Statistics and Its Interface Volume 8 (2015) 227–237

Objective Bayesian analysis for masked data under symmetric assumption

Ancha Xu*, Yincai Tang, and Dongchu Sun

In this paper, we consider an exponential model with masked data. We show that the parameters are nonidentifiable under a general masking probability assumption, and under symmetric assumption find a prior based on which the posterior means of parameters coincide with their MLEs. The Jeffreys prior and the reference prior are also derived under symmetric assumption. Propriety of the posteriors under the Jeffreys prior and the reference prior is assessed. When the hazard function of the series system is of interest, a reparametrization is considered, and we derive Jeffreys prior and the reference prior under the reparametrization. Then the frequentist coverage probabilities of the α-quantiles of the marginal posterior distributions of the parameters are obtained. The simulation study shows that the reference prior performs better than the Jeffreys prior in meeting the target coverage probabilities.

AMS 2000 subject classifications: Primary 62F15; secondary 62F12.
Keywords and phrases: Masked data, Reference prior, Jeffreys prior, Series system, Exponential distribution.

1. INTRODUCTION

Masked data, a complex variant of competing risk data, is becoming more prominent in reliability studies, medical diagnostic studies and biological systems. With a competing risk data, each failure time is associated to a known cause of failure, whereas for masked data, the causes for a failure may be unknown (masked) for a group of subjects. In this paper, we assume that these masked failure times are known up to a subset of all causes of failures, the so-called Minimum Random Subset (MRS) ([10]). Denote MRS as M. When M is a singleton, the masked data reduces to a competing risk data. When M is the set of all causes of failures, the data is known to be completely masked ([3]). Note that the set M varies from subject to subject. There are many reasons that lead to masking in the data. The most common reasons are: (i) the lack of proper diagnostic equipments, (ii) the cost and time constraints associated to the data collection, (iii) recording errors, and (iv) the destructive nature of certain failed components that prevents an exact diagnostic to take place.

*Corresponding author.

Consider n identical series systems, each with J components. Let the random variable X_{ij} be the lifetime of the j-th component of the i-th series system. Then the lifetime of the i-th series system is given by $Z_i = \min\{X_{i1}, \cdots, X_{iJ}\}$. Due to censoring, the observed data from the i-th series system reduces to $(t_i;\ M_i;\ C_i)$, $i = 1, \ldots, n$, where t_i is the failure time of the i-th series system, M_i denote the MRS corresponding to the i-th system, and the binary variable C_i captures whether the observed failure time of the i-th system is censored ($C_i = 0$) or not ($C_i = 1$). The setting of the censoring scheme is very general, including type-II censoring and progressively type-II censoring. We denote the observed data as $(\boldsymbol{t}, \boldsymbol{M}, \boldsymbol{C})$. Then the likelihood function is

$$
\begin{aligned}
(1)\quad & L((\boldsymbol{t}, \boldsymbol{M}, \boldsymbol{C})|\boldsymbol{\theta}) \\
= & \prod_{i=1}^{n}\left\{\left[\sum_{j\in M_i}\Pr(M = M_i|T = t_i, K = j)\right.\right. \\
& \left.\left. \times f_j(t_i|\boldsymbol{\theta}_j)\prod_{k\neq j}R_k(t_i|\boldsymbol{\theta}_k)\right]^{C_i}\right\}\prod_{i=1}^{n}\left[\prod_{k=1}^{J}R_k(t_i|\boldsymbol{\theta}_k)\right]^{1-C_i} \\
= & \prod_{i=1}^{n}\left\{\left[\sum_{j\in M_i}\Pr(M = M_i|T = t_i, K = j)h_j(t_i|\boldsymbol{\theta}_j)\right]^{C_i}\right. \\
& \left. \times\prod_{k=1}^{J}R_k(t_i|\boldsymbol{\theta}_k)\right\},
\end{aligned}
$$

where $h_j(t_i|\boldsymbol{\theta}_j) = f_j(t_i|\boldsymbol{\theta}_j)/R_j(t_i|\boldsymbol{\theta}_j)$ is the hazard rate function of X_{ij}, K is a latent variable describing the true cause of failure of the system, and we assume the observed MRS always includes the true cause of failure. Thus, when M_i is a singleton, $K = M_i$. $P(M = M_i|T = t_i, K = j)$ in (1) is the masking probability. The most used assumption of the masking probability is symmetric assumption (also called equiprobable assumption by some authors), that is, $\forall j \in M_i$,

$$(2)\quad P(M = M_i|T = t_i, K = j) = p(M_i),\ i = 1, \ldots, n.$$

In other words, under the symmetric assumption, the masking probability does not depend on the failure time and the true cause of failure. And the likelihood (1) reduces to

$$\begin{aligned}(3)\qquad & L_1((\boldsymbol{t},\boldsymbol{M},\boldsymbol{C})|\boldsymbol{\theta},\boldsymbol{P}_1) \\ = & \prod_{i=1}^{n} p(M_i)\left\{\left[\sum_{j\in M_i} h_j(t_i|\boldsymbol{\theta}_j)\right]^{C_i} \prod_{k=1}^{J} R_k(t_i|\boldsymbol{\theta}_k)\right\},\end{aligned}$$

where $\boldsymbol{P}_1$ is a vector with distinct $p(M_i)$, $i = 1, 2, \ldots, n$. Most of the earlier works in masked data analysis subject the masking probability to this symmetric assumption. [15] was the first one who considered the problem of a two-component series system when the lifetime of the system's components followed exponential distributions and the maximum likelihood estimators for the parameters were obtained based on the masked data. [20] extended Miyakawa's results to a three-component series system under the same assumption. [10], Lin et al. (1993, 1996) further extended these results when Weibull distribution was assumed. Reiser et al. (1995) provided a Bayesian analysis for the case of [20]. [3] discussed Bayesian inference with Weibull distributions for the system's components under complete masking and an extension to partial masking cases was studied by [17], Basu et al. (1999) and Basu et al. (2003). [23] utilized a nonparametric Bayesian method to estimate the survival function of the series system when the data was masked.

A generalization of symmetric assumption is to assume that the masking probability is independent of failure time, but depends on the cause of failure, that is

$$\begin{aligned}(4)\qquad & P(M = M_i|T = t_i, K = j) = p_j(M_i), \\ & j \in M_i,\ i = 1, 2, \ldots, n.\end{aligned}$$

Thus (1) can be written as

$$\begin{aligned}(5)\qquad & L_2((\boldsymbol{t},\boldsymbol{M},\boldsymbol{C})|\boldsymbol{\theta},\boldsymbol{P}_2) \\ = & \prod_{i=1}^{n}\left\{\left[\sum_{j\in M_i} p_j(M_i)h_j(t_i|\boldsymbol{\theta}_j)\right]^{C_i} \prod_{k=1}^{J} R_k(t_i|\boldsymbol{\theta}_k)\right\},\end{aligned}$$

where $\boldsymbol{P}_2$ is a vector with distinct $p_j(M_i)$, $j \in M_i$, $i = 1, 2, \ldots, n$. Under this assumption, [12] developed a Bayesian analysis for two-component systems with both independent exponential and Weibull component lifetimes. [16] used the maximum likelihood method to estimate both the lifetime parameters and masking probabilities via an EM algorithm, and constructed approximate confidence intervals, further corrected them by bootstrap method. [22] proposed a Bayesian analysis for two-component systems with Pareto distribution lifetime. Xu et al. (2014b) considered Bayesian analysis of masked data in step-stress accelerated life testing. Xu et al. (2014a) proposed a full Bayesian method for analyzing masked data in step-stress accelerated life testing.

All the literature referenced so far can be split into two broad methodologies: the classical approach and the Bayesian approach, each having its advantages and drawbacks. In the classical approach, large-sample asymptotic methods are heavily relied upon to construct confidence intervals for the parameters of interest. Subjective Bayesian methods do not rely on the normal approximation and are known to work very well. However, the process of eliciting the prior distribution may not be easy to determine even in presence of historical data or the experience of experts. With limited time and little knowledge about the hyperparameters, the obtained priors could be quite bad. See [4]. Instead, the objective Bayesian approach is an alternative. The main spirit of the objective Bayesian approaches is the use of the noninformative prior distributions, and the Jeffreys prior and the reference prior are the two most used often noninformative priors. For more details, see [11], [6], [5] and [9]. To the best of our knowledge, all the literature about masked data is not devoted to objective Bayesian method. Thus, we will consider objective Bayesian method to analyze masked data in this paper, and the lifetime of each component is assumed to be an exponential distribution. In Section 2, we prove that the parameters of the likelihood function in (5) are nonidentifiable. Under the symmetry property, we present an improper prior under which the posterior means of the model parameters coincide with the MLEs. However, this prior can lead a posterior that is not proper. Thus, in Section 3, we derive the Jeffreys and the reference priors and show that their corresponding posterior distributions are always proper. Then we derive the frequentist coverage probabilities of the α posterior quantiles of parameters in the model. Small sample comparison of the noninformative priors is performed in Section 4. Finally, some concluding remarks and discussions are made in Section 5.

2. MAXIMUM LIKELIHOOD ESTIMATORS

2.1 Model

Suppose that a series system has two exponential components, that is, $X_1 \sim Exp(\lambda_1)$, $X_2 \sim Exp(\lambda_2)$. This model is considered by [15] and [12]. [15] derived maximum likelihood estimates of λ_1 and λ_2 under the symmetric assumption, while [12] proposed Bayesian method to obtain estimation of model parameters under the assumption (4). The likelihood function under (4) simplifies to

$$\begin{aligned}(6)\quad L \;=\; & (1-p_1)^{r_1}(1-p_2)^{r_2}\lambda_1^{r_1}\lambda_2^{r_2}(p_1\lambda_1 + p_2\lambda_2)^{r_3} \\ & \times \exp\{-(\lambda_1+\lambda_2)T\},\end{aligned}$$

where $p_1 = P(M = \{1, 2\} \mid K = 1)$, $p_2 = P(M = \{1, 2\} \mid K = 2)$, $T = \sum_{i=1}^{n} t_i$, r_1 and r_2 are the number of system failures due to component one and two, respectively, r_3 denotes the number of failures masked and $r_1 + r_2 + r_3 = r$.

Theorem 2.1. *Under the assumption* (4)*, the parameters in the* L *are nonidentifiable.*

See the proof in the Appendix. Thus, the frequentist method cannot provide a viable solution to this problem. For the Bayesian approach, it will work well if there is precise prior information of the parameters available. However, the precise prior information of the parameters are hard to collect. Thus, under the general assumption (4), both the frequentist and Bayesian methods fail to deal with the problem well. Flehinger et al. (2002) added second stage information (providing definitive diagnosis for part of the masked causes), and successfully estimated all the parameters in the model by maximum likelihood method. Sen et al. (2010) used covariates as additional information, and assumed a logistic structure between masking probability and covariates, then obtained the estimates of the parameters in the model. However, additional information is not always available. Another way to estimate these parameters is under symmetric assumption. Under symmetric assumption, the likelihood function (6) is reduced to

$$(7)\qquad L_1 = p^{r_3}(1-p)^{r_1+r_2}\lambda_1^{r_1}\lambda_2^{r_2}(\lambda_1+\lambda_2)^{r_3} \times \exp\{-(\lambda_1+\lambda_2)T\},$$

where $p = P(M = \{1,2\})$. (7) is the model we consider in this paper. Without loss of generality, we assume a type-II censoring mechanism for the failure times and that $r_3 > 0$, since $r_3 = 0$ means that there is no data masked.

2.2 Connections between MLEs and Bayesian estimators

The MLEs of the parameters in (7) are

$$(8)\qquad \hat{p} = \frac{r_3}{r},\quad \hat{\lambda}_1 = \frac{r}{T}\frac{r_1}{r_1+r_2},\quad \hat{\lambda}_2 = \frac{r}{T}\frac{r_2}{r_1+r_2}.$$

The MLEs are very intuitive and can be obtained when the population means are replaced by the sample means, and they also have some nice properties.

Proposition 2.1. *1.*

$$(r_1,r_2,r_3) \sim \textit{Multinomial}\left(r, \frac{(1-p)\lambda_1}{\lambda_1+\lambda_2}, \frac{(1-p)\lambda_2}{\lambda_1+\lambda_2}, p\right).$$

2. T follows the gamma distribution with mean $r/(\lambda_1+\lambda_2)$ and variance $r/(\lambda_1+\lambda_2)^2$.
3. (r_1,r_2,r_3) is independent of T.

Thus $1/T$ follows the inverse gamma distribution with mean $(\lambda_1+\lambda_2)/(r-1)$ and variance $(\lambda_1+\lambda_2)^2/[(r-1)^2(r-2)]$ $(r > 2)$, and $rr_1/(r_1+r_2)$ follows $Binomial(r, \lambda_1/(\lambda_1+\lambda_2))$.

Proposition 2.2. *$\hat{p}$ is unbiased, and $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are asymptotically unbiased.*

See the proof in the Appendix.

From Proposition 2.2, we can obtain that the unbiased estimates (UEs) of λ_1 and λ_2 are

$$\tilde{\lambda}_1 = \frac{r-1}{T}\frac{r_1}{r_1+r_2} \text{ and } \tilde{\lambda}_2 = \frac{r-1}{T}\frac{r_2}{r_1+r_2},$$

respectively. And

$$\mathrm{Var}(\tilde{\lambda}_1) = \frac{\lambda_1[(r-1)\lambda_2 + r\lambda_1]}{r(r-2)},$$

$$\mathrm{Var}(\tilde{\lambda}_2) = \frac{\lambda_2[(r-1)\lambda_1 + r\lambda_2]}{r(r-2)}.$$

There are two drawback of the frequentist method: (i) the constraint of the number of failure $(r > 2)$, (ii) $r_1, r_2 > 0$. When $r_1 = 0$ or $r_2 = 0$, the MLEs and UEs of λ_1 or λ_2 will be 0, which always underestimates the parameter. As an alternative, Bayesian method can be used in this problem. To perform Bayesian analysis, we should assign prior for $(p, \lambda_1, \lambda_2)$. A natural prior for p is $U(0,1)$. Besides, we choose $1/(\lambda_1\lambda_2)$ as the prior of λ_1 and λ_2. Thus the prior of $(p, \lambda_1, \lambda_2)$ is

$$\pi_1(p, \lambda_1, \lambda_2) \propto 1/(\lambda_1\lambda_2).$$

$\pi_1(p, \lambda_1, \lambda_2)$ is an improper prior. Thus, the posterior propriety needs to be justified. See the following results.

Theorem 2.2. *Under the prior $\pi_1(p, \lambda_1, \lambda_2)$, if $r_1 > 0$ and $r_2 > 0$,*

1. *Then the posterior distribution of $(p,\ \lambda_1,\ \lambda_2)$ is proper.*
2. *The posterior of p is $B(r_3+1, r_1+r_2+1)$.*
3. *The posterior means of λ_1 and λ_2 are exact the same as their MLEs.*

See the proof in the Appendix. In Theorem 2.2, a condition is needed to make the posterior distribution proper. However, the condition is not always guaranteed, especially in the case of small sample size. Thus the prior $\pi_1(p, \lambda_1, \lambda_2)$ is not recommended, though the posterior means coincide with MLEs. Also, if $r_1 = 0$ or $r_2 = 0$, the MLEs will fail, since $\hat{\lambda}_1 = 0$ or $\hat{\lambda}_2 = 0$, which significantly underestimate λ_1 or λ_2. In the next section, two noninformative priors will be derived to overcome this problem.

3. NONINFORMATIVE PRIORS AND POSTERIOR ANALYSIS

It is not difficult to show that the Fisher information matrix of $(p, \lambda_1, \lambda_2)$ is

$$(9)$$

$$I(p,\lambda_1,\lambda_2) = \begin{bmatrix} \frac{r}{p(1-p)} & 0 & 0 \\ 0 & \frac{r(\lambda_1+\lambda_2) - rp\lambda_2}{\lambda_1(\lambda_1+\lambda_2)^2} & \frac{rp}{(\lambda_1+\lambda_2)^2} \\ 0 & \frac{rp}{(\lambda_1+\lambda_2)^2} & \frac{r(\lambda_1+\lambda_2) - rp\lambda_1}{\lambda_2(\lambda_1+\lambda_2)^2} \end{bmatrix}.$$

Then we have

$$|I(p,\lambda_1,\lambda_2)| = \frac{r^3}{p\lambda_1\lambda_2(\lambda_1+\lambda_2)^2}.$$

Thus the Jeffreys prior is

$$\pi_J(p,\lambda_1,\lambda_2) \propto p^{-1/2}\lambda_1^{-1/2}\lambda_2^{-1/2}(\lambda_1+\lambda_2)^{-1}.$$

Theorem 3.1. *(i) When the group order is $\{(\lambda_1, \lambda_2), p\}$, the reference prior is*

$$\pi_R(p, \lambda_1, \lambda_2) \propto p^{-1/2}(1-p)^{-1/2}\lambda_1^{-1/2}\lambda_2^{-1/2}(\lambda_1+\lambda_2)^{-1}.$$

(ii) When the group order is $\{p, (\lambda_1, \lambda_2)\}$, the reference prior is identical to the Jeffreys prior.

See the proof in the Appendix. Since the Jeffreys prior and the reference prior are improper, we should justify the posterior propriety of the parameters.

Theorem 3.2. *The posterior distributions of $(p, \lambda_1, \lambda_2)$ based on $\pi_R(p, \lambda_1, \lambda_2)$ or $\pi_J(p, \lambda_1, \lambda_2)$ are always proper.*

See the proof in the Appendix. The results of Theorem 3.2 are very interesting, since no matter how small the number of the failures, we can always use the two noninformative priors to do statistical inference. Based on $\pi_R(p, \lambda_1, \lambda_2)$ or $\pi_J(p, \lambda_1, \lambda_2)$, the joint posterior density functions of $(p, \lambda_1, \lambda_2)$ are

$$\begin{aligned}\pi_R(p, \lambda_1, \lambda_2|\boldsymbol{t}) &\propto (1-p)^{r_1+r_2-1/2}p^{r_3-1/2}\lambda_1^{r_1-1/2}\lambda_2^{r_2-1/2}\\ &\quad\times(\lambda_1+\lambda_2)^{r_3-1}\exp\{-(\lambda_1+\lambda_2)T\},\\ \pi_J(p, \lambda_1, \lambda_2|\boldsymbol{t}) &\propto (1-p)^{r_1+r_2-1}p^{r_3-1/2}\lambda_1^{r_1-1/2}\lambda_2^{r_2-1/2}\\ &\quad\times(\lambda_1+\lambda_2)^{r_3-1}\exp\{-(\lambda_1+\lambda_2)T\}.\end{aligned}$$

Theorem 3.3. *Under the priors $\pi_R(p, \lambda_1, \lambda_2)$ and $\pi_J(p, \lambda_1, \lambda_2)$, both the marginal posterior cumulative distribution of λ_1 are*

(10)

$$\begin{aligned}&F_1(\lambda_1|\boldsymbol{t})\\ &= \frac{\int_0^{\lambda_1}\int_0^\infty x^{r_1-1/2}\lambda_2^{r_2-1/2}(x+\lambda_2)^{r_3-1}\exp\{-(x+\lambda_2)T\}d\lambda_2\,dx}{\int_0^\infty\int_0^\infty x^{r_1-1/2}\lambda_2^{r_2-1/2}(x+\lambda_2)^{r_3-1}\exp\{-(x+\lambda_2)T\}d\lambda_2\,dx}\\ &= \frac{\int_0^1 y^{r_1-1/2}(1-y)^{r_2-1/2}\Gamma(\lambda_1/y; r, T)\,dy}{beta(r_1+1/2, r_2+1/2)},\end{aligned}$$

where $\Gamma(\lambda_1/y; r, T)$ is the cumulative distribution function of the gamma distribution with mean r/T evaluated at λ_1/y. Both the marginal posterior cumulative distribution of λ_2 are

(11)

$$\begin{aligned}&F_2(\lambda_2|\boldsymbol{t})\\ &= \frac{\int_0^{\lambda_2}\int_0^\infty x^{r_2-1/2}\lambda_1^{r_1-1/2}(x+\lambda_1)^{r_3-1}\exp\{-(x+\lambda_1)T\}d\lambda_1\,dx}{\int_0^\infty\int_0^\infty x^{r_2-1/2}\lambda_1^{r_1-1/2}(x+\lambda_1)^{r_3-1}\exp\{-(x+\lambda_1)T\}d\lambda_1\,dx}\\ &= \frac{\int_0^1 y^{r_2-1/2}(1-y)^{r_1-1/2}\Gamma(\lambda_2/y; r, T)\,dy}{beta(r_2+1/2, r_1+1/2)}.\end{aligned}$$

Proof. The second equality of (10) is due to the transformation $y = x/(x+\lambda_2)$, $z = x+\lambda_2$, and (11) is because of the transformation $y = x/(x+\lambda_1)$, $z = x+\lambda_1$. □

Sometimes, the hazard rate of the series system $\lambda_1+\lambda_2$ may be of the most interest. We reparametrize p, λ_1 and λ_2 as

$$p = p,\ \nu = \frac{\lambda_1}{\lambda_1+\lambda_2},\ \mu = \lambda_1+\lambda_2.$$

In this setting, ν is the probability that the failure of series system is due to the first component. Then the likelihood function becomes

$$L(p, \nu, \mu|\boldsymbol{t}) = p^{r_3}(1-p)^{r_1+r_2}\nu^{r_1}(1-\nu)^{r_2}\mu^r\exp\{-T\mu\},$$

and the Fisher information matrix of p, ν and μ is

$$I(p, \nu, \mu) = \begin{bmatrix}\dfrac{r}{p(1-p)} & 0 & 0\\ 0 & \dfrac{r(1-p)}{\nu(1-\nu)} & 0\\ 0 & 0 & \dfrac{r}{\mu^2}\end{bmatrix}.$$

Then the Jeffreys prior is $\pi_J(p, \nu, \mu) \propto p^{-1/2}\nu^{-1/2}(1-\nu)^{-1/2}\mu^{-1}$. The reference priors of the different orders are

$$\pi_R(p, \nu, \mu) \propto p^{-1/2}(1-p)^{-1/2}\nu^{-1/2}(1-\nu)^{-1/2}\mu^{-1}$$

or

$$\pi_J(p, \nu, \mu) \propto p^{-1/2}\nu^{-1/2}(1-\nu)^{-1/2}\mu^{-1}.$$

Theorem 3.4. *Under the priors $\pi_R(p, \nu, \mu)$ or $\pi_J(p, \nu, \mu)$,*

(a) Both the marginal posterior distributions of ν are $B(r_1+1/2, r_2+1/2)$.

(b) Both the marginal posterior distributions of μ are $G(r, T)$, where $G(r, T)$ denotes the gamma distribution with mean r/T and variance r/T^2.

(c) The marginal posterior distributions of p are $B(r_3+1/2, r_1+r_2+1/2)$ based on $\pi_R(p, \nu, \mu)$, and the marginal posterior distributions of p are $B(r_3+1/2, r_1+r_2+1)$ based on $\pi_J(p, \nu, \mu)$.

From Theorem 3.4, we can easily obtain the posterior means and interval estimates of p, ν and μ. The estimates of the original parameters λ_1 and λ_2 can also be obtained using the results of Theorem 3.4. The procedure is as follows:

1. Generate $\nu^{(i)}$ from $B(r_1+1/2, r_2+1/2)$, $\mu^{(i)}$ from $G(r, T)$, $i = 1, 2, \ldots, m$.
2. Let $\lambda_1^{(i)} = \nu^{(i)}\mu^{(i)}$ and $\lambda_2^{(i)} = (1-\nu^{(i)})\mu^{(i)}$. Then we can obtain that the posterior means of λ_1 and λ_2 are

$$\frac{1}{m}\sum_{i=1}^m\lambda_1^{(i)}\quad\text{and}\quad\frac{1}{m}\sum_{i=1}^m\lambda_2^{(i)},$$

respectively. The $100(1-\alpha)\%$ credible intervals of λ_1 and λ_2 are $[\lambda_1^{(\alpha/2)}, \lambda_1^{(1-\alpha/2)}]$ and $[\lambda_2^{(\alpha/2)}, \lambda_2^{(1-\alpha/2)}]$, where $\lambda_1^{(\alpha/2)}$ and $\lambda_2^{(\alpha/2)}$ are the $\alpha/2$-quantile of $\lambda_1^{(i)}$, $i = 1, 2, \ldots, m$ and $\lambda_2^{(i)}$, $i = 1, 2, \ldots, m$, respectively.

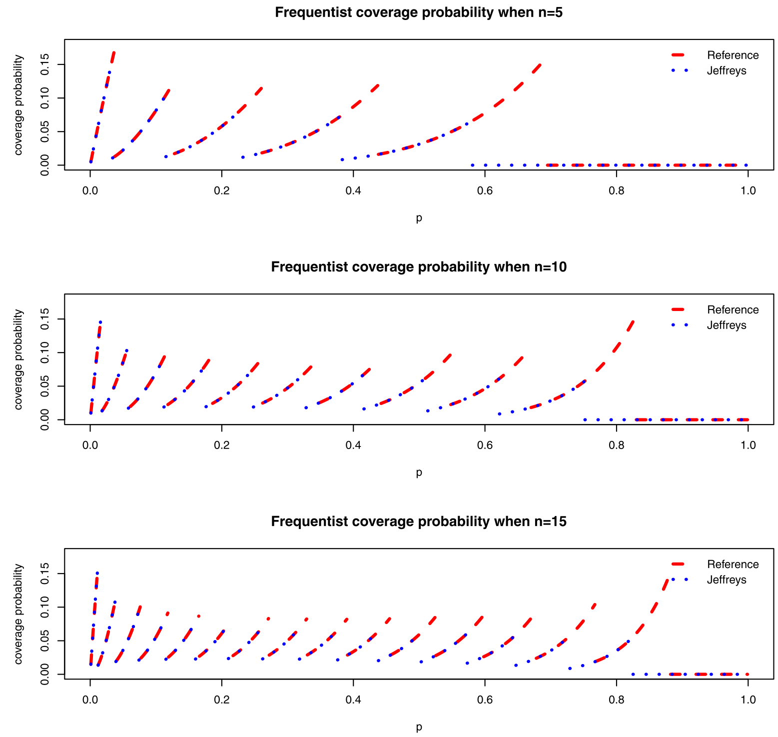

Figure 1. 5% frequentist coverage probability of $F_p^{-1}(\alpha)$ under different sample sizes.

Denote $qb(\alpha, i, j)$ as the α-quantile of $B(i, j)$. Let $F_\phi^{-1}(\alpha)$ be the α-quantile of the marginal posterior distribution of ϕ, $F_{Rp}^{-1}(\alpha)$ be the α-quantile of the marginal posterior distribution of p based on the reference prior, and $F_{Jp}^{-1}(\alpha)$ be the α-quantile of the marginal posterior distribution of p based on the Jeffreys prior.

Theorem 3.5. *(a) The frequentist coverage probability of $F_\mu^{-1}(\alpha)$ is α.*

(b) The frequentist coverage probability of $F_\nu^{-1}(\alpha)$ is $\sum_{k_{ij}} \frac{r!}{i!j!(r-i-j)!} p^{r-i-j}(1-p)^{i+j}\nu^i(1-\nu)^j$, where $S_{ij} = \{(i,j): qb(\alpha, i+1/2, j+1/2) \geq \nu, i+j \leq r\}$. If the set S_{ij} is empty, then the frequentist coverage probability is 0.

(c) The frequentist coverage probability of $F_{Rp}^{-1}(\alpha)$ is $\sum_{j=k_1}^{r} \binom{r}{j} p^j (1-p)^{r-j}$, where $S_1 = \min\{i: qb(\alpha, i+1/2, r-i+1/2) \geq p\}$. If S_1 does not exist, then the frequentist coverage probability is 0.

(d) The frequentist coverage probability of $F_{Jp}^{-1}(\alpha)$ is $\sum_{j=k_2}^{r} \binom{r}{j} p^j (1-p)^{r-j}$, where $S_2 = \min\{i: qb(\alpha, i+1/2, r-i+1) \geq p\}$. If S_2 does not exist, then the frequentist coverage probability is 0.

See the proof in the Appendix. From Theorem 3.5, we know that the frequentist coverage probabilities of the α-quantile of the marginal posterior distributions of p under both the Jeffreys prior and the reference prior are related to p. Under the reference prior, the frequentist coverage probability has a symmetric property in the sense of Corollary 1 below. However, it does not hold under the Jeffreys prior. Besides, given p, the frequentist coverage probabilities of the α-quantile of the marginal posterior distributions of ν also has the symmetric property.

Corollary 3.1. *(a) If $p \notin \{qb(\alpha, j+1/2, r-j+1/2), j = 0, \ldots, r\}$, then $P(F_{Rp}^{-1}(\alpha) \geq p) + P(F_{Rp}^{-1}(1-\alpha) \geq 1-p) = 1$.*

(b) If $p \notin \{qb(\alpha, i+1/2, j+1/2),\ i+j \leq r\}$, then given p, $P(F_\nu^{-1}(\alpha) \geq v) + P(F_\nu^{-1}(1-\alpha) \geq 1-v) = 1$.

See the proof in the Appendix.

4. SMALL SAMPLE COMPARISON

In this section simulation studies are performed to see the frequentist coverage of the α-quantiles of the marginal posterior distributions of p, ν, μ, λ_1 and λ_2. We take $\lambda_1 = 0.001$, $\lambda_2 = 0.002$, and $p = 0.1(0.2, ..., 0.9)$. The sample size $n = 5(10, 15)$ and $\alpha = 0.05(0.95)$. Since the coverage probability of the $F_\mu^{-1}(\alpha)$ is exact, we do not list here. The relationships between p and the frequentist coverages of $F_{Rp}^{-1}(\alpha)$ and $F_{Jp}^{-1}(\alpha)$ are drawn in Figures 1 and 2. We see that the reference prior performs much better than the Jeffreys prior, and the coverage probabilities based on the reference prior have symmetric property, just as Corollary 1 indicated. Besides, from Theorem 3.5, we know that the frequentist coverage probability of $F_\nu^{-1}(\alpha)$ is related to ν and p, and we do not list the result because the resolution of the figures are not high. The R codes are available from the authors, upon request.

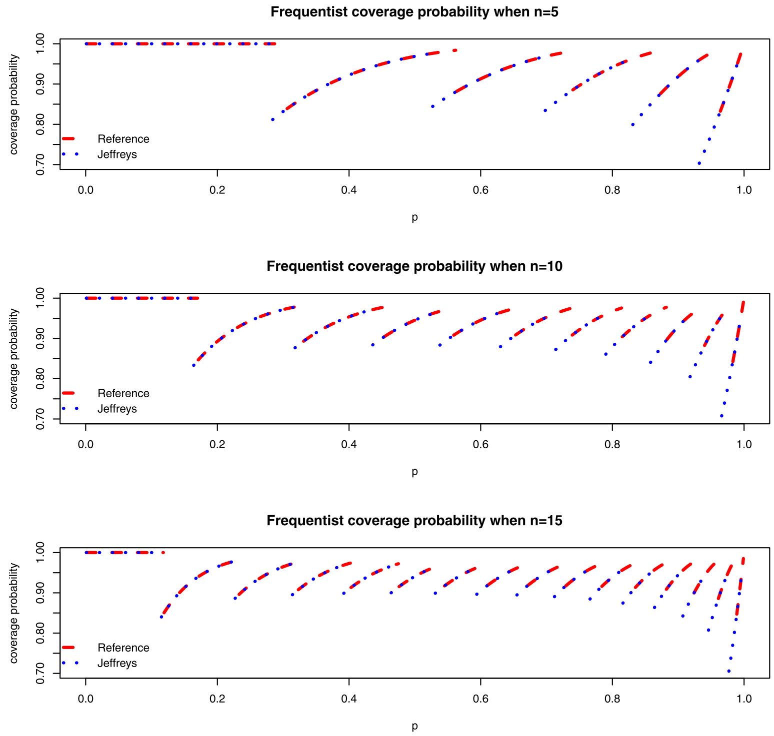

Figure 2. 95% frequentist coverage probability of $F_p^{-1}(\alpha)$ under different sample sizes.

Table 1. 5% frequentist coverage probability of $F_{\lambda_1}^{-1}(\alpha)$ and $F_{\lambda_2}^{-1}(\alpha)$

$n=5$	$p=.1$	$p=.2$	$p=.3$	$p=.4$	$p=.5$	$p=.6$	$p=.7$	$p=.8$	$p=.9$
λ_1	0.0370	0.0378	0.0359	0.0355	0.0395	0.0368	0.0351	0.0264	0.0154
λ_2	0.0339	0.0310	0.0312	0.0249	0.0214	0.0162	0.0126	0.0096	0.0038
$n=10$	$p=.1$	$p=.2$	$p=.3$	$p=.4$	$p=.5$	$p=.6$	$p=.7$	$p=.8$	$p=.9$
λ_1	0.0409	0.0424	0.0401	0.0405	0.0444	0.0461	0.0465	0.0418	0.0281
λ_2	0.0349	0.0339	0.0331	0.0325	0.0303	0.0284	0.0203	0.0139	0.0046
$n=15$	$p=.1$	$p=.2$	$p=.3$	$p=.4$	$p=.5$	$p=.6$	$p=.7$	$p=.8$	$p=.9$
λ_1	0.0463	0.0452	0.0443	0.0454	0.0440	0.0438	0.0496	0.0495	0.0394
λ_2	0.0392	0.0359	0.0351	0.0352	0.0326	0.0337	0.0277	0.0188	0.0072

Table 2. 95% frequentist coverage probability of $F_{\lambda_1}^{-1}(\alpha)$ and $F_{\lambda_2}^{-1}(\alpha)$

$n=5$	$p=.1$	$p=.2$	$p=.3$	$p=.4$	$p=.5$	$p=.6$	$p=.7$	$p=.8$	$p=.9$
λ_1	0.9344	0.9424	0.9483	0.9586	0.9665	0.9758	0.9855	0.9921	0.9964
λ_2	0.9239	0.9261	0.9226	0.9241	0.9226	0.9232	0.9280	0.9346	0.9538
$n=10$	$p=.1$	$p=.2$	$p=.3$	$p=.4$	$p=.5$	$p=.6$	$p=.7$	$p=.8$	$p=.9$
λ_1	0.9331	0.9304	0.9300	0.9326	0.9358	0.9479	0.9629	0.9807	0.9961
λ_2	0.9342	0.9318	0.9319	0.9276	0.9250	0.9218	0.9248	0.9281	0.9431
$n=15$	$p=.1$	$p=.2$	$p=.3$	$p=.4$	$p=.5$	$p=.6$	$p=.7$	$p=.8$	$p=.9$
λ_1	0.9392	0.9406	0.9392	0.9368	0.9367	0.9357	0.9470	0.9688	0.9913
λ_2	0.9381	0.9351	0.9363	0.9349	0.9313	0.9301	0.9306	0.9248	0.9364

Tables 1 and 2 give the numerical values of the frequentist coverage probabilities of $F_{\lambda_1}^{-1}(\alpha)$ and $F_{\lambda_2}^{-1}(\alpha)$. We can not obtain the explicit form of the coverage probabilities of $F_{\lambda_1}^{-1}(\alpha)$ and $F_{\lambda_2}^{-1}(\alpha)$, but they are related to p, λ_1 and λ_2, and have some jump points. That is why the coverage probabilities are close to α in some areas of p, while in other areas, the coverage probabilities are much different from α. To assess the effects of λ_1 and λ_2 on the coverage probabilities, we

Table 3. 5% frequentist coverage probability of $F^{-1}_{\lambda_1}(\alpha)$ and $F^{-1}_{\lambda_2}(\alpha)$ when $\lambda_1 = 0.0010,\ \lambda_2 = 0.0011$

$n=5$	$p=.1$	$p=.2$	$p=.3$	$p=.4$	$p=.5$	$p=.6$	$p=.7$	$p=.8$	$p=.9$
λ_1	0.0350	0.0340	0.0346	0.0318	0.0336	0.0277	0.0195	0.0185	0.0095
λ_2	0.0321	0.0340	0.0303	0.0290	0.0277	0.0237	0.0160	0.0111	0.0049
$n=10$	$p=.1$	$p=.2$	$p=.3$	$p=.4$	$p=.5$	$p=.6$	$p=.7$	$p=.8$	$p=.9$
λ_1	0.0397	0.0366	0.0391	0.0390	0.0394	0.0396	0.0354	0.0285	0.0146
λ_2	0.0387	0.0417	0.0408	0.0369	0.0354	0.0359	0.0317	0.0239	0.0094
$n=15$	$p=.1$	$p=.2$	$p=.3$	$p=.4$	$p=.5$	$p=.6$	$p=.7$	$p=.8$	$p=.9$
λ_1	0.0454	0.0450	0.0455	0.0450	0.0415	0.0443	0.0467	0.0462	0.0258
λ_2	0.0414	0.0426	0.0429	0.0416	0.0420	0.0410	0.0395	0.0338	0.0180

Table 4. 95% frequentist coverage probability of $F^{-1}_{\lambda_1}(\alpha)$ and $F^{-1}_{\lambda_2}(\alpha)$ when $\lambda_1 = 0.0010,\ \lambda_2 = 0.0011$

$n=5$	$p=.1$	$p=.2$	$p=.3$	$p=.4$	$p=.5$	$p=.6$	$p=.7$	$p=.8$	$p=.9$
λ_1	0.9228	0.9250	0.9259	0.9326	0.9381	0.9471	0.9627	0.9738	0.9867
λ_2	0.9284	0.9239	0.9258	0.9230	0.9285	0.9322	0.9427	0.9580	0.9737
$n=10$	$p=.1$	$p=.2$	$p=.3$	$p=.4$	$p=.5$	$p=.6$	$p=.7$	$p=.8$	$p=.9$
λ_1	0.9370	0.9361	0.9347	0.9317	0.9317	0.9321	0.9414	0.9605	0.9825
λ_2	0.9369	0.9358	0.9346	0.9311	0.9343	0.9325	0.9321	0.9427	0.9687
$n=15$	$p=.1$	$p=.2$	$p=.3$	$p=.4$	$p=.5$	$p=.6$	$p=.7$	$p=.8$	$p=.9$
λ_1	0.9326	0.9359	0.9352	0.9334	0.9304	0.9353	0.9337	0.9435	0.9722
λ_2	0.9433	0.9409	0.9400	0.9385	0.9341	0.9337	0.9351	0.9394	0.9590

Table 5. NRVD data for 58 female mice

Disease Status	M_i	Time to death
incidental	$\{1\}$	231,444,468,473,527,550,559,593,595,596,600,603,610,650,655,660,715,720,752,765,783,785,794,811,832, 838,856,859,870,883,891,896,897,904,931,952,975,978,991,998,1005,1023,1026,1053
fatal	$\{2\}$	500,591,713,751,778,784,786,796
unknown	$\{1,2\}$	593,735,816,848,850,1046

also compute these frequentist coverage probabilities when $\lambda_1 = 0.0010,\ \lambda_2 = 0.0011$. The results are listed in Tables 3 and 4. We find that the coverage probabilities are a little better than before, but not significantly.

5. REAL DATA STUDY

[7] gave the data that reported the death time and nonrenal vascular disease (NRVD) status at death for 58 female mice. The disease status is classified as "absent", "incidental", "unknown" or "fatal", according to whether the animal died without disease, with the disease present but not responsible for the death, with the disease present but its role in causing death unknown, or as a result of the disease. Therefore, the status "unknown" is referred to as masking of "incidental" and "fatal". To assess the role of the disease, we combine both the absent and the incidental status into one called the incidental status, which is also done by [12]. Let M_i indicate the disease status. We denote $M_i = \{1\}$ if the disease status is "incidental" and $M_i = \{2\}$ if the disease status is "fatal". Thus $M_i = \{1, 2\}$ if the disease status is "unknown". The data is listed in Table 5.

Based on $\pi_J(p, \lambda_1, \lambda_2)$ and $\pi_R(p, \lambda_1, \lambda_2)$, we compute the posterior means and 95% credible intervals (CIs) of $p, \lambda_1, \lambda_2, \nu$ and μ. The results are listed in Table 6. As a comparison, we also list the results of [12]. They utilized the uniform prior for p and $G(1, 0.000001)$ for λ_1 and λ_2. The result based on the three priors are very close to each other, because the sample size is large enough, so that the prior information can be ignored. From Table 6, we see that λ_1 is significantly greater than λ_2, because their 95% CIs are not overlapped. This means the probability of "incidental" status is greater than that of "fatal" status. Such a result can also be reflected by the 95% CI of ν.

6. DISCUSSION

In this paper, we indicate that the parameters may be nonidentifiable in the masked data model, and take the exponential distribution as an example to avoid the nonidentifiable problem by the symmetric assumption. Following the results of this paper, it can be shown that the unidentified problem also exists when the lifetimes of components are Weibull distribution with common shape parameter. Thus,

Table 6. Posterior means and 95% CI of the parameters

Parameters	Estimates	$\pi_J(p, \lambda_1, \lambda_2)$	$\pi_R(p, \lambda_1, \lambda_2)$	Kuo and Yang's
p	mean	0.1092	0.1102	0.1170
	95%CI	(0.0439, 0.1992)	(0.0443, 0.2008)	(0.0491, 0.208)
λ_1	mean	1.09×10^{-3}	1.09×10^{-3}	1.13×10^{-3}
	95%CI	$(8.09 \times 10^{-4}, 1.43 \times 10^{-3})$	$(8.09 \times 10^{-4}, 1.43 \times 10^{-3})$	$(8.24 \times 10^{-4}, 1.45 \times 10^{-3})$
λ_2	mean	2.10×10^{-4}	2.10×10^{-4}	2.25×10^{-4}
	95%CI	$(9.26 \times 10^{-5}, 3.74 \times 10^{-4})$	$(9.26 \times 10^{-5}, 3.74 \times 10^{-4})$	$(1.02 \times 10^{-4}, 3.85 \times 10^{-4})$
ν	mean	0.8395	0.8395	0.8328
	95%CI	(0.7288, 0.9247)	(0.7288, 0.9247)	(0.7251, 0.9203)
μ	mean	1.31×10^{-3}	1.31×10^{-3}	1.33×10^{-3}
	95%CI	$(9.95 \times 10^{-4}, 1.67 \times 10^{-3})$	$(9.95 \times 10^{-4}, 1.67 \times 10^{-3})$	$(1.01 \times 10^{-3}, 1.69 \times 10^{-3})$

our results can be extended to the Weibull case. Besides, for the case of the number of the components $J > 2$, the derivation of the noninformative priors will be more complicated. However, it can be easily proved that the result of item (a) of the theorem 7 still holds.

APPENDIX

Proof of Theorem 2.1. We make the following transformation

$$\begin{cases} u_1 = \lambda_1 - \lambda_1 p_1, \\ u_2 = \lambda_2 - \lambda_2 p_2, \\ u_3 = \lambda_1 p_1 + \lambda_2 p_2, \\ u_4 = g(p_1, p_2, \lambda_1, \lambda_2), \end{cases}$$

where $g()$ is an any function, such that the transformation from $(p_1, p_2, \lambda_1, \lambda_2)$ to $(u_1, \ldots, u_4)$ is one-to-one. For example, $g(p_1, p_2, \lambda_1, \lambda_2) = \lambda_1$. Thus, the result is equivalent to show that $(u_1, \ldots, u_4)$ is nonidentifiable. Based on the above transformation, the likelihood function (6) becomes

$$L = u_1^{r_1} u_2^{r_2} u_3^{r_{12}} \exp\left[-(u_1 + u_2 + u_3)T_r\right].$$

For any two different points (u_1, u_2, u_3, u_4') and (u_1, u_2, u_3, u_4'') in the support of the parameters, where $u_4' \neq u_4''$, the values of L are the same. Thus, the result holds. □

Proof of Proposition 2.2. From the first result of Proposition 2.1, we have

$$\mathrm{E}(\hat{p}) = \mathrm{E}\left(\frac{r_3}{r}\right) = \frac{rp}{r} = p, \ \mathrm{Var}(\hat{p}) = \frac{p(1-p)}{r}.$$

Thus, the first result holds. Furthermore,

$$\begin{aligned} \mathrm{E}(\hat{\lambda}_1) &= \mathrm{E}\left(\frac{1}{T}\right) \cdot \mathrm{E}\left(\frac{rr_1}{r_1 + r_2}\right) \\ &= \frac{(\lambda_1 + \lambda_2)}{r-1} \cdot \frac{r\lambda_1}{\lambda_1 + \lambda_2} = \frac{r}{r-1}\lambda_1, \\ \mathrm{Var}(\hat{\lambda}_1) &= \mathrm{E}\left[\left(\frac{1}{T}\right)^2 \cdot \left(\frac{rr_1}{r_1 + r_2}\right)^2\right] \\ &\quad - \left[\mathrm{E}\left(\frac{r}{T}\right) \cdot \mathrm{E}\left(\frac{r_1}{r_1 + r_2}\right)\right]^2 \\ &= \mathrm{Var}\left(\frac{1}{T}\right) \cdot \mathrm{Var}\left(\frac{rr_1}{r_1 + r_2}\right) + \mathrm{Var}\left(\frac{1}{T}\right) \\ &\quad \times \left(\mathrm{E}\left(\frac{rr_1}{r_1 + r_2}\right)\right)^2 \\ &\quad + \mathrm{Var}\left(\frac{rr_1}{r_1 + r_2}\right) \cdot \left(\mathrm{E}\left(\frac{1}{T}\right)\right)^2 \\ &= \frac{r\lambda_1[(r-1)\lambda_2 + r\lambda_1]}{(r-1)^2(r-2)}. \end{aligned}$$

Similarly, $\mathrm{E}(\hat{\lambda}_2) = \frac{r}{r-1}\lambda_2$ and $\mathrm{Var}(\hat{\lambda}_2) = \frac{r\lambda_2[(r-1)\lambda_1 + r\lambda_2]}{(r-1)^2(r-2)}$. Thus, the second result holds. □

Proof of Theorem 2.2. Let $\boldsymbol{t} = (T, r_1, r_2, r_3)$. Then the posterior density function of $(p, \lambda_1, \lambda_2)$ is

(12)
$$\begin{aligned} \pi(p, \lambda_1, \lambda_2 | \boldsymbol{t}) &= \frac{L_1/(\lambda_1\lambda_2)}{m(\boldsymbol{t})} \\ &= p^{r_3}(1-p)^{r_1+r_2}\lambda_1^{r_1-1}\lambda_2^{r_2-1} \\ &\quad \times (\lambda_1 + \lambda_2)^{r_3} \exp\{-(\lambda_1 + \lambda_2)T\}/m(\boldsymbol{t}), \end{aligned}$$

where

$$\begin{aligned} m(\boldsymbol{t}) &= \int\int\int L_1/(\lambda_1\lambda_2) \mathrm{d}p\mathrm{d}\lambda_1\mathrm{d}\lambda_2 \\ &= \frac{beta(r_1 + r_2 + 1, r_3 + 1)}{T^r} \\ &\quad \times \sum_{i=0}^{r_3} \binom{r_3}{i} \Gamma(r_1 + i)\Gamma(r_2 + r_3 - i), \end{aligned}$$

where $\Gamma(\cdot)$ denotes the gamma function, and $beta(\cdot, \cdot)$ is the beta function. The third equality holds only if $r_1 > 0$ and $r_2 > 0$; otherwise, the integration would be infinity for the cases $i = 0$ and $i = r_3$. Thus the posterior distribution of $(p, \ \lambda_1, \ \lambda_2)$ is proper only if $r_1 > 0$ and $r_2 > 0$. Thus, the first result holds. The second result can be easily obtained from (12).

To obtain a more simpler expression of $m(\boldsymbol{t})$, we make the following transformation

$$p = p,\ \nu = \frac{\lambda_1}{\lambda_1+\lambda_2},\ \mu = \lambda_1+\lambda_2. \tag{13}$$

Then

$$\begin{aligned} m(\boldsymbol{t}) &= \int\int\int p^{r_3}(1-p)^{r_1+r_2}\lambda_1^{r_1-1}\lambda_2^{r_2-1}(\lambda_1+\lambda_2)^{r_3} \\ &\quad \times \exp\{-(\lambda_1+\lambda_2)T\}\mathrm{d}p\mathrm{d}\lambda_1\mathrm{d}\lambda_2 \\ &= \int\int\int p^{r_3}(1-p)^{r_1+r_2}\nu^{r_1-1}(1-\nu)^{r_2-1}\mu^{r-1} \\ &\quad \times \exp\{-\mu T\}\mathrm{d}p\mathrm{d}\nu\mathrm{d}\mu \\ &= \frac{beta(r_1+r_2+1, r_3+1)beta(r_1,r_2)\Gamma(r)}{T^r} \\ &= \frac{r_3!(r_1-1)!(r_2-1)!(r_1+r_2)}{r(r+1)T^r}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} m_1(\boldsymbol{t}) &= \int\int\int L_1/\lambda_2\mathrm{d}p\mathrm{d}\lambda_1\mathrm{d}\lambda_2 = \frac{r_3!r_1!(r_2-1)!}{(r+1)T^{r+1}}, \\ m_2(\boldsymbol{t}) &= \int\int\int L_1\lambda_1/\lambda_2\mathrm{d}p\mathrm{d}\lambda_1\mathrm{d}\lambda_2 = \frac{r_3!r_1!(r_2-1)!}{(r_1+r_2+1)T^{r+2}}. \end{aligned}$$

Thus, the posterior mean and variance of λ_1 are

$$\mathrm{E}(\lambda_1|\boldsymbol{t}) = \frac{m_1(\boldsymbol{t})}{m(\boldsymbol{t})} = \frac{r}{T}\frac{r_1}{r_1+r_2}$$

and

$$\begin{aligned} \mathrm{Var}(\lambda_1|\boldsymbol{t}) &= \frac{m_2(\boldsymbol{t})}{m(\boldsymbol{t})} - [\mathrm{E}(\lambda_1|\boldsymbol{t})]^2 \\ &= \frac{r}{T^2}\frac{r_1(1+r_1)(r_1+r_2)+r_1r_2r}{(r_1+r_2)^2(r_1+r_2+1)}, \end{aligned}$$

respectively. Similarly, the posterior mean and variance of λ_2 are

$$\mathrm{E}(\lambda_2|\boldsymbol{t}) = \frac{m_1(\boldsymbol{t})}{m(\boldsymbol{t})} = \frac{r}{T}\frac{r_2}{r_1+r_2}$$

and

$$\mathrm{Var}(\lambda_2|\boldsymbol{t}) = \frac{r}{T^2}\frac{r_2(1+r_2)(r_1+r_2)+r_1r_2r}{(r_1+r_2)^2(r_1+r_2+1)}.$$

Thus the results hold. □

Proof of Theorem 3.1. (i) According to [5], we need to find h_1 and h_2. From (9), we have the fisher information matrix for (λ_1,λ_2,p) has the following form

$$I((\lambda_1,\lambda_2),p) = \begin{pmatrix} I_{11} & 0 \\ 0 & I_{22} \end{pmatrix},$$

where

$$I_{11} = \begin{pmatrix} \frac{r(\lambda_1+\lambda_2)-rp\lambda_2}{\lambda_1(\lambda_1+\lambda_2)^2} & \frac{rp}{(\lambda_1+\lambda_2)^2} \\ \frac{rp}{(\lambda_1+\lambda_2)^2} & \frac{r(\lambda_1+\lambda_2)-rp\lambda_1}{\lambda_2(\lambda_1+\lambda_2)^2} \end{pmatrix},\ I_{22} = \frac{r}{p(1-p)}.$$

It is not difficult to calculate that

$$\begin{aligned} h_1 &= |I_{11}| = \frac{r^2(1-p)}{\lambda_1\lambda_2(\lambda_1+\lambda_2)^2}, \\ h_2 &= \frac{|I((\lambda_1,\lambda_2),p)|}{|I_{11}|} = \frac{r}{p(1-p)}. \end{aligned}$$

Choose $\Omega_k = \Omega_{12k}\times\Omega_{3k} = \{(\lambda_1,\lambda_2)|a_{1k}<\lambda_1<b_{1k}, a_{2k}<\lambda_2<b_{2k}\}\times\{p|a_{3k}<p<b_{13k}\}$, such that a_{1k}, a_{2k}, $a_{3k}\to 0$, b_{1k}, $b_{2k}\to\infty$ and $b_{3k}\to 1$. Then the conditional prior of p given (λ_1,λ_2) is

$$\xi_2^k(p|\lambda_1,\lambda_2) = \frac{\sqrt{h_2}A_{\Omega_{3k}}(p)}{\int_{\Omega_{3k}}\sqrt{h_2}dp},$$

where $A_{\Omega_{3k}}(p)$ is the indicator function of p. The marginal prior of (λ_1,λ_2) is

$$\begin{aligned} \xi_1^k(\lambda_1,\lambda_2) &\propto \exp\left\{\frac{1}{2}\int_{\Omega_{3k}}\xi_2^k(p|\lambda_1,\lambda_2)\log(h_1)dp\right\} \\ &\quad\times A_{\Omega_{12k}}(\lambda_1,\lambda_2) \\ &\propto \frac{1}{\sqrt{\lambda_1\lambda_2(\lambda_1+\lambda_2)^2}}A_{\Omega_{12k}}. \end{aligned}$$

Taking any point in the support of (λ_1,λ_2,p), say (0.3,0.3,0.2), then the reference prior of $\{(\lambda_1,\lambda_2),p\}$ is

$$\begin{aligned} \pi_R(p,\lambda_1,\lambda_2) &= \lim_{k\to\infty}\frac{\xi_1^k(\lambda_1,\lambda_2)\xi_2^k(p|\lambda_1,\lambda_2)}{\xi_1^k(0.3,0.3)\xi_2^k(0.2|0.3,0.3)} \\ &= p^{-1/2}(1-p)^{-1/2}\lambda_1^{-1/2}\lambda_2^{-1/2}(\lambda_1+\lambda_2)^{-1}. \end{aligned}$$

(ii) The proof is similar, and is omitted. □

Proof of Theorem 3.2. We just prove the result for $\pi_R(p,\lambda_1,\lambda_2)$, since the result for $\pi_J(p,\lambda_1,\lambda_2)$ is similar. It suffices to show that the marginal distribution of $\boldsymbol{t}$ is

$$m_R(\boldsymbol{t}) = \int L_1\pi_R(p,\lambda_1,\lambda_2)\mathrm{d}p\mathrm{d}\lambda_1\mathrm{d}\lambda_2 < \infty.$$

First, we have

$$\begin{aligned} m_R(\boldsymbol{t}) &= \int(1-p)^{r_1+r_2-1/2}p^{r_3-1/2}\lambda_1^{r_1-1/2}\lambda_2^{r_2-1/2} \\ &\quad\times(\lambda_1+\lambda_2)^{r_3-1}\exp\{-(\lambda_1+\lambda_2)T\}\mathrm{d}p\mathrm{d}\lambda_1\mathrm{d}\lambda_2. \end{aligned}$$

Using the transformation (13), we have

$$\begin{aligned} m_R(\boldsymbol{t}) &= \int(1-p)^{r_1+r_2-1/2}p^{r_3-1/2}\nu^{r_1-1/2}(1-\nu)^{r_2-1/2} \\ &\quad\times\mu^{r-1}\exp\{-\mu T\}\mathrm{d}p\mathrm{d}\nu\mathrm{d}\mu \end{aligned}$$

$$= beta(r_1 + r_2 + 1/2, r_3 + 1/2) \times beta(r_1 + 1/2, r_2 + 1/2)\Gamma(r)/T^r < \infty. \quad \square$$

Proof of Theorem 3.5. (a) From Theorem 4, we know that $\mu|\boldsymbol{t} \sim G(r, T)$. Then

$$\begin{aligned} P\left(\mu \le F_\mu^{-1}(\alpha)\right) &= \mathrm{P}(\Gamma(\mu, r, T) \le \alpha) \\ &= P\left(\int_0^\mu \frac{T^r}{\Gamma(r)} x^{r-1} \exp\{-Tx\}\mathrm{d}x \le \alpha\right) \\ &= P\left(\int_0^{T\mu} \frac{1}{\Gamma(r)} y^{r-1} \exp\{-y\}\mathrm{d}y \le \alpha\right) \\ &= P\left(\int_0^{T} \frac{\mu^r}{\Gamma(r)} z^{r-1} \exp\{-z\mu\}\mathrm{d}z \le \alpha\right) \\ &= P(\Gamma(T; r, \mu) \le \alpha). \end{aligned}$$

The third equality and the fourth equality hold due to transformations $y = Tx$ and $z = y/\mu$, respectively. Notice that the statistics T follows $G(r, \mu)$, hence $\Gamma(T; r, \mu)$ follows $U(0, 1)$. Thus $P(\Gamma(T; r, \mu) \le \alpha) = \alpha$.

(b) Notice that

$$(r_1, r_2, r_3) \sim Multinomial\left(r, \frac{(1-p)\lambda_1}{\lambda_1 + \lambda_2}, \frac{(1-p)\lambda_2}{\lambda_1 + \lambda_2}, p\right).$$

Then $F_\nu^{-1}(\alpha)$ follows a discrete distribution, taking values in the set $\{qb(\alpha, i + 1/2, j + 1/2), i + j \le r\}$, and the corresponding probabilities are $\frac{r!}{i!j!(r-i-j)!} p^{r-i-j}(1-p)^{i+j}\nu^i(1-\nu)^j$, $i + j \le r$. If all the values of $F_\nu^{-1}(\alpha)$ are less than ν, the frequentist coverage probability is 0. Otherwise,

$$\begin{aligned} &P(F_\nu^{-1}(\alpha) \ge \nu) \\ = &\sum_{k_{ij}} \frac{r!}{i!j!(r-i-j)!} p^{r-i-j}(1-p)^{i+j} v^i (1-v)^j, \end{aligned}$$

where $S_{ij} = \{(i, j) : \ qb(\alpha, i + 1/2, j + 1/2) \ge \nu, i + j \le r\}$.

(c) Notice that $r_3 \sim Binomial(r, p)$. Then $F_{Rp}^{-1}(\alpha)$ follows a discrete distribution, taking values in the set $\{qb(\alpha, j + 1/2, r - j + 1/2), \ j = 0, \dots, r\}$, and the corresponding probabilities are $\binom{r}{j} p^j (1-p)^{r-j}, j = 0, \dots, r$. If all the values of $qb(\alpha, r_3 + 1/2, r_1 + r_2 + 1/2)$ are less than p, $P(F_{Rp}^{-1}(\alpha) \ge p) = 0$. Otherwise,

$$\begin{aligned} P(F_{Rp}^{-1}(\alpha) \ge p) &= P(qb(\alpha, r_3 + 1/2, r_1 + r_2 + 1/2) \le p) \\ &= \sum_{j=k_1}^{r} \binom{r}{j} p^j (1-p)^{r-j}, \end{aligned}$$

where $S_1 = \min\{i : \ qb(\alpha, i + 1/2, r - i + 1/2) \ge p\}$.

(d) The proof is similar to (c). $\square$

Proof of Corollary 3.1. (a) From Theorem 5,we have

$$P(F_{Rp}^{-1}(\alpha) \ge p) = \sum_{j=k_1}^{r} \binom{r}{j} p^j (1-p)^{r-j},$$

$$P(F_{Rp}^{-1}(1-\alpha) \ge 1-p) = \sum_{j=k_0}^{r} \binom{r}{j} (1-p)^j p^{r-j},$$

where k_1 is the same as in Theorem 5, $k_0 = \min\{i : \ qb(1 - \alpha, i + 1/2, r - i + 1/2) \ge 1 - p\}$.

Notice that $qb(\alpha, a, b) + qb(1 - \alpha, b, a) = 1$, for any $a, \ b > 0$, thus we know that the set $\{1 - qb(1 - \alpha, i + 1/2, r - i + 1/2), i = r, r - 1, \dots, 0\}$ has the same elements as that in the set $\{qb(\alpha, j + 1/2, r - j + 1/2), j = 0, \dots, r\}$. Hence, from the definition of k_0 and k_1, and $p \notin \{qb(\alpha, j + 1/2, r - j + 1/2), j = 0, \dots, r\}$, we have $k_0 = r + 1 - k_1$. Thus we obtain

$$\begin{aligned} \sum_{j=k_0}^{r} \binom{r}{j} (1-p)^j p^{r-j} &= \sum_{j=r-k_1+1}^{r} \binom{r}{j} p^{r-j} (1-p)^j \\ &= \sum_{j=0}^{k_1 - 1} \binom{r}{j} p^j (1-p)^{r-j}. \end{aligned}$$

Then the result follows immediately.

(b) The proof is similar to (a). $\square$

ACKNOWLEDGMENT

The research is supported by the Natural Science Foundation of China (11201345, 11271136), Program of Shanghai Subject Chief Scientist (14XD1401600) and the 111 Project (B14019). The authors thank the two referees for their constructive suggestions that greatly improved the article.

Received 19 October 2013

REFERENCES

[1] Basu, S., Basu, A., and Mukhopadhyay, C. (1999). Bayesian analysis for masked system failure data using nonidentical weibull models. *J. Statist. Plann. Inference*, 78:255–275. MR1705552

[2] Basu, S., Sen, A., and Banerjee, M. (2003). Bayesian analysis of competing risks with partially masked cause of failure. *Appl. Statist.*, 52:77–93. MR1959084

[3] Berger, J. and Sun, D. (1993). Bayesian analysis for the poly-weibull distribution. *J. Amer. Statist. Assoc.*, 88:1412–1418. MR1245378

[4] Berger, J. O. (2006). The case for objective bayesian analysis (with discussion). *Bayesian Analysis*, 1:386–402. MR2221271

[5] Berger, J. O. and Bernardo, J. M. (1992). Ordered group reference priors with application to the multinomial problem. *Biometrika*, 79:25–37. MR1158515

[6] Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference (C/R p128-147). *Journal of the Royal Statistical Society, Series B, Methodological*, 41:113–128. MR0547240

[7] Dinse, G. (1986). Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause-of-death data. *J. Amer. Statist. Assoc.*, 81:328–336.

[8] Flehinger, B. J., Reiser, B., and Yashchin, E. (2002). Parametric modeling for survival with competing risks and masked failure causes. *Lifetime Data Analysis*, 8:177–203. MR1906342

[9] Guan, Q., Tang, Y., and Xu, A. (2013). Objective bayesian analysis for bivariate marshall–olkin exponential distribution. *Computational Statistics and Data Analysis*, 64:299–313. MR3061905

[10] Guess, F., Usher, J., and Hodgson, T. (1991). Estimating system and component reliabilities under partial information on cause of failure. *J. Statist. Plann. Inference*, 29:75–85. MR1133692

[11] Jeffreys, H. (1961). *Theory of probability*. 3rd ed. Oxford: Oxford University Press. MR0187257

[12] Kuo, L. and Yang, T. (2000). Bayesian reliability modeling for masked system lifetime data. *Statist. Probab. Lett.*, 47:229–241. MR1747483

[13] Lin, D., Usher, J., and Guess, F. (1993). Exact maximum likelihood estimation using masked system data. *IEEE Trans. Reliab.*, 42:631–635.

[14] Lin, D., Usher, J., and Guess, F. (1996). Bayes estimation of component-reliability from masked system-life data. *IEEE Trans. Reliab.*, 45:233–237.

[15] Miyakawa, M. (1984). Analysis of incomplete data in a competing risks model. *IEEE Trans. Reliab.*, 33:293–296.

[16] Mukhopadhyay, C. (2006). Maximum likelihood analysis of masked series system lifetime data. *J. Statist. Plann. Inference*, 136:803–838. MR2181978

[17] Mukhopadhyay, C. and Basu, A. (1997). Bayesian analysis of incomplete time and cause of failure data. *J. Statist. Plann. Inference*, 59:79–100. MR1450778

[18] Reiser, B., Guttman, I., Lin, D., Usher, J., and Guess, F. (1995). Bayesian inference for masked system lifetime data. *Appl. Statist.*, 44:79–90. MR1323353

[19] Sen, A., Banerjee, M., Li, Y., and Noone, A. (2010). A bayesian approach to competing risks analysis with masked cause of death. *Statist. Med.*, 29:1681–1695. MR2752176

[20] Usher, J. and Hodgson, T. (1988). Maximum likelihood analysis of component reliability using masked system life data. *IEEE Trans. Reliab.*, 37:550–555.

[21] Xu, A., Basu, S., and Tang, Y. (2014a). A full bayesian approach for masked data in step-stress accelerated life testing. *IEEE Trans. Reliab.*, 63:798–806.

[22] Xu, A. and Tang, Y. (2009). Bayesian analysis of pareto reliability with dependent masked data. *IEEE Trans. Reliab.*, 58:583–588.

[23] Xu, A. and Tang, Y. (2011). Nonparametric bayesian analysis of competing risks problem with masked data. *Communications in Statistics-Theory and Methods*, 40:2326–2336. MR2863207

[24] Xu, A., Tang, Y., and Guan, Q. (2014b). Bayesian analysis of masked data in step-stress accelerated life testing. *Communications in Statistics-Simulation and Computation*, 43:2016–2030.

Ancha Xu
College of Mathematics and Information Science
Wenzhou University
Zhejiang 325035
School of Natural and Applied Sciences
Northwestern Polytechnical University
Xi'an 710072
China
E-mail address: xuancha@wzu.edu.cn

Yincai Tang
School of Finance and Statistics
East China Normal University
Shanghai 200241
China
E-mail address: yctang@stat.ecnu.edu.cn

Dongchu Sun
Department of Statistics
University of Missouri
Columbia, MO 65211
USA
E-mail address: sund@missouri.edu