Biometrika (1993), **80**, 1, *pp.* 193–201
Printed in Great Britain

Stopping rules and estimation for recapture debugging with unequal failure rates

BY ANNE CHAO, M.-C. MA
Institute of Statistics, National Tsing Hua University, Hsin-Chu 30043, Taiwan
AND MARK C. K. YANG
Department of Statistics, University of Florida, Gainesville, Florida 32605, U.S.A.

SUMMARY

A theoretical optimal stopping rule based on the minimization of the testing cost and expected penalty due to the unremoved bugs is derived without the equal failure rate assumption under Nayak's (1988) recapture debugging procedure. Two adaptive stopping rules suitable for practical use are constructed and tested by simulation. The estimation of the number of undetected bugs is also considered via remaining failure rate estimation.

Some key words: Optimal stopping; Poisson process; Remaining failure rate; Sample coverage; Software reliability.

1. INTRODUCTION

We consider the problem of finding stopping rules for software testing, and estimating the number of remaining errors, i.e. faults and bugs, in the software. Assume that there are N errors which are indexed by $1, \ldots, N$. Most previous models have assumed that (i) the occurrences of the ith error ($i = 1, \ldots, N$) follow a Poisson process with a common failure, or detection, rate, and (ii) an error is removed when detected and no additional errors are added. Previous work includes that by Jelinski & Moranda (1972), Blumenthal & Marcus (1975), Forman & Singpurwalla (1977), Goudie & Goldie (1981), Joe & Reid (1985) and Langberg & Singpurwalla (1985). A Bayesian approach is discussed by Dalal & Mallows (1988).

The common rate assumption has been criticized and discussed by Littlewood & Verrall (1973) and Goel & Okumoto (1979). In a departure from the second assumption, Nayak (1988) proposed a recapture debugging process in which each detected error is still corrected but a counter is inserted to record each time the error would have recurred. Nayak showed that extra information can be obtained from this recapture debugging procedure compared to the traditional debugging procedure. The assumption of an equal failure rate among bugs was, however, still made by Nayak. He also discussed maximum likelihood and moment estimators as well as a stopping rule. Goudie (1990) recently presented a likelihood-based stopping rule for recapture debugging and generalized it to a situation where there are two different detection rates. The jackknife estimator proposed by Burnham & Overton (1978, 1979) from the capture-recapture literature can be applied to estimate the number of remaining bugs for unequal failure rates.

Nayak's recapture debugging procedure is considered for the heterogeneous case in this paper. For technical convenience, the estimation procedure proposed in § 2 is discussed before the stopping rules of § 3. Simulation results are finally given in § 4.

2. ESTIMATION

The testing 'time' is assumed to be measured in some way, e.g. working hour or CPU time. The occurrences of the ith error follow a Poisson process with rate λ_i and the software is tested until a predetermined time τ. Detection times for different bugs are also assumed to be independent. Let X_i $(i=1, \ldots, N)$ denote the number of occurrences of the ith error and let D be the number of distinct errors discovered by time τ. Also let $I(.)$ be the indicator function, and $f_k = \Sigma\, I(X_i = k)$ the number of errors which were detected exactly k times $(k=0, 1, \ldots, m)$, where m denotes the maximum frequency. Here $f_0 = N - D$ is the unobservable number of undetected bugs. Obviously, $\Sigma\, kf_k = \Sigma\, X_k, \Sigma\, f_k = D$.

If the usual likelihood function is used for statistical inference, then the above model contains too many parameters. Fortunately, we do not need to know all the λ_i. The essentially relevant parameters are the mean $\bar{\lambda} = \Sigma\, \lambda_i / N$ and coefficient of variation $\gamma = \{\Sigma\, (\lambda_i - \bar{\lambda})^2 / N\}^{\frac{1}{2}} / \bar{\lambda}$. A key quantity is the sample coverage, which is defined as

$$C = \sum_{i=1}^{N} \lambda_i I(X_i > 0) \Big/ \sum_{i=1}^{N} \lambda_i. \qquad (2\cdot1)$$

Note that C varies with the sample and is invariant to the choice of scale of the time unit. The random quantity $1-C$ is the relative remaining failure rate after testing. This is an important criterion in evaluating software reliability. Our basic motivation here is to estimate the number of remaining bugs via the estimation of the expected relative remaining failure rate, which can be well estimated for the case of unequal failure rates as will be shown in (2·9). The earliest references on sample coverage can be traced back to Good (1953) and Good & Toulmin (1956). Also refer to Esty (1985) and Chao & Lee (1992) for a discrete-time analogue of the formulation of this paper.

In the special case of equal failure rate $(\gamma = 0)$, C reduces to D/N. A natural estimator of N in this case is then

$$\hat{N}_0 = D/\hat{C}, \qquad (2\cdot2)$$

where $\hat{C}$ will be derived in (2·9). A similar type of estimator was introduced by Darroch & Ratcliff (1980) under a discrete-time species problem.

For unequal λ_i $(\gamma > 0)$, we first evaluate the discrepancy between $E(D)/E(C)$ and N. Note that

$$E(D) = N - \sum \exp(-\lambda_i \tau), \quad E(C) = 1 - \sum \lambda_i \exp(-\lambda_i \tau) / \sum \lambda_i. \qquad (2\cdot3)$$

It can be shown after an expansion that

$$\frac{E(D)}{E(C)} = N - \frac{N\tau\bar{\lambda}e^{-\tau\bar{\lambda}}}{E(C)}\gamma^2 + R_1, \qquad (2\cdot4)$$

where R_1 denotes the remainder term. Also note that

$$E(f_1) = \sum \tau\lambda_i \exp(-\tau\lambda_i) \qquad (2\cdot5)$$

$$= N\tau\bar{\lambda} \exp(-\tau\bar{\lambda}) + R_2, \qquad (2\cdot6)$$

where R_2 is the remainder term. It follows from (2·4) and (2·6) that

$$N = \frac{E(D)}{E(C)} + \frac{E(f_1)}{E(C)}\gamma^2 - R, \qquad (2\cdot7)$$

where $R = R_1 + R_2\gamma^2/E(C)$ is a function of the third and fourth central moments of the λ_i. The remainder term R is usually negligible when the variation between the λ_i is not relatively large, say $\gamma < 1$. The main result will then be based on the following approximation ignoring R in (2·7):

$$N \simeq \frac{E(D)}{E(C)} + \frac{E(f_1)}{E(C)}\gamma^2. \tag{2·8}$$

From the right-hand equation in (2·3), (2·5) and $E(\Sigma X_i) = \Sigma \lambda_i \tau$, an estimator for $E(C)$ and a predictor of C is

$$\hat{C} = 1 - f_1/\sum X_i = 1 - f_1/\sum if_i. \tag{2·9}$$

Nayak (1991) showed that $\Sigma X_i/\tau$ and f_1/τ are the best unbiased estimators of $\Sigma \lambda_i$ and $\Sigma \lambda_i \exp(-\lambda_i\tau)$ respectively, since $(X_1, \ldots, X_N)$ or equivalently $(f_1, \ldots, f_m)$ are complete and sufficient statistics for the parameters $(N, \lambda_1, \ldots, \lambda_N)$. It follows from (2·9) that the proportion of singletons is the proposed predictor for the relative remaining failure rate. Since

$$E\{\sum X_i(X_i - 1)\} = \sum i(i-1)E(f_i) = \tau^2 \sum \lambda_i^2,$$

we have

$$\gamma^2 = N \sum \lambda_i^2/(\sum \lambda_i)^2 - 1 = N \sum_{i=1}^{m} i(i-1)E(f_i) \Big/ \left\{E\left(\sum_{i=1}^{m} if_i\right)\right\}^2 - 1.$$

The estimator $\hat{N}_0$ in (2·2) leads to the following estimator of the nonnegative parameter γ^2:

$$\hat{\gamma}^2 = \max\{\hat{N}_0 \sum i(i-1)f_i/(\sum if_i)^2 - 1, 0\}. \tag{2·10}$$

The following estimator can then be proposed from (2·8) for unequal failure rates conditional on $\hat{C} > 0$, that is not all bugs are singletons:

$$\hat{N}_1 = \frac{D}{\hat{C}} + \frac{f_1}{\hat{C}}\hat{\gamma}^2. \tag{2·11}$$

If all bugs are singletons, we have $\hat{C} = 0$ and $\hat{N}_0 = \hat{N}_1 = \infty$. No method seems to work in this extreme case.

In order to obtain a variance estimator for the proposed estimator, $\lambda_1, \ldots, \lambda_N$ are assumed to be a random sample from an unknown distribution $F(\lambda)$. Unconditionally, $(f_0, f_1, \ldots, f_m)$ is approximately multinomially distributed and both $\hat{N}_0$ and $\hat{N}_1$ are functions of $(f_1, \ldots, f_m)$. The asymptotic normality then follows directly and an approximate variance estimator for $\hat{N}_1$ is

$$\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial \hat{N}_1}{\partial f_i}\frac{\partial \hat{N}_1}{\partial f_j}\hat{c}(f_i, f_j), \tag{2·12}$$

where

$$\hat{c}(f_i, f_j) = \begin{cases} f_i(1 - f_i/\hat{N}_1) & (i = j), \\ -f_i f_j/\hat{N}_1 & (i \neq j). \end{cases}$$

The adequacy of this variance approximation is evaluated in a simulation study in § 4.

Our experiences in simulation have suggested that the above procedure generally yields reasonable estimates when $\gamma < 1$. When $\gamma \geq 1$, we can slightly modify the above procedure as follows. Since the large bugs can be easily detected and removed, they may be ignored from a practical point of view; i.e. we may consider only those bugs with no more than κ detections. A suitable value for κ might be 10. The number of bugs detected more than κ times is then added to the resulting estimate. Restated, we concentrate on only a subset of the bugs so that the value of γ for these bugs is smaller than the original. The number of distinct bugs detected for this subset is

$$D^* = D - \sum_{i=1}^{N} I(X_i > \kappa) = \sum_{i=1}^{N} I(0 < X_i \leqslant \kappa).$$

The sample coverage can similarly be defined here and its estimator can be obtained as

$$\hat{C}^* = 1 - f_1 \Big/ \sum_{i=1}^{\kappa} if_i.$$

The modified estimators for equal and unequal rates respectively then become

$$\hat{N}_0^* = \sum_{i=1}^{N} I(X_i > \kappa) + D^*/\hat{C}^*, \quad \hat{N}_1^* = \sum_{i=1}^{N} I(X_i > \kappa) + \frac{D^*}{\hat{C}^*} + \frac{f_1}{\hat{C}^*}\hat{\gamma}^{*2}, \tag{2·13}$$

where

$$\hat{\gamma}^{*2} = \max\left\{\frac{D^*}{\hat{C}^*}\sum_{i=1}^{\kappa} i(i-1)f_i \Big/ \left(\sum_{i=1}^{\kappa} if_i\right)^2 - 1, 0\right\}. \tag{2·14}$$

Variance estimators of $\hat{N}_0^*$ and $\hat{N}_1^*$ can be analogously obtained as previously derived in (2·12).

3. Stopping rules

Let c_1 be the cost per unit of testing time, c_2 be the cost of incurring one error in the field, and T be the expected total user-time that the software will be used in the field or the total user-time until the next revision. Assume that the test terminates at time $t > 0$. The expected number of field errors is then $\Sigma\, T\lambda_i I\{X_i(t) = 0\}$, where $X_i(t)$ is the number of occurrences of error i in the time interval $[0, t]$. The following cost or loss function is then considered:

$$L^*(t) = c_1 t + c_2 T \sum_{i=1}^{N} \lambda_i I\{X_i(t) = 0\}.$$

Equivalently, we can use

$$L(t) = t + (c_2 T/c_1) \sum_{i=1}^{N} \lambda_i I\{X_i(t) = 0\}. \tag{3·1}$$

The penalty here, unlike most previous criteria, is not assessed by the number of remaining bugs, but by the chance that they are encountered in the field, since a small number of bugs with higher failure rates can cost more than a large number of bugs with very low failure rates. Two stopping schemes are used as follows.

Scheme 1. The testing procedure can stop at any time $t \geqslant 0$. This is a continuous-time stopping problem.

Scheme 2. The stopping procedure only stops at epochs when an error is found. This becomes a discrete-time stopping problem.

We first consider the continuous-time stopping problem. Let $\mathscr{F}_t$ be the σ-field generated by recapture history up to time t. The purpose is now to find the optimal stopping rule τ_{opt}, for which

$$E\{-L(\tau_{\text{opt}})\} = \max E\{-L(\tau^*)\}$$

for any stopping rule τ^* such that $(\tau^* > t) \in \mathscr{F}_t$.

Following the approach of Kramer & Starr (1990) and Starr, Wardrop & Woodroofe (1976), we first let $R(t) = -\Sigma\, \lambda_i I\{X_i(t)=0\}$ and evaluate the infinitesimal operator

$$\alpha(t) = \lim_{h \to 0^+} h^{-1} E\{R(t+h) - R(t) \mid \mathscr{F}_t\}.$$

It is easily seen that

$$E\{R(t+h) - R(t) \mid \mathscr{F}_t\} = h \sum \lambda_i^2 I\{X_i(t)=0\} + o(h).$$

Obviously $\alpha(t) = \Sigma\, \lambda_i^2 I\{X_i(t)=0\}$ is nonincreasing in t, and the optimal stopping rule is

$$\begin{aligned} \tau_{\text{opt}} &= \inf\{t \geq 0 \mid \alpha(t) \leq c_1/(c_2 T)\} \\ &= \inf[t \geq 0 \mid \textstyle\sum \lambda_i^2 I\{X_i(t)=0\} \leq c_1/(c_2 T)]. \end{aligned} \tag{3·2}$$

The stopping rule τ_{opt} stops only at the epochs since $\Sigma\, \lambda_i^2 I\{X_i(t)=0\}$ is a step function with jump points at the epochs. An adaptive rule is now constructed since the λ_i's are unknown in practice. Let

$$f_k(t) = \sum_{i=1}^{N} I\{X_i(t) = k\},$$

which is the number of errors discovered exactly k times in $[0, t]$. Since

$$E\{f_k(t)\} = \sum_{i=1}^{N} (\lambda_i t)^k \exp(-\lambda_i t)/k!, \tag{3·3}$$

$$E[I\{X_i(t)=0\}] = \exp(-\lambda_i t),$$

$2t^{-2} f_2(t)$ is a predictor of $\Sigma\, \lambda_i^2 I\{X_i(t)=0\}$ in the sense that both have the same expected values. The adaptive stopping time is then chosen as

$$\hat{\tau}_c = \inf\{t \geq t_0 \mid 2t^{-2} f_2(t) \leq c_1/(c_2 T)\}. \tag{3·4}$$

An initial time t_0 is used here because the number of doubletons before the second epoch is by definition 0. Obviously, we do not wish to stop this early when the information on bugs has not been revealed by the testing. Note that, in (3·4), $\hat{\tau}_c$ can stop at any time, not necessarily at epochs.

The test which only terminates at the epochs when an error is found is now considered. Suppose we terminate at the nth epoch. Let $t_1, \ldots, t_n$ denote the occurrence times of the epochs. A discrete type cost function can then be formulated as

$$L(n) = t_n + (c_2 T/c_1) \sum_{i=1}^{N} \lambda_i I\{X_i(t_n) = 0\}.$$

Define $e_i = j$ if the ith epoch discovers the jth bug, and $\mathscr{F}_n = \sigma(t_i, e_i;\, i = 1, \ldots, n)$. The well-known result of Chow, Robbins & Siegmund (1971, p. 55) for the monotone case can then be applied. The optimal stopping is the smallest n such that

$$E\{-L(n+1) \mid \mathscr{F}_n\} \leq -L(n). \tag{3·5}$$

Since $t_{n+1}-t_n$ is an exponential random variable with mean $(\Sigma\,\lambda_i)^{-1}$, we have

$$E(t_{n+1}\,|\,\mathscr{F}_n)=t_n+(\Sigma\,\lambda_j)^{-1}.$$

Let $S_i(t)=X_i(t_{n+1})-X_i(t_n)$. Conditioning on $\mathscr{F}_n, (S_1,\ldots,S_N)$ is multinomially distributed with index 1 and cell probabilities $\lambda_i/\Sigma\,\lambda_j$ $(i=1,\ldots,N)$. Hence

$$E[I\{X_i(t_{n+1})=0\}\,|\,\mathscr{F}_n]=(1-\lambda_i/\Sigma\,\lambda_j)I\{X_i(t_n)=0\}.$$

It is easy to check that $A_n=[E\{-L(n+1)\,|\,\mathscr{F}_n\}\le -L(n)]$ is monotonically increasing with respect to n. The optimal stopping rule for discrete-time is then exactly the same as the optimal rule τ_{opt} given for continuous-time. Similarly an adaptive rule which stops at epochs only is

$$\hat{\tau}_d=\inf\{t_n\ge t_0\,|\,2t_n^{-2}f_2(t_n)\le c_1/(c_2T)\}. \qquad (3{\cdot}6)$$

The closeness in stopping times for the two adaptive rules $\hat{\tau}_c$ and $\hat{\tau}_d$ is revealed in the simulation results of § 4. It can be shown analytically that $E|\hat{\tau}_c-\hat{\tau}_d|\le 1/(\Sigma\,\lambda_i)$.

4. Simulation study

A simulation study was carried out to investigate the performance of the proposed stopping rules and estimation procedure. The following seven cases of failure rates were considered; CV is the coefficient of variation of the failure rates. The number of bugs was fixed to be 100 and the constant c in each case is a normalizing constant such that $\Sigma\,\lambda_i=1$; this normalizing constant can be adjusted to the scale of the time unit.

Case 1: (CV = 0). $\lambda_i=c=0{\cdot}01$ $(i=1,2,\ldots,100)$.
Case 2: (CV = 0·54). $\lambda_i=c/(i+20)$ $(i=1,2,\ldots,100)$.
Case 3: (CV = 0·57). $\lambda_i=cu_i$, $u_1,u_2\ldots,u_N$ a random sample from $U(0,1)$.
Case 4: (CV = 0·75). $\lambda_i=c/(i+10)$ $(i=1,2,\ldots,100)$.
Case 5: (CV = 0·99). $\lambda_i=c/(i+5)$ $(i=1,2,\ldots,100)$.
Case 6: (CV = 1·34). $\lambda_i=c/(i+2)$ $(i=1,2,\ldots,100)$.
Case 7: (CV = 2·25). $\lambda_i=c/i$ $(i=1,2,\ldots,100)$.

All cases except for Case 3, which was considered by Ross (1985), are a form of Zipf's law, which is widely prevalent in natural frequency data. We present the results only for the case $c_1/(c_2T)=10^{-3}$. Results for other values generally support the same conclusions. If we set $\Sigma\,\lambda_i=\psi$ instead of $\Sigma\,\lambda_i=1$, the expected stopping time would be $1/\psi$ times longer. All the results would then be valid for $c_1/(c_2T)=10^{-3}\psi^2$. For example, the expected stopping times become 10 times longer if $\Sigma\,\lambda_i=0{\cdot}1$. The results are then for the constant $c_1/(c_2T)=10^{-5}$.

For each case, 200 trials were generated and each trial ended when the stopping times were reached. For each generated data set, the theoretical optimal stopping time, τ_{opt}, and two adaptive stopping times, $\hat{\tau}_c$ and $\hat{\tau}_d$, were obtained. Their corresponding costs defined in (3·1) were also recorded. The minimal test time t_0 was fixed to be 50. Table 1 presents the average stopping times and costs over 200 trials. The proposed adaptive rules, $\hat{\tau}_c$ and $\hat{\tau}_d$, perform well, in the sense of their stopping times and costs being close to those of the theoretical optimum, when γ is small, but less satisfactorily when γ is large.

After obtaining the stopping rules, we reran the simulation for the same cases to investigate the performance of the proposed estimation procedure at various stopping times. Four stopping times 100, 200, 300 and the average value of $\hat{\tau}_d$, listed in Table 1, were selected. None of the trials yielded the extreme case that all bugs were singletons

Table 1. *Comparisons of optimal and adaptive stopping times, $t_0 = 50$*

Case		τ_{opt}	$\hat{\tau}_c$	$\hat{\tau}_d$
1	Mean stopping time	225	228	228
	Mean cost $\times 10^{-2}$	3·25	3·29	3·30
2	Mean stopping time	208	201	202
	Mean cost $\times 10^{-2}$	3·30	3·43	3·43
3	Mean stopping time	195	192	192
	Mean cost $\times 10^{-2}$	3·08	3·18	3·18
4	Mean stopping time	197	191	191
	Mean cost $\times 10^{-2}$	3·35	3·47	3·46
5	Mean stopping time	183	177	176
	Mean cost $\times 10^{-2}$	3·36	3·49	3·51
6	Mean stopping time	168	153	154
	Mean cost $\times 10^{-2}$	3·33	3·53	3·53
7	Mean stopping time	143	123	124
	Mean cost $\times 10^{-2}$	3·26	3·48	3·47

and hence the results may be interpreted as referring to the conditional distribution given $\hat{C} > 0$. Only the results for the stopping time $\hat{\tau}_d$ and Cases 1, 3, 5 and 7 are given in Table 2. Complete tables are given in a technical paper by the authors. The maximum likelihood, moment estimates proposed by Nayak (1988), the interpolated jackknife proposed by Burnham & Overton (1978), $\hat{N}_0$, $\hat{N}_1$, $\hat{N}_0^*$ and $\hat{N}_1^*$ as well as all estimated

Table 2. *Comparison of various estimators*

	Stopping time, $\hat{\tau}_d$	Method	Mean (median) estimate	Mean bias	Estimated SE	Sample SE	Sample RMSE
Case 1	228	$\hat{N}_{MLE}$	100 (100)	0	3·9	3·9	3·9
$\gamma = 0$,		$\hat{N}_j$	108 (113)	8	7·4	18·9	20·7
$D = 90$,		$\hat{N}_0$, $\hat{N}_0^*$	100 (100)	0	4·2	4·1	4·1
$\hat{C} = 0·90$		$\hat{N}_1$, $\hat{N}_1^*$	101 (100)	1	4·7	4·5	4·6
Case 3	192	$\hat{N}_{MLE}$	84 (83)	−16	3·6	4·7	17·0
$\gamma = 0·57$,		$\hat{N}_j$	97 (98)	−3	6·8	16·3	16·5
$D = 76$,		$\hat{N}_0$, $\hat{N}_0^*$	86 (86)	−14	4·4	5·5	14·8
$\hat{C} = 0·88$		$\hat{N}_1$, $\hat{N}_1^*$	90 (89)	−10	6·4	7·4	12·7
Case 5	176	$\hat{N}_{MLE}$	76 (76)	−24	3·4	5·5	24·2
$\gamma = 0·99$,		$\hat{N}_j$	100 (99)	0	8·6	21·7	21·7
$D = 69$,		$\hat{N}_0$	83 (83)	−17	5·5	6·5	18·1
$\hat{C} = 0·83$		$\hat{N}_1$	106 (105)	6	14·5	15·2	16·3
		$\hat{N}_0^*$	84 (84)	−16	5·8	6·8	17·3
		$\hat{N}_1^*$	101 (100)	1	12·4	13·5	13·5
Case 7	124	$\hat{N}_{MLE}$	51 (51)	−49	2·6	5·9	49·5
$\gamma = 2·25$,		$\hat{N}_j$	77 (78)	−23	9·4	25·1	33·7
$D = 47$,		$\hat{N}_0$	59 (59)	−41	6·3	7·4	41·3
$\hat{C} = 0·79$		$\hat{N}_1$	145 (139)	45	43·6	39·0	59·5
		$\hat{N}_0^*$	66 (66)	−34	8·3	9·5	35·3
		$\hat{N}_1^*$	93 (89)	−7	20·8	21·5	22·7

SE, standard error; RMSE, root mean squared error.
$\hat{N}_{MLE}$, maximum likelihood estimator (Nayak, 1988); $\hat{N}_j$, interpolated jackknife (Burnham & Overton, 1978); $\hat{N}_0$, $\hat{N}_0^*$, proposed estimator for equal failure rate case; $\hat{N}_1$, $\hat{N}_1^*$, proposed estimator for unequal failure rate case.

standard errors were calculated for each generated data set. The value of κ was chosen as 10 in obtaining $\hat{N}_0^*$ and $\hat{N}_1^*$. The inverse of the observed Fisher information was used here as a variance estimator of the maximum likelihood estimator whereas variance estimates of the proposed estimators were obtained from (2·12). The formula for the estimated standard errors of the interpolated jackknife was provided by Burnham & Overton (1978). These 200 estimates and their estimated standard errors were averaged and the median estimate was also calculated. Based on these 200 estimates, sample standard errors as well as sample root mean squared errors were then obtained. The average of D, the number of distinct bugs observed, and $\hat{C}$, the coverage estimate, are also listed. The estimator of the expected relative remaining failure rate performs very well because the value of $E(C)$ for each case is identical with the average of $\hat{C}$ to the second decimal place. Only the maximum likelihood estimator is tabulated here for comparison since the moment estimator has similar performance. For Cases 1–3, $\hat{N}_i$ and $\hat{N}_i^*$ $(i=0, 1)$ are nearly identical; i.e. all errors were found no more than 10 times. The sample standard error and root mean squared error for the median estimate differ only slightly from those for the corresponding mean estimate. The former results are thus not shown in the table.

The maximum likelihood estimator, as expected, performs best in the sense of having the smallest root mean squared error when the failure rates are all equal, in Case 1, $\text{CV}=0$. When γ is increased, it becomes negatively biased and the magnitude of the bias increases with γ. The proposed $\hat{N}_1^*$ clearly outperforms all the others with respect to root mean squared error in the unequal failure rate cases under consideration. The interpolated jackknife estimator generally works better than the maximum likelihood but performs worse than the proposed $\hat{N}_1^*$. The estimator $\hat{N}_1$ yields reasonable estimates for $\gamma<1$ as previously mentioned. The proposed $\hat{N}_1^*$ improves significantly over $\hat{N}_1$ when $\hat{N}_1$ becomes positively biased if $\gamma \geqslant 1$, as shown in Case 7.

The estimated standard error based on (2·12) is tabulated in the sixth column of Table 2. The theoretical standard error formulae for the proposed estimators are shown by the numerical results to be generally satisfactory since their estimates are close to the sample standard errors given in the seventh column.

In summary, the usual maximum likelihood estimator works well only for the equal failure rate cases; for nonequal failure rates, the proposed $\hat{N}_1^*$ which treats higher and low frequencies separately is recommended for practical use.

Acknowledgements

The authors would like to thank the referees for pointing out a calculation error and also for insightful suggestions and comments which have greatly improved the paper. This research was supported by the National Science Council of Taiwan.

References

BLUMENTHAL, S. & MARCUS, R. (1975). Estimation population size with exponential failure. *J. Am. Statist. Assoc.* **70**, 913–22.

BURNHAM, K. P. & OVERTON, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**, 625–33.

BURNHAM, K. P. & OVERTON, W. S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology* **60**, 927–36.

CHAO, A. & LEE, S.-M. (1992). Estimating the number of classes via sample coverage. *J. Am. Statist. Assoc.* **87**, 210–7.

CHOW, Y. S., ROBBINS, H. & SIEGMUND, D. (1971). *Great Expectation: The Theory of Optimal Stopping.* Boston: Houghton Miffin.

DALAL, S. R. & MALLOWS, C. L. (1988). When should one stop testing software? *J. Am. Statist. Assoc.* **83**, 872–9.

DARROCH, J. N. & RATCLIFFE, D. (1980). A note on capture-recapture estimation. *Biometrics* **36**, 149–53.

ESTY, W. W. (1985). Estimation of the number of classes in a population and the coverage of a sample. *Math. Scientist* **10**, 41–50.

FORMAN, E. H. & SINGPURWALLA, N. D. (1977). An empirical stopping rule for debugging and testing computer software. *J. Am. Statist. Assoc.* **72**, 750–7.

GOEL, A. L. & OKUMOTO, K. (1979). Time-dependent error detection rate model for software reliability and other performance measures. *IEEE Trans. Rel.* **R-28**, 206–11.

GOOD, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237–64.

GOOD, I. J. & TOULMIN, G. (1956). The number of new species and the increase of population coverage when a sample is increased. *Biometrika* **43**, 45–63.

GOUDIE, I. B. J. (1990). A likelihood-based stopping rule for recapture debugging. *Biometrika* **77**, 203–6.

GOUDIE, I. B. J. & GOLDIE, C. M. (1981). Initial size estimation for the linear pure death process. *Biometrika* **68**, 543–50.

JELINSKI, Z. & MORANDA, P. M. (1972). Software reliability research. In *Statistical Computer Performance Evaluation*, Ed. W. Freiberger, pp. 465–84. New York: Academic Press.

JOE, H. & REID, N. (1985). Estimating the number of faults in a system. *J. Am. Statist. Assoc.* **80**, 222–6.

KRAMER, M. & STARR, N. (1990). Optimal stopping in a size dependent search. *Sequential Anal.* **9**, 59–80.

LANGBERG, N. & SINGPURWALLA, N. D. (1985). A unification of some software reliability models. *SIAM J. Sci. Statist. Comp.* **6**, 781–90.

LITTLEWOOD, B. & VERRALL, J. L. (1973). A Bayesian reliability growth model for computer software. *Appl. Statist.* **22**, 332–46.

NAYAK, T. K. (1988). Estimating population size by recapture sampling. *Biometrika* **75**, 113–20.

NAYAK, T. K. (1991). Estimating the number of component processes of a superimposed process. *Biometrika* **78**, 75–81.

ROSS, S. M. (1985). Software reliability: the stopping rule problem. *IEEE Trans. Software Eng.* **SE-11**, 1472–6.

STARR, N., WARDROP, R. & WOODROOFE, M. (1976). Estimating a mean from delayed observations. *Z. Wahr. verw. Geb.* **35**, 103–13.

[*Received December* 1990. *Revised September* 1992]