Second Seminar on
Reliability Theory and its Applications
18-19, May 2016

Analysis of Masked Data with non-ignorable missing mechanism

Misaii, H. [1] and Eftekhari, S. [2]

[1,2] Department of Statistics, Faculty of Mathematics and Statistics, University of Tehran, Tehran, Iran

Abstract

In this article we consider series system products with independent lifetimes for each component in which causes of system failures might be masked Due to cost considerations or environmental restrictions. As a new approach we have defined a missing indicator for the true cause of failure which might be observed or masked. The likelihood with missing not at random (MNAR) mechanism is derived and compared with the usual likelihood for masked data via some simulation studies. The results show superior performance of our approach when non-ignorable missing mechanism is occurred.

Keywords: Masked Data , Non-ignorable missing data, Reliability Analysis.

1 Introduction

In general, failure time data for a series system contain the time to failure along with information on the exact component responsible for the system failure. Which can be used to estimate system and component reliabilities. In many cases, however, due to lack of proper diagnostic equipment or cost and time constraints, the exact component causing the system failure is not identified, but the cause of failure is only narrowed down to a smaller set of components. These Kinds of data are called to be masked (Miyakawa;1984, Basu, et al;1999). In the analysis of masked data, often it is assumed that the masking probability is independent of the component which caused the failure or in the case of dependency some prior probabilities are assumed for these conditional events. Also masking probability is assumed to be dependent on the observed failure time. In this article, as a new approach we will introduce a missing indicator according to the masking status of each

[1]hasanmisaii14@gmail.com
[2]s.eftekhari@khayam.ut.ac.ir

observed failure time. Also a MNAR mechanism (Little Rubin 2002) is assumed for this variable which allow masking to be dependent on the component index along with the observed failure time using generalized linear model. We have performed some simulation studies under exponentially distributed failure times which show superior results for our proposed model when masking has a non-ignorable mechanism.

2 Assumption and Likelihood function

Consider series systems with J components. Assume the observed failure data are $t_1, t_2, ..., t_r$,(n-r systems are censored) but the exact associated cause of failure might be unknown, which is only known to belong to the Minimum Random Subset (MRS) of $\{1, 2, ..., J\}$. Let M_i be the observed MRS corresponding to the failure time $t_i, i = 1, 2, ..., r$. The set M_i essentially consists of components, narrowed down to be the possible causes responsible for the system failure. When M_i is a singleton set, then the corresponding system has an observed cause of failure. While if $M_i = \{1, ..., J\}$ the system is called to be completely masked. We define the binary variable R_i which takes the value 1, when M_i is a singleton and has a zero value for masked data(when M_i has more than one element). Thus, the observed data are

$$(t_1, M_1, R_1), (t_2, M_2, R_2), ..., (t_r, M_r, R_r) \tag{2.1}$$

The model considered in this article is based on the following assumptions.

- The failure of a system occurs due to one of the J independent components with lifetimes, $T_1, T_2, ..., T_J$, and the failure time of the system (T) is the shortest of $T_l, l = 1, ..., J$.

- T_l, the failure time of the lth component , follows a continuous distribution with density and reliability functions denoted by $f_l(t), R_l(t)$.

- $Pr(M = M_i|T = t_i, K_i = l)$ is called the masking probability, where K_i denotes the index of the component actually causing the ith system to fail. $p_l(M_i)$s have some constraints. Let M_i be the collection of all 2^{J-1} possible nonempty subsets of $\{1, ..., J\}$ and $M_l = \{M_0 \in M : l \in M_0, l \in \{1, ..., J\}\}$ and $p_l(M_i) = P(M = M_i|K_i = l) = 0, \forall M_i \in M_l^c = M - M_l$ and thus $\sum_{M_i \in M} p_l(M_i) = \sum_{M_i \in M_l} p_l(M_i) = 1, l = 1, ..., J$. Denote $p_l = \{P_l(M_i) : M_i \in M_l\}$ and $p = (p_1, ..., p_J)$.

- R_i is a Bernoulli variable with success probability $p(R_i = 1|k_i = j, M_i, t_i; j \in M_i) = h(\beta_0 + \beta_1 k_i + \beta_2 t_i)$, where h(.) is some appropriate link function (e.g. logit, probit, clog-log,...). When $\beta_1 = 0$ the missing is ignorable.

Let τ be the end time of the test and $I_{mask} = \{1 \leq i \leqslant r; R_i = 0\}$ denotes the set of indices for maked systems. Therefor the complete likelihood function for data 2.1 is as follows:

$$\begin{aligned} L(\theta) &= \prod_{i \in I^c_{mask}} [p(R_i = 1|k_i = j, t_i, M_i = \{j\})p(M_i = \{j\}|k_i = j, t_i) \\ &\times \quad f_j(t_i)p(k_i = j)] \prod_{i \in I_{mask}} [p(R_i = 0|k_i = j, t_i, M_i) \\ &\times \quad p(M_i|k_i = j, t_i)f_j(t_i)p(k_i = j)[\prod_{l=1}^{J} R_l(\tau)]^{n-r} \end{aligned} \tag{2.2}$$

If the missing mechanism is at random $(\beta_1 = 0)$ the above likelihood reduces to

$$L(\theta) \propto \prod_{i=1}^{r} \left[f(R_i|t_i) \sum_{j \in M_i} f(t_i|k_i, t_i)p(M_i|k_i, t_i)p(k_i = j) \right] \left[\prod_{l=1}^{J} R_l(\tau) \right]^{n-r}$$

Where R_is could be ignored and simple masked data analysis be used.

3 Simulation Study

In this section we assume 100 series systems with two components where the lifetimes of components follow some exponential distribution with parameters α_1 and α_2 respectively for first and second component. We have generated non-ignorable missing mechanism according to the logistic regression $logit(p(R_i = 1|k_i = j)) = \beta_0 + \beta_1 k_i$. The results of maximum likelihood estimation of the parameters α_1 and α_2 are presented in Table 1. In this table the true value of parameters along with the percent of failure due to second component ,CP, and masking rate,MP, are given. Also

Table 1: The results of simulation analysis

	α_1	α_2	β_1	β_0	CP	MP	Bα_1	Bα_2
MAR	0.3	0.7	2.5	-2	70	47	0.062	0.059
MNAR							0.033	0.036
MAR	0.3	0.7	2.5	-3	70	27	0.04	0.024
MNAR							0.011	0.024
MAR	0.3	0.7	2.5	-4	70	13	0.011	0.02
MNAR							0.012	0.002
MAR	0.3	0.7	3.5	-3.5	69	35	0.081	0.073
MNAR							0.002	0.006
MAR	0.3	0.7	3.5	-4	70	26	0.041	0.044
MNAR							0.007	0.005
MAR	0.3	0.7	3.5	-5	70	13	0.02	0.009
MNAR							0.002	0.008
MAR	0.3	0.5	2.5	-4	62	12	0.034	0.018
MNAR							0.007	0.009
MAR	0.3	0.5	2.5	-3	63	25	0.041	0.043
MNAR							0.021	0.019
MAR	0.3	0.5	2.5	-2	63	44	0.082	0.076
MNAR							0.041	0.047
MAR	0.3	0.5	3.5	-5	61	12	0.034	0.021
MNAR							0.005	0.007
MAR	0.3	0.5	3.5	-4	63	25	0.053	0.037
MNAR							0.009	0.025
MAR	0.3	0.5	3.5	-3	62	41	0.117	0.108
MNAR							0.007	0.016

the last two columns present the amount of biasness for α_1 and α_2 (respectively denoted by $B\alpha_1$ and $B\alpha_2$) in 100 iterations of each simulation study. According to the results, MNAR modeling leads to less biased estimators compared with the usuall MAR model.

References

[1] Basu, S. Basu, A. P. and Mukhopadhyay, C. (1999). Bayesian analysis for masked system failure data using nonidentical weibull models, *J.Statist. Plann. Inference*, **78**, 255275.

[2] Little, R. J. A., Rubin, D. B. (2002). Statistical Analysis With Missing Data, 2nd Edition, *Joh Wiley Sons: New York.*

[3] Miyakawa, M. (1984), Analysis of incomplete data in a competing risks model, *IEEE Transactions on Reliability*, **33**, 293-296.