Published in volume: 65 Issue: 3 pages: 1571-1583 september 2016

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Maximum-Likelihood Estimation of Parameters of NHPP Software Reliability Models Using Expectation Conditional Maximization Algorithm

Panlop Zeephongsekul, *Member, IEEE*, Chathuri L. Jayasinghe, Lance Fiondella, *Member, IEEE*, and Vidhyashree Nagaraju, *Graduate Student Member, IEEE*

***Abstract*—Since its introduction in 1977, the expectation maximization (EM) algorithm has been one of the most important and widely used estimation method in estimating parameters of distributions in the presence of incomplete information. In this paper, a variant of the EM algorithm, the expectation conditional maximization (ECM) algorithm, is introduced for the first time and it provides a promising alternative in estimating the parameters of nonhomogeneous poisson (NHPP) software reliability growth models (SRGM). This algorithm circumvents the difficult M-step of the EM algorithm by replacing it by a series of conditional maximization steps. The utility of the ECM approach is demonstrated in the estimation of parameters of several well-known models for both time domain and time interval software failure data. Numerical examples with real-data indicate that the ECM algorithm performs well in estimating parameters of NHPP SRGM with complex mean value functions and can produce a faster rate of convergence.**

***Index Terms*—Expectation conditional maximization (EMC) algorithm, expectation Maximization (EM) algorithm, nonhomogeneous poisson process (NHPP), software reliability growth models (SRGM).**

Acronyms

ECM	Expectation conditional maximization.
EM	Expectation maximization.
GEM	Generalized expectation maximization.
SRGM	Software reliability growth model.
NHPP	Nonhomogeneous poisson process.
MLE	Maximum-likelihood estimation.
pdf	Probability density function.
LLF	Log likelihood function.
MSE	Mean squared error.
r.v.	Random variable.

Notation

$N(t)$	Cumulative number of failures detected at time t.
$\mathbf{Y}$	Complete set of random data $(Y_1, Y_2, \ldots, Y_p)$.
$\mathbf{X}$	Incomplete set of random data $(X_1, X_2, \ldots, X_n)$, $n < p$.
$\mathbf{Z}$	Set of unobserved random data $(Z_1, Z_2, \ldots, Z_m)$, $m = p - n$.
R^d	d dimensional Euclidean space.
$f(\mathbf{y};\theta)$	pdf of Y.
$f(\mathbf{z}\|\mathbf{x};\theta)$	Conditional pdf of $\mathbf{Z}\|\mathbf{X}$.
$Q(\theta\|\theta_0, \mathbf{x})$	$\int \ln(f(\mathbf{x}, \mathbf{z}; \theta))\, f(\mathbf{z}\|\mathbf{x}; \theta_0) d\mathbf{z}$.
$G = \{g_s(\theta) : s = 1, 2, \ldots, S\}$	S preselected vector–valued Functions of θ.
$\Theta_s(\theta_{t+(s-1)/S})$	$\{\theta : g_s(\theta) = g_s(\theta_{t+(s-1)/S})\}$.
$\mathcal{X} = R^+$	Nonnegative half line $[0, \infty)$.
$\mathcal{X}^n$	n–fold Cartesian product of $\mathcal{X}$.
$\mathcal{S}$	Family of all subsets of $\mathcal{X}$.
$\mathcal{S}_k$	Family of events generated by all rectangular sets $A_1 \times A_2 \times \ldots A_k$.
$\mathcal{B}$	$\cup_{k=0}^{\infty} \mathcal{S}_k$.
$\mathbf{P}$	Probability measure on $(\mathcal{X}, \mathcal{B})$.
$\Pr(N(A_i) = k_i, 1 \le i \le m)$	Probability of k_i failures observed in A_i, $1 \le i \le m$.
$l(\cdot)$	Length measure on $\mathcal{S}$.
$l^{\otimes k}$	k-volume measure on $\mathcal{S}_k$.
l_k	$\frac{l^{\otimes k}}{k!}$.
$\Gamma(\cdot)$	$\oplus_{k=0}^{\infty} l_k(\cdot)$.
$\mathbf{P} \ll \Gamma$	$\mathbf{P}$ is absolutely continuous with respect to Γ.
$p_n(\cdot)$	Janossy density.
1_A	Indicator function of the set A.
$\mu(.)$	Mean measure of the NHPP.
$Q_\mu(.)$	NHPP with mean measure $\mu(.)$.
$q_\mu(.)$	Likelihood function of $Q_\mu(.)$.
$G(t;\varphi)$	Distribution function of r.v. T:. the time of occurrence of a software failure.
$\bar{G}(t;\varphi)$	Survival function of r.v. T.
ω	Parameter of the Poisson distribution.

Manuscript received October 15, 2015; revised February 25, 2016 and April 05, 2016; accepted April 27, 2016. The work of L. Fiondella and V. Nagaraju was supported in part by the Naval Air Systems Command (NAVAIR) through the Systems Engineering Research Center (SERC), a Department of Defense (DoD) University Affiliated Research Center (UARC) under Research Task 139: Software Reliability Modeling and in part by the National Science Foundation (NSF) (#1526128). Associate Editor: S. Li.

P. Zeephongsekul and C. L. Jayasinghe are with the School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, VIC 3000, Australia (e-mail: panlop.zeephongsekul@rmit.edu.au; chathuri.jayasinghe@rmit.edu.au).

L. Fiondella and V. Nagaraju are with the Department of Electrical and Computer Engineering, University of Massachusetts, Dartmouth, MA 02747 USA (e-mail: lfiondella@umassd.edu; vnagaraju@umassd.edu).

Digital Object Identifier 10.1109/TR.2016.2570557