1991

M.C.J. van Pul

Simulations on the Jelinski-Moranda model of software reliability

Department of Operations Research, Statistics, and System Theory Report BS-R9122 September

CWI, nationaal instituut voor onderzoek op het gebied van wiskunde en informatica

CWI is the research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a non-profit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands organization for scientific research (NWO).

Simulations on the Jelinski-Moranda Model of Software Reliability

Mark C. van Pul

CWI
P.O. Box 4079, 1009 AB Amsterdam
The Netherlands

In software reliability theory many different models have been proposed and investigated. Some of these models intuitively match better with reality than others. The properties of certain statistic estimation procedures in connection with these models, on the contrary, are also model dependent. In this paper we will investigate how well the maximum likelihood estimation procedure behaves in case of the very well-known software reliability model suggested by Jelinski-Moranda (1972). For this study we will make use of simulated data.

1980 Mathematics Subject Classification: 62N05, 62M99.
Key Words and Phrases: counting process, intensity function, the Musa model, maximum likelihood estimation, asymptotic behaviour, point estimates and confidence intervals, Wilks likelihood ratio test statistic, coverage probabilities.

1. Introduction

Several investigators have built statistical models in order to estimate the evolution in reliability of computer software during the debugging phase. We refer to Musa et al. (1987) for a complete overview. They all, more or less, consider the following test experiment.

A computer program has been executed during a specific exposure period τ and the failure times $T_i, i = 1,2,...$ are observed. The repairing of a fault takes place immediately after it produces a failure and no new faults are introduced. By using the information obtained from the experiment one can estimate the parameters of the underlying model, especially N, the total number of faults initially present in the software. We will use the maximum likelihood estimation (MLE) procedure for this purpose.

In Van Pul (1990) theoretical results are derived for the asymptotic behaviour of the MLE-procedure for a large class of software reliability models. In this paper we will study how well these theoretical results appear in practice for a specific, very well-known software reliability model, namely the Jelinski-Moranda model. The model is very popular, although practical application has often been disappointing. We will not consider the question, how well the Jelinski-Moranda model matches with reality, however, but we will study how well we can estimate the parameters of the this model with the MLE-method, given some simulated data according to the Jelinski-Moranda model. We will compute point estimates, (upper) confidence bounds and confidence intervals for the model parameters and study their asymptotic behaviour. For the application of the Jelinksi-Moranda model (and other models) to real data we refer to Moek (1983, 1984) and Andersen et al. (1991).

In the next section we will give a description of the Jelinski-Moranda model and give some backgrounds on counting processes and martingale theory. In section 3 we describe the simulation experiments and give the computations of some asymptotic expected variances. In section 4 we will discuss the simulation results. We end this paper with some concluding remarks in section 5.

Report BS-R9122
ISSN 0924-0659
CWI
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

2. The Jelinski-Moranda (JM) model

We start with some definitions. Let N be the unknown number of faults initially present in the software. Let the exposure period be τ 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 failure. Finally, we define $t_{n(\tau)+1} := \tau - T_{n(\tau)}$

In the *Jelinski-Moranda model* (1972) — introduced by Jelinski & Moranda and later generalized by Musa (1975) — when a failure occurs, the corresponding fault is immediately removed with probability one. 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 λ_i becomes

$$\lambda_i = \phi_0 \left[N_0 - (i-1) \right], \tag{2.1}$$

where ϕ_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

$$\lambda(t) = \phi_0 \left[N_0 - n(t-) \right], \tag{2.2}$$

where $\lambda(t)$, $t[0,\tau]$ denotes the failure rate at time t. See figure 1. The interfailure times t_i are independent and exponentially distributed

$$t_i \sim F_i(t), \quad F_i(t) = Pr(t_i < t) = 1 - e^{-\lambda_i t}, \tag{2.3}$$

where the parameter λ_i is given by (2.1).

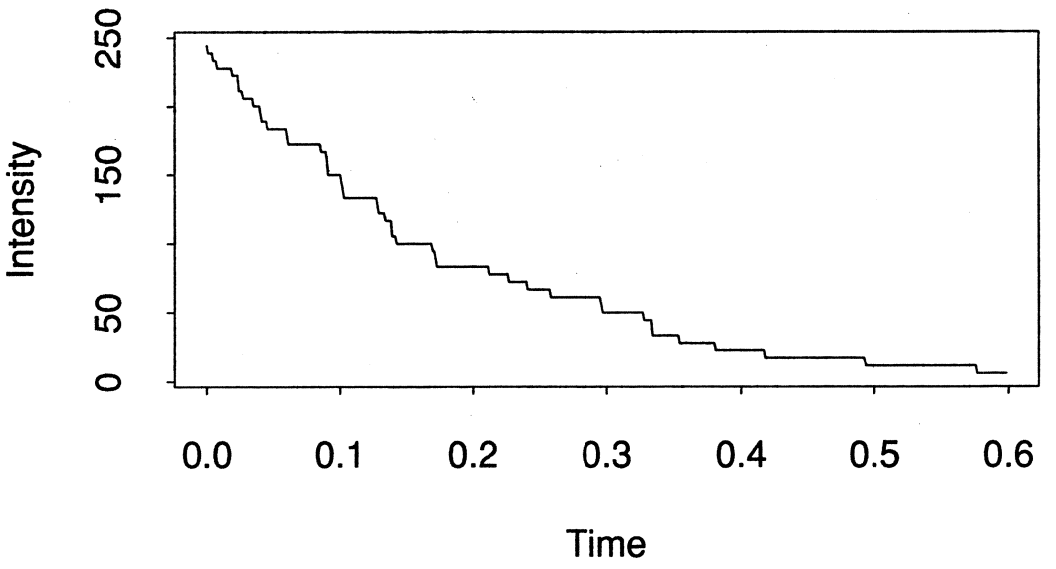

Figure 1: The estimated intensity function for the JM-model.

The joint pdf of the interfailure times $t_1, t_2, \ldots, t_{n(\tau)}, t_{n(\tau)+1}$ is given by (see Aalen (1978)):

$$L(N_0,\phi_0 \mid t_1,\ldots,t_{n(\tau)+1}) = \left[\prod_{i=1}^{n(\tau)} f_i(t_i)\right]\left[1-F_{n(\tau)+1}(t_{n(\tau)+1})\right]$$
$$= \left[\prod_{i=1}^{n(\tau)} \lambda_i\right]\exp\left[-\sum_{i=1}^{n(\tau)+1}\lambda_i t_i\right]. \tag{2.4}$$

With (2.1), the log-likelihood function is

$$\log L(N_0,\phi_0) = \sum_{i=1}^{n(\tau)} \log \phi_0(N_0-i+1) - \sum_{i=1}^{n(\tau)+1} \phi_0(N_0-i+1)t_i; \tag{2.5}$$

hence the likelihood equations become

$$\frac{\partial}{\partial N}\log L(\hat{N},\hat{\phi}) = \sum_{i=1}^{n(\tau)} \frac{1}{\hat{N}-i+1} - \hat{\phi}\tau = 0; \tag{2.6}$$

$$\frac{\partial}{\partial \phi}\log L(\hat{N},\hat{\phi}) = \frac{n(\tau)}{\hat{\phi}} - \sum_{i=1}^{n(\tau)+1}(\hat{N}-i+1)t_i = 0. \tag{2.7}$$

If we define

$$c(n) := \frac{1}{\tau}\sum_{i=1}^{n(\tau)+1}(i-1)t_i, \tag{2.8}$$

we get from (2.6)−(2.7):

$$g(\hat{N}) := \sum_{i=1}^{n(\tau)} \frac{1}{\hat{N}-i+1} - \frac{n(\tau)}{\hat{N}-c(n)} = 0; \tag{2.9}$$

$$\hat{\phi} = \frac{n(\tau)}{(\hat{N}-c(n))\tau}. \tag{2.10}$$

It is easy to see that $n(\tau)$ and $c(n)$ are sufficient statistics in the statistical sense that we can calculate all likelihood based quantities from $n(\tau)$ and $c(n)$ only. Given the numbers $n(\tau)$ and $c(n)$ as the outcomes of a simulation (or as real life data), we can solve

$$g(\hat{N}) = 0 \tag{2.11}$$

numerically and then compute $\hat{\phi}$ with (2.10). Moek (1983) showed that there exists a unique solution of (2.11) if and only if

$$c(n) > \frac{n(\tau)-1}{2}. \tag{2.12}$$

To make this intuitively more likely, note that for large values of N, $g(N)$ can be approximated by:

$$g(N) = \sum_{i=1}^{n(\tau)}\left[\frac{1}{N-i+1} - \frac{1}{N-c(n)}\right]$$
$$= \sum_{i=1}^{n(\tau)}\left[\frac{(N-c(n)) - (N-i+1)}{(N-i+1)(N-c(n))}\right]$$
$$\sim \frac{1}{N^2}\sum_{i=1}^{n(\tau)}[i-1-c(n)]$$
$$= \frac{n(\tau)}{N^2}\left[\frac{n(\tau)-1}{2} - c(n)\right]. \tag{2.13}$$

In figure 2 we have illustrated this criterion.

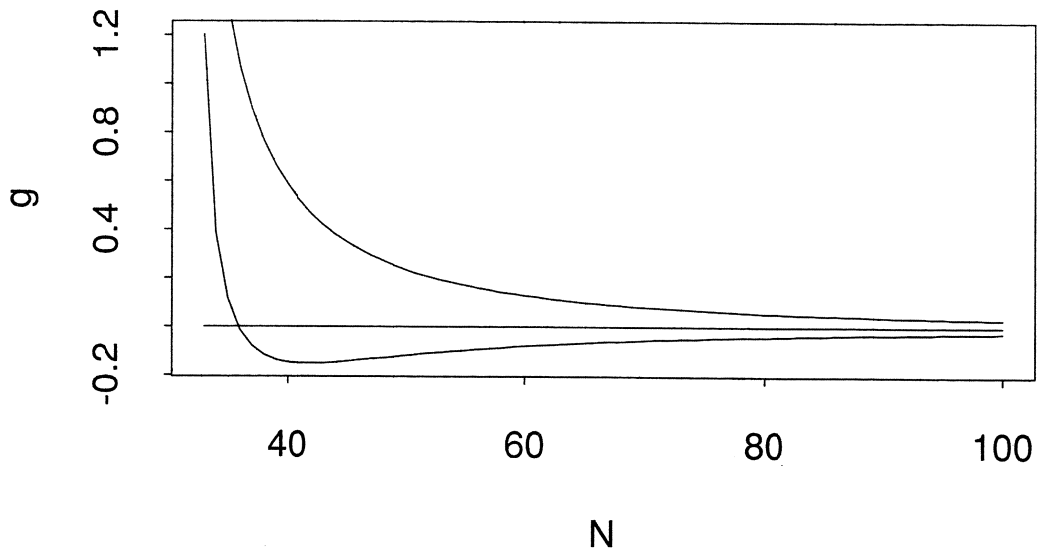

Figure 2: Moek's criterion.

In Van Pul (1990) the asymptotic behaviour of the ML-estimators $\hat{N}$ and $\hat{\phi}$ was investigated when N_0 is large. An important novel concept is the way in which asymptotics is treated. It does not make sense to let τ 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 then, that asymptotics should be relevant in the practical situation in which N_0 is large and $n(\tau)/N_0$ is not close to zero or one.

We therefore introduce a sequence of counting processes $n_\nu(t)$, $t \in [0,\tau]$, $\nu = 1,2,...$ and let N conceptually increase. Let $N = N_\nu \to \infty$ for $\nu \to \infty$. By the reparametrization

$$N_\nu = \nu\gamma_\nu \tag{2.14}$$

with a dummy variable γ_ν we can denote the associated intensity functions by

$$\lambda_\nu(t;\gamma,\phi) := \lambda(t;\nu\gamma,\phi) = \phi\left[\nu\gamma - n_\nu(t-)\right]. \tag{2.15}$$

We define the stochastic processes

$$x_\nu(t) := \nu^{-1} n_\nu(t), \quad t \in [0,\tau] \tag{2.16}$$

and we further define:

$$\begin{aligned}\beta(t;\gamma,\phi;x_\nu) &:= \frac{\lambda_\nu(t;\gamma,\phi)}{\nu} \\ &= \frac{\phi}{\nu}\left[\nu\gamma - n_\nu(t-)\right] \\ &= \phi\left[\gamma - x_\nu(t-)\right].\end{aligned} \tag{2.17}$$

It is shown (Kurtz (1983), van Pul (1990)) that under weak smoothness and boundedness conditions on the function β, the sequence of stochastic processes x_ν converges uniformly on $[0,\tau]$ in probability

as $\nu \to \infty$ to a deterministic function x_0, which is the solution of

$$x(t) = \int_0^\tau \beta(s;\gamma_0,\phi_0;x)\mathrm{d}s. \tag{2.18}$$

In case of the Jelinski-Moranda model, it is easy to show that this deterministic function x_0 is given by:

$$x_0(t) := \gamma_0(1-\mathrm{e}^{-\phi_0 t}). \tag{2.19}$$

In the real-life situation we have $\nu = N_0$ and $\gamma_0 = 1$. It is further shown in Van Pul (1990), that the normalized ML-estimators $(\sqrt{\nu}(\hat{\gamma}_\nu - \gamma_0), \sqrt{\nu}(\hat{\phi}_\nu - \phi_0))$ in the Jelinski-Moranda model are consistent and asymptotically normal with asymptotic mean zero and asymptotic covariance matrix

$$\Sigma = \begin{bmatrix} \frac{1}{\gamma_0}(\mathrm{e}^{\phi_0\tau}-1) & \tau \\ \tau & \frac{\gamma_0}{\phi_0^2}(1-\mathrm{e}^{-\phi_0\tau}) \end{bmatrix}^{-1}. \tag{2.20}$$

From this it follows, that we can define centered and normalized quantities $\hat{X}$ and $\hat{Y}$, satisfying:

$$\hat{X} := \frac{\hat{N}-N_0}{\sqrt{N_0}} \xrightarrow{D} N\left(0, \frac{1-\mathrm{e}^{-\phi_0\tau}}{\mathrm{e}^{\phi_0\tau}+\mathrm{e}^{-\phi_0\tau}-\phi_0^2\tau^2-2}\right), \tag{2.21}$$

$$\hat{Y} := \sqrt{N_0}(\hat{\phi}-\phi_0) \xrightarrow{D} N\left(0, \frac{\phi_0^2(\mathrm{e}^{\phi_0\tau}-1)}{\mathrm{e}^{\phi_0\tau}+\mathrm{e}^{-\phi_0\tau}-\phi_0^2\tau^2-2}\right), \tag{2.22}$$

as $N_0 \to \infty$. For more details, see Van Pul (1990).

3. Description of the simulations and computations

In our experiments we will keep the exposure period $\tau := 1$, the sample-size $K_0 := 10000$ and one of the model parameters, namely the true occurrence rate $\phi_0 := 1$ fixed. In the simulations we will vary the other model parameter N_0, the number of faults initially present in the software: $N_0 \in \{ 50, 500, 5000 \}$. These three cases should represent very small, normal and very large computer programs.

We define for $i = 1,2,\ldots,N_0$:

$$\lambda_i := \phi_0(N_0 - i + 1) = N_0 - i + 1, \tag{3.1}$$

$$F_i(t) := 1-\mathrm{e}^{-\lambda_i t}, \quad t \in [0,\tau]. \tag{3.2}$$

We simulate interfailure times $t_1 \sim F_1$, $t_2 \sim F_2$, . . . , $t_m \sim F_m$, $t_{m+1} \sim F_{m+1}$ until

$$\sum_{i=1}^{m} t_i < \tau, \tag{3.3}$$

$$\sum_{i=1}^{m+1} t_i \geqslant \tau. \tag{3.4}$$

Then we set

$$t_{m+1} := \tau - \sum_{i=1}^{m} t_i. \tag{3.5}$$

The characteristic quantities of each simulation are $n = n(\tau)$, the number of faults detected up to time τ, and $c = c(n)$, given by (2.8), which is a measure for the change of the occurrence rate in time. Since $n(\tau)$ and $c(n)$ are sufficient statistics, we can spare a lot of memory space during sampling.

Repeating the above described simulation K_0 times, yields two K_0-vectors: $\vec{n} := (n_1, \ldots, n_{K_0})$ and $\vec{c} := (c_1, \ldots, c_{K_0})$. Given these simulated vectors $\vec{n}$ and $\vec{c}$, we restrict ourselves to those simulations for which Moek's criterion is satisfied, which induces the existence and uniqueness of the solution of the ML-equations (2.6)−(2.7). We define

$$K := \#\{ i : 2c_i - n_i - 1 > 0 \}. \tag{3.6}$$

We now solve $g(\hat{N}) = 0$ (see (2.9)), with a variant of binary search. Then we compute the estimator $\hat{\phi}$ with (2.10). Hence this procedure yields two K-vectors of ML-estimators: $\hat{N} = (\hat{N}_1, \ldots, \hat{N}_K)$ and $\hat{\phi} = (\hat{\phi}_1, \ldots, \hat{\phi}_K)$.

We will construct upper confidence bounds and two-sided confidence intervals for N, based on the ML-estimator $\hat{N}$. For this purpose, we will make use of an approximate normal test statistic z and of a likelihood ratio test statistic w. From (2.21) it follows directly that we have approximately

$$z := \frac{\hat{N} - N_0}{\sigma\sqrt{\hat{N}}} \sim N(0,1), \tag{3.7}$$

where the asymptotic variance σ^2 is given by

$$\sigma^2 := \frac{1 - e^{-\phi_0 \tau}}{e^{\phi_0 \tau} + e^{-\phi_0 \tau} - \phi_0^2 \tau^2 - 2}; \tag{3.8}$$

σ can be estimated consistently by using either the expected information matrix Σ (see 2.20), which yields

$$\begin{aligned} \hat{\sigma}_{EXP} &:= \left[\frac{1 - e^{-\hat{\phi}\tau}}{e^{\hat{\phi}\tau} + e^{-\hat{\phi}\tau} - \hat{\phi}^2\tau^2 - 2} \right]^{1/2} \\ &= \left[(e^{\hat{\phi}\tau} - 1) - \frac{\hat{\phi}^2\tau^2}{1 - e^{-\hat{\phi}\tau}} \right]^{-1/2} \end{aligned} \tag{3.9}$$

or by using the observed information matrix I (see Van Pul (1990)), which yields

$$\hat{\sigma}_{OBS} := \left[\hat{N} \sum_{i=1}^{n(\tau)} \frac{1}{(\hat{N} - i + 1)^2} - \frac{\hat{N}\hat{\phi}^2\tau^2}{n(\tau)} \right]^{-1/2}. \tag{3.10}$$

Let k_α and $k_{\alpha/2}$, $\alpha \in [0, \frac{1}{2}]$, be quantiles of the standard normal distribution, such that:

$$\Pr(|z| \leq k_\alpha) = 1 - \alpha, \tag{3.11}$$

$$\Pr(|z| \leq k_{\alpha/2}) = 1 - \alpha/2. \tag{3.12}$$

If $\hat{\sigma}$ denotes either $\hat{\sigma}_{EXP}$ or $\hat{\sigma}_{OBS}$ an approximate $(1-\alpha)$-upper-bound U_{normal} is given by

$$U_{\text{normal}} := \hat{N} + k_\alpha \hat{\sigma}\sqrt{\hat{N}} \tag{3.13}$$

and $(LB_{\text{normal}}, UB_{\text{normal}})$ is an approximate $(1-\alpha)$ confidence interval where:

$$LB_{\text{normal}} := \hat{N} - k_{\alpha/2}\hat{\sigma}\sqrt{\hat{N}}, \tag{3.14}$$

$$UB_{\text{normal}} := \hat{N} + k_{\alpha/2}\hat{\sigma}\sqrt{\hat{N}}. \tag{3.15}$$

Another way of constructing upper-bounds and confidence intervals for $\hat{N}$ can be done by making

use of the Wilks likelihood ratio test (WLRT) statistic

$$w := 2[\max_{\phi} \log L(\hat{N},\phi) - \max_{\phi} \log L(N_0,\phi)], \tag{3.16}$$

where the log-likelihood function $\log L$ is given by (2.5). As indicated in Van Pul (1990), the WLRT statistic w is asymptotically chi-squared distributed with one degree of freedom. Let c_α and $c_{2\alpha}$ denote the $(1-\alpha)$ and $(1-2\alpha)$ quantiles of the $\chi^2(1)$-distribution. Then an alternative $(1-\alpha)$-upper-bound U_{Wilks} is the largest solution x of

$$G(x) := 2[\max_{\phi} \log L(\hat{N},\phi) - \max_{\phi} \log L(x,\phi)] = c_\alpha \tag{3.17}$$

and if LB_{Wilks} and UB_{Wilks} are the solutions of

$$G(x) := 2[\max_{\phi} \log L(\hat{N},\phi) - \max_{\phi} \log L(x,\phi)] = c_{2\alpha}, \tag{3.18}$$

with $LB_{\text{Wilks}} < \hat{N} < UB_{\text{Wilks}}$, then $(LB_{\text{Wilks}}, UB_{\text{Wilks}})$ is an alternative $(1-\alpha)$-confidence interval for the parameter of most interest N. See figure 3.

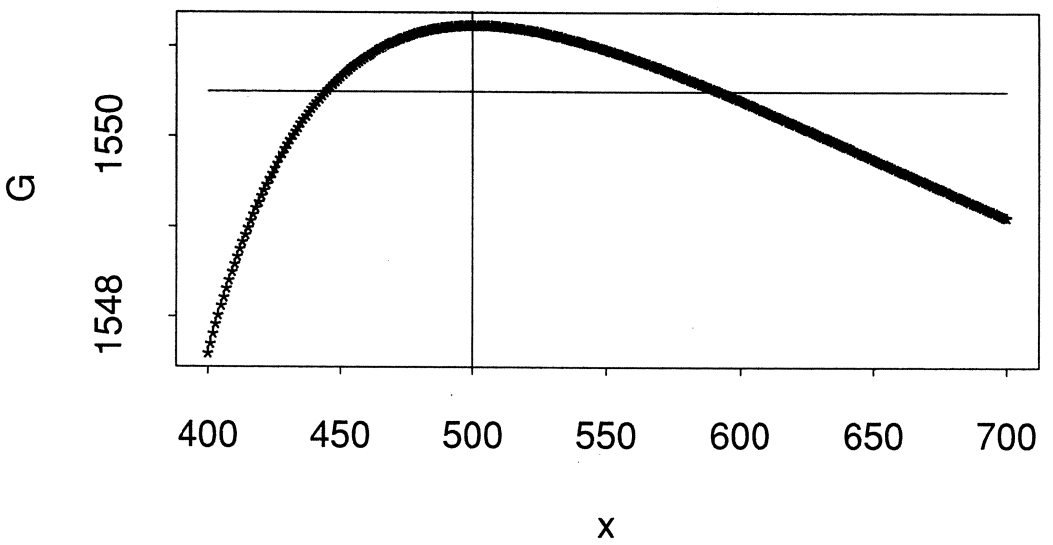

Figure 3: The function G(x).

The computation of approximate normal upper confidence bounds and confidence intervals is much easier, but those calculated from the WLRT statistic will be much more satisfying. To spare computer time in the latter case we will approximate sums (with 500-5000 terms) by standard integrals with well-known primitives.

The empirical coverage probability or hitting percentage (hit%) of a confidence interval is defined to be the fraction of the K constructed confidence intervals, which contain the true number of faults N_0. The miss-under percentage (mu%) is the fraction of confidence intervals whose upper-bound is lower than N_0. In an analogous manner we define the miss-over percentage (mo%).

4. Discussion of the results of the experiment

Once again, we mention the fact that during the simulation experiment we will keep the exposure period $\tau=1$ and the true occurrence rate $\phi_0=1$ fixed. As we have studied some theoretical properties of the ML-estimators as $N_0\to\infty$ in Van Pul (1990), we are very interested for which size of N_0 this asymptotic behaviour appears in practice. Therefore, we do the simulations and computations, described in the last section for various values of N_0 in the range from 50 to 5000. Simulations with values of N_0, essentially bigger then 5000, were not possible (and not interesting) with the available computing facilities. For each value of N_0, we repeat the experiment $K_0=10000$ times.

In table 1 we give the number of the finite estimators. That is, the number of estimators which corresponding simulation characteristics $n(\tau)$ and $c(n)$, satisfy Moek's criterion (2.12). In fact this number K is the realized number of replicates. In table 2 means and mean square errors for the ML (point-) estimators $\hat{N}$ and $\hat{\phi}$ are given. In table 3 means and variances of the centered and normalized quantities $\hat{X}$ and $\hat{Y}$ are given. In tables 4 we compared the approximate normal and Wilks likelihood ratio upper confidence bound for N_0 at various values of the confidence level α. Using the same simulation data as in tables 1−3, we computed the realized hitting percentages. In table 5 we did the same for two-sided confidence intervals derived from the approximate normal and the Wilks ratio likelihood test statistics.

N_0	K
50	9027
500	9996
5000	10000

Table 1: Number of finite estimators.

N_0	$E\hat{N}$	$E\hat{\phi}$	$E(\frac{\hat{N}}{N_0}-1)^2$	$E(\hat{\phi}-\phi_0)^2$
50	48.3947	1.3095	0.0909	0.3982
500	506.1666	1.0228	0.0194	0.0406
5000	5008.0997	1.0015	0.0015	0.0039

Table 2: Means and mean square errors of the MLE's $\hat{N}$ and $\hat{\phi}$.

N_0	$E\hat{X}$	$Var\hat{X}$	$E\hat{Y}$	$Var\ \hat{Y}$
50	-0.2270	4.4943	2.1886	15.1202
500	0.2758	9.6089	0.5100	20.0434
5000	0.1145	7.4131	0.1071	19.7273
∞	0.0000	7.3365	0.0000	19.9423

Table 3: Means and variances of centered and normalized quantities $\hat{X}$ and $\hat{Y}$.

For small values of N_0, we see in table 1 that $K<K_0$. When N_0 increases, K/K_0 converges quickly to 1. For small values of N_0 the variation in n can be relatively large. So possibly a number of simulations will exist, which doesn't satisfy Moek's criterion: $2c-n-1>0$. On the other hand we have as $N_0\to\infty$:

$$\frac{n}{N_0} \xrightarrow{P} x_0(\tau) = 1-e^{-1}, \tag{4.1}$$

$$\frac{c}{N_0} = \frac{1}{N_0}\sum_{i=1}^{n(\tau)+1}(i-1)t_i = \frac{1}{N_0}\int_0^\tau n(s-)ds \xrightarrow{P} \int_0^\tau x_0(s)ds = \int_0^\tau (1-e^{-s})ds = e^{-1}. \tag{4.2}$$

Hence $n/c \xrightarrow{P} e-1$ as $N_0 \to \infty$ and Moek's criterion will be satisfied with a probability tending to one, when N_0 grows larger.

For small values of N_0 we see in table 2, that $\hat{N} \ll N_0$ and $\hat{\phi} \gg \phi_0$. As N_0 increases, we find however, by looking at the mean square errors, confirmation of the theoretical results:

$$\frac{\hat{N}}{N_0} \xrightarrow{P} 1, \tag{4.3}$$

$$\hat{\phi} \xrightarrow{P} \phi_0 = 1, \tag{4.4}$$

derived in Van Pul (1990). The under-estimation of N_0 can be explained by the fact that for small values of N_0, K is smaller than K_0 and the biggest estimates for N_0 are omitted. In the same way the over-estimation of ϕ_0 by $\hat{\phi}$ can be explained, because $\hat{\phi} \sim 1/\hat{N}$ (see (2.9)).

The distribution of $\hat{N}$ is even for large values of N_0 not symmetric at all, but has a very long tail on the right. The skewness of the distribution of $\hat{N}$ is related to stability problems of the numerical approximation procedure. A vertical perturbation of the curve of the likelihood function would only cause a small deviation of the solution $\hat{N}$ to the left, but easily cause a large one to the right. See figure 2. This skewness disappears slowly as N_0 increases. We see in table 3 that the that the convergence of $\hat{X}$ and $\hat{Y}$ to normal distributions with means zero and asymptotic variances as expected (in (2.21) and (2.22)) is rather slow. The difference in the asymptotic behaviour of $\hat{N}$ and $\hat{\phi}$ is illustrated by the histograms and qq-plots given in figures 4 and 5. 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.

We will know consider the construction of confidence intervals for N_0. For the construction of confidence intervals based on the approximate normal test statistic we estimated the asymptotic variance with use of the expected information (see (3.9)). If the observed information had been used, the results would not have been significantly different. As the distribution of $\hat{N}$ is skew, however, 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 (3.18). In table 4 we see that at the 50% confidence level the approximate normal and the Wilks likelihood ratio upper-bounds both are equal to the maximum likelihood estimator $\hat{N}$. For higher confidence levels α, the Wilks upper confidence bound is larger than the corresponding approximate normal one. The same will then of course hold for the coverage probabilities. All these observations are clearly understood by the way of constructing the upper-bounds. 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 less skew coverage probabilities, than the approximate normal ones. Note that the sample-size $K_0 = 10000$.

Finally, we consider two-sided confidence intervals. In table 5 we observe a skewness in the hitting probabilities, which appears to be significantly larger in case of the approximate normal confidence intervals for N_0. This skewness decreases with decreasing confidence level α and with increasing N_0. The size of an approximate normal confidence interval, which is by definition symmetric around $\hat{N}$, increases of course with increasing α, but it is also proportional to $\hat{N}$. This explains the extreme skewness of the hitting probabilities of the approximate normal confidence intervals. Wilks confidence intervals behave better for two reasons: they are bigger and shifted to the right in comparison with the approximate normal confidence intervals. In table 5 we can see that for high levels of confidence the Wilks confidence intervals are significantly better and less skew than the approximate normal ones.

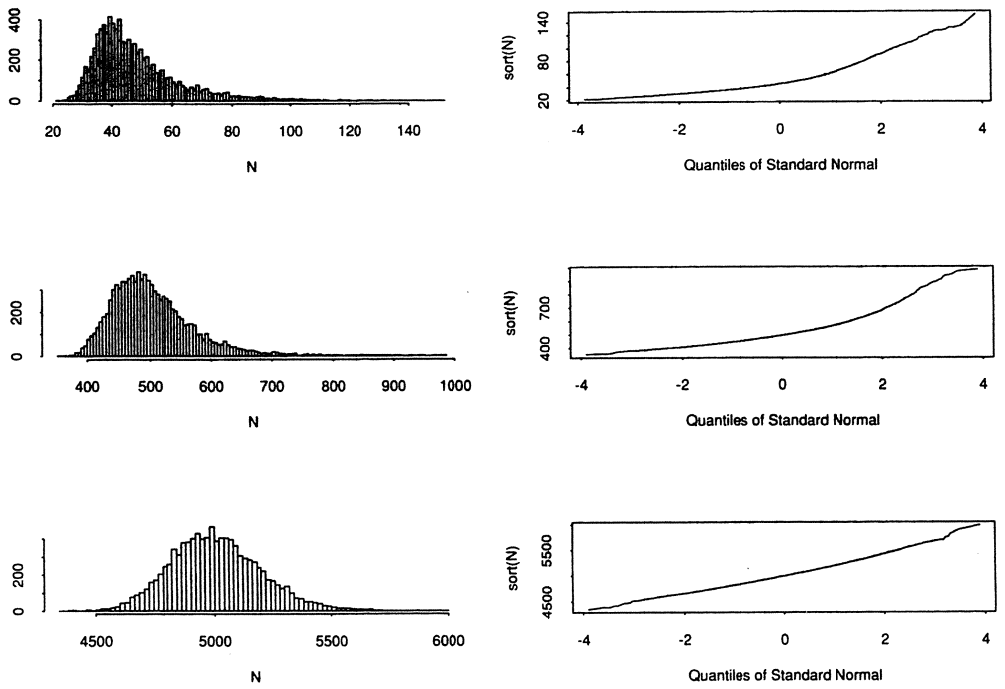

Figure 4: Histograms and qq-plots of $\hat{N}$.
(4a) $N_0 = 50$ (4b) $N_0 = 500$ (4c) $N_0 = 5000$

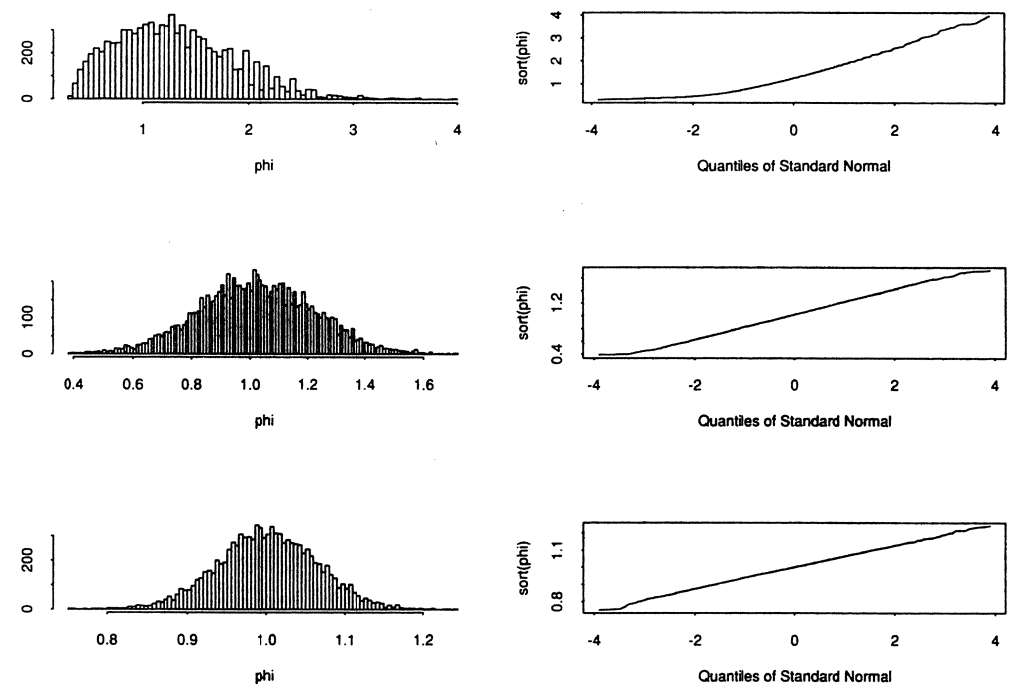

Figure 5: Histograms and qq-plots of $\hat{\phi}$.
(5a) $N_0 = 50$ (5b) $N_0 = 500$ (5c) $N_0 = 5000$

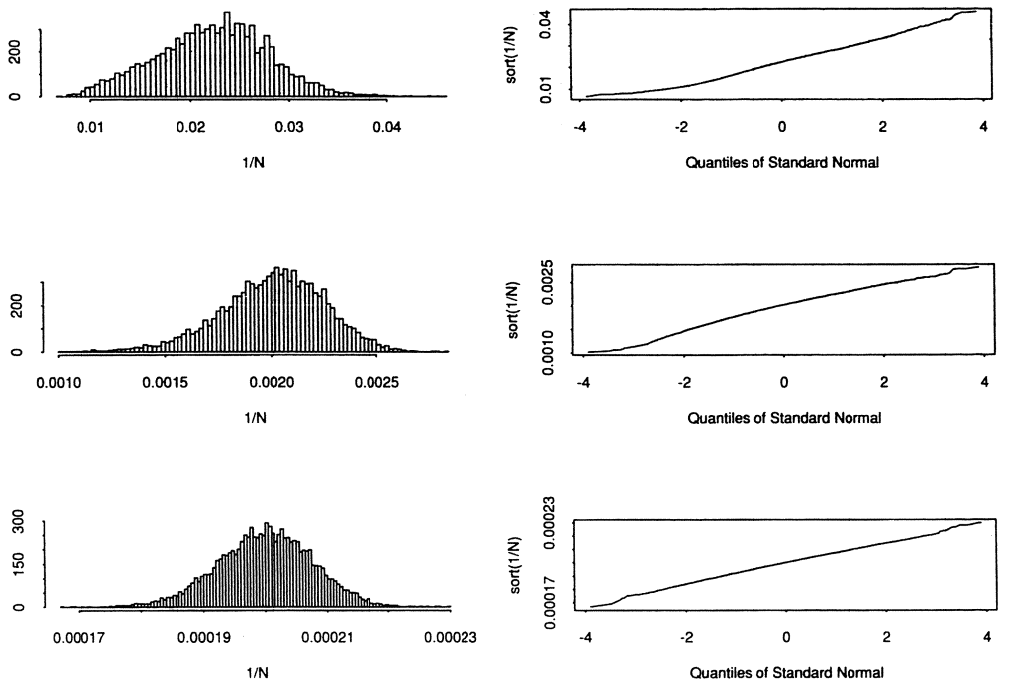

Figure 6: Histograms and qq-plots of $1/\hat{N}$.
(6a) $N_0 = 50$ (6b) $N_0 = 500$ (6c) $N_0 = 5000$

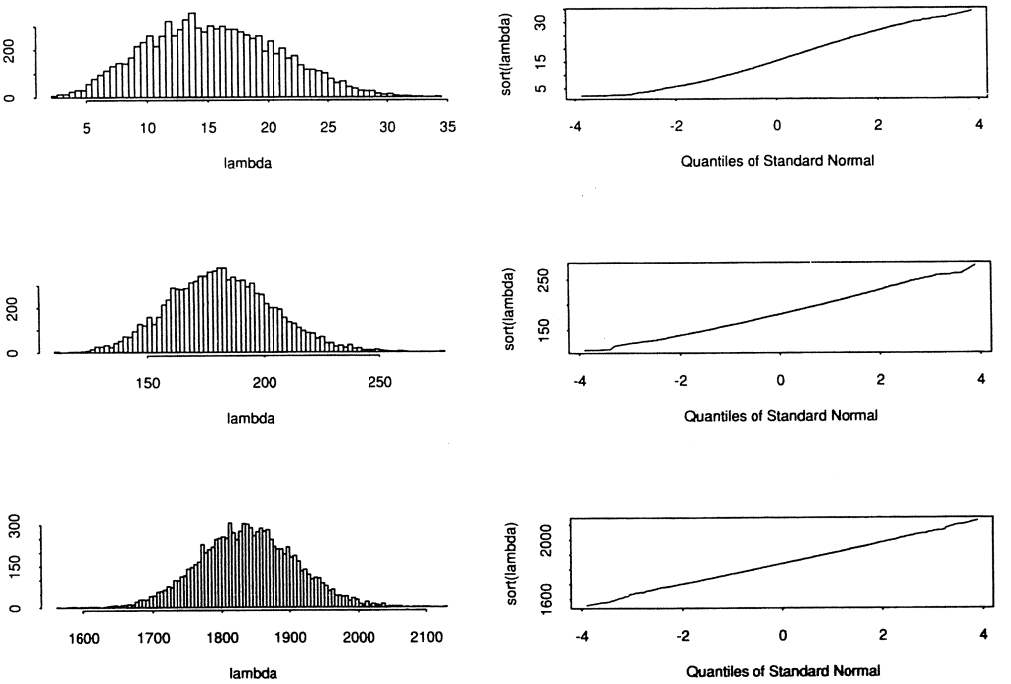

Figure 7: Histograms and qq-plots of $\hat{\lambda} := \hat{\phi}(\hat{N} - n)$.
(7a) $N_0 = 50$ (7b) $N_0 = 500$ (7c) $N_0 = 5000$

	normal		Wilks	
α	hit%	E(U_{norm})	hit%	E(U_{Wilks})
0.50	34	48	35	49
0.55	39	50	38	51
0.60	44	53	44	55
0.65	48	56	49	60
0.70	52	59	55	72
0.75	56	62	61	∞
0.80	60	65	68	∞
0.85	64	69	75	∞
0.90	69	74	82	∞
0.95	75	82	90	∞

Table 4a: $N_0 = 50$.

	normal		Wilks	
α	hit%	E(U_{norm})	hit%	E(U_{Wilks})
0.50	46	506	46	507
0.55	51	514	51	514
0.60	56	523	56	523
0.65	60	532	61	534
0.70	65	541	66	545
0.75	69	551	71	558
0.80	73	562	76	575
0.85	78	575	82	596
0.90	83	591	88	627
0.95	88	615	94	685

Table 4b: $N_0 = 500$.

	normal		Wilks	
α	hit%	E(U_{norm})	hit%	E(U_{Wilks})
0.50	49	5008	49	5009
0.55	54	5032	54	5032
0.60	59	5056	59	5057
0.65	64	5082	64	5084
0.70	69	5109	69	5112
0.75	74	5138	75	5144
0.80	79	5170	80	5180
0.85	84	5208	85	5222
0.90	88	5255	90	5278
0.95	93	5326	95	5364

Table 4c: $N_0 = 5000$.

Table 4: Hitting percentages and mean upper-bounds of one-sided confidence intervals.

	normal					Wilks				
α	mu%	hit%	mo%	E(LB_{norm})	E(UB_{norm})	mu%	hit%	mo%	E(LB_{Wilks})	E(UB_{Wilks})
0.50	44	56	0	34	62	39	52	9	39	∞
0.55	42	58	0	32	64	36	57	7	39	∞
0.60	40	60	0	30	65	32	62	6	38	∞
0.65	38	62	0	29	67	29	67	4	37	∞
0.70	36	64	0	26	69	25	72	3	37	∞
0.75	34	66	0	24	72	22	76	2	36	∞
0.80	31	69	0	21	74	18	81	1	36	∞
0.85	28	72	0	18	78	14	85	1	35	∞
0.90	25	75	0	14	82	10	90	0	34	∞
0.95	22	78	0	7	88	6	94	0	33	∞

Table 5a: $N_0 = 50$.

	normal					Wilks				
α	mu%	hit%	mo%	E(LB_{norm})	E(UB_{norm})	mu%	hit%	mo%	E(LB_{Wilks})	E(UB_{Wilks})
0.50	31	51	18	461	551	28	51	21	466	588
0.55	29	56	15	456	556	26	55	19	462	566
0.60	27	61	12	450	562	23	60	17	458	575
0.65	25	66	9	444	568	20	65	15	454	585
0.70	22	72	6	437	575	18	70	12	449	596
0.75	20	77	3	430	582	15	75	10	445	610
0.80	17	82	1	421	591	12	80	8	439	627
0.85	14	86	0	410	602	9	85	6	433	650
0.90	12	88	0	397	615	6	90	4	426	685
0.95	9	91	0	376	636	3	95	2	415	780

Table 5b: $N_0 = 500$.

	normal					Wilks				
α	mu%	hit%	mo%	E(LB_{norm})	E(UB_{norm})	mu%	hit%	mo%	E(LB_{Wilks})	E(UB_{Wilks})
0.50	26	51	23	4877	5138	26	50	24	4883	5143
0.55	24	44	21	4862	5154	23	55	22	4869	5160
0.60	21	61	18	4845	5170	20	60	20	4854	5179
0.65	19	66	15	4827	5188	18	65	17	4837	5199
0.70	16	71	13	4807	5208	15	71	14	4820	5221
0.75	14	76	10	4785	5230	12	76	12	4801	5247
0.80	12	81	7	4760	5255	10	80	10	4779	5247
0.85	9	86	5	4729	5286	8	85	7	4753	5314
0.90	7	90	3	4690	5326	5	90	5	4721	5363
0.95	4	95	1	4629	5386	3	95	2	4672	5441

Table 5c:

Table 5: Hitting/miss-percentages and mean lower/upper-bounds of two-sided confidence intervals.

5. Concluding remarks

In this paper we have only made a beginning with the study of the behaviour of the ML-estimators (in practice), computed from simulated data. As stated before, it was not our intention to prove or disprove the validity of the Jelinski-Moranda model for real data sets. We wanted to get some more insight in the asymptotic behaviour of the ML-estimators, assuming the model is correct. The simulation results discussed in the previous section, confirm the consistency and asymptotic normality of the maximum likelihood estimators (as derived in Van Pul (1990)), but also show that asymptotic convergence in distribution is appearing very slowly and that for small values of N_0 the distribution of $\hat{N}$ can be very skew. On the other hand, asymptotic confidence intervals, constructed with the Wilks likelihood test statistic, have coverages probabilities close to the corresponding confidence levels, even at low sample-sizes.

For $N_0=50$ we have further modified the estimated coverage probabilities for two-sided Wilks confidence intervals by taking into account also data-points $(n(\tau),c(n))$ where Moek's criterion is not satisfied. This corresponds with $\hat{N}=\infty$ (see figure 2). Note that the Jelinski-Moranda model has as a special case the Poisson model (with constant failure intensity λ_0); this is the limit-case letting $N_0\to\infty$ and $\phi_0\to 0$ such that $\lambda_0:=\phi_0 N_0$ is a constant. We can easily derive and maximize the (log) likelihood function for this special case. Using (3.20) we have constructed confidence intervals of the form (lower-bound, infinity) and considered them as two-sided confidence intervals for N_0. In table 6 we give the original values P_{old} (see table 5a), correction terms $P_{infinite}$ and the corrected values P_{new} for the estimation of the coverage probabilities. We computed P_{new} from

$$P_{new}=\frac{K}{K_0}P_{old}+(1-\frac{K}{K_0})P_{infinite}, \tag{5.1}$$

where $K=9027$ and $K_0=10000$ are respectively the number of finite estimators (see table 1) and the sample-size. The figures of table 6 show that the new hitting percentages are a little bit worse than the old ones; the miss-rates mu% and mo%, however, are more symmetrically distributed. Probably the old values P_{old} (totally ignoring infinite estimators $\hat{N}$) were (a bit) too optimistic.

	P_{old}			$P_{infinite}$			P_{new}		
α	mu%	hit%	mo%	mu%	hit%	mo%	mu%	hit%	mo%
0.50	39	52	9	0	15	85	35	49	16
0.55	36	57	7	0	17	83	32	54	14
0.60	32	62	6	0	21	79	29	58	13
0.65	29	67	4	0	27	73	26	63	11
0.70	25	72	3	0	33	67	23	68	9
0.75	22	76	2	0	42	59	20	73	7
0.80	18	81	1	0	49	51	16	78	6
0.85	14	85	1	0	61	39	13	83	4
0.90	10	90	0	0	72	28	9	88	3
0.95	6	94	0	0	89	11	5	94	1

Table 6: Corrected hitting and miss percentages of two-sided confidence intervals for $N_0=50$.

We plan to investigate various ways of improving the asymptotic approximation. In Van Pul (1991) we studied the asymptotic behaviour of parametric *bootstrap methods*. We have proved that the parametric bootstrap (in the context of counting processes) works, that is, is asymptotically consistent. Computer simulations showed that with a studentized version of the parametric bootstrap we can construct one-sided confidence intervals with hitting percentages even better than those constructed with Wilks LRT statistic. For two-sided confidenced intervals, constructed with a studentized version of

the parametric bootstrap, we obtained less satisfying figures. We are trying to improve this. With use of *Edgeworth expansions* we can probably determine the second order term of the limit distribution of $\hat{N}$.

Another direction for further research seems to be the investigation of *certain functions of the model parameters.* Using the delta-method (Gill (1989)) we can derive immediately asymptotic results (convergence to normal distribution) for decent functions of the parameters N and ϕ. Entities like $1/\hat{N}$ and $\hat{\lambda}:=\hat{\phi}(\hat{N}-n)$ behave already for small values of N_0 remarkably well: they have smaller variances, are less skewed in comparison with $\hat{N}$ and converge faster to normality (see table 7 and figures 6 and 7).

N_0	Var $\hat{N}$	Var $\hat{\lambda}$
50	225	183
500	4804	2028
5000	37066	20395

Table 7: Comparison of the variances of $\hat{N}$ and $\hat{\lambda}:=\hat{\phi}(\hat{N}-n)$.

Of course, we can use asymptotic results of for instance $1/\hat{N}$ to construct alternative confidence intervals for N. It is, however, perhaps even more interesting to study and estimate quantities like $\hat{\lambda}$ that have a clear interpretation, also when the assumed underlying model is not true.

Acknowledgement

This research was carried out under a grant of the Netherlands Foundation for Applied Technology (STW).

References

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

Andersen,P.K., Borgan, Ø., Gill, R.D. & Keiding, N. (1991), *Statistical Methods Based on Counting Processes.* Springer-Verlag, New York. To appear.

Gill,R.D. (1989), *Non- and semi-parametric maximum likelihood estimators and the von Mises method 1622 (Part 1). Scandinavian Journal of Statistics* **16,** 97-128.

Jelinski,Z. & Moranda, P. (1972), Software Reliability Research. *Statistical Computer Performance Evaluation* 465-484.

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

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. *International Journal of Modelling & Simulation,* Volume 4, No. 1, 29-41.

Musa,J.D. (1975), A theory of software reliability and its application. *IEEE Transactions on Software Engineering,* **3,** 312-327.

Musa,J.D., Iannino, A. & Okumoto, K. (1987), *Software Reliability: Measurement, Prediction, Application.* McGraw-Hill Book Company, 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. (To appear in SJS 1992).

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