An. Şt. Univ. Ovidius Constanţa Vol. 16(1), 2008, 81–90

Statistical simulation and prediction in software reliability

Alexei Leahu and Carmen Elena Lupu

Abstract

On the base of statistical simulation (Monte Carlo method), in this paper it was investigated the rate (relative frequencies) of "success" in predicting the number of initial (remained) errors by means of Maximum Likelihood Principle. Some numerical results it will be discussed.

1. Model's description

As an extension of the paper [1], the aim of this paper is to use the statistical simulation for some Jelinski-Moranda's software reliability models in order to check the efficiency of the well known maximum likelihood statistical estimators for some parameters. More exactly, we consider the Jelinski-Moranda (**JM**) models based on the following hypotheses:

1. *The total number N of errors existing initially in the software is unknown constant.*

2. *Each error is eliminated with probability $p = 1$, independently of the past trials, repair of the error being snapshot and without introduction of the new errors or*

$2'$.*Each error is eliminated with probability p, $0 < p < 1$, independently of the past trials, repair of the error being snapshot and without introduction of the new errors.*

Key Words: Reliability; Maximum likelihood principle; Statistical simulation.
Mathematics Subject Classification: 2N05, 62N02, 90B25.
Received: November, 2007
Accepted: February, 2008

3. *The time intervals between two successive failures of the software are independent identically exponentially distributed random variables with parameter* $\mu > 0$.

4. *Distribution's parameter of the interval between two successive failures of the software, i.e. rate or intensity of the failures, is directly proportional with the number of non eliminated errors in the software at the beginning of this interval.*

In this way, we have $\mathbf{(JM)}_1$ model if the hypotheses $1, 2, 3, 4$ are valid and $\mathbf{(JM)}_2$ model if the hypotheses $1, 2', 3, 4$ are valid.

In order to verify experimentally the maximum likelihood procedure based on the statistical estimation of the probability of the successful prediction for the number of initial (remained) software's errors in the model $\mathbf{(JM)}_2$ we use Monte-Carlo simulation based on the following assertions.

Proposition 1. *If* $(X_k)_{k\geq 1}$ *are independent identically exponentially distributed random variables with parameter* $\mu > 0$*, and* K *is a random variable geometrically distributed with parameter* p*,* $0 < p \leq 1$*, which is independent of* $(X_k)_{k\geq 1}$*, then* $X_1 + X_2 + ... + X_K$ *is an exponentially distributed random variable with parameter* $\mu \cdot p$.

Proof. We apply a result in [2], for integer k, $k \geq 1$, the sum $X_1+X_2+...+X_k$ of independent identically exponentially distributed random variables with parameter $\mu > 0$ is Erlang distributed random variable (r.v.) with parameter μ, and with k degrees of freedom, i.e.,

$$X_1 + X_2 + ... + X_k \sim Erlang(k, \mu).$$

That means

$$\mathbf{P}(X_1 + X_2 + ... + X_k \ \leq t) = (1 - \sum_{i=0}^{k-1} \frac{(\mu t)^i}{i!} e^{-\mu t}) I_{[0,\infty)}(t),$$

where

$$I_{[0,\infty)}(t) = \begin{cases} 0, & if\ t < 0, \\ 1, & if\ t \geqslant 0. \end{cases}$$

So, using the Formula of total probability, we have that distribution function (d.f.) of r.v. $X_1 + X_2 + ... + X_K$ is

$$F(t) = \mathbf{P}(X_1 + X_2 + ... + X_K \ \leq t) =$$

$$\sum_{k\geq 1} \mathbf{P}(X_1 + X_2 + ... + X_k \ \leq t \ / \ K = k) \mathbf{P}(K = k) =$$

$$\left[\sum_{k \geq 1} (1 - \sum_{i=0}^{k-1} \frac{(\mu t)^i}{i!} e^{-\mu t}) p (1-p)^{k-1} \right] I_{[0,\infty)}(t).$$

By derivation of d.f. $F(t)$, we find that density function (d.f.) f of r.v.$X_1 + X_2 + ... + X_K$ corresponds to the exponential distribution. More exactly,

$$f(t) = \mu p \cdot e^{-\mu \cdot p \cdot t} I_{[0,\infty)}(t). \qquad \square$$

Let us consider that during the time interval T of error detections and their eliminations,$T > 0$, we observes software's lifetimes, i.e., intervals of length t_1, t_2, ..., t_n, where n is the total number of eliminated errors until the moment T. In this case, as a consequence of the Proposition 1 we obtain the following result

Proposition 2. *The likelihood function* $L(t_1, t_2, ..., t_n; \mu_0, N)$ *for* **(JM)**$_2$ *model is the same as the likelihood function for* **(JM)**$_1$ *model with the parameter* μ_0, *where* $\mu_0 = \mu \cdot p$.

According to the [3], that means that likelihood equations to be solve in the model **(JM)**$_2$ for the prediction of initial number of errors N are the same as for model **(JM)**$_1$, i.e.,

$$\begin{cases} \frac{\partial \ln L}{\partial N} = \sum\limits_{i=1}^{n} \frac{1}{N-i+1} - \mu_0 \sum\limits_{i=1}^{n} t_i = 0, \\ \frac{\partial \ln L}{\partial \mu_0} = \frac{n}{\mu_0} - \sum\limits_{i=1}^{n} t_i (N - i + 1) = 0. \end{cases}$$

2. Numerical results

In the context of validation of maximum likelihood procedure we have to calculate the statistical probability of the successful prediction of N in a number of M trials, i,e., M repetitions of Monte-Carlo simulations for the different values of N, μ, p and T. The confidence levels $(1-\alpha)100\%$ for a given error $\varepsilon = 0.01$ are

- 47% for $M = 1000$;
- 84% for $M = 5000$;
- 95% for $M = 10000$.

Remark. In order to find out the confidence level $1-\alpha$ for a given error ε and for a given number of trials M, such that $\mathbf{P}\left(\left| f_n\left(succes\right) - \mathbf{P}\left(succes\right)\right| < \varepsilon\right) \geqslant 1 - \alpha$ for $\forall n \geqslant M$, where $\mathbf{P}\left(succes\right)$ is the theoretical probability of the successful prediction of N, $f_n\left(succes\right)$ is the frequential (statistical) probability of the successful prediction of N in a number of n trials and $\alpha \in (0,1)$, we apply essentially the

Central Limit Theorem (in the Moivre-Laplace form). *If* $(Y_n)_{n\geqslant 1}$ *are i.i.d.r.v such that* $\mathbf{P}(Y_n = 1) = \mathbf{P}\,(succes) = q$, $\mathbf{P}(Y_n = 0) = 1 - q$, $q \in (0,1)$, *then*

$$\mathbf{P}\left(\frac{\sum\limits_{i=1}^{n}}{Y\ _i} - nq\sqrt{nq\,(1-q)}\leq x\right) \to \Phi\,(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}}du.$$

As a result of statistical simulations on the base of the models $\mathbf{(JM)}_1$ and $\mathbf{(JM)}_2$ we have the following Histogram and Tables.

Model $\mathbf{(JM)}_1$

Table 1

$\mu = 2, p = 1, T = 3$

$N\backslash M$	1000	5000	10000
5	0.550	0.5778	0.5903
10	0.664	0.6781	0.6827
15	0.691	0.6926	0.7036

Table 2

$\mu = 3, p = 1, T = 3$

$N\backslash M$	1000	5000	10000
5	0.550	0.5778	0.5903
10	0.664	0.6781	0.6827
15	0.691	0.6926	0.7036

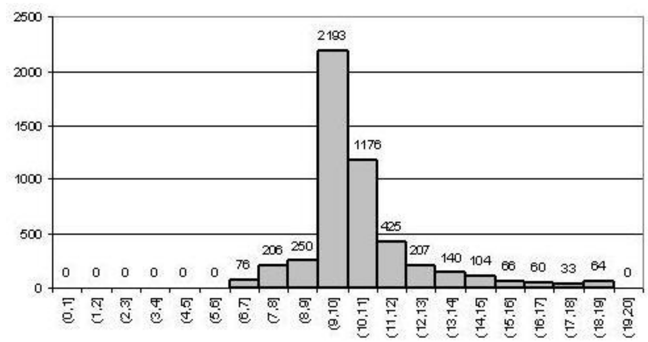

Figure 1. Histogram for the number of predicted errors when $N = 10, \mu = 2$, $p = 1$, $T = 3$ for $M = 5000$ in the model $\mathbf{(JM)}_1$

The sample mean value of the numbers of predicted errors calculated according to the Histogram (fig.1) is equal to $11,3954$, true number of errors being equal to 10.

Table 3

$\mu = 2, p = 1, T = 5$

$N \backslash M$	1000	5000	10000
5	0.672	0.6792	0.6789
10	0.694	0.6936	0.7071
15	0.698	0.7108	0.7291

Table 4

$\mu = 3, p = 1, T = 5$

$N \backslash M$	1000	5000	10000
5	0.708	0.7019	0.6992
10	0.721	0.7102	0.7093
15	0.733	0.7398	0.7346

Model $\mathbf{(JM)}_2$

Table 5

$\mu = 2, p = 1/2, T = 3$

$N \backslash M$	1000	5000	10000
5	0.554	0.5492	0.5501
10	0.541	0.5294	0.5305
15	0.513	0.5193	0.5229

Table 6

$\mu = 3, p = 1/2, T = 3$

$N \backslash M$	1000	5000	10000
5	0.594	0.6082	0.6025
10	0.661	0.6215	0.6523
15	0.673	0.6854	0.6769

Table 7

$\mu = 2, p = 1/2, T = 5$

$N \backslash M$	1000	5000	10000
5	0.492	0.4886	0.4902
10	0.478	0.4794	0.4761
15	0.398	0.3896	0.3864

Table 8

$\mu = 3, p = 1/2, T = 5$

$N \backslash M$	1000	5000	10000
5	0.675	0.6724	0.6753
10	0.684	0.6802	0.6897
15	0.696	0.6966	0.6941

Table 9

$\mu = 4, p = 1/2, T = 3$

$N \backslash M$	1000	5000	10000
5	0.687	0.6907	0.6925
10	0.703	0.7113	0.7083
15	0.729	0.7328	0.7349

Table 10

$\mu = 4, p = 1/2, T = 5$

$N \backslash M$	1000	5000	10000
5	0.702	0.6996	0.7088
10	0.716	0.7215	0.7196
15	0.741	0.7397	0.7427

Conclusion 1. *From the above mentioned tables we observes that, by increasing of initial number N of errors, the statistical probability of the successful prediction of N based on the maximum likelihood procedure grows and this happens at the same time with the increasing of intensity μ and of duration T for checking time.*

In the same context of the validation of maximum likelihood method it was calculated the rate of successful prediction of the error's intensity value μ with exactitudes $\varepsilon = 0,01$ and $\varepsilon = 0,02$.

For the error $\varepsilon = 0,02$ the following confidence levels have been obtained:
- 74% for $M = 1000$;
- 96% for $M = 5000$;
- 99% for $M = 10000$.

In order to validate the maximum likelihood estimator $\hat{\mu}$ of parameter μ we find, for a given number M of trials (Monte-Carlo simulations), the relative frequency of the cases, when the difference, in the absolute value, between the true value of this parameter and the value of its estimator $\hat{\mu}$ doesn't go beyond the given error ε.

Case 1: $\varepsilon = 0,01$

Table 11

$N = 10, p = 1, T = 3$

$\mu \backslash M$	1000	5000	10000
2	0.131	0.1391	0.1395
3	0.085	0.0843	0.0854
4	0.079	0.0801	0.0789

Table 12

$N = 15, p = 1, T = 5$

$\mu \backslash M$	1000	5000	10000
2	0.195	0.1963	0.1949
3	0.139	0.1401	0.1387
4	0.107	0.1024	0.1085

Table 13
$N = 10, p = 1/2, T = 3$

$\mu\backslash M$	1000	5000	10000
2	0.054	0.0538	0.0554
3	0.032	0.0324	0.0331
4	0.028	0.0294	0.0289

Table 14
$N = 15, p = 1/2, T = 5$

$\mu\backslash M$	1000	5000	10000
2	0.048	0.0479	0.0490
3	0.021	0.0207	0.0211
4	0.018	0.0185	0.0183

Case 2: $\varepsilon = 0,02$

Table 15
$N = 10, p = 1, T = 3$

$\mu\backslash M$	1000	5000	10000
2	0.204	0.2081	0.2074
3	0.165	0.1582	0.1743
4	0.105	0.1142	0.1258

Table 16
$N = 15, p = 1, T = 5$

$\mu\backslash M$	1000	5000	10000
2	0.276	0.2549	0.2653
3	0.195	0.1948	0.1919
4	0.138	0.1379	0.1347

Table 17

$N = 10, p = 1/2, T = 3$

$\mu \backslash M$	1000	5000	10000
2	0.122	0.1483	0.1542
3	0.076	0.0721	0.0773
4	0.049	0.0473	0.0498

Table 18

$N = 15, p = 1/2, T = 5$

$\mu \backslash M$	1000	5000	10000
2	0.060	0.0603	0.0602
3	0.049	0.0486	0.4904
4	0.034	0.0351	0.0343

From the tables 11-18 we draw the

Conclusion 2. *For the two above studied models, the maximum likelihood method is not efficient to predict the intensity parameter μ of software's errors.*

References

[1] Leahu, A., Lupu, E.C., *Software's Reliability with the Random Number of Errors,* Proceedings of Workshops Mathematical Modeling of Environmental and Life Sciences Problems, Ed. Acad. Romane, 2004, 167-172.

[2] Feller, W., *Introduction in Probability Theory and its Applications,* Vol. II, Ed. Wiley&Sons, New-York, 1970.

[3] Vaduva, I., *Fiabilitatea programelor*, Ed. Univ. Bucuresti, 2003, pp. 160 (in Romanian).

Faculty of Mathematics and Informatics,
"Ovidius" University of Constanta,
Mamaia Ave. 124,
P.O.Box 900527, Constanta, Romania
e-mail:aleahu@univ-ovidius.ro

Faculty of Dentar Medicine,
"Ovidius" University of Constanta,
Ilarie Voronca Str.,
P.O.Box 900527, Constanta, Romania
e-mail: celena@yahoo.com