Journal of Probability and Statistical Science
6(1), 61-68, Feb. 2008

Semi-parametric Time-weighted Population-unbiased Estimates for Software Error Intensity

Show-Long Patrick Koh
Columbia University

ABSTRACT For software error intensity, this paper suggests the semi-parametric time-weighted population-unbiased estimate, which sets the integration of time-weighted estimating errors to zero. Two previous estimates, obtained from different data-collecting requirements, are found to be of this kind. These estimates are compared by the expected time-weighted square errors and its integration; accordingly, one of them is suggested.

Keywords Software error intensity; Time-weighted population-unbiased estimate; Mean square error; Expected square error; Integration.

1. Introduction

Software usually contains faults which cause errors to occur. In order to reduce the error intensity (or the error rate), the software is often put through a testing procedure to detect errors and to remove faults. Let $\Lambda(t)$ denote the software's error intensity after being tested for t units of time. A problem of great importance to software manufacturing is to estimate $\Lambda(t)$. This problem has been studied for more than thirty years. It remains an active field of research due to its diversity in results. Most works are of parametric estimation which is difficult due to the infinite possible models for selection. Chen and Singpurwalla [1] give a good partial review on this problem through the effort of model unification.

This paper adopts a semi-parametric model put forth by Robbins [6] and Ross [7]. The model assumes there are constant m faults in the software and each fault is independently causing errors to occur in accordance with Poisson process with unknown error intensity. The model is first suggested by Robbins [6] in a different context and is brought into the software testing by Ross [7]. Since m is a constant parameter but the distribution of the error intensity of individual

---

Received September 2007, revised November 2007, in final form December 2007.
Show-Long Patrick Koh (Ph.D, Columbia University, 1982) was a former Ph.D student of Cyrus Derman; email: slkoh.patrick@msa.hinet.net.

 ISSN 1726-3328

fault is unknown it is classified as semi-parametric. In this setting, there are two cases assumed for the mechanism of fault-removing: 1. the fault is immediately removed at the earliest detection of its error, or 2. the fault is removed only at time t and the number of errors caused by each fault prior to t is observed. Let us call the first case the immediate removal and the second the delayed removal. These two cases differ by the requirement of data-collecting. For the immediate removal, one only need to know the occurrence time of the first error caused by each detected fault, but for the delayed removal, the numbers of errors caused by those faults are also needed. The author feels that, in terms of data-collecting, the immediate removal case is much more practical than the delayed removal case.

An unbiased estimate $\theta_r(t)$ for $\Lambda(t)$ is found by Robbins [6] and Ross [7] for the delayed removal and an asymptotical unbiased estimate $\theta_k(t)$ is found by Koh [3] for the immediate removal. The exact mathematical expressions are as follows.

$$E(\theta_r(t)-\Lambda(t))=0 \quad \text{and} \quad \lim_{t\to\infty} t^2 E_\lambda E(\theta_k(t)-\Lambda(t))=0\,. \qquad \text{(1) and (2)}$$

(2) requires an assumption that the distribution F of the error intensity of each individual fault is identically independently distributed with the property of $F'(0+)>0$. (2) indicates the expected bias converges to zero faster than the reciprocal of t square. In general, $\theta_k(t)$ is not unbiased for the given t. The review of $\theta_r(t)$ and $\theta_k(t)$ will be given in section 2.

It is found in this paper that

$$\int_0^\infty tE[\theta(t)-\Lambda(t)]dt=0\,. \tag{3}$$

holds for both $\theta_r(t)$ and $\theta_k(t)$. (3) says while the bias of $\theta(t)$ is allowed to fluctuate around zero, the integration (or the sum) of the time-weighted bias for the whole population is zero. Let us call $\theta(t)$ in (3) the time-weighted population-unbiased estimate. Since the estimate which satisfies (1) must satisfy (3), (3) holds for $\theta_r(t)$. In section 2, it would be proved that (3) also holds for $\theta_k(t)$ without using the assumption $F'(0+)>0$.

In the practice of software testing, the estimating error of large t should be weighted heavier than that of small t, not only it is closer to the time of releasing or finalizing the software but also it is built-in smaller due to less faults remained in the software. Thus, the weighting on the estimating error by t makes the estimate more relevant to the purpose of software testing and more comparable for all t. By these sense, (3) seems a good criterion.

Weighted estimate for software error intensity is first attempted by Koh [4] for parametric models; however, the weight there is the square root of the reciprocal of the expected variance, instead of the testing time.

The expected square error is often used to evaluate estimates. For estimates in this paper, the expected time-weighted square error and its integration would be analyzed in Sections 3 and 4.

2. $\theta_r(t)$ and $\theta_k(t)$

Let $I_i(t)$ = 0 or 1 according to the i-th fault being removed prior to t or not, and λ_i the error intensity of the i-th fault. The model of Ross [7] is

$$\Lambda(t) = \sum_{i=1}^{m} \lambda_i I_i(t) \quad \text{and} \quad E(\Lambda(t)) = \sum_{i=1}^{m} \lambda_i e^{-\lambda_i t}. \tag{4) and (5}$$

If λ_i's are independently identically distributed as F (see Koh [3]), then

$$E_\lambda E(\Lambda(t)) = m\int_0^\infty \lambda e^{-\lambda t}\, dF(\lambda). \tag{6}$$

For the delayed removal, the Robbins's estimate is

$$\theta_r(t) = M_1(t)/t, \tag{7}$$

where $M_1(t) =$ the number of faults, each causing exactly one error to occur in the time interval [0, t]. (1) holds for $\theta_r(t)$. This can be seen by taking the expectation of

$$M_1 = \sum_{i=1}^{m} J_i(t), \tag{8}$$

where $J_i(t) = 1$ or 0, according to the i-th fault causing exactly one error to occur prior to t or not. Derman and Koh [2] compare $\theta_r(t)$ with another estimate put forth by Ross [7] and conclude that $\theta_r(t)$ is a pretty good estimate by the sense of expected square error (or equivalently, mean square error), especially for large m and large t. Note that, in [2], t is set to be one, so the concept of large t here is equivalent to the small $\sum \lambda_i / m$ in the context of [2]. Nayak [5] further shows that $\theta_r(t)$ is actually the minimum variance unbiased estimate.

For the immediate removal, Koh [3] suggests a family of estimates as follows:

$$\begin{aligned}\theta(t, \beta t) &= \left(\text{number of errors detected in } [0, t] - \text{number of errors detected in } [0, \beta t]\right)/t \\ &= \frac{1}{t}\left(\sum_{i=1}^{m}[1 - I_i(t)] - \sum_{i=1}^{m}[1 - I_i(\beta t)]\right) \\ &= \frac{1}{t}\sum_{i=1}^{m}[I_i(\beta t) - I_i(t)],\end{aligned} \tag{9}$$

where $0 < \beta < 1$. By taking the expectation of (9),

$$E[\theta(t, \beta t)] = \frac{1}{t}\sum_{i=1}^{m}(e^{-\beta\lambda_i t} - e^{-\lambda_i t}), \tag{10}$$

and together with (5),

$$E(\theta(t,\beta t)-\Lambda(t))=\sum_{i=1}^{m}\frac{e^{-\beta\lambda_i t}-e^{-\lambda_i t}}{t}-\lambda_i e^{-\lambda_i t}=\sum_{i=1}^{m}B_t(\beta,\lambda_i), \tag{11}$$

where $B_t(\beta,\lambda)=[(e^{-\beta\lambda t}-e^{-\lambda t})/t]-\lambda e^{-\lambda t}$, which is the bias of a single fault for the given λ. In Koh [3], it is proved that if λ_i's are independently identically distributed as F, then

$$\lim_{t\to\infty} t^2 E_\lambda B_t(\beta,\lambda)=(\frac{1}{\beta}-2)F^{'}(0+). \tag{12}$$

Thus, by assuming $F^{'}(0+)>0$, $\beta=1/2$ and letting $\theta_k(t)=\theta(t,t/2)$, (2) holds.

By (11), the time-weighted bias and its integration are

$$E(t(\theta(t,\beta t)-\Lambda(t)))=\sum_{i=1}^{m} tB_t(\beta,\lambda_i) \tag{13}$$

and

$$\int_0^\infty E(t(\theta(t,\beta t)-\Lambda(t)))dt=\sum_{i=1}^{m}\int_0^\infty tB_t(\beta,\lambda_i)dt\ . \tag{14}$$

Each term in (14) is

$$\int_0^\infty tB_t(\beta,\lambda_i)\,dt=\int_0^\infty (e^{-\beta\lambda_i t}-e^{-\lambda_i t}-\lambda_i te^{-\lambda_i t})dt=\frac{1}{\lambda_i}(\frac{1}{\beta}-2). \tag{15}$$

Thus, (3) holds for $\theta_k(t)=\theta(t,\beta t)$ when $\beta=1/2$. One of the advantages of the criterion (3) and (15) is that $F'(0+)>0$ is no longer needed. Recall that (3) also holds for $\theta_r(t)$. It then concludes both $\theta_k(t)$ and $\theta_r(t)$ are time-weighted population-unbiased.

It is always desirable to see the expected square error (or mean square error) of any estimate. For the time-weighted population-unbiased estimates $\theta_k(t)$ and $\theta_r(t)$, the expected time-weighted square error and its integration over testing time are both considered as follows.

3. The Expected Time-weighted Square Error

For a given t, let $MS_k(t,\beta)$ and $MS_r(t)$ denote the expected time-weighted square error of $\theta(t,\beta t)$ and $\theta_r(t)$ respectively.

$$\begin{aligned}MS_k(t,\beta)&=E(t^2(\theta(t,\beta t)-\Lambda(t))^2)\\&=\sum_{i=1}^{m}t^2E((\theta_i(t,\beta t)-\Lambda_i(t))^2)+2\sum_{i>j}t^2E(\theta_i(t,\beta t)-\Lambda_i(t))E(\theta_j(t,\beta t)-\Lambda_j(t))\end{aligned} \tag{16}$$

in which

$$\begin{aligned}E(t^2(\theta_i(t,\beta t)-\Lambda_i(t))^2)&=E((I_i(\beta t)-I_i(t)-\lambda_i tI_i(t))^2)\\&=E(I_i(\beta t)-I_i(t))+\lambda_i^2t^2E(I_i(t))\quad (\text{by } I_i(\beta t)I_i(t)=I_i(t)\end{aligned}$$

and $I_i(t)^2 = I_i(t).$)

$$= e^{-\beta\lambda_i t} - e^{-\lambda_i t} + \lambda_i{}^2 t^2 e^{-\lambda_i t}$$

$$= tB_t(\beta,\lambda_i) + \lambda_i t e^{-\lambda_i t} + \lambda_i{}^2 t^2 e^{-\lambda_i t}. \tag{17}$$

Since

$$\begin{aligned} MS_r(t) &= E(t^2(\theta_r(t) - \Lambda(t))^2) \\ &= t^2 E((\theta_r(t) - \Lambda(t))^2) \\ &= t^2 (\sum_{i=1}^{m} \frac{\lambda_i}{t} e^{-\lambda_i t} + \lambda_i^2 e^{-\lambda_i t}) \text{ (see Robbins [6] and Ross[7])} \\ &= \sum_{i=1}^{m} \lambda_i t e^{-\lambda_i t} + \lambda_i^2 t^2 e^{-\lambda_i t} \end{aligned} \tag{18}$$

(16), (17) and (18) together shows

$$MS_k(t,\beta) = MS_r(t) + t\sum_{i=1}^{m} B_t(\beta,\lambda_i) + 2\sum_{j>i}[tB_t(\beta,\lambda_i)][tB_t(\beta,\lambda_j)]. \tag{19}$$

By taking expectation of (19) with respect to λ ,

$$E_\lambda MS_k(t,\beta) = E_\lambda MS_r(t) + mE_\lambda tB_t(\beta,\lambda) + m(m-1)[E_\lambda tB_t(\beta,\lambda)]^2 \tag{20}$$

$$\le E_\lambda MS_r(t) + mE_\lambda tB_t(\beta,\lambda) + m(m-1)E_\lambda[t^2 B_t(\beta,\lambda)^2] \tag{21}$$

(by Cauchy Schwartz Inequality)

Since

$$\lim_{t\to\infty} E_\lambda tB_t(\beta,\lambda) = 0\,, \tag{22}$$

(20) indicates that if $E_\lambda tB_t(\beta,\lambda) < 0$ then $E_\lambda MS_k(t,\beta) < E_\lambda MS_r(t)$ for large *t*. This property holds for some *F*; for example, if *F* is uniformly distributed on [0, *L*], then

$$E_\lambda tB_t(\beta,\lambda) = (\frac{-e^{L(1-\beta)t}}{\beta tL} + \frac{2}{tL} + 1)e^{-Lt} - (2 - \frac{1}{\beta})\frac{1}{tL}\,. \tag{23}$$

Letting $\beta = 1/2$,

$$E_\lambda tB_t(1/2,\lambda) = \frac{-\frac{2}{2!}(\frac{tL}{2})^2 - \frac{2}{3!}(\frac{tL}{2})^3 - \dots}{tL} e^{-Lt} < 0. \tag{24}$$

Thus, by (20), (22) and (24), when *t* is large and *F* is uniform, the expected time-weighted square error of $\theta_k(t)$ is smaller than that of $\theta_r(t)$. Koh [3] already proves that, under the same condition, the un-weighted mean square error of $\theta_k(t)$ is also smaller than that of $\theta_r(t)$.

This section concludes that even for large *t* under which $\theta_r(t)$ is known to be good (see Derman and Koh [2]), $\theta_k(t)$ is still competitive.

4. The Integrated Expected Time-weighted Square Error

The time integration of the first term of (20) is

$$\int_0^\infty E_\lambda MS_r(t)dt = 3m\int_0^\infty \frac{1}{\lambda}dF(\lambda), \tag{25}$$

and of the second term is

$$(\frac{1}{\beta}-2)m\int_0^\infty \frac{1}{\lambda}dF(\lambda)\,, \tag{26}$$

but of the third term is hard to evaluate. Note that the integration in (25) and (26) is finite if F is empirical and m is finite; that is,

$$\int_0^\infty \frac{1}{\lambda}dF(\lambda) = \sum_{i=1}^m \frac{1}{\lambda_i}\frac{1}{m} < \infty.$$

(21) is obtained by taking an upper bound of the third term of (20). By taking the time integration of the third term in (21),

$$\int_0^\infty\int_0^\infty t^2 B_t(\beta,\lambda)^2\, dF(\lambda)dt$$

$$= \int_0^\infty\int_0^\infty (e^{-2\beta\lambda t} + e^{-2\lambda t} + \lambda^2 t^2 e^{-2\lambda t} - 2e^{-(\beta+1)\lambda t} - 2\lambda t e^{-(\beta+1)\lambda t} + 2\lambda t e^{-2\lambda t})dF(\lambda)dt$$

$$= \int_0^\infty [\int_0^\infty (e^{-2\beta\lambda t} + e^{-2\lambda t} + \lambda^2 t^2 e^{-2\lambda t} - 2e^{-(\beta+1)\lambda t} - 2\lambda t e^{-(\beta+1)\lambda t} + 2\lambda t e^{-2\lambda t})dt\,]dF(\lambda)$$

$$= (\frac{1}{2\beta} - \frac{2}{(\beta+1)} - \frac{2}{(\beta+1)^2} + \frac{5}{4})\int_0^\infty \frac{dF(\lambda)}{\lambda} \tag{27}$$

and

$$\frac{d}{d\beta}\int_0^\infty\int_0^\infty t^2 B_t(\beta,\lambda)^2\, dF(\lambda)dt = \frac{-1-3\beta+9\beta^2+3\beta^3}{2\beta^2(1+\beta)^3}\int_0^\infty \frac{dF(\lambda)}{\lambda}$$

$$= 0,\ \text{if}\ \beta = -3.2743, -0.2101\ \text{or}\ 0.4845\,. \tag{28}$$

The only zero of the above derivative located in the allowable range is 0.4845. Together with

$$\frac{d^2}{d\beta^2}\int_0^\infty E_\lambda(t^2 B_t(0.4845,\lambda)^2)\,dt > 0\ ,$$

$\int_0^\infty\int_0^\infty t^2 B_t(\beta,\lambda)^2\, dF(\lambda)dt$ is minimized at $\beta = 0.4845,$ which is close to $\beta = 0.5.$ In fact,

$$\int_0^\infty \int_0^\infty t^2 B_t(0.4845,\lambda)^2\, dF(\lambda)dt = 0.0272 \int_0^\infty \frac{dF(\lambda)}{\lambda} \tag{29}$$

and

$$\int_0^\infty \int_0^\infty t^2 B_t(0.5,\lambda)^2\, dF(\lambda)dt = 0.0278 \int_0^\infty \frac{dF(\lambda)}{\lambda}. \tag{30}$$

Thus, by (21), (25), (26), (29) and (30),

$$\int_0^\infty E_\lambda MS_k(t,\, 0.5)\, dt \le (3m + 0.0278m(m-1)) \int_0^\infty \frac{dF(\lambda)}{\lambda}, \tag{31}$$

and

$$\int_0^\infty E_\lambda MS_k(t,\, 0.4845)\, dt \le (3.064m + 0.0272m(m-1)) \int_0^\infty \frac{dF(\lambda)}{\lambda}. \tag{32}$$

If the upper bounds of (31) and (32) are believed to be good enough, then $\theta_k(t) = \theta(t, t/2)$ is favored when m < 107.7 or $\theta(t,\, 0.4845t)$ is favored when m > 107.7.

5. Conclusion

Two estimates in literature, $\theta_r(t)$ and $\theta_k(t)$ are proved to be time-weighted population-unbiased and an assumption originally required by $\theta_k(t)$ is relieved. Between these two estimates, $\theta_k(t)$ is favored for two reasons:

1) data-collecting requirement of $\theta_k(t)$ is much milder than that of $\theta_r(t)$,
2) when the distribution of the error intensity of individual faults is uniform the expected time-weighted square error of $\theta_k(t)$ is better than that of $\theta_r(t)$ for large t.

$\theta_k(t)$ is chosen from the family of $\theta(t, \beta t)$, $0 < \beta < 1$ with $\beta = 1/2$. By the criterion of integrated expected time-weighted square error, $\beta = 0.4845$ is favored if m is expected to be larger than 107.7, or else $\beta = 1/2$ is favored; however, since the difference in β is small and (31) and (32) are merely upper bounds, 1/2 seems the most reasonable value for β.

References

[1] Chen, Y. and Singpurwalla, N. (1997). Unification of software reliability models by self-exciting point process, *Advanced Applied Probability*, **29**, 337-352.

[2] Derman, C. and Koh S. P. (1988). On the comparison of two software reliability estimates, *Probability in the Engineering and Informational Sciences*, **2**, 15-21.

[3] Koh, S. P. (1994). Software reliability model: a simple estimate for error intensity, *Asia-Pacific Journal of Operational Research*, **11**, 67-77.

[4] Koh, S. P. (2008). Weighted least-square estimate for software error intensity, *Journal of the Chinese Institute of Industrial Engineers*, 25, 162-173.

[5] Nayak, T. K. (1989). Minimum variance unbiased estimation of software reliability, *Probability in the Engineering and Informational Sciences*, **3,** 335-340.

[6] Robbins, H. (1968). Estimating the probability of unbounded events, *Annals of Mathematical Statistics*, **39**, 256-257.

[7] Ross, S. M. (1985), Statistical estimation of software reliability, *IEEE Transactions on Software Engineering*, **SE-11**, 1479-1483.