Software Reliability based on Renewal Process Modeling for Error Occurrence due to Each Bug with Periodic Debugging Schedule

Technical Report No. ASU/2019/2
Dated 30 September, 2019

Sudipta Das

Anup Dewanji

and

Subrata Kundu

Applied Statistics Unit
Indian Statistical Institute
Kolkata 700 108

Software Reliability based on Renewal Process Modeling for Error Occurrence due to Each Bug with Periodic Debugging Schedule

Sudipta Das

e-mail: jusudipta@gmail.com

Anup Dewanji

e-mail: dewanjia@isical.ac.in

Subrata Kundu

e-mail: kundu@gwu.edu

Abstract

In this article, we discuss continuous time testing of a software with periodic debugging in which bugs are corrected, instead of at the instants of their detection, at some pre-specified time points. Under the assumption of renewal distribution for the time between successive occurrence of a bug, maximum likelihood estimation of the initial number of bugs in the software is considered when the renewal distribution belongs to a parametric family or is arbitrary. The asymptotic properties of the estimated model parameters are also discussed. Finally, we investigate the finite sample properties of the estimators, specially that of the number of initial number of bugs, through simulation.

Keywords– Software testing, Periodic debugging, Renewal process, Parametric estimation, Non-parametric estimation, Asymptotic distribution.

1 Introduction

After Jelinski and Moranda[10] introduced the pioneer model of software reliability, several varieties of software reliability models were proposed by Singpurwalla and Wilson

[14], Lyu [11], Gokhale et al. [7], Dalal [1] and many other researchers. Among the different assumptions made by the different researchers, one assumption that the software bugs were removed as and when they are detected has been common. However, in reality, there exists a time gap between a software bug detection and its removal. Hence, the software reliability modeling with delayed debugging also attract interests and studied by Hwang and Pham [9], Huang and Lin [8], Sun et. al. [15] and others. In the context of delayed debugging, in many practical in-house testing situations, there are some prefixed time points when debugging of bugs, which are detected since the previous debugging, takes place. The software testing goes on till the last debugging time. Therefore, in this kind of situations the bugs are not removed as and when they are detected causing errors/failures of the software during testing; only a record of these detected bugs is maintained and these are removed, with certainty and introducing no other new bug, at the subsequent scheduled debugging time. This kind of software testing and data collection technique, named as 'periodic debugging schedule', was first studied by Dewanji et al. [6]. It is to be noted that, this kind of periodic debugging schedule can be commonly seen when subsequent versions of the software are released at different times and testing continues with the most recent version.

The software reliability with periodic debugging has got scanty attention in the literature. Das et. al [2] has considered such periodic debugging schedule under the homogeneous Poisson process (HPP) assumption. Later, the discrete software reliability model is also developed by Das et.al. [3]. The main objective of this work is to study the software reliability model, under the periodic debugging schedule, considering that the bug detection process follows a general renewal process. This work, therefore, generalizes the HPP assumption (that is, the Exponential renewal distribution) of Das et al. [2] by considering any general renewal distribution. In this work, we focus on estimating ν, the number of initial bugs in the software. Under the renewal process assumption, the maximum likelihood (ML) estimator of ν as well as its asymptotic distribution have been obtained under both the parametric and non-parametric frameworks. Therefore, this work can be viewed as a generalization, in the context of periodic debugging, of Dewanji et. al. [5] and Dewanji et. al. [4], who have developed parametric and non-parametric methods, respectively, for estimating ν based on software testing data with instantaneous debugging and recapture sampling, thus allowing more information than our periodic debugging schedule. In fact, the data configuration of this work reduces to that of [5] and [4] when there is only one single debugging period.

The article is organized as follows. In Section 2, we describe the data arising from the periodic debugging schedule under a continuous time scale and construct the likelihood function under the renewal process assumption. Computational methods for obtaining a parametric and a non-parametric estimate of ν along with the corresponding asymptotic properties are provided in Sections 3 and 4, respectively. Section 5 reports results of a simulation study to investigate the properties of the estimators developed in Sections 3 and 4. Section 6 contains some concluding remarks.

2 Modeling and Likelihood

We assume that there are initially ν bugs in the software. The failure times of these bugs are independent and identical, following a renewal process with common renewal density function $f(\cdot)$. In periodic debugging schedule, there are k prefixed time points, $0 < \tau_1 < \cdots < \tau_k < \infty$, at which debugging is scheduled to take place. Suppose we observe the total number of error occurrence, hereafter referred to as failures, $M(\geq 0)$ between $\tau_0 = 0$ and τ_k along with the identities of the corresponding bugs. The M failures may correspond to fewer than M distinct bugs, since a particular bug may trigger more than one failure. Let $M^{(d)}(\geq 0)$ denote the total number of distinct detected bugs, resulting in M failures. Suppose that the i^{th} distinct bug appears for the first time in the l_ith interval $(\tau_{l_i-1}, \tau_{l_i}]$, for $i = 1, \ldots, M^{(d)}$; in addition, suppose that it appears

$m_i - 1$ more times (totaling m_i times) in that interval before being debugged at τ_{l_i}, with certainty and in a negligible amount of time without introducing any other new bug. We also observe t_{i_j}, the inter-event (renewal) times of these m_i failures, for $j = 1, \ldots, m_i$, with t_{i_1} being the first appearance time in $(\tau_{l_i-1}, \tau_{l_i}]$. Therefore, $\sum_{j=1}^{m_i} t_{i_j} \leq \tau_{l_i}$, and the difference $\tau_{l_i} - \sum_{j=1}^{m_i} t_{i_j}$ is denoted by c_i, the last censored event time for the i^{th} detected bug. It is to be noted that $\sum_{i=1}^{M^{(d)}} m_i = M$.

Following Nayak [12], the likelihood function can be expressed as

$$L(\nu, f) \quad = \quad \frac{\nu!}{(\nu - M^{(d)})!} \left[\bar{F}(\tau_k)\right]^{\nu - M^{(d)}} \prod_{i=1}^{M^{(d)}} \left[\left(\prod_{j=1}^{m_i} f(t_{i_j}) \right) \bar{F}(c_i) \right], \tag{1}$$

where f and $\bar{F}$ are the pdf and the survival function, respectively, of an arbitrary renewal distribution. It is to be noted that, ν is the parameter of primary interest while the renewal distribution may be treated as nuisance. In the following two sections, we develop a parametric and a non-parametric method, respectively, for estimating ν.

3 Parametric Estimation

Let us assume that the renewal distribution has a parametric form with the associated parameters $\underline{\phi}$, (possibly, vector valued). Therefore, the likelihood function in (1) will be an explicit function of $\nu, \underline{\phi}$ and can be written as

$$L(\nu, \underline{\phi}) \quad = \quad \frac{\nu!}{(\nu - M^{(d)})!} \left[\bar{F}(\tau_k, \underline{\phi})\right]^{\nu - M^{(d)}} \prod_{i=1}^{M^{(d)}} \left[\left(\prod_{j=1}^{m_i} f(t_{i_j}, \underline{\phi}) \right) \bar{F}(c_i, \underline{\phi}) \right]. \tag{2}$$

In order to find the maximum likelihood estimator (MLE) of ν the likelihood function in (2) needs to be maximized w.r.t. ν and $\underline{\phi}$. For a fixed value of $\underline{\phi}$, it can be seen that (2) is maximized over ν at

$$\nu(\underline{\phi}) = \lfloor \frac{M^{(d)}}{1 - \bar{F}(\tau_k, \underline{\phi})} \rfloor, \tag{3}$$

where $\lfloor z \rfloor$ is the largest integer less than or equal to z. However, for a fixed value of ν, maximization of the likelihood function in (2) is needed to be maximized numerically, for most of the renewal distributions. We propose the following algorithm to find the MLEs of ν and $\underline{\phi}$.

Algorithm 1:

- Step 1: Start with an initial estimate ν_0 of ν. Since $\nu \geq M^{(d)}$, we take $\nu_0 = M^{(d)}$.
- Step 2: Obtain $\underline{\phi}_0 = \hat{\underline{\phi}}(\nu_0)$ by numerically maximizing (2) at $\nu = \nu_0$.
- Step 3: Obtain $\nu_1 = \hat{\nu}(\underline{\phi}_0)$ using (3).
- Step 4: Go to Step 2 with ν_1 replacing ν_0 and iterate until it converges.

It is to be noted that the MLE of ν and $\underline{\phi}$ may not always exist. As an example, for Exponential renewal distribution, a necessary and sufficient condition for existence and uniqueness of MLE is $M \neq M^{(d)}$ or $M \leq 1 + \frac{2}{\tau_k} \sum_{i=1}^{k} \left(\sum_{j=1}^{i-1} \sum_{l=1}^{\nu} I_{\{t_{l_1} \in (\tau_{j-1}, \tau_j]\}} \right) (\tau_i - \tau_{i-1})$ (See Das. et. al. [2]). Clearly, by the nature of it, the likelihood function (2) is non-decreasing over the successive steps of Algorithm 1. However, it is to be noted that this algorithm often fails to converge to the MLE due to the discrete nature of the parameter ν. In particular, in the iterative step, sometime the change in the value of $\frac{M^{(d)}}{1-\bar{F}(\tau_k, \underline{\phi})}$ in (3) is too small to make a change in its integer part. As a result, the updated value of ν in (3) remains unchanged. Thus, the above procedure gets stuck at some value of ν and the corresponding $\underline{\phi}$, which are not the true MLE. Therefore, one can incorporate a slight modification to the algorithm, as described next. Specifically, in step 3 of *Algorithm 1*, one can use the actual value of $\frac{M^{(d)}}{1-\bar{F}(\tau_k, \underline{\phi})}$, instead of its integer part, to update the estimate of ν. When we stop the process, based on convergence, we take the integer part of the latest estimate of ν for the MLE of ν. Alternatively, one can maximize (2) by direct search by checking it for all $\nu \geq M^d$.

In order to derive the asymptotic properties of the parametric MLE of $\underline{\phi}$ and ν, denoted by $\hat{\underline{\phi}}$ and $\hat{\nu}_{pa}$, respectively, obtained by the method described above, as $\nu \to \infty$, we follow the general results of Dewanji et al. [5]. In particular, the primary interest lies in the asymptotic distribution of $\hat{\nu}_{pa}$. In the software reliability context, this limiting context is reasonable as most of the software with practical significance contain thousands of lines of code with large number of bugs.

Let the ν bugs be labeled as $1, \cdots, \nu$ and let X_j denote the observation from the jth bug (possibly unobservable). If the jth bug is not detected, let us write $X_j = 0$; otherwise, X_j consists of the debugging time τ_{l_j} (say) of the jth bug and the number m_j of times it appears in $[\tau_{l_j-1}, \tau_{l_j})$ along with the times of failures $\{t_{j_1} < \cdots < t_{j_{m_j}}\}$. The labeling of X_j's are clearly hypothetical since the labeling $1, \cdots, \nu$ is not observed. However, the X_j's are independent and identically distributed with the common probability distribution given by

$$p_X(x_j; \underline{\phi}) = \begin{cases} \bar{F}(\tau_k, \underline{\phi}), & \text{if } x_j = 0 \\ \bar{F}(c_j, \underline{\phi}) \prod_{s=1}^{m_j} f(t_{j_s}, \underline{\phi}), & \text{otherwise,} \end{cases}$$

where $c_j = \tau_{l_j} - t_{j_{m_j}}$. The joint distribution of $(X_1, \cdots, X_\nu)$ is, therefore, $\prod_{j=1}^{\nu} p_X(x_j; \underline{\phi})$, which is proportional to $L(\nu, \underline{\phi})$ in (2).

For a finite dimensional $\underline{\phi} = [\phi_1, \ldots, \phi_p]^T$, similar to the technique used in Dewanji et al.[5], we define $\underline{V_j} = [V_{1j}, V_{2j}, \ldots, V_{p+1,j}]^T$ where

$$V_{1j} = \begin{cases} -\frac{1-\bar{F}(\underline{\phi}, \tau_k)}{\bar{F}(\underline{\phi}, \tau_k)}, & \text{if } x_j = 0 \\ 1, & \text{otherwise;} \end{cases}$$

and

$$V_{h+1,j} = \frac{\partial}{\partial \phi_h} \log \ p_X(x_j; \underline{\phi})$$

for $h = 1, \ldots, p$ and $j = 1, \ldots, \nu$.

It can be verified that $E[\underline{V}_j] = \underline{0}$ and their variances are as follows

$$Var[V_{1j}] = \frac{1 - \bar{F}(\underline{\phi}, \tau_k)}{\bar{F}(\underline{\phi}, \tau_k)}$$

and

$$Var[V_{h+1,j}] = E\left[-\frac{\partial^2}{\partial \phi_h^2} \ \log \ p_X(x_j; \underline{\phi})\right] = I_{h+1,h+1}(\underline{\phi}), \text{ say},$$

for $h = 1, \ldots, p$ and $j = 1, \ldots, \nu$. It can also be verified that

$$Cov[V_{1j}, V_{h+1,j}] \ = \ \frac{\partial}{\partial \phi_h} \log \bar{F}(\underline{\phi}, \tau_k) = I_{1,h+1}(\underline{\phi}), \text{ say},$$

and

$$Cov[V_{g+1,j}, V_{h+1,j}] \ = \ E\left[-\frac{\partial^2}{\partial \phi_g \partial \phi_h} \log \ p_X(x_j; \underline{\phi})\right] = I_{g+1,h+1}(\underline{\phi}), \text{ say},$$

for $g, h = 1, \ldots, p$ and $j = 1, \ldots, \nu$. Therefore, writing

$$u_{h,\nu} = \nu^{-1/2} \sum_{j=1}^{\nu} V_{h,j}, \text{ for } h = 1, \ldots, p+1,$$

we have, by the central limit theorem,

$$\underline{u}_\nu \ = \ \begin{pmatrix} u_{1\nu} \\ \vdots \\ u_{p+1,\nu} \end{pmatrix} = \nu^{-1/2} \sum_{j=1}^{\nu} \begin{pmatrix} V_{1j} \\ \vdots \\ V_{p+1,j} \end{pmatrix} \xrightarrow{L} N(\underline{0}, \Sigma^{-1}), \text{ as } \nu \to \infty,$$

where

$$\Sigma = \begin{pmatrix} \frac{1-\bar{F}(\underline{\phi},\tau_k)}{\bar{F}(\underline{\phi},\tau_k)} & I_{12}(\underline{\phi}) & \ldots & I_{1,p+1}(\underline{\phi}) \\ I_{12}(\underline{\phi}) & I_{22}(\underline{\phi}) & \ldots & I_{2,p+1}(\underline{\phi}) \\ \vdots & \vdots & \ddots & \vdots \\ I_{1,p+1}(\underline{\phi}) & I_{2,p+1}(\underline{\phi}) & \ldots & I_{p+1,p+1}(\underline{\phi}) \end{pmatrix}^{-1}.$$

For bounded $(a_1, \underline{a})$, where $\underline{a} = [a_2, \ldots, a_{p+1}]^T$, following the technique of Dewanji et al. [5] and writing $l(\nu, \underline{\phi}) = \log L(\nu, \underline{\phi})$, we consider

$$l(\nu + \nu^{1/2} a_1, \underline{\phi} + \nu^{-1/2} \underline{a}) - l(\nu, \underline{\phi}),$$

which can be reduced to

$$\sum_{h=1}^{p+1} a_h u_{h,\nu} - \sum_{h=2}^{p+1}\sum_{g=1}^{h-1} a_g a_h I_{h,g}(\underline{\phi}) - \frac{a_1^2}{2}\left(\frac{1-\bar{F}(\underline{\phi},\tau_k)}{\bar{F}(\underline{\phi},\tau_k)}\right) - \sum_{h=2}^{p+1}\frac{a_h^2}{2} I_{h,h}(\underline{\phi}) + o_p(1).$$

Then, using the argument of Sen and Singer ([13], p207), as in Dewanji et al. [5], we have the following result.

Result 1: As $\nu \to \infty$,

$$[\nu^{-1/2}(\hat{\nu}_{pa} - \nu), \nu^{1/2}(\hat{\underline{\phi}} - \underline{\phi})] \xrightarrow{L} N(0, \Sigma),$$

where the covariance matrix Σ can be consistently estimated by

$$\hat{\Sigma} = \begin{pmatrix} \frac{1-\bar{F}(\hat{\underline{\phi}},\tau_k)}{\bar{F}(\hat{\underline{\phi}},\tau_k)} & -\frac{\partial}{\partial \underline{\phi}^T}\log \bar{F}(\hat{\underline{\phi}},\tau_k) \\ -\frac{\partial}{\partial \underline{\phi}}\log \bar{F}(\hat{\underline{\phi}},\tau_k) & -\hat{\nu}_{pa}^{-1}\frac{\partial^2}{\partial \underline{\phi}\partial \underline{\phi}^T}\log L(\hat{\nu}_{pa},\hat{\underline{\phi}}) \end{pmatrix}^{-1}.$$

In particular, the asymptotic variance of $\hat{\nu}_{pa}$ can be consistently estimated by

$$\hat{\nu}_{pa}\left[\frac{1-\bar{F}(\hat{\underline{\phi}},\tau_k)}{\bar{F}(\hat{\underline{\phi}},\tau_k)} - \hat{\nu}_{pa}\frac{\partial}{\partial \underline{\phi}^T}\log \bar{F}(\hat{\underline{\phi}},\tau_k)\left(\frac{\partial^2}{\partial \underline{\phi}\partial \underline{\phi}^T}\log L(\hat{\nu}_{pa},\hat{\underline{\phi}})\right)^{-1}\frac{\partial}{\partial \underline{\phi}}\log \bar{F}(\hat{\underline{\phi}},\tau_k)\right]^{-1}. \quad (4)$$

4 Nonparametric Estimation

To find the non-parametric MLE of ν one needs to maximize the likelihood function (1) with respect to ν and the density function f. Toward this, one notices that the renewal distribution f is not involved in the first factor $\frac{\nu!}{(\nu-M^{(d)})!}$ in (1). Therefore, in order to estimate f for a given ν, the product of all terms, excluding $\frac{\nu!}{(\nu-M^{(d)})!}$, should be considered. It can be seen that the product of those terms has the form of the Kaplan and Meier likelihood function with censored data. Thus, as in the case of Kaplan–Meier estimator, it is sufficient to consider all renewal distributions with mass concentrated at the observed renewal times $\{t_{i_j}, j = 1, \ldots, m_i, i = 1, \ldots, M_i^{(d)}\}$, and possibly one extra time point greater than the largest debugging time, τ_k. Let $y_1 < \ldots < y_n$ denote the distinct ordered values of the time points t_{i_j}s and f_l denote the frequency of y_l, for $l = 1, \ldots, n$. We shall consider all the probability distributions with sample space $\{y_1, \ldots, y_n, y_{n+1}\}$, where $y_{n+1} > \tau_k$ is a suitably chosen time point.

Now, by putting $\pi_l = P(X = y_l)$, $l = 1, \ldots, n+1$, with $\sum_{l=1}^{n+1} \pi_l = 1$, the likelihood function (1) can be written as a function of ν and $\underline{\pi}$ as

$$L(\nu, \underline{\pi}) = \frac{\nu!}{(\nu - M^{(d)})!} [\pi_{n+1}]^{\nu - M^{(d)}} \prod_{l=1}^{n} (\pi_l)^{f_l} \left[\prod_{i=1}^{M^{(d)}} \left(\sum_{x_h > c_i} \pi_h \right) \right], \tag{5}$$

where $\underline{\pi} = (\pi_1, \cdots, \pi_{n+1})^T$. This likelihood function (5) can be maximized to obtain the MLE of ν and $\underline{\pi} = [\pi_1, \ldots, \pi_{n+1}]^T$. To avoid the maximization with the constraint $\sum_{l=1}^{n+1} \pi_l = 1$, however, it is more convenient to work with the discrete hazard components

$$\lambda_l = \pi_l / \left(\sum_{j=l}^{n+1} \pi_j \right), \; l = 1, \ldots, n,$$

instead of the probability masses. Note that the transformation from $[\pi_1, \ldots, \pi_{n+1}]$ to $[\lambda_1, \ldots, \lambda_n]$ is one-to-one, with $\pi_1 = \lambda_1$,

$$\pi_l = \lambda_l \prod_{j=1}^{l-1} (1 - \lambda_j), l = 2, \ldots, n \text{ and } \pi_{n+1} = \prod_{j=1}^{n} (1 - \lambda_j).$$

Also note that, $\pi_l + \ldots + \pi_{n+1} = \prod_{j=1}^{l-1} (1 - \lambda_j)$. Hence, likelihood function (5) can be written in terms of the discrete hazards $\lambda_1, \lambda_2, \ldots, \lambda_n$ as

$$\begin{aligned} L(\nu, \underline{\lambda}) &= \frac{\nu!}{(\nu - M^{(d)})!} \left[\prod_{j=1}^{n} (1 - \lambda_j) \right]^{(\nu - M^{(d)})} \prod_{l=1}^{n} \left[\lambda_l \prod_{j=1}^{l-1} (1 - \lambda_j) \right]^{f_l} \left[\prod_{i=1}^{M^{(d)}} \left(\prod_{j=1}^{[c_i]} (1 - \lambda_j) \right) \right] \\ &= \frac{\nu!}{(\nu - M^{(d)})!} \left[\prod_{l=1}^{n} (\lambda_l)^{f_l} \right] \left[\prod_{l=1}^{n} (1 - \lambda_l)^{c_l(\nu)} \right], \end{aligned} \tag{6}$$

where $\lfloor c_i \rfloor$ is the largest integer less than or equal to c_i, $c_l(\nu) = (\nu - M^{(d)}) + \sum_{u=l+1}^{n} f_u + k_l$ with $k_l = \{i : \lfloor c_i \rfloor \geq y_l\}$.

For a fixed ν, the likelihood function (6) is maximized by

$$\hat{\lambda}_l(\nu) = \frac{f_l}{f_l + c_l(\nu)}, \quad l = 1, \ldots, n, \tag{7}$$

and by substituting (7) in (6), we get the profile likelihood function $L_1(\nu) = L(\nu, \hat{\underline{\lambda}}(\nu)))$ as

$$L_1(\nu) \quad = \quad \frac{\nu!}{(\nu - M^{(d)})!} \left[\prod_{l=1}^{n} (\hat{\lambda}_l(\nu))^{f_l} \right] \left[\prod_{l=1}^{n} \left(1 - \hat{\lambda}_l(\nu)\right)^{c_l(\nu)} \right]. \tag{8}$$

The MLE of ν, to be denoted by $\hat{\nu}_{np}$, can be obtained by maximizing $L_1(\nu)$ with respect to ν. An iterative procedure, as described next, is used for finding $\hat{\nu}_{np}$. Note that, for a fixed $\underline{\lambda}$,

$$\frac{L(\nu + 1, \underline{\lambda})}{L(\nu, \underline{\lambda})} = \frac{\nu + 1}{\nu + 1 - M^{(d)}} \prod_{l=1}^{n} (1 - \lambda_l) \geq \text{ or } < 1$$

if and only if

$$\nu \leq \text{ or } > M^{(d)} \left[1 - \prod_{l=1}^{n} (1 - \lambda_l) \right]^{-1} - 1$$

This gives the maximum likelihood estimate of ν, for a given $\underline{\lambda}$, as

$$\nu(\underline{\lambda}) = \left\lfloor M^{(d)} \left[1 - \prod_{l=1}^{n} (1 - \lambda_l) \right]^{-1} \right\rfloor. \tag{9}$$

The estimates (7) and (9) together suggests the following algorithm for estimating ν and $\underline{\lambda}$.

Algorithm 2:

- Step 1: Start with an initial estimate ν_0 of ν. Since $\nu \geq M^{(d)}$, we may take $\nu_0 = M^{(d)}$.
- Step 2: Obtain $\underline{\lambda}_0 = \hat{\underline{\lambda}}(\nu_0)$ using (7).
- Step 3: Obtain $\nu_1 = \hat{\nu}(\underline{\lambda}_0)$ using (9).
- Step 4: Go to Step 2 with ν_1 replacing ν_0 and iterate until it converges.

Similar to the parametric estimation in the previous section, while updating ν in step 3 of *Algorithm 2*, one can use the actual value of $M^{(d)}\left[1-\prod_{l=1}^{n}(1-\lambda_l)\right]^{-1}$, instead of its integer part. Also, when we stop the process, based on convergence, we take the integer part of the latest estimate of ν as the non-parametric MLE of ν. Alternatively, as in the previous section, one can use a direct search method by maximizing the likelihood (6) for all values of $\nu \geq M^{(d)}$.

In an attempt to investigate the asymptotic properties of the non-parametric MLEs of ν and $\underline{\lambda}$ obtained by the above method, denoted by $\hat{\nu}_{np}$ and $\hat{\underline{\lambda}}$, respectively, as $\nu \to \infty$, let us initially assume that the true renewal distribution is a discrete probability distribution with a finite sample space $\Omega = \{w_1, w_2, \ldots, w_N\}$, $N > n$, where $w_1 < w_2 < \ldots < w_{N_1} < \tau_k < w_{N_1+1} < \cdots < w_N$ $(N_1 < N)$ with the corresponding probabilities being $p_1, \ldots, p_N$. Transforming the probabilities to discrete hazards, as described earlier, we obtain a parametric model with $(\nu, \underline{\lambda})$ as the parameters. The general asymptotic results of Dewanji et al. (1995) is then applied to this model. Let $\left(\hat{\nu}_w, \hat{\underline{\lambda}}_w\right)$ denote the MLE of $(\nu, \underline{\lambda})$ under the assumed discrete model. Note that, under this model, the observed renewal times must be a subset of Ω and our non-parametric MLE $\left(\hat{\nu}_{np}, \hat{\underline{\lambda}}\right)$ coincides with $\left(\hat{\nu}_w, \hat{\underline{\lambda}}_w\right)$.

Then, using the argument of Sen and Singer ([13], p207), as in Dewanji et al. [5], we conclude that, as $\nu \to \infty$, $[\nu^{-1/2}(\hat{\nu}_w - \nu), \nu^{1/2}(\hat{\underline{\lambda}} - \underline{\lambda})] \sim N(0, \Sigma^*)$, where the covariance matrix Σ^* can be consistently estimated by

$$\hat{\Sigma}^* = \begin{pmatrix} F(\tau_k, \hat{\underline{\lambda}})/\bar{F}(\tau_k, \hat{\underline{\lambda}}) & -\frac{\partial}{\partial \underline{\lambda}^T} \log \bar{F}(\tau_k, \hat{\underline{\lambda}}) \\ -\frac{\partial}{\partial \underline{\lambda}} \log \bar{F}(\tau_k, \hat{\underline{\lambda}}) & -\hat{\nu}_{np}^{-1} \frac{\partial^2}{\partial \underline{\lambda} \partial \underline{\lambda}^T} \log L(\hat{\nu}_{np}, \hat{\underline{\lambda}}) \end{pmatrix}^{-1}.$$

It is to be noted that, since no observed renewal time can be larger than τ_k, the distribution of $\hat{\nu}_w$ is affected only by $F(t), 0 \le t \le \tau_k$. The tail of F, over (τ_k, ∞), has no effect on any statistical property of $\hat{\nu}_w$. In addition, one can approximate a cdf $F(t)$ by a distribution with finite sample space at least within the finite interval $[0, \tau_k]$. Hence, heuristically, for any arbitrary renewal distribution, the asymptotic distribution of $\nu^{-\frac{1}{2}}(\hat{\nu}_{np} - \nu)$, as $\nu \to \infty$, can be approximated by that of $\nu^{-1/2}(\hat{\nu}_w - \nu)$, for some suitable choice of N_1, N and the mass points $w_1, \cdots, w_N$, which is normal with mean zero and variance $\left[F(\tau_k, \hat{\underline{\lambda}})/\bar{F}(\tau_k, \hat{\underline{\lambda}}) - \hat{\nu}_w \sum_{l=1}^n \left(f_l \left(\frac{1}{\hat{\lambda}_l} - 1 \right)^2 + c_l(\hat{\nu}_w) \right)^{-1} \right]^{-1}$, from the above result. Thus, as in (4), the asymptotic variance of $\hat{\nu}_{np}$ can be consistently estimated by

$$\hat{\nu}_{np} \left[F(\tau_k, \hat{\underline{\lambda}})/\bar{F}(\tau_k, \hat{\underline{\lambda}}) - \hat{\nu}_{np} \sum_{l=1}^{n} \left(f_l \left(\frac{1}{\hat{\lambda}_l} - 1 \right)^2 + c_l(\hat{\nu}_{np}) \right)^{-1} \right]^{-1}. \qquad (10)$$

The simulation results, reported in Section 5, agree with this heuristic conclusion.

5 A Simulation Study

In this section, some simulation results are reported to assess the performance of the parametric and non-parametric MLEs of ν. We consider three values of ν, namely, $\nu = 100, 500$ and 1000. The time between consecutive debugging is *one* and the number of debugging time points k is equal to 10, resulting in $\tau_k = 10$. We consider three types of renewal distributions, namely, Exponential, Weibull and Gamma.

For the Exponential distribution, we take three choices of the rate parameter (λ) as $0.2303, 0.1609$ and 0.1204 resulting in $\bar{F} = \bar{F}(\tau_k, \lambda) = e^{-\lambda \tau_k} = 0.1$, 0.2 and 0.3, respectively, reflecting different extent of non-detection. As a result, for the Exponential distribution, we have $3 \times 3 = 9$ different parameter configurations. For each of the 9 configurations, we generate 10000 data sets. For each such simulated data set, we compute the parametric and non-parametric MLEs, $\hat{\nu}_{pa}$ and $\hat{\nu}_{np}$, respectively, of ν along with the corresponding standard errors, obtained from (4) and (11) and denoted by $\hat{s}(\hat{\nu}_{pa})$ and $\hat{s}(\hat{\nu}_{np})$, respectively. Since ν is the parameter of primary interest, for the sake of convenience in presenting the results, let us denote its MLE by $\hat{\nu}$ and the corresponding standard error as $\hat{s}(\hat{\nu})$, while the kind of MLE (parametric or non-parametric) will be clear from the context. We take the average and median of $\hat{\nu}$'s and the average of the corresponding standard errors $\hat{s}(\hat{\nu})$ over the 10000 simulations and denote them by $\overline{\hat{\nu}}, \tilde{\hat{\nu}}$ and $\overline{s(\hat{\nu})}$, respectively. We also obtain the sample standard error of $\hat{\nu}$ from the 10000 values of $\hat{\nu}$ and denote it by $sse(\hat{\nu})$. The estimated coverage probability, denoted by CP, is computed as the proportion of times (out of 10000 simulations) the asymptotic 95% confidence interval, obtained through the normal approximation of $\hat{\nu}$, contains the true ν. For the purpose of comparison, we also provide relative bias and relative standard error defined as $(\bar{\hat{\nu}} - \nu)/\nu$ and $\overline{s(\hat{\nu})}/\nu$, respectively. The results are reported in Table 1.

Table 1: Empirical evaluation of the estimators for Exponential distribution.

Model for Estimation	ν	$\bar{F}$	$\bar{\hat{\nu}}$	$\frac{(\bar{\hat{\nu}}-\nu)}{\nu}$	$\tilde{\hat{\nu}}$	$\overline{s(\hat{\nu})}$	$\frac{\overline{s(\hat{\nu})}}{\nu}$	$s(\tilde{\hat{\nu}})$	$sse(\hat{\nu})$	CP
Exponential	100	0.1	99.79	-0.0021	99.68	5.69	0.0569	5.31	5.36	0.8929
		0.2	99.56	-0.0044	98.51	11.25	0.1125	10.47	10.78	0.8891
		0.3	108.54	0.0854	107.57	64.89	0.6489	62.86	55.82	0.8630
	500	0.1	499.72	-0.0006	498.22	11.73	0.0235	11.61	11.26	0.9458
		0.2	499.48	-0.0010	497.10	22.79	0.0456	22.49	22.18	0.9349
		0.3	503.22	0.0064	502.33	60.16	0.1203	58.97	59.13	0.9295
	1000	0.1	999.71	-0.0003	998.87	16.09	0.0161	16.07	16.35	0.9589
		0.2	999.46	-0.0005	997.85	30.89	0.0309	30.92	31.13	0.9518
		0.3	1002.32	0.0023	1001.19	102.33	0.1023	101.14	101.09	0.9416
Non-parametric	100	0.1	119.97	0.1997	104.75	38.28	0.3828	37.62	42.14	0.8627
		0.2	122.43	0.2243	103.70	54.77	0.5477	46.47	59.89	0.8029
		0.3	134.23	0.3423	102.01	86.68	0.8668	57.21	94.26	0.7524
	500	0.1	516.50	0.0330	503.10	70.51	0.1410	81.01	74.19	0.9306
		0.2	521.04	0.0421	497.89	99.89	0.1998	102.25	99.03	0.8512
		0.3	530.99	0.0620	506.53	133.33	0.2667	127.90	141.65	0.7664
	1000	0.1	1018.96	0.0190	1002.80	105.28	0.1053	114.71	102.92	0.9486
		0.2	1020.69	0.0207	1005.12	143.44	0.1434	144.91	147.47	0.9330
		0.3	1026.71	0.0267	996.26	183.44	0.1834	177.88	197.81	0.8831

It is to be noted that the ML estimates are nearly unbiased in all the cases and, also, the average standard error under $\overline{s(\hat{\nu})}$ and the sample standard error under $sse(\hat{\nu})$ are very close specially for large ν and small $\bar{F}(\tau_k, \lambda)$. As expected, the estimator $\hat{\nu}$ seems to perform better with respect to relative bias, relative standard error and CP with increasing ν and decreasing $\bar{F}(t_k, \lambda)$. Also, the CP values are close to 0.95, specially for large ν, suggesting convergence to normality. However, in all aspects, the performance of the parametric estimator is better than the performance of the non-parametric estimator, as expected.

Similar simulation setup is also used for Weibull and Gamma distribution. In addition, we also consider both the increasing failure rate (IFR) and decreasing failure rate (DFR) type parameter set-up for these two distributions. For IFR Weibull distribution, the shape and scale ($\frac{1}{\lambda}$) parameters are chosen as $(1.1, 4.685), (1.1, 6.488)$ and $(1.1, 8.477)$

and, for DFR Weibull distribution, those are selected as $(0.9, 3.959), (0.9, 5.893)$ and $(0.9, 8.136)$ such that the resulting $\bar{F}$ is $0.1, 0.2$ and 0.3, respectively. Results of simulation for both the IFR and DFR Weibull distributions are reported in Table 2A and Table 2B, respectively.

From Table 2A and Table 2B, it can be seen that, for both the DFR and IFR Weibull distributions, the relative bias and relative standard error decrease and CP becomes closer to 0.95 as ν increases and $\bar{F}(t_k, \lambda)$ decreases. Besides, the parametric estimator performs better than the non-parametric estimator, as expected. However, it can be seen that the non-parametric estimator performs better than the estimators under Exponential and Gamma distributions when data follows Weibull distribution. Therefore, under mis-specified distribution, non-parametric estimator is superior than all the wrongly specified estimators, as expected.

Table 2A: Empirical performance of the estimators for DFR Weibull distribution.

Model for Estimations	ν	$\bar{F}$	$\bar{\hat{\nu}}$	$\frac{(\bar{\hat{\nu}}-\nu)}{\nu}$	$\tilde{\hat{\nu}}$	$\overline{s(\hat{\nu})}$	$\frac{\overline{s(\hat{\nu})}}{\nu}$	$s(\tilde{\hat{\nu}})$	$sse(\hat{\nu})$	CP
Exponential	100	0.1	93.22	-0.0678	92.58	6.62	0.0662	5.14	8.56	0.7172
		0.2	90.49	-0.0951	89.88	15.18	0.1518	9.83	20.71	0.6417
		0.3	88.73	-0.1127	87.03	23.85	0.2385	13.76	32.79	0.6182
	500	0.1	475.23	-0.0495	472.42	15.69	0.0314	14.92	16.03	0.7551
		0.2	463.49	-0.0730	453.82	33.33	0.0667	30.84	35.67	0.7378
		0.3	452.40	-0.0952	445.19	58.40	0.1168	52.77	61.13	0.7136
	1000	0.1	954.26	-0.0457	951.84	22.72	0.0227	22.10	22.56	0.7889
		0.2	939.76	-0.0602	922.22	48.44	0.0484	46.29	49.37	0.7727
		0.3	909.54	-0.0905	892.48	81.31	0.0813	77.99	82.14	0.7579
Weibull	100	0.1	98.60	-0.0140	96.94	5.16	0.0516	3.88	7.37	0.8242
		0.2	98.33	-0.0167	93.85	8.82	0.0882	5.59	9.47	0.7914
		0.3	96.01	-0.0399	87.85	16.62	0.1662	12.02	15.58	0.7766
	500	0.1	497.13	-0.0057	494.99	8.88	0.0178	7.32	9.76	0.9427
		0.2	495.85	-0.0083	489.96	15.38	0.0308	12.75	16.28	0.9212
		0.3	494.80	-0.0104	486.93	34.38	0.0688	24.66	34.75	0.8769
	1000	0.1	997.26	-0.0027	994.93	12.65	0.0127	10.98	11.89	0.9522
		0.2	995.52	-0.0045	990.99	21.78	0.0218	18.69	21.12	0.9472
		0.3	991.57	-0.0084	983.96	47.39	0.0474	40.96	47.22	0.9065
Gamma	100	0.1	113.47	0.1347	101.38	29.58	0.2958	31.74	32.75	0.7613
		0.2	120.35	0.2035	103.77	46.81	0.4681	41.89	55.37	0.6931
		0.3	128.70	0.2870	104.00	70.39	0.7039	49.79	79.96	0.6577
	500	0.1	511.08	0.0222	496.35	59.31	0.1186	68.15	62.36	0.8215
		0.2	516.64	0.0333	503.04	87.90	0.1758	89.02	87.56	0.8111
		0.3	520.00	0.0400	504.22	112.72	0.2254	108.69	116.10	0.7939
	1000	0.1	984.42	-0.0156	972.29	89.42	0.0894	96.73	89.89	0.8461
		0.2	974.02	-0.0260	967.48	124.25	0.1243	124.91	125.57	0.8237
		0.3	965.11	-0.0349	959.56	155.84	0.1558	151.50	154.90	0.8101
Non-parametric	100	0.1	97.68	-0.0232	92.08	5.30	0.0530	4.09	18.05	0.7850
		0.2	96.19	-0.0381	90.79	12.03	0.1203	10.59	16.49	0.7259
		0.3	95.74	-0.0426	85.99	22.84	0.2284	19.95	26.46	0.6682
	500	0.1	491.39	-0.0172	488.38	12.40	0.0248	10.68	15.06	0.8973
		0.2	485.42	-0.0292	480.38	26.49	0.0530	23.95	28.16	0.8825
		0.3	482.80	-0.0344	471.19	47.95	0.0959	44.56	45.02	0.8393
	1000	0.1	1011.58	0.0116	1000.62	17.89	0.0179	15.84	18.59	0.9355
		0.2	1013.79	0.0138	1003.81	37.96	0.0380	35.67	38.78	0.9159
		0.3	1015.43	0.0154	993.63	66.49	0.0665	65.12	66.72	0.8660

Table 2B: Empirical performance of the estimators for IFR Weibull distribution.

Model for Estimation	ν	$\bar{F}$	$\bar{\hat{\nu}}$	$\frac{(\bar{\hat{\nu}}-\nu)}{\nu}$	$\tilde{\hat{\nu}}$	$\overline{s(\hat{\nu})}$	$\frac{\overline{s(\hat{\nu})}}{\nu}$	$s(\tilde{\hat{\nu}})$	$sse(\hat{\nu})$	CP
		0.1	106.65	0.0665	101.86	7.49	0.0749	6.84	6.68	0.7765
	100	0.2	113.72	0.1372	108.55	26.39	0.2639	21.20	30.57	0.7071
		0.3	120.28	0.2028	120.74	47.93	0.4793	42.16	61.39	0.6829
		0.1	532.10	0.0642	521.87	15.98	0.0320	14.32	15.38	0.8213
Exponential	500	0.2	560.23	0.1205	553.22	36.74	0.0735	35.76	38.21	0.8093
		0.3	600.16	0.2003	631.39	100.54	0.2011	99.91	116.03	0.7765
		0.1	1040.08	0.0401	1051.53	23.78	0.0238	24.38	25.89	0.8599
	1000	0.2	1101.55	0.1016	1109.39	50.61	0.0506	50.60	50.02	0.8437
		0.3	1150.91	0.1509	1252.28	123.84	0.1238	126.59	125.97	0.8148
		0.1	99.66	-0.0034	97.11	6.09	0.0609	5.08	6.92	0.9036
	100	0.2	99.34	-0.0066	94.65	14.02	0.1402	9.56	16.62	0.8409
		0.3	100.74	0.0074	91.47	22.00	0.2200	13.31	24.24	0.7467
		0.1	499.33	-0.0013	493.83	14.42	0.0288	13.87	14.67	0.9645
Weibull	500	0.2	501.31	0.0026	498.62	31.00	0.0620	29.02	30.58	0.9442
		0.3	498.15	-0.0037	490.59	55.86	0.1117	47.44	59.66	0.9427
		0.1	1000.00	0.0000	997.03	20.48	0.0205	20.01	20.20	0.9719
	1000	0.2	999.59	-0.0004	994.19	43.91	0.0439	42.25	45.39	0.9694
		0.3	998.53	-0.0015	990.66	77.03	0.0770	71.95	80.08	0.9616
		0.1	123.14	0.2314	100.02	46.19	0.4619	42.50	59.62	0.8396
	100	0.2	130.26	0.3026	99.48	77.68	0.7768	52.95	73.79	0.7829
		0.3	135.40	0.3540	101.33	95.94	0.9594	64.67	93.17	0.7059
		0.1	521.65	0.0433	491.69	78.76	0.1575	92.16	85.27	0.9234
Gamma	500	0.2	530.92	0.0618	502.98	117.19	0.2344	119.92	124.58	0.8679
		0.3	538.47	0.0769	504.69	158.21	0.3164	145.65	161.35	0.7749
		0.1	1026.54	0.0265	1006.24	117.11	0.1171	133.45	117.26	0.9499
	1000	0.2	1026.78	0.0268	997.61	165.33	0.1653	166.34	171.39	0.9066
		0.3	1027.42	0.0274	991.71	212.36	0.2124	204.44	211.58	0.8817
		0.1	98.83	-0.0117	97.92	6.54	0.0654	6.34	12.30	0.8889
	100	0.2	97.79	-0.0221	94.81	16.29	0.1629	14.97	19.11	0.8107
		0.3	94.67	-0.0533	88.93	27.70	0.2770	23.47	25.22	0.7202
		0.1	504.33	0.0087	497.88	15.83	0.0317	16.50	19.44	0.9369
Non-parametric	500	0.2	507.98	0.0160	499.92	34.11	0.0682	34.48	36.60	0.9210
		0.3	511.20	0.0224	490.98	60.42	0.1208	61.04	61.33	0.8916
		0.1	1008.21	0.0082	997.86	22.83	0.0228	21.11	23.98	0.9646
	1000	0.2	1015.79	0.0158	994.00	48.45	0.0485	45.30	50.97	0.9251
		0.3	1020.20	0.0202	990.93	85.12	0.0851	84.18	88.75	0.9105

We also consider DFR and IFR Gamma distributions for the simulation model. For DFR Gamma distribution, we choose $(0.9, 4.702), (0.9, 6.849)$ and $(0.9, 9.307)$ as the three pairs of shape and scale parameters resulting in $\bar{F} = 0.1, 0.2$ and 0.3, respectively. Similarly, for IFR Gamma distribution, corresponding pairs of parameters are $(1.1, 4.042), (1.1, 5.695)$ and $(1.1, 7.507)$. The simulation results corresponding to IFR Gamma and DFR Gamma can be seen in Table 3A and Table 3B, respectively. The performance of these estimators is qualitatively similar to that of the previous estimators.

Table 3A: Empirical performance of the estimators for DFR Gamma distribution.

Model for Estimation	ν	$\bar{F}$	$\bar{\hat{\nu}}$	$\frac{(\bar{\hat{\nu}}-\nu)}{\nu}$	$\tilde{\hat{\nu}}$	$\overline{s(\hat{\nu})}$	$\frac{\overline{s(\hat{\nu})}}{\nu}$	$s(\tilde{\hat{\nu}})$	$sse(\hat{\nu})$	CP
Exponential	100	0.1	88.67	-0.1133	86.66	7.52	0.0752	6.58	19.02	0.7233
		0.2	81.12	-0.1888	79.48	17.01	0.1701	14.79	38.29	0.6593
		0.3	70.81	-0.2919	65.55	26.41	0.2641	24.43	56.12	0.6108
	500	0.1	448.21	-0.1036	441.92	17.85	0.0357	15.87	18.05	0.7981
		0.2	407.47	-0.1851	400.53	38.19	0.0764	35.12	40.36	0.7532
		0.3	386.12	-0.2278	375.31	65.49	0.1310	62.02	67.49	0.7055
	1000	0.1	897.82	-0.1022	893.17	25.70	0.0257	22.51	26.42	0.8296
		0.2	822.84	-0.1772	817.41	54.46	0.0545	50.06	56.39	0.7939
		0.3	776.11	-0.2239	770.40	91.16	0.0912	88.72	87.25	0.7678
Weibull	100	0.1	115.71	0.1571	100.21	35.79	0.3579	33.78	43.08	0.8112
		0.2	118.70	0.1870	98.84	52.01	0.5201	42.36	64.00	0.7754
		0.3	133.94	0.3394	103.16	83.43	0.8343	54.33	96.27	0.7056
	500	0.1	511.46	0.0229	498.95	69.32	0.1386	75.55	74.35	0.9092
		0.2	516.74	0.0335	502.59	98.27	0.1965	96.84	102.68	0.8922
		0.3	527.47	0.0549	504.05	127.06	0.2541	118.70	147.48	0.8804
	1000	0.1	1011.92	0.0119	1007.19	103.97	0.1040	108.68	107.79	0.9364
		0.2	1017.98	0.0180	1010.86	137.67	0.1377	136.24	137.11	0.9200
		0.3	1022.35	0.0224	1013.49	171.95	0.1720	167.05	180.39	0.9044
Gamma	100	0.1	100.40	0.0040	97.98	5.42	0.0542	6.12	20.46	0.8321
		0.2	100.70	0.0070	98.21	6.19	0.0619	7.28	16.33	0.7950
		0.3	98.18	-0.0182	96.46	10.29	0.1029	12.68	12.51	0.7775
	500	0.1	498.52	-0.0030	497.87	12.80	0.0256	11.07	19.34	0.9329
		0.2	497.77	-0.0045	493.93	19.26	0.0385	17.90	22.94	0.9320
		0.3	497.63	-0.0047	485.97	30.60	0.0612	29.87	28.89	0.9217
	1000	0.1	997.66	-0.0023	996.94	18.21	0.0182	16.25	18.29	0.9662
		0.2	997.53	-0.0025	993.95	32.19	0.0322	30.27	31.98	0.9473
		0.3	997.43	-0.0026	988.87	46.81	0.0468	43.64	44.14	0.9381
Non-parametric	100	0.1	99.52	-0.0048	97.16	6.15	0.0615	5.15	10.09	0.8277
		0.2	99.06	-0.0094	94.92	14.28	0.1428	9.35	20.80	0.7813
		0.3	97.06	-0.0294	88.98	23.44	0.2344	14.08	30.39	0.7181
	500	0.1	502.29	0.0046	501.14	14.31	0.0286	13.86	14.75	0.9245
		0.2	503.44	0.0069	498.79	30.61	0.0612	28.53	32.34	0.9228
		0.3	511.10	0.0222	506.88	54.49	0.1090	47.77	58.24	0.8869
	1000	0.1	1004.49	0.0045	999.34	20.34	0.0203	19.98	20.97	0.9524
		0.2	1006.07	0.0061	1001.29	43.43	0.0434	42.10	43.53	0.9435
		0.3	1020.87	0.0209	1000.87	75.58	0.0756	72.08	76.31	0.9139

Table 3B: Empirical performance of the estimators for IFR Gamma distribution.

Model for Estimation	ν	$\bar{F}$	$\bar{\hat{\nu}}$	$\frac{(\bar{\hat{\nu}}-\nu)}{\nu}$	$\tilde{\hat{\nu}}$	$\overline{s(\hat{\nu})}$	$\frac{\overline{s(\hat{\nu})}}{\nu}$	$s(\tilde{\hat{\nu}})$	$sse(\hat{\nu})$	CP
Exponential	100	0.1	91.36	-0.0864	90.02	6.50	0.0650	5.20	7.27	0.7019
		0.2	82.21	-0.1779	78.46	17.09	0.1709	9.70	25.13	0.6243
		0.3	69.88	-0.3012	64.59	27.56	0.2756	12.77	43.33	0.5619
	500	0.1	464.11	-0.0718	461.26	15.89	0.0318	15.42	15.52	0.7201
		0.2	420.31	-0.1594	414.31	35.36	0.0707	31.89	37.64	0.6912
		0.3	394.70	-0.2106	379.08	63.80	0.1276	54.17	67.53	0.6565
	1000	0.1	939.20	-0.0608	938.08	23.27	0.0233	22.86	22.21	0.7662
		0.2	889.37	-0.1106	881.72	50.85	0.0509	48.39	52.70	0.7010
		0.3	791.39	-0.2086	780.27	89.95	0.0900	83.51	89.03	0.6787
Weibull	100	0.1	122.03	0.2203	105.14	39.97	0.3997	39.30	48.56	0.7407
		0.2	130.06	0.3006	102.50	63.93	0.6393	47.18	82.95	0.6877
		0.3	139.59	0.3959	100.11	92.45	0.9245	61.33	100.70	0.6024
	500	0.1	525.09	0.0502	497.93	67.49	0.1350	84.51	75.47	0.8454
		0.2	526.83	0.0537	504.85	100.73	0.2015	109.99	100.48	0.8209
		0.3	537.56	0.0751	507.95	139.80	0.2796	135.70	149.09	0.8000
	1000	0.1	1024.94	0.0249	1000.93	98.80	0.0988	120.51	105.16	0.8706
		0.2	1032.03	0.0320	1005.39	144.06	0.1441	153.16	148.77	0.8445
		0.3	1032.95	0.0330	1007.31	191.91	0.1919	186.95	200.27	0.8332
Gamma	100	0.1	100.28	0.0028	99.93	3.14	0.0314	7.66	15.67	0.7985
		0.2	98.43	-0.0157	92.96	8.29	0.0829	12.85	28.41	0.7475
		0.3	96.29	-0.0371	88.97	10.06	0.1006	20.92	37.72	0.7001
	500	0.1	499.51	-0.0010	496.85	11.92	0.0238	10.79	13.58	0.8980
		0.2	495.61	-0.0088	490.96	18.17	0.0363	16.88	20.88	0.8765
		0.3	494.47	-0.0111	489.99	29.10	0.0582	25.72	35.84	0.8602
	1000	0.1	1000.55	0.0005	998.82	19.61	0.0196	18.93	19.95	0.9426
		0.2	993.85	-0.0061	987.93	32.71	0.0327	30.08	33.26	0.9379
		0.3	992.11	-0.0079	979.98	47.22	0.0472	45.85	49.14	0.8978
Non-parametric	100	0.1	97.40	-0.0260	98.12	6.12	0.0612	8.85	11.38	0.7728
		0.2	95.75	-0.0425	97.08	14.19	0.1419	17.40	24.26	0.7100
		0.3	90.70	-0.0930	91.79	22.07	0.2207	26.52	50.06	0.6388
	500	0.1	493.80	-0.0124	491.59	15.06	0.0301	14.83	18.48	0.8647
		0.2	487.50	-0.0250	484.73	32.58	0.0652	31.74	39.38	0.8413
		0.3	484.85	-0.0303	479.96	59.52	0.1190	57.62	69.04	0.8155
	1000	0.1	989.63	-0.0104	983.66	21.63	0.0216	21.12	22.08	0.9281
		0.2	979.72	-0.0203	970.93	46.82	0.0468	45.39	46.11	0.9097
		0.3	973.79	-0.0262	962.39	83.84	0.0838	81.76	84.90	0.8842

To study the effect of the number and size of debugging intervals, we have carried out another simulation study with fixed $\tau_k = 10$, while the number k of debugging is varied as $1, 2, 5$ and 10 with the corresponding time between successive debugging being $10, 5, 2$ and 1, respectively. The probability of non-detection $\bar{F}(t_k)$ is kept fixed at 0.2 and ν is fixed as 1000. The same simulation exercise as before is carried out in 10000 repetitions. The results for Weibull and Gamma distributions are presented in Table 4 and Table 5, respectively.

Table 4: Simulation results with varying number of debugging for Weibull distribution.

Model for Estimation	Failure Rate	k	$\bar{\hat{\nu}}$	$\frac{(\bar{\hat{\nu}}-\nu)}{\nu}$	$\overline{s(\hat{\nu})}$	$\frac{\overline{s(\hat{\nu})}}{\nu}$	CP
Weibull	DFR	1	999.74	-0.0003	16.85	0.0169	0.9539
		2	998.37	-0.0016	18.58	0.0186	0.9526
		5	996.49	-0.0035	20.51	0.0205	0.9516
		10	995.52	-0.0045	21.78	0.0218	0.9472
	IFR	1	1000.05	5E-05	23.45	0.0235	0.9747
		2	1000.24	0.0002	33.19	0.0332	0.9716
		5	1000.37	0.0004	36.66	0.0367	0.9697
		10	999.59	-0.0004	43.91	0.0439	0.9694
Non-parametric	DFR	1	1000.82	0.0008	26.03	0.0260	0.9383
		2	1002.92	0.0029	28.08	0.0280	0.9351
		5	1007.84	0.0078	32.63	0.0326	0.9249
		10	1013.79	0.0138	37.96	0.0380	0.9159
	IFR	1	1000.96	0.0010	29.25	0.0293	0.9546
		2	1005.90	0.0059	37.33	0.0373	0.9344
		5	1013.89	0.0139	42.10	0.0421	0.9293
		10	1015.79	0.0158	48.45	0.0485	0.9251

Table 5: Simulation results with varying number of debugging for Gamma distribution.

Model for Estimation	Failure Rate	k	$\bar{\hat{\nu}}$	$\frac{(\bar{\hat{\nu}}-\nu)}{\nu}$	$\overline{s(\hat{\nu})}$	$\frac{\overline{s(\hat{\nu})}}{\nu}$	CP
Gamma	DFR	1	1000.48	0.0005	21.69	0.0216	0.9612
		2	1000.98	0.0010	25.75	0.0258	0.9548
		5	998.03	-0.0020	27.67	0.0277	0.9493
		10	997.53	-0.0025	32.19	0.0322	0.9473
	IFR	1	1000.14	0.0001	22.26	0.0223	0.9615
		2	999.24	-0.0008	27.12	0.0271	0.9501
		5	999.01	-0.0010	29.55	0.0296	0.9389
		10	993.85	-0.0062	32.71	0.0327	0.9379
Non-parametric	DFR	1	1001.62	0.0016	26.72	0.0267	0.9672
		2	1002.17	0.0022	32.90	0.0329	0.9623
		5	1004.47	0.0045	39.28	0.0393	0.9512
		10	1006.07	0.0061	43.43	0.0434	0.9435
	IFR	1	999.50	-0.0005	28.60	0.0286	0.9517
		2	1003.17	0.0032	36.14	0.0361	0.9424
		5	1013.19	0.0132	41.57	0.0416	0.9239
		10	979.72	-0.0203	46.82	0.0468	0.9097

The average standard error and the sample standard error turn out to be very close, as before, and so only the $\overline{s(\hat{\nu})}$s are reported. The estimator $\hat{\nu}$ seems to perform better with respect to relative bias, relative standard error and CP with decreasing number k of debugging intervals. Therefore, a single debugging schedule at the end of testing at time τ_k seems to be the most efficient design (See [2] for similar result). However, as remarked earlier, a schedule of more than one debugging intervals may be necessary due to the market demand for software release.

6 Concluding Remarks

In this work, we suggest both parametric and non-parametric method to estimate the initial number of bugs present in a software, assuming that the successive times of appearances of the bugs follow independent and identically distributed renewal processes, where the common renewal distribution may belong to a specific parametric family or is arbitrary. Overall, it is seen through simulation studies that the estimate of ν performs the best if the assumed parametric model is correct; however, the performance of non-parametric estimate is better compared to that under mis-specified parametric models. It is also seen that, when ν is small and the $\bar{F}(\tau_k)$ is high, estimate of ν sometimes diverges to infinity. In our previous study with Exponential renewal distribution (see [2]), we have shown that the condition $M^{(d)} \neq M$ is one of the sufficient conditions for finite estimate of ν; hence, a necessary condition for estimate of ν being infinity is $M^{(d)} = M$. In our simulation study with arbitrary renewal distribution, we have seen that, whenever the estimate $\hat{\nu}$ diverges to infinity, the value of $M^{(d)}$ equals M, in line with the result obtained for Exponential renewal distribution.

It is to be noted that, other than applying Algorithm 1 and Algorithm 2 to estimate ν, one can also resort to some direct search method as indicated in Sections 3 and 4. It is seen that the estimates of ν by both the methods turn out to be very close.

In studying the asymptotic properties of the proposed estimates of ν, we have considered the limiting situation with $\nu \to \infty$. These results are useful in software reliability context, as most programs of practical significance contain thousands of lines of code and contain numerous bugs. It is to be noted that general asymptotic results, as $\tau_k \to \infty$, are not meaningful in this context, since there is usually a fixed and finite time period for software testing and also because the situation $\tau_k \to \infty$ will detect all the bugs in principle rendering the estimation to be useless.

From the efficiency point of view, we have seen that the choice of $k = 1$ minimizes the variance of $\hat{\nu}$. However, in reality, the design issue will force more than one debugging period (i.e. $k > 1$) due to other practical considerations and cost aspects. In addition, for a fixed k, an interesting design consideration may deal with choosing the values of the τ_i's optimally depending upon the cost incurred in testing as well as the cost of bug arrivals after software release.

References

[1] S. R. Dalal. *Software reliability models: A selective survey and new directions.* Handbook of Reliability Engineering, 2003.

[2] Sudipta Das, Anup Dewanji, and Ashis Kumar Chakraborty. Software reliability modeling with periodic debugging. *IEEE Transactions On Reliability*, 65(3):1449–14556, September 2016.

[3] Sudipta Das, Anup Dewanji, and Debasis Sengupta. Discrete time software reliability modeling with periodic debugging schedule. *Statistical Methodology*, 33:147–159, 2016.

[4] Anup Dewanji, Subrata Kundu, and Tapan K. Nayak. Nonparametric estimation of the number of components of a superposition of renewal processes. *Journal of Statistical Planning and Inference*, 142:2710–2718, 2012.

[5] Anup Dewanji, Tapan K. Nayak, and Pranab K. Sen. Estimating the number of components of a system of superimposed renewal processes. *Sankhyā: The Indian Journal of Statistics Series A*, pages 486–499, 1995.

[6] Anup Dewanji, Debasis Sengupta, and Ashis Kumar Chakraborty. A discrete time model for software reliability with application to a flight control software. *Applied Stochastic Models in Business and Industry*, 27:723–731, 2011.

[7] Swapna S. Gokhale, Peter N. Marinos, and Kishor S. Trivedi. Important milestones in software reliability modeling. *SEKE*, pages 345–352, 1996.

[8] Chin-Yu Huang and Chu-Ti Lin. Software reliability analysis by considering fault dependency and debugging time lag. *IEEE Transactions on Reliability*, 55(3):436–450, September 2006.

[9] Seheon Hwang and Hoang Pham. Quasi-renewal time-delay fault-removal consideration in software reliability modeling. *IEEE Transactions on Systems Man and Cybernetics*, 39(1):200–209, January 2009.

[10] Z. Jelinski and P. Moranda. Software reliability research. *Statistical Computer Performance Evaluation, Academic Press*, pages 465–484, 1972.

[11] Michael R. Lyu. *Handbook of software reliability engineering.*, volume 3. CA: IEEE computer society press, 1996.

[12] Tapan K. Nayak. Estimating population size by recapture sampling. *Biometrika*, 75:113–120, 1988.

[13] P. K. Sen and J. M. Singer. *Large sample methods in statistics an introduction with application.* Chapman and Hall, London, 1993.

[14] Nozer D. Singpurwalla and Simon P. Wilson. Software reliability modeling. *International Statistical Review*, 62(3):289–317, December 1994.

[15] Haiyan Sun, Jia Heand Bingfeng Xie, and Ji Wu. Analysis of software release problems based on fault detection and correction processes. *International Conference on Computer, Communications and Information Technology*, pages 174–177, 2014.