<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Estimating the number of errors in a system using a Martingale approach</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Yip, Paul</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>IEEE Transactions On Reliability, 1995, v. 44 n. 2, p. 322-326</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>1995</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/43502">http://hdl.handle.net/10722/43502</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.; ©1995 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.</td>
</tr>
</tbody>
</table>
Estimating the Number of Errors In a System Using a Martingale Approach

Paul Yip
University of Hong Kong, Pokfulam

Key Words — Martingale difference, Removal experiment, Time-dependent failure intensity, Weight function, Zero-mean martingale

Summary & Conclusions — A new, efficient procedure estimates the number of errors in a system. A known number of seeded errors are inserted into a system. The failure intensities of the seeded and real errors are allowed to be different and time dependent. When an error is detected during the test, it is removed from the system. The testing process is observed for a fixed amount of time \( t \). Martingale theory is used to derive a class of estimators for the number of seeded errors in a continuous time setting. Some of the estimators and their associated standard deviations have explicit expressions. An optimal estimator among the class of estimators is obtained. A simulation study assesses the performance of the proposed estimators.

1. INTRODUCTION

Mills [15] proposed a capture-recapture sampling method which allows estimation of the number of errors in a system by randomly inserting a known number of errors and then testing the system for both inserted & indigenous errors. Duran & Wiorkowski [5], and Yip & Fong [21] derived maximum likelihood estimates of the indigenous errors and showed methods to obtain \( s \)-confidence limits. They assumed that the seeded errors behave as if randomly selected from the distribution of possible real errors in the system. The error seeding method was also discussed in [2, 16].

Martingale theory is used here for an alternative estimation procedure which allows different failure intensities for the seeded & real errors. In order to avoid the identifiability problem, a known constant proportionality is assumed between the seeded & real errors. This problem can be related to a removal experiment for a closed population of a certain type of animal in wildlife studies. Several authors have addressed this problem [4, 6, 9 - 11, 13, 14]. Martingale theory is applied in reliability studies in [12, 19].

Martingale theory is used to derive a class of estimators for the population size of the real errors. Some of the estimators and their associated standard deviations have explicit expressions. An optimal estimator among the class estimators for the number of real errors is obtained.

Acronyms & Abbreviations

RMSE (square) root of mean square deviation

StdDev standard deviation

ZMM zero-mean martingale.

Notation

\( \mathcal{F}_t \) history of the process during \([0,t]\)

\( \nu \) number of real errors in the system (parameter of interest)

\( D \) number of seeded errors (known)

\( \lambda_r, \beta_r \) failure intensity for [real, seeded] errors

\( U_{1r}, M_{1r} \) number of [real, seeded] errors detected/removed in \([0, u]\)

\( U_t, M_t \) number of [real, seeded] errors detected/removed in \([0, t]\)

\( \Theta_1, \Theta \) zero-mean martingales

see (3)

\( \theta \) \( \lambda_r/\beta_r \), a known constant

\( \text{Av}(\hat{\theta}), \text{SD}(\hat{\theta}) \) [average, StdDev] of the 500 simulated values \( \hat{\theta} \) distributed as.

Other, standard notation is given in "Information for Readers & Authors" at the rear of each issue.

Assumptions

1. A known number of errors, \( D \), is seeded in the system at the beginning of the experiment.
2. \( \lambda_r, \beta_r \), may be time dependent.
3. \( \theta \) is a known constant; it need not be 1.
4. The same failure intensity is applied to each type of error in the system.
5a. Errors are removed (without introducing new errors or affecting old errors) immediately after detection.
5b. Errors are detected/removed one at a time. (This allows a continuous time formulation.)
6. \( U_1, M_1 \) are measurable w.r.t. \( \mathcal{F}_u \).

2. CONTINUOUS-TIME MARTINGALE

The methods here are based on results for continuous martingales and follow from the work of [1]. The appendix informally explains martingales. For this paper, ZMM is a stochastic process \( \{\Omega_t, t \geq 0\} \) such that:

\[ E[\Omega_0] = 0, \]

\[ E[|\Omega_t|] < \infty, \text{ for all } t \geq 0, \]
YIP: ESTIMATING THE NUMBER OF ERRORS IN A SYSTEM. USING A MARTINGALE APPROACH

\[ E\{\mathfrak{M}_{t+h}\mid F_t\} = \mathfrak{M}_t, \text{ for all } t \geq 0, h > 0. \]

We can construct ZMM with respect to the counting processes \( U_t \) & \( M_t \). The following 2 statements are true:

\[ dU_t\mid F_t = \text{binm}(\lambda_t \cdot dt, v - U_t) \]
\[ dM_t\mid F_t = \text{binm}(\beta_t \cdot dt, D - M_t). \]

By the Doob-Meyer decomposition, (1) & (2) are ZMM with respect to the increasing family of \( \sigma \)-fields or filtration \( \{F_t, t \geq 0\} \).

\[ \mathcal{U}_t = U_t - \int_0^t \lambda_u \cdot (v - U_u) \, du, \quad (1) \]
\[ \mathcal{M}_t = M_t - \int_0^t \beta_u \cdot (D - M_u) \, du. \quad (2) \]

In the presence of \( \lambda_t \), identifiability problems occur when we want to estimate \( v \) by (1) only. Obviously, the information of \( v \) & \( \lambda_t \) in (1) is confounded. The extra effort by inserting \( D \) errors into the system in the beginning of the testing provides an extra equation to estimate \( v \). However, an identifiability problem can still occur with an extra unknown parameter \( \beta \); thus assumption 3 is needed. Sacks & Chiang [18] and Wolter [21] made a similar assumption of constant proportionality between two intensities in a competing risk model and a capture-recapture experiment respectively.

We now use the ZMM of (1) & (2) to get an estimating equation which can generate a class of estimators for the population size of real errors, \( v \). Define a martingale-difference:

\[ \mathcal{R}^*_t = \int_0^t W_u \cdot (D - M_u) \, du - \theta \cdot (v - U_u) \cdot dM_u. \]

Then \( E\{\mathcal{R}^*_t\mid F_u\} = 0 \).

Notation

\( W_u \) : a measurable function w.r.t. \( F_u \)
\( \mathcal{R}^*_t = \int_0^t \mathcal{R}_u \cdot (D - M_u) \, dU_u - \theta \cdot \int_0^t \mathcal{R}_u \cdot (v - U_u) \, dM_u. \)

\[ \mathcal{R}^*_t = \{\mathcal{R}^*_t, t \geq 0\}, \text{ is a ZMM.} \]

Equate (3b) to zero and evaluate it at time \( t \); then a class of estimators for \( v \) is obtained:

\[ \hat{v}_v = \left[ \int_0^t W_u \cdot (D - M_u) \, dU_u + \theta \cdot \int_0^t W_u \cdot U_u \, dM_u \right]^{\frac{1}{2}} \left[ \int \theta \cdot \int_0^t W_u \, dM_u \right]^{-\frac{1}{2}}. \]

(4)

which depends on the choice of \( W_v \). The conditional variance of \( \mathcal{R}_v \) is:

\[ \text{Var}\{\mathcal{R}^*_v\} = E\left\{ \int_0^t W^2_u \cdot (D - M_u)^2 \, dU_u + \theta^2 \cdot \int_0^t W^2_u \, dM_u \right\}. \]

The terms for variance follow from a standard result, eg, [3]; the covariance is zero by virtue of the orthogonality result of martingales since \( dM_u \) & \( dU_u \) cannot jump simultaneously. Use a result from [17]:

\[ \mathcal{R}^*_v \sqrt{\mathcal{R}^*_v} \sim N(0,1) \text{ as } v, D \to \infty. \]

StdDev\{\hat{v}_v\} = \left[ \int_0^t W^2_u \cdot (\hat{v}_v - U_v)^2 \, dM_u \right]^{\frac{1}{2}} \left[ \int \theta \cdot \int_0^t W_u \, dM_u \right]^{-\frac{1}{2}}.

(5)

Any \( W_v \) which is measurable w.r.t. \( F_v \) can be used. Consider some choices of \( W_v \) which gives explicit expression for \( v \).

**Choice 1:** \( W_v = 1 \)

\[ \hat{v}_1 = \left[ \int_0^t (D - M_u)^2 \, dU_u + \theta \cdot \int_0^t U_u \, dM_u \right]^{\frac{1}{2}} \left[ \int \theta \cdot \int_0^t W_u \, dM_u \right]^{-\frac{1}{2}}. \]

StdDev\{\hat{v}_1\} = \left[ \int_0^t (D - M_u)^2 \, dU_u + \theta^2 \cdot \int_0^t U_u \, dM_u \right]^{\frac{1}{2}} \left[ \int \theta \cdot \int_0^t W_u \, dM_u \right]^{-\frac{1}{2}}.

(6)

(7)

**Choice 2:** Quasi-Score \( W_v \)

An optimal estimating function within such a class of martingale estimating functions in (3) is discussed in [7, 8]. The best choice, quasi-score, is:

\[ W_v = (d\mathcal{R}_u)/(d\langle \mathcal{R} \rangle_u) \]
\[ d\mathcal{R}_u = E\{d\mathcal{R}_u/d\langle \mathcal{R} \rangle_u\} = E\{\theta \cdot dM_u\mid F_v\} = \theta \cdot (D - M_u) \cdot \beta_u \cdot du. \]
\[ d\langle \mathcal{R} \rangle_u = E\{(d\mathcal{R}_u)^2\mid F_v\}. \]

\[ = E\{(D - M_u)^2 \cdot dU_u + \theta^2 \cdot (v - U_u)^2 \cdot dM_u\mid F_v\} \]
\[ = (D - M_u)^2 \cdot (v - U_u) \cdot \lambda_u \cdot du + \theta^2 \cdot (v - U_u)^2 \cdot (D - M_u) \cdot \beta_u \cdot du. \]

Hence the optimal weight corresponding to (3) is:

\[ W^*_v = [(v - U_v) - (D - M_v) + (v - U_v) \cdot \theta]^{-1}. \]
Accordingly, the optimal estimating equation (gives the tightest
s-confidence limits for \( P \)) is, using (9a):

\[
\theta* = \int_0^1 W^\theta_\omega \cdot (D - M^\omega) \, dU_\omega - \theta \cdot \int_0^1 W^\theta_\omega \cdot \nu \cdot U^\omega_\cdot \, dM^\omega.
\]

(9b)

\( \theta* \) is the solution of (9b). An explicit expression is not available;

an iterative procedure is required. From (5),

\[
\text{StdDev} \{ \nu \} = \left[ \int_0^1 (D - M^\omega)^2 / \psi^2_\omega \, dU_\omega + \theta^2 \right]^{1/2} \cdot \left[ \theta \cdot \int_0^1 dM^\omega / \psi^\theta_\omega \right]^{1/2}
\]

\[
\psi^\theta_\omega = \left( \hat{\theta}^* \cdot U^\omega_\cdot \right) \cdot \left( (D - M^\omega) + (\hat{\theta}^* - U^\omega_\cdot) \cdot \theta \right)
\]

3. SIMULATION

A Monte Carlo simulation was performed to evaluate

the performance of \( \hat{\nu}_1 \) from (7) and \( \hat{\theta}^* \) from (9b). Various values

of \( \theta \) have been used. An arbitrary stopping time could be used

for \( \tau \). Here we assume that the stopping time is determined by

the removed proportion of the seeded errors. We investigate

the effects of \( \theta \), the stopping time, and the proportion of seed-

ed errors placed in the system, on the performance of the

estimators.

The simulation results are in two tables. Table 1 lists all

the sets of parameter values in the simulation study. Table 2

lists the results of the 7 trials in table 1. There were 500 repetitions

each trial.

Notation

\[ \text{Av}_2(\text{SD}(\hat{\theta})) \] average StdDev \{ \hat{\nu} \} of the 500 simulated trials

RMSE \{ \hat{\nu} \} \quad [\text{Bias}(\hat{\theta})^2 + \text{SD}(\hat{\theta})^2]^{1/2}

Bias \{ \hat{\nu} \} \quad \hat{\theta} - \nu

Coverage proportion of the estimates between the 95%

s-confidence limits

\( P \) proportion of seeded errors removed: the stopping
criterion.

The statistics computed were:

\[ \text{Av}_1(\theta), \text{SD}(\hat{\theta}), \text{Av}_2(\text{SD}(\hat{\theta})), \text{RMSE}, \text{Coverage}. \]

- Trials 4, 6, 7 examine the effort of \( \nu \) on the performance when

\( D, P \) are kept the same.

- Trials 1 – 3 confirmed that when \( P \) is large: the performance

of \( \hat{\nu}_1 \) & \( \hat{\theta}^* \) improve, and the RMSE & SD(\( \hat{\theta} \)) decrease

appreciably.

- Trials 2, 4, 5 showed that the performance of \( \hat{\nu}_1 \) & \( \hat{\theta}^* \) improve

when \( \theta > 1 \).

- Trials 4, 6, 7 showed that: \( \hat{\theta}^* \) underestimates \( \nu \) when \( \nu \) is

large.

For all trials,

- \( \hat{\nu}_1 \) is satisfactory though RMSE(\( \hat{\nu}_1 \)) > RMSE(\( \hat{\theta}^* \)),

- Coverage(\( \hat{\nu}_1 \)) is approx 80% – 100%, Coverage(\( \hat{\theta}^* \)) is approx

91% – 99%.

- performance of \( \hat{\theta}^* \) is uniformly better than \( \hat{\nu}_1 \),

- \( \hat{\nu}_1 \) is easier to compute and can be used as an initial value

to search for \( \hat{\theta}_1 \).
ACKNOWLEDGMENT

I am pleased to thank the Associate Editor and referees for their comments which improved the presentation of the paper. The work was supported by the Hong Kong Research Grant Council.

APPENDIX

Counting Process, Martingales, and Stochastic Integrals

\( N = \{ N_t; t \in \mathbb{R} \} \) is a counting process if it begins at 0 and increases only by integer-valued jumps, where \( \mathbb{R} = [0, \infty) \).

The observed process can include one or more counting processes, such as the process counting the number that have failed and the process counting the number that have been censored (lost to follow-up). The increasing family of histories \( \mathcal{F} = \{ \mathcal{F}_s; t \leq 3 \} \) is a filtration.

A process \( \mathcal{N} = \{ \mathcal{N}_t; t \in \mathbb{R} \} \) is a martingale (with respect to \( \mathcal{F} \)) if, for all \( t \in \mathbb{R} \):

\[
E[\mathcal{N}_t | \mathcal{F}_s] = \mathcal{N}_s
\]

A consequence of the martingale property is that \( E[\mathcal{N}_t] = E[\mathcal{N}_0] \) for all \( t \in \mathbb{R} \). For a ZMM, \( E[\mathcal{N}_0] = 0 \).

This appendix introduce some properties of martingales which are useful to reliability analysts and gives results for continuous-time martingales in the context of independent continuous failures times. These results are most relevant to applications in this paper. These are not always the most general results. A more rigorous approach is in [3].

Notation

\( N \) number of the failure times falling in \((0, t]\)

\( Y \) number still at risk just prior to time \( s \).

Let \( N_0 = 0 \). Partition the interval \((0, t]\) into many very small increments. Then write \( N_t \) as \( \int_0^t dN_s \). We are concerned with the conditional distribution of \( dN_t \), given \( \mathcal{F}_s \). This conditional distribution is binomial:

\[
dN_t | \mathcal{F}_s \overset{d}{=} \text{bim}(\alpha_s \cdot ds, Y_s)
\]

Most time increments contain 0 failures and a small fraction of the increments contain 1 failure; the probability of an increment containing more than 1 failure is negligible. As,

\[
E[dN_t | \mathcal{F}_s] = Y_s \cdot \alpha_s \cdot ds
\]

\( d\mathcal{M}_t = dN_t - Y_s \cdot \alpha_s \cdot ds \)

can be treated as a martingale-difference. It follows that the process,

\[ \mathcal{M} = \{ \mathcal{M}_t; t \in \mathbb{R} \}, \]

specified by,

\[
\mathcal{M}_t = \int_0^t d\mathcal{M}_s = N_t - \int_0^t Y_s \cdot \alpha_s \cdot ds
\]

is a ZMM.

Let \( W = \{ W_t; t \in \mathbb{R} \} \) be any process such that \( W_s \) is determined by \( \mathcal{F}_s \) for each \( s \geq 0 \), then each,

\[
W_s \cdot d\mathcal{M}_t
\]

is also a martingale-difference. It follows that \( \mathcal{M} \) specified by,

\[
\mathcal{M}_t = \int_0^t W_s \cdot d\mathcal{M}_s = \int_0^t W_s \cdot dN_s = \int_0^t W_s \cdot Y_s \cdot \alpha_s \cdot ds
\]

is a ZMM. This property is very useful for generating estimating equations. In a similar manner, using various choices of \( W \), one obtains estimating equations for other quantities.

The next step is to associate a standard deviation with such an estimate.

\[
\text{Var} \{ W_s \cdot d\mathcal{M}_t \} = \text{E} \{ \text{Var} \{ W_s \cdot d\mathcal{M}_t | \mathcal{F}_s \} \} + \text{Var} \{ E \{ W_s \cdot d\mathcal{M}_t | \mathcal{F}_s \} \}
\]

\[
= \text{E} \{ W_s^2 \cdot \text{Var} \{ d\mathcal{M}_t | \mathcal{F}_s \} \} = \text{E} \{ W_s^2 \cdot \text{Var} \{ dN_s | \mathcal{F}_s \} \}
\]

\[
= \text{E} \{ W_s^2 \cdot Y_s \cdot \alpha_s \cdot ds \cdot (1 - \alpha_s \cdot ds) \} \approx \text{E} \{ W_s^2 \cdot Y_s \cdot \alpha_s \cdot ds \}
\]

Hence,

\[
\text{Var} \{ \mathcal{M}_t \} = E \left[ \int_0^t W_s^2 \cdot Y_s \cdot \alpha_s \cdot ds \right] = E \left[ \int_0^t W_s^2 \cdot dN_s \right]
\]

A central limit theorem also applies, indicating that inference can be made by using the \( s \)-normal distribution when the amount of data is sufficiently large [3].

REFERENCES

OPINION ITEMS
OPINION ITEMS
OPINION ITEMS

Do you want to sound off on a reliability topic?
Then DO IT!
Write a “Letter to the Editor”.
The Editorial Staff welcomes your comments on material published in this Transactions (eg, editorials and technical papers) as well as on other topics of interest to Transactions readers. If you comment on the work of someone else, that person will be invited to present a reply or clarification. We will publish controversy, but we don’t want to publish misunderstandings.

We encourage interaction between you and the original author to minimize misunderstandings and clarify the areas of disagreement.

Material judged unsuitable as a refereed paper can sometimes be abbreviated and published as opinion. Opinions will be edited as to length and propriety — at the sole discretion of the Editorial staff.

IEEE Log Number 94-11254