Analysis of Sequential Quality Improvement Plans to Obtain Confidence Bounds

Technical Report No.ASD/2013/2
Dated: 29 April, 2013

Vignesh T Subrahmaniam
GE Global Research

And

Anup Dewanji and Bimal K Roy
Indian Statistical Institute

Indian Statistical Institute
Applied Statistics Unit
Kolkata 700 108

Submitted to the Annals of Applied Statistics

ANALYSIS OF SEQUENTIAL QUALITY IMPROVEMENT PLANS TO OBTAIN CONFIDENCE BOUNDS

By Vignesh T Subrahmaniam*, Anup Dewanji† and Bimal K Roy†

GE Global Research and Indian Statistical Institute* †

Sequential quality improvement plans create a series of product versions with improving reliability by addressing the failure modes of the previous versions. They represent a widely used reliability improvement strategy. Reliability metrics for the current product version needs to be estimated from the failure times of the previous versions. Confidence bounds on the reliability metrics are of more importance to decision makers than point estimates as they allow for an evaluation of a margin of reliability in the product. The availability of exactly one failure time for each product version poses a challenge for obtaining confidence bounds. In this article, we consider a model in which the time to failure distributions belong to a family of distributions indexed by a real valued parameter, whose ordering determines the stochastic ordering of the distributions. We do not make any assumptions about the nature of improvement in reliability, allow for no improvement in reliability for some consecutive versions and relax the requirement of independence of the failure times. We propose a novel statistic, based upon the maximum of the observed failure times, which is proved to be a confidence bound for the parameter of interest with a minimum coverage probability. The methods developed are applied to datasets relating to software debugging and a new dataset derived from an automated error logger. The analysis reveals some surprising insights.

1. Introduction.

A common method for improving the reliability of a product is to build a sequence of product versions, test or use each version till failure and create the next version by addressing the cause of failure of the current version. The very nature of creating product versions, which address the failure modes of the previous versions, imposes a constraint that the reliability of the product must increase, or remain the same. Such sequential quality improvement plans are not only used for industrial product development, but also for improving the reliability of critical procedures such as air traffic control (ATC) protocols, where after a major incident an investigation is launched to determine if existing protocols can be improved to prevent similar incidents in the future, leading to a revised and safer protocol. In an

AMS 2000 subject classifications: Primary 62P30; secondary 62N05

Keywords and phrases: Software Reliability, Isotonic Regression, Error Loggers, Coverage Probability, Bug-Database

Internet banking context, after a major fraudulent credit card transaction, the bank usually imposes stricter measures for fraud prevention intended to make future credit card transactions safer. Sequential quality improvement procedures are most famously used in software testing, where significant defects are fixed as and when they are discovered. It is important to estimate the time to failure distribution of the most recent version of the product, with such estimates being used to compute reliability metrics. For example, one might compute metrics such as the expected time to failure, or the probability that the next use case of the product/procedure will result in a failure. The data for such analysis is the time to failure for each of the previous versions of the product. In the ATC context, one may consider the number of successful take-off's achieved through the use of the protocol before a major incident. In the Internet banking example, the available data could be the number of legitimate credit card transactions that were processed before a fraudulent transaction was erroneously processed. Sometimes it is more important to obtain a one-sided bound for the reliability than a point estimate and the methodology must be applicable when the number of failures is small. The method must also consider the possibility that for some consecutive versions there might be no improvement in reliability at all.

In this article, we propose a novel and simple method for constructing an *exact* one-sided confidence bound, with a minimum coverage probability, for the single index parameter of the failure time distribution. The method is based on the distribution of $max(X_1, \ldots, X_n)$, where X_k's denote the observed time to failures of, say, n versions of the product. The proposed bound is valid even for small data sizes, can be modified to consider non-independence of the X_k's and used when there is no improvement in reliability for some consecutive versions of the product. Additionally, the bound can be used even when the time to failure X_n of the current version of the product is right censored. Section 2 formulates the problem and presents the main result in a general framework. Section 3 studies the properties of the proposed upper-bound through simulation. Section 4 presents the application of the proposed upper-bound to two publicly available software reliability datasets and a new dataset retrieved from the automatic error-logger of a workstation. Although we expect the reliability of the product not to decrease with every subsequent version, this might not happen in some situations. To address this issue, we demonstrate how the method can be modified to consider a bounded decrease in reliability of the subsequent product versions (See Section 4.2). Section 5 ends with some concluding remarks.

Sequential quality improvement procedures are typically used on products whose failure can be catastrophic. It is more important to provide confidence intervals for the reliability estimates than point estimates as it allows for an evaluation of the margin of safety in the product. In particular, one-sided bounds which have a minimum coverage for the parameter of interest of the failure time distribution can be invaluable for decision makers. The failure time distribution of each version can be modeled parametrically while the nature of the improvement in reliability is in general unknown. While computing confidence bounds, it is important to note that asymptotic procedures may not be appropriate since the size of the dataset comprising the failure times for all versions can be small. Also, there is exactly one observation on the time to failure for each version of the product. Existing procedures for computing confidence bounds usually incorporate some modeling of the parameters of failure time distributions of successive versions, as in Jelinski and Moranda (1972), representing the nature of improvement in reliability. Independence of the failure times of different versions can be assumed in many situations, but may be questioned in some others.

Software quality improvement, through instant debugging of any defects discovered under controlled testing, represents the most studied example of the problem considered. Each time a defect is corrected, a new version of the software is released and tested till another defect is detected. The time to failure between each debugging provides the data for reliability modeling. Jelinski and Moranda (1972) provide the first model for analyzing such data. They assume $X_k \sim exponential(\theta_k)$, with $\theta_k = \lambda(N - k)$, for $k = 1, \ldots, n$, where X_k is the time to failure of the kth version of the software (or, alternatively after fixing k defects in the software). The parameter N is interpreted as the unknown number of defects in the software. Since then a number of parametric models with increasing complexity have been considered for modeling the θ_k's (see Moranda 1975, Musa and Okumoto 1976 and Singpurwalla and Wilson 1999, for some examples). In some examples, such as the dataset provided by Jelinski and Moranda, the assumption that θ_k must decrease with k has been questioned and has led to the development of Bayesian models where such strict decrease is not necessary (Littlewood and Verall 1973, Kuo and Yang 1996 and Basu and Ebrahimi 2003). However, their selection of the priors has been motivated towards improving computational ease of estimating the model.

Isotonic regression provides a method of estimating the reliability distribution of the final version of the software, under the constraint of non-decreasing reliability in the successive versions. The pool adjacent violator (PAV) algorithm (Ayers 1955) can be used to obtain maximum likelihood estimates of parameters of the underlying failure time distributions under isotonic constraints; however, computing confidence intervals in small samples is a challenge. This is because of (i) the non-linear nature of the PAV estimator and (ii) the possibility of equality in some or all of the failure time distributions. Recently, there has been increased interest in characterizing the asymptotic distribution of PAV estimator in the context of dosage studies (Jewel and Kalbfleisch 2003, Tebbs and Swallow 2004, Bhattacharya and Kong 2007). The inconsistency of bootstrap procedures when the parameter space has non-strict inequality constraints has been noted by Andrew (2000). Li, Taylor and Nan (2010) propose a modified bootstrap estimator for estimating two binomial proportions in the presence of order restrictions in small samples. Similar bootstrap method seems difficult for the problem being considered here since there is exactly one observed failure time for every version. Every additional version introduces an additional unknown parameter, whereas in the method of Li, Taylor and Nan (2010) the number of parameters is fixed at two. The method of constructing a conservative one-sided confidence bound proposed in this article is an alternative solution.

2. The Main Result. Let $X_k \sim F_k(.)$, for $k = 1, \ldots, n$, where $F_k(.)$ is the cumulative distribution function for X_k. Since successive versions of the product have non-decreasing reliability, we will impose the constraint that $F_k(x) \leq F_{k-1}(x)$ for all x. An estimate of $F_n(x)$, for any $x \in \mathbb{R}$, along with an upper confidence bound is of interest. If $F_1, \ldots, F_n$ are assumed to be arbitrary CDF's with the above stochastic ordering, then estimation of $F_n(x)$, for all $x \in \mathbb{R}$, may not be feasible, specifically for $x > max(X_1, \ldots, X_n)$. However, if F_k's belong to a family of distributions indexed by a real-valued single parameter, i.e $F_k(x) = F(x; \theta_k)$, then under the assumption of mutual independence, maximum likelihood estimation of θ_k's, and in particular θ_n, may be possible for certain families of distribution. In order to ensure a margin of safety, it is more important to obtain a confidence bound with a specified minimum coverage probability for θ_n than obtaining a point estimate for θ_n. As in the context of software reliability (Jelinski and Moranda 1972), we assume $\theta_1 \geq \ldots \geq \theta_n$, leading to $F(x; \theta_1) \geq \cdots \geq F(x; \theta_n)$ for all x. The assumption that θ_k may possibly be equal to θ_{k-1} implies that the number of unknown parameters is not specified. Such an observation indicates that asymptotic methods for computing confidence bounds for θ_n as $n \to \infty$ may not be appropriate. The possibility of $X_1, \ldots, X_n$ being a dependent sequence of observations, with the nature of the dependence also being unknown, only complicates the problem. It is in these contexts that we propose Theorems 2.1 and 2.2, which can be used to obtain confidence bounds for the parameter θ_n with a minimum coverage probability. Theorem 2.1 is formulated under the assumption of $X_1, \ldots, X_n$ being independent, while Theorem 2.2 is formulated without the assumption of independence.

THEOREM 2.1. *Define $\mathcal{F} = \{F(.,\theta), \theta \in \mathbb{R}\}$ to be a family of cumulative distribution functions with support in $(-\infty, \infty)$ indexed by the parameter $\theta \in \Theta \subseteq \mathbb{R}$, with the property that $F(x,\theta) > F(x,\theta')$, if $\theta > \theta'$, $\forall\, x \in (-\infty, \infty)$. Let $X_1 \sim F(.,\theta_1), X_2 \sim F(.,\theta_2), \ldots, X_n \sim F(.,\theta_n), \ldots$ with $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_n \geq \ldots$, be an independent sequence of observations. Define $F^{-1}(p,\theta) = inf\ \{x : F(x,\theta) \geq p\}$. Then the statistic*

$$\hat{\theta}_n^p = min\left\{\theta : max(X_1, \ldots, X_n) > F^{-1}\left(p^{1/n}, \theta\right), \theta \in \Theta\right\} \tag{2.1}$$

based on $X_1, \ldots, X_n$, has the property $P(\hat{\theta}_n^p < \theta_n) \leq 1 - p$. Note that, since $F^{-1}(.;\theta)$ is a non-increasing function of θ, this minimum $\hat{\theta}_n^p$ exists. This implies that $\hat{\theta}_n^p$ is an at least $100 \times p\%$ upper-bound for the parameter θ_n, that is, $P(\hat{\theta}_n^p \geq \theta_n) \geq p$. In other words, $\left\{\theta : \theta \leq \hat{\theta}_n^p\right\}$ is an atleast $100 \times p\%$ one-sided confidence interval.

PROOF. Note that $P\left[max\ (X_1, \ldots, X_n) \leq \lambda\right] = \prod_{j=1}^n F(\lambda, \theta_j) \geq [F(\lambda, \theta_n)]^n$, since $F(\lambda, \theta_n) \leq F(\lambda, \theta_j)$, for $j = 1, \ldots, n$. Define $A_k = \left\{max(X_1, \ldots, X_n) > F^{-1}\left(p^{1/n}, \theta_n - \frac{1}{k}\right)\right\}$, for $k = 1, 2, \ldots$. Then,

$$\begin{aligned} P(A_k) &\leq 1 - \left[F\left(F^{-1}\left(p^{1/n}, \theta_n - 1/k\right), \theta_n\right)\right]^n \\ &\leq 1 - \left[F\left(F^{-1}\left(p^{1/n}, \theta_n\right), \theta_n\right)\right]^n \\ &\leq 1 - p \end{aligned}$$

since $F^{-1}\left(p^{1/n}, \theta_n - \frac{1}{k}\right) \geq F^{-1}\left(p^{1/n}, \theta_n\right)$ and $F\left(F^{-1}\left(p^{1/n}, \theta_n\right), \theta_n\right) \geq p^{1/n}$.
Now, $P(\hat{\theta}_n^p < \theta_n) =$
$P\left(\exists\, \epsilon > 0\ s.t.\ max(X_1, \ldots, X_n) > F^{-1}\left(p^{1/n}, \theta_n - \delta\right), \forall\, 0 < \delta \leq \epsilon\right) =$
$P\left(\bigcup_{j=1}^\infty \bigcap_{k=j}^\infty A_k\right) = P\left(\underset{k}{limsup}\, A_k\right) \leq \underset{k}{limsup}\, P\left(A_k\right) \leq 1 - p.$

□

THEOREM 2.2. *Define $\mathcal{F} = \{F(.,\theta), \theta \in \mathbb{R}\}$ to be a family of cumulative distribution as in Theorem 2.1 . Let $X_1 \sim F(.,\theta_1), X_2 \sim F(.,\theta_2), \ldots, X_n \sim F(.,\theta_n), \ldots$ with $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_n \geq \ldots$. be a possibly dependent sequence of observations. Then the statistic*

$$(2.2) \quad \tilde{\theta}_n^p = min\left\{\theta :\ max(X_1,\ldots,X_n) > F^{-1}\left(1 - \frac{1-p}{n}, \theta\right), \theta \in \Theta\right\}$$

has the property $P(\tilde{\theta}_n^p < \theta_n) \leq 1-p$*, for* $k = 1,\ldots,n$*. This implies that* $\tilde{\theta}_n^p$ *is an at least* $100 \times p\%$ *upper-bound for the parameter* θ_n*, even if* $X_1,\ldots,X_n$ *are dependent.*

Proof. Note that $P\left(max\ \{X_1,\ldots,X_n\} > \lambda\right) = P\left(\bigcup_j \{X_j > \lambda\}\right) \leq \sum_{j=1}^n 1 - F(\lambda,\theta_j) \leq n[1 - F(\lambda,\theta_n)]$, since $F(\lambda,\theta_n) \leq F(\lambda,\theta_j)$, for $j = 1,\ldots,n$. The rest of the proof follows the arguments of Theorem 2.1. □

In situations when X_n is right censored at x_0, the proposed bounds $\hat{\theta}_n^p$ and $\tilde{\theta}_n^p$ may be computed by replacing X_n with x_0. The statistics will still be a valid upper-bound for the parameter with a minimum coverage probability, since $max(X_n,\ldots,X_n) \geq max(X_n,\ldots,X_{n-1},x_0)$. This is particularly useful as the time to failure for the latest version of the product may often be right censored. Note that, with X_n being right censored, the maximum likelihood estimate of θ_n does not exist. However, $\hat{\theta}_n^p$ can still be obtained as an at least $100 \times p$ % upper-bound for θ_n.

3. A Simulation study. In order to study the performance of the proposed conservative upper-bounds numerically, we consider a model in which $X_k \sim exponential(\theta_k)$, for $k = 1,\ldots,n = 10$. Theorems 2.1 and 2.2 can be used to provide a conservative $100 \times p\%$ confidence bounds for θ_{10}. We consider four possible decreasing patterns of the θ_k's, namely, constant, linear, convex and concave, as presented with values in Figure 1. The conservative upper-bound $\hat{\theta}_{10}^p$ for θ_{10}, under the assumption of independence, is obtained by simulating $X_1,\ldots,X_{10}$ independent of each other with θ_k's as per each of the four patterns presented in Figure 1. To consider the possibility of dependence between the X_k's, we simulate correlated X_i and X_j, for $i \neq j$, using a bivariate Gaussian Copula with a copula correlation of 0.5 (Nelson 1999), and then obtain $\tilde{\theta}_{10}^p$. The actual coverage probabilities for $\hat{\theta}_{10}^p$ and $\tilde{\theta}_{10}^p$ are estimated from 1000 such simulations with $p = 0.90$, 0.95 and 0.99 and are presented in Table 1.

The results of the simulation study are in concordance with the statement of Theorems 2.1 and 2.2 as the upper-bounds achieve their intended minimum coverage probabilities. Note that, when all the θ_k's are equal, for $k = 1,\ldots,10$, and the X_k's are independent, the coverage probability of $\hat{\theta}_{10}^p$ is close to p, for $p =$ 0.90,0.95 and 0.99, indicating that, in such a case,

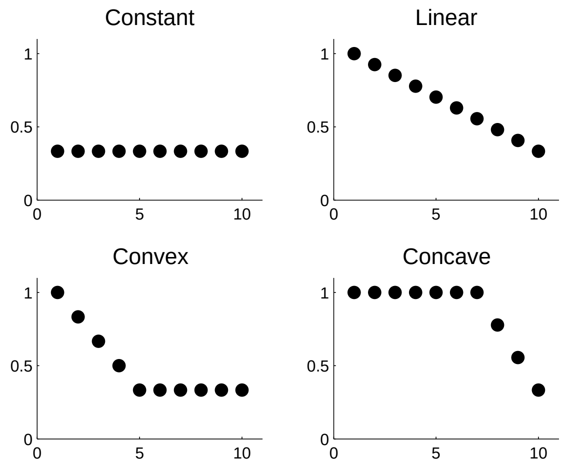

FIG 1. *Plots of* θ_k *vs* k *for* $k = 1, \ldots, 10$ *considered for the simulation study*

TABLE 1
Estimated coverage probability of the upper-bounds for the four patterns of decrease in θ_k

	Independent			Dependent		
Pattern	$\hat{\theta}_{10}^{0.90}$	$\hat{\theta}_{10}^{0.95}$	$\hat{\theta}_{10}^{0.99}$	$\tilde{\theta}_{10}^{0.90}$	$\tilde{\theta}_{10}^{0.95}$	$\tilde{\theta}_{10}^{0.99}$
Constant	0.90	0.95	0.99	0.92	0.96	0.99
Linear	0.99	0.99	0.99	0.98	0.99	1.00
Concave	0.94	0.97	0.99	0.95	0.97	0.99
Convex	0.98	0.99	1.00	0.98	0.99	1.00

the bounds will be tight, as is expected from the proof of Theorem 2.1. We will next study the average size of the conservative upper-bound which we define as the ratio of the upper-bound to the true parameter value, given by $\hat{\theta}_n^p/\theta_n$ and $\tilde{\theta}_n^p/\theta_n$, respectively. The results are presented in Table 2 for $p =$ 90%,95% and 99%. From the results, the average size seems to increase with p. It is also to be noted that the upper-bounds for the dependent model appear to be more conservative.

4. Applications. In this section, we consider three application areas for the proposed conservative upper-bound and illustrate with the analysis of a dataset in each area.

4.1. *A Sequential quality improvement plan with non-decreasing exponential failure times.* Consider a sequential quality improvement plan for a product with n quality improvement iterations. First, a prototype of the product is tested till its failure with corresponding failure time X_1 following

Table 2
Average values of $\hat{\theta}_{10}^{p}/\theta_{10}$ and $\tilde{\theta}_{10}^{p}/\theta_{10}$ for the four patterns of decrease in θ_k.

	Independent			Dependent		
Model	$\hat{\theta}_{10}^{0.90}$	$\hat{\theta}_{10}^{0.95}$	$\hat{\theta}_{10}^{0.99}$	$\tilde{\theta}_{10}^{0.90}$	$\tilde{\theta}_{10}^{0.95}$	$\tilde{\theta}_{10}^{0.99}$
Constant	1.82	2.08	2.72	2.92	3.14	3.80
Linear	3.16	3.42	4.28	3.92	4.25	4.78
Concave	2.16	2.56	3.28	3.12	3.42	4.13
Convex	3.82	4.11	4.85	4.37	4.69	5.06

exponential(θ_1) distribution with $E(X_1) = 1/\theta_1$. After the failure, the production process goes through a quality improvement exercise resulting in the second prototype which is again tested till its failure, after time X_2 following *exponential*(θ_2) distribution, is observed. We assume that due technical diligence is followed during the quality improvement exercise (i.e., product's quality does not deteriorate) so that $\theta_1 \geq \theta_2$ can be assumed. The n such successive improvement exercises give rise to n failure times $X_1, \dots, X_n$, with $X_k \sim exponential(\theta_k)$, for $k = 1, \dots, n$, and $\theta_1 \geq \dots \geq \theta_n$. A statistician is expected to estimate θ_n and provide an at least 95% upper-bound (one-sided confidence interval) for θ_n. Such an upper-bound can be used to obtain an at least 95% lower bound for the probability that the nth improved version does not fail before a certain time t_0 (that is, the reliability). Theorem 2.2 can be used to compute an at least $100 \times p$ % upper-bound for θ_n *without making any assumption about the nature of dependence* between X_k's as

$$\tilde{\theta}_n^p = \frac{log(n) - log(1-p)}{max(X_1, \dots, X_n)}. \tag{4.1}$$

Under the assumption of independence of X_k's, the MLE of θ_n can be obtained through the PAV algorithm (Ayers 1955) as given by

$$\hat{\theta}_n = \frac{1}{max\left\{X_n, \frac{X_n + X_{n-1}}{2}, \dots, \frac{X_n + \dots + X_1}{n}\right\}}. \tag{4.2}$$

However, since the asymptotic distribution of $\hat{\theta}_n$ is difficult to obtain (See Li, Taylor and Nan 2010 for an example relating to binomial distribution), finding a $100 \times p\%$ confidence interval for θ_n can be difficult too. On the other hand, the proposed upper-bound $\hat{\theta}_n^p$ as given by

$$\hat{\theta}_n^p = \frac{-log\left(1 - p^{1/n}\right)}{max(X_1, \dots, X_n)} \tag{4.3}$$

provides a conservative one-sided confidence interval, when the X_k's can be assumed to be independent.

The Dataset of Musa (2012) consists of software failure times for 136 iterations of debugging for a subsystem of a commercial software. In reality, failure times from this many iterations are hardly available and hence we consider the more realistic problem of estimating the reliability of the software using the observed failure times in the last 10 iterations only. The data is provided in Table 3.

TABLE 3
The inter-failure times (CPU seconds) for the last 10 iterations

Iteration (k)	1	2	3	4	5	6	7	8	9	10
Failure Time (X_k)	40	2	86	221	6	891	23	4	437	66

Two popular models for software failure data are the Jelinski and Moranda (1972) model and Moranda (1975) model which assume the failure time X_k to follow $exponential(\theta_k)$ distribution with $\theta_k = \lambda(N-k)$ and $\theta_k = exp(\alpha - \beta k)$, respectively. Assuming independence of the X_k's, the parameters (λ, N) or (α, β) can be estimated through the maximum likelihood principle and the upper-bound of the asymptotic 95 % one-sided confidence interval for θ_{10} may be obtained. Alternatively, an at least 95% upper-bound $\hat{\theta}_{10}^{0.95}$ for θ_{10}, without making any assumption about the nature of decrease in θ_k's can be obtained through (4.3). Moreover, the same $\left\{\tilde{\theta}_n^p\right\}$ can be obtained through (4.1) without making the independence assumption. The corresponding conservative lower bounds for θ_{10} are given in the third and fourth rows of Table 4. Those against Jelinski and Moranda (first row) and Moranda (second row) are obtained through parametric bootstrap of the respective models.

TABLE 4
Comparison of upper-bounds for θ_{10}

	Confidence		
Model	90%	95%	99%
Jelinski and Moranda	0.0072	0.0089	0.0123
Moranda	0.0065	0.0083	0.0141
$\hat{\theta}_n^p$	0.0051	0.0059	0.0077
$\tilde{\theta}_n^p$	0.0052	0.0060	0.0078

Interestingly, the upper-bounds of θ_{10} obtained through the two parametric modeling assumptions are more conservative than the proposed conservative upper-bounds, casting doubts on these models. This indicates that the reliability of the software could be increasing much faster than postulated by either the Jelinski and Moranda or the Moranda model. In the face of uncertainty regarding the parametric assumptions and small sample size, the proposed conservative bounds may be acceptable. Also, similar values of $\hat{\theta}_n^p$ and $\tilde{\theta}_n^p$ for different values of p give evidence for independence of the X_k's.

4.2. *Sequential quality improvement plans with bounded decrease in reliability.* In certain reliability applications, the assumption of non-decreasing reliability of subsequent iterations cannot be justified and hence the assumption $\theta_1 \geq \ldots \geq \theta_n$ may not be valid. In such a situation, an assumption such as $\beta\theta_{k-1} \geq \theta_k$, for a known $\beta > 0$ might be appropriate. If $\beta > 1$, then the parameter sequence need not be non-increasing and would allow bounded increase of the parameter sequence. For example, if failure times $X_1, \ldots, X_n$ are independently distributed as exponential variables, but the order restriction of $E(X_1) \leq \ldots \leq E(X_n)$ is hard to justify, one may consider a restriction of the form $E(X_{k-1}) \leq \beta E(X_k)$, with a known $\beta > 1$. Then, consider the parameter sequence $\gamma_1 = \theta_1, \gamma_2 = \theta_2/\beta, \ldots, \gamma_n = \theta_n/\beta^{n-1}$. Note that the parameter sequence $\gamma_1, \ldots, \gamma_n$ is non-increasing and $\beta^{k-1}X_k \sim exp(\gamma_k)$. Hence, Theorem 2.1 can be applied to the sequence $X_1, \beta X_2, \ldots, \beta^{n-1}X_n$ to obtain an at least $100 \times p\%$ confidence upper-bound $\hat{\gamma}_n^p$ for γ_n. The at least $100 \times p\%$ upper-bound for θ_n can be obtained as $\beta^{n-1}\hat{\gamma}_n^p$. Also, the PAV estimator (4.2) can be used to compute the MLE of θ_n as

$$(4.4) \qquad \hat{\theta}_n = \frac{\beta^{n-1}}{max\left\{\beta^{n-1}X_n, \frac{\beta^{n-1}X_n+\beta^{n-2}X_{n-1}}{2}, \ldots, \frac{\beta^{n-1}X_n+\ldots+X_1}{n}\right\}}.$$

Jelinski and Moranda (1972) present the time to failures of a software subsequent to fixing the defect which caused the previous failure. They provide the failure time data for 26 iterations of debugging and testing. As with the previous analysis in Section 4.1, we consider only the last 10 failure times. The data is presented in Table 5. Jelinski and Moranda assume that the kth failure time $X_k \sim exp(\theta_k)$ with $\theta_k = \lambda(N-k)$. The linear decrease model and the assumption that the parameters θ_k must decrease with k has often been questioned. For this example, considering the failure times in Table 5, the assumption of successive X_k's being stochastically larger does not seem to be right. Therefore, the assumption of bounded increase of the θ_k's seems more appropriate. Theorem 2.1 enables us to compute a conservative 95% upper-bound for θ_n, if we assume $\theta_k \leq \beta\theta_{k-1}$. For the illustration, we consider $\theta_k \leq 2\theta_{k-1}$. The application of Theorem 2.1 results in a conservative 95% upper-bound for θ_{10}, as $\hat{\theta}_{10}^{0.95} = 0.23$, which is suprisingly close to the value of 0.21 (obtained from the graphically presented posterior mean and posterior standard deviation for θ_{10}) as reported in Basu and Ebrahimi (2003).

TABLE 5
The inter-failure times (days) for the last 10 iterations

Iteration (k)	1	2	3	4	5	6	7	8	9	10
Failure Time (X_k)	3	3	6	1	11	33	7	91	2	1

4.3. *Software Reliability from bug-databases and error loggers.* Software bug-databases that record user-reported defects provide an increasingly important source of data for software reliability assessment. A bug-database is often used to track the set of known defects in the software up to a calender time t. Error loggers perform a similar function but the errors are not user-reported, but system-reported. We denote the set of known defects up to time t by $D(t)$, with each defect identified by a unique defect id. Software updates are often based upon fixing the set of known defects in the bug-database. When a defect is reported at time t, the defect is compared with the known set of defects $D(t-)$ to determine whether it is already known and, if it is a new defect, then $D(t)$ is updated with the new defect. Define $p(t)$ to be the probability that a defect reported at time t is not contained in the known set of defects $D(t-)$. In order to minimize the number of software updates, it is important to release an update by fixing the defects in $D(t)$ only if the probability $p(t)$ is less than a certain threshold (i.e., the probability of recording an unknown defect is small). A simple and generic model for $p(t)$ is to assume that $p(t) = p(s)$ if $D(t) = D(s)$ and $p(t) \leq p(s)$ if $D(s) \subseteq D(t)$. Let $T_1, \ldots, T_n$ be the calender times when the first n "new" defects are observed and let X_k be the number of defects that are observed between T_{k-1} and T_k, for $k = 1, \cdots, n$, with $T_0 = 0$. Let $p_k = p(T_k)$, for $k = 1, \ldots, n$. Then, clearly, $p_1 \geq p_2 \geq \ldots \geq p_n$. It may be reasonable to assume that $X_k \sim Geometric(p_k)$, for $k = 1, \ldots, n$, and X_i is independent of X_j for $i \neq j$. The PAV algorithm can be used to obtain the MLE of p_n, but as argued before, finding a confidence interval for p_n is difficult. Theorem 2.1 can be used to obtain an at least $100 \times p\%$ upper-bound for p_n as given by

$$\hat{p}_n^p = 1 - exp\left(\frac{log(1-p^{1/n})}{max(X_1,\ldots,X_n)}\right). \tag{4.5}$$

If the X_k's cannot be assumed to be independent, then Theorem 2.2 gives an at least $100 \times p$ % upper-bound for p_n as

$$\tilde{p}_n^p = 1 - exp\left(\frac{log((1-p)/n)}{max(X_1,\ldots,X_n)}\right). \tag{4.6}$$

Many operating systems provide software for monitoring system errors when they arise. After observing the error logs till n distinct errors are reported, it may be necessary to compute the probability that a subsequent error that is logged will not belong to the observed set of errors. For an operating system observed by the authors, the data consisting of the number of additional errors observed before the kth new error is observed, for $k = 1, \ldots, n = 10$, is presented in Table 6. Using (4.5), an at least 90%, 95% and 99% upper-bounds for p_{10} are obtained as 0.047, 0.054 and 0.071, respectively. These bounds using (4.6), without assuming independence, are 0.048, 0.055 and 0.072, respectively. As expected, the bounds for the dependent model are more, although slightly, conservative.

TABLE 6
Number of known defects reported from the operating system under use.

Iteration (k)	1	2	3	4	5	6	7	8	9	10
Number of defects (X_k)	7	97	1	9	54	87	5	14	48	49

5. Concluding remarks. Computing confidence bounds for reliability with a minimum coverage probability is important for assessing the risk of failure in a product whose quality has been improved sequentially. A non-parametric model for the increase in reliability may provide a worst-case scenario for failure, which in turn may be used for risk management. As is demonstrated in Section 4.2, the proposed methodology can be modified for applications which may allow the reliability to decrease with a subsequent version of the product. The proposed method can be used as an alternative to Bayesian procedures proposed by various authors for considering possibilities of a decrease in reliability after fixing a defect. The generality of Theorems 2.1 and 2.2 can be used for computing one-sided confidence bounds for any parametric family of failure time distributions indexed by a single unknown parameter.

Often the failure time distributions $F(x;\theta)$ is continuous and monotonic in the single index parameter θ. In case $F(x;\theta)$ is decreasing in θ and successive θ_k's are non-decreasing, unlike the conditions in Theorem 2.1, one can re-parametrize $F(x;\theta)$ to satisfy the conditions and apply Theorem 2.1. Also, when the distribution involves more than one parameter, often there is one parameter of interest which keeps changing over different iterations while the others remain unchanged. Theorem 2.1 then gives bounds for the parameter of interest.

When X_k's are independent and identically distributed (IID), one may use likelihood methods to obtain asymptotic confidence bounds. Alternatively, Theorem 2.1 can be used to obtain a conservative upper-bound even for small sample sizes. Moreover, Theorem 2.2 gives such a bound even when X_k's are not independent. From the proofs of Theorems 2.1 and 2.2, the conservative bound in the IID case is expected to be tighter with more accurate coverage probability and relatively smaller size, possibly because of sharing of more information. Our simulation study (See Tables 1 and 2) also indicates this.

References.

[1] Andrew, D. W. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space, Econometrica, 68: 399-405.

[2] Ayer,M., Brunk, H.D., Ewing, G.M., Reid,W.T., Silverman, E., 1955. An empirical distribution function for sampling with incomplete information. Ann. Math. Statist. 26, 641–647.

[3] Basu, A. and Ebrahimi, N. (2003). Bayesian Software Reliability Models Based on Martingale Processes, Technometrics, Vol. 54, No. 2, 150-158.

[4] Bhattacharya, R. and Kong, M. (2007). Consistency and asymptotic normality of the estimated effective doses in bioassay, Journal of Statistical Planning and Inference, 137, 643 – 658

[5] Efron, B. and Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals and Other Measures of Statistical Accuracy, Statistical Science, Vol. 1, No. 1, 54-77.

[6] Goel, A.L. and Okumoto, K. (1978). An Analysis of Recurrent Software Failures on a Real-time Control System, . Proc. ACM Annu. Tech. Conf., Washington D.C., 496-500.

[7] Jelinsky, Z. and Moranda P. B. (1972). Software Reliability Research. In Statistical Computer Performance Evaluation, ed. Freiberger W., 465-484.

[8] Jewel, N. P. and Kalbfleisch, J. D. (2004), Maximum likelihood estimation of ordered multinomial parameters, Biostatistics, 5, 291–306.

[9] Kuo, L., and Yang, T. Y. (1996), Bayesian Computation for Nonhomogeneous Poisson Processes in Software Reliability, Journal of American Statistical Association, 91, 763–773.

[10] Littlewood, B. and Verall, J. L. (1973), A Bayesian Reliability Growth Model for Computer Software, Applied Statistics, 22, 332-46.

[11] Moranda, P. Predictions of software reliability during debugging, Proceedings of the Annual Reliability and Maintainability Symposium, Washington, D. C., January 1975, 327-332.

[12] Musa, J. D. (2012), Dataset 6, https://sw.thecsiac.com/databases/sled/swrel.php.

[13] Musa, J. D. and Okumoto, K. (1975), A Logarithmic Poisson Execution Time Model for Software Reliability Measurement, Proceedings of the 7th IEEE Conference on Software Engineering, 230-238.

[14] Nelsen, Roger B. (1999), An Introduction to Copulas, New York: Springer.

[15] Singpurwalla, N. D. and Wilson, S. P. (1999), Statistical Methods in Software Reliability, Springer, New York.

[16] Tebbs, J. M. and Swallow, W. H. (2003), Ordered Binomial Proportions with the Use of Group Testing, Biometrika, 90, 471-477.

[17] Zhiguo Li, Jeremy M. G. Taylor and Bin Nan (2010). Construction of Confidence Intervals and Regions for Ordered Binomial Probabilities, The American Statistician, 64:4, 291-298.

GE Global Research
Bangalore 560066, India
E-mail: vignesh.ts@ge.com

Indian Statistical Institute
Calcutta 700108, India
E-mail: dewanji@isical.ac.in; bimal@isical.ac.in