Inference for the first order statistic distribution based on masked Exponential distributions for two independent competing risks

Miguel Henry-Osorio[a1] and Francis Pascual[b]

[a] School of Economic Sciences, Washington State University, Pullman, WA, USA; [b] Department of Statistics, Washington State University, Pullman, WA, USA

This paper considers the scenario of independent risks which have times to occurrence described by Exponential distributions. The problem involves masking, that is, the time of occurrence of the earlier risk (first order statistic) is observed but not the specific risk. Nonidentifiability is a consequence of this limitation, and a Bayesian procedure is implemented to address this issue. More specifically, independent Gamma priors are placed on the Exponential parameters which allow identification of the Exponential parameters. Confidence intervals (CIs) for the distribution function $F_{(1)}(x)$ of the time to first occurrence are studied. Rigorous computer simulations are carried out to assess the relative performance of CIs based on Maximum Likelihood methods, Bayesian, and distribution-free asymptotics in terms of empirical coverage probability and average interval length. All the approaches presented in this study are for two independent Exponential risks, which can be generalized to three or more risks.

Keywords: Bayesian method; confidence intervals; credible intervals.

Subject Classification: 62N05, 62F15, 62G30

1. Introduction

This paper considers the study of CI procedures for the cumulative distribution function (CDF) of the first order statistic failure-time of an item, where it is assumed that the failure times are observed (i.e. uncensored), and the competing risks (i.e. the *failure modes*) are both masked and statistically independent exponentials.

"Competing risks" are an extension of classical survival analysis and refer to a situation in which a unit is at risk of experiencing more than one distinct cause or type of failure during its lifetime. For instance, a government may collapse due to dissolution (e.g. takeover) or replacement (new elections). The competing risks literature defines latent random failure/survival times X_1, X_2,…,X_k, where X_k denotes a non-negative duration time in which a unit would fail due to a specific k^{th} cause and the X_k's can be independent or dependent among them. Since we observe only the earliest of these failure modes (David and Moeschberger [1]), we are considering only time-to-first-event data.

One of the important problems in competing risks analysis is to uniquely identify the parameters associated with the parametric distribution that the failure time of a unit follows. When the cause of failure is masked, that is, the exact cause of the failure is unknown, then there may be an identifiability problem such that we are unable to uniquely estimate the model parameters. Modeling failure times of latent independent competing risks has been already discussed in the literature as well as alternative methods to the conventional parametric procedures (e.g. see Miyakawa [2]; Usher and Hodgson [3]; Bergera and Sun [4]; Reiser *et al.*

[1] Corresponding author: mhenryo@wsu.edu

[5]; Flehinger, Reiser and Yashchin [6]; Basu, Basu and Mukhopadhyay [7]; Jiang and Murthy [8]). Basu, Basu and Mukhopadhyay [7] point out that the only separation between parameters of the component failure time distributions is "the one inherent in the prior" so that they argue that identifiability "is formally not a problem for Bayesian analyses", unless the effect of the priors do not remove a non-identifiable scenario, even in cases of large samples.

This study aims at illustrating the performance of CIs for the CDF of the first order statistic based on different estimation procedures. Comparison of their relative performance is undertaken in terms of average length and empirical coverage probability. In this study, the term empirical coverage is defined as the proportion or percentage of times that the estimated simulated results for which the $100(1-\alpha)\%$ CIs for $F_{(1)}(x)$ cover the true value of interest. The first procedure is based on a parametric approach using the exact Exponential distribution when competing risks are assumed to be independent Exponentials. The second and third approaches belong to Bayesian analysis and to the distribution-free class of estimators (e.g. the Kaplan-Meier estimator), respectively. In the Bayesian context, credible intervals for the parameters where constructed using independent Gamma priors. A comparison of the three methods is undertaken through a simulation study. To our knowledge the only study that systematically compares different estimators to assess their performance is by Kundu, Kannan and Balakrishnan [9]. However, they address it using progressive Type-II censoring data.

Section 2 gives a brief characterization of the methods utilized as it relates to competing risks. Section 3 illustrates the performance of the three CI methods using a simulation study. Finally, Section 4 shows the empirical results achieved when applying the three procedures to a real dataset, followed by concluding remarks in Section 5.

2. Statistical Model Assumptions and Notation

Let X_1, $X_2,\ldots,X_k$ denote the times of occurrences of k competing risks, respectively. As stated earlier, many times we are only able to observe a very first of multiple causes of failure, known in the literature as *minimum random subset*. If $X_{(1)}$ denotes the time to first occurrence, then $X_{(1)} = \min\{X_1, X_2,\ldots,X_k\}$. To simplify the exposition in this paper, we consider k=2 causes of failure and assume that they are statistically independent Exponentials. Because we cannot identify the specific risk associated with the first occurrence (masking), then we cannot uniquely estimate the individual Exponential parameters. Here, we adopt the convention that capital letters imply random variables and lower case letters denote actual values or data. All the statistical models presented in this study and their results can be generalized to the case of more than two competing risks.

2.1 The exact parametric estimator

For unit j, let X_{ij} be the time to failure due to risk *i* for i=1,2 and j=1,…,n. Assume that X_{ij} has a one-parameter Exponential distribution with mean η_i (or hazard rate $\lambda_i=1/\eta_i$), and write $X_{ij}\sim\text{Exp}(\eta_i)$. Also assume that all the X_{ij}s are independent and X_{ij} has a probability density function (PDF) and CDF given by $f_i(x)=\frac{1}{\eta_i}e^{(-x/\eta_i)}=\lambda_i e^{(-\lambda_i x)}$ and $F_i(x)=1-e^{(-x/\eta_i)}=1-e^{-\lambda_i x}$, respectively, for $x \ge 0$.

Let $X_{(1)j} = \min\{X_{1j}, X_{2j}\}$ be the time to failure of unit j. Note that $X_{(1)1}, X_{(1)2}, \ldots, X_{(1)n}$ are independent and identically distributed (iid) with common CDF $F_{(1)}$. We have $F_{(1)}(x) = 1 - P[X_{(1)j} > x] = 1 - \prod_{i=1}^{2} P[X_{ij} > x] = 1 - \prod_{i=1}^{2}[1 - F_i(x)]$. Thus, $F_{(1)}(x) = 1 - \prod_{i=1}^{2} e^{-\frac{x}{\eta_i}} = 1 - e^{-\left(\frac{1}{\eta_1}+\frac{1}{\eta_2}\right)x} = 1 - e^{-(\lambda_1+\lambda_2)x}$ and $X_{(1)j} \sim Exp\left(\eta = \left(\frac{1}{\eta_1}+\frac{1}{\eta_2}\right)^{-1} = \frac{1}{\lambda_1+\lambda_2}\right)$ for j=1,…,n.

Since we assume that competing risks are completely masked η_1, η_2 (or, equivalently, λ_1, λ_2), they are not individually identifiable (i.e., uniquely estimable). However, we can estimate $\eta = \left(\frac{1}{\eta_1}+\frac{1}{\eta_2}\right)^{-1} = \frac{1}{\lambda_1+\lambda_2}$. For example, the Maximum Likelihood estimator (MLE) of η is given by $\bar{X}_{(1)} = \sum_{j=1}^{n} X_{(1)j} / n$. Note that $\sum_{j=1}^{n} X_{(1)j}$ has a Gamma distribution with scale parameter η and shape parameter *n* and, hence, $\bar{X}_{(1)}$ has a Gamma distribution with scale η / n and shape *n*. Furthermore, $\bar{X} / \eta$ is Gamma with scale *1/n* and scale *n*. Let y_γ be the γ quantile of this distribution. If $0 < \alpha < 1$, then,

$$P\left(y_{\alpha/2} < \frac{\bar{X}_{(1)}}{\eta} < y_{1-\alpha/2}\right) = 1 - \alpha$$

and

$$P\left(\frac{\bar{X}_{(1)}}{y_{\alpha/2}} > \eta > \frac{\bar{X}_{(1)}}{y_{1-\alpha/2}}\right) = 1 - \alpha \tag{1}$$

This suggests that a $100(1-\alpha)\%$ CI for η is

$$\left(\frac{\bar{X}_{(1)}}{y_{\alpha/2}}, \frac{\bar{X}_{(1)}}{y_{1-\alpha/2}}\right)$$

and a $100(1-\alpha)\%$ CI for $F_{(1)}(x)$ is given by

$$\left(1 - exp\left(\frac{-x}{\left(\bar{X}_{(1)} / y_{\frac{\alpha}{2}}\right)}\right), 1 - exp\left(\frac{-x}{\left(\bar{X}_{(1)} / y_{1-\frac{\alpha}{2}}\right)}\right)\right) \tag{2}$$

2.2 The Bayesian parametric estimator

Our data consist of realizations of $X_{(1)1},...,X_{(1)n}$, which are iid $Exp\left(\eta=\left(\frac{1}{\eta_1}+\frac{1}{\eta_2}\right)^{-1}=\frac{1}{\lambda_1+\lambda_2}\right)$. Because the individual scale parameters λ_i are non-identifiable under the scenario of masking (Basu, Basu and Mukhopadhyay [7], we placed Gamma priors on λ_1 and λ_2 to address the non-identifiability problem and find Bayesian estimates. According to Kundu, Kannan and Balakrishnan [9], *failure rate* parameters λ_i are reasonably modeled using gamma priors in the context of exponential lifetimes. The Gamma prior distribution are commonly chosen due to their mathematical tractability and are employed when "prior information about the model is quite limited" (Robert and Casella [10]; section 1.6.1). Some guidelines about how to choose the model parameters for the priors are specified at the end of this section.

Assume that λ_1 and λ_2 are independent with λ_i following a $\text{Gamma}\left(\alpha_i,\beta_i\right)$ prior distribution with PDF

$$p_i\left(\lambda_i\right)=\frac{1}{\beta_i^{-\alpha_i}\Gamma\left(\alpha_i\right)}\lambda_i^{\alpha_i-1}\exp\left(-\beta_i\lambda_i\right),\ \lambda_i\geq 0$$

for i=1,2 where $\Gamma(\cdot)$ is the gamma function, $\alpha_i>0$ is the shape parameter and $\beta_i>0$ is the scale parameter. We have $E\left[\lambda_i\right]=\frac{\alpha_i}{\beta_i}$ and $V\left[\lambda_i\right]=\frac{\alpha_i}{\beta_i^2}$. When α_i is close to 0, the prior on λ_i becomes non-informative (Kundu, Kannan and Balakrishnan [9]). It is important to note that under this approach, λ_i are treated as if they are random variables; however, as Mittelhammer*,* Judge and Miller [11] (section 22.2) point out "this does not imply that a parameter is actually a random variable. Rather, this is a conceptual device that allows the analyst to identify the possible values considered as candidates for the generally fixed unknown parameter value as well as the corresponding probabilities, or degrees of belief, relating to the validity of these values."

The (conditional) joint PDF of $\underline{X}_{(1)}=\left[X_{(1)1},...,X_{(1)n}\right]^T$ given $\underline{\lambda}=\left[\lambda_1,\lambda_2\right]^T$ is:

$$f\left(\underline{x},\underline{\lambda}\right)=\left(\lambda_1+\lambda_2\right)^n\exp\left(-\left(\lambda_1+\lambda_2\right)x_{.}\right) \tag{3}$$

where $x_{.}=\sum_{j=1}^{n}x_{(1)j}$ and $\underline{x}=\left[x_{(1)1},\ \ldots,\ x_{(1)n}\right]^T$. In general, Bayes' theorem for densities follows:

$$f\left(\underline{\lambda}\mid\underline{x}\right)=\frac{p\left(\underline{\lambda}\right)f\left(\underline{x}\mid\underline{\lambda}\right)}{f_{\underline{X}}\left(\underline{x}\right)} \tag{4}$$

where $f\left(\underline{\lambda}\mid\underline{x}\right)$ is referred to as a posterior density, $p\left(\underline{\lambda}\right)$ is the prior joint density of $\underline{\lambda}$, and $f_{\underline{X}}\left(\underline{x}\right)$ is the marginal pdf of $\underline{X}_{(1)}$. By performing algebraic manipulation, the posterior distribution of $\underline{\lambda}$ given $\underline{x}$ is given by:

$$f\left(\underline{\lambda}\mid\underline{x}\right)=\frac{\sum_{i=0}^{n}\binom{n}{i}\lambda_1^{n+\alpha_1-i-1}\lambda_2^{i+\alpha_2-1}e^{-\lambda_1\left(\beta_1+x_.\right)-\lambda_2\left(\beta_2+x_.\right)}}{\sum_{i=0}^{n}\binom{n}{i}\left(\beta_1+x_.\right)^{-(n+\alpha_1-i)}\Gamma\left(n+\alpha_1-i\right)*\left(\beta_2+x_.\right)^{-(i+\alpha_2)}\Gamma\left(i+\alpha_2\right)} \quad (5)$$

See appendix A for a complete derivation of (5). Posterior means of λ_1 and λ_2 given $\underline{x}$ are provided in appendix B.

Letting $S=(\lambda_1+\lambda_2)\mid\underline{x}$, the posterior CDF of the convolution S is:

$$\begin{aligned}F_S\left(s\right)&=P\left(\lambda_1+\lambda_2\le s\mid\underline{x}\right)\\&=\int_0^\infty\int_0^{s-\lambda_1}f\left(\underline{\lambda}\mid\underline{x}\right)d\lambda_2 d\lambda_1\\&=\frac{\sum_{i=0}^{n}\binom{n}{i}}{f_{\underline{x}}\left(\underline{x}\right)}\int_0^\infty\int_0^{s-\lambda_1}\frac{1}{\beta_1^{-\alpha_1}\Gamma\left(\alpha_1\right)}\lambda_1^{n+\alpha_1-i-1}e^{-\lambda_1\left(\beta_1+x_.\right)}*\frac{1}{\beta_2^{-\alpha_2}\Gamma\left(\alpha_2\right)}\lambda_2^{i+\alpha_2-1}e^{-\lambda_2\left(\beta_2+x_.\right)}d\lambda_2 d\lambda_1\end{aligned}$$

where

$$D=f_{\underline{X}}\left(\underline{x}\right)*\beta_1^{-\alpha_1}\beta_2^{-\alpha_2}\Gamma\left(\alpha_1\right)\Gamma\left(\alpha_2\right)=\sum_{i=0}^{n}\binom{n}{i}\left(\beta_1+x_.\right)^{-(n+\alpha_1-i)}\Gamma\left(n+\alpha_1-i\right)*\left(\beta_2+x_.\right)^{-(i+\alpha_2)}\Gamma\left(i+\alpha_2\right)$$

Then,

$$\begin{aligned}&=\frac{\sum_{i=0}^{n}\binom{n}{i}}{D}\int_0^\infty\lambda_1^{n+\alpha_1-i-1}e^{-\lambda_1\left(\beta_1+x_.\right)}\left[\int_0^{s-\lambda_1}\lambda_2^{i+\alpha_2-1}e^{-\lambda_2\left(\beta_2+x_.\right)}d\lambda_2\right]d\lambda_1\\&=\frac{1}{D}\left\{\sum_{i=0}^{n}\binom{n}{i}\varnothing\left(n+\alpha_1-i,\beta_1+x_.,i+\alpha_2,\beta_2+x_.S\right)\right\}\end{aligned}$$

where:

$\varnothing\left(n+\alpha_1-i,\beta_1+x_.,i+\alpha_2,\beta_2+x_.S\right)=\beta_2^{-\alpha_2}\Gamma\left(\alpha_2\right)\int_0^\infty\lambda_1^{n+\alpha_1-i-1}e^{-\lambda_1\left(\beta_1+x_.\right)}G_2\left(s-\lambda_1\right)d\lambda_1$ and

$G_2\left(s-\lambda_1\right)=\int_0^{s-\lambda_1}\frac{1}{\beta_2^{-\alpha_2}\Gamma\left(\alpha_2\right)}\lambda_2^{i+\alpha_2-1}e^{-\lambda_2\left(\beta_2+x_.\right)}d\lambda_2$ (i.e. G_2 is the CDF of Gamma($i+\alpha_2,\beta_2+x_.$)).

If s_γ is the γ quantile of $S = (\lambda_1 + \lambda_2) \mid \underline{x}$ (i.e., $\gamma = F_S(s_\gamma)$), then a $100(1-\alpha)\%$ "*credible*" interval for $(\lambda_1 + \lambda_2) = 1/\eta$ is $\left(s_{\alpha/2}, s_{1-\alpha/2}\right)$. Thus, a $100(1-\alpha)\%$ "*credible*" interval for $F_{(1)}(x)$ is given by:

$$\left(1-e^{-x^* s_{1-\alpha/2}}, 1-e^{-x^* s_{\alpha/2}}\right) \qquad (6)$$

Posterior quantiles of S are numerically obtained using the *uniroot* function implemented in the programming language *R* [12] interfaced with a *FORTRAN* quadrature subroutine and a dynamically loadable object for computing the function $\emptyset(\cdot)$.

Guidelines for choosing model parameters for the priors

As noted previously, it was assumed that λ_i follows a Gamma(α_i, β_i) prior distribution despite some formal rules (e.g. Jeffreys's method) illustrated by Kass and Wasserman [13] and some methods provided by Berger [14; Section 3.2] could be used for selecting priors. However, many of these rules are subject to criticism (see Koop *et al.* [15]; Section 8).

An optional choice is to get the practitioner expertise about a rough guess of what λ_i are. Nevertheless, this information may not be available. Koop *et al.* [15] point out that the analyst should "*let the data speak for themselves*". In this sense, some researchers tend to choose *diffuse priors* with large dispersion when there is no prior information. Agresti and Min [16] state that "intervals based on diffuse priors performed considerably better than more informative priors." However, Guikema [17] argues that "the amount of information assumed in the prior make a critical difference in the accuracy of the posterior inferences." Thus, using more informative priors can lead to more accurate posterior inferences. A standard measure of dispersion or diffuseness of a probability distribution is the well known coefficient of variation (CV), which is defined as the ratio of the standard deviation over the mean. For the Gamma distributions considered above, we have $E[\lambda_i] = \frac{\alpha_i}{\beta_i}$, $V[\lambda_i] = \frac{\alpha_i}{\beta_i^2}$, and, hence, $CV = \frac{1}{\sqrt{\alpha_i}}$. Thus, if the prior means are relatively similar, then the variance is a measure of diffuseness. Otherwise, we may use *CV* in which case a more diffuse prior distribution is associated with a smaller α_i.

We consider the following guidelines in choosing a Gamma prior on λ_i:

- If there is a rough guess for $E(\lambda_i)$, then α_i and β_i are chosen so that $E[\lambda_i] = \alpha_i / \beta_i$ where larger α_i values correspond to stronger faith in the prior.
- Otherwise, a diffuse prior is placed on λ_i.

2.3 The Kaplan-Meier estimator

Suppose that we have no information regarding the distribution of time to failure X. Let $x_1, \ldots, x_m \ (m \le n)$ denote the observed unique failure times in increasing order of the n units. Also, let $d_1, \ldots, d_j$ and $n_1, \ldots n_j$ denote the number of failures and number alive, respectively, immediately before t_i. Then, for $x \in [x_i, x_{i+1})$, an estimate of the survival function at x is:

$$\hat{S}(x)=\prod_{j=1}^{i}\left(1-\frac{d_j}{n_j}\right)=\prod_{j=1}^{i}\left(1-\hat{p}_j\right).$$

The estimate $\hat{S}(x)$ is known as the Kaplan-Meier (K-M) or product-limit estimator of the survival function at time *x*. The function $\hat{S}(x)$ is a right-continuous step function with hazard components $\hat{p}_1,\ldots,\hat{p}_j$ at $x_1,\ldots,x_i$, respectively. For an extended overview of this estimator see Kaplan and Meier [13].

To compute CIs, Greenwood's formula (Kalbfleisch and Prentice [14]) offers an asymptotic consistent estimate of the standard error (se) of $\hat{S}(x)$ as follows:

$$se=\hat{S}(x)*\sqrt{\sum_{j=1}^{i}\frac{d_j}{n_j\left(n_j-d_j\right)}}.$$

Then, a pointwise normal-approximation $100(1-\alpha)\%$ CI for the CDF $F(x)$ of X may be constructed by assuming that $\hat{F}(x)$ is approximately normal with the above standard error (See Greenwood [15]). However, CIs based on the logit transformation of $F(x)$ results in coverage probabilities closer to nominal than the above CI procedure in both small and large-sample situations (see Weston and Meeker [16]). Thus, we construct an approximate CI for F(x) by assuming $logit(\hat{F})=logit(1-\hat{S})$ is approximately normal where $logit(p)=log\left[\frac{p}{1-p}\right]$ (see Meeker and Escobar [17], chapter 3) and applying the error-propagation (delta) method with Greenwood's standard error. Thus, an asymptotic two-sided nonparametric $100(1-\alpha)\%$ CI for $F(x)$ is given by:

$$\left(\underline{F},\overline{F}\right)=\left(\frac{\hat{F}}{\left(\hat{F}+\left(\left(1-\hat{F}\right)w_i\right)\right)},\frac{\hat{F}}{\left(\hat{F}+\left(\left(1-\hat{F}\right)\big/w_i\right)\right)}\right) \tag{7}$$

where $w_i=\exp\left(z_{\left(1-\alpha/2\right)}*\left(se\big/\left(\hat{F}\left(1-\hat{F}\right)\right)\right)\right)$ and $z_{\left(1-\alpha/2\right)}$ is the $\left(1-\alpha/2\right)$ quantile of the standard normal distribution. Note that we dropped *x* in $\hat{F}(x)$ to simplify the expressions.

3. Simulation Results

As mentioned in section 2.2, we choose the Gamma prior on λ_i so that $E(\lambda_i)=\alpha_i/\beta_i$ or alternatively $\alpha_i=E(\lambda_i)\beta_i$. Thus, if $\{E(\lambda_1),E(\lambda_2)\}=\{1,1/3\}$ then $\beta_1=\alpha_1$ and $\beta_2=3\alpha_2$. In this section, we study the effects of incorrectly specifying the values of α_i and β_i. That is to say, $\alpha_i = E(\lambda_i)\beta_i$ is not satisfied for at least one of i=1,2. In pursuit of this task, we considered 95% credible intervals for $F_{(1)}(x)$ when $\{E(\lambda_1), E(\lambda_2)\} = \{1, 1/3\}$ for the sets of priors $\text{Gamma}(\alpha_1 = 1.25, \beta_1 = 2.5)$, $\text{Gamma}(\alpha_1 = 5, \beta_1 = 4)$, $\text{Gamma}(\alpha_2 = 1, \beta_2 = 5)$, and $\text{Gamma}(\alpha_2 = 10, \beta_2 = 10)$. Here, we compare to 95% the empirical coverage probabilities (ECPs) of the 95% credible intervals under these priors. We also report results for a candidate of correct priors $(\text{Gamma}(\alpha_1 = 2, \beta_1 = 2)$ and $\text{Gamma}(\alpha_2 = 1, \beta_2 = 3))$ as a reference set. The best priors should yield ECPs close to the nominal ones of 95% with narrow credible intervals.

In order to estimate the ECPs of the credible intervals within a 1% margin of error, uncensored Exponential random samples of size n = 20, 40, 60, 80 and 100 were generated 10,000 times. Column 1 in Table 1 summarizes ECPs under the aforementioned sets of priors. Columns 2, 3, 4 and 5 from Table 1 report estimates of interval lengths and average bounds.

Table 1. Simulated coverage probabilities of 95% credible intervals for $F_{(1)}(x)$ when p = 0.5

Gamma ($\alpha_1 = 1.25$, $\beta_1 = 2.5$) and Gamma ($\alpha_2 = 1$, $\beta_2 = 5$)

N	Empirical Coverage Probability (%)	Average Interval Length	Variance Interval Length	Average Lower Limit	Average Upper Limit
20	96.09	0.283	0.00041	0.573	0.856
40	95.29	0.208	0.00013	0.621	0.829
60	95.50	0.172	0.00006	0.642	0.815
80	95.42	0.150	0.00003	0.655	0.806
100	95.38	0.135	0.00002	0.665	0.799

Gamma ($\alpha_1 = 5$, $\beta_1 = 4$) and Gamma ($\alpha_2 = 10$, $\beta_2 = 10$)

N	Empirical Coverage Probability	Average Interval Length	Variance Interval Length	Average Lower Limit	Average Upper Limit
20	88.13	0.215	0.00033	0.669	0.885
40	90.25	0.177	0.00012	0.674	0.851
60	91.90	0.154	0.00006	0.678	0.832
80	92.40	0.138	0.00003	0.683	0.821
100	93.00	0.126	0.00002	0.686	0.812

Gamma ($\alpha_1 = 2$, $\beta_1 = 2$) and Gamma ($\alpha_2 = 1$, $\beta_2 = 3$)

N	Empirical Coverage Probability	Average Interval Length	Variance Interval Length	Average Lower Limit	Average Upper Limit
20	96.65	0.278	0.00038	0.577	0.855
40	95.60	0.206	0.00012	0.624	0.829
60	95.19	0.171	0.00005	0.643	0.814
80	95.42	0.149	0.00003	0.656	0.806
100	95.20	0.134	0.00002	0.664	0.799

From Table 1, a number of findings can be derived. First, it is clear that the ECPs depend upon the sample size regardless of the priors we finally choose. This is expected because inference based on posterior distributions will rely less on the assumed priors with more observations. Second, it is seen that the precision, in terms of interval lengths, of the credible intervals tends to improve as the sample size increases. Third, the higher the confidence level is, the wider the average interval lengths are. Moreover, it is also quite clear that as the sample size gets large enough, the effect on the credible intervals and coverage probabilities become negligible. On the other hand, Table 1 shows that under the Gamma($\alpha_1 = 5,\ \beta_1 = 4$) and Gamma($\alpha_2 = 10,\ \beta_2 = 10$) priors, the ECPs are evidently below the 95% nominal coverage. Also, the average credible interval lengths for $F_{(1)}(x)$, which have been constructed based on (6) and a type I error equal to 0.05, get shorter compared to those when using more informative ($\text{Gamma}(\alpha_1 = 1.25, \beta_1 = 2.5)$, $\text{Gamma}(\alpha_1 = 5, \beta_1 = 4)$) and correct priors. As expected, correct prior specification provides a balance between efficiency and nominal coverage probabilities.

Lastly, for illustrative purposes, we studied the effect of diffuseness on the interval lengths and ECPs under the assumption of $\alpha_1 = \alpha_2$, $\eta_1 = 1$ and $\eta_2 = 3$. As shown in Figure 1, the (average) credible interval lengths for $F_{(1)}(x)$ become narrower as α_i gets larger. This figure, which is based on a sample size equal to 100 and 1,000 repetitions, suggests that the less informative the α prior is, the shorter the 95% empirical Bayes confidence interval is. On the other hand, the more diffuse the α prior is, the more inaccurate the ECP is. It should be mentioned that when the alpha hyperparameters (prior parameters) go beyond 50, the numerical computation of the ECPs breaks down.

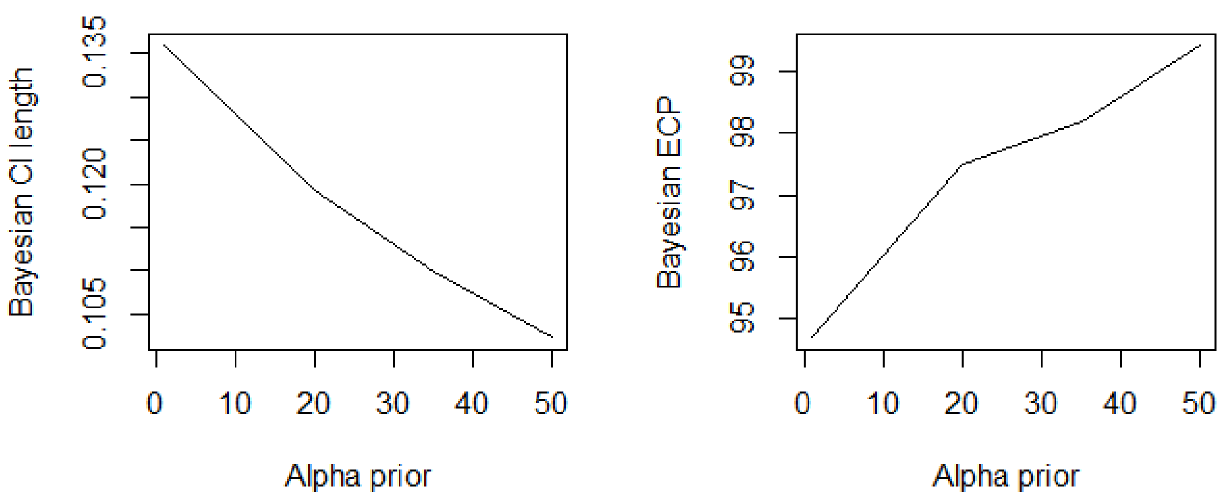

Figure 1. Effect of diffuseness on 95% credible intervals for $F_{(1)}(x)$ and ECP

4. An Illustrative Example

In this section, the three CI procedures discussed earlier are applied to a failure-time dataset with two masked competing risks. The values of the first failures from the dataset from Jiang and Murthy (2003) are given below:

$\underline{X}_{(1)}$=[68.58,63.10,61.54,52.18,50.25,48.47,47.65,43.71,41.21,33.08,31.10,28.91,28.47,27.20,26.58,26.23,25.72,22.92,19.82,18.72,17.29,16.93,16.76,15.91,15.38,14.68,14.36,11.80,10.84,10.30,7.71,4.80,4.78,4.46,4.44,4.18,3.82,2.78,2.65,2.54,1.83,1.81,1.8,1.52,1.18,1.02,0.81,0.66,0.5,0.46,0.39,0.22,0.17,0.165,0.155,0.14,0.09,0.04,0.0014,0.001]

Table 2 presents the 99%, 95% and 90% CIs for $F_{(1)}(x)$ using the exact parametric and Bayesian approaches based on (2) and (6), respectively (see also appendix C), whereas Figure 2 shows the K-M approach with CIs for the CDF around each of the $\hat{F}(t_i)$ estimates based on (7).

To fit the Bayesian credible intervals, the data were assumed to be realizations of the minimum occurrence times of two independent risks described by Exponential distributions. Based on practitioners' inputs, $E[X_1] \doteq 40$ and $E[X_2] \doteq 45.1373$. Thus, for the Gamma priors, we have $\frac{\alpha_1}{\beta_1} = \frac{1}{40}$ or $\frac{\alpha_1}{0.025} = \beta_1$ and $\frac{\alpha_2}{\beta_2} = \frac{1}{45.1373}$ or $\frac{\alpha_2}{0.0222} = \beta_2$. Finally, diffuse priors with large dispersion (i.e. relatively small α_i) as well as informative priors were chosen to obtain credible intervals for $F_{(1)}(x)$. For this application, we choose the following informative and diffuse priors $\text{Gamma}(\alpha_1 = 1, \beta_1 = 40)$, $\text{Gamma}(\alpha_2 = 10, \beta_2 = 450.450)$ and $\text{Gamma}(\alpha_1 = .1, \beta_1 = 4)$, $\text{Gamma}(\alpha_2 = 1, \beta_2 = 45.045)$, respectively.

Table 3. $100(1-\alpha)\%$ upper (UCL) and lower (LCL) confidence limits for $F_{(1)}(x)$ using the exact parametric and Bayesian methods

	Confidence Interval Method								
	Exact Parametric			**Bayesian (informative priors)**			**Bayesian (diffuse priors)**		
1 - cl	LCL	ULC	Length	LCL	ULC	Length	LCL	ULC	Length
99%	.0425	.0813	.0388	.0417	.0793	.0376	.0419	.0800	.0381
95%	.0463	.0759	.0296	.0454	.0740	.0286	.0457	.0747	.0290
90%	.0484	.0731	.0247	.0474	.0714	.0240	.0477	.0720	.0243

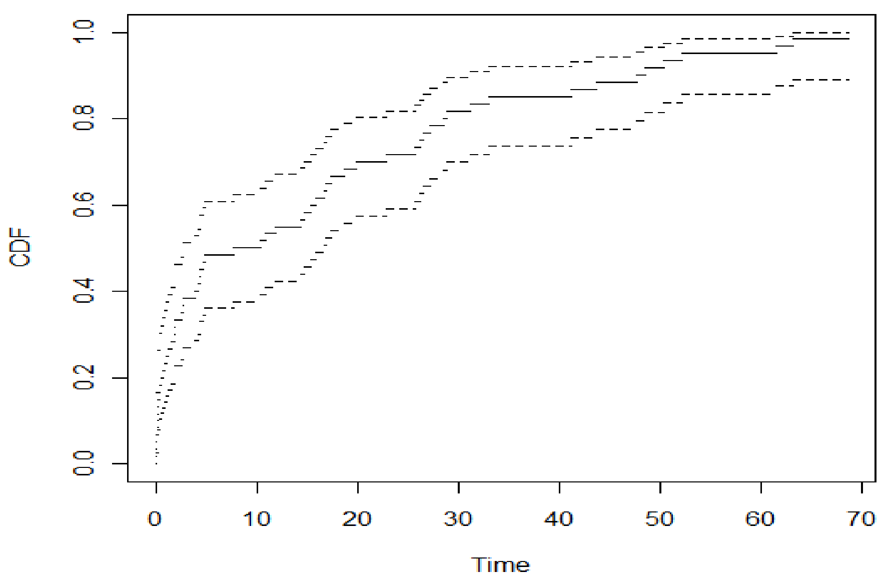

Figure 3. Plot of the non-parametric estimate with pointwise logit-transformation normal-approximation 95% confidence intervals for $F_{(1)}(x)$.

Confidence limits using the exact parametric and Bayesian procedures are quite similar. However, the Bayesian technique reports more precise estimates, especially when using diffuse priors with large dispersion. Comparing the K-M CIs obtained in Figure 3 to those 95% CIs reported in Table 3, we can see that the K-M CIs are wider (see also Figure 2). This is an expected result since even though the K-M estimation method incorporates more flexibility than the other two approaches it relies on weaker assumptions.

5. Concluding remarks

This paper examined three different methods for computing CIs for $F_{(1)}(x)$ in terms of empirical coverage probabilities and average interval lengths. Even though non-parametric estimation techniques are more robust to wrong functional forms of the underlying lifetime sampling distribution and it can be implemented more easily than other techniques, the exact fully parametric approach performed better than the non-parametric asymptotic approach. Regarding the Bayesian method, credible intervals based on non-informative priors get very close to the exact approach and both tend to converge as the sample size increases. Hence, diffuse priors with large dispersion rather than more informative ones should be considered when choosing priors, unless the information that the expert provides is reliable. Then, more informative priors can be used to yield shorter CIs. Rather, if there is some doubt on the quality information, more diffuse priors should be chosen. That is, the belief we have in the expert information should be reflected in the priors we finally choose. The conclusions drawn from the simulation study are supported by the findings obtained in the empirical application. Regarding the solution time needed to acquire the empirical coverage probabilities and other estimate results, the Bayesian approach, as compared to the fully parametric approach and non-parametric technique, involves computational challenges and it takes considerably more computational time especially when conducting simulations. CI lengths for the unit's first-failure-time distribution using random $\lambda_i's$ could be computed and compared to the current results, however, more work is needed.

References

Agresti, A. and Y. Min. 2005. Frequentist Performance of Bayesian Confidence Intervals for Comparing Proportions in 2x2 Contingency Tables. *Biometrics*. 61: 515 – 523.

Basu, S., Basu, A.P. and C.Mukhopadhyay. 1999. Bayesian analysis for masked system failure data using non-identical Weibull models. *Journal of Statistical Planning and Inference*. 78: 255 – 275.

Berger, J.O. and D. Sun. 1993. Bayesian analysis for the Poly-Weibull distribution. *Journal of the American Statistical Association*. 88(424): 1412 – 1418.

Cox, D.R. 1959. The Analysis of Exponentially Distributed Life-Times with Two Types of Failure. *Journal of the Royal Statistical Society, Series B*. 21(2): 411 – 421.

David, H.A. and M.L. Moeschberger. 1978. The Theory of Competing Risks. Griffin's Statistical Monographs & Courses No. 39. Series Editor: Professor Alan Stuart.

Flehinger, B. J., Reiser, B. and , E. Yashchin. 1996. Inference about defects in the presence of masking. *Technometrics*. 38: 247 - 255.

Guikema, S.D. 2007. Formulating informative, data-based priors for failure probability estimation in reliability analysis. *Reliability Engineering & System Safety*. 92: 490 – 502.

Jiang, R. and D.N.P. Murthy. 2003. Study of *n*-Fold Weibull Competing Risk Model. *Mathematical and Computer Modelling*. 38: 1259 – 1273.

Kalbfleisch, J.D. and R.L. Prentice. 1980. The Statistical Analysis of Failure Time Data.

Kaplan, E.L. and P. Meier. 1958. Nonparametric Estimation for Incomplete Observations. *Journal of the American Statistical Association*. 53(282): 457 – 481.

Kass, R.E. and L.Wasserman. 1996. The Selection of Prior Distributions by Formal Rules. *Journal of the American Statistical Association*. 91(435): 1343 – 1370.

Koop, G., Poirier, D.J. and J.L. Tobias. 2007. Bayesian Econometric Methods. Econometric Exercises 7. Cambridge University Press.

Kundu, D., Kannan, N. and N. Balakrishnan. 2004. Analysis of Progressively Censored Competing Risks Data. In: Handbook of Statistics 23, Advances in Survival Analysis. Edited by N. Balakrishnan and C.R. Rao. Part VI. Competing Risks and Analysis.

Meeker, W.Q. and L.A. Escobar. 1998. Nonparametric Estimation. *In*: Statistical Methods for Reliability Data.

Mittelhammer, R., Judge, G. and D. Miller. 2000. Econometric Foundation, New York: Cambridge University Press.

Miyakawa, M. 1984. Analysis of Incomplete Data in Competing Risks Model. *IEEE Transactions Reliability*. R-33(4): 293 – 296.

Reiser, B., Guttman, I., Lin, D. K. J., Usher, J. S. and , F. M. Guess. 1995. Bayesian inference for masked system life time data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. 44(1): 79 - 90.

Robert, C.P. and G. Casella. 2004. Monte Carlo Statistical Methods. Second Edition. Springer.

Sarhan, A.M., Hamilton, D.C. and B. Smith. 2010. Statistical analysis of competing risks models. *Reliability Engineering & System Safety*. 95(9): 953 – 962.

Usher, J.S. and T.J. Hodgson. 1988. Maximum likelihood analysis of component reliability using masked system life-test data. *IEEE Transactions Reliability*. R-37: 550 – 555.

Weston, S.A. and W.Q. Meeker. 1990. Coverage Probability of Nonparametric Simultaneous Confidence Bands for a Survival Function. *Journal of Statistical Computation and Simulation*. 32: 83-97.

Appendix A. Derivation of the posterior distribution of $\underline{\lambda}$ given $\underline{x}$

Let's first derive the numerator of Eq. (2); that is, the joint pdf of $\left(\underline{X}_{(1)}, \underline{\lambda}\right)$. Based on (1) and the prior density results:

$$f\left(\underline{X}_{(1)}, \underline{\lambda}\right) = f\left(\underline{x}|\underline{\lambda}\right) * \prod_{i=1}^{2} g_i(\lambda_i)$$

$$= (\lambda_1 + \lambda_2)^n e^{-(\lambda_1+\lambda_2)x_.} * \frac{1}{\beta_1^{-\alpha_1}\Gamma(\alpha_1)} \lambda_1^{\alpha_1-1} e^{-\beta_1\lambda_1} * \frac{1}{\beta_2^{-\alpha_2}\Gamma(\alpha_2)} \lambda_2^{\alpha_2-1} e^{-\beta_2\lambda_2}$$

$$= (\lambda_1 + \lambda_2)^n * \frac{1}{\beta_1^{-\alpha_1}\beta_2^{-\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} * e^{-\beta_1\lambda_1+-\beta_2\lambda_2+-(\lambda_1+\lambda_2)x_.} * \lambda_1^{\alpha_1-1} * \lambda_2^{\alpha_2-1}$$

$$= (\lambda_1 + \lambda_2)^n * \frac{1}{\beta_1^{-\alpha_1}\beta_2^{-\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} * e^{-\lambda_1(\beta_1+x_.)-\lambda_2(\beta_2+x_.)} * \lambda_1^{\alpha_1-1} * \lambda_2^{\alpha_2-1}$$

(Hint: $(a + b)^n = \sum_{i=0}^{n}\binom{n}{i} a^{n-i} b^i$)

$$= \sum_{i=0}^{n} \binom{n}{i} \lambda_1^{n-i} \lambda_2^{i} \frac{\lambda_1^{\alpha_1-1} * \lambda_2^{\alpha_2-1}}{\beta_1^{-\alpha_1}\beta_2^{-\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda_1(\beta_1+x_.)-\lambda_2(\beta_2+x_.)}$$

$$= \frac{\sum_{i=0}^{n}\binom{n}{i}\lambda_1^{n+\alpha_1-i-1}\lambda_2^{i+\alpha_2-1} e^{-\lambda_1(\beta_1+x_.)-\lambda_2(\beta_2+x_.)}}{\beta_1^{-\alpha_1}\beta_2^{-\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)}$$

The marginal (unconditional) pdf of $\underline{X}_{(1)}$ is denoted by:

$$f_{\underline{x}}\left(\underline{x}\right) = \iint_0^{\infty} \frac{\sum_{i=0}^{n}\binom{n}{i}\lambda_1^{n+\alpha_1-i-1}\lambda_2^{i+\alpha_2-1} e^{-\lambda_1(\beta_1+x_.)-\lambda_2(\beta_2+x_.)}}{\beta_1^{-\alpha_1}\beta_2^{-\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} d\lambda_1 d\,\lambda_2$$

$$= \frac{\sum_{i=0}^{n}\binom{n}{i}}{\beta_1^{-\alpha_1}\beta_2^{-\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} \iint_0^{\infty} \lambda_1^{n+\alpha_1-i-1}\lambda_2^{i+\alpha_2-1} e^{-\lambda_1(\beta_1+x_.)-\lambda_2(\beta_2+x_.)} d\lambda_1 d\,\lambda_2$$

(Hint: $\int_0^{\infty} x^{\alpha-1}\, e^{-\beta x} dx = \beta^{-\alpha}\Gamma(\alpha)$)

$$= \frac{\sum_{i=0}^{n}\binom{n}{i}\,(\beta_1 + x_.)^{-(n+\alpha_1-i)}\Gamma(n + \alpha_1 - i) * (\beta_2 + x_.)^{-(i+\alpha_2)}\Gamma(i + \alpha_2)}{\beta_1^{-\alpha_1}\beta_2^{-\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)}$$

Finally, the posterior distribution of $\underline{\lambda}$ given $\underline{x}$ is given by:

$$f\left(\underline{\lambda}|\underline{x}\right) = \frac{\dfrac{\sum_{i=0}^{n}\binom{n}{i}\lambda_1^{n+\alpha_1-i-1}\lambda_2^{i+\alpha_2-1}e^{-\lambda_1(\beta_1+x_.)-\lambda_2(\beta_2+x_.)}}{\beta_1^{-\alpha_1}\beta_2^{-\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)}}{\dfrac{\sum_{i=0}^{n}\binom{n}{i}\,(\beta_1+x_.)^{-(n+\alpha_1-i)}\Gamma(n+\alpha_1-i)*(\beta_2+x_.)^{-(i+\alpha_2)}\Gamma(i+\alpha_2)}{\beta_1^{-\alpha_1}\beta_2^{-\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)}}$$

$$= \frac{\sum_{i=0}^{n}\binom{n}{i}\lambda_1^{n+\alpha_1-i-1}\lambda_2^{i+\alpha_2-1}e^{-\lambda_1(\beta_1+x_.)-\lambda_2(\beta_2+x_.)}}{\sum_{i=0}^{n}\binom{n}{i}\,(\beta_1+x_.)^{-(n+\alpha_1-i)}\Gamma(n+\alpha_1-i)*(\beta_2+x_.)^{-(i+\alpha_2)}\Gamma(i+\alpha_2)}$$

Appendix B. Posterior means of $\boldsymbol{\lambda_1}$ and $\boldsymbol{\lambda_2}$ given $\underline{\boldsymbol{x}}$

Bayesian estimate for λ_1:

$$E\left(\lambda_1|\underline{x}\right) = \iint_0^\infty \lambda_1\ f\left(\underline{\lambda}|\underline{x}\right)d\lambda_1 d\lambda_2$$

$$= \iint_0^\infty \lambda_1\left[\frac{\sum_{i=0}^{n}\binom{n}{i}\lambda_1^{n+\alpha_1-i-1}\lambda_2^{i+\alpha_2-1}e^{-\lambda_1(\beta_1+x_.)-\lambda_2(\beta_2+x_.)}}{\sum_{i=0}^{n}\binom{n}{i}\,(\beta_1+x_.)^{-(n+\alpha_1-i)}\Gamma(n+\alpha_1-i)*(\beta_2+x_.)^{-(i+\alpha_2)}\Gamma(i+\alpha_2)}\right]d\lambda_1 d\lambda_2$$

(Hint: $\int_0^\infty x^{\alpha-1}\,e^{-\beta x}dx = \beta^{-\alpha}\Gamma(\alpha)$)

$$= \frac{\sum_{i=0}^{n}\binom{n}{i}(\beta_1+x_.)^{-(n+\alpha_1-i+1)}\Gamma(n+\alpha_1-i+1)*(\beta_2+x_.)^{-(i+\alpha_2)}\Gamma(i+\alpha_2)}{\sum_{i=0}^{n}\binom{n}{i}\,(\beta_1+x_.)^{-(n+\alpha_1-i)}\Gamma(n+\alpha_1-i)*(\beta_2+x_.)^{-(i+\alpha_2)}\Gamma(i+\alpha_2)}$$

Bayesian estimate for λ_2:

$$E\left(\lambda_2|\underline{x}\right) = \iint_0^\infty \lambda_2\ f\left(\underline{\lambda}|\underline{x}\right)d\lambda_1 d\lambda_2$$

$$= \iint_0^\infty \lambda_1\left[\frac{\sum_{i=0}^{n}\binom{n}{i}\lambda_1^{n+\alpha_1-i-1}\lambda_2^{i+\alpha_2-1}e^{-\lambda_1(\beta_1+x_.)-\lambda_2(\beta_2+x_.)}}{\sum_{i=0}^{n}\binom{n}{i}\,(\beta_1+x_.)^{-(n+\alpha_1-i)}\Gamma(n+\alpha_1-i)*(\beta_2+x_.)^{-(i+\alpha_2)}\Gamma(i+\alpha_2)}\right]d\lambda_1 d\lambda_2$$

(Hint: $\int_0^\infty x^{\alpha-1}\,e^{-\beta x}dx = \beta^{-\alpha}\Gamma(\alpha)$)

$$= \frac{\sum_{i=0}^{n}\binom{n}{i}(\beta_1+x_.)^{-(n+\alpha_1-i)}\Gamma(n+\alpha_1-i)*(\beta_2+x_.)^{-(i+\alpha_2+1)}\Gamma(n+\alpha_1-i+1)}{\sum_{i=0}^{n}\binom{n}{i}\,(\beta_1+x_.)^{-(n+\alpha_1-i)}\Gamma(n+\alpha_1-i)*(\beta_2+x_.)^{-(i+\alpha_2)}\Gamma(i+\alpha_2)}$$

Appendix C. Plots of the exact parametric and Bayesian 95% confidence intervals for $F_{(1)}(x)$.