | 41 | 0.03780 | 0.41791 |
| 20 | 0.00098 | 0.10271 | 42 | 0.07422 | 0.49213 |
| 21 | 0.02430 | 0.12701 | 43 | 0.08444 | 0.57657 |
| 22 | 0.00175 | 0.12876 | 44 | 0.02343 | 0.60000 |

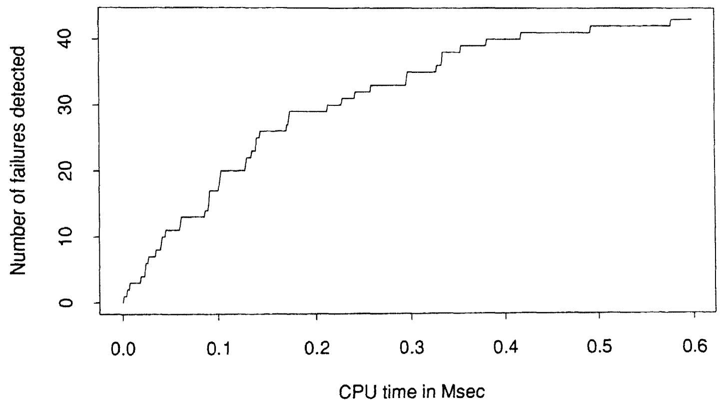


*Fig. 1.* Counting process belonging to failure data of project A.

functions are given by respectively (4.1), (4.2) and (4.3). To determine MLE's for the (three-parameter) models L and GL we used a standard optimization program, written in Pascal. This program, called *Amoeba* and described in Vetterling *et al.* (1985), carries out a down-hill simplex method. The results are given in Table 2.

We find that conditioned on $\rho \geqslant 0$, the log likelihood function of the Littlewood model is maximal for $\rho = 0$. In this case the Littlewood model reduces to the Jelinski–Moranda model with $\phi = 1/\mu = 5.5463$. In the generalized Littlewood model (4.3), also allowing small negative values for $\rho$, the log likelihood function is maximized for $N = 43.0000$ and hence $\lambda = 0$. This seems not to make much sense. It should be noticed that the number of bugs is too small to make accurate predictions. This will be pointed out in the next example. The

Table 2. *Comparison of maximum likelihood estimators for the Jelinski-Moranda (JM), Littlewood (L) and Generalized Littlewood (GL) model with use of data from Table 1*

| max log $L_\tau$ | (JM) 156.2290 | (L) 156.2298 | (GL) 156.8618 |
|---|---|---|---|
| $\hat{N}$ | 44.0734 | 44.0742 | 43.0000 |
| $\hat{\phi}$ | 5.5465 | 5.5463 | – |
| $\hat{a}$ | – | $\infty$ | −3.9246 |
| $\hat{b}$ | – | $\infty$ | −0.8191 |
| $\hat{\mu}$ | – | 0.1803 | 0.2087 |
| $\hat{\rho}$ | – | 0.0000 | −0.2548 |
| $\hat{\lambda}(\tau)$ | 5.9536 | 5.9578 | 0.0000 |

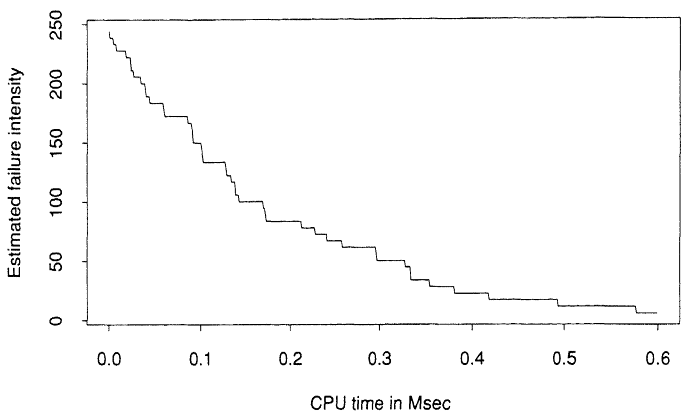


*Fig. 2.* Estimated failure intensity of Jelinski–Moranda and Littlewood model.

estimated failure intensity of the Jelinski–Moranda and Littlewood model ($\rho \geqslant 0$) are identical and given in Fig. 2.

More results on this case study, including standard deviations etc. can be found in Andersen *et al.* (1991). □

*Example 4. Simulation of the Jelinski–Moranda model.* In our simulation experiments we generated failure times according to the Jelinski–Moranda model with $\phi_0 = 1$, $\tau = 1$ and various values for $N_0$. From the asymptotic theory developed in sections 3 and 4, it follows, that we can define centered and normalized quantities $X$ and $Y$, satisfying:

$$\hat{X} := \frac{\hat{N} - N_0}{\sqrt{N_0}} \xrightarrow{D} N\left[0, \frac{1 - e^{-\phi_0\tau}}{e^{\phi_0\tau} + e^{-\phi_0\tau} - \tau^2\phi_0^2 - 2}\right], \tag{5.1}$$

$$\hat{Y} := \sqrt{N_0}(\hat{\phi} - \phi_0) \xrightarrow{D} N\left[0, \frac{\phi_0^2(e^{\phi_0\tau} - 1)}{e^{\phi_0\tau} + e^{-\phi_0\tau} - \tau^2\phi_0^2 - 2}\right], \tag{5.2}$$

as $N_0 \to \infty$. We generated and studied sets of 10,000 replicates of $\hat{X}$ and $\hat{Y}$. The figures for $N_0 = 50$, 500 and 5000 are given in Table 3.

We see that the convergence of $\hat{X}$ and $\hat{Y}$ to normal distributions with mean zero and asymptotic variances as expected (in (5.1) and (5.2)) is rather slow. The difference in the

Table 3. *Means, variances and skewness coefficients of simulated approximate normal quantities X and Y. Number of replicates K = 10,000,* $\phi = 1$, $\tau = 1$

| $N_0$ | Mean $\hat{X}$ | Var $\hat{X}$ | Skew $\hat{X}$ | Mean $\hat{Y}$ | Var $\hat{Y}$ | Skew $\hat{Y}$ |
|---|---|---|---|---|---|---|
| 50 | −0.2270 | 4.4943 | 1.6352 | 2.1887 | 15.1202 | 0.6342 |
| 500 | 0.2884 | 10.0716 | 0.9480 | 0.5038 | 20.1352 | 0.0351 |
| 5,000 | 0.1145 | 7.4131 | 0.3837 | 0.1071 | 19:7273 | 0.0178 |
| $\infty$ | 0.0000 | 7.3365 | 0.0000 | 0.0000 | 19.9423 | 0.0000 |

asymptotic behaviour of $N$ and $\phi$ is illustrated by the histograms and qq plots given in Figs 3 and 4. Both tables and figures give the same impression, namely that the distribution of $\hat{N}$ shows a severe skewness and that the distribution of $\hat{\phi}$ is rather biased for small $N_0$. Both defects slowly disappear as $N_0$ increases.

As the distribution of $\hat{N}$ is skew, the coverage percentages of confidence intervals based on the asymptotic normal statistic could be expected to be disappointing. In Table 4 we compare these percentages with those based on the Wilks test statistic (3.10).

As the Wilks confidence intervals are larger, shifted to the right (and hence not symmetric around $N_0$) in comparison with the approximate normal confidence intervals, for high levels of confidence the Wilks intervals are significantly better (and have coverage probabilities that are less skew) than the approximate normal ones (see van Pul, 1991a). □

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

As stated in Remark 2, theorems 1 and 2 remain valid if we replace (G1)–(G2) and (L1)–(L3) by the weaker set of conditions (C1)–(C4). Conditions comparable to these ones are also given by Cramér (1946) and Kulldorff (1957), using classical statistical techniques to prove consistency and asymptotic normality of maximum likelihood estimators. Nowadays modern methods have been developed by Ibragimov & Has'minskii (1979), Jacod & Shiryaev (1988), Dzhaparidze & Valkeila (1988) and Le Cam & Yang (1990) among others leading to the same results (and even more), without requiring the existence of higher derivatives of the intensity function and so weakening condition (L1) (and (C2) considerably. Ibragimov & Has'minskii (1979) consider the parametric case, but no theory for counting processes is developed, while Jacod & Shiryaev (1988) and Dzhaparidze & Valkeila (1988) study only binary experiments for counting processes. Also the work of Gill (1980) and van der Vaart (1987) 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, intensity functions tend to be very smooth and determining the existence of derivatives (with respect to $\theta$) is relatively easy.

Moreover, we think we can improve the construction of confidence intervals by making use of parametric bootstrap methods. The validity of the parametric bootstrap method will follow by standard arguments on contiguity, regular estimators and the Skorohod–Dudley–Wichura almost sure representation theorem (see Gill, 1989). Asymptotic consistency of the parametric bootstrap is proved in van Pul (1991c) and some numerical results in software reliability are presented.

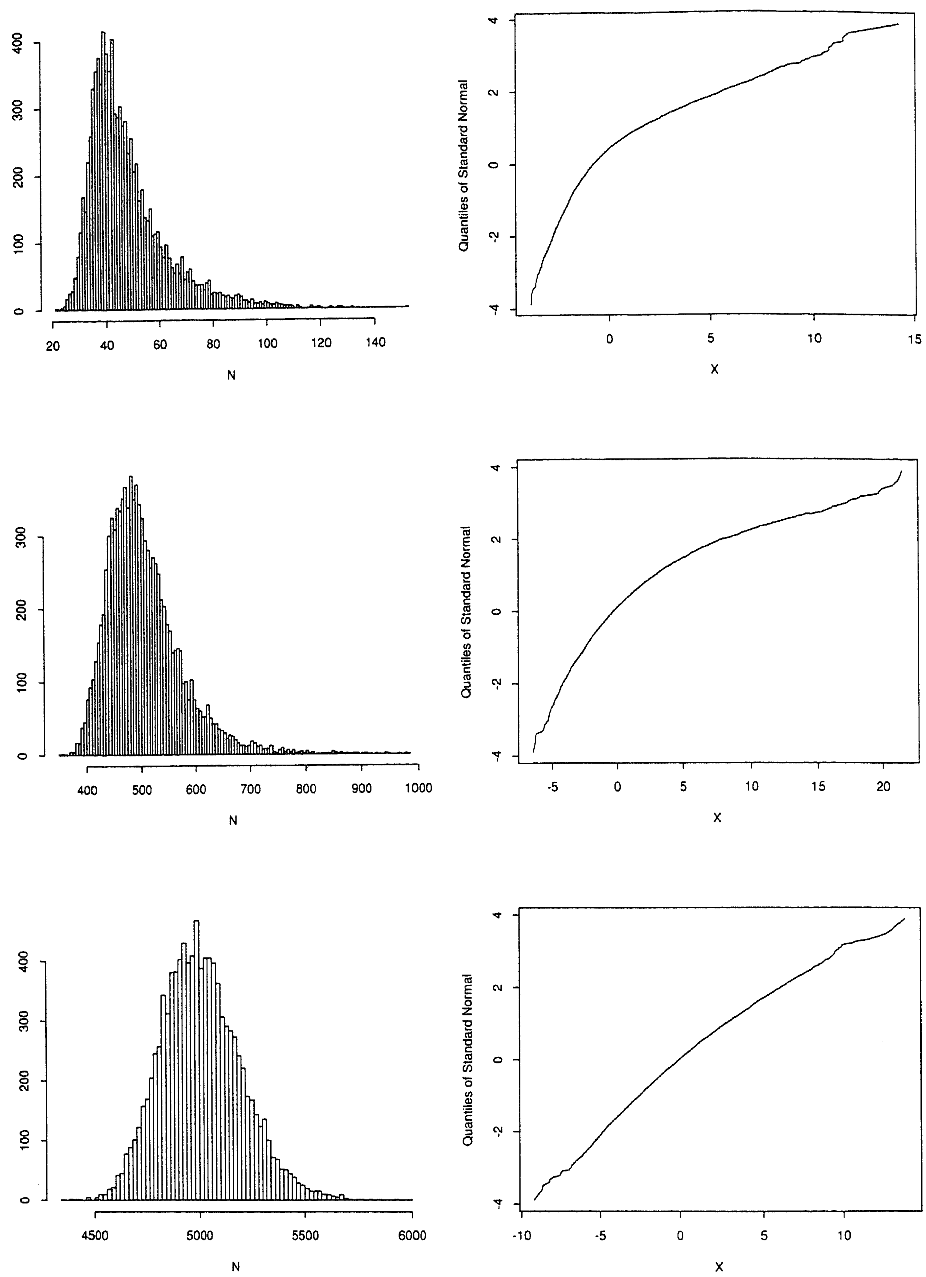


*Fig. 3.* Histogram of $\hat{N}$ and qq plot of $\hat{X}$: (a) $N_0 = 50$ (b) $N_0 = 500$ (c) $N_0 = 5000$.

Furthermore, another topic of future investigation will be the study of goodness of fit tests. We intend to follow the Martingale approach of Khmaladze (1981). See also Geurts *et al.* (1988) and Hjort (1990).

Of course, our ultimate goal will be to study more realistic models, incorporating imperfect repair and software growth and taking account of covariate measurements. We have recently

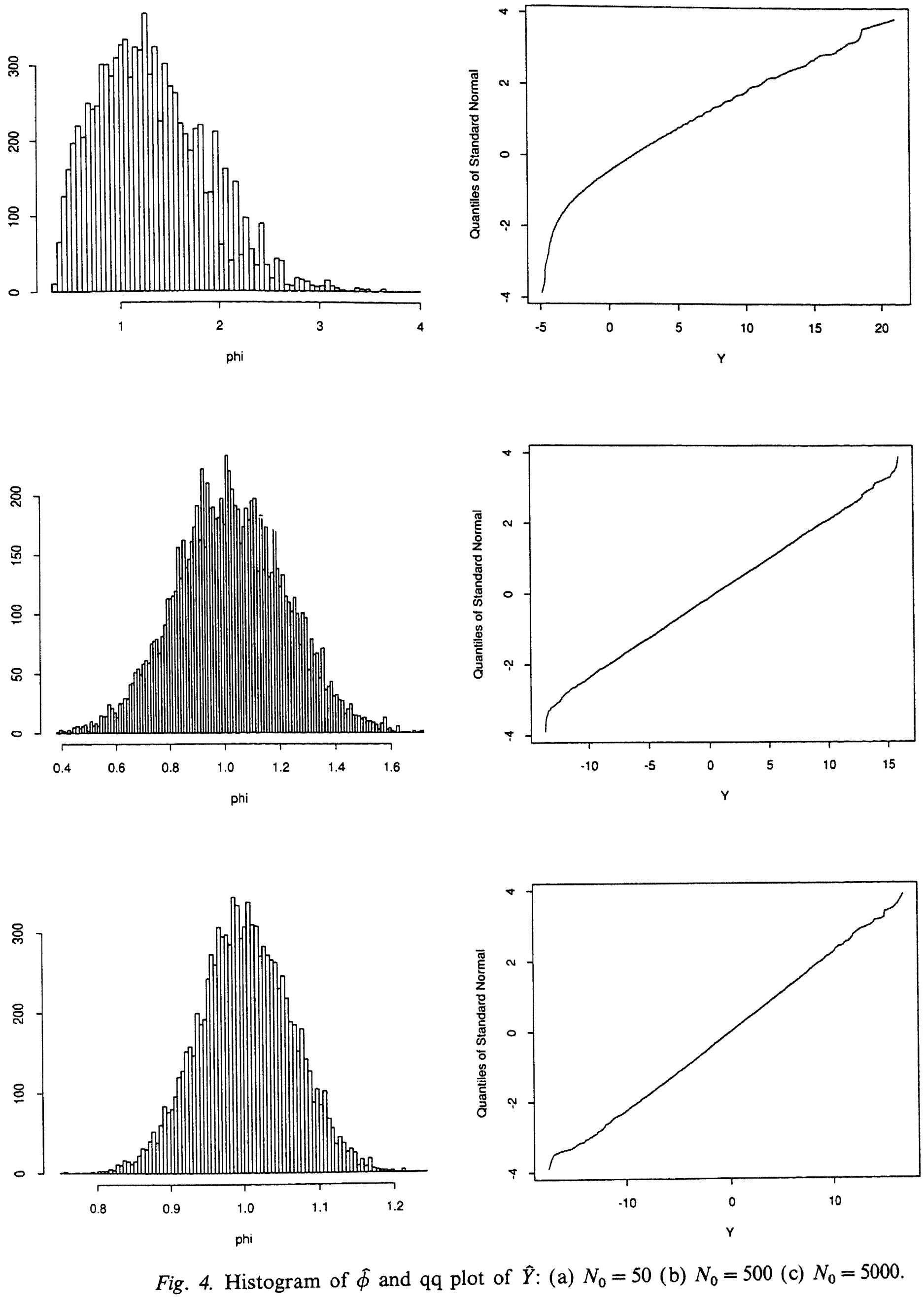


*Fig. 4.* Histogram of $\hat{\phi}$ and qq plot of $\hat{Y}$: (a) $N_0 = 50$ (b) $N_0 = 500$ (c) $N_0 = 5000$.

constructed such a model (van Pul, 1991b) and are now investigating whether this model fits in the theory developed so far.

Finally, we note that theorem 1 does not 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 (1983) developed an easy criterion, satisfied with probability tending to one, for the existence of a unique solution of the ML

Table 4. *Coverage probabilities of the two-sided approximate normal and the two-sided Wilks confidence intervals based on simulated data of Table 3. Number of replicates $K = 10{,}000$, $\phi = 1$, $\tau = 1$*

| | Approximate normal | | | Wilks | | |
|---|---|---|---|---|---|---|
| $\alpha$ | $N_0 = 50$ | $N_0 = 500$ | $N_0 = 5000$ | $N_0 = 50$ | $N_0 = 500$ | $N_0 = 5000$ |
| 50 | 55.58 | 50.90 | 50.43 | 52.56 | 50.43 | 50.39 |
| 60 | 60.03 | 61.28 | 60.62 | 62.38 | 59.95 | 60.12 |
| 70 | 64.15 | 71.81 | 70.95 | 71.76 | 70.27 | 70.75 |
| 80 | 68.86 | 81.82 | 80.71 | 80.45 | 80.26 | 80.48 |
| 90 | 74.52 | 88.42 | 90.30 | 89.64 | 89.79 | 90.32 |
| 95 | 77.62 | 91.30 | 94.99 | 94.16 | 94.82 | 95.23 |

equations for the Jelinski–Moranda model. The problem in case of the Littlewood model, however, is much harder and is in fact still an open question. Instead of finding such a criterion in the Littlewood case, Barendregt & van Pul (1991) developed an algorithm in order to determine the consistent one from a set of solutions from the ML equations by choosing the nearest solution to a consistent estimator. A more general approach may be possible, probably with use of compactification ideas (see Bahadur, 1967).

**Acknowledgements**

This research was carried out under a grant of the Netherlands, Foundation for Applied Technology (STW). I would like to thank Richard D. Gill, Kacha Dzhaparidze and the referees for their many helpful comments.

**References**

Aalen, O. O. (1978). Non-parametric inference for a family of counting processes. *Ann. Statist.* **6,** 701–726.

Aalen, O. O. (1980). A model for non-parametric regression analysis of counting process. *Lecture Notes in Statistics* **2,** 1–25. Springer, New York.

Andersen, P. K. & Borgan, Ø. (1985). Counting processes models for life history data: a review. *Scand. J. Statist.* **12,** 97–158.

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1992). *Statistical methods based on counting processes.* Springer, New York (to appear).

Bahadur, R. R. (1967). Rates of convergence of estimates and some test statistics. *Ann. Math. Statist.* **38,** 303–324.

Barendregt, L. & van Pul, M. C. (1991). *An analytical solution to the consistency problem for the Littlewood model in software reliability.* Report BS-R9129, Centre for Mathematics and Computer Science, Amsterdam.

Borgan, Ø. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scand. J. Statist.* **1,** 1–16 (corrigendum, *ibid.* 275).

Cramér, H. (1946). *Mathematical methods of statistics.* Princeton University Press, Princeton.

Dzhaparidze, K. & Valkeila, E. (1988). On the Hellinger type distances for filtered experiments. *Probab. Theory Related Fields* **85,** 105–117.

Geurts, W. A. J., Hasselaar, M. M. A. & Verhagen, J. H. (1988). *Large sample theory for statistical inference in several software reliability models.* Report MS-R8807, Centre for Mathematics and Computer Science, Amsterdam.

Gill, R. D. (1980). *Censoring and stochastic integrals.* Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

Gill, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part 1). *Scand. J. Statist.* **16,** 97–128.

Goel, A. L. & Okumoto, K. (1980). *A time dependent error-detection rate model for software performance assessment with applications.* Rome Air development Center, Griffiths Air Force Base, New York.

Hjort, N. L. (1986). Bayes estimators and asymptotic efficiency in parametric counting process models. *Scand. J. Statist.* **13,** 63–85.

Hjort, N. L. (1990). Goodness of fit test in models for life history data based on cumulative hazard rates. *Ann. Statist.* **18,** 1221–1258.

Ibragimov, I. A. & Has'minskii, R. Z. (1979). *Statistical estimation—asymptotic theory.* Springer, New York.

Jacod, J. & Shiryaev, A. N. (1988). *Limit theorems for stochastic processes.* Springer, New York.

Jelinski, Z. & Moranda, P. (1972). Software reliability research. *Statist. Comp. Perform. Eval.* 465–484.

Khmaladze, E. V. (1981). Martingale approach in the theory of goodness of fit tests. *Theory Probabil. Applicat.* **16,** 240–257.

Koch, G. & Sprey, P. J. C. (1983). A martingale approach to software reliability. *IEEE Trans. Rel.* **32,** 342–345.

Kulldorff, G. (1957). On the conditions for consistency and asymptotic efficiency of maximum likelihood estimates. *Skand. Aktuarietids.* **40,** 129–144.

Kurtz, T. G. (1983). Gaussian approximations for Markov chains and counting processes. *Bull. Int. Statist. Inst.* **50,** 361–375.

Le Cam, L. & Lo Yang, G. (1990). *Asymptotics in statistics.* Springer, New York.

Littlewood, B. (1980). Theories of software reliability: how good are they and how can they be improved? *IEEE Trans. Software Eng.* **6,** 489–500.

Moek, G. (1983). *Software reliability models on trial: selection, improved estimation, and practical results.* Report MP 83059 U, National Aerospace Laboratory NLR, Amsterdam.

Moek, G. (1984). Comparison of some software reliability models for simulated and real failure data. *Int. J. Model. & Simulat.* **4,** 29–41.

Musa, J. D. (1975). A theory of software reliability and its application. *IEEE Trans. Software Eng.* **3,** 312–327.

Musa, J. D., Iannino, A. & Okumoto, K. (1987). *Software reliability: measurement, prediction, application.* McGraw-Hill, New York.

van Pul, M. C. (1990). *Asymptotic properties of statistical models in software reliability.* Report BS-R9011, Centre for Mathematics and Computer Science, Amsterdam.

van Pul, M. C. (1991a). *Software reliability simulations on the Musa model.* Report BS-R9122, Centre for Mathematics and Computer Science, Amsterdam.

van Pul, M. C. (1991b). *Modelling imperfect repair and software growth.* Proceedings, Interface '91, Seattle.

van Pul, M. C. (1991c). *Software reliability and the bootstrap.* Report BS-R9123, Centre for Mathematics and Computer Science, Amsterdam.

Rao, C. R. (1973). *Linear statistical interference and its applications.* Wiley, New York.

van der Vaart, A. W. (1987). *Statistical estimation in large parameter spaces.* Mathematical Centre Tracts 44, Mathematisch Centrum, Amsterdam.

Vetterling, W. T., Teukolsky, S. A., Press, W. H. & Flannery, B. D. (1985). *Numerical recipes: the art of scientific computing.* Cambridge University Press, Cambridge.

*Received June 1990, in final form December 1991*

Mark C. van Pul, CWI, P.O. Box 4079, 1009 AB Amsterdam, The Netherlands.