

Towards a filter-based EM-algorithm for parameter estimation of Littlewood's software reliability model

James Ledoux

INSA & IRMAR
20 avenue des Buttes de Cöesmes, CS 14315,
35043 Rennes Cedex, France
(e-mail: james.ledoux@insa-rennes.fr)

Abstract. In this paper, we deal with a continuous-time software reliability model designed by Littlewood. This model may be thought of as a partially observed Markov process. The EM-algorithm is a standard way to estimate the parameters of processes with missing data. The E-step requires the computation of basic statistics related to observed/hidden processes. In this paper, we provide finite-dimensional non-linear filters for these statistics using the innovations method. This allows to plan the use of the filter-based EM-algorithm developed by Elliott.

Keywords: Filtering, Hidden Markov process, Point process, Innovations method.

1 Introduction

A major issue in software reliability modeling is the calibration of the models from data. This is well documented in the so-called “black-box approach”. We refer to [Ledoux, 2003] and references therein for details. To the best of our knowledge, no statistical procedure has been proposed in the architecture-based approach for assessing the reliability of a software. A standard model in this context was provided by Littlewood [Littlewood, 1975]. It has inspired most other works [Goseva-Popstojanova and Trivedi, 2001]. Littlewood proposed a Markov-type reliability model for modular softwares. For a software with a finite number of modules:

- the structure of the software is represented by a finite continuous time Markov chain $X = (X_t)_{t \geq 0}$ where X_t is the active module at time t . The generator of X is denoted by Q and its state space is assumed to be $\mathcal{U} := \{e_i, i = 1, \dots, n\}$.
- When module e_i is active, the failure times are part of a homogeneous Poisson Process with intensity $\mu(i)$.
- When control switches from module e_i to module e_j , a failure may happen with probability $\mu(i, j)$.
- When a failure appears, the time to recover a safe state is neglected. A failure does not affect the execution dynamics of the software.
- All failure processes are assumed to be independent, given a sequence of activated modules.

Let us denote the number of observed failures over $[0, t]$ by N_t . It can be seen that $(N_t, X_t)_{t \geq 0}$ is a Markov process with state space $\mathbb{N} \times \mathcal{U}$. It has the following generator

$$A = \begin{pmatrix} D_0 & \mathbf{0} & \mathbf{0} & \cdots \\ D_1 & D_0 & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix} \tag{1}$$

when the states are listed in lexicographic order and the matrices D_0 and D_1 are defined by

$$\begin{aligned} \text{if } j \neq i : D_0(j, i) &:= Q(j, i)(1 - \mu(j, i)) & D_1(j, i) &:= Q(j, i)\mu(j, i), \\ D_0(i, i) &:= -\sum_{j \neq i} Q(j, i) - \mu(i) & D_1(i, i) &= \mu(i). \end{aligned}$$

The nonnegative number $D_0(j, i)$ ($j \neq i$) represents the rate at which X jumps from state e_i to e_j with no failure event. The entry $D_1(j, i)$ is the rate at which X jumps from state e_i to state e_j with the occurrence of one failure. Note that $Q = D_0 + D_1$. The distribution function of the counting variable N_t may be numerically evaluated using the uniformization technique. But this requires the knowledge of the non-negative parameter vector

$$\theta = \{D_k(j, i), \quad k = 0, 1 \quad i, j = 1, \dots, n\}.$$

In general, we can obtain a priori estimates for θ using procedures reported in [Goseva-Popstojanova and Trivedi, 2001]. They are based on data collected at earlier phases of the software life cycle (validation phases, integration tests, ...). Sometimes, these estimates might appear to be rough estimates when the software is in operation. The only available data is the observation of failure events. In that perspective, the process (N, X) should be thought of as a partially observed Markov process or a hidden Markov process. The observed process is the failure point process $(N_t)_{t \geq 0}$ and the state or hidden process is the finite Markov process $(X_t)_{t \geq 0}$. The EM-algorithm is a standard way to estimate the parameters of hidden Markov processes. Elliott proposed a filter-based EM-algorithm in [Elliott *et al.*, 1995]. That is, the standard forward-backward form of the E-step of the algorithm is replaced by a single-pass procedure that involves finite-dimensional filters for various statistics related to the observed/hidden processes. The aim of this paper is to provide such finite-dimensional filters for Littlewood’s model.

We point out that we deal with a failure point process that is a Markovian Arrival Process (MAP) as defined by Neuts [Neuts, 1989]. The Littlewood model has the (doubly stochastic) Poisson process (driven) modulated by a Markov process as a special instance (setting parameters $\mu(\cdot, \cdot)$ to 0). Statistical estimation for the MAP has been recently developed in the continuous-time context (see [Asmussen, 2000, Klemmm *et al.*, 2003, and references therein]). All these works use the forward-backward procedure. The numerical experiments reported in their studies show that EM-algorithm works well in general.

2 Finite-dimensional filters

Main notation and convention

- Vectors are column vectors. Row vectors are denoted by means of the transpose operator $(\cdot)^\top$.
- $\mathbf{1}_k$ is a k -dimensional vector with each entry equals to one.
- We denote the left limit of function f at t by f_{t-} .
For any function $t \mapsto f_t$, $\Delta f_t := f_t - f_{t-}$ for $t > 0$ is the jump of the function at time t . We set $\Delta f_0 := f_0$.
- The state space of X is $\mathcal{X} := \{e_i, i = 1, \dots, n\}$, where e_i is the i th vector of the canonical basis of \mathbb{R}^n . With this convention,

$$\mathbf{1}_{\{X_t=e_i\}} = \langle X_t, e_i \rangle, \quad \mathbf{1}_n^\top X_t = 1$$

where $\langle \cdot, \cdot \rangle$ is the usual scalar product in \mathbb{R}^n .

- All processes are assumed to be defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The internal filtrations of processes N and (N, X) are denoted by $\mathbb{F}^N = (\mathbb{F}_t^N)_t$ and $\mathbb{F} = (\mathbb{F}_t)_t$ respectively. These filtrations are assumed to be complete.
- For any integrable adapted random process $(Z_t)_{t \geq 0}$, the conditional expectation $\mathbb{E}[Z_t | \mathbb{F}_t^N]$ is denoted by \widehat{Z}_t .

2.1 Basic material on the observed/hidden processes

We report here a semi-martingale representation of the basic statistics of the Littlewood’s model for which filters will be derived. Due to the special structure of generator A of (N, X) (see (1)), N and the following counting processes are easily interpreted to be counters of specific transitions in (N, X)

$$\begin{aligned} N_t^{X,ji} &:= \sum_{0 < s \leq t} \langle X_{s-}, e_i \rangle \langle X_s, e_j \rangle = \int_0^t \langle X_{s-}, e_i \rangle \langle e_j, dX_s \rangle \\ \mathcal{L}_t^{1,ji} &:= \sum_{0 < s \leq t} \langle X_{s-}, e_i \rangle \langle X_s, e_j \rangle \Delta N_s = \int_0^t \langle X_{s-}, e_i \rangle \langle X_s, e_j \rangle dN_s \\ j \neq i \quad \mathcal{L}_t^{0,ji} &:= \sum_{0 < s \leq t} \langle X_{s-}, e_i \rangle \langle X_s, e_j \rangle (1 - \Delta N_s) = N_t^{X,ji} - \mathcal{L}_t^{1,ji} \end{aligned}$$

and $N_t(x, y)$ is the number of transitions of (N, X) from state x to state y at time t . It is well known that [Bremaud, 1981]

$$M_t(y, x) := N_t(y, x) - \int_0^t A(y, x) \mathbf{1}_{\{(N_{s-}, X_{s-})=x\}} ds$$

is a \mathbb{F} -martingale. In other words, the \mathbb{F} -semi-martingale (or Doob-Meyer here) decomposition of $N(y, x)$ is $N_t(y, x) = \int_0^t \mathbf{1}_{\{(N_{s-}, X_{s-})=x\}} A(y, x) ds +$

$M_t(y, x)$. Then, it is easily seen that the \mathbb{F} -semi-martingale decomposition of the counting processes above are

$$\begin{aligned}
 N_t &= \int_0^t \lambda_s ds + \mathcal{M}_t \quad \text{with } \lambda_s := \mathbf{1}_n^T D_1 X_{s-} & (2) \\
 N_t^{X,ji} &:= \int_0^t Q(j, i) \langle X_{s-}, e_i \rangle ds + \mathcal{M}_t^{N^X(j,i)} \\
 \mathcal{L}_t^{k,ji} &:= \int_0^t D_k(j, i) \langle X_{s-}, e_i \rangle ds + \mathcal{M}_t^{\mathcal{L}^{k,ji}} \quad k = 0, 1 & (3)
 \end{aligned}$$

where $\mathcal{M}, \mathcal{M}^{N^X(j,i)}, \mathcal{M}^{\mathcal{L}^{k,ji}}$ are \mathbb{F} -martingales.

The last statistics that we need, is the sojourn time of X in any state e_i in the interval $[0, t]$

$$\mathcal{O}_t^{(i)} := \int_0^t \langle X_{s-}, e_i \rangle ds$$

The basic semi-martingale decomposition of the Markov process X is [Brenaud, 1981]

$$X_t = \int_0^t Q X_{s-} ds + \mathcal{M}_t^X. \tag{4}$$

We recognize in (4) and (2) a standard representation of a continuous-time hidden Markov process, with X as the state process and N the observed process. The observation and state “noises” are correlated here.

2.2 The EM-algorithm

We briefly explain the EM-algorithm for our continuous-time hidden Markov model. We refer to [Klemmm *et al.*, 2003] for full details. For a fixed parameter vector θ , we denote the underlying probability measure and associated expectation respectively by \mathbb{P}_θ and \mathbb{E}_θ . X_0 or its probability distribution x_0 is assumed to be known. The observed data are supposed to be the inter-failure durations $\{t_1, \dots, t_K\}$ where $t_K = t$. The likelihood function for the complete data (N, X) up to time t under \mathbb{P}_θ is

$$L_t(\theta; N, X) := \prod_{i,j=1,}^n D_1(j, i)^{\mathcal{L}_t^{1,ji}} \prod_{i,j=1,j \neq i}^n D_0(j, i)^{\mathcal{L}_t^{0,ji}} \prod_{i=1}^n e^{D_0(i,i)\mathcal{O}_t^{(i)}} \prod_{i=1}^n x_0(i)^{\langle X_0, e_i \rangle}.$$

The formulas for estimating θ from the observations $N_s, s \leq t$ are obtained using the following iterative procedure:

- i*) Initialization : Choose θ_0
- ii*) E-step. Set $\theta := \theta_t$. Compute the so-called pseudo-log-likelihood $Q(\cdot | \theta)$ defined by

$$Q(\theta^* | \theta) := \mathbb{E}_\theta [\log L_t(\theta^*; N, X) | \mathbb{F}_t^N] \tag{5}$$

where $\theta^* := \{D_k^*(j, i), i, j = 1, \dots, n; k = 0, 1\}$.

- iii) **M-step.** Determine θ_{l+1} maximizing the function (5).
- iv) Return in 2 until a stopping criterion is satisfied.

For the M-step, it is easily seen that

$$i, j = 1, \dots, n \quad D_1^*(j, i) = \frac{\widehat{\mathcal{L}}_t^{1,ji}}{\widehat{\mathcal{O}}^{(i)}_t}, \quad D_0^*(j, i) = \frac{\widehat{\mathcal{L}}_t^{0,ji}}{\widehat{\mathcal{O}}^{(i)}_t}, \quad i \neq j. \quad (6)$$

An appealing property of the EM-algorithm is that the sequence of estimates $\{\theta_l, l \geq 0\}$ gives a nondecreasing values of the likelihood function with equality iff $\theta_{l+1} = \theta_l$ (under mild conditions). Note that the zero entries of D_k s are preserved by the procedure above.

As a result of the procedure above, we have to compute the estimates in (6). The standard way is to use the Baum-Welch implementation of the EM-algorithm (also referred to as the “forward-backward” technique). This is what is done in the previously mentioned works [Asmussen, 2000, Klemmm *et al.*, 2003]. Using the filter-based approach pioneering by Elliott [Elliott *et al.*, 1995], the estimates in (6) are computed from the filters given in Theorem 1. The basic difference with the standard Baum-Welch method is that only one pass through the data set is needed for the filter-based method.

2.3 The results

We use a trick proposed by Elliott. We compute the following filters

$$\widehat{N^{X,ji} X}_t, \widehat{\mathcal{O}^{(i)} X}_t \text{ and } \widehat{\mathcal{L}^{1,ji} X}_t$$

which turn to be finite-dimensional. Then, we have

$$\widehat{N^{X,ji} X}_t = \mathbf{1}_n^\top \widehat{N^{X,ji} X}_t, \quad \widehat{\mathcal{O}^{(i)} X}_t = \mathbf{1}_n^\top \widehat{\mathcal{O}^{(i)} X}_t, \quad \text{and} \quad \widehat{\mathcal{L}^{1,ji} X}_t = \mathbf{1}_n^\top \widehat{\mathcal{L}^{1,ji} X}_t.$$

A filter equation for $\widehat{\mathcal{L}^{0,ji} X}_t$ ($j \neq i$) can be derived as that of Theorem 1 or using the fact that $\widehat{\mathcal{L}^{0,ji} X}_t = \widehat{N^{X,ji} X}_t - \widehat{\mathcal{L}^{1,ji} X}_t$.

Theorem 1 Let $\widehat{\lambda}_t := \mathbf{1}_n^\top D_1 \widehat{X}_{t-}$. The fundamental \mathbb{F}^N -martingale $(N_t - \int_0^t \widehat{\lambda}_s ds)_{t \geq 0}$ is denoted by $(\widehat{\mathcal{M}}_t^N)_{t \geq 0}$.

- i) *Estimator for the state.* We have for any $t \geq 0$

$$\widehat{X}_t = \widehat{X}_0 + \int_0^t Q \widehat{X}_{s-} ds + \int_0^t \frac{D_1 \widehat{X}_{s-} - \widehat{X}_{s-} \widehat{\lambda}_s}{\widehat{\lambda}_s} d\widehat{\mathcal{M}}_s^N. \quad (7)$$

- ii) *Estimator for the number of jumps of X from e_i to e_j .* We have for any $t \geq 0$

$$\begin{aligned} \widehat{N^{X,ji} X}_t &= \int_0^t Q \widehat{N^{X,ji} X}_{s-} ds + \int_0^t Q(j, i) \langle \widehat{X}_{s-}, e_i \rangle ds e_j \\ &\quad + \int_0^t \frac{D_1 \widehat{N^{X,ji} X}_{s-} - \widehat{N^{X,ji} X}_{s-} \widehat{\lambda}_s}{\widehat{\lambda}_s} d\widehat{\mathcal{M}}_s^N. \end{aligned} \quad (8)$$

iii) *Estimator for the sojourn time to e_i . We have for any $t \geq 0$*

$$\begin{aligned} \widehat{\mathcal{O}}^{(i)} X_t &= \int_0^t Q \widehat{\mathcal{O}}^{(i)} X_{s-} ds + \int_0^t \langle \widehat{X}_{s-}, e_i \rangle ds e_i \\ &+ \int_0^t \frac{D_1 \widehat{\mathcal{O}}^{(i)} X_{s-} - \widehat{\mathcal{O}}^{(i)} X_{s-} \widehat{\lambda}_s}{\widehat{\lambda}_s} d\widehat{\mathcal{M}}_s^N. \end{aligned} \tag{9}$$

iv) *Estimator for the number of joint transitions. We have for $t \geq 0$*

$$\begin{aligned} \widehat{\mathcal{L}}^{1,ji} X_t &= \int_0^t Q \widehat{\mathcal{L}}^{1,ji} X_{s-} ds + \int_0^t D_1(j, i) \langle \widehat{X}_{s-}, e_i \rangle ds e_j \\ &+ \int_0^t \frac{D_1(j, i) \langle \widehat{X}_{s-}, e_i \rangle e_j + D_1 \widehat{\mathcal{L}}^{1,ji} X_{s-} - \widehat{\mathcal{L}}^{1,ji} X_{s-} \widehat{\lambda}_s}{\widehat{\lambda}_s} d\widehat{\mathcal{M}}_s^N. \end{aligned} \tag{10}$$

Remark 1 *The filters for the statistics of an MMPP may be obtained from the previous theorem. We have $D_1 = \text{Diag}(\mu(i))$.*

Proof. A proof of (7) may be found in [Gravereaux and Ledoux, 2004]. In the sequel, \mathcal{M} (resp. $\widehat{\mathcal{M}}$) will denote a generic \mathbb{F} (resp. \mathbb{F}^N)-martingale. The proof of (9) is as follows. An integration by parts gives

$$\begin{aligned} \mathcal{O}_t^{(i)} X_t &= \int_0^t \mathcal{O}_{s-}^{(i)} dX_s + \int_0^t X_{s-} d\mathcal{O}_s^{(i)} + \underbrace{[\mathcal{O}^{(i)}, X]_t}_0 \\ &= \int_0^t Q \mathcal{O}_s^{(i)} X_{s-} ds + \int_0^t \langle X_{s-}, e_i \rangle e_i ds + \mathcal{M} \quad \text{from (4)}. \end{aligned} \tag{11}$$

The \mathbb{F}^N -optional projection of the equation above, is

$$\widehat{\mathcal{O}}^{(i)} X_t = \int_0^t Q \widehat{\mathcal{O}}^{(i)} X_{s-} ds + \int_0^t \langle \widehat{X}_{s-}, e_i \rangle e_i + \widehat{\mathcal{M}}. \tag{12}$$

The integral representation of \mathbb{F}^N -martingales says that $\widehat{\mathcal{M}}$ in the right hand side member above, has the form [Bremaud, 1981]

$$\int_0^t G_s^{(i)} d\widehat{\mathcal{M}}_s^N.$$

Thus, the proof will be complete if we prove that

$$G_s^{(i)} = \frac{D_1 \widehat{\mathcal{O}}^{(i)} X_{s-} - \widehat{\mathcal{O}}^{(i)} X_{s-} \widehat{\lambda}_s}{\widehat{\lambda}_s}. \tag{13}$$

The product $N_t \widehat{\mathcal{O}^{(i)} X}_t$ has the form from an integration by parts

$$\begin{aligned} N_t \widehat{\mathcal{O}^{(i)} X}_t &= \int_0^t N_{s-} d\widehat{\mathcal{O}^{(i)} X}_s + \int_0^t \widehat{\mathcal{O}^{(i)} X}_{s-} dN_s + [N, \widehat{\mathcal{O}^{(i)} X}]_t \\ &= \int_0^t N_{s-} [Q\widehat{\mathcal{O}^{(i)} X}_s + \langle \widehat{X}_{s-}, e_i \rangle e_i] ds + \widehat{\mathcal{M}} \text{ from (12)} \\ &\quad + \int_0^t \widehat{\mathcal{O}^{(i)} X}_{s-} \widehat{\lambda}_s ds + \widehat{\mathcal{M}} \\ &\quad + \int_0^t G_s^{(i)} \widehat{\lambda}_s ds + \widehat{\mathcal{M}} \end{aligned} \tag{14}$$

Note that $\mathcal{O}^{(i)}$ has continuous paths so that $\Delta \mathcal{O}_s^{(i)} = 0$. Next, the product $N_t(\mathcal{O}_t^{(i)} X_t)$ is with an integration by parts

$$\begin{aligned} N_t \mathcal{O}_t^{(i)} X_t &= \int_0^t N_{s-} d(\mathcal{O}^{(i)} X)_s + \int_0^t \mathcal{O}_{s-}^{(i)} X_s dN_s \\ &= \int_0^t N_{s-} [Q\mathcal{O}_{s-}^{(i)} X_{s-} + \langle \widehat{X}_{s-}, e_i \rangle e_i] ds + \int_0^t \mathcal{O}_{s-}^{(i)} X_s dN_s \text{ from (11)}. \end{aligned}$$

Let us compute the last term in the equality above:

$$\begin{aligned} \int_0^t \mathcal{O}_{s-}^{(i)} X_s dN_s &= \sum_{0 < s \leq t} \mathcal{O}_{s-}^{(i)} X_s \Delta N_s = \sum_j e_j \sum_k \int_0^t \mathcal{O}_{s-}^{(i)} d\mathcal{L}_s^{1,jk} \\ &= \int_0^t \mathcal{O}_{s-}^{(i)} \sum_j e_j \sum_k D_1(j, k) \langle X_{s-}, e_k \rangle ds + \mathcal{M} \text{ from (3)} \\ &= \int_0^t \mathcal{O}_{s-}^{(i)} D_1 X_{s-} ds + \mathcal{M}. \end{aligned}$$

Then, we deduce from the last equality that

$$N_t \mathcal{O}_t^{(i)} X_t = \int_0^t N_{s-} [Q\mathcal{O}_{s-}^{(i)} X_{s-} + \langle X_{s-}, e_i \rangle e_i] ds + \int_0^t \mathcal{O}_{s-}^{(i)} D_1 X_{s-} ds + \mathcal{M}.$$

The \mathbb{F}^N -optional projection of the previous formula leads to a second decomposition of the special semi-martingale $N_t \widehat{\mathcal{O}^{(i)} X}_t$

$$N_t \widehat{\mathcal{O}^{(i)} X}_t = \int_0^t N_{s-} [Q\widehat{\mathcal{O}^{(i)} X}_{s-} + \langle \widehat{X}_{s-}, e_i \rangle e_i] ds + \int_0^t D_1 \widehat{\mathcal{O}^{(i)} X}_{s-} ds + \widehat{\mathcal{M}}. \tag{15}$$

We know that the bounded variations part of the decomposition of a special semi-martingale is unique. Then, we can identify the corresponding terms in the decompositions (14) and (15), that is the Lebesgue integrals. The expression (13) of the gain $G^{(i)}$ follows easily.

Formulas (10) and (8) are shown in the same way. Their proofs are not reported here for saving space.

3 Conclusion

The Littlewood's software reliability model may be thought of as a partially observed Markov process. The contribution of this paper is to provide finite-dimensional non-linear filters for various statistics associated with this model. These filters are the first step in view of implementing the filter-based form of the EM-algorithm proposed by Elliott [Elliott *et al.*, 1995]. We mention that basic extensions may be obtained following the guidelines of this paper. We can derive filters for the general class of MAP's. The case of occurrences of failures in clusters can also be included in the discussion. We just have to consider N as a multivariate point process of failures. From the numerical point of view, a second step in implementing the filter-based approach would be to find the so-called robust versions of the non-linear filtering equations obtained here. Then, robust numerical algorithms may be expected. We refer to [James *et al.*, 1996] for a detailed discussion of such a time discretization approach. In this perspective, we mention that it should be desirable to obtain Zakai form for our filters. Such a form can be derived from our results. The details will be reported elsewhere.

References

- [Asmussen, 2000]S. Asmussen. Matrix-analytic models and their analysis. *Scandinavian Journal of Statistics*, pages 193–226, 2000.
- [Bremaud, 1981]P. Bremaud. *Point Processes and Queues*. Springer, 1981.
- [Elliott *et al.*, 1995]R.J. Elliott, L. Aggoun, and J.B. Moore. *Hidden Markov Models*. Springer, 1995.
- [Goseva-Popstojanova and Trivedi, 2001]K. Goseva-Popstojanova and K.S. Trivedi. Architecture-based approach to reliability assessment of software systems. *Performance Evaluation*, pages 179–204, 2001.
- [Gravereaux and Ledoux, 2004]J.-B. Gravereaux and J. Ledoux. Poisson approximation for some point processes in reliability. *Advances in Applied Probability*, pages 455–470, 2004.
- [James *et al.*, 1996]M.R. James, V. Krishnamurthy, and Le Gland F. Time discretization of continuous-time filters and smoothers for HMM parameter estimation. *IEEE Trans. Information Theory*, pages 593–605, 1996.
- [Klemmm *et al.*, 2003]A. Klemmm, C. Lindemann, and M. Lohmann. Modeling IP traffic using the Batch Markovian Arrival Process. *Performance Evaluation*, pages 149–173, 2003.
- [Ledoux, 2003]J. Ledoux. Chap. 12: Software reliability modeling. In H. Pham, editor, *Handbook of Reliability Engineering*, pages 213–234. Springer, London, 2003.
- [Littlewood, 1975]B. Littlewood. A reliability model for systems with Markov structure. *Appl. Statist.*, pages 172–177, 1975.
- [Neuts, 1989]M. F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. Marcel Dekker Inc., New-York and Basel, 1989.