Statistica Sinica **12**(2002), 1001-1013

A FRAILTY MODEL FOR DETECTING NUMBER OF FAULTS IN A SYSTEM

Yan Wang, Paul S. F. Yip and Y. Hayakawa

University of South Australia, University of Hong Kong and University of Wellington

Abstract: A frailty model for failure data is proposed to estimate the total number of faults in a system. The Littlewood model and Jelinski-Moranda are the two particular cases of the proposed formulation. The two-stage estimating procedure, a conditional likelihood and a Horvitz-Thompson estimator, is found to be efficient. Simulation studies are given to assess the performance of the estimator. Two examples are also presented.

Key words and phrases: Capture-recapture, conditional likelihood, fault detection, Littlewood, Jelinski-Moranda, removal experiment.

1. Introduction

We consider the problem of estimating the number of faults in a system (van Pul (1993); Yip, Xi, Fong and Hayakawa (1999)). Let ν denote the unknown number of faults initially present in a system, let $N_i(t)$ be the counting process for fault i and ϕ_i be its occurrence rate as a stochastic variable, i.e., the intensity process for the counting process depends on an unobservable random variable. Let $Y_i(t)$ indicate, by the value of 1 or 0, whether fault i has been removed before t, thus $Y_i(t)$ is an observable non-negative predictable process. Let τ be the duration of the study and $\mathcal{F}_{t-}$ denote the smallest σ-algebra generated by $\{N_i(s), Y_i(s), 0 \leq s < t\}$.

Suppose that ϕ_i has a gamma distribution with parameters α and β, denoted $G(\alpha, \beta)$, with density function $f(\phi_i|\alpha,\beta) = \{e^{-\beta\phi_i}\phi_i^{\alpha-1}\beta^\alpha\}/\Gamma(\alpha) \quad (\phi_i > 0)$. For a given ϕ_i, the inter-failure times are independent and exponentially distributed with rate ϕ_i. Let $f(\phi_i|\mathcal{F}_0)$ and $L(\phi_i|\mathcal{F}_{t-})$ denote the prior for ϕ_i and the likelihood function for ϕ_i given $\mathcal{F}_{t-}$, respectively. It can be shown that the posterior for ϕ_i, denoted by π, is

$$\pi(\phi_i|\mathcal{F}_{t-}) \propto f(\phi_i|\mathcal{F}_0)L(\phi_i|\mathcal{F}_{t-}) = \frac{\beta^\alpha}{\Gamma(\alpha)}\phi_i^{\alpha-1}e^{-\beta\phi_i}\phi_i^{N_i(t-)}e^{-t\phi_i} \propto \phi_i^{\alpha+N_i(t-)-1}e^{-(\beta+t)\phi_i}.$$

Thus the posterior distribution for ϕ_i is gamma distributed: $\phi_i|\mathcal{F}_{t-} \sim G(\alpha + N_i(t-), \beta + t)$. The expectation of a $\mathrm{G}(a, b)$ variable is a/b. Conditional on

$\mathcal{F}_{t-}$ the expectation of ϕ_i is $\{\alpha + N_i(t-)\}/(\beta + t)$. Hence, using the innovation theorem (Aalen (1978)), the compensator of the counting process $N_i(t)$ is given by

$$\lambda_i(t) = Y_i(t)\Big\{\frac{\alpha + N_i(t-)}{\beta + t}\Big\}\ . \tag{1}$$

The intensity function $\lambda_i(t)$ depends on time and the number of times of the ith individual fault has been detected. The formulation in (1) includes both the removal and recapture sampling methods. For the case of removal, in which $N_i(t-)$ equals zero since the ith fault is removed from the system after being detected. The model becomes the *Littlewood model* (Littlewood (1980); Andersen, Borgan, Gill and Keiding (1993)). In addition, (1) also includes the recapture experiment in which a counter is inserted at the location after the fault is detected and the counter registers the number of revisit of a particular fault without causing the system failure (Nayak (1988)). The revisit information has been shown to be important in determining the performance in estimating the number of faults in a system (Lloyd, Yip and Chan (1999)). Furthermore, the proposed formulation allows random removals which could happen in a recapturing process.

If we reparameterise the intensity in (1) by letting $\epsilon = 1/\alpha$, $\omega = \beta/\alpha$, then we have

$$\lambda_i(t) = Y_i(t)\Big\{\frac{1 + \epsilon N_i(t-)}{\omega + \epsilon t}\Big\}\ . \tag{2}$$

This extension allows the case of $\epsilon \leq 0$ to have meaningful interpretation (Nielsen, Gill, Andersen and Sørensen (1992)). Note that when $\epsilon = 0$ (i.e., $\alpha = \infty$) for a removal experiment, the model reduces to Jelinski-Moranda model in software reliability studies (Jelinski and Moranda (1972)).

The full likelihood function is given by

$$L(\theta) = \frac{\nu!}{(\nu - n)!}\prod_{0\leq t\leq\tau}\Big\{\prod_{i=1}^{\nu}\lambda_i(t)^{dN_i(t)}(1 - \lambda_i(t)dt)^{1-dN_i(t)}\Big\}, \tag{3}$$

where n denotes the number of distinct faults being detected. This can be reduced to

$$L(\theta) = \frac{\nu!}{(\nu - n)!}\prod_{i=1}^{\nu}\Big\{\Big(\prod_{0\leq t\leq\tau}\lambda_i(t)^{dN_i(t)}\Big)\cdot\exp\Big(-\int_0^{\tau}\lambda_i(t)dt\Big)\Big\}. \tag{4}$$

For the removal experiment, Littlewood (1980) suggested use of the maximum likelihood (ML) to estimate the parameters ω, ϵ (or α, β) and ν. However, due to the complexity of the three highly non-linear likelihood equations, it is difficult if not impossible to determine consistent estimates from the possible multiple solutions of the likelihood equations. The same difficulty exists for alternative estimation methods such as M-estimation. In fact, M-estimators using optimal weight functions are equivalent to the MLE. The simple M-estimators for α, β and ν can be obtained by solving the equations

$$\sum_{i=1}^{\nu} \int_0^{\tau} k_j(t) \Big[dN_i(t) - Y_i(t) \Big\{ \frac{\alpha + N_i(t)}{\beta + t} \Big\} dt \Big] = 0, \tag{5}$$

$j = 1, 2, 3$, with weight function $k_j(t) = (\beta + t)t^{j-1}$, see Andersen et al.(1993). The results, however, are highly unstable. Furthermore, for the recapture case, the parameter ν cannot be separated from the estimating equation (5) and so we are unable to solve (5) for $\hat{\nu}$.

In this paper we propose a two-step estimation procedure. First, with the specified form of intensity (1) or (2), we are able to compute the conditional likelihood of ω and ϵ (or α and β) using the observed failure information. The second stage employs a Horvitz-Thompson (1952) type estimator which is the minimum variance unbiased estimator if the failure intensity is known (Chen (2001)). The score functions in the first stage reduce to two dimensional equations that are tractable, and more applicable in practice.

2. Inference Procedure

Let $p_i = Pr(\delta_i = 1)$ denote the probability of the ith fault being detected during the course of the experiment. These probabilities are the same for all faults under the model assumption and, with the application of Laplace transform, it can be shown that $p_i = p(\omega, \epsilon) = 1 - (\omega/(\omega + \epsilon\tau))^{1/\epsilon}$. The likelihood function given in (4) can be rewritten as $L(\theta) = L_1 * L_2$, where

$$L_1 = \prod_{i=1}^{\nu} \left\{ \frac{\prod_{0 \le t \le \tau} \lambda_i(t)^{dN_i(t)} \cdot \exp\big(- \int_0^{\tau} \lambda_i(t)dt \big)}{p_i} \right\}^{\delta_i} \tag{6}$$

$$L_2 = \frac{\nu!}{(\nu - n)!} \prod_{i=1}^{\nu} \left\{ \exp \left[- \int_0^{\tau} \lambda_i(t)dt \right]^{1-\delta_i} \cdot p_i^{\delta_i} \right\}, \tag{7}$$

and δ_i indicates, by the value of 1 versus 0, whether or not the ith fault has ever been detected during the experiment. Since the marginal likelihood L_2 depends on the unknown parameter ν, we propose to make inference about $\theta = (\omega, \epsilon)'$ based on the conditional likelihood L_1 which does not depend on ν. The corresponding score function of the conditional likelihood L_1 is given by

$$\begin{aligned} U(\theta) &= \partial \log(L_1)/\partial\theta \\ &= \sum_{i=1}^{\nu} \delta_i \Big\{ \int_0^{\tau} \frac{\partial \log \lambda_i(t)}{\partial\theta} dN_i(t) - \int_0^{\tau} \frac{\partial \lambda_i(t)}{\partial\theta} dt - \frac{\partial \log(p_i)}{\partial\theta} \Big\}, \end{aligned} \tag{8}$$

where $\lambda_i(t) = Y_i(t)\{(1+\epsilon N_i(t-))/(\omega+\epsilon t)\}$.

Let $\hat{\theta} = (\hat{\omega}, \hat{\epsilon})'$ be the solution to $U(\theta) = 0$. The usual arguments ensure consistency and the convergence of $\nu^{\frac{1}{2}}(\hat{\theta}-\theta)$ to $N(0, I^{-1}(\theta))$, where $I(\theta) = E\{\nu^{-1}A(\theta)\}$, $A(\theta) = -\partial U(\theta)/\partial\theta'$. The variance of $\hat{\theta}$ can be estimated by the inverse of the negative of the first derivative of the score function, $A^{-1}(\hat{\theta})$.

Since the probability of being detected is $p(\theta)$, it is natural to estimate the population size ν by the Horvitz-Thompson estimator (Horvitz and Thompson (1952)):

$$\hat{\nu} = \sum_{i=1}^{\nu} \frac{\delta_i}{p(\hat{\theta})} = \frac{n}{p(\hat{\theta})}, \tag{9}$$

where n denotes the number of distinct faults detected. To compute the variance of $\hat{\nu}$, we have $\hat{\nu}(\theta) - \nu = \sum_{i=1}^{\nu}[\delta_i/(p-1)]$. By a Taylor series expansion and some simple probabilistic arguments,

$$\nu^{-\frac{1}{2}}\{\hat{\nu}(\hat{\theta}) - \nu\} = \nu^{-\frac{1}{2}} \sum_{i=1}^{\nu} \left[\frac{\delta_i}{p} - 1\right] + H'(\theta)\nu^{\frac{1}{2}}(\hat{\theta}-\theta) + o_p(1), \tag{10}$$

where $H(\theta) = E\{\hat{H}(\theta)\}$, $\hat{H}(\theta) = \nu^{-1}\frac{\partial\hat{\nu}(\theta)}{\partial\theta} = -\nu^{-1}\left\{n\frac{\partial p/\partial\theta}{p^2}\right\}$.

It follows from the Multivariate Central Limit Theorem and the Cramér-Wold device that $\nu^{-\frac{1}{2}}\{\hat{\nu}(\hat{\theta}) - \nu\}$ converges in distribution to a zero-mean normal random variable. The variance for the first term on the right side of (10) is

$$\nu^{-1}\sum_{i=1}^{\nu}\frac{\mathrm{Var}\,(\delta_i)}{p^2} = \nu^{-1}\sum_{i=1}^{\nu}\frac{p(1-p)}{p^2},$$

which can be estimated by $\nu^{-1}\sum_{i=1}^{\nu}\frac{\delta_i(1-p)}{p^2} = \nu^{-1}\frac{n(1-p)}{p^2}$. The variance for the second term is $H'(\theta)I^{-1}(\theta)H(\theta)$. The covariance between the two terms is zero. A consistent variance estimator for $\nu^{-\frac{1}{2}}\{\hat{\nu}(\hat{\theta}) - \nu\}$ is then given by

$$\hat{s}^2 = \hat{\nu}^{-1}(\hat{\theta})\frac{n(1-\hat{p})}{\hat{p}^2} + \hat{H}'(\hat{\theta})\hat{\nu}(\hat{\theta})A^{-1}(\hat{\theta})\hat{H}(\hat{\theta}), \tag{11}$$

where $\hat{H}(\theta) = \nu^{-1}(\partial\hat{\nu}/\partial\theta)$ and $\hat{p} = (1 - \hat{\omega}/(\hat{\omega}+\hat{\epsilon}\tau))^{1/\hat{\epsilon}}$.

3. Simulation Studies

A number of simulation studies were conducted to examine the performance of the estimator. For the selected values of the parameters ω and ϵ, we generated detection time $T_{ij}, i = 1,\ldots,\nu$ of each fault i according to different distributions as follows:

$$\phi_i, i = 1,\ldots,\nu, \text{ i.i.d. } G(1/\epsilon, \omega/\epsilon),$$

$T_{ij}|\phi_i(j = 1, \ldots, n_i)$, i.i.d. $\exp(\phi_i)$, until $\sum_{j=1}^{n_i+1} T_{ij} > \tau$, where n_i denotes the number of times fault i being detected during the experiment.

Different combinations of population sizes and failure intensities were considered for removal and recapture experiments. The results are given in Tables 1 and 2 respectively. The population size varies from 100 to 5000. In the simulations, we used the gamma distributions $G(1, 1)$ for the occurrence rate ϕ, that is, $\omega = 1$ and $\epsilon = 1$. In the removal experiment, the probabilities of being detected are 0.8 and 0.9, which correspond to $\tau = 4$ and $\tau = 9$, respectively. The settings are the same as those in van Pul (1993) for removal studies, while for the recapture experiment, with more information available, the detection probability can be smaller to achieve comparably good estimates. In our simulations, we set $p = 0.7$ and $p = 0.9$. For each combination of ν and p, 5000 replications were simulated to estimate the mean, the sample standard deviation (SD), and the mean of the estimated standard error (AV(SE)) of population size ν and the parameters ω and ϵ. The behavior of $\hat{\nu}$ for removal and recapture experiments are illustrated by the histograms in Figures 1 and 2, respectively.

For the removal studies shown in Table 1, we cannot get a comparably good estimator of $\hat{\nu}$ with small number of faults in the system; ω can be estimated relatively well in contrary with $\hat{\epsilon}$, which has a larger range. As shown in Figure 1, the distributions of $\hat{\nu}$ are not symmetric with a long right tail. However, with the increase of the total number of faults and detection probability, the performance of the proposed method improves and appears satisfactory.

Table 2 gives similar results for a recapture study. The results in the recapture study are better than those described above. This is not surprising because more information is available in recapture experiment, see Lloyd et al. (1999), Yip (1998), and Yip et al. (1999). As shown in Figure 2, the distributions of $\hat{\nu}$ (also $\hat{\omega}$ and $\hat{\epsilon}$) are positively biased for small ν. As ν and the detection probability increase, this bias slowly disappears: the means of estimated parameters are close to the true values and the estimated standard errors seem to agree reasonably well with sample standard deviations. For recapture experiments, when detection probabilities are high, both $\hat{\nu}$ and $\hat{s}$ are almost unbiased, even for $\nu = 50$.

Simulation experiments were also conducted to study the effects of ignoring heterogeneity. We generated data from the homogeneous Jelinski-Moranda model. The results of estimating ν are also given in Table 2. The $\hat{\nu}$ derived from assuming a Jelinski-Moranda are substantially underestimated. The estimated standard errors are also biased downward and the coverages are unsatisfactory.

Simulation studies were performed to compare the proposed method with the full ML estimators for removal experiment, i.e., the Littlewood model. Table 3 gives the estimation results of the proposed two-step method and Method II (van

Table 1. Summary of simulation results for removal studies with $\omega = 1.0$, $\epsilon = 1.0$.

$\nu = 100$

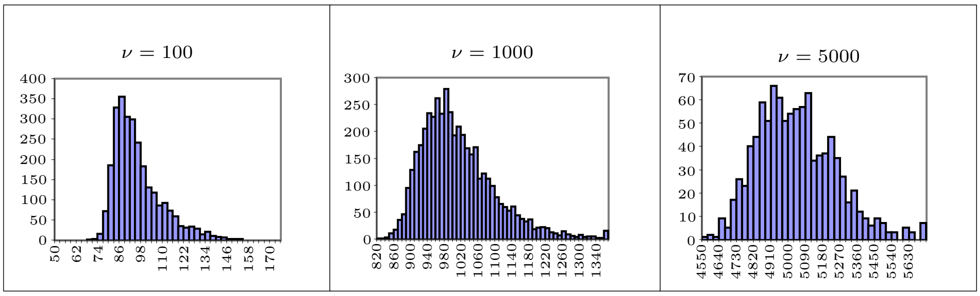
	$p = 0.8$			$p = 0.9$		
	Mean	SD	AV(SE)	Mean	SD	AV(SE)
$\hat{\nu}$	93.26	12.76	20.40	98.84	8.25	10.03
$\hat{\omega}$	1.06	0.1668	0.2516	1.12	0.1872	0.2318
$\hat{\epsilon}$	0.59	0.5248	0.8923	0.83	0.4689	0.6176
	Breakdown Prob. 42.5%			Breakdown Prob. 29.3%		

$\nu = 1000$

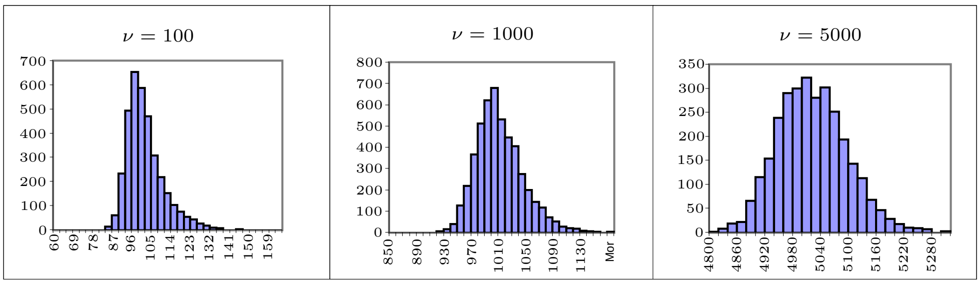
	$p = 0.8$			$p = 0.9$		
	Mean	SD	AV(SE)	Mean	SD	AV(SE)
$\hat{\nu}$	1005.33	88.81	98.71	1002.31	34.57	34.42
$\hat{\omega}$	1.02	0.0869	0.0902	1.01	0.0656	0.0676
$\hat{\epsilon}$	1.00	0.3799	0.4341	1.00	0.2110	0.2157
	Breakdown Prob. 8.2%			Breakdown Prob. 2.4%		

$\nu = 5000$

	$p = 0.8$			$p = 0.9$		
	Mean	SD	AV(SE)	Mean	SD	AV(SE)
$\hat{\nu}$	5012.95	204.96	209.61	5002.58	72.88	75.03
$\hat{\omega}$	1.00	0.0372	0.0376	1.00	0.0303	0.0299
$\hat{\epsilon}$	1.01	0.1835	0.1870	1.00	0.0935	0.0948
	Breakdown Prob. 0%			Breakdown Prob. 0%		



$p = 0.8$



$p = 0.9$

Figure 1. Histogram of $\hat{\nu}$ for removal experiment with $\omega = 1.0$ and $\epsilon = 1.0$.

Table 2. Summary of simulation results for recapture studies with $\omega = 1.0$, $\epsilon = 1.0$.

$\nu = 100$

	p=0.7			p=0.9		
	Mean	SD	AV(SE)	Mean	SD	AV(SE)
$\hat{\nu}$	103.02	19.48	21.92	99.93	5.68	5.65
	(73.21)	(5.09)	(2.12)	(89.99)	(2.99)	(0.07)
$\hat{\omega}$	1.03	0.2327	0.2612	1.10	0.1250	0.1223
$\hat{\epsilon}$	1.02	0.5538	0.6348	0.97	0.2410	0.2514

$\nu = 500$

	$p = 0.7$			$p = 0.9$		
	Mean	SD	AV(SE)	Mean	SD	AV(SE)
$\hat{\nu}$	505.45	45.62	44.30	496.95	11.69	11.94
	(365.09)	(10.70)	(4.64)	(449.95)	(6.73)	(0.15)
$\hat{\omega}$	1.01	0.1073	0.1049	1.01	0.0524	0.0529
$\hat{\epsilon}$	1.03	0.2662	0.2595	0.95	0.0999	0.1074

$\nu = 1000$

	$p = 0.7$			$p = 0.9$		
	Mean	SD	AV(SE)	Mean	SD	AV(SE)
$\hat{\nu}$	1005.03	62.42	60.42	993.02	16.96	17.07
	(730.04)	(15.03)	(6.53)	(899.77)	(9.61)	(0.21)
$\hat{\omega}$	1.01	0.0726	0.0717	1.01	0.0363	0.0371
$\hat{\epsilon}$	1.01	0.1823	0.1773	0.95	0.0694	0.0753

Note: Numbers inside the brackets are the estimates obtained ignoring frailty in the model.

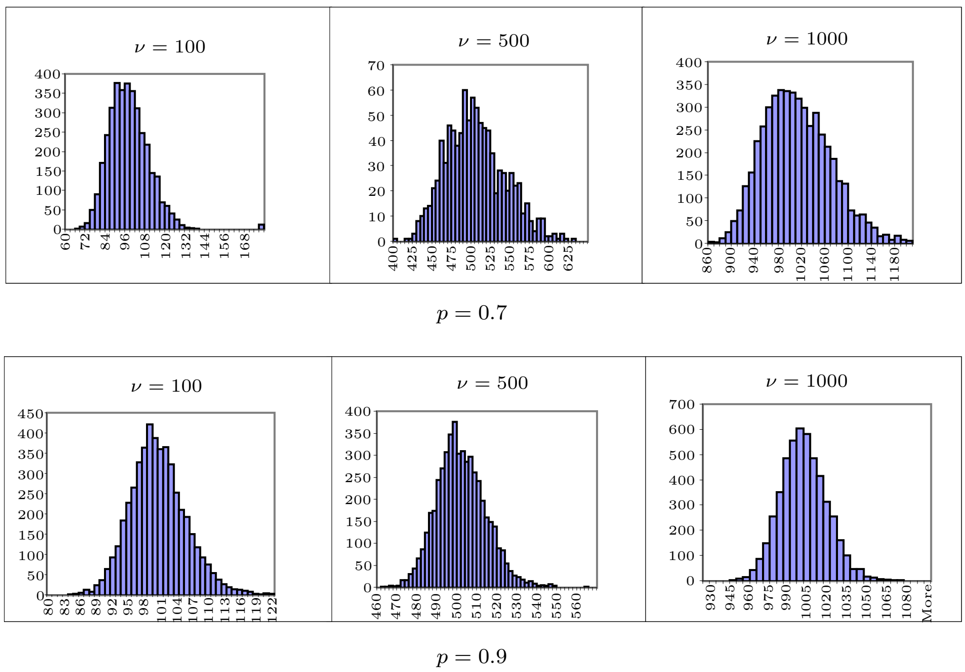


Figure 2. Histogram of $\hat{\nu}$ for recapture experiment with $\omega = 1.0$ and $\epsilon = 1.0$.

Pul (1993), p.94) for solving the full likelihood. For convenience of comparison, the set-up for the simulations is identical with that of van Pul (1993). We generated failure times according to the Littlewood model with $\omega = 1$, $\epsilon = 1$, and different values of ν (100, 1000 and 10,000). In the simulations we carried out the Newton-Raphson iterative procedure to solve the estimating equations (8), while van Pul (1993) took much more effort to search for the roots of the three non-linear equations from the full likelihood equations. Comparison of the two types of estimators show that the proposed estimators do not differ much from the full MLE even when the sample size is not large.

Furthermore, we performed simulations with other frailty models of varying degree of heterogeneity, with the results summarized in Table 4. The failure intensities are generated from gamma distributions with different coefficients of variation (CV). The sample size is 500. The estimation results of Chao's sample coverage method (Chao and Lee (1992)) are also given for comparison. One sees that the non-parametric sample coverage estimator underestimates the population size with a large relative mean square error, while the present proposal performs satisfactorily.

Table 3. Mean square errors for $\tilde{\theta}$ and $\hat{\theta}$

	(a) ν=100	(b) ν=1000	(c) ν=10000
$\tilde{\nu}$	0.0237	0.0060	0.0006
$\tilde{\omega}$	0.0461	0.0052	0.0007
$\tilde{\epsilon}$	0.5767	0.1322	0.0101
$\hat{\nu}$	0.0208	0.0079	0.0008
$\hat{\omega}$	0.0318	0.0079	0.0007
$\hat{\epsilon}$	0.4432	0.1441	0.0175

Note: $\hat{\theta}$ - the estimator of the two-step method
$\tilde{\theta}$ - the estimator of method II (van Pul (1993)).

Table 4. Simulation results for heterogeneous models with ν=500

Estimator	CV	Ave($\hat{\nu}$)	sd($\hat{\nu}$)	Ave($\hat{se}(\hat{\nu})$)	RMSE
Chao & Lee	0.58	486.6	13.9	13.8	19.2
	0.71	483.8	13.5	13.7	21.2
	1.00	477.7	14.0	14.5	26.6
	1.12	474.7	14.3	15.0	29.4
Two-step	0.58	500.3	18.6	17.9	17.8
	0.71	500.3	18.8	18.6	18.6
	1.00	500.9	21.5	20.8	20.9
	1.12	500.5	22.9	23.6	23.6

Note: $\text{Ave}(\hat{\nu}) = \Sigma\, \hat{\nu}_i / R$; $\widehat{sd}(\hat{\nu})^2 = \Sigma\, (\hat{\nu}_i - \text{Ave}(\hat{\nu}))^2/(R-1)$; $\text{Ave}\{\widehat{se}(\hat{\nu})\} = \{\Sigma\ \widehat{se}(\hat{\nu}_i)\}/R$; $\text{RMSE}^2 = \Sigma\ \left(\frac{\hat{\nu}_i - \nu}{\nu}\right)^2/R$; $R = 1000$ is the number of simulations.

4. Examples on Reliability Data

The proposed method is applied to two data sets. The first is from a reliability project concerning an information system for registering aircraft movements Moek (1983,1984). Failure data collected during the testing stage can be found in Table 5. The proposed conditional likelihood is maximized at $\hat{\omega} = 0.36(0.0001)$ and $\hat{\epsilon} = 0$, which indicates that the Jelinski-Moranda model is the best model. The corresponding estimate for the total number of faults is $\hat{\nu}$=45.53 (2.37), which is close to the full likelihood estimator $\nu = 44.5(1.98)$. The estimated detection probability is thus around 0.95. Using the M-estimates in (5), suggested in Andersen et al. (1993), we obtain $\hat{\omega} = 0.196(0.007)$, $\hat{\epsilon} = 0.089(0.102)$ and $\hat{\nu} = 46.4(10.14)$ based on the 43 failures reported. The M-estimates without the optimal weights obviously perform worse than the proposed conditional likelihood estimates.

For the recapture experiment, each fault is detected, corrected and a counter is inserted to record the number of revisits of a particular fault. To use the dataset, we generated interfailure times for each of the detected faults in Table 5 from an exponential distribution with the maximum likelihood estimate of the failure intensity $\hat{\phi} = 5.4$ (Andersen et al., 1993), the data on their recapture times are given in the third column of Table 5. The estimated number of total faults in the system using the recapture data is $\hat{\nu} = 45.24(1.60)$.

The second example considers the dataset in van Pul ((1993), Appendix). It comprised 165 distinct faults, recorded in Table 6. The time of the last observed failure, 1.0472, was chosen as the stopping time of the experiment, τ. Figure 3 gives a total time on test (TTT) plot of the data, which graphically suggests an exponential failure intensity (van Pul (1993)). We fitted the proposed removal method, giving the estimates $\hat{\omega} = 0.377(0.029)$, $\hat{\epsilon} = 0$ and $\hat{\nu} = 175.92(4.22)$. The estimated parameter $\hat{\epsilon}$ turned out to be zero, which suggests the Jelinski-Moranda model is the appropriate model within a larger class, with the failure rate $\hat{\phi} = 2.654(0.204)$. The estimated detection probability is $\hat{p} = 0.95$.

5. Discussion

The proposed model in this paper is a unified one which comprises both removal and recapture models. Our motivation of the recapture debugging model is to obtain extra information on the failure intensity ϕ and the remaining number of faults ν, so that better estimation can be achieved. The conditional likelihood with a Horvitz-Thompson estimator is proposed. The resulting estimators are asymptotically equivalent to the full likelihood estimators, confirmed by simulation studies. Sanathanan (1972) also showed the asymptotic equivalence of the unconditional and conditional MLE. The proposed estimators perform estimators perform well in small samples when capture probabilities are high, especially for

Table 5. Real and simulated failure times of the Moek's data (1984)

fault #	first	subsequent simulated failure times					
1	880	64653	329685	520016			
2	4310	305937	428364	432134	576243		
3	7170	563910					
4	18930	186946	195476	473206			
5	23680	259496	469180				
6	23920	126072	252204	371939			
7	26220	251385					
8	34790	353576					
9	39410	53878	147409	515884			
10	40470	371824	466719				
11	44290	83996	327296	352035	395324	494037	
12	59090	61435	222288	546577			
13	60860	75630	576386				
14	85130						
15	89930	205224	292321	294935	342811	536573	553312
16	90400	228283	334152	360218	368811	377529	547048
17	90440	511836	511967				
18	100610	367520	429213				
19	101730	162480	534444				
20	102710	194399	294708	295030	360344	511025	
21	127010	555065					
22	128760						
23	133210	167108	370739				
24	138070	307101	451668				
25	138710	232743					
26	142700	215589					
27	169540	299094	428902	520533			
28	171810	404887					
29	172010	288750					
30	211190						
31	226100	378185	446070	449665			
32	240770	266322	459440				
33	257080	374384					
34	295490	364952					
35	296610						
36	327170	374032	430077				
37	333380						
38	333500	480020					
39	353710						
40	380110	433074					
41	417910	422153	479514	511308			
42	492130						
43	576570						

Table 6. Failure dataset from van Pul ((1993), Appendix)

0.0020	0.0052	0.0066	0.0106	0.0146	0.0189	0.0231	0.0253	0.0286	0.0309	0.0315
0.0323	0.0332	0.0337	0.0377	0.0446	0.0460	0.0533	0.0547	0.0578	0.0628	0.0656
0.0719	0.0739	0.0740	0.0779	0.0792	0.0802	0.0810	0.0848	0.0852	0.0863	0.0872
0.1021	0.1040	0.1043	0.1061	0.1087	0.1125	0.1178	0.1180	0.1191	0.1192	0.1196
0.1212	0.1215	0.1220	0.1221	0.1260	0.1261	0.1262	0.1291	0.1303	0.1331	0.1370
0.1375	0.1394	0.1406	0.1436	0.1437	0.1452	0.1460	0.1464	0.1481	0.1482	0.1494
0.1496	0.1497	0.1504	0.1523	0.1533	0.1534	0.1536	0.1538	0.1539	0.1552	0.1618
0.1713	0.1750	0.1767	0.1773	0.1856	0.1871	0.1930	0.1963	0.1994	0.2015	0.2029
0.2040	0.2053	0.2062	0.2141	0.2143	0.2152	0.2199	0.2250	0.2302	0.2304	0.2314
0.2323	0.2763	0.2864	0.2888	0.2959	0.3016	0.3017	0.3029	0.3048	0.3055	0.3057
0.3154	0.3156	0.3211	0.4036	0.4144	0.4188	0.4196	0.4207	0.4214	0.4296	0.4298
0.4300	0.4302	0.4379	0.4471	0.4569	0.4688	0.4821	0.4975	0.5013	0.5176	0.5295
0.5611	0.5687	0.5712	0.5744	0.5783	0.5834	0.6032	0.6323	0.6602	0.6761	0.6911
0.7036	0.7105	0.7461	0.7570	0.7596	0.7626	0.7655	0.8052	0.8133	0.8204	0.9009
0.9059	0.9220	0.9296	0.9605	0.9752	0.9928	1.0063	1.0139	1.0289	1.0375	1.0472

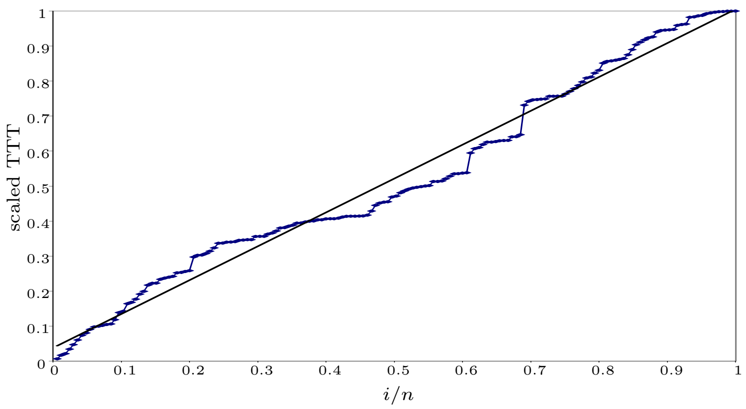


Figure 3. Scaled TTT plot of the failure data set from van Pul (1993)

recapture experiments. Ignoring frailty would seriously underestimate the population size and give a misleading smaller standard error (Chao (1987), Yip et al. (1999)).

The proposed two-step estimator further provides an algorithm for solving the Littlewood model. Different methods have been discussed to solve the likelihood equations (Moek (1984); Geurts, Hasselaar and Verhagen (1988); van Pul (1993)). However, these procedures are generally complicated. The proposed score equations (8) are more tractable: standard methods like the Newton-Raphson procedure or downhill simplex methods can be used to search the equation roots. In practice we suggest using the method I estimators of van Pul (1993), $\sqrt{n}$-consistent, as the initial value for ω and ϵ separately.

Acknowledgements

This work is supported by the Research Grant Council of Hong Kong. The authors are grateful for the suggestions from the Editor, an Associate Editor and a referee that improved the paper.

References

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701-726.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal capturability. *Biometrics* **43**, 783-791.

Chao, A. and Lee, S. M. (1992). Estimating the number of classes via sample coverage. *J. Amer. Statist. Assoc.* **87**, 210-217.

Chen, K. (2001). Parametric and semiparametric models for recapture and removal studies: a likelihood approach. *J. Roy. Statist. Soc. Ser. B* **63**, 607-619.

Geurts, W. A. J., Hasselaar, M. M. A. and Verhagen, J. H. (1988). Large sample theory for statistical inference in several software reliability models. Report MS-R8807, Centre for Mathematics and Computer Science, Amsterdam.

Hayakawa, Y. and Yip, P. S. F. (2000). A Gibbs sampler approach to estimation of the number of faults in a system using capture-recapture sampling. *IEEE Trans. Reliability* **49**, 342-351.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47**, 663-685.

Jelinski, Z. and Moranda, P. (1972). Software reliability research. *Statist. Comput. Performance Evaluation*, 465-484.

Littlewood, B. (1980). Theories of software reliability: How good are they and how can they be improved? *IEEE Trans. on Software Engineering* **6**, 489-500.

Lloyd, C., Yip, P. S. F and Chan, K. S. (1999). Estimating the number of errors in a system: the relative efficiency of removal, recapture and seeded errors methods. *IEEE Trans. Reliability* **48**, 369-376.

Moek, G. (1983). Software reliability models on trial: selection, improved estimation, and practical results. Report MP 83059 U, National Aerospace Laboratory NLR, Amsterdam.

Moek, G. (1984). Comparison of some software reliability models for simulated and real failure data. *Internat. J. Modeling and Simulation* **4**, 29-41.

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

Nielsen, G. G., Gill, R. D., Andersen, P. K. and Sørensen, T. I. A. (1992). A counting precess approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19**, 25-43.

Sanathanan, L. (1972) Estimating the size of a multinomial population. *Ann. Math. Statist.* **43,** 142-52.

van Pul M. C. (1992). Asymptotic properties of a class of statistical models in software reliability. *Scand. J. Statist.* **19**, 235-253.

van Pul M. C. (1993). *Statistical Analysis of Software Reliability models*. CWI Tract, Amsterdam.

Yip, P. S. F. (1998). An efficiency comparison of estimating the number of errors in a system using recapture and removal methods. *Stochastic Anal. Appl.* **16**, 1129-1145.

Yip, P. S. F., Xi, L., Fong, D. Y. T. and Hayakawa, Y. (1999). Sensitivity analysis and estimating number of faults in removal debugging. *IEEE Trans. Reliability* **48**, 300-305.

School of Mathematics, University of South Australia, Adelaide, Australia.

E-mail: yan.wang@unisa.edu.au

Department of Statistical and Actuarial Sciences, the University of Hong Kong, Hong Kong.

E-mail: sfpyip@hku.hk

Institute of Statistics and Operation Research, Victoria University of Wellington, Wellington, New Zealand.

E-mail: Yu.Hayakawa@mcs.vum.ac.nz

(Received January 2001; accepted February 2002)