11-64-CR
254808
14P

DEPARTMENT OF MATHEMATICS AND STATISTICS
COLLEGE OF SCIENCES
OLD DOMINION UNIVERSITY
NORFOLK, VIRGINIA 23529

ESTIMATION IN A DISCRETE TAIL RATE FAMILY OF RECAPTURE SAMPLING MODELS

By

Rajan Gupta, Graduate Research Assistant

and

Larry D. Lee, Principal Investigator

Final Report
For the period ended February 28, 1990

Prepared for
National Aeronautics and Space Administration
Langley Research Center
Hampton, VA 23665

Under
Research Grant NAG-1-835
Dr. Dave E. Eckhardt Jr., Technical Monitor
ISD-Systems Architecture Branch

Submitted by the
Old Dominion University Research Foundation
P.O. Box 6369
Norfolk, Virginia 23508-0369

February 1990

(NASA-CR-180452) ESTIMATION IN A DISCRETE TAIL RATE FAMILY OF RECAPTURE SAMPLING MODELS Final Report, period ending 28 Feb. 1990 (Old Dominion Univ.) 14 p CSCL 12A

N90-18197
356791

Unclas
G3/64 0264808

ESTIMATION IN A DISCRETE TAIL RATE FAMILY OF RECAPTURE SAMPLING MODELS

Rajan Gupta and Larry Lee

Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA 23529-0077

ABSTRACT

In the context of a recapture sampling design for debugging experiments we consider the problem of estimating the error or hitting rate of the faults remaining in a system. Moment estimators are derived for a family of models in which the rate parameters are assumed proportional to the tail probabilities of a discrete distribution on the positive integers. The estimators are shown to be asymptotically normal and fully efficient. Their fixed sample properties are compared, through simulation, with those of the conditional maximum likelihood estimators.

Key words and phrases: software reliability; asymptotic efficiency; conditional likelihood; average information; interval truncated sampling.

1. INTRODUCTION

Nayak (1988) recently proposed a recapture sampling design for estimating the number of errors or faults remaining in a system. As is common in debugging experiments, a system is tested for a time period of length τ, the failure (i.e., error detection) times $T_1, T_2, \ldots, T_r$ are observed, and repair takes place immediately after a fault produces an error. By using standard error detection techniques (e.g., Avizienis and Chen, 1977) the hitting frequency $M_i = M_i(T_i, \tau)$ of the fault detected at time T_i is observed as the number of times the region (i.e., path in software) containing the fault is accessed during the interval (T_i, τ). Nayak's (1988) discussion concerns the Jelinski-Moranda (1972) model $\lambda_i = (\nu - i + 1)\phi$, $\phi > 0$, $i = 1, 2, \ldots, \nu$ where ν, a parameter, is the initial number of faults in a system.

The purpose of the present paper is to study estimation procedures related to the following model. The spacings $Y_i = T_i - T_{i-1}$ $(T_0 \equiv 0)$, $i = 1, 2, \ldots$ are assumed independent and exponentially distributed with rate parameters λ_i given by

$$\lambda_i = \alpha\, \bar{G}(i-1, \phi),\ \alpha > 0, \quad \xi_i = \alpha\, g(i, \phi) \tag{1}$$

That is, λ_i is the rate of encountering the remaining faults after $i-1$ faults have been removed and $\xi_i = \lambda_i - \lambda_{i+1}$ are the hitting rates of the first, second, etc., detected faults. In (1), $G(x, \phi) = 1 - \bar{G}(x, \phi)$, is the distribution function of a discrete positive random variable and $g(x, \phi)$ is the density function or probability mass function of $G(x, \phi)$. The quantity ξ_i can be interpreted as the amount by which λ_i decreases when repairing the fault detected at time T_i. Counts $\{M_i(t)\}$ of repeated error occurrences are assumed to be independent homogeneous Poisson processes with rate parameters ξ_i.

In this context the J-M model is given by a discrete uniform distribution with mass at $1, 2, \ldots, \nu$. This model, however, assumes that faults have a common rate $\xi_i = \phi$ whereas experimental investigations (e.g., Nagel, Scholz, and Skrivan,

1984) indicate that faults may have different hitting rates. The log linear rate model $\lambda_i = \alpha e^{-\beta(i-1)}$ (Cox and Lewis 1966) corresponds to a geometric distribution and describes the case $\xi_1 > \xi_2 > \ldots$ in which faults having the highest hitting rates are detected early in the debugging process. Other models that seem to be related to (1) are those of Sandland and Cormack (1984) and Miller (1986). For our purpose it suffices to take $g(i,\phi)$ to be the discrete exponential family of densities $g(i,\phi) = \exp[a_i\phi - \psi(\phi) + b_i]$. These models yield a sufficient statistic of smaller dimension than obtained in general, although for most families the likelihood function (Section 2) does not have exponential family structure.

The main problem we consider is that of estimating the error (or hitting) rate λ_{r+1} for a system in its final state; i.e., a system for which $R = r$ faults have been removed. Moment estimators of (α, ϕ) are presented in Section 3. Their bias and asymptotic variances are compared with those of the maximum likelihood estimators in Section 4. In Section 3 we show that functions of the form $r^{-1/2}\ln(\tilde{\lambda}_{r+1}/\lambda_{r+1})$ have a limiting ($r \to \infty$) normal distribution under various models. The conditional likelihood function given in Section 2 defines the setting of our discussion.

2. A CONDITIONAL LIKELIHOOD

We assume that a system is tested until no errors are detected for a time period of length s. Data is obtained through interval truncated sampling by which we observe $T_1, T_2, \ldots, T_R$ and $R = r$ providing $Y_i \le s, i = 1, 2, \ldots, r$ and $Y_{r+1} > s$.

With $Y_1, Y_2, \ldots,$ being independent exponential random variables, the conditional density of $Y_1, Y_2, \ldots, Y_R$ given $R = r$, is

$$f(y_1, y_2, \ldots, y_r \mid r) = \prod_{i=1}^{r} \lambda_i \exp(-\lambda_i y_i)[1 - \exp(-\lambda_i s)]^{-1}, \quad 0 < y_i < s; \; i = 1, 2, \ldots, r \qquad (2)$$

The total test time $\tau = \sum_1^R Y_i + s$ is random while in Nayak's (1988) discussion, τ is a fixed quantity.

The full data vector can be represented in terms of the vector quantities Z_k defined by

$$Z_1 = (Y_1), \quad Z_k = (Y_k, M_{1k}, M_{2k}, \ldots, M_{(k-1)k}) \quad (k = 2, 3, ..., r+1) \tag{3}$$

Here $M_{ik}, i < k$, is the number of times the system encounters the ith detected fault during the interval $(T_{k-1}, T_k]$. The last interval $(T_r, \tau]$ has fixed length s while the remaining intervals $(T_{k-1}, T_k]$, $k \leq r$, have random length Y_k. For notational convenience, we let $Y_{r+1} = s$.

Our earlier assumption that $\{M_i(t)\}$ are independent homogeneous Poisson processes together with $Y_1, Y_2, \ldots, Y_r$ being independent implies that $Z_1, Z_2, \ldots$, Z_{r+1} are conditionally, given $R = r$, independent with densities

$$\begin{aligned} &g_k(y_k; m_{1k}, m_{2k}, \ldots, m_{(k-1)k}) \\ &= \lambda_k e^{-\lambda_k y_k}(1 - e^{-\lambda_k s})^{-1} \prod_{i=1}^{k-1} (\xi_i y_k)^{m_{ik}} e^{-\xi_i y_k} / m_{ik}! \quad (k = 1, 2, \ldots, r) \\ &= \prod_{i=1}^{r} (\xi_i s)^{m_{i(r+1)}} e^{-\xi_i s} / m_{i(r+1)}! \quad (k = r+1) \end{aligned} \tag{4}$$

Substituting $\lambda_i = \alpha \bar{G}(i-1, \phi)$ and $\xi_i = \alpha g(i, \phi)$ in (4), the log likelihood is $l_c = \sum_1^{r+1} l_k$ where

$$l_k = C_k(\alpha, \phi) - \alpha y_k + \ln \alpha \sum_{i=1}^{k-1} m_{ik} + \sum_{i=1}^{k-1} m_{ik} \ln g(i, \phi) + C \tag{5}$$

Here $Y_{r+1} = s$, C does not depend upon (α, ϕ), and

$$\begin{aligned} C_k(\alpha, \phi) &= \ln[\lambda_k (1 - e^{-\lambda_k s})^{-1}], \quad k = 1, 2, \ldots, r \\ &= \lambda_{r+1} s, \quad k = r+1 \end{aligned}$$

Let $V_k' = (V_{1k}, V_{2k}, V_{3k})$, $k = 1, 2, \ldots, r+1$ (prime denotes vector transpose) be defined by

$$V_{1k} = Y_k, \quad V_{2k} = \sum_{i=1}^{k-1} M_{ik}, \quad V_{3k} = \sum_{i=1}^{k-1} a_i M_{ik} \tag{6}$$

where a_i are constants.

The following moments are needed to obtain the average information matrix and also in Section 3, to study the asymptotic distribution of $S_r = \sum_{k=1}^{r} V_k$.

THEOREM 1. *Let the means, variances and covariances of the elements of V_k be denoted by $\mu_{ik} = E(V_{ik})\,; i = 1,2,3$ and $\sigma_{ijk} = Cov(V_{ik}, V_{jk})$; $i,j = 1,2,3$, with $\gamma_{1k} = \sum_{i=1}^{k-1} a_i \xi_i$ and $\gamma_{2k} = \sum_{i=1}^{k-1} a_i^2 \xi_i$.*

Then

$$\mu_{1k} = \lambda_k^{-1} - s(e^{\lambda_k s} - 1)^{-1}$$

$$\mu_{2k} = \mu_{1k}(\lambda_1 - \lambda_k)$$

$$\mu_{3k} = \mu_{1k}\gamma_{1k}$$

$$\sigma_{11k} = \lambda_k^{-2} - s^2 e^{\lambda_k s}(e^{\lambda_k s} - 1)^{-2}$$

$$\sigma_{12k} = \sigma_{11k}(\lambda_1 - \lambda_k)$$

$$\sigma_{13k} = \sigma_{11k}\gamma_{1k}$$

$$\sigma_{22k} = \mu_{1k}(\lambda_1 - \lambda_k) + \sigma_{11k}(\lambda_1 - \lambda_k)^2$$

$$\sigma_{23k} = \mu_{1k}\gamma_{1k} + \sigma_{11k}(\lambda_1 - \lambda_k)\gamma_{1k}$$

$$\sigma_{33k} = \mu_{1k}\gamma_{2k} + \sigma_{11k}\gamma_{1k}^2$$

These moments can be obtained by noting that Y_k has an exponential distribution truncated over the interval (0,s) and that $\{M_{ik}\}$ are conditionally, given Y_k, independent Poisson random variable with means $\xi_i Y_k$. Since V_{3k} is a linear function of $M_{ik}, i < k$, these moment are similar to those given in [3].

By taking derivatives and expectations, the Fisher information matrix $A_k = (a_{ijk})$, based on l_k, can be obtained as follows:

$$a_{11k} = -E(\frac{\partial^2 l_k}{\partial \alpha^2}) = \alpha^{-1}\mu_{1k} + \alpha^{-2}(\lambda_k^2 \sigma_{11k} - \lambda_k \mu_{1k})$$

$$a_{12k} = -E(\frac{\partial^2 l_k}{\partial \alpha \partial \phi}) = \gamma_{1k}'(\mu_{1k} - \lambda_k \sigma_{11k})$$

$$a_{22k} = -E(\frac{\partial^2 l_k}{\partial \phi^2}) = \alpha \mu_{1k}\gamma_{2k}' + \alpha^2 \sigma_{11k}(\gamma_{1k}')^2$$

$$\gamma'_{1k} = \sum_{i=1}^{k-1} \frac{\partial}{\partial\phi} \ln g(i,\phi).g(i,\phi) \qquad \gamma'_{2k} = \sum_{i=1}^{k-1} [\frac{\partial}{\partial\phi} \ln g(i,\phi)]^2.g(i,\phi)$$

Since Y_k converges (as $k \to \infty$) in distribution to a uniform distribution on the interval $(0,s)$ the moments of Y_k converge to the corresponding moments of the limiting uniform distribution (Serfling, 1980, p.14). We thus have $\lim \mu_{1k} = s/2$, $\lim \sigma_{11k} = s^2/12$, and $\lim \lambda_k = 0$ as $k \to \infty$ Assuming that $g(i,\phi)$ is a regular family with support not depending upon ϕ, the limits $\gamma_i = \lim \gamma'_{ik}$, $i = 1,2$ are given by

$$\gamma_1 = \sum_{i=1}^{\infty} \frac{\partial}{\partial\phi} \ln g(i,\phi).g(i,\phi) \quad \gamma_2 = \sum_{i=1}^{\infty} [\frac{\partial}{\partial\phi} \ln g(i,\phi)]^2.g(i,\phi)$$

where γ_2 is the Fisher information about the parameter ϕ based on a single observation from $g(i,\phi)$. Thus $\gamma_1 = 0$ and the limiting average information matrix $A = \lim(1/r)(A_1 + A_2 + \ldots + A_{r+1})$ is $A = (a_{ij})$ where $a_{11} = s/2$, $a_{12} = 0$, and $a_{22} = \alpha\gamma_2 s/2$.

3. ESTIMATION

We now consider exponential family rate models given by

$$\xi_i = \alpha \exp[\phi a(i) - \psi(\phi) + b(i)], i = 1,2,\ldots. \tag{7}$$

where ϕ varies over the natural parameter set $\{\phi : \sum_{i=1}^{\infty} \exp[\phi a(i) + b(i)] < \infty\}$. This family includes the Poisson ($\xi_i = \alpha\theta^{i-1}e^{-\theta}/(i-1)!, \theta > 0, i = 1,2,\ldots.$) and log linear model as well as other models.

Let V_k be defined as in (6) where a_i is the coefficient of ϕ in (7). In reference to $l_c = \sum_1^{r+1} l_k$ defined by (5), we have the following:

(i) $V_1, V_2, \ldots, V_{r+1}$ are independent.

(ii) $S_r = \sum_1^{r+1} V_k$ is a sufficient statistic for the family defined by l_c.

(iii) $S'_r = (S_{1r}, S_{2r}, S_{3r})$ is given by $S_{1r} = r$, $S_{2r} = \sum_1^r M_i$, $S_{3r} = \sum_1^r a_i M_i$

THEOREM 2. *Under (7) where $a_i \geq 0$ is nondecreasing in $i = 1, 2, \ldots$, $(1/r)S_r$ has a limiting (as $r \to \infty$) normal distribution with mean vector $\mu' = (\mu_1, \mu_2, \mu_3)$ and covariance matrix $(1/r)C$ given by*

$$\begin{array}{lll} \mu_1 = s/2 & \mu_2 = \alpha s/2 & \mu_3 = \alpha s \psi' \\ c_{11} = s^2/12 & c_{12} = \alpha s^2/12 & c_{13} = (\alpha s^2/12)\psi' \\ c_{22} = \alpha s/2 + \alpha^2 s^2/12 & c_{23} = c_{22}\psi' & c_{33} = (\alpha s/2)\psi'' + c_{22}(\psi')^2 \end{array}$$

The proof is given in Section 5.

Note that $\alpha = g_1(\mu_1, \mu_2, \mu_3)$ and $\psi'(\phi) = g_2(\mu_1, \mu_2, \mu_3)$ where $g_1(z_1, z_2, z_3) = z_2/z_1$, $g_2(z_1, z_2, z_3) = z_3/z_2$, Applying the δ-method, the estimates $(\tilde{\alpha}, \tilde{\phi})$ given by

$$\tilde{\alpha} = \sum_{i=1}^{r} M_i/r, \quad \psi'(\tilde{\phi}) = \sum_{i=1}^{r} a_i M_i / \sum_{i=1}^{r} M_i$$

have a limiting normal distribution with mean vector (α, ϕ), and are asymptotically independent with variances $\sigma_{\tilde{\alpha}}{}^2 = 2\alpha/rs$, $\sigma_{\tilde{\phi}}{}^2 = 2[r\alpha s\psi''(\phi)]^{-1}$.

In estimating λ_{r+1} by $\tilde{\lambda}_{r+1} = \tilde{\alpha}\bar{G}(r; \tilde{\phi})$ we must account for the fact that r increases as $r^{1/2}(\tilde{\alpha} - \alpha)$ and $r^{1/2}(\tilde{\phi} - \phi)$ converge to their limiting distributions. For the log linear rate model $\lambda_i = \alpha e^{-\beta(i-1)}$ with $\phi = -\beta, \beta > 0$, we have $r^{-1/2}\ln(\tilde{\lambda}_{r+1}/\lambda_{r+1}) = r^{1/2}(\tilde{\phi} - \phi)$ so that by Theorem 2, $r^{-1/2}\ln(\tilde{\lambda}_{r+1}/\lambda_{r+1})$ has a limiting $(r \to \infty)$ normal distribution with mean zero and variance $2[\alpha s\psi''(\phi)]^{-1} = 2e^{\phi}(e^{-\phi} - 1)^2(\alpha s)^{-1}$.

To deal with the other models in Table 1, we apply Taylor's formula to $H(\phi) = -\ln \bar{G}(r; \phi)$ and obtain

$$\begin{aligned} r^{-1/2}\ln(\tilde{\lambda}_{r+1}/\lambda_{r+1}) = r^{-1/2}(\ln\tilde{\alpha} - \ln\alpha) &- r^{-1/2}(\tilde{\phi} - \phi)H'(\phi*) \\ &- (1/2)r(\tilde{\phi} - \phi)^2 r^{-3/2}H''(\phi*) \quad (8) \end{aligned}$$

where $|\phi^* - \phi| < |\tilde{\phi} - \phi|$. The first term of (8) converges in probability to zero. Under the Poisson and logarithmic series models, $r^{-1}|H'(\phi*)|$ converges to 1, and

$r^{-3/2}H''(\phi^*)$ converges in probability to zero. Thus for all of the models of Table 1, the limiting distribution of $r^{-1/2}\ln(\tilde{\lambda}_{r+1}/\lambda_{r+1})$is identical to that of $r^{1/2}(\tilde{\phi}-\phi)$ and is a normal distribution with mean zero and variance $2[\alpha s\psi''(\phi)]^{-1}$.

4. Efficiency and Bias

Since $\sigma_{\tilde{\alpha}}{}^2 = 2\alpha/(rs)$ and $\sigma_{\tilde{\phi}}{}^2 = 2[r\alpha s\psi''(\phi)]^{-1}$ where $\psi''(\phi) = \gamma_2$ is defined at the end of Section 2, it follows that $\tilde{\alpha}$ and $\tilde{\phi}$ are asymptotically fully efficient.

To study the fixed sample properties of $\tilde{\alpha}$ and $\tilde{\phi}$, we simulated their values for the Poisson rate model under the conditional likelihood defined in (6). This was done by generating 200 replicates of $(T_1, T_2, \ldots, T_r, M_1, M_2, \ldots, M_r)$ for the values of r shown in Table 2. In addition the conditional maximum likelihood estimates $\hat{\alpha}_c$ and $\hat{\phi}_c$ were calculated for each replicate by maximizing $l_c = \sum_1^{r+1} l_k$ where l_k is defined in (5) .

r		$\hat{\alpha}_c$	$\hat{\phi}_c$	$\tilde{\alpha}$	$\tilde{\phi}$
15	Bias	-.012464	.020852	-.001241	-.050126
	MSE	.000155	.000435	.000001	.002512
20	Bias	-.010878	.004578	-.002542	-.025457
	MSE	.000018	.000021	.000006	.000648
25	Bias	-.008613	-.028581	-.001379	-.046221
	MSE	.000074	.000817	.000002	.002136
30	Bias	-.007036	.012544	-.001484	.022037
	MSE	.000050	.000157	.000002	.000485

Table 2. Bias and mean square error (MSE) of the conditional maximum likelihood estimators $\hat{\alpha}_c$ and $\hat{\phi}_c$ and moment estimators $\tilde{\alpha}$ and $\tilde{\phi}$ based on 200 simulations with $\alpha = .10$ and $\phi = -2.00$.

Table 2 shows the bias and mean square error (MSE) for $\tilde{\alpha}$ and $\tilde{\phi}$ and also the bias and MSE for $\hat{\alpha}_c$ and $\hat{\phi}_c$. Although the conditional MLE $\hat{\phi}_c$ has smaller bias than $\tilde{\phi}$, the moment estimator $\tilde{\alpha}$ seems to generally have smaller bias than $\hat{\alpha}_c$.

5. PROOF OF THEOREM 2

To prove Theorem 2, note that the elements of μ and C are given by $\mu_i = \lim(1/r)\sum_{k=1}^{r+1}\mu_{ik}$, $i = 1,2,3$ and $c_{ij} = \lim(1/r)\sum_{k=1}^{r+1}\sigma_{ijk}$ as r tends to infinity, where the terms in these sums are the moments given in Theorem 1. Since μ_{ik} and σ_{ijk} converge to finite limits as k tends to infinity, we have $\mu_i = \lim \mu_{ik}$ and $c_{ij} = \lim \sigma_{ijk}$. Thus the calculations are similar to those discussed at the end of Section 2. The remainder of the proof requires showing (Serfling, 1980, p.30) that

$$\lim(1/r)\sum_{k=1}^{r+1} h_{kr} = 0 \tag{9}$$

where

$$\begin{aligned} h_{kr} &= E[U_k I(U_k > \epsilon^2 r)] \\ U_k &= \sum_{i=1}^{3}(V_{ik} - \mu_{ik})^2 \end{aligned} \tag{10}$$

and $I(.)$ is the indicator function. Since $h_{kr} \leq (\epsilon^2 r)^{-1}E(U_k^2)$, the limit in (9) can be established by examining the fourth moments of the $V_{ik}, i = 1,2,3$.

To obtain bounds for these moments, we replace Z_k by

$$Z_k' = (Y_k, N_k,\ X_{1k}, X_{2k}, \ldots, X_{Nk}) \quad (k = 1,2,\ldots,r+1)$$

where $N_k = \sum_{i=1}^{k-1} M_{ik}$ and X_{jk} takes the value $X_{jk} = i$ if the jth event occurring in the interval (T_{k-1}, T_k) corresponds to the occurrence of the ith detected fault, $i = 1,2,\ldots,k-1$. Given that $N_k = n$, $X_{1k}, X_{2k}, \ldots, X_{nk}$ are i.i.d with truncated density $g(i\,;\phi)/G(k-1\,;\phi), i = 1,2,\ldots,k-1$.

In terms of Z_k', the vectors V_k defined in (6) can be written

$$V_{1k} = Y_k, \quad V_{2k} = N_k \quad V_{3k} = \sum_{j=1}^{Nk} a(X_{jk}) \tag{11}$$

where $a(i) = a_i$.

Since the distribution of N_k is conditionally, given Y_k, Poisson with mean $(\lambda_1 - \lambda_k)Y_k$ and since $Y_k \leq s$, it follows that N_k is stochastically smaller than the Poisson random variable N that has mean $\lambda_1 s = \alpha s$. Thus for any positive integer p we have $E(Y_k^p) \leq s^p$ and $E(N_k{}^p) \leq E(N^p) < \infty$.

For any nonnegative quantities $w_1, w_2, \ldots, w_n$ and positive integer p we have $(\sum_{i=1}^n w_i)^p \leq n^p \max(w_i^p) \leq n^p \sum_{i=1}^n w_i{}^p$ and thus

$$(\sum_{i=1}^n w_i)^p \leq n^p \sum_{i=1}^n w_i{}^p \tag{12}$$

By applying (12) to the form of V_{3k} given in (11), we obtain

$$V_{3k}{}^p \leq N_k{}^p \sum_{j=1}^{N_k} [a(X_{jk})]^p$$

.

$$\begin{aligned} E(V_{3k}{}^p) &\leq E(N_k{}^{p+1}) E\{[a(X_{1k})]^p\} \\ &\leq E(N^{p+1}) E\{[a(X)]^p\} \end{aligned}$$

where N has a Poisson distribution with mean αs and X has the density $g(i;\phi)$. All positive moments of $a(x)$ exist for the family of densities in (7). In summary we have $E(V_{ik}^p) \leq B_{ip}$ where $B_{ip}, i = 1,2,3$ do not depend on k.

To complete the proof of Theorem 2, we again use (12) to obtain $U_k{}^2 \leq 9 \sum_{i=1}^3$ $(V_{ik} - \mu_{ik})^4$. Since $(V_{ik} - \mu_{ik})^4 \leq (V_{ik} + \mu_{ik})^4$, the binomial expansion can be used to show that $E(U_k^2) \leq B$ where B is finite and does not depend on k. Thus the limit in (9) is zero, which proves Theorem 2.

6. Final Remarks

Software testing counters (Huang, 1977) will tend to over count the number of times a fault produces an error in the output. An alternative method, which will accurately give the fault hitting frequencies, has been described in the literature on multiversion programming (Avizienis and chen, 1977). To describe this method in the context of error recapture experiments, let $P_1, P_2, \ldots$ denote successive versions of the original program P_0, where P_i is the result of correcting the fault detected in P_{i-1} at time T_i. A copy of P_{i-1} is made before correcting this fault and all of the versions $(P_0, P_1, \ldots, P_i)$ are run on the same input series during the interval (T_i, τ). To determine the hitting frequency of the fault detected at time T_i, the outputs of P_{i-1} are compared with those of P_i. Any difference in the outputs is due to the fault that resides in P_{i-1} which has been corrected in P_i. Similarly, comparing the outputs of all pairs (P_{i-1}, P_i), $i = 1, 2, \ldots, r$ will yield the total set of fault hitting frequencies.

ACKNOWLEDGMENT

This research was supported by the National Aeronautics and Space Administration, under contract NAG-1-835.

Model	Estimator	Variance
Poisson,		
$\xi_i = \alpha \exp[\phi(i-1) - \psi(\phi)]/(i-1)!$,	$\tilde{\alpha} = \sum_{i=1}^{r} M_i/r$	$2\alpha(rs)^{-1}$
$\alpha > 0,\ -\infty < \phi < \infty,\ \psi(\phi) = e^{\phi}$	$\tilde{\phi} = \ln(\sum_{i=1}^{r}(i-1)M_i / \sum_{i=1}^{r} M_i)$	$2[\alpha r s \psi'']^{-1}$
$\lambda_{r+1} = \alpha \int_0^{e^{\phi}} u^{r-1} e^{-u} du/(r-1)!$		
Geometric,		
$\xi_i = \alpha \exp[\phi(i-1) - \psi(\phi)]$	$\tilde{\alpha} = \sum_{i=1}^{r} M_i/r$	$2\alpha(rs)^{-1}$
$\alpha > 0,\ \phi < 0,\ \psi(\phi) = -\ln(1-e^{\phi})$	$\tilde{\phi} = -\ln(1 + \sum_{i=1}^{r} M_i / \sum_{i=1}^{r}(i-1)M_i)$	$2[\alpha r s \psi'']^{-1}$
$\lambda_{r+1} = \alpha e^{\phi r}$		
Logarithmic Series,		
$\xi_i = \alpha \exp[\phi i - \psi(\phi)]/i$,		$2\alpha(rs)^{-1}$
$\alpha > 0,\ \phi < 0$	$\tilde{\alpha} = \sum_{i=1}^{r} M_i/r$	$2[\alpha r s \psi'']^{-1}$
$\psi(\phi) = -\ln[-\ln(1-e^{\phi})]$	$(1-e^{-\tilde{\phi}})\ln(1-e^{-\tilde{\phi}})$	
$\lambda_{r+1} = \alpha e^{-\psi(\phi)} \int_0^{e^{\phi}} u^{r-1} e^{-u} du$	$= \sum_{i=1}^{r} M_i / \sum_{i=1}^{r} i M_i$	

Table 1. Moment estimators of (α, ϕ) for the Poisson, Geometric and Logarithmic series rate models.

REFERENCES

[1] Avizienis and Chen (1977. On the implementation of N-version programming for software fault tolerance during program execution. Proc. COMPSAC 77, 149-55.

[2] Cox, D.R. and Lewis, P.A.W. (1966). The Statistical Analysis of Series of Events Methuen, London.

[3] Gupta, R. & Lee, L. (1989), Estimating Fault Hitting rates by Recapture Sampling. Communication in Statistics, 18 (6) 2139-53.

[4] Huang, J.C. (1977), Error detection through program testing. In Current Trends in Programming Methodology, 2, Ed. R.T. Yeh, pp. 16-43. Prentice Hall, Engelwood Cliffs, N.J.

[5] Miller (1986), Exponential order statistics models of software reliability growth. IEEE Trans. Software Engrg., SE-12, 12-24.

[6] Jelinski, Z. and Moranda, P.M. (1972). Software reliability research. In *Statistical* Computer Performance Evaluation, Ed. W. Freiberger, pp. 465-84. Academic Press, New York.

[7] Nagel, P.M., Scholz, F.W., and Skrivan, J.A. (1984). Software reliability; additional investigations into modeling with replicated experiments. NASA Langley Research Center, Hampton, VA, NASA Contractor Report 172378, June.

[8] Nayak, T.K. (1988). Estimating population size by recapture sampling. *Biometrika*, 75, 113-120.

[9] Sandland, R.L. and Cormack R.M. (1984), Statistical inference for Poisson and multinomial models for capture-recapture experiments. Biometrika 71, 27-33.

[10] Serfling, R.J. (1980).Approximation Theorems of Mathematical Statistics. John Wiley & Sons, New York.