Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 62 (2015) 360 – 369

Procedia Computer Science

The 2015 International Conference on Soft Computing and Software Engineering (SCSE 2015)

A Parametric Empirical Bayes Model to predict Software Reliability Growth

Néstor R. Barraza*

Department of Engineering, Universidad Nacional de Tres de Febrero, Caseros, Argentina

Abstract

A new software reliability model based on the empirical Bayes estimate is developed. The number of failures estimated up to a given time is used in order to estimate the probability of failure appearance during the next time interval. Instead of a non homogeneous in time failure rate as it is usually used to model reliability growth, a failure rate depending non linearly on the previous number of failures is obtained from our model. The estimate is obtained from a mixed Poisson model where the mixing probability density function models the reliability growth. The model can be used either to simulate the cumulative failures curve or to estimate the time between failures. Data of a similar project can be used to estimate the parameters of a given project. Results of simulations and estimated mean time between failures comparing well with experimental data are also shown.

Peer-review under responsibility of organizing committee of The 2015 International Conference on Soft Computing and Software Engineering (SCSE 2015)

Keywords: Software Reliability; Empirical Bayes; Simulation of software failures; Probability of failure

1. Introduction

The interest in Software Reliability has been increasing since the first models were proposed in the 70's. Several statistical models involve Non-homogeneous Poisson, Compound Poisson, clusters and Markov chains among others, see References [1,2,3,4]. There were also analyzed particular methods of estimation. These researches show that Software Reliability has become an important application of statistics.

Software Reliability models are generally stochastic processes that intend to introduce metrics on the failure detection process like MTBF and MTTF or the number of remaining failures at any time. Those metrics allow to evaluate the software development and testing processes in order to assign testing resources or to predict the release time. The characteristic of the failure detection process as a function of time shows a decreasing rate as testing time progresses. In order to follow this behavior, several software reliability models were proposed in the literature, the so called software reliability growth models. Non homogeneous and Compound Poisson processes among others were proposed decades ago as software reliability growth models.

* Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: nbarraza@untref.edu.ar

Peer-review under responsibility of organizing committee of The 2015 International Conference on Soft Computing and Software Engineering (SCSE 2015)
doi:10.1016/j.procs.2015.08.416

On the other hand, besides statistical software reliability models, simulations of the stochastic software failures detection process is an important research issue related to testing and debugging, see References[5,6,7].

The Bayes estimation framework has also been an important statistical method of estimation. Bayesian estimation has been largely applied to Software Reliability since the pioneer paper listed in Reference[8]. Estimation of the mean time between failures using several priors distribution for the parameters were proposed in Reference[9,10]. Other approaches involve the use of combinatorial optimization algorithms like Metropolis and Gibbs sampler in order to calculate the posterior distribution, see References[9,11]. Since the Bayes method lies on the number of previous reported failures, this procedure leads to the Empirical Bayes estimate.

The empirical Bayes estimate is obtained from the Bayes rule applied to the conditional expectation of the unknown parameter given the samples. The conditional expectation minimizes the mean square error function as it is well known. Several applications can be found in different areas of engineering, specially in speech recognition and word processing, Refs[12,13]. Many simple forms of this estimate were proposed decades ago.

In this work, a new software reliability growth model to predict the failure cumulative curve and the mean time between failures behavior using information at the very start of the project is proposed. We propose to estimate the probability of failure in the next interval time given r failures are detected before from the Empirical Bayes method. Our model allows on one hand to simulate the stochastic failure detection process and to estimate the mean time between failures based on the number of previous reported failures on the other. From a Montecarlo simulation of the estimated probability, our model allows to simulate the stochastic failure process. Prediction of Software Reliability at an early stage is quite important in order to adjust testing resources or to estimate the release time.

Since a nonlinear dependence on r is required for software failures in order to follow the software failure cumulative curve, we use the expression introduced in Reference[14]. Parameters of the model are estimated in two ways, from previous reported data of the same project and using a mixing probability density function that fits well the cumulative failures curve of a similar project. The use of software reliability information of similar projects in order to estimate the reliability of a given project is a common practice, see References[15,16,17,18,19].

This paper is organized as follows: The Empirical Bayes estimate with a nonlinear characteristic is shown in section (2), the theoretical foundations that supports our main motivation are presented in section (3), discussions on choosing the mixing distribution are presented in section (4), a simulation and comparison with experimental data is shown in section (5), an application of our model to predict the mean time between failures is presented in section (6), conclusions are presented in section (7).

2. The Empirical Bayes estimate using mixed distributions

A well known non parametric estimate was originally proposed in Reference[20] as an alternative to the maximum likelihood, the so called Good-Turing estimate. An Empirical Bayes form of the non parametric Good-Turing estimate based on a mixed binomial distribution was proposed in Reference[12]. From a general expression of the Empirical Bayes estimate using mixed distributions, either mixed binomial or mixed Poisson, that method was later generalized in[14] in order to get a given family of estimates.

In order to get the estimate, we start to consider the distribution of r statistically independent single events in n outcomes. This distribution is given by the binomial.

$$P(r) = \binom{n}{r}\theta^r \, (1-\theta)^{n-r} \tag{1}$$

From the Poisson approximation valid for large values of n and $r \ll n$, we obtain a Poisson distribution with parameter:

$$\lambda = \theta \, n \tag{2}$$

Thus, using (2), we can get an estimate for θ from an estimate for λ. Being $S(\lambda)$ the prior probability density function of λ, we get for r a mixed Poisson distribution:

$$P(r) = \int \frac{\lambda^r}{r!} \exp(-\lambda) \, dS(\lambda) \, . \tag{3}$$