Title	**Estimating the number of faults: efficiency of removal, recapture, and seeding**
Author(s)	**Lloyd, CJ; Yip, PSF; Chan, KS**
Citation	**IEEE Transactions on Reliability, 1999, v. 48 n. 4, p. 369-376**
Issued Date	**1999**
URL	**http://hdl.handle.net/10722/43501**
Rights	**Creative Commons: Attribution 3.0 Hong Kong License**

Estimating the Number of Faults: Efficiency of Removal, Recapture, and Seeding

Christopher J. Lloyd
Australian Graduate School of Management, Sydney

Paul S.F. Yip
The University of Hong Kong, Hong Kong

Kin Sun Chan
American International Assurance, Hong Kong

Summary & Conclusions — Various experimental designs for estimating the number of faults in a system are studied including:

i. removal of each fault as it is detected,

ii. marking of each fault as it is detected,

iii. introduction of a known number of faults into the system followed by #i or #ii.

A unified framework is developed for comparing these designs; it also produces simplified estimators having high efficiency relative to maximum likelihood estimators. The designs are compared in terms of

· statistical accuracy,

· the number of failures that need to occur to achieve a given accuracy.

On the basis of these comparisons, some general recommendations are made on the level of seeding as well as the choice of removal or recapture designs. When the testing effort is sufficient so that roughly two thirds of the faults are detected, the removal-design is preferred over the recapture-design, and there are no gains from seeding. However, this conclusion depends on assigning unit cost to all fault detections, which might not always be reasonable.

1. INTRODUCTION

Acronyms

ML	maximum likelihood
MLE	ML estimator
MG	martingale
MGE	MG estimator
ZMMG	zero-mean MG

Notation

ν	unknown number of undetected faults
N_t	number of failures during $[0, t]$
λ_t	failure intensity at time t
	(for a homogeneous population)

$\lambda_i(t)$	failure intensity for fault i at time t
t_j	time of failure j
θ	proportion of seeded-faults relative to ν in the system
S	$\theta \cdot \nu$: number of 'seeded' faults added to the code
$N_i(t)$	number of times fault i is detected in $[0, t]$
$dN_i(t)$	increment of $N_i(t)$ in $[t, t + dt]$
U_t	number of new faults detected in $[0, t]$
K_t	number of known faults detected in $[0, t]$
M_t	number of known faults in the program at time t
λ	$\dfrac{1}{\tau} \cdot \displaystyle\int_0^\tau \lambda_t \, dt$: average failure intensity
f_t	$\dfrac{\lambda_t}{\lambda}$, a known function
F_t	$\displaystyle\int_0^t f_s \, ds$; note that $F_\tau == \tau$
P_t	Pr{a real fault is detected in time t}: $\mathrm{E}\left[\dfrac{U_t}{\nu}\right]$
	$\lambda \cdot F_t$: for a recapture experiment
	$1 - \exp(-\lambda \cdot F_t)$: for a removal experiment
P_τ	'capture effort'
r_t	$(1 + \theta) \cdot P_t$: s-expected number of failures divided by ν
m_t	$\mathrm{E}\left[\dfrac{M_t}{\nu + S}\right]$
$\mathcal{F}_t$	'history' of the experiment up to time t
E_t	s-expectation, conditional on $\mathcal{F}_t$
G_t	a ZMMG stochastic process
D_t	'conditional mean-derivative' matrix of dG_t
V_t	'conditional variance' matrix of dG_t
I_t	information matrix obtained by integrating $D_t^T \times V_t^{-1} \times D_t$
W_t	a conditionally non-random matrix
$\sigma[x]$	standard deviation of r.v. x
$\bar{x}$	$1 - x$

Nomenclature

• Failure: An incorrect output.

• Fault: Some defect in the code which, under at least one set of conditions, produces a failure.

• Removal experiment: As each fault is detected, it is corrected. Only detection-times are recorded.

• Recapture experiment: Both 'detection times' and 'location of the fault' are recorded. Thus, revisits of the same fault can be identified. Faults are corrected, and a counter is inserted to record the number of re-visits

• Fault seeding or fault injection: A known number of faults are inserted into the program. Both 'detection times' and 'location of the fault' are recorded. This is first suggested in [16]; see also [23].

• Relative efficiency: Reciprocal of the ratio of the standard deviations of two estimators.   ◄

Let there be an unknown number, ν, of faults in a computer program. The program is executed for a total time τ, perhaps under a range of conditions, to simulate the operating environment. Failures occur on N_τ occasions at times:

$0 < t_1 < t_2 < \ldots < t_{N_\tau} < \tau$.

The failure times until detection of fault i are positive r.v. with hazard rate $\lambda_i(t)$.

This paper estimates ν in the presence of the nuisance parameters $\lambda_i(t)$, and examines how efficiently this can be achieved under the experimental designs in section 2.

The $\lambda_i(t)$ being unknown makes ν difficult to estimate. Recapture & seeding experiments are aimed at enhancing knowledge of $\lambda_i(t)$ by establishing a baseline population of known faults against which the number of failures can be calibrated.

The main purpose of this paper is to compare these designs in terms of statistical & practical efficiency. Throughout, we concentrate on estimation of ν for given testing effort; rather than, say, the mean test-time until a given number of faults remain. Estimating ν is an important and separate problem, *eg*, in the clean-room software testing strategy [7, 17, 21].

Some faults will always remain, and the level of faults in released software are typically between 1 and 10 faults per 1000 lines of code [16].

2. SOME EXPERIMENTAL DESIGNS

Assumptions

A. Times between failures caused by faults are *s*-independent and exponentially distributed.

B. Each fault is equally likely to be detected: $\lambda_i(t) = \lambda_t$.

C. No new faults are created when a fault is corrected.

D. The failure intensity is uniform in time[1]: $\lambda_t = \lambda$.

Various devices are available to the tester to make these assumptions more likely to be satisfied. For instance, restarting the code from a different part of the program makes #A more likely, while if #B is questionable, faults

---
[1]This is a common assumption.

can be classified, either before or after testing, into more homogeneous groups. Assumptions #A – #C are used throughout. Assumption #D is used only where stated explicitly.

For a removal experiment, the available data are $\{N_\tau, t_1, \ldots, t_{N_\tau}\}$. Under assumptions #A – #D, the MLE ν is derived in [12] and asymptotic properties are treated mathematically in [22]. Because the MLE is often unstable, alternative estimates have been suggested [3, 13]. Other references are [4, 8, 11, 14]. The underlying difficulty with the removal method is that information about ν and λ is largely confounded in the data, rather similar to the way n and p are partly confounded in the binomial(n, p) distribution [19]. This suggests modifying the experiment so that separate information on λ becomes available.

In the *recapture* approach, as the experiment progresses, the number of revisited faults provides separate information on λ since, at any moment, the number of previously visited faults is known. Both MLE and moment-based estimators for ν are derived in [18], assuming equal failure intensity for detected & undetected faults. However, in a different context, this estimator for ν had been given 30 years earlier in [6]. Simulation studies [26, 27] showed that infinite estimates of ν occur with very small probability.

For the fault-injection method [16], detection of the known number of injected faults provides separate information about λ. In its original formulation, both seeded & real faults are corrected & removed as they are detected. MLE based only on the total numbers of seeded & real faults detected, not on their times, have been derived [5]. If the failure intensities for seeded & real errors have a known, constant ratio then estimators based on an MGE equation perform satisfactorily [25].

The recapture & seeded faults approaches can be combined [10]: faults, either seeded or real, are marked (and possibly removed) and the numbers of first-visits and revisits to both types of faults are recorded. Basically, think of the marked real faults as being added to the known number of seeded faults in the program. Estimation & minimization of anticipated debugging time was studied in [10], but not estimation of ν.

It is interesting to explore the difference in the performance of the recapture & removal methods for various amounts of seeded faults. For each type of experiment, estimators are studied which minimize standard error within a derived classes of estimators for ν in a continuous time setting. The statistical accuracy of some simpler MGE are compared to MLE for the same design. The relative efficiencies of the different designs are compared by the information in the optimal MGE equations [15]. The details are given in the appendix. 'Comparison by information' is equivalent to 'comparison by standard deviation' for large ν. Of course, in practice ν might not be sufficiently large for the asymptotic results to apply; thus the results of a simulation study are reported to confirm the results for finite parameter values.

3. INFERENCE VIA OPTIMAL ESTIMATING EQUATIONS

This section uses assumptions #A – #C, but not #D.

The framework is general enough to include removal, recapture, and seeding experiments. The total faults comprise 'S injected or seeded faults' and 'ν real faults'. Each $\{N_i(s) : 0 \le s \le \tau\}$ is a continuous time counting process with jumps of size $+1$.

$$\Pr\{dN_i(t) = 1 \mid \mathcal{F}_t\} = \lambda_t \, dt. \tag{1}$$

Also $\lambda_t = \lambda \cdot f_t$, where $\int_0^\tau f_t \, dt = 1$.

The N_t failures observed in $[0, t]$ comprise 'U_t previously unknown real faults' and '$K_t = N_t - U_t$ known (seeded or previously detected) faults'. Then

$$E_t[dU_t] = \lambda_t \cdot (\nu - U_t) dt, \quad E_t[dK_t] = \lambda_t \cdot M_t \, dt, \tag{2}$$

dU_t, dK_t denote the increments of U_t, K_t in $[t, t + dt]$. Increments are generated from the '$\nu - U_t$ undetected' and 'M_t detected' (marked or seeded) faults in the program and thus are s-independent, conditional on $\mathcal{F}_t$. Define

$$dG_1 = dU_t - \lambda_t \cdot (\nu - U_t) \, dt,$$
$$dG_2 = dK_t - \lambda_t \cdot M_t \, dt;$$

it follows that $dG_t = (dG_1, dG_2)^T$ are MG *differences*, and functions of the unknown parameter $\beta = (\nu, \lambda_t)$. For any conditionally non-random 2×2 matrices W_t, the stochastic integral

$$H_t = \int_0^t W_s \, dG_s \tag{3}$$

is a 2-dimensional ZMMG with respect to $\mathcal{F}_t$.

The matrices (A-5) in the general theory are:

$$D_t = - \begin{pmatrix} \lambda & \nu - U_t \\ 0 & M_t \end{pmatrix} \cdot f_t \, dt,$$
$$V_t = \begin{pmatrix} \nu - U_t & 0 \\ 0 & M_t \end{pmatrix} \cdot \lambda \cdot f_t \, dt. \tag{4}$$

Substituting optimal weights (A-4) into (3), gives the optimal estimating equations:

$$\int_0^\tau (\nu - U_s)^{-1} \, dU_s = \lambda \cdot \tau \tag{5a}$$

$$N_\tau = \lambda \cdot \int_0^\tau (\nu - U_s + M_s) \cdot f_s \, ds \tag{5b}$$

from observing the system up to time τ. The l.h.s. of (5a) is simply written

$$\frac{1}{\nu} + \frac{1}{\nu - 1} + \ldots + \frac{1}{\nu - U_\tau + 1}$$

which depends only on U_τ and not on the actual times that these events occur. This is approximated by $\log(\nu) - \log(\nu - U_\tau)$ as $\nu \to \infty$.

For a counting process X_t with intensity function γ_t, the log-likelihood function is [1]

$$\int_0^t \log(\gamma_s) \, dX_s - \int_0^t \gamma_s \, ds.$$

The failure intensity for the counting process U_t is $\gamma_t = \lambda_t \cdot (\nu - U_t)$, and for K_t is $\gamma_t = \lambda_t \cdot M_t$.

It is then simple to show that the optimal estimating equations (5) are identical to the likelihood score equations. Conditions under which such models are asymptotically Gaussian are in [22]; these conditions are easily established for the U_t and K_t, as $\nu \to \infty$. The estimators solving (5) are MLE.

The information (A-3) in the optimal equation is

$$I_\tau^* = \tag{6}$$
$$\begin{pmatrix} \lambda \int_0^\tau (\nu - U_s)^{-1} \cdot f_s \, ds & \tau \\ \tau & \lambda^{-1} \int_0^\tau (\nu - U_s + M_s) \cdot f_s \, ds \end{pmatrix}$$

By the Law of Large Numbers,

$$\frac{\nu - U_t}{\nu} \to \exp(-\lambda \cdot F_t)$$

and so

$$I_{1,1}^* \to \frac{\exp(\lambda \cdot \tau) - 1}{\nu}.$$

Similarly,

$$\frac{M_t}{\nu + S} \to \text{the (design dependent) } s\text{-expected proportion}$$

m_t of marked faults at time t; thus

$$I_{2,2}^* \to \frac{\nu}{\lambda^2} \cdot (1 - \exp(-\lambda \cdot \tau))$$
$$+ \frac{S}{\lambda} \cdot \int_0^\tau f_s \cdot m_s \, ds \quad \text{as } \nu \to \infty. \tag{7}$$

The Var$[\hat\nu]$ is the top left element of the inverse of (6).

4. A SIMPLER CLASS OF ESTIMATORS

Eq (5b) involves f_t. The linear combination

$$dG_t = \lambda_t^{-1} \cdot M_t \, dG_1 - \lambda_t^{-1} \cdot (\nu - U_t) \, dG_2$$
$$= M_t \, dU_t - (\nu - U_t) \, dK_t \tag{8}$$

depends only on ν. Eliminating an unknown nuisance parameter has been used successfully in several closely related contexts [24, 27]. Apply W_t as in (A-2), and equate to 0; the implied estimator is

$$\hat\nu_{MG} = \frac{\int_0^\tau W_s \cdot (M_s \, dU_s + U_s \, dK_s)}{\int_0^\tau W_s \, dK_s}. \tag{9}$$

This is an MG estimator. When the weights are functions of ν, (9) can be used recursively to solve the estimating equation.

By judicious choice of weight function, the stochastic integrals in (9) can be made independent of the capture-times, and depend only on the final U_τ, K_τ, M_τ. This implies that the estimator, and its distribution, are completely robust to time heterogeneity in λ_t. In other words,

assumption #D is not required for this family of estimators. The cost of this robustness is measured by comparing the asymptotic variance of $\hat{\nu}_{\mathrm{MG}}$ with that of the MLE. Using the general theory in the appendix with $p = 1$, the optimal 1×1 weights W_t are

$$W_t^* = \frac{1}{(\nu - U_t) \cdot (\nu - U_t + M_t)}.$$

The optimal information (A-3) for this optimal simplified estimator is

$$I_\tau^* = \int_0^\tau \frac{\lambda \cdot M_s}{(\nu - U_s) \cdot (\nu - U_s + M_s)} \cdot f_s \, ds \qquad (10)$$

$$\to \frac{\lambda \cdot (1 + \theta)}{\nu} \cdot \int_0^\tau \frac{\exp(\lambda \cdot F_s) \cdot m_s}{\exp(-\lambda \cdot F_s) + (1 + \theta) \cdot m_s} \cdot f_s \, ds$$

The asymptotic variance, the reciprocal of (12), is again of order ν.

5. INFERENCE FROM SEEDED REMOVAL EXPERIMENTS

For the removal approach, $M_t = S - K_t$. The optimal estimating (5b) becomes

$$\frac{N_\tau}{\lambda} = (\nu + S) \cdot \tau - \int_0^\tau N_s \cdot f_s \, ds \qquad (11)$$

which depends on the f_s and t_j. This not only complicates the computation of $\hat{\nu}_{\mathrm{REM}}$ but means that a form for f_s must be assumed (eg, assumption #D imposes $f_t \equiv 1$). Nevertheless, if f_t does not systematically increase with t then the integral in (11) is little affected by the precise form of f_t. When $f_s \equiv 1$ and $S = 0$, then this estimator (12) was given in [3]. It may be infinite with appreciable probability when $S = 0$ or when S is small (simulations not reported).

Under assumptions #A & #B,
$\mathrm{E}[M_t] = S \cdot \exp(-\lambda \cdot F_t)$ which implies that

$$m_t = \exp(-\lambda \cdot F_t) \cdot \frac{\theta}{1 + \theta}.$$

Substituting into (6) & (7), this leads to

$$\mathrm{Var}(\hat{\nu}_{\mathrm{REM}}) = \frac{\nu \cdot (1 + \theta) \cdot P_\tau \cdot \bar{P}_\tau}{(1 + \theta) \cdot P_\tau^2 - \bar{P}_\tau \cdot \log^2(\bar{P}_\tau)} \qquad (12)$$

$$P_t \equiv 1 - \exp(-\lambda \cdot F_t)$$
$$= \Pr\{\text{a fault leads to failure by time } t\}.$$

For the simplified estimator $\hat{\nu}_{\mathrm{MG}}$, two possibilities for the weights are:

$$W_t' = \frac{1}{(S - K_t) \cdot (\nu - U_t)},$$

$$W_t^* = \frac{1}{(\nu - U_t) \cdot (\nu + S - N_t)}.$$

W_t' leads to the explicit estimator

$$\hat{\nu}_{\mathrm{MG}} = U_\tau \cdot K_\tau \cdot \frac{S}{S - K_\tau}.$$

W_t^* gives the optimally weighted estimator (9) which depends on the event times. Neither requires assumption #D. The variance (10) of the optimal MG estimator is

$$\mathrm{Var}[\hat{\nu}_{\mathrm{MG}}] = \frac{\nu \cdot (1 + \theta)}{\theta \cdot (\exp(\lambda \cdot \tau) - 1)}$$
$$= \frac{\nu \cdot (1 + \theta) \cdot \bar{P}_\tau}{\theta \cdot P_\tau} \qquad (13)$$

Since θ appears in the denominator of (13), but not in (12), it follows that for small θ, even the optimal MG estimator is extremely inefficient compared to MLE. Indeed, when $\theta = 0$, the estimator (9) is undefined because $dK_t = 0$ for all t.
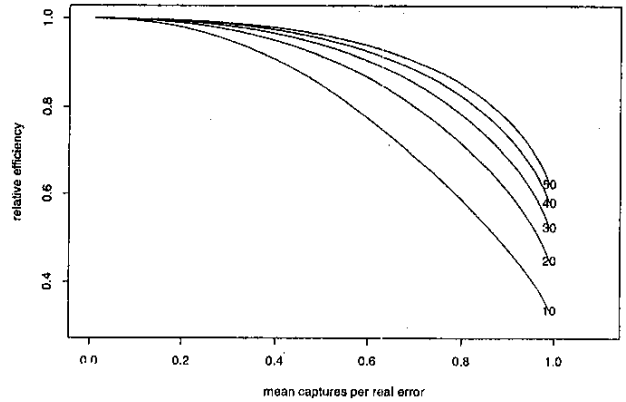

Figure 1: Efficiency of MG relative to MLE vs Capture Effort
For removal experiment with various seeding proportions.

Figure 1 shows the 'relative efficiency of $\hat{\nu}_{\mathrm{REM}}$ compared to $\hat{\nu}_{\mathrm{MG}}$' against P_τ, which measures the capture effort (mean fraction of faults removed). For $P_\tau \le 0.5$, and for more than $\theta = 0.2$ seeded faults introduced, the simple estimator is greater than 90% efficient compared to ML.

6. INFERENCE FROM SEEDED RECAPTURE EXPERIMENTS

In the recapture approach, the number of marked faults increases with each real fault: $M_t = S + U_t$. Hence the mean proportion of marked faults at time t is

$$m_t = \frac{1 - \exp(-\lambda \cdot F_t)}{1 + \theta}$$

which leads to

$$\mathrm{Var}[\hat{\nu}_{\mathrm{REC}}] = \frac{\nu \cdot (1 + \theta)}{(1 + \theta) \cdot (\exp(\lambda \cdot \tau) - 1) - \lambda \cdot \tau}. \qquad (14)$$

The ML equations (5) are easy to solve since, on substituting $M_s = S + U_s$, the second equation becomes $N_\tau = \lambda \cdot (\nu + S) \cdot \tau$, and substituting this into the first equation gives

$$\int_0^\tau (\nu - U_s)^{-1} \, dU_s = \frac{N_\tau}{\nu + S}. \qquad (15)$$

When $S = 0$, (15) reduces to the standard recapture estimator [6, 18].

For the simplified MGE, (9), consider optimal weights $W_t^* = \dfrac{1}{\nu - U_t}$ for which the optimal estimating equation $G_t = 0$ simplifies to

$$\int_0^\tau \frac{S + U_s}{\nu - U_s} \, dU_s = K_\tau. \tag{16}$$

This equation is identical to the MLE (16). A similar result was found in [2] for the experiment with no seeding. The estimator (16) does not depend on the failure times nor on f_t. It follows that the MLE for a seeded recapture experiment is valid without assumption #D.

7. REMOVAL OR RECAPTURE?

Special Assumption

E. Each failure has a unit cost, whether the corresponding fault is real, injected, known, or unknown.

From (12) & (14), it is easily verified that
· $\mathrm{Var}[\nu_{\mathrm{REM}}] \geq \mathrm{Var}[\hat{\nu}_{\mathrm{REC}}]$,
· both reduce with more seeding.
However, different designs can involve different costs. In general, the cost is a function of the
· total time τ,
· number of failures observed,
· unavoidable cost of correcting faults.
Estimation accuracies of the designs are compared, adjusting for assumption #E. Other cost functions could lead to different quantitative conclusions. But since 'recapture' and 'high levels of seeding' imply that 'faults known already' will continually be identified, some kind of tradeoff of cost and accuracy can always be anticipated.

Recapture experiments improve precision by providing marked faults from which λ is separately estimated. If no seeds are present, this is the only source of separate estimation of λ and we anticipate the gains from recapture to be large. If there is already a large proportion of seeds, then adding further seeds by 'not removing known faults' does not gain much. The cost of recapture is that there are more fault detections per unit time. For a sufficiently high seeding rate then, the costs of recapture could outweigh the benefits.

The mean number of detections is
$\nu \cdot (1 + \theta) \cdot P_\tau$, for removal experiment,
$\nu \cdot (1 + \theta) \cdot \lambda \cdot \tau$, for recapture experiment.
Let r_t be these quantities divided by ν: the mean number of failures per real fault. We could compare estimators by plotting standard error against r_t on the same plot. With the same intention, the asymptotic variances in (12), (14) can be expressed as explicit functions of r_t; then
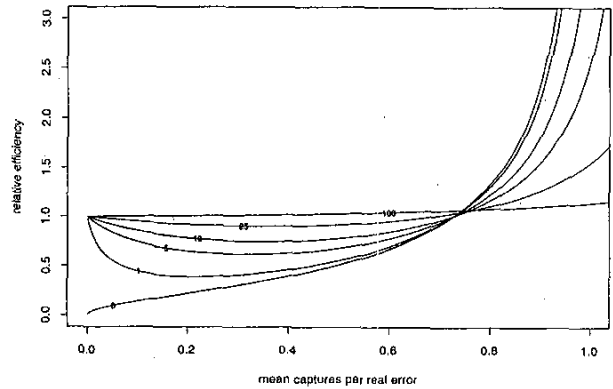

Figure 2: Effort-Adjusted Relative Efficiency of 'Removal Compared to Recapture' Experiment for Various Seeding Proportions

$$\frac{\sigma[\hat{\nu}_{\mathrm{REC}}]}{\sigma[\hat{\nu}_{\mathrm{REM}}]} = \tag{17}$$

$$\sqrt{\frac{(1 + \theta) \cdot \left[r_t^2 - (\theta + \bar{r}_t) \cdot \log^2\left(1 - \dfrac{r_t}{1 + \theta}\right)\right]}{r_t \cdot (\theta + \bar{r}_t) \cdot \left[\left(\exp\left(\dfrac{r_t}{1 + \theta}\right) - 1\right) \cdot (1 + \theta)^2 - r_t\right]}}$$

Figure 2 shows the relationship (17), over $0 < r_t < 1$; note that r_t can be larger than 1. Each curve corresponds to a specific seeding proportion, θ, labeled as percentages. The confluence near $r_t = 0.75$ is interesting and unanticipated. When $r_t < 0.75$, recapture-sampling is preferred to removal-sampling, for small amounts of seeding. As the seeding increases, this superiority reduces; when $\theta = 50\%$ the relative efficiency is very close to 1 for all values of r_t.

8. THE EFFECTS OF SEEDING

After expressing (12) in terms of r_t,

$$\frac{\sigma[\hat{\nu}_{\mathrm{REM}}; \theta)}{\sigma[\hat{\nu}_{\mathrm{REM}}; \theta = 0]} = \tag{18}$$

$$\sqrt{\frac{(\theta + \bar{r}_t) \cdot [r_t^2 - \bar{r}_t \cdot \log^2(\tilde{r}_t)]}{\bar{r}_t \cdot \left[r_t^2 - (\theta + \bar{r}_t) \cdot \log^2\left(1 - \dfrac{r_t}{1 + \theta}\right)\right]}}$$

is the relative efficiency for a removal experiment with seeding θ compared to no seeding, adjusted for equal mean number of failures. The corresponding expression for the recapture experiment is

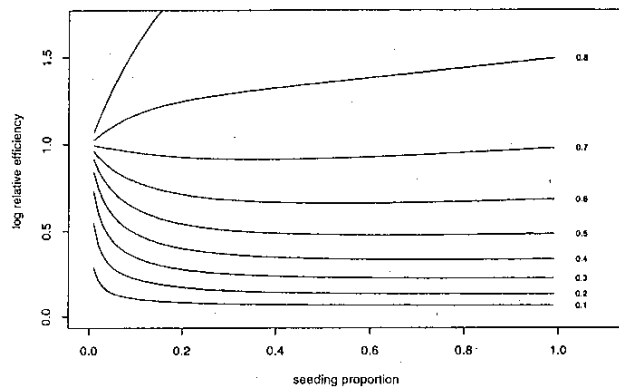$$\frac{\exp(r_t) - 1 - r_t}{\exp\left(\dfrac{r_t}{1 + \theta}\right) - 1 - \dfrac{r_t}{(1 + \theta)^2}}. \tag{19}$$

Figure 3 plots the logarithm of these relative efficiencies (18) for various proportions of θ; figure 3a shows the

removal experiment, and figure 3b shows the recapture experiment. In each case, the relative efficiency of seeding compared to no seeding is plotted against θ, each curve corresponding to a particular r_t. For small capture effort, extreme efficiency gains are achieved by only a small amount of seeding.
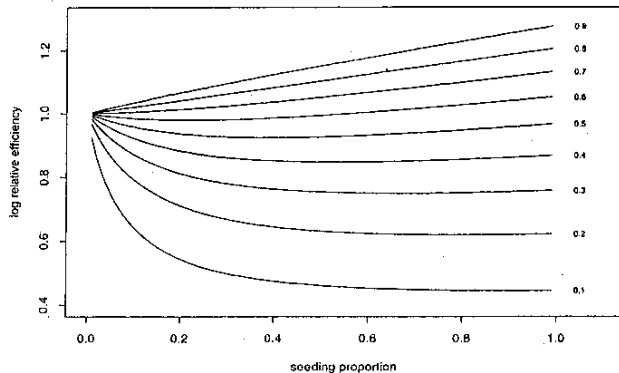
For the removal experiment, when the capture effort is around 70%, seeding has very little effect — the extra precision of seeding is almost exactly outweighed by the extra detection costs.

For the recapture experiment, the main differences are that the:

· efficiency gains are not nearly as large (note the different scales on the y-axis)
· point where seeding has little effect is at the lower cost of about 60%.

a. Removal Experiment

b. Recapture Experiment

Figure 3: Logarithm of Relative Efficiency of Seeded vs Unseeded Experiment For Various Seeding Proportions and Various Total Design Costs

9. SIMULATION STUDY

A simulation was performed to validate the asymptotic variance formulae (12), (14). All the asymptotic efficiency comparisons are based on these formulae, validation of which also validate the efficiency comparisons.

Table 1 relates to recapture and removal experiments with $\nu = 400$ and various values of θ and $\lambda \cdot t$. Each cell compares the theoretical standard error computed from (14) with the observed standard error of the 500 simulated realizations of $\hat{\nu}_{\text{REC}}$. There is good agreement between theoretical and observed for both recapture and removal experiments, except for small values of θ and $\lambda \cdot t$. Under the latter conditions, the distribution of $\hat{\nu}_{\text{REC}}$ is more skewed; thus we anticipate information (based on a linear approximation to the estimating equation) to be a less accurate estimator of the true variability. There is a decreasing trend in standard error as θ increases and as $\lambda \cdot t$ increases. For large capture effort $\lambda \cdot t$, increasing seeding from $\theta = 0.5$ to $\theta = 1.0$ produces only modest reduction in standard errors, but would involve the detection of many more faults.

The MLE of ν can break down when $\theta = 0$: the likelihood is maximized at infinity. For the recapture experiment this occurs only when there are no recaptures at all, which is extremely improbable. For the removal experiment, this occurs whenever the removal-rate appears to increase in time rather than decrease; see [13, 20] for sampling in discrete time. The possibility of such breakdown persists in the seeded experiments but occurs with smaller frequency. The proportions of simulations where breakdown occurred are given in parentheses. Further simulation results (not shown here) suggest that the asymptotic results for a smaller ν are less accurate.

All these comparisons depend on $\theta = \dfrac{S}{\nu}$ as well as r_t which is a function of the total hazard $\lambda \cdot t$. In the context of continuing software development, there is some prior information on both ν and λ before a cycle of testing begins. There is no reason why λ should change from cycle to cycle; thus previous estimates could be used. The remaining faults could be estimated by the previous estimate minus the number of corrections reported by the repair team. The $(\hat{\nu}, \hat{\lambda})$ can suggest an optimal design strategy for the testing team.

Another measure of system reliability is the mean time to failure, $\mu = \nu \cdot \lambda$. It would useful to measure the efficiency of the different designs for inference on μ; however the precision of inference on μ is much higher than for ν, because of the negative correlation of estimates of ν & λ.

APPENDIX

Martingale Estimating Functions

Consider a stochastic process that develops in time. G_t is any r.v. depending only on data up to time t; dG_t measures the change in G_t in $[t, t + dt)$. G_t is ZMMG if

$$\text{E}_t[dG_t] = 0, \quad \text{E}_t[dG_t^2] < \infty; \qquad (\text{A-1})$$

E_t denotes s-expectation, conditional on $\mathcal{F}_t$. The process G_t is an accumulation of mean-zero finite-variance r.v. More generally, let $dG_t = (dG_1, \ldots, dG_k)^T$ be a martingale difference of dimension k. For each t let W_t be a $p \times k$ 'weight' matrix which depends only on data up to time t.

Table 1: Comparison of Theoretical & Observed Standard Faults in the Recapture & Removal Experiments, $\nu = 400$

θ	$\lambda \cdot t$	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8
						Recapture				
0	$\sqrt{I_t^{-1}}$	137	66.0	42.4	30.7	23.6	18.9	15.5	13.0	11.1
	se($\hat{\nu}$)	107	67.6	44.4	31.1	23.5	18.9	15.88	13.6	11.2
		(35%)	(2%)							
0.1	$\sqrt{I_t^{-1}}$	100	55.9	38.0	28.3	22.2	18.0	15.0	12.6	10.8
	se($\hat{\nu}$)	99.8	61.8	38.9	28.4	22.3	17.9	15.4	13.2	11.0
		(6%)								
0.5	$\sqrt{I_t^{-1}}$	67.4	42.1	30.8	24.0	19.5	16.2	13.7	11.8	10.2
	se($\hat{\nu}$)	71.0	43.6	30.5	24.8	19.8	16.3	14.1	12.0	10.5
1.0	$\sqrt{I_t^{-1}}$	57.4	37.0	27.7	22.0	18.1	15.2	13.0	11.3	9.8
	se($\hat{\nu}$)	58.1	37.1	27.3	22.8	18.3	15.0	13.2	11.4	9.9
						Removal				
0	$\sqrt{I_t^{-1}}$	737	248	128	79.5	54.2	39.3	29.7	23.2	18.5
	se($\hat{\nu}$)	111	98.8	82.9	70.4	54.2	42.2	30.7	24.3	19.6
		(54%)	(48%)	(37%)	(17%)	(6%)	(1%)			
0.1	$\sqrt{I_t^{-1}}$	139	88.9	64.3	48.6	37.8	29.9	24.1	19.6	16.2
	se($\hat{\nu}$)	108	87.6	65.4	48.8	38.6	29.6	24.9	20.8	17.1
		(32%)	(11%)	(3%)	(1%)	(1%)				
0.5	$\sqrt{I_t^{-1}}$	73.4	48.8	37.1	29.8	24.5	20.6	17.4	14.8	12.7
	se($\hat{\nu}$)	75.0	50.8	37.2	30.0	25.2	20.4	17.7	15.0	12.9
1.0	$\sqrt{I_t^{-1}}$	60.0	40.1	30.7	24.9	20.8	17.6	15.1	13.0	11.3
	se($\hat{\nu}$)	61.3	39.6	30.4	25.4	20.8	16.9	15.0	12.7	11.3

Note: The % shown in parentheses are the proportion of the 500 simulations which failed to provide a finite estimate. The se($\hat{\nu}$) was based on successful simulations.

Then, conditional on $\mathcal{F}_t$, these matrices are non-random; thus

$$H_t = \int_0^t W_s \, dG_s \tag{A-2}$$

is a p-dimensional ZMMG. When dG_t depends on an unknown parameter β of dimension p, then the solution $\hat{\beta}_t$ of $H_t = 0$ is a generalization of the 'method of moments' estimation. Details of both finite sample and asymptotic theory of such estimators are in [9]. They establish mild conditions under which $I_t^{1/2}(\hat{\beta}_t - \beta)$ converges to multivariate standard Gaussian variable where the $p \times p$ random matrix I_t is the information in H_t. In the partial ordering of matrices, I_t is dominated by

$$I_t^* = \int_0^t D_s^T \times V_s^{-1} \times D_s. \tag{A-3}$$

The W_t which achieves this bound is

$$W_t^* = D_t^T \times V_t^{-1}; \tag{A-4}$$

D_t and V_t are respectively $k \times p$ and $k \times k$ random matrices with entries

$$D_{i,j} = \mathrm{E}_t\left[\frac{\partial dG_i}{\partial \beta_j}\right], \quad V_{i,j} = \mathrm{E}_t\left[dG_i \, dG_j\right]. \tag{A-5}$$

The optimal set of equations is

$$H_t^* = \int_0^t D_s^T \times V_s^{-1} \times dG_s = 0.$$

Amongst the family of estimating functions given by (A-2) for given dG_t, the H_t^* has maximal s-correlation with the true likelihood score function, and in some of these applications will actually recover the score function. The asymptotic variance matrix of $\hat{\beta}_t^*$ is the inverse of (A-3). The D_t and V_t may be replaced by their s-expectations in (A-3) whenever

$$\frac{D_t}{\mathrm{E}_t[D_t]} \text{ and } \frac{V_t}{\mathrm{E}_t[V_t]}$$

converge to 1. Estimators can then be compared via comparisons of their (non-random) asymptotic information matrices.

REFERENCES

[1] O.O. Aalen, "Weak convergence of stochastic integrals related to counting processes", *Zeitschrifft fur Wahrverw. Geb.*, vol 38, 1977, pp 261 - 277.

[2] N.G. Becker, C.C. Heyde, "Estimating population size from multiple recapture experiments", *Stochastic Processes and Their Applications*, vol 36, 1990, pp 77 – 83.

[3] S. Blumenthal, R. Marcus, "Estimating population size with exponential failure", *J. American Statistical Assoc*, vol 70, 1975, pp 913 – 922.

[4] R.J. Carroll, F. Lombard, "A note on N estimators from the binomial distribution", *J. American Statistical Assoc*, vol 80, 1985, pp 423 – 426.

[5] J.W. Duran, J.J. Wiorkowski, "Capture-recapture sampling for estimating software content", *IEEE Trans. Software Engineering*, vol SE-7, 1981, pp 147 – 148.

[6] J.N. Darroch, "The multiple recapture census I. Estimation of a closed population", *Biometrika*, vol 45, 1958, pp 343 – 359.

[7] M. Dyer, "Cleanroom software development method", 1982; IBM Federal Systems Div, Bethesda.

[8] E.H. Forman, N.D. Singpurwalla, "An empirical stopping rule for debugging and testing computer software", *J. American Statistical Assoc*, vol 72, 1977, pp 750 – 757.

[9] V.P. Godambe, C.C. Heyde, "Quasi-likelihood and optimal estimation," *Int'l Statistical Review*, vol 55, 1987, pp 231 – 244.

[10] I.B.J. Goudie, "A plant-recapture approach for achieving complete coverage of a population", *Communications in Statistics, A*, vol 24, 1995, pp 1293 – 1305.

[11] I.B.J. Goudie, C.M. Goldie, "Initial size estimation for the linear pure death process", *Biometrika*, vol 68, 1981, pp 543 – 550.

[12] Z. Jelinski, P.M. Moranda, "Software reliability research", *Statistical Computer Performance Evaluation* (W. Freiberger, *Ed*), 1972, pp 465 – 484; Academic Press.

[13] H. Joe, N. Reid, "Estimating the number of faults in a system", *J. American Statistical Assoc*, vol 80, 1985, pp 222 – 226.

[14] B. Littlewood, J.L. Verrall, "Likelihood function of a debugging model for computer software reliability", *IEEE Trans. Reliability*, vol R-30, 1981 Jun, pp 145 – 148.

[15] C.J. Lloyd, "Efficiency of martingale methods in recapture studies", *Biometrika*, vol 81, 1994, pp 305 – 315.

[16] H.D. Mills, "On the statistical validation of computer programs", 1972; IBM Federal Systems Div.

[17] H.D. Mills, M. Dyer, R.C. Linger, "Cleanroom software engineering", *IEEE Trans. Software*, vol S-36, 1987, pp 19 – 25.

[18] T.N. Nayak, "Estimating population size by recapture sampling", *Biometrika*, vol 75, 1988, pp 113 – 120.

[19] I. Olkin, A.I. Petkau, J.U. Zidak, "A comparison of n estimators for the binomial distribution", *J. American Statistical Assoc*, vol 76, 1981, pp 637 – 642.

[20] G.A.F. Seber, J.F. Whale, "The removal method for two and three samples", *Biometrics*, vol 26, 1970, pp 393 – 400.

[21] R.W. Selby, V.R. Basili, F.T. Baker, "Cleanfoorm software development: An empirical evaluation," *IEEE Trans. Software Eng*, vol SE-9, 1987, pp 1027 – 1036.

[22] M.C. van Pul, "Asymptotic properties of a class of statistical models in software reliability", *Scandinavian J. Statistics*, vol 19, 1992, pp 235 – 253.

[23] J.M. Voas, G. McGraw, *Software Fault Injection*, 1998; John Wiley & Sons.

[24] P. Yip, "A martingale estimating equation for capture-recapture experiment in discrete time", *Biometrics*, vol 47, 1991, pp 1081 – 1088.

[25] P. Yip, "Estimating the number of faults in a system using a martingale approach", *IEEE Trans. Reliability*, vol 44, 1995 Jun, pp 322 – 326.

[26] P. Yip, D.Y.T. Fong, "Estimating population size from a removal experiment in discrete time", *Statistics and Probability Letters*, vol 16, 1993, pp 129 – 135.

[27] P. Yip, D.Y.T. Fong, K. Wilson, "Estimating population size by recapture sampling via estimating function", *Communications in Statistics, A*, vol 16, 1993, pp 129 – 135.

AUTHORS

Dr. Chris J. Lloyd; Australian Graduate School of Management; Univ. of New South Wales; Kensington, 2052 AUSTRALIA.
Internet (e-mail): chrisl@agsm.unsw.edu.au

Chris Lloyd received his PhD (1986) from the University of Melbourne. He has worked at La Trobe University, University of Waterloo, Imperial College, and University of Hong Kong. He is Reader at the Australian Graduate School of Management. His main research interests are estimation theory and ecological statistics, but he has recently become interested in reliability statistics.

Dr. Paul S.F. Yip; Dep't of Statistics and Actuarial Science; Univ. of Hong Kong; Pokfulam HONG KONG.
Internet (e-mail): sfpyip@hkucc.hku.hk

Paul Yip received his PhD (1989) from La Trobe University. He has wide research interests including the application of capture-recapture methods in ecology and epidemiology, and the analysis of infectious disease and suicide data. He is a Senior Lecturer in the Department of Statistics and Actuarial Science at the University of Hong Kong.

Kin Sun Chan; Dep't of Statistics and Actuarial Science; Univ. of Hong Kong; Pokfulam HONG KONG.
Kin Sun Chan is an actuarial trainee with American International Assurance. The material in this paper was part of his Master degree in statistics, and his Master degree supervisor was Dr. Paul Yip. He has research interests in survival and reliability data.