On the estimation of the parameters for the Littlewood model in software reliability

L. G. Barendregt

Graaf Janstraat 57, 2713 CH Zoetermeer, The Netherlands

M. C. van Pul[1]

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

A very well-known model in software reliability theory is that of Littlewood (1980). The (three) parameters in this model are usually estimated by means of the maximum likelihood method. The system of likelihood equations can have more than one solution. Only one of them will be consistent, however. In this paper we present a different, more analytical approach, exploiting the mathematical properties of the log-likelihood function itself. Our belief is that the ideas and methods developed in this paper could also be of interest for statisticians working on the estimation of the parameters of the generalised Pareto distribution. For those more generally interested in maximum likelihood the paper provides a 'practical case', indicating how complex matters may become when only three parameters are involved. Moreover, readers not familiar with counting process theory and software reliability are given a first introduction.

Key Words & Phrases: generalised Pareto distribution, maximum likelihood estimation, consistency.

1 Introduction

To-day it is hard to think of any area of modern society in which computer systems do not play a dominant role. In space- and air-navigation, defence, telecommunication and health-care, to name a few, computers have taken over the most life-critical tasks. Since the early seventies, many researchers have therefore paid attention to the problem of estimation and prediction of software reliability. They used various starting-points, assumptions, and techniques; all aiming at the same goal. In this paper we will use the concepts of *(error) counting models*. Other approaches to software reliability are: *fault seeding* (MILLS, 1972), *iterated testing* (NAGEL and SKRIVAN, 1982) and *static complexity analysis* (AKIYAMA, 1971, McCABE, 1976, HALSTEAD, 1977).

---

[1] The author's current address is: Measurement and Computational Applications, Koninklijke/ Shell-Laboratorium Amsterdam, P.O. Box 3003, 1003 AA Amsterdam, The Netherlands.

Following the dynamic approach of counting processes, we consider the test experiment described below. A computer program is tested for a specified length of time. Inputs are selected randomly from the input-space in a way that is representative for the operational profile. Either the program produces the correct output, or a software failure occurs. That is, the software produces the wrong answer or no answer at all. After the detection of a failure, the program is sent to a team of debuggers. When the fault is found and fixed, available data concerning fault and failure are gathered in a database and testing continues with a new input until the end of the test period is reached.

The class of *Error-Counting and Debugging Models* consists of relatively simple models, considering the test experiment as described above, characterised by the fact that they are only based on certain test data, such as the occurrence times of failures. These error-counting and debugging models do not explicitly depend on factors like the length and the structure of the program, the language in which it is written, the skill of the programmer, etcetera. By using the information obtained from the experiment one can estimate the parameters of the underlying model, in particular the total number of faults initially present in the software. Certain functions of these model-parameters will yield estimates of other interesting quantities (such as the failure intensity, the reliability, the mean time between failures and the release time). In practice, however, decisions about when to stop testing are rarely based solely on critical values for such quantities. More often, to find an optimal stopping time, the reliability model is extended by associating cost functions, modelling the cost of testing versus the costs of faults in the field. An optimal stopping rule will tell to stop testing as soon as the cost of discovering and fixing the remaining faults is greater than the cost of repairs in the field.

Efforts in describing the evolution of the reliability of computer software during testing resulted in the proposal of dozens of error-counting and debugging models over the past twenty years. Each individual model is completely characterised by a certain set of assumptions. Sometimes, we assume that failures in the software will occur independently and that when a failure is detected, the fault is fixed immediately with no new faults introduced. This is the case for some very well-known models: the Jelinski-Moranda model (JELINSKI and MORANDA, 1972), the Goel-Okumoto model (GOEL and OKUMOTO, 1979) and the Littlewood model (LITTLEWOOD, 1980).

The *Jelinski-Moranda model* is the oldest and one of the most elementary software reliability models introduced so far. In this model 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. In the *Goel-Okumoto model* the failures occur according to a non-homogeneous Poisson process. The failure rate does not depend on the debugging process; it is a simple deterministic function which decreases exponentially in time. Both the Jelinski-Moranda model and the Goel-Okumoto model are in some sense special cases of a more general model, the *Littlewood model*. The main difference with respect to the two previous models is the fact that Littlewood does not assume that each fault makes the same contribution to the

failure intensity. He allows each fault to have its own occurrence probability. Littlewood's argument for this is that larger faults will produce failures earlier than small ones. The way in which Littlewood assigns occurrence rates to the faults, seems rather *ad hoc*. For some real data-sets, Littlewood's model showed a better fit than other models, however.

Of course we are interested in estimators for the model parameters. The maximum likelihood estimation method is generally used for this purpose. With use of Jacod's theory AALEN (1980) derived an expression for the log-likelihood function in a counting process context. Maximisation of the log-likelihood is usually done by setting partial derivatives of this expression to zero and solving the resulting system of highly non-linear likelihood (or score) equations. Apart from numerical problems, one should keep in mind that in many cases the system of likelihood equations will have more than one solution. Classical theorems state that under suitable conditions exactly one of these solutions of the likelihood equations will be consistent and that this consistent solution will be asymptotically normally distributed and efficient. Here the problem arises how to choose from a couple of candidates (solutions of the likelihood equations) the consistent one. More seriously, if one has laboriously numerically determined one solution, how can one be sure there are no others? LE CAM (1990) addresses those kind of problems too. His advice is to just apply one-step of the Newton-Raphson method to an initial estimator, which is $\sqrt{n}$-consistent. It is a well-known result that such an one-step estimator will be asymptotically equivalent to the maximum likelihood estimator. Practical results obtained with the one-step Newton-Raphson method, however, often turn out to be very disappointing. Moreover, it is often actually rather difficult to construct a suitable initial estimator.

In this paper we present a different, more analytical approach, exploiting the mathematical properties of the log-likelihood function itself. The following questions may arise: (1) How to find a global maximiser of the log-likelihood function? (2) Can this likelihood function be multi-peaked? (3) Is the likelihood-maximiser usually the (or an) solution of the likelihood equations? and (4) What can be said about the asymptotic properties of the global maximiser (consistency etcetera)? Our belief is that the ideas and methods developed in this paper could also be of interest for statisticians who try to estimate the parameters of the generalised Pareto distribution (a subject that arises more especially when estimating the extreme value index).

In the next section we introduce Littlewood's parametric software reliability model in more detail. In Section 3 we discuss the maximum likelihood estimation method for counting processes and show that for the software reliability model of Littlewood the likelihood equations can have more than one solution. Hence we face here the problem mentioned earlier. In Section 4 we describe an analytical approach to the problem, exploiting the mathematical properties of the log-likelihood function of the Littlewood model. We will prove that the global maximiser of the likelihood is consistent, and discuss how to find it numerically. Recently we found an alternative way of constructing consistent estimators for the parameters of the Littlewood model. This method, which deals with modified score equation solutions, seems more

generally applicable and will be sketched in Section 5. Finally, in Section 6 we compare some numerical results of the different methods, when applied to simulated data, generated by the Littlewood model. In the appendix we have included for the interested reader a heuristic description of the underlying theory of counting processes.

2   The Littlewood model

We assume that a computer program contains a finite but unknown number of N faults initially. The repairing of a fault takes place immediately after it produces a failure and no new faults are introduced with probability one. Let $n(t)$ denote the number of faults detected at time t. If testing could go on indefinitely, all the N faults would be detected with probability one. Define $T_0 := 0$, let T_i, $i = 1, 2, \ldots, N$, the unordered failure times, while $T_{(i)}$, $i = 1, 2, \ldots, N$, denote the ordered failure times.

In the *Jelinski-Moranda model*, introduced in 1972 and a few years later generalised by MUSA (1975), the failure rate of the program is at any time proportional to the number of remaining faults. So if at some time $t > 0$ already $(i - 1)$ faults have been detected, the failure rate for the ith occurring failure, λ_i, becomes

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

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

$$\lambda^{JM}(t) = \phi_0(N_0 - n(t-)), \quad t \in [0, \tau],\tag{2}$$

where $\lambda(t)$, $t \in [0, \tau]$ denotes the failure rate at time t. As a consequence the unordered failure times T_i, $i = 1, 2, \ldots, N$, are i.i.d exponentials with parameter ϕ_0. The interfailure times, that is the differences $t_i := T_{(i)} - T_{(i-1)}$, are also exponentially distributed, with parameter λ_i given by (1), and mutually independent.

If the computer program is only tested during a specified exposure period $[0, \tau]$, then those failure times T_i that exceed τ will be unobserved. The remaining data-set can be considered as a truncated data-set, since we do not know how large $N_0 - n(\tau)$ is. For the interfailure times $t_i := T_{(i)} - T_{(i-1)}$, $i = 1, \ldots, n(\tau)$, we still have the same distribution (exponential with parameter $\lambda_i = \phi_0(N_0 - i + 1)$), but the unordered T_i follow a truncated exponential distribution.

In the *Littlewood model*, introduced by LITTLEWOOD (1980), it is again 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 $\lambda(t)$. He treats ϕ_j, the failure rate of fault j, as a stochastic variable and suggests a Gamma distribution. So Littlewood assumes that each of the unordered T_i, $i = 1, 2, \ldots, N$ follow a compound distribution. For $j = 1, 2, \ldots, N$, T_j has an exponential distribution with parameter ϕ_j, where ϕ_j itself is random and distributed like $\Gamma(a_0, b_0)$.

One can deduce that the marginal distribution of T_j is given by

$$\Pr(T_j > t) = \left(\frac{b_0}{b_0 + t}\right)^{a_0},$$

that is, the T_j have a generalised Pareto distribution. Furthermore

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

and hence the failure rate of fault j is

$$\frac{a_0}{b_0 + t}.$$

If the computer program is tested during $[0, \tau]$, only $T_{(i)}$, $i = 1, 2, \ldots, n(\tau)$, are observed. 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

$$\lambda^{L}(t) = \frac{a_0(N_0 - n(t-))}{b_0 + t}. \tag{3}$$

By a simple reparametrisation, namely:

$$\alpha_0 = \frac{a_0}{b_0}, \quad \epsilon_0 = \frac{1}{b_0},$$

we get from (3):

$$\lambda^{GL}(t) = \frac{\alpha_0(N_0 - n(t-))}{1 + \epsilon_0 t}, \quad t \in [0, \tau]. \tag{4}$$

Actually formula (4) provides an extension of the Littlewood model (3), allowing also small values of $\epsilon_0 \leqslant 0$. Restricting ourselves to the conventional parameter space

$$\Theta := \{(N, \alpha, \epsilon) \in \mathbb{R}^3 \mid N \geqslant 0, \alpha \geqslant 0, \epsilon \geqslant 0\},$$

which is not compact, we investigate the model behaviour at the boundary of the parameter-set. We will see that letting certain combinations of N, α, ϵ converge to their boundary limits (zero or infinity) at various speeds, this may lead to different limiting models.

Apart from the *null-model (0)* (where, as for instance $N = 0$, nothing happens) we can roughly distinguish four non-trivial boundary models:

(1) *The explosion model* (E^δ):

    If for instance $0 < N < \infty$, $\alpha = \infty$ and $0 < \epsilon < \infty$, the failure-intensity at time zero becomes infinite. Thus the expected number of bugs detected makes a jump from zero to δ at time zero and remains constant for $t > 0$. The failure-intensity drops to zero for $t > 0$. Special cases are E^0 (the null-model), E^N and E^∞. E^4 represents the class of all explosion models $\{E^\delta \mid 0 \leqslant \delta \leqslant \infty\}$.

Table 1. Boundary cases of the Littlewood model (LW)

	$N = 0$			$0 < N < \infty$			$N = \infty$		
	$\epsilon = 0$	$0 < \epsilon < \infty$	$\epsilon = \infty$	$\epsilon = 0$	$0 < \epsilon < \infty$	$\epsilon = \infty$	$\epsilon = 0$	$0 < \epsilon < \infty$	$\epsilon = \infty$
$\alpha = 0$	0	0	0	0	0	0	HP	IP	E^Δ
$0 < \alpha < \infty$	0	0	0	JM	LW	0	E^∞	E^∞	E^Δ
$\alpha = \infty$	0	0	0	E^N	E_N	0 or E^N	E^∞	E^∞	E^Δ

(2) *The Jelinski-Moranda model (JM):*

If $0 < N < \infty$, $0 < \alpha < \infty$ and $\epsilon = 0$, we are dealing with the model we discussed earlier, namely the *Jelinski-Moranda model* (2) with occurrence rate parameter ϕ equal to α. We can therefore treat the Jelinski-Moranda model as a special (limit-)case of the Littlewood model.

(3) *The inhomogeneous Poisson model (IP):*

If $N \to \infty$, $\alpha \to 0$ and $0 < \epsilon < \infty$ such that $N\alpha \to \delta$, the influence of the past of the counting-process $n(t-)$ is eliminated from expression (4) and the general model reduces to an *inhomogeneous Poisson model* with intensity function $\lambda(t) = \delta/(1 + \epsilon t)$, $t \in [0, \tau]$.

(4) *The homogeneous Poisson model (HP):*

If $N \to \infty$, $\alpha \to 0$ and $\epsilon \to 0$ such that $N\alpha \to \delta$, all time-dependence is eliminated from the expression (4) and the general model reduces to a *homogeneous Poisson model* with constant failure intensity δ.

Table 1 shows which of the above mentioned limiting models occur for which (N, α, ϵ). Note that for instance for $0 < N < \infty$, $\alpha \to \infty$ and $\epsilon \to \infty$ the limiting model heavily depends on the way we let α and ϵ increase. In Fig. 1 we plotted expected number of faults detected versus time for various choices of parameter triples (N, α, ϵ) approaching the boundary of the parameter-set Θ. The bold curves represent the limit models.

3 Maximum likelihood estimation for counting processes

By using the information obtained from the test experiment one can estimate the parameters of the underlying model. Mostly maximum likelihood estimation is used for this purpose. AALEN (1978) showed using a theorem of JACOD (1975) that the likelihood function of the vector-parameter θ of the stochastic intensity $\lambda(t)$, associated with a counting process $n(t)$ observed on $[0, t]$, is given by

$$L_t(\theta) := \exp\left(\int_0^t \log \lambda(s; \theta) \, dn(s) - \int_0^t \lambda(s; \theta) \, ds \right).$$

(1) Explosion

(2) Jelinski-Moranda
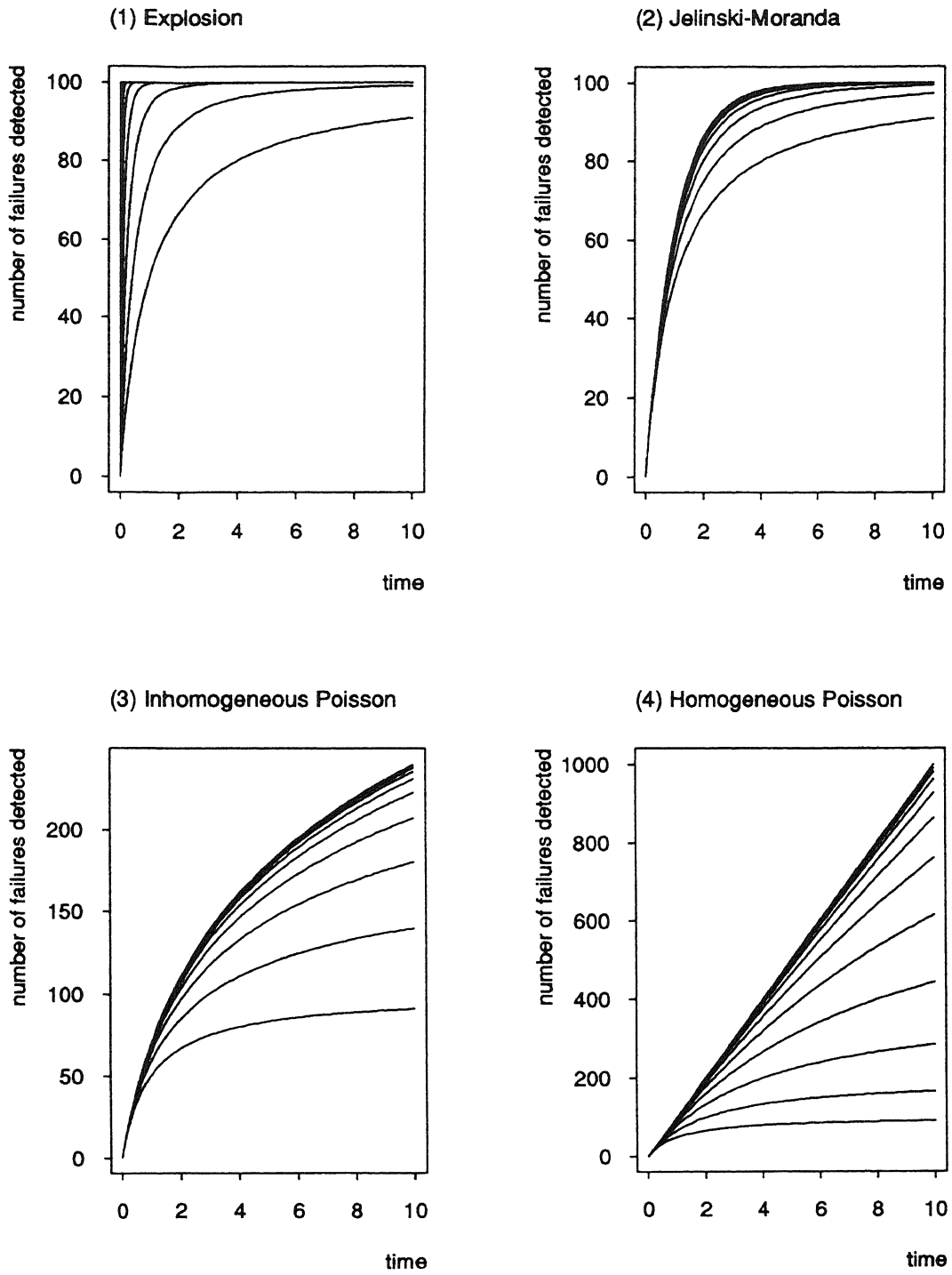

(3) Inhomogeneous Poisson

(4) Homogeneous Poisson

Fig. 1. Boundary cases of the Littlewood model (LW).

For the Littlewood model with intensity (4), parameter $\theta = (N, \alpha, \epsilon)$ and exposure period $t = \tau$, the log-likelihood function becomes

$$\log L_\tau(N, \alpha, \epsilon) := \sum_{i=1}^{n(\tau)} \log\left(\frac{\alpha(N-i+1)}{1+\epsilon T_i}\right) - \sum_{i=1}^{n(\tau)+1} \frac{\alpha}{\epsilon}(N-i+1)\log\left(\frac{1+\epsilon T_i}{1+\epsilon T_{i-1}}\right)$$

$$= n(\tau)\log(\alpha) - \alpha(N-n(\tau))\frac{\log(1+\epsilon\tau)}{\epsilon}$$

$$+ \sum_{i=1}^{n(\tau)} \log(N-i+1) - (\alpha+\epsilon)\sum_{i=1}^{n(\tau)} \frac{\log(1+\epsilon T_i)}{\epsilon}, \tag{5}$$

where $T_0 := 0$ and $T_{n(\tau)+1} := \tau$. Hence the log-likelihood equations are

$$\frac{\partial}{\partial N}\log L_\tau(N, \alpha, \epsilon) = \sum_{i=1}^{n(\tau)} \frac{1}{N-i+1} - \alpha\frac{\log(1+\epsilon\tau)}{\epsilon} = 0, \tag{6}$$

$$\frac{\partial}{\partial\alpha}\log L_\tau(N, \alpha, \epsilon) = \frac{n(\tau)}{\alpha} - (N-n(\tau))\frac{\log(1+\epsilon\tau)}{\epsilon} - \sum_{i=1}^{n(\tau)}\frac{\log(1+\epsilon T_i)}{\epsilon} = 0, \tag{7}$$

$$\frac{\partial}{\partial\epsilon}\log L_\tau(N, \alpha, \epsilon) = \frac{\alpha(N-n(\tau))}{\epsilon^2}\left(\log(1+\epsilon\tau) - \frac{\epsilon\tau}{1+\epsilon\tau}\right)$$

$$+ \frac{\alpha}{\epsilon^2}\sum_{i=1}^{n(\tau)}\left(\log(1+\epsilon T_i) - \frac{\epsilon T_i}{1+\epsilon T_i}\right) - \sum_{i=1}^{n(\tau)}\frac{T_i}{1+\epsilon T_i} = 0. \tag{8}$$

REMARK 1. Natural questions to ask are whether $\log L_\tau(N, \alpha, \epsilon)$, as defined by (5), might have more than one (local) maximum, and whether the system of likelihood equations (6)–(8) might have more than one solution, which is a different problem. The answer to both questions is affirmative. In Remark 4 we will present a data-set for which $\log L_\tau(N, \alpha, \epsilon)$ has a global maximum at the boundary ($N = n(\tau)$ and $\epsilon = 0$) and both a local maximum and a saddle-point in the interior of the parameter-set Θ. □

REMARK 2. We shall investigate the asymptotic behaviour of the ML-estimators for (α, N, ϵ) when the parameters of the counting process evolve in the following way: α_0 and ϵ_0 remain fixed, while $\delta_0 := N_0\alpha_0$ (the intensity of the process at time $t = 0$) goes to infinity. □

For the Jelinski-Moranda model MOEK (1983) gave a criterion on the data, satisfied with probability one when the model is true, under which there exists an unique solution of the maximum likelihood equations. For the model of Littlewood such a criterion, however, is not known and probably will not exist. With use of asymptotic theory it is proved (VAN PUL, 1992) that in case of one or more solutions of the likelihood equations exactly one of them will be consistent, if we can make this choice using the data only. Here the earlier mentioned problem arises: how to

choose a consistent MLE, when the likelihood equations have more than one solution. Also alternative estimation methods do have the same difficulties with the Littlewood model. In case of M-estimation, nonsingularity of the resulting matrix has unfortunately not been proved yet (see GEURTS et al. (1988)).

As the system of the three highly non-linear likelihood equations (6)–(8) obviously cannot be solved analytically, MOEK (1984) and GEURTS et al. (1988), using the parametrisation (3), suggest to simplify the problem by fixing one of the parameters $a = 1$. Furthermore, Moek was able to express both N and a as functions of b and used this to derive an equation $f(b) = 0$ from which b can be solved numerically relatively easy. Criteria were formulated which ensure the existence of at least one solution. The problem of the possibility of multiple solutions is ignored. GEURTS et al. signal this problem but do not provide a solution. They state that in case of multiple solutions exactly one of them will be consistent; moreover this consistent solution is asymptotically normal distributed and efficient.

In BARENDREGT and VAN PUL (1991) various approaches are presented to construct estimates that are indeed consistent. A first attempt describes how an initial estimator can be obtained that is $\sqrt{n}$-consistent. Starting with this initial estimator an application of the one-step Newton-Raphson method (or an other suitable iterative optimisation procedure) will then provide us with an estimator that is asymptotically equivalent with the MLE. Practical results obtained with these likelihood equations based estimators turned out to be disappointing, especially for small N_0. In Section 4 of this paper an other, more analytical approach is presented which reduces the problem of maximising the log-likelihood function $\log L_\tau(N, \alpha, \epsilon)$, given in (5), to a one-dimensional one by eliminating first the parameter α explicitly and then the parameter N implicitly. A third approach, described in Section 5, is based on finding solutions of a slightly modified system of score equations.

4 Exploiting the mathematical properties of the likelihood function

For sake of convenience we apply the parameter-transformation

$$M := N - n(\tau).$$

The log-likelihood, as expressed in the new parameters, is given by

$$\log L_\tau(\alpha, M, \epsilon) = n(\tau) \log \alpha - (\epsilon + \alpha) \sum_{i=1}^{n(\tau)} \frac{\log(1 + \epsilon T_i)}{\epsilon} - \alpha M \frac{\log(1 + \epsilon\tau)}{\epsilon}$$

$$+ \sum_{i=1}^{n(\tau)} \log(M + i) \tag{9}$$

with parameter-domain $\{(\alpha, M, \epsilon): \alpha > 0, M \geqslant 0, \epsilon \geqslant 0\}$. When M and ϵ are kept fixed, it is quite easy to maximise $\log L_\tau(\alpha, M, \epsilon)$; it can be done analytically. If we define the profile-likelihood

$$R_2(M, \epsilon) := \sup_{\alpha > 0} \log L_\tau(\alpha, M, \epsilon);$$

then

$$\sup_{\alpha > 0, M \geqslant 0, \epsilon \geqslant 0} \log L_\tau(a, M, \epsilon) = \sup_{M \geqslant 0, \epsilon \geqslant 0} R_2(M, \epsilon).$$

Now

$$\frac{\partial}{\partial \alpha} \log L_\tau(\alpha, M, \epsilon) = \frac{n(\tau)}{\alpha} - \sum_{i=1}^{n(\tau)} \frac{\log(1 + \epsilon T_i)}{\epsilon} - M \frac{\log(1 + \epsilon \tau)}{\epsilon}$$

and it is easy to check that $\log L_\tau(\alpha, M, \epsilon)$ for M, ϵ fixed, is maximal for

$$\alpha = n(\tau) \left[\sum_{i=1}^{n(\tau)} \frac{\log(1 + \epsilon T_i)}{\epsilon} + M \frac{\log(1 + \epsilon \tau)}{\epsilon} \right]^{-1}; \tag{10}$$

so

$$R_2(M, \epsilon) = n(\tau) \log n(\tau) - n(\tau) \log \left[\sum_{i=1}^{n(\tau)} \frac{\log(1 + \epsilon T_i)}{\epsilon} + M \frac{\log(1 + \epsilon \tau)}{\epsilon} \right]$$

$$- \sum_{i=1}^{n(\tau)} \log(1 + \epsilon T_i) + \sum_{i=1}^{n(\tau)} \log(M + i). \tag{11}$$

In order to maximise $R_2(M, \epsilon)$ we will first maximise with respect to M, keeping ϵ fixed, and then maximise with respect to ϵ. The first maximisation is relatively easy because we can make use of the following theorem:

THEOREM 1. *For n integer and $\geqslant 2$, and η real and > 0, the function z, where*

$$z(M) := -n \log(M + \eta) + \sum_{i=1}^{n} \log(M + i)$$

has precisely one local supremum, which is realised at:

- $M = 0$, *for* $\eta \leqslant n[\sum_{i=1}^{n} 1/i]^{-1}$,
- $M = \infty$, *for* $\eta \geqslant (n + 1)/2$,
- $M \in (0, \infty)$, *for intermediate values of η.*

PROOF. It is easy to check that

$$\dot{z}(M) = \frac{d}{dM} g(M) = \frac{-n}{M + \eta} + \sum_{i=1}^{n} \frac{1}{M + i},$$

$$\ddot{z}(M) = \frac{d^2}{dM^2} g(M) = \frac{n}{(M + \eta)^2} - \sum_{i=1}^{n} \left(\frac{1}{M + i} \right)^2.$$

If for a finite value M' of M

$$\dot{z}(M') = 0, \tag{12}$$

then

$$\ddot{z}(M') = \frac{n}{(M' + \eta)^2} - \sum_{i=1}^{n} \left(\frac{1}{M' + i} \right)^2 = \frac{1}{n} \left(\sum_{i=1}^{n} \frac{1}{M' + i} \right)^2 - \sum_{i=1}^{n} \left(\frac{1}{M' + i} \right)^2 < 0, \tag{13}$$

since $n \geqslant 2$. So there is at most one finite value M' for M for which (12) holds, as if there were two of them, say $M'_1 < M'_2$, then we would have $\ddot{z}(M) \geqslant 0$ for either one of them, or there would be an intermediate value M'_3, $M'_1 < M'_3 < M'_2$ for which $\dot{z}(M'_3) = 0$, $\ddot{z}(M'_3) \geqslant 0$. Both options, however, are excluded by (13).

- *Case 1:* $\eta \leqslant n[\Sigma^n_{i=1} 1/i]^{-1}$.

It is easy to check that $\dot{z}(0) \leqslant 0$. From this and (13) it follows immediately that $\dot{z}(M) < 0$ for all $M > 0$. Hence $M = 0$ provides the unique local maximum of $z(M)$.

- *Case 2:* $\eta \geqslant (n + 1)/2$.

We find that for all $M \geqslant 0$:

$$\dot{z}(M) = \frac{-n}{M + \eta} + \sum_{i=1}^{n} \frac{1}{M + i} \geqslant \frac{-n}{M + \dfrac{n + 1}{2}} + \sum_{i=1}^{n} \frac{1}{M + i} \geqslant 0.$$

The last inequality follows from Jensen's inequality applied on the function $x \to 1/x$. So the supremum is reached at $M = \infty$.

- *Case 3:* $n[\Sigma^n_{i=1} 1/n]^{-1} < \eta(< n + 1)/2$.

We have $\dot{z}(0) > 0$, but

$$\lim_{M \to \infty} M^2 \dot{z}(M) = n\eta - \sum_{i=1}^{n} i < 0.$$

So there exists a M_0, $0 < M_0 < \infty$ such that for all $M > M_0$ we have $M^2\dot{z}(M) < 0$ and hence also $\dot{z}(M) < 0$. Thus again there exists a finite positive value M' for which $\dot{z}(M') = 0$ and the unique local maximum of the function $z(M)$ is realised. This completes the proof of Theorem 1. $\qquad\square$

Now $R_2(M, \epsilon)$ can be rewritten as

$$R_2(M, \epsilon) = n(\tau) \log n(\tau) - n(\tau) \log\left(\frac{\log(1 + \epsilon\tau)}{\epsilon}\right) - n(\tau) \log(M + \eta(\epsilon))$$

$$+ \sum_{i=1}^{n(\tau)} \log(M + i) - \sum_{i=1}^{n(\tau)} \log(1 + \epsilon T_i), \tag{14}$$

where

$$\eta(\epsilon) := \sum_{i=1}^{n(\tau)} \frac{\log(1 + \epsilon T_i)}{\log(1 + \epsilon\tau)}.$$

Only the third and the fourth of the terms in the right-hand side of (14) depend on M. Their sum has a form similar to the function $h(M)$ of Theorem 1, and so $R_2(M, \epsilon)$ has for ϵ fixed, exactly one local supremum, which occurs for $M = 0$ if $\eta(\epsilon) \leqslant n[\Sigma^{n(\tau)}_{i=1} 1/i]^{-1}$, for $M = \infty$ if $\eta(\epsilon) \geqslant (n + 1)/2$ and for a finite, positive value of M otherwise.

REMARK 3. $\tilde{M}(\epsilon)$, the value of M which maximises $R_2(M, \epsilon)$, cannot be expressed explicitly, contrary to the α of (10). If ϵ increases, $\eta(\epsilon)$ increases as well, and so will $M(\epsilon)$.   □

Finally then, if we define

$$R_1(\epsilon) := R_2(\tilde{M}(\epsilon), \epsilon),$$

we have to maximise $R_1(\epsilon)$. While doing this, one must bear in mind that $R_1(\epsilon)$ might reach its maximum for $\epsilon = 0$. As a matter of fact, we have encountered this feature several times when analyzing real data-sets. All maximisation results are easily obtained by a standard optimisation procedure for one-dimensional functions, called *Golden section search* (see for instance VETTERLING et al., 1985).

REMARK 4. There exist data-sets for which $R_1(\epsilon)$ has more than one local maximum and as a consequence $\log L(\alpha, M, \epsilon)$ has more than one local maximum. An example of such a data-set is:

$$\tau = 709.5, \quad n(\tau) = 3;$$

$$T_1 = 1, \quad T_2 = 399.9, \quad T_3 = 400.1.$$

The function $R_1(\epsilon)$ takes its global maximum for $\epsilon = 0$, a second local maximum for $\epsilon = 1.152$ and a local minimum for $\epsilon = 0.023$. This corresponds to a global maximum for $R_2(M, \epsilon)$ for $\epsilon = 0$, $M = 0$, a second local maximum $\epsilon = 1.152$, $M = 9.8$, and a saddle-point at $\epsilon = 0.023$, $M = 0.0015$. The data-set is of course in extremely bad agreement with the Littlewood model. It was constructed after mathematical analysis rather than encountered while analyzing data.   □

THEOREM 2. *The global maximiser of the log-likelihood function for the Littlewood model (see expressions (5) or (9)) is consistent, when* $\delta_0 = N_0 \alpha_0$ *goes to infinity (see Remark 2).*

PROOF. It will turn out to be of great advantage to apply the parameter-transformation

$$\zeta := \frac{\alpha \log (1 + \epsilon \tau)}{\epsilon}. \tag{15}$$

With the new parameters N, ϵ and ζ, we can rewrite the log-likelihood as follows:

$$\log L_\tau(N, \epsilon, \zeta) = n(\tau) \log \frac{\zeta}{1 - e^{-\zeta}} - \zeta \sum_{i=1}^{n(\tau)} \frac{\log (1 + \epsilon t_i)}{\log (1 + \epsilon \tau)}$$

$$+ n(\tau) \log \frac{\epsilon}{\log (1 + \epsilon \tau)} - \sum_{i=1}^{n(\tau)} \log (1 + \epsilon t_i)$$

$$+ n(\tau) \log (1 - e^{-\zeta}) - (N - n(\tau))\zeta + \sum_{i=1}^{n(\tau)} \log (N - i + 1)$$

$$= \sum_{i=1}^{n(\tau)} \log h(t_i; \epsilon, \zeta) + n(\tau) \log (1 - e^{-\zeta})$$

$$- (N - n(\tau))\zeta + \sum_{i=1}^{n(\tau)} \log (N - i + 1), \tag{16}$$

where

$$h(t; \epsilon, \zeta) := \frac{\zeta}{1 - e^{-\zeta}} \exp\left(-\zeta \frac{\log (1 + \epsilon t)}{\log (1 + \epsilon \tau)} \right) \frac{\epsilon}{\log (1 + \epsilon \tau)} \frac{1}{1 + \epsilon t}.$$

It is easy to check that $h(t; \epsilon, \zeta)$ is continuous on $[0, \tau]$ and for all $\epsilon \geqslant 0$, $\zeta \geqslant 0$:

$$\int_0^\tau h(s; \epsilon, \zeta) \, ds = 1.$$

Hence $h(t; \epsilon, \zeta)$ can be considered as the continuous density function of a stochastic variable with values in $[0, \tau]$. We are now going to eliminate N from the log-likelihood by maximising (16) with respect to N for fixed ϵ and ζ. Let $\hat{N}(\zeta)$ be that value of $N \geqslant n(\tau)$, for which

$$R(N, \zeta) := n(\tau) \log (1 - e^{-\zeta}) - (N - n(\tau))\zeta + \sum_{i=1}^{n(\tau)} \log (N - i + 1)$$

is maximal and define

$$R(\zeta) := n(\tau) \log (1 - e^{-\zeta}) - (\hat{N}(\zeta) - n(\tau))\zeta + \sum_{i=1}^{n(\tau)} \log (\hat{N}(\zeta) - i + 1).$$

For large values of $n(\tau)$, $R(\zeta)$ is approximately $n(\tau) \log n(\tau) - n(\tau) + \zeta/2$, but we will not need this. We will only use that:

$$0 \leqslant \frac{\partial}{\partial \zeta} R(\zeta) \leqslant 1. \tag{17}$$

Now the log-profile-likelihood for (ϵ, ζ) becomes

$$LPL(\epsilon, \zeta) := \sum_{i=1}^{n(\tau)} \log h(t_i; \epsilon, \zeta) + R(\zeta).$$

As in our asymptotic approach $N_0 \to \infty$ implies $n(\tau) \to \infty$ with probability 1, we consider the sequence $(T_i)_{i \geqslant 1}$ of i.i.d. random variables with density $h(t, \epsilon_0, \zeta_0)$. We will prove that the corresponding sequence of profile-log-likelihood maximisers $(\hat{\epsilon}_n, \hat{\zeta}_n)$ converges to (ϵ_0, ζ_0) with probability 1. With $\Omega = (\epsilon, \zeta): \epsilon \geqslant 0, \zeta \geqslant 0\}$, the density function $h(t; \epsilon, \zeta)$ satisfies all eight assumptions of WALD (1949). The verification of the validity of these assumptions is straightforward, except perhaps for Assumptions II and V. Now it follows from a theorem of WALD (1949) that for all closed subsets ω of Ω, not containing the true value (ϵ_0, ζ_0), we have

$$P\left(\lim_{n \to \infty} \sup_{(\epsilon, \zeta) \in \omega} \left(\sum_{i=1}^{n} \log h(t_i; \epsilon, \zeta) - \sum_{i=1}^{n} \log h(t_i; \epsilon_0, \zeta_0) \right) = -\infty \right) = 1.$$

Some calculations yield that with probability 1 there exists a N_1 such that for $n > N_1$:

$$\hat{\zeta}_n < \frac{4\tau}{\mathrm{E}T},$$ (18)

where $\mathrm{IE}T$ denotes the expectation of the stochastic variable with density $h(t; \epsilon_0, \zeta_0)$. As $LPL(\hat{\epsilon}_n, \hat{\zeta}_n) \geqslant LPL(\epsilon_0, \zeta_0)$ it follows that for $n > N_1$:

$$\sum_{i=1}^{n} \log h(t_i; \hat{\epsilon}_n, \hat{\zeta}_n) - \sum_{i=1}^{n} \log h(t_i; \epsilon_0, \zeta_0) \geqslant R(\zeta_0) - R(\hat{\zeta}_n)$$

$$\geqslant \min (0, \zeta_0 - \hat{\zeta}_n) \geqslant -\frac{4\tau}{\mathrm{E}T}.$$

Let $\eta > 0$ be arbitrary. Suppose the sequence $(\hat{\epsilon}_n, \hat{\zeta}_n)$ has an accumulation point at a distance larger than η of (ϵ_0, ζ_0). So for infinitely many n:

$$\sup_{|(\epsilon,\zeta) - (\epsilon_0,\zeta_0)| > \eta} \sum_{i=1}^{n} \log h(t_i; \epsilon, \zeta) \geqslant \sum_{i=1}^{n} \log h(t_i; \hat{\epsilon}_n, \hat{\zeta}_n)$$

and hence

$$\sup_{|(\epsilon,\zeta) - (\epsilon_0,\zeta_0)| > \eta} \sum_{i=1}^{n} \log h(t_i; \epsilon, \zeta) \geqslant \sum_{i=1}^{n} \log h(t_i; \epsilon_0, \zeta_0) - \frac{4}{\mathrm{E}T}$$

for infinitely many n. But this event has according to Wald's theorem probability zero. Therefore all accumulation points of $(\hat{\epsilon}_n, \hat{\zeta}_n)$ are within distance η of (ϵ_0, ζ_0). As η was chosen arbitrarily, it follows that

$$\mathrm{P}\left(\lim_{n \to \infty} (\hat{\epsilon}_n, \hat{\zeta}_n) = (\epsilon_0, \zeta_0)\right) = 1.$$

This yields the consistency of $(\hat{\epsilon}_n, \hat{\zeta}_n)$. Of course $\hat{N}(\hat{\zeta}_n)$ itself is not consistent. We can prove, however, as

$$0 < \frac{n(\tau)}{1 - \mathrm{e}^{-\zeta}} - \hat{N}(\zeta) < 1,$$

that

$$\frac{\hat{N}(\hat{\zeta}_n)}{N_0} \to_{a.s} \frac{\mathrm{E}(n(\tau))}{N_0} \frac{1}{1 - \mathrm{e}^{-\zeta}} = 1.$$

Full details of the verification of (17), (18) and Assumptions II and V of Wald can be found in BARENDREGT and VAN PUL (1993). This completes the proof of Theorem 2.   ☐

5 Investigating solutions of modified score equations

Recently, the proof of Theorem 2 inspired us to an alternative way of constructing estimators for the model parameters of the Littlewood model. We believe this method, which is in the spirit of Le Cam's idea to investigate slightly modified but more easily tractable score equations, will be more generally applicable.

We consider again the counting process $n(t)$ with failure intensity according to the Littlewood model:

$$\lambda(t) := \frac{\alpha_0(N_0 - n(t-))}{1 + \epsilon_0 t}.$$

The log-likelihood function in theorems of N, α and ϵ is given by (see also (5))

$$\log L_\tau(N, \alpha, \epsilon) = \sum_{i=1}^{n(\tau)} \log(\alpha(N - i + 1)) - \left(1 + \frac{\alpha}{\epsilon}\right) \sum_{i=1}^{n(\tau)} \log(1 + \epsilon t_i)$$

$$- \frac{\alpha(N - n(\tau))}{\epsilon} \log(1 + \epsilon\tau). \tag{19}$$

In this section we will construct approximate maximum likelihood estimators $(\bar{N}, \bar{\alpha}, \bar{\epsilon})$, sometimes called *modified score equation solutions*, and prove the consistency of $(\bar{\alpha}, \bar{\epsilon})$ and the convergence (in probability) to zero of var $(\bar{N})/\bar{N}^2$, when N_0 (the true value of N) tends to infinity. We rewrite the log-likelihood as follows:

$$\log L_\tau(N, \alpha, \epsilon) = \sum_{i=1}^{n(\tau)} \log g(t_i; \alpha, \epsilon) + n(\tau) \log\left(\frac{1 - (1 + \epsilon\tau)^{-\alpha/\epsilon}}{\alpha}\right)$$

$$+ \sum_{i=1}^{n(\tau)} \log(\alpha(N - i + 1)) - \frac{\alpha(N - n(\tau))}{\epsilon} \log(1 + \epsilon\tau), \tag{20}$$

$$g(t; \alpha, \epsilon) := \frac{\alpha(1 + \epsilon t)^{-\alpha/\epsilon - 1}}{1 - (1 + \epsilon\tau)^{-\alpha/\epsilon}}, \quad 0 \leqslant t \leqslant \tau, \tag{21}$$

Thus $g(t; \alpha, \epsilon)$ is the density function of a continuous stochastic variable with values in $[0, \tau]$. This is true for all parameter combinations (α, ϵ) with $\alpha \geqslant 0$ and $\epsilon \geqslant 0$. Degeneration of the probability distribution only occurs if at least one of the parameters α and ϵ takes the value ∞. For $\alpha > 0$, $g(t; \alpha, \epsilon)$ can be viewed as the density function of a generalised Pareto distributed variable, that is truncated at τ. The three remaining terms on the right-hand side of (20) can be compared with $R(N, \zeta)$, defined in the proof of theorem 2. We found that $R(N, \zeta)$ was of secondary importance once N was replaced by $\hat{N}(\zeta)$, because of its limited variability (see (17)). This led us to estimate (α, ϵ) by maximising the leading term of the log-likelihood (20) only:

$$\sum_{i=1}^{n(\tau)} \log g(t_i; \alpha, \epsilon). \tag{22}$$

The approximate estimator $(\bar{\alpha}, \bar{\epsilon})$ we obtain in this way is asymptotically consistent, even if we take for the parameter-set $\Omega = \{(\alpha, \epsilon): \alpha \geqslant 0, \epsilon \geqslant 0\}$. This follows directly as the truncated Pareto density $g(t; \alpha, \epsilon)$ in (21) satisfies all eight assumptions of WALD (1949). Verification of this is, again apart from conditions (II) and (V) perhaps, straightforward. Estimation of N could be performed by substituting $\bar{\alpha}$ and $\bar{\epsilon}$ in the remaining terms of the log-likelihood (20):

$$n(\tau) \log\left(\frac{1 - (1 + \epsilon\tau)^{-\alpha/\epsilon}}{\alpha}\right) + \sum_{i=1}^{n(\tau)} \log(\alpha(N - i + 1)) - \frac{\alpha(N - n(\tau))}{\epsilon} \log(1 + \epsilon\tau) \tag{23}$$

and maximisation of (23) with respect to N. This would yield

$$\bar{N} := \frac{n(\tau)}{1 - (1 + \bar{\epsilon}\tau)^{-\bar{\alpha}/\bar{\epsilon}}} - \frac{1}{2} + O\left(\frac{1}{n(\tau)}\right). \tag{24}$$

We will estimate N by neglecting the last two terms in the right hand side of (24). Stated more generally, the idea is to regard the failure times T_i, conditional on $n(\tau)$, as the ordered outcomes of $n(\tau)$ independent variables from some parametric distribution function $F(t, \psi)$. For the Littlewood model this distribution function is the generalised *Pareto* (t, α, ϵ) distribution. Conditioned on $\psi_0 = (\alpha_0, \epsilon_0)$, $n(\tau)$ has a binomial distribution with parameters N_0 and $F(\tau, \psi_0)$. We define $\bar{\psi}$ and $\bar{N}$ as the roots of

$$\frac{\partial}{\partial \psi} \sum_{i=1}^{n(\tau)} \log \frac{f(T_i, \psi)}{F(\tau, \psi)} = 0, \tag{25}$$

$$n(\tau) - NF(\tau, \bar{\psi}) = 0. \tag{26}$$

Although this choice seems rather *ad hoc*, a closer examination yields that the system of equations (25)–(26) is asymptotically equivalent to the system of likelihood equations. Let $\hat{U}_v$ be the system of likelihood equations (6)–(8) and let $\bar{U}_v$ be the vector of left-hand sides of (25)–(26). So $\hat{U}_v(\hat{\theta}_v) = 0$ and $\bar{U}_v(\bar{\theta}_v) = 0$ and furthermore we have

$$\bar{U}_v(\theta) = A_v(\theta)\hat{U}_v(\theta) + B_v(\theta),$$

where $A_v \to A$, A non-singular and $\sqrt{n}B_n \to 0$. Under weak boundedness and smoothness conditions on A and B the asymptotic properties of $\hat{\theta}$ are transferred to $\bar{\theta}$.

6 Comparison of some numerical results

In this section we will discuss the results of some simulation experiments. We generated failure times according to the Littlewood model (4) with $\alpha_0 = 1$, $\epsilon_0 = 1$ and different values of N_0 (100, 1000 and 10000). The exposure period was kept fixed: $\tau = 4$. We compare the global maximiser of the likelihood $\hat{\theta}$ (described in section 4) and the modified score equations solution $\bar{\theta}$ (suggested in section 5). For each value of N_0 we repeated the simulation experiments and parameter estimations $K_0 = 200$ times and computed $\hat{\theta}$ and $\bar{\theta}$. Mean square errors for these estimators are given in Table 2.

Table 2.   Mean square errors for $\hat{\theta}$ and $\bar{\theta}$ with (a) $\theta_0 = (100, 1, 1)$, (b) $\theta_0 = (1000, 1, 1)$, (c) $\theta_0 = (10000, 1, 1)$

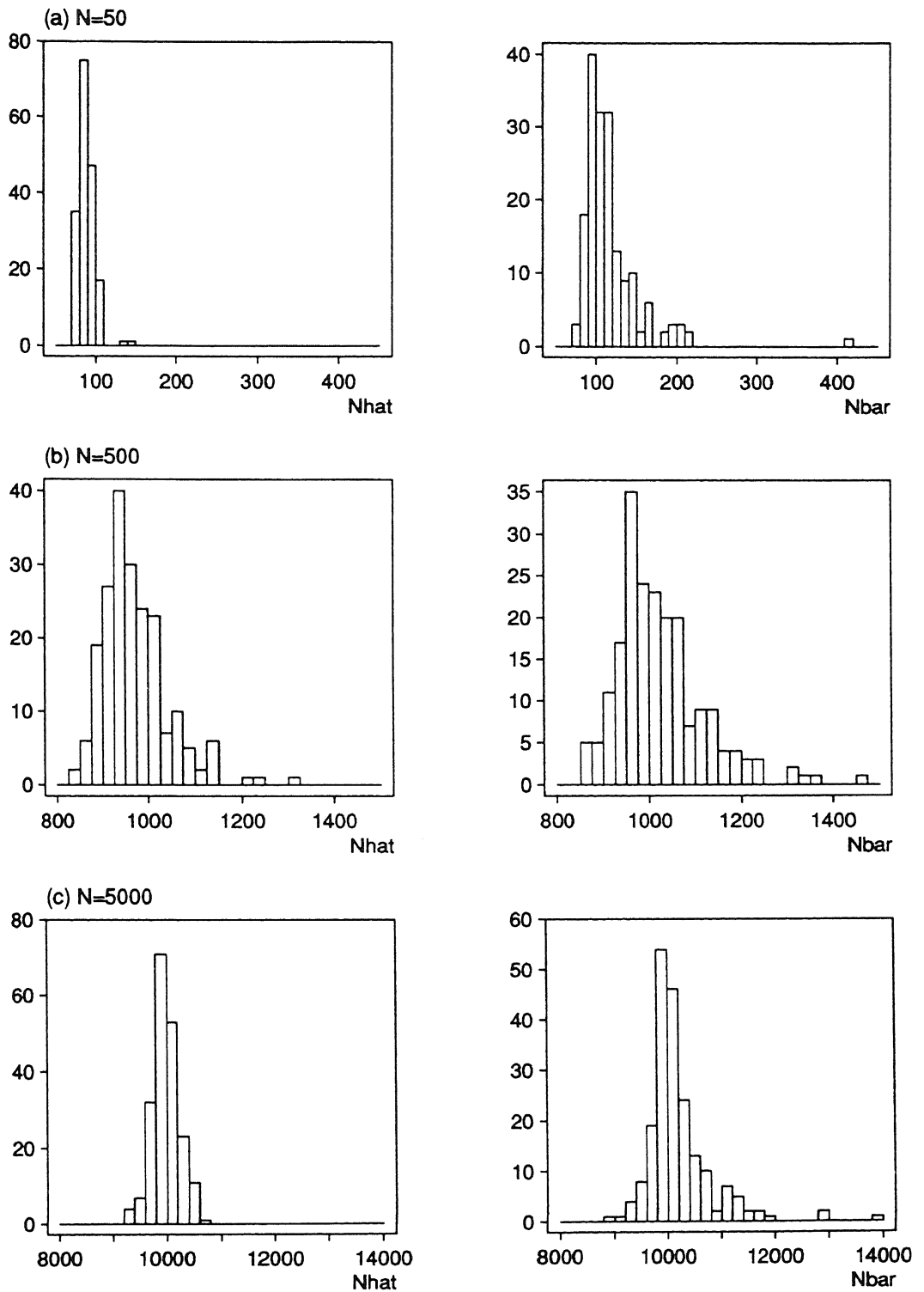|	(a) $N_0 = 100$	(b) $N_0 = 1000$	(c) $N_0 = 10000$
$\hat{N}$	0.0237	0.0060	0.0006
$\hat{\alpha}$	0.0461	0.0052	0.0007
$\hat{\epsilon}$	0.5767	0.1322	0.0101
$\bar{N}$	0.1601	0.0099	0.0041
$\bar{\alpha}$	0.0644	0.0075	0.0017
$\bar{\epsilon}$	0.5473	0.1517	0.0545

Fig. 2. Histograms for $\hat{N}$ and $\bar{N}$ with (a) $N_0 = 100$, (b) $N_0 = 1000$, (c) $N_0 = 10000$.

From these figures we may draw several conclusions. All estimators become better N_0 gets larger. The is not surprising as we have more statistical information. α can be estimated relatively well already for small N_0; this in contrary to ϵ which has for $N_0 = 10000$ still a large range. The differences in the value of the likelihood function are in some cases very small, indicating the extreme flatness of the likelihood in the neighbourhood of its maximum. Different triples can lead to almost the same curve (up to τ) and likelihood. Predictions of events after τ will differ, of course. In a few cases the likelihood function takes its absolute maximum on the boundary $\epsilon = 0$. We checked that in these cases there are no other (local) maxima. The data here falsely suggests the Jelinski-Moranda model.

Appendix: Counting process theory and maximum likelihood estimation

We are going to model the occurrence of discrete, random events in continuous time. We fix $\mathcal{T} = [0, \tau]$ for a given finite terminal time τ, $0 < \tau < \infty$. Note that we are observing a non-deterministic process through the fixed time window $\mathcal{T}$. The fact that the number of faults detected in $\mathcal{T}$ will be stochastic is the reason why we cannot use classical maximum likelihood theory for i.i.d. observations in deriving asymptotic results. Therefore we introduce a powerful mathematical instrument which we will use to solve these problems: the theory of counting processes and martingales. For a complete summary we refer to ANDERSEN and BORGAN (1985), JACOD and SHIRYAEV (1987) or ANDERSEN et al. (1993).

A counting process n is a stochastic process which can be thought of as registering the occurrences in time of a number of discrete events. More formally, a *counting process* is a stochastic process, zero at time zero, with piecewise constant and non-decreasing paths, having jumps of size one only. We say that n has *intensity process* λ, if λ is a predictable process and the process m, defined by

$$m(t) := n(t) - \int_0^t \lambda(s)\, ds,\tag{A1}$$

satisfies the martingale property (A2):

$$\mathrm{E}(m(t)\,|\,m(u), 0 \leqslant u \leqslant s) = m(s), \quad s \leqslant t.\tag{A2}$$

That is, the increment of the stochastic process $m(t)$ over an arbitrary time interval $(t, t + h]$ given the past has zero expectation. The integral in the right-hand side of (A1) is often referred to as the *cumulative intensity process* or *compensator* of n. We can consider a martingale as being a pure noise process. The systematic part of a counting process is its compensator, a smoothly varying and predictable process, which, if subtracted from the counting process, leaves unpredictable zero-mean noise.

Martingales have been studied intensively during the past few decades and a lot of nice mathematical properties have been derived by now. Some very important martingale results are Kurtz' theorems, Lenglart's inequality and the Martingale Central Limit Theorem (MCLT), which can be seen as analogues of the law of large

numbers and the usual Central Limit Theorem in the classical i.i.d. case. These results will be essentially in the proofs of convergence in probability and weak convergence for the non-i.i.d. case. For a comprehensive treatment of these and other martingale results we refer to ANDERSEN et al. (1993).

We observe the counting process $n(t)$ on $[0, \tau]$ with underlying intensity process λ. We assume that this stochastic intensity function, depending on the past of n, is a member of some parametric family:

$$\lambda(t) = \lambda(t, \theta, n(t-)), \quad \theta \in \Theta \subseteq R^k. \tag{A3}$$

We assume the true parameter-value θ_0 is contained in the interior of Θ. In all typical cases $\theta_0 = (N_0, \psi_0)$, where N_0, the parameter of most interest, represents the scale or the size of the problem (sometimes $N_0 = n(\infty)$), while ψ_0 denotes a nuisance vector parameter.

The question is now, of course, how to find estimators for N and ψ. We will use the method of Maximum Likelihood Estimation (MLE) for this purpose. Using the fact that $\lambda(t)\, dt$ represents the conditional probability given the strict past that the counting process $n(t)$ jumps in the interval $[t, t + dt]$, we can write for the likelihood:

$$L_\tau(\theta) \propto \prod_{0 < t < \tau} ((\lambda(t, \theta)\, dt)^{dn(t)}(1 - \lambda(t, \theta)\, dt)^{1 - dn(t)})$$

$$\propto \exp\left(\int_0^\tau \log \lambda(t, \theta)\, dn(t) - \int_0^\tau \lambda(t, \theta)\, dt\right). \tag{A4}$$

For a standard definition of the product integral in the upper expression of (A4) we refer to GILL and JOHANSEN (1990). The lower expression in (A5) is also known as Jacod's formula (ANDERSEN et al., 1993).

Maximization of expression (A4) is usually done by setting partial derivatives of the log-likelihood to zero and solving the resulting system of highly non-linear likelihood (or score) equations:

$$\frac{\partial}{\partial N} \log L_\tau(N, \psi) = 0, \tag{A5}$$

$$\frac{\partial}{\partial \psi} \log L_\tau(N, \psi) = 0. \tag{A6}$$

We have assumed in (A5) that the model is also meaningful for non-integer N. The direct algebraic solution of the system of non-linear equations (A5)–(A6) is usually impossible. The best we can realistically hope for is to solve these equations for a subset of the parameters in terms of the remaining parameters. The remaining parameters are then estimated using numerical methods.

References

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

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

AKIYAMA, F. (1971), An example of software system debugging, *Information Processing 71,* North-Holland, New York, 353–359.

ANDERSEN, P. K. and Ø. BORGAN (1985), Counting processes models for life history data: a review, *Scandinavian Journal of Statistics 12,* 97–158.

ANDERSEN, P. K., Ø. BORGAN, R. D. GILL and N. KEIDING (1993), *Statistical methods based on counting processes,* Springer-Verlag, New York.

BARENDREGT, L. and M. C. VAN PUL (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.

BARENDREGT, L. and M. C. VAN PUL (1993), The Littlewood model in software reliability revisited, Private communication.

GEURTS, W. A. J., M. M. A. HASSELAAR and 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.

GILL, R. D. and S. JOHANSEN (1990), A survey of product-integration with a view towards application in survival analysis, *Annals of Statistics 18,* 1501–1555.

GOEL, A. L. and K. OKUMOTO (1979), Time dependent error-detection rate model for software reliability and other performance measures, *IEEE Transactions on Reliability 28,* 206–211.

HALSTEAD, M. H. (1977), *Elements of software science,* Elsevier, New York.

JACOD, J. (1975), Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 31,* 235–253.

JACOD, J. and A. N. SHIRYAEV (1987), *Limit theorems for stochastic processes,* Springer-Verlag, Berlin.

JELINSKI, Z. and P. MORANDA (1972), Software reliability research, *Statistical Computer Performance Evaluation,* 465–484.

LE CAM, L. (1990), Maximum likelihood: an introduction, *International Statistical Review 58,* 153–171.

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

McCABE, T. J. (1976), A complexity measure, *IEEE Transactions on Software Engineering 2,* 308–320.

MILLS, H. D. (1972), On the statistical validation of computer programs, IBM Federal Systems Division, Report FSC-72-6015, Gaithersburg, MD.

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 and Simulation 4,* 29–41.

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

NAGEL, P. M. and J. A. SKRIVAN (1982), Software reliability: repetitive run experimentation and modelling, Boeing Computer Service Company, BSC-40366, Seattle, WA.

NELDER, J. A. and R. MEAD (1965), A simplex method for function minimisation, *Computer Journal 7,* 308–313.

VAN PUL, M. C. (1992), Asymptotic properties of statistical models in software reliability, *Scandinavian Journal of Statistics 19,* 235–253.

VETTERLING, W. T., S. A. TEUKOLSKI, W. H. PRESS and B. D. FLANNERY (1985), *Numerical recipes: the art of scientific computing,* Cambridge University Press, Cambridge.

WALD, A. (1949), Note on the consistency of the maximum likelihood estimate, *Annals of Mathematical Statistics 20,* 595–601.