IJRRS
International Journal of Reliability, Risk Saftey: Theory and Application

DOI : 10.30699/IJRRS.3.2.1 Vol. 3/ Issue 2/ 2020
pp. 1-7

Masked Data Analysis based on the Generalized Linear Model

H. Misaii[1*], F. Haghighi[1], S. EftekhariMahabadi[1]

1. School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Tehran, Iran

Abstract

In this paper, we consider the estimation problem in the presence of masked data for series systems. A missing indicator is proposed to describe masked set of each failure time. Moreover, a Generalized Linear model (GLM) with appropriate link function is used to model masked indicator in order to involve masked information into likelihood function. Both maximum likelihood and Bayesian methods were considered. The likelihood function with both missing at random (MAR) and missing not at random (MNAR) mechanisms are derived. Using an auxiliary variable, a Bayesian approach is expanded to obtain posterior estimations of the model parameters. The proposed methods have been illustrated through a real example.

Keywords: Bayesian Modeling; Markov chain Monte Carlo Method; Masked Data; Non-ignorable Missing Mechanism.

Introduction

In a series system, the failure time and the exact component that causes system failure are important and can be used to estimate the reliability of component and system. However, in many cases, for reasons such as lack of diagnostic equipment, cost and time constraints, the exact component that causes the system to fail is not known and we only know that it belongs to a smaller set of components. Data with this feature is called masked data [1, 2, 3].

The problem of maximum likelihood estimates (MLE) in the presence of masked data has been considered by some authors such as Miyakawa [1], Usher and Hodgson [4] and Lin et al. [5], while Reiser [6], Berger and Sun [7],Mukhopadhyay and Basu[8] and Cai et al. [9] studied Bayesian statistical inference under masked data. Sen et al. [10] provided comprehensive details of the statistical analysis of system failure data under competing risks with possibly masked failure causes. Basu et al. [11] developed a Bayesian analysis for masked competing risks data from engineering systems and presented a general parametric framework for any number of competing risks and any distribution. Above mentioned works have been done under equiprobableassumption, that is, the masking probabilities do not relate to cause of failure(called the symmetry assumption by some authors).However, many authors have not considered this assumption, some of them are referred as follows. Lin and Guess [12] considered reliability estimation when the masking probability is related to the particular cause of failure. Guttman et al. [13] developed a Bayesian method to estimate component reliabilities from masked system lifetime data when the masking probability is related to the true cause of system failure. Kuo and Yang [14] considered different probability model for the conditional masking probabilities, along with exponential and Weibull distributions for the component lifetimes. Mukhopadhyay and Basu[15] developed a Bayesian analysis for s-independent exponentials without the symmetry assumption using s-independent priors for the component failure rates and masking probabilities. Craiu and Duchesne [16] considered the maximum likelihood estimation of the cause-specific hazard functions and the masking probabilities via an EM algorithm. Mukhopadhyay [17] developed the maximum likelihood method to estimate the lifetime parameters and masking probabilities via an EM algorithm, and constructed approximate confidence intervals, also presented bootstrap confidence intervals. Xu and Tang [18] considered a Bayesian analysis for series systems with two components with Pareto distribution lifetime where masking probabilities are independent of time. Xu et al [19] presented a Bayesian approach for masked data in step stress accelerated life testing and considered log-location-scale distribution family for their study.

There is another type of incomplete data called missing data. Missing data occur when no data value is stored for the variable in an observation and have different mechanisms with respect to missingness reasons. If missingness depends only on observed values, missing mechanism is called missing at random (MAR), while if missingness depends on both observed

* Corresponding Author Email: hmisaii@ut.ac.ir

Received: 2020.12.07, Accepted: 2020.12. 20, Online ISSN: 2676-3346

and missing values, missing mechanism is called missing not at random (MNAR) (Little \& Rubin, 2002).

In this work, both classic and Bayesian statistical inference in the presence of masked data has been studied. Novelty of work lies on the definition a missing indicator for masking set of each observed failure time. That is, if the masked set is singleton set, missing indicator takes one, otherwise takes zero. Then, a generalized linear model (GLM)with appropriate link function is used to model missing indicator and it is involved into the likelihood function. This method allows to analyse masked data in a new manner which is more flexible than existing approach, specially when using of Bayesian method is desired.

The rest of the paper is as follow. In Section 2, model assumptions are introduced, and the general formulation of the likelihood function is given. In Section 3, the auxiliary variables are introduced, and the Bayesian analysis is discussed. In section 4, The proposed methodology is represented by a numerical example. Finally, a conclusion is given in Section 5.

Model Assumptions and Likelihood Function

Assumptions

Suppose that we have r series systems under the test such that all of them have equal components, say J components. Assume that at the end of the test we observe failure data, $t_1, t_2, \dots, t_r$, but the exact cause of failure might be unknown, and only we know that belongs to the Minimum Random Subset (MRS) of {1, 2, ..., J}. Let M_i be the observed MRS corresponding to the failure time t_i; $i = 1, 2, \dots, r$ for ith system. The set M_i essentially includes components that are possible to be cause for system failure. If M_i be a singleton set, then the data are competing risks data. While if $M_i = \{1, 2, \dots, J\}$ then 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 set and has zero value for masked data (when M_i has more than one element). Thus, the observed data are

$$(t_1, M_1, R_1), (t_2, M_2, R_2), \dots, (t_r, M_r, R_r). \quad (1)$$

The model used in this paper is based on the following assumptions:

- Let$T_1, T_2, \dots, T_J$be the lifetimes of independent components, also assume that the system fails only due to one of the components, therefore system failure time is$T = \min(T_1, T_2, \dots, T_J)$.
- T_l, the failure time of the first component, follows a distribution in continuous distribution family 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 exact cause offailure ofi[th] system. In this article, we assume$\Pr(M = M_i | T = t_i, K_i = l) = \Pr(M = M_i | K_i = l) = p_l(M_i)$, that is, the masking probability is independent of failure time, but is dependent to the cause of failure.
- $p_l(M_i)$s have some constraints. SupposeMbe the all of nonempty subsets of{1,...,J}thathave$2^J - 1$members. Define$M_l = \{M_0 \in M : l \in M_0,\ l \in \{1, \dots, J\}\}$thus
  $$p_l(M_i) = \Pr(M = M_i | K_i = l) = 0, \forall M_i \in M_l^c = M - M_l$$
  And
  $$\sum_{M_i \in M} p_l(M_i) = \sum_{M_i \in M_l} p_l(M_i) = 1, \quad l = 1, \dots, J. \quad (2)$$
  Denote $p_l = \{p_l(M_i) : M_i \in M_l\}, l = 1, 2, \dots, J$ then $p = (p_1, \dots, p_J)$.
- Let Tbe the system failure time, the reliability function is given by
  $$R(t) = R(t; \theta) = P(T > t) = \prod_{l=1}^{J} [1 - F_l(t)] \quad (3)$$
  Where $\theta = (\theta_1, \dots, \theta_J)$ and θ_l is parameters set related to componentl.
- Let K be a random variable which indicates the indicator for the failure cause. Then the jointprobability distribution function of (T, K) is given by
  $$f_{T,K}(t, l) = f_l(t) \prod_{j \neq l} [1 - F_j(t)] \quad (4)$$
- $R_i$is a Bernoulli variable with success probability $p(R_i = 1 | k_i = j I_{\{j \in M_i\}}, M_i, t_i) =$ where $h(\beta_0 + \beta_1 k_i + \beta_2 t_i)$ h(.) is some appropriate link function (e.g. logit, probit, clog-log,...). When$\beta_1 = 0$the missing is ignorable and missing mechanism is MAR.

Likelihood Function

The likelihood function for data (1) can be written as follow:

$$L(\theta, p, \beta) = \prod_{i=1}^{r} [\sum_{j \in M_i} P(R_i | t_i, M_i, K_i = j) P(M_i | t_i, K_i = j) f_{T,K}(t_i, j)]$$
$$= \prod_{i=1}^{r} [\sum_{j \in M_i} P(R_i | t_i, M_i, K_i = j) P_j(M_i) f_{T,K}(t_i, j)] \quad (5)$$

Where $\beta = (\beta_0, \beta_1)$, and $\theta$is the vector of parameters related to lifetime distributions.

For simplify let$I_{mask} = \{1 \le i \le r;\ R_i = 0\}$denotes the set of indices for masked data. Therefore, the complete likelihood function for data (1) is rewritten as follows:

$$L_s(\theta) = \prod_{i \in I^c_{mask}} [P(R_i = 1|k_i = j, t_i, M_i = \{j\}) p(M_i = \{j\}|k_i = j, t_i) f_{T,K}(t_i, j)] \times \prod_{i \in I_{mask}} \sum_{j \in M_i} [P(R_i = 0|k_i = j, t_i, M_i) p_j(M_i) f_{T,K}(t_i, j)] \quad (6)$$

If the missing mechanism is at random $\beta_1 = 0$then the above likelihood is reduced to:

$$L_R(\theta) \propto \prod_{i=1}^{r} \left[f(R_i \mid t_i) \sum_{j \in M_i} p_j(M_i) f_{T,K}(t_i, j) \right] \quad (7)$$

Where the part related to$R_i's$ could be ignored and simple masked data analysis could be used.

Bayesian Analysis

Here, we define an auxiliary variable to simplify Bayesian likelihood function. Consider $I_{ij} = I(T_j = t_i)$ for $1 \le i \le r$and $1 \le j \le J$, where$I(.)$ is the indicator variable such that shows the exactcauseoffailure.$I_{ij} = 1$ means that thei^{th} system has been failed due to component where$j \in M_i$.Note that, if$M_i = \{j\}$ is a singleton set, that is the failure cause is known,then$I_{ij} = 1$ and$I_{ij'} = 0;\ j' \neq j$. Therefore, likelihood functions (5), (6) and (7) can be rewritten as follows (8), (9) and (10), respectively:

$$L(\theta, p, \beta \mid t, M, R) = \prod_{i=1}^{r} \left[\prod_{j \in M_i} (P(R_i|t_i, M_i, K_i = j) p_j(M_i) f_{T,K}(t_i, j))^{I_{ij}} \right] = \prod_{i=1}^{r} \left[\prod_{j=1}^{J} (P(R_i|t_i, M_i, K_i = j) p_j(M_i) f_{T,K}(t_i, j))^{I_{ij}} \right] \quad (8)$$

$$L_S(\theta) = \prod_{i \in I^c_{mask}} [P(R_i = 1|k_i = j, t_i, M_i = \{j\}) p(M_i = \{j\}|k_i = j, t_i) f_{T,K}(t_i, j)] \times \prod_{i \in I_{mask}} \prod_{j=1}^{J} [P(R_i = 0|k_i = j, t_i, M_i = \{j\}) p(M_i = \{j\}|k_i = j, t_i) f_{T,K}(t_i, j)]^{I_{ij}} \quad (9)$$

$$L_R(\theta) \propto \prod_{i=1}^{r} \left[f(R_i|t_i) \prod_{j=1}^{J} [p_j(M_i) f_{T,K}(t_i, j)]^{I_{ij}} \right] \quad (10)$$

Because of conjugacy a suitable prior for$p_j$is the Dirichletdistribution,$D(\gamma_j)$,where$\gamma_j$isa2^{J-1}dimensional vector. The choice of prior distributions for other parameters will be s-dependent on theCDF that is considered forT_l.

If $\pi(\theta)$,$\pi(\beta)$ and$\pi_l(p_l)$ be the priors for parametersθ,$\beta$and$p_l$respectively, then the joint density function of (t,M,I,R) isresulted as

$$p(t, M, I, R) = p(\theta, p, \beta|t, M, I, R) \pi(\theta) \pi(\beta) \prod_{l=1}^{J} \pi_l(p_l) \quad (11)$$

The full conditional posterior distribution of$p_l$is also a Dirichlet distribution, but its parameters depend on the observations. The full conditional distribution of$I_{ij} = 1, j \in M_i$is

$$P(I_{ij} = 1 \mid t, M, R, \beta, \theta, p, I_{(-ij)}) = \frac{P(R_i|t_i, K_i = j) p_j(M_i) f_{T,K}(t_i, j)}{\sum_{m \in M_i} P(R_i|t_i, K_i = m) p_m(M_i) f_{T,K}(t_i, m)} = \frac{P(R_i|t_i, K_i = j) p_j(M_i) f_j(t_i) \prod_{l \neq j} R_l(t_i)}{\sum_{m \in M_i} P(R_i|t_i, K_i = m) p_m(M_i) f_m(t_i) \prod_{l \neq m} R_l(t_i)} = \frac{P(R_i|t_i, K_i = j) p_j(M_i) h_j(t_i)}{\sum_{m \in M_i} P(R_i|t_i, K_i = m) p_m(M_i) h_m(t_i)} \quad (12)$$

where$I_{(-ij)}$is$I$excluding$I_{(ij)}$and$h_j(t_i) = \frac{f_j(t_i)}{1-F_j(t_i)};\ j = 1, 2, \ldots, J$.

As a especial case, the likelihood function for exponential distribution with parameter$\alpha_l$forl^{th} component based on (8) isobtained as follows:

$$L(\theta, p, \beta \mid t, M, R) = \prod_{i=1}^{r} \left[\prod_{j=1}^{J} P(R_i|t_i, M_i, K_i = j)^{I_{ij}} p_j(M_i)^{I_{ij}} \alpha_j^{I_{ij}} \right] \times exp\{-r \sum_{i=1}^{r} \sum_{l=1}^{J} \alpha_l t_i\} \quad (13)$$

And the likelihood function for the Weibull distribution with parameters (μ_l, β_l)forl^{th} component based on (8) is as follows:

$$L(\theta, p, \beta \mid t, M, R) = \prod_{i=1}^{r} \left[\frac{\prod_{j=1}^{J} P(R_i|t_i, M_i, K_i = j)^{I_{ij}} p_j(M_i)^{I_{ij}} b_j^{I_{ij}} t_i^{(b_j - 1) I_{ij}}}{\mu_j^{b_j I_{ij}}} \right] \times exp\{-\sum_{i=1}^{r} \sum_{l=1}^{J} (\frac{t_i}{\mu_l})^{b_l}\} \quad (14)$$

Numerical Example

In this section, we illustrate the application of the proposed methods by two simulation data sets and a real data.

Exponential Distribution

We consider 100 series systems with two components where the lifetime of components follows the exponential distribution with parameters$\alpha_1$and$\alpha_2$for the first and second component, respectively. 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 masking probabilities of the data are$p_1$andp_2,where$p_1 = p_1(\{1,2\})$ and$p_2 = p_2(\{1,2\})$. Let$\alpha_1 = 0.3, \alpha_2 = 0.7, \beta_0 = -0.1, \beta_1 = 0.5, p_1 = 0.1, p_2 = 0.2$. The simulated data are generated by thefollowing steps.

1. Generate$(T_1^{(i)}, T_2^{(i)});\ i = 1, 2, \ldots, r$ independently,from$exp(\alpha_1)$ and$exp(\alpha_2)$, respectively.
2. Set $T_i = \min(T_1^{(i)}, T_2^{(i)});\ i = 1, 2, \ldots, r$as the failure time of thei^{th} system and then specify the cause of failure such that if the first

component causes to failure putk=1 if not putk=2.

3. Generate$R_i$from Bernoulli distribution with success probability$p(R_i = 1|k_i = j)$, such thatlogit$(p(R_i = 1|k_i = j)) = \beta_0 + \beta_1 k_i$.
4. If$R_i = 0$, that isith masked set have more than one element, and if failure cause is first (second)component, we randomly masked $100p_1\%(100p_2\%)$ of the observations.

The simulated data are listed in Table 1.

Table1.Thesimulated data

(t,k,R)
(1.501,1,1),(2.386,1,1),(0.849,2,1),(2.223,1,0),(3.055,1,1),(0.444,0,0),(1.878,1,0),(4.101,2,1),(1.965,2,0), (0.252,1,0),(1.938,2,1),(0.221,1,0),(0.282,1,1),(2.006,2,1),(0.511,2,1),(3.057,1,1),(0.563,2,1),(1.308,2,1), (0.160,1,0),(1.341,2,1),(2.546,2,1),(0.040,0,0),(1.863,1,0),(1.207,2,1),(1.497,1,1),(1.717,0,0),(0.705,2,1), (3.170,1,1),(0.067,1,1),(1.808,1,1),(0.319,2,0),(3.444,2,1),(0.676,2,1),(0.566,2,1),(0.960,1,1),(0.299,0,0), (2.111,0,0),(0.210,1,1),(0.433,2,1),(0.868,2,1),(0.275,2,1),(2.029,2,0),(3.218,2,1),(0.584,1,1),(1.221,2,1), (0.224,0,0),(0.485,1,1),(0.333,0,0),(0.919,2,1),(0.209,2,1),(0.816,1,1),(1.488,2,1),(1.234,2,1),(1.792,0,0), (1.681,2,1),(0.291,2,1),(0.815,1,0),(0.444,2,1),(2.776,2,1),(0.718,1,0),(0.847,2,1),(1.362,2,1),(2.438,2,0), (1.735,2,1),(1.481,2,1),(0.471,2,1),(0.545,0,0),(0.688,1,1),(1.489,2,1),(2.274,1,0),(1.095,1,1),(0.265,2,1), (0.166,2,1),(0.557,1,1),(0.181,2,1),(0.544,2,0),(2.207,2,1),(0.246,2,1),(0.645,1,1),(0.095,1,0),(0.090,2,1), (0.195,2,0)(0.486,2,1),(0.203,2,1),(0.215,2,1),(0.248,2,1),(1.310,2,1),(0.826,2,1),(0.198,2,1),(1.634,0,0), (0.689,2,1),(0.357,1,0),(3.419,0,0),(1.148,2,0),(0.607,2,1),(1.249,2,1),(1.259,1,1),(0.921,2,1),(0.071,2,1),(2.169,2,1)

The results of MLEs of the parameters α_1 and $\alpha_2$have been presented in Table 2. The true values of the parameters as well as the corresponding bias ofα_1 and α_2 (denoted by $B\alpha_1$ and $B\alpha_2$, respectively) based on the 1000 iterations have been reported in the Table 2.

Table 2. The MLE results of simulation analysis

	α_1	α_2	β_0	β_1	$B\alpha_1$	$B\alpha_2$
MAR MNAR	0.3	0.7	-0.1	0.5	0.038 0.023	0.033 0.018

According to the results, MNAR model leads to less biased estimators compared with the usual MAR model. Now we consider some proper priors for parameters at the MNAR model and obtain Bayesian estimates for the parameters using MCMC method with masking probabilities$p_1 = 0.1, p_2 = 0.2$ and true values $\alpha_1 = 0.3$, $\alpha_2 = 0.7$. We consider the following prior setting

$$\alpha_1 \sim gamma(0.9, 3),\ \alpha_2 \sim gamma(0.49, 0.7),\ \beta_0 \sim norm(-0.1, 1000),$$
$$\beta_1 \sim norm(0.5, 1000),\ p_1 \sim Beta(0.8, 7.2),\ p_2 \sim Beta(0.01, 0.05) \quad (15)$$

Using 10,000 iterations of Gibbs sampling with burn-in 2,000 iterations and length of the thinning interval 5, the posterior estimates of the parameters based on (15) and 1,600 posterior samples are listed in Table 3.

Table 3. posterior estimates of parameters

Parameter	True Value	mean	SD	LCI	UCI
p_1	0.1	0.388	0.079	0.241	0.545
p_2	0.2	0.340	0.058	0.224	0.460
α_1	0.3	0.295	0.056	0.195	0.415
α_2	0.7	0.606	0.075	0.474	0.756
b_0	-0.1	-0.138	0.301	-0.199	-0.077
b_1	0.5	0.418	0.028	0.364	0.471

Also, standard deviation (SD), lower bound (LCI) and upper bound of credible interval are calculated. To avoid of randomness effects the simulation has been repeated 200 times with masking probabilities (p_1, p_2) = (0.3,0.5), (0.7,0.3), and (0.8,0.8). The results are given in Table 4. In Table 4, Mean is referred to the average posterior estimates of model parameters and SRMSE is referred to the square root of the mean squared errors. As we expected, it is observed that as masking probability increases the SRMSE becomes larger.

Table 4. The posterior estimates based on non-informative priors with 200 replications

Masking probability	Statistics	p_1	p_2	α_1	α_2	b_0	b_1
(0.1,0.2)	Mean	0.367	0.274	0.317	0.653	-0.143	0.412

	SRMSE	0.278	0.091	0.058	0.084	0.043	0.088
(0.3,0.5)	Mean	0.345	0.274	0.312	0.651	-0.141	0.415
	SRMSE	0.093	0.233	0.059	0.093	0.041	0.085
(0.7,0.3)	Mean	0.214	0.349	0.246	0.714	-0.137	0.419
	SRMSE	0.492	0.073	0.076	0.083	0.037	0.081
(0.8,0.8)	Mean	0.160	0.359	0.229	0.737	-0.137	0.419
	SRMSE	0.643	0.445	0.084	0.086	0.037	0.082

Weibull Distribution

Similar to the previous subsection, here we consider100 series systems with two components where the lifetime of components follows the Weibull distribution with parameters (μ_1, b_1) and (μ_2, b_2) for the first and second components, respectively. Suppose$b_1 = 1.5, b_2 = 0.8, \mu_1 = \mu_2 = 1, \beta_0 = -0.5, \beta_1 = 1, p_1 = 0.1,$ and $p_2 = 0.2.$ The results of MLEs of the parametersb_1, b_2, μ_1 and μ_2 have been presentedin Table 5. True values of parameters and the corresponding bias of b_1, b_2, μ_1 and $\mu_2$based on 1000 iterations are also given in Table 5. According tothe results, MNAR model has less bias compared with the usual MAR model.

Table 5. The MLE results of simulation analysis

	b_1	b_2	μ_1	μ_2	β_0	β_1	Bb_1	Bb_2	$B\mu_1$	$B\mu_2$
MAR	1.5	0.8	4	2	-0.5	1	0.114	0.013	0.509	0.451
MNAR							0.102	0.009	0.455	0.411

For Bayesian inference, we consider the following prior setting

$$b_1 \sim gamma(0.002, 0.001),\ b_2 \sim gamma(0.0006, 0.0008),\ \beta_0 \sim norm(-0.5, 1000),$$
$$\beta_1 \sim norm(1,1000), p_1 \sim Beta(0.8,7.2), p_2 \sim Beta(0.01,0.05), \qquad (16)$$
$$\mu_1^{b_1} \sim IG(0.001, 0.001),\ \mu_2^{b_2} \sim IG(0.001, 0.001)$$

The simulation has been repeated 200 times in order to avoid of randomness effects. Different cases of masking probabilities have been considered such as, $(p_1, p_2) = (0.3,0.5)$, (0.7,0.3), and (0.8,0.8). Based on the obtained results in Table 7, as the masking probability increases SRMSE becomes larger.

Table 6. Posterior estimates of parameters

Parameter	True Value	mean	SD	LCI	UCI
b_1	1.5	1.514	0.158	1.215	1.829
b_2	0.8	1.03	0.144	0.770	1.343
μ_1	1	0.799	0.072	0.674	0.963
μ_2	1	1.28	0.249	0.925	1.894
β_0	-0.5	-0.497	0.031	-0.559	-0.435
β_1	1	0.999	0.029	0.943	1.059
p_1	0.1	0.201	0.033	0.140	0.269
p_2	0.2	0.251	0.032	0.192	0.316

Table 7. The posterior estimates based on non-informative priors with 200 replicationsReal Data

Masking probability	Statistics	b_1	b_2	μ_1	μ_2	β_0	β_1	p_1	p_2
(0.1,0.2)	Mean	1.521	0.795	0.975	1.161	-0.498	0.998	0.184	0.271
	SRMSE	0.162	0.093	0.098	0.333	0.004	0.003	0.087	0.073

(0.3,0.5)	Mean	1.523	0.799	0.985	1.169	-0.500	0.998	0.300	0.493
	SRMSE	0.177	0.107	0.111	0.374	0.004	0.003	0.011	0.013
(0.7,0.3)	Mean	1.538	0.836	1.083	0.971	-0.498	1.002	0.623	0.330
	SRMSE	0.210	0.107	0.167	0.191	0.004	0.004	.078	.033
(0.8,0.8)	Mean	1.579	0.848	1.135	0.902	-0.495	1.004	0.683	0.743
	SRMSE	0.230	0.108	0.186	0.192	0.006	0.005	0.118	0.059

To motivating our study, we consider the real dataset given in Levuliene[8] from a test recorded bus tire failure times (T) and corresponding cause of failure (V). In this data, we ignored soft failures and randomly masked $100p_1$ percent and $100p_2$ percent of those that failed due to first and second competing risks, such that p_1= 0.1 and p_2= 0.2. A Weibull distribution was fitted to these data, also for implementation logit model we considered$\beta_0 = -0.5, \beta_1 = 1$ and MLEs of parameters based on (14) have been presented in Table 8.

Table 8. The MLE results of simulation analysis

	b_1	b_2	μ_1	μ_2	β_0	β_1	p_1	p_2
MAR	7.013	8.999	67.711	71.395	-	-	0.388	0.557
MNAR	7.003	8.947	67.907	71.160	-0.286	0.796	0.375	0.571

Using bellow non-informative priors, posterior estimates are presented in Table 9.

Table 9. Posterior estimates of parameters

Parameter	True Value	mean	SD	LCI	UCI
b_1	7	6.361	0.644	5.215	7.754
b_2	9	8.963	1.043	6.947	10.93
μ_1	70	68.64	2.008	65.04	72.63
μ_2	70	71.04	1.789	67.96	75.05
β_0	-0.5	-0.497	0.031	-0.555	-0.432
β_1	1	1.001	.032	0.0940	1.066
p_1	0.1	0.303	.073	0.172	0.458
p_2	0.2	0.592	0.103	0.380	0.776

$$b_1 \sim gamma(0.049, 0.007),\ b_2 \sim gamma(0.081, 0.009),\ \beta_0 \sim norm(-0.5, 1000),$$
$$\beta_1 \sim norm(1,1000), p_1 \sim Beta(0.8,7.2), p_2 \sim Beta(0.01,0.05), \quad (17)$$
$$\mu_1^{b_1} \sim IG(1.474E-13, 1.214),\ \mu_2^{b_2} \sim IG(6.141E-17, 2.478)$$

Conclusion

In this paper, we have introduced a new approach for handle masked data. We proposeda generalizedlinear model to conduct relationship between masking probability and exact cause of failure using a binary variable. The simulation results show that the proposed method provides good estimations for modelparameters under both maximum likelihood and Bayesian methods.

References

[1] Miyakawa, M. “Analysis of incomplete data in a competing risks model,” IEEE Trans. Rel. 33(4): pp. 293–296, 1984.

[2] Basu, S. Basu, A. P. and Mukhopadhyay, C. “Bayesian analysis for masked system failuredata using nonidenticalweibull models,” J. Statist. Plann. Inference, vol. 78, pp. 255–275, 1999.

[3] AziziF. and Haghighi,F. Joint modeling of linear degradation and failure time data with masked causes of failure under simple step-stress test. J Stat Comput Simul. 88(8):1603–1615, 2018. doi: 10.1080/00949655.2018.1442468.

[4] Usher, J. S. and Hodgson, T. J. “Maximum likelihood analysis of component reliability usingmasked system life data,” IEEE Trans. Rel., vol. 37, no. 5, pp. 550–555, 1998.

[5] Lin, D. K. J. Usher, J. S. and Guess, F. M. “Exact maximum likelihood estimation usingmasked system data”, IEEE Trans. Rel., vol. 42, no. 4, pp. 631–635, 1993.

[6] Reiser, B. Guttman, I. Lin, D. K. J. Usher, J. S. and Guess, F. M. “Bayesian inference formasked system lifetime data,” Appl. Statist., vol. 44, pp. 79–90, 1995.

[7] Berger, J. O. and Sun, D. “Bayesian analysis for the poly–Weibull distribution,” J. Amer.Statist. Assoc., vol. 88, pp. 1412–1418, 1993.

[8] Mukhopadhyay, C. and Basu, A. P. “Bayesian analysis of incomplete time and cause offailure data,” J. Statist. Plann. Inference, vol. 59, pp. 79–100, 1997.

[9] Cai, J., Shi, Y., Zhang, Y. Robust Bayesian analysis for parallel system with masked data under inverse Weibull lifetime distribution. Commun Stat-Theory Methods. 49:1–13, 2019.

[10] Sen, A. Banerjee, M. and Basu, S. Balakrishnan, N. and Rao, C. R. Eds. “Analysis ofmasked failure data under competing risks,” in Handbook of Statistics. Amsterdam, The Nether-lands: North-Holland, vol. 20, pp. 523–540, 2001.

[11] Basu, S. Sen, A. and Banerjee, M. “Bayesian analysis of competing risks with partiallymasked

cause of failure," Appl. Statist., vol. 52, pp. 77–93, 2003.

[12] Lin D. K. J. and Guess, F. M. "System life data analysis with dependent partial knowledgeon the exact cause of system failure", Microelectron. Rel., vol. 34, pp. 535–544, 1994.

[13] Guttman, I. Lin, D. K. J. Reiser, B. and Usher, J. S. "Dependent masking and systemlife data analysis: Bayesian inference for two-component systems," Lifetime Data Anal., vol. 1, pp.87–100, 1995.

[14] Kuo, L. and Yang, T. E. "Bayesian reliability modeling for masked system lifetime data",Statist. Probab. Lett., vol. 47, pp. 229–241, 2000.

[15] Mukhopadhyay, C. and Basu, A. P. "Masking without the symmetry assumption: A Bayesianapproach," in Proc. Abstract Book 2nd Int. Conf. Math. Methods Rel., Bordeaux, France, vol. 2,pp. 784–787, 2000,Universite Victor Segalen.

[16] Craiu, R. V. and Duchesne, T. Gelman, A. and. Meng, Eds. X. L. "Using EM and DA for thecompeting risk model," in Applied Bayesian Modeling and Causal Inference from an Incomplete-Data Perspective. New York, NY, USA: Wiley, pp. 234–245, 2004.

[17] [17] Mukhopadhyay, C. "Maximum likelihood analysis of masked series system lifetime data," J.Statist. Plann. Inference, vol. 136, pp. 803–838, 2006.

[18] Xu, A. and Tang, Y. "Bayesian analysis of Pareto reliability with dependent masked data,"IEEE Trans. Rel., vol. 58, no. 4, pp. 583–588, 2009.

[19] Xu, A. Basu, S. and Tang, Y." A Full Bayesian Approach for Masked Data in Step-StressAccelerated Life Testing," IEEE Trans. Rel., vol. 63, no. 3, pp. 798-806, 2014.