Proceedings of Machine Learning Research 103:49–58, 2019 ISIPTA 2019

Possibility Measures for Valid Statistical Inference Based on Censored Data

Joyce Cahoon JYU21@NCSU.EDU
Ryan Martin RGMARTI3@NCSU.EDU
Department of Statistics, North Carolina State University, USA

Abstract

Inferential challenges that arise when data are corrupted by censoring have been extensively studied under the classical frameworks. In this paper, we provide an alternative approach based on a generalized inferential model whose output is a data-dependent possibility distribution. This construction is driven by an association between the censored data, parameter of interest, and unobserved auxiliary variable that takes the form of a relative likelihood. The possibility distribution then emerges from the introduction of a nested random set designed to predict that unobserved auxiliary variable and is calibrated to achieve certain frequentist guarantees. The performance of the proposed method is investigated using real and simulated data.

Keywords: Inferential model; Kaplan–Meier estimator; likelihood; Monte Carlo; random set.

1. Introduction

Data are said to be *censored* when at least one of the observations is incomplete, i.e., only an interval that contains the actual value is available. For example, in clinical trials or other time-to-event studies, it may happen that only a lower bound for the event time is observed because subjects drop out of the study, or the study ends before the event takes place. This is called right-censoring. Alternatively, in environmental applications, it may happen that only an upper bound on a chemical content is observed because the available device is limited to a certain detection level. This is called left-censoring. Of course, a combination of left- and right-censoring, or interval-censoring, is possible as well. Beyond censoring direction, there are also Type I and Type II classifications, but we refer the reader to Klein and Moeschberger (2003) for these details. For concreteness, we focus on Type I right-censored data in a time-to-event setting, but it is easy to apply the same ideas for left- or interval-censored data and for contexts other than time.

Let X_i denote the event time and C_i the censoring time for unit $i = 1, \ldots, n$. Under right censoring, the observed data consists of the pair

$$T_i = \min(X_i, C_i), \quad D_i = 1(X_i \le C_i), \quad i = 1, \ldots, n, \quad (1)$$

where $1(\cdot)$ is the indicator function, so that D_i identifies whether T_i is an event time or a censoring time. Let $Y = \{(T_i, D_i) : i = 1, \ldots, n\}$ denote the observable data.

A common assumption that we will adopt here is that of *random censoring*, where $X_1, \ldots, X_n$ are independent and identically distributed (iid) with continuous distribution function F_θ, depending on a parameter $\theta \in \Theta$; $C_1, \ldots, C_n$ are iid with distribution function G; and the X_i's and C_i's are independent of one another (Lawless, 2011). Since the variables are time (or some other "amount"), the statistical models, F_θ, considered here and throughout the literature on this topic are supported on subsets of $(0, \infty)$ and are typically right-skewed. The goal is to make inference on the unknown parameter θ of the time-to-event distribution; G is an unknown nuisance parameter assumed to have no dependence whatsoever on θ.

For data $y = \{(t_i, d_i)\}$ observed from a random, Type I, right-censored data generating process, Klein and Moeschberger (2003, Sec. 3.5) gives the likelihood function

$$L_y(\theta) = \prod_{i=1}^{n} f_\theta(t_i)^{d_i} \bar{F}_\theta(t_i)^{1-d_i}, \quad \theta \in \Theta, \quad (2)$$

where $f_\theta = F_\theta'$ and $\bar{F}_\theta = 1 - F_\theta$ are the density and survival functions corresponding to F_θ, respectively. From the likelihood in (2), it is relatively straightforward to produce point estimates, asymptotic confidence regions, or even Bayesian posterior distributions (Ibrahim et al., 2001). These results, however, are not fully satisfactory as their coverage probabilities can be far from the target in finite samples.

In this paper, we take an alternative approach to construct an *inferential model* whose output takes the form of a data-dependent possibility distribution (Dubois and Prade, 2012). This construction relies on a particular connection between the data, parameter, and an unobservable auxiliary variable. Here, following the recommendations in Martin (2015, 2018), we make use of an association driven by the relative likelihood partially determined in (2). The possibility distribution arises from the introduction of a (nested) random set aimed to predict that unobserved auxiliary variable. An important consequence of this particular construction is that the possibility distribution output inherits a calibration or *validity* property. A precise statement is given in Section 2, but validity implies that the confidence, or possibility, regions derived from the inferential model achieve nominal frequentist coverage probability.

Unfortunately, the presence of censoring complicates the basic inferential model construction and validity properties described in the references above. Here we propose an extension of that basic approach, one that makes novel use of the Kaplan–Meier estimator (e.g., Kaplan and Meier, 1958) for the censoring distribution G. In particular, this estimator is embedded naturally in an algorithm for evaluating the possibility distribution via Monte Carlo. From this, one can immediately evaluate the necessity and possibility of any hypothesis about θ, and inference drawn from these values is valid, at least approximately, in the sense described in Section 2. Details of this construction and its properties are presented in Section 3 and numerical examples comparing the proposed solution to that of more traditional methods are given in Section 4. Finally, some concluding remarks are given in Section 5.

2. Inferential Models

For observable data $Y \in \mathbb{Y}$, consider a statistical model $\{\mathsf{P}_{Y|\theta} : \theta \in \Theta\}$ that contains candidate probability distributions for Y, indexed by a parameter space Θ; throughout we write $Y \sim \mathsf{P}_{Y|\theta}$ to mean "Y is distributed according to $\mathsf{P}_{Y|\theta}$." As presented in Martin and Liu (2013, 2015), an inferential model is a map from the available inputs, including observed data and posited statistical model, to a data-dependent function, $b_y : 2^\Theta \to [0,1]$, where $b_y(A)$ denotes the data analyst's degree of belief about the hypothesis $A \subseteq \Theta$ based on the observed data $Y = y$. Naturally, inferences would be drawn from b_y. This definition of inferential model encompasses many different approaches, including those based on additive beliefs, e.g., Bayes, fiducial, and others, as well as non-additive beliefs like those discussed below, among others. It is up to the data analyst to specify their own degrees of belief, these cannot be derived uniquely from the data and posited statistical model. Note that, while additive beliefs are the most common in statistical applications, if the data is not especially informative, then perhaps both $b_y(A)$ and $b_y(A^c)$ should be small, and a non-additive b_y would facilitate this.

What other properties should b_y have? If the goal is just individual decision-making, then the beliefs based on b_y are what they are, hence nothing is left to be done. But in scientific applications, like we have in mind here, if it is desired that large $b_y(A)$ be interpreted as support for the claim that A^c is false, then it becomes essential that the degrees of belief be calibrated so that we know what a "large" b_y means, and consequently avoid making "systematically misleading conclusions" (Reid and Cox, 2015). We formalize this need for an inferential model to be calibrated in terms of the following validity constraint: that b_y satisfies

$$\sup_{\theta \notin A} \mathsf{P}_{Y|\theta}\{b_Y(A) > 1-\alpha\} \leq \alpha, \quad \begin{cases} \forall\, \alpha \in (0,1) \\ \forall\, A \subseteq \Theta. \end{cases} \tag{3}$$

That is, if the hypothesis A is false, so that $A \not\ni \theta$, the degree of belief $b_Y(A)$, as a function of $Y \sim \mathsf{P}_{Y|\theta}$, will be stochastically no larger than $\mathsf{Unif}(0,1)$. This validity condition can equivalently be expressed in terms of the plausibility function, $p_y(A) = 1 - b_y(A^c)$, the belief function's dual (Shafer, 1976). This dual inferential model output is valid if

$$\sup_{\theta \in A} \mathsf{P}_{Y|\theta}\{p_Y(A) \leq \alpha\} \leq \alpha, \quad \begin{cases} \forall\, \alpha \in (0,1) \\ \forall\, A \subseteq \Theta. \end{cases} \tag{4}$$

Following this constraint, the plausibility output is placed on an objective $\mathsf{Unif}(0,1)$ scale and is said to be valid. In other words, uniform quantiles can be used to interpret the observed degrees of plausibility (or belief) magnitudes, and decisions based on such an interpretation will control frequentist error rates (Martin, 2018).

Based on the *false confidence theorem* in Balch et al. (2017), Martin (2019) argues that validity as in (3) requires that the degrees of belief be non-additive. Since we take this validity property to be fundamental to the logic of statistical inference, we focus here on genuinely non-additive degrees of belief, in particular, necessity/possibility functions (e.g., Dubois and Prade, 2012; Dubois, 2006).

How to construct a valid inferential model? The original construction in Martin and Liu (2013), starts with an association, i.e., a characterization of the statistical model based on what is called an auxiliary variable. The prototype for this takes the form

$$Y = a(\theta, U), \quad U \sim \mathsf{P}_U, \tag{5}$$

where a is a given function and P_U is a distribution for $U \in \mathbb{U}$ that does not depend on any unknown parameters. This describes an algorithm for simulating from $\mathsf{P}_{Y|\theta}$ but also guides our intuition about inference. That is, *if U were observable, along with Y, then the best possible inference follows by simply solving* (5) *for θ*, as in (6). Since U is actually unobservable, it is tempting to create a sort of "posterior distribution" for θ by taking draws from P_U, plugging them into (5), with the observed $Y = y$, and solving for θ. This is basically Fisher's fiducial argument (e.g., Fisher, 1973; Dempster, 1963; Hannig et al., 2016), which generally leads to additive beliefs that necessarily fail to meet the validity condition. Non-additivity can be introduced by stretching fiducial's draws from P_U into a random set designed to hit the unobserved value of U in (5) that corresponds to the observed $Y = y$ and the true value of θ. The following three steps summarize this construction.

A-step Given the *association* (5) and the observed $Y = y$, define the focal elements

$$\Theta_y(u) = \{\theta : y = a(\theta, u)\}, \quad u \in \mathbb{U}. \tag{6}$$

P-step Introduce a random set $\mathcal{S} \sim \mathsf{P}_\mathcal{S}$, taking values in $2^\mathbb{U}$, designed to *predict* the unobserved value of U in (5).

C-step *Combine* the output of the A- and P-steps to get a new random set

$$\Theta_y(\mathcal{S}) = \bigcup_{u \in \mathcal{S}} \Theta_y(u), \quad \mathcal{S} \sim \mathsf{P}_{\mathcal{S}},$$

and define the *belief function,*

$$b_y(A) = \mathsf{P}_{\mathcal{S}}\{\Theta_y(\mathcal{S}) \subseteq A\}, \quad A \subseteq \Theta,$$

and its dual, the *plausibility function,* $p_y(A) = 1 - b_y(A^c)$.

Under very mild conditions on the user-specified random set $\mathcal{S}$, the corresponding inferential model is valid in the sense of (3). Indeed, the only requirement is that $\mathcal{S}$ be calibrated in a certain sense to predicting unobserved draws from P_U. This is relatively easy to arrange because P_U is known and $\mathcal{S} \sim \mathsf{P}_{\mathcal{S}}$ is user-specified. More specifically, let $\gamma(u) = \mathsf{P}_{\mathcal{S}}(\mathcal{S} \ni u)$, an ordinary function on $\mathbb{U}$, be determined implicitly by $\mathsf{P}_{\mathcal{S}}$; note that γ is the plausibility contour corresponding to $\mathcal{S}$. Then validity as in (3) corresponds to a stochastic dominance property, namely, $\gamma(U) \geq_{\text{st}} \mathsf{Unif}(0,1)$. For example, in what follows, we work with a random set $\mathcal{S}$ of the form

$$\mathcal{S} = [\tilde{U}, 1], \quad \tilde{U} \sim \mathsf{P}_U := \mathsf{Unif}(0,1), \tag{7}$$

so that $\gamma(u) = u$ and, hence, $\gamma(U) = U \sim \mathsf{Unif}(0,1)$. Though not strictly necessary for validity, efficiency considerations suggest that $\mathcal{S}$ be nested, like in (7), which makes the belief function consonant; the validity property together with consonance is reminiscent of the confidence structure developments in Balch (2012). With this construction, the inferential model output becomes a necessity and possibility function pair, which we will henceforth denote as nec_y and pos_y, respectively.

As Martin (2018) argued, the above formulation can be rather rigid; greater flexibility and, in some cases, improved performance can be gained by working with a so-called generalized association, one that does not fully characterize the posited statistical model. We describe this generalized association in the present context below and propose a Monte Carlo strategy to overcome the challenges that arise when data are corrupted by censoring.

3. A Valid Inferential Model under Censoring

Consider a situation like the censored-data problem described in Section 1, where a likelihood function is straightforward, but a formal association that encodes the data-generation process is difficult to ascertain. One can consider a *generalized association* of the form

$$R_{Y,\theta} = H_\theta^{-1}(U), \quad U \sim \mathsf{P}_U = \mathsf{Unif}(0,1), \tag{8}$$

where $R_{y,\theta}$ is some real-valued function of (y, θ) and H_θ is its distribution function,

$$H_\theta(r) = \mathsf{P}_{Y|\theta,G}\{R_{Y,\theta} \leq r\}, \quad r \in \mathbb{R}.$$

The distribution $\mathsf{P}_{Y|\theta,G}$ is that of Y defined according to the rule (1) with (θ, G) as the unknown parameters. Unlike (5), (8) does not describe the data-generation process, it only establishes a relationship between data, parameter, and auxiliary variable, which is all that was needed for the inferential model construction described above.

The advantage of this generalized association is that we have directly reduced the dimension of the auxiliary variable, from at least the dimension of θ down to 1. This greatly simplifies the construction of a (good) random set $\mathcal{S}$ for predicting that unobservable quantity. But what is an appropriate choice of $R_{y,\theta}$? The options are indeed unlimited. Since dimension reduction would generally result in loss of information, and since we prefer to retain as much information as possible, we opt to take $R_{y,\theta}$ as the *relative likelihood*

$$R_{y,\theta} = L_y(\theta)/L_y(\hat{\theta}), \tag{9}$$

where $\hat{\theta}$ is the maximum likelihood estimator, i.e., $\hat{\theta} = \arg\max_\vartheta L_y(\vartheta)$. Extensive studies have explored the use of relative likelihood to define degrees of belief (e.g., Shafer, 1976; Wasserman, 1990), but they focus on examples where the likelihood cannot be normalized or where a normalized likelihood is misleading (Shafer, 1982). Our approach differs in the sense that we can evaluate the distribution of the relative likelihood by Monte Carlo. From here, the inferential model construction is conceptually straightforward.

A-step Set $\Theta_y(u) = \{\theta : R_{y,\theta} = H_\theta^{-1}(u)\}$ for $u \in [0,1]$.

P-step Define $\mathcal{S} = [\tilde{U}, 1]$, where $\tilde{U} \sim \mathsf{Unif}(0,1)$ like in (7); so that the distribution, $\mathsf{P}_{\mathcal{S}}$, is fully determined by the uniform distribution.

C-step Combine the two sets above to get

$$\begin{aligned}\Theta_y(\mathcal{S}) &= \bigcup_{u \in \mathcal{S}} \Theta_y(u) \\ &= \{\theta : H_\theta(R_{y,\theta}) \geq \tilde{U}\}, \quad \tilde{U} \sim \mathsf{Unif}(0,1).\end{aligned}$$

Then the possibility contour is

$$\mathsf{pos}_y(\theta) := \mathsf{P}_{\mathcal{S}}\{\Theta_y(\mathcal{S}) \ni \theta\} = H_\theta(R_{y,\theta}), \quad \theta \in \Theta,$$

which determines the possibility and necessity functions.

An important observation, as discussed in Martin (2018), is that the generalized association above can relate the data Y and parameter θ to a *scalar* auxiliary variable U. In the basic inferential model construction of Martin and Liu (2013), usually the dimension of U is determined by that of θ, which subsequently requires the specification of a "good" multi-dimensional random set. While we are not inhibited by such a challenge in the present context, our reviewers have suggested p-boxes from Troffaes and Destercke (2011) as a useful strategy in future extensions. Here, however, the baseline formulation is in terms of a scalar auxiliary variable and the structure immediately suggests a good random set $\mathcal{S}$ in (7).

That the corresponding inferential model satisfies the validity property follows immediately from the arguments presented in Martin (2018). Since our predictive random sets are tailored such that the possibility contours are stochastically no larger than uniform, i.e., $H_\theta(R_{Y,\theta}) \leq_{\rm st} \mathsf{Unif}(0,1)$ when $Y \sim \mathsf{P}_{Y|\theta,G}$, and therefore

$$\sup_{\theta \in A} \mathsf{P}_{Y|\theta,G}\{\mathsf{pos}_Y(A) \leq \alpha\} \leq \alpha, \quad \begin{cases} \forall\, \alpha \in (0,1) \\ \forall\, A \subseteq \Theta. \end{cases} \tag{10}$$

A desirable consequence of validity is that confidence regions having the nominal frequentist coverage probability can be constructed immediately based on the possibility function output. Indeed, the set

$$\{\theta : \mathsf{pos}_y(\theta) > \alpha\} \tag{11}$$

is a nominal $100(1-\alpha)\%$ confidence region for any $\alpha \in (0,1)$. This follows since the probability that the above region contains the true parameter value θ equals the probability that $\mathsf{pos}_Y(\theta) > \alpha$ which, in turn, equals $1-\alpha$.

Putting the above inferential model construction into practice requires that the distribution function H_θ be evaluated, at least approximately, for every θ. This is straightforward to do, albeit potentially tedious computationally, when data are not censored. This is similarly straightforward if data are censored but the censoring distribution G is known. Indeed, a simple Monte Carlo approximation is available:

$$H_\theta(r) \approx \frac{1}{M} \sum_{m=1}^{M} 1\{R_{Y^{(m)},\theta} \leq r\}, \tag{12}$$

where $\{Y^{(m)} : m = 1,\ldots,M\}$ are independent copies of $Y^* = \{(T_i^*, D_i^*) : i = 1,\ldots,n\}$ and (T_i^*, D_i^*) as in (1), with X_i^* iid from F_θ and C_i^* iid from the known censoring distribution G. However, in our present context, H_θ depends (implicitly) on the unknown distribution G of censoring times, so something more sophisticated than that simple strategy just described is needed. Here we recommend using a plug-in estimator of G.

The Kaplan–Meier estimator was not originally designed to estimate the censoring distribution function, but it is straightforward to simply reverse the event/censored classification. That is, we still observe $T_i = \min(X_i, C_i)$ but now we think of C_i as the "event time" and X_i is the "censoring time." Then we construct the Kaplan–Meier estimator, $\widehat{G}$, of G based on this alternative perspective.

After swapping the observed/censored classifications, obtaining the Kaplan–Meier estimate is straightforward; we use the built-in functions in R's `survival` package (Therneau, 2015). But there are a few technical points worth making about the estimation process. Recall that, in typical applications of the Kaplan–Meier estimator of a survival function $S(t)$, if the largest observation corresponds to a "censored" outcome, then $\hat{S}(t)$ *does not* vanish as $t \to \infty$, which amounts to putting some positive amount of mass at ∞. In our context, since we interpret the original event times as censored times, our estimate $\widehat{G}$ will put positive mass at ∞ when the largest observation is an event. Therefore, when the censoring level is relatively high, we expect our $\widehat{G}$ to put mass at ∞, which means that some C_i^*'s drawn from $\widehat{G}$ will equal ∞ and, consequently, T_i^* corresponds to an event time as $X_i^* < C_i^*$. This ensures that the Monte Carlo sample, Y^*, reflects the censoring level in the original data.

Can anything be said about validity of the inferential model derived from the above algorithm with the plug-in estimator $\widehat{G}$? That is, can we conclude that

$$\mathsf{P}_{Y|\theta,G}\{\mathsf{pos}_Y(\theta;\widehat{G}) \leq \alpha\} \leq \alpha,$$

at least approximately? Here $\mathsf{pos}_y(\theta;\widehat{G})$ denotes the possibility function obtained by applying the above algorithm with $\widehat{G}$ plugged in for the unknown G, i.e., simulating C_i^*'s iid from $\widehat{G}$ instead of the unknown G. The dependence of $\mathsf{pos}_y(\theta;\widehat{G})$ on the Kaplan–Meier estimator, an infinite-dimensional quantity, is quite complicated, so a precise mathematical result is not yet available. But we can give some strong heuristics and numerical results to support a conjecture of (approximate) validity. First, the Kaplan–Meier estimator itself is known to have certain desirable asymptotic properties such as consistency (e.g., Meier, 1967) and a fast rate of convergence (e.g., Fleming and Harrington, 1991, Chap. 6), which suggests that

$$\mathsf{pos}_y(\theta;\widehat{G}) \approx \mathsf{pos}_y(\theta;G) =: \mathsf{pos}_y(\theta)$$

and, hence, $\mathsf{pos}_Y(\theta;\widehat{G})$ is approximately $\mathsf{Unif}(0,1)$ because $\mathsf{pos}_Y(\theta;G)$ is. Therefore, consistency of $\widehat{G}$ and validity in the known-G case together suggest approximate validity of the proposed plug-in approach. Second, numerical experiments support our claim of approximate validity. In one example, we take 10,000 samples of size $n = 50$ in which X_i's are generated from a standard exponential subject to random right censoring from the $\mathsf{Unif}(0,5)$. A Monte Carlo estimate of the distribution function $\alpha \mapsto \mathsf{P}_{Y|\theta,G}\{\mathsf{pos}_Y(\theta;\widehat{G}) \leq \alpha\}$ is shown in Figure 1 is approximately uniform, with further confirmation from the Kolmogorov–Smirnov test, hence our motivation for approximate validity. Simulated- and real-data examples in Section 4 below demonstrate the proposed method's strong performance compared to others, and provide further support for our approximate validity claim.

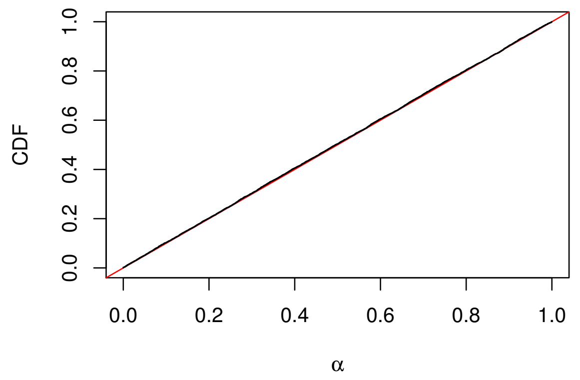

Figure 1: Distribution of $\alpha \mapsto \mathsf{P}_{Y|\theta,G}\{\mathsf{pos}_Y(\theta;\widehat{G}) \leq \alpha\}$ (black) compared with that of $\mathsf{Unif}(0,1)$ (red) based on Monte Carlo samples from a standard exponential distribution subject to random right censoring. The average censoring level among all 10,000 replications at this setting is 19.9%.

4. Examples

We compare our proposed approach against frequentist and Bayesian methods with simulated and real data. The exponential and Weibull examples are taken from the `survival` package in R, while the last log-normal example is taken from Krishnamoorthy and Xu (2011). We consider these three parametric distributions that are commonly used in time-to-event analyses, and we generate 10,000 replications of censored data under various settings of these distributions. We repeat each set of simulations at four sample sizes of $n \in \{15, 20, 25, 50\}$. As our following results suggest, possibility measures consistently outperform the more familiar methods, achieving nearly the nominal $100(1-\alpha)\%$ coverage rate across different distributions, parameter settings, and sample sizes.

4.1. Exponential

The classic time-to-event distribution is exponential, characterized by a constant hazard rate $\theta > 0$, in which the density function is $f_\theta(t) = \theta e^{-\theta t}$. For n items, independently subject to random right censoring, summarized by $y = \{(t_i, d_i)\}$ as above, the maximum likelihood estimate is $\hat{\theta} = \sum_{i=1}^n d_i / \sum_{i=1}^n t_i$. From its asymptotic normality, a 95% confidence interval is easily obtained as $\hat{\theta} \pm 1.96 I(\hat{\theta})^{-1/2}$, where $I(\hat{\theta})$ is the observed information. From a Bayesian standpoint, the censoring mechanism can be safely ignored as the likelihood can be formed from (2) and combined with a conjugate $\mathsf{Gamma}(\alpha_0, \beta_0)$ prior to arrive at the posterior $\mathsf{Gamma}(\alpha_0 + \sum_{i=1}^n d_i, \beta_0 + \sum_{i=1}^n t_i)$. Posterior credible intervals are then easily obtained.

From an inferential model perspective, we begin with the baseline association of the relative likelihood for $\theta \in \Theta$,

$$\frac{\theta^{\sum_i D_i} e^{-\theta \sum_i T_i}}{\hat{\theta}^{\sum_i D_i} e^{-\hat{\theta} \sum_i T_i}} = H_\theta^{-1}(U), \quad U \sim \mathsf{P}_U = \mathsf{Unif}(0,1). \quad (13)$$

As described above, we write $R_{Y,\theta}$ for the left-hand side of the above display. For fixed data y, we follow through our A-step with the singleton-valued map

$$\Theta_y(u) = \{\theta : R_{y,\theta} = H_\theta^{-1}(u)\}, \quad u \in [0,1].$$

Next, the P-step requires introducing a predictive random set $\mathcal{S}$ in (7) for U. We then combine our A- and P-steps

$$\Theta_y(\mathcal{S}) = \bigcup_{u \in \mathcal{S}} \Theta_y(u) = \{\theta : H_\theta(R_{y,\theta}) \geq \tilde{U}\}, \quad \tilde{U} \sim \mathsf{Unif}(0,1).$$

And we summarize the distribution of this random set $\Theta_y(\mathcal{S})$ by a possibility function

$$\mathsf{pos}_y(\theta) = H_\theta(R_{y,\theta}), \quad \theta > 0.$$

A $100(1-\alpha)\%$ "confidence interval" can be obtained as the upper level set of the possibility function as in (11). Evaluating this possibility function requires Monte Carlo procedure discussed in Section 3.

For numerical illustration, we simulate 10,000 replications of lifetimes arising from nine different θ settings in the exponential distribution. For each of these 90,000 simulations, the lifetimes $X_1, \ldots, X_n \sim F_\theta$ generated were subject to random right censoring from $C_1, \ldots, C_n \sim \mathsf{Unif}(0,5)$, allowing us to compare the coverage of our inference procedure under a wide range of censoring levels. Results shown in Figure 2 demonstrate that the nominal $100(1-\alpha)\%$ coverage is attainable, thus supporting our validity conjecture.

For a real-data illustration, we consider the primary biliary cirrhosis (PBC) data from a clinical trial at the Mayo Clinic from 1974 to 1984. The data consists of $n = 312$ recorded survival times for patients involved in the randomized trial, along with a corresponding right censoring indicator; there are 168 censored cases, more than 50% of total observations. Figure 3 shows the point possibility function $\mathsf{pos}_y(\theta)$ for a range of parameter values, along with the corresponding 95% possibility interval (11). For comparison, 95% confidence intervals based on asymptotic normality of the maximum likelihood estimate are also displayed. The intervals derived from the possibility function are almost indistinguishable from the likelihood-based intervals, which is a sign of our proposed approach's efficiency, since the latter are the asymptotically "best" intervals.

4.2. Weibull

One of the most widely used time-to-event distributions is the Weibull, with applications in manufacturing, health,

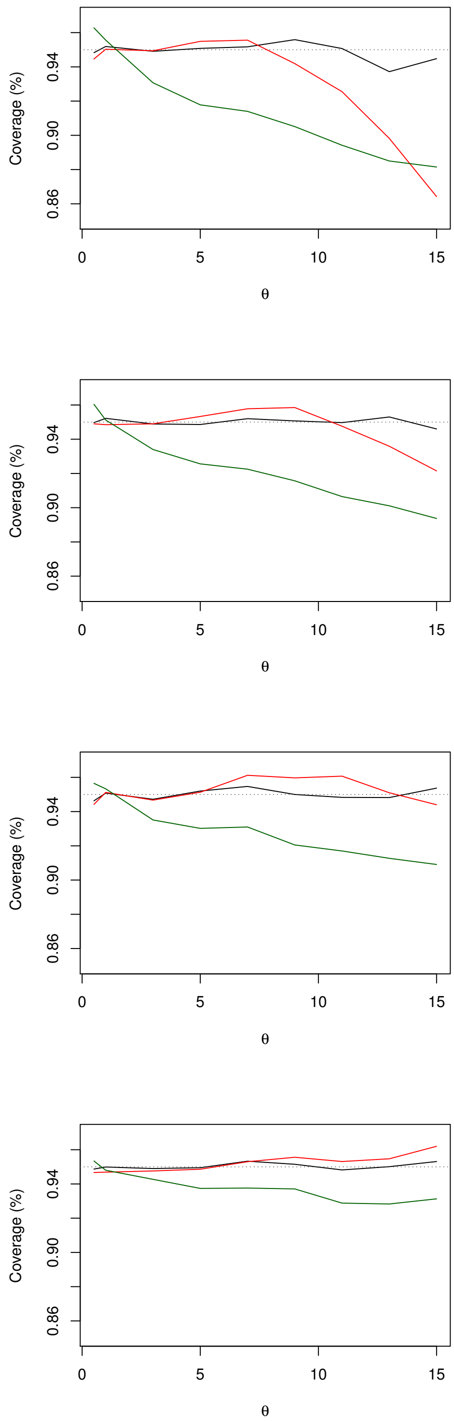

Figure 2: Coverage probability of the 95% possibility region for θ in the exponential model (black). Results compared to those based on maximum likelihood (red) and Bayesian with a $\mathsf{Gamma}(2,1)$ prior (green). From top to bottom, data are generated with sample size $n = 15, 20, 25, 50$.

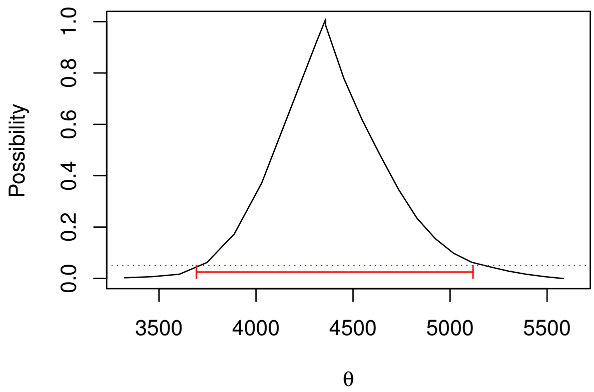

Figure 3: Point possibility functions for the mean in the PBC example under an exponential model (black). Reference line at $\alpha = 0.05$ (dotted) and approximate 95% confidence intervals based on maximum likelihood (red).

etc., as it has sufficient flexibility to capture changes in the hazard rate (Lawless, 2011). Exponential is a special case of Weibull when the shape parameter $\beta = 1$. The density and survival functions, indexed by $\theta = (\beta, \lambda)$, are

$$f_\theta(t) = \lambda \beta t^{\beta-1} \exp(-\lambda t^\beta), \quad \bar{F}_\theta(t) = \exp(-\lambda t^\beta).$$

Similar to the setup as described for the exponential example, we compare the performance of our proposed approach against that of a more traditional frequentist or objective Bayesian approach. An inferential model requires that we simulate the distribution of $R_{Y,\theta}$; so for a finite grid of $\theta = (\beta, \lambda)$ values, for each pair, 500 Monte Carlo samples of Y^* are obtained by taking the minimum between realizations of $X^* \sim \mathsf{Weib}(\beta, \lambda)$ and $C^* \sim \hat{G}$, the modified Kaplan–Meier estimate. We implement this procedure for 10,000 replications of lifetimes arising from six different settings of the Weibull distribution. These 60,000 replications were each subject to random right censoring from $G \sim \mathsf{Unif}(0,4)$. As shown in Figure 5, we demonstrate good agreement between nominal and actual coverage from our new approach in finite sample settings.

For a real-data example, we consider survival data on ovarian cancer patients from a clinical trial that took place from 1974 to 1977. This data set has $n = 26$ survival times for patients that entered the study with stage II or IIIA cancer and were treated with cyclophosphamide alone or cyclophosphamide with adriamycin. Of this patient group, 14 survived (or was censored) by the end of the study, while 12 died (Edmonson et al., 1979). Despite the small sam-

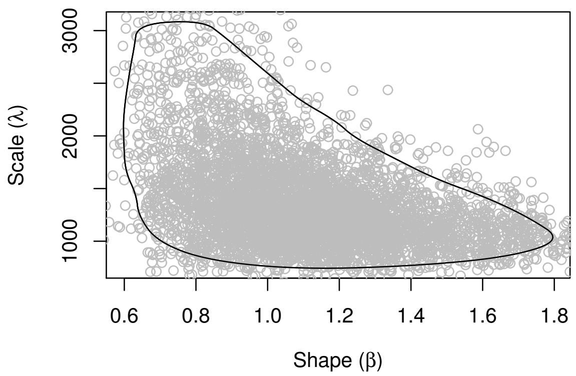

Figure 4: Possibility contour (black) for $\theta = (\beta, \lambda)$, the shape and scale parameter pair, in the ovarian cancer data under a Weibull model subject to Type I right censoring. Bayesian posterior samples based on a $\mathsf{Gamma}(1, 0.1)$ prior for the shape and $\mathsf{N}(0, 10)$ prior for the log transformed scale parameter (gray).

ple size and high censoring level, our possibility contours capture the non-elliptical shape as shown by the Bayesian posterior in Figure 4.

4.3. Log-Normal

Within environmental science, the log-normal distribution is often used to approximate data that are censored to the left, e.g., chemical pollutants that can only be detected above some minimal threshold (Krishnamoorthy and Xu, 2011). The density function, indexed by $\theta = (\mu, \sigma)$, is

$$f_\theta(t) = \frac{1}{(2\pi)^{1/2}\sigma t} \exp\Big\{-\frac{1}{2}\Big(\frac{\log t - \mu}{\sigma}\Big)^2\Big\}.$$

Similar to our examples above, we compare the coverage performance of our possibility contours against that of ellipses based on asymptotic normality of the maximum likelihood estimator and posterior credible regions based on a $\mathsf{Gamma}(1, 0.1)$ prior on the precision $\tau = \sigma^{-2}$ and $\mathsf{N}(0, 1000/\tau)$ prior on the mean. Again, 10,000 replications of censored data were generated from 11 different settings of the log-normal distribution, subject to left censoring under $G \sim \mathsf{Unif}(0, 1)$. In order to approximate the distribution of $R_{Y,\theta}$, however, our modified Kaplan–Meier estimate $\hat{G}$ now requires putting positive mass at 0 when the smallest observation corresponds to an actual event record, so the challenges we encountered under right censoring are simply reversed. As shown in Figure 8, under various

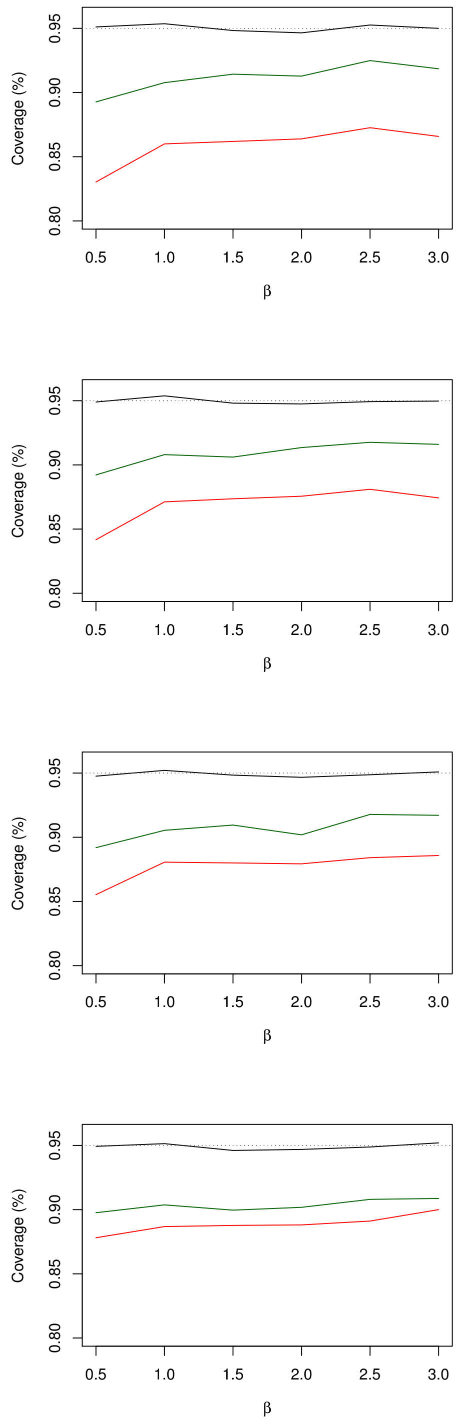

Figure 5: Coverage probability of the 95% possibility region for $\theta = (\beta, \lambda)$ in the Weibull model (black). Results compared to maximum likelihood (red) and Bayesian intervals based on a $\mathsf{Gamma}(0.1, 1)$ prior on the shape and $\mathsf{N}(0, 10)$ prior on the log transformed scale (green). From top to bottom, data are generated from a fixed sample size of $n = 15, 20, 25, 50$.

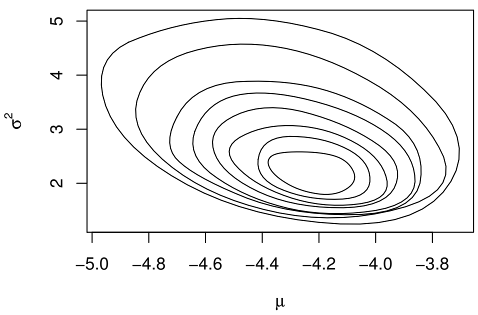

Figure 6: Possibility contours at each $\alpha = 10\%$ increment level beginning at 20% for the Atrazine example under a log-normal model with Type I left censoring.

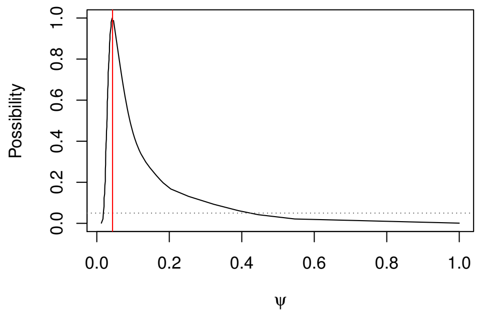

Figure 7: Marginal possibility function for inference on the mean $\psi = \exp(\mu + \sigma^2/2)$ in the Atrazine example. Reference line at $\alpha = 0.05\%$ (dotted) and the maximum likelihood estimate (red).

censoring levels, our proposed method outperforms with nominal $100(1-\alpha)\%$ coverage.

We use Atrazine concentration data collected from a well in Nebraska as an example. This set of 24 observations were randomly subject to two lower detection limits of 0.01 and 0.05 μg/l of which 11 observations were censored. Despite this censoring level of 45.8%, previous studies indicate the log-normality assumption holds (Helsel, 2005). We apply our Monte Carlo approach to determine the joint possibility contours for $\theta = (\mu, \sigma^2)$ in Figure 6, along with the marginal possibility function for the log-normal mean, $\psi = \exp(\mu + \sigma^2/2)$, in Figure 7. The point at which we assign the highest possibility aligns with the maximum

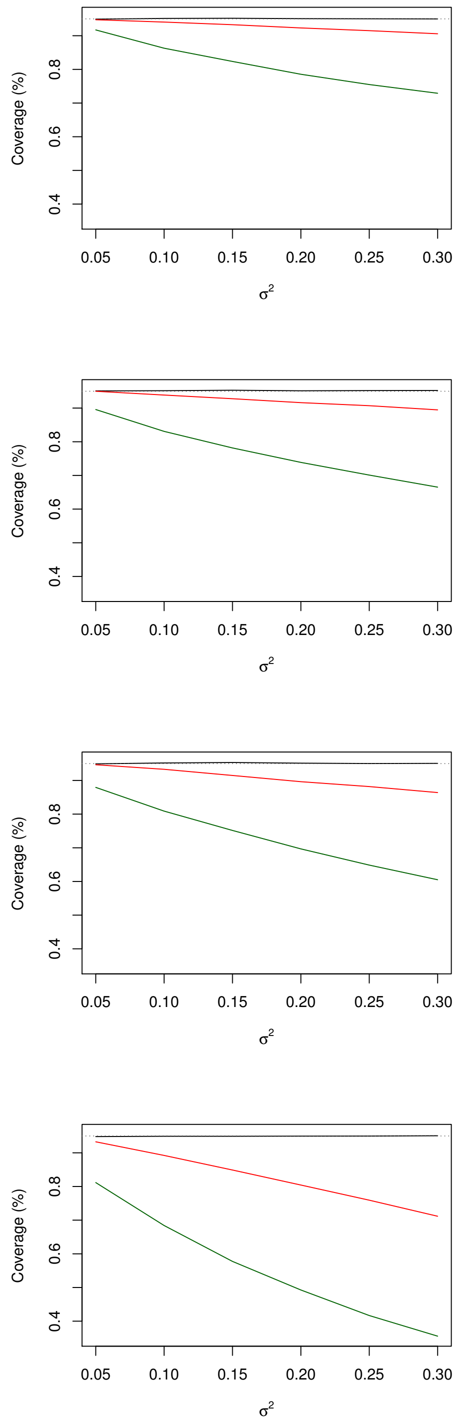

Figure 8: Coverage probability of the 95% possibility interval for $\psi = \exp(\mu + \sigma^2/2)$ in the log-normal model (black). Results compared to maximum likelihood (red) and Bayesian intervals based on a $\mathsf{Gamma}(1, 0.1)$ prior on the precision $\tau = \sigma^{-2}$ and $\mathsf{N}(0, 1000/\tau)$ prior on the mean (green). From top to bottom, data are generated from a fixed sample size of $n = 15, 20, 25, 50$.

likelihood estimator, $\hat{\mu} = -4.206$ and $\hat{\sigma} = 1.462$ (Krishnamoorthy and Xu, 2011).

5. Conclusion

In this paper, we proposed a specific inferential model construction for contexts in which the data are corrupted via censoring. The main obstacle is that the censoring time distribution is a unknown; despite not being of scientific interest, the presence of an infinite-dimensional nuisance parameter complicates the inferential model construction. To overcome this challenge, we extend the generalized inferential model framework in Martin (2018) to cover the case of censoring according to a distribution G, and then we propose a plug-in approximation to the known-G inferential model construction with one that relies on a modified version of the classical Kaplan–Meier estimator, swapping the roles of event and censoring times. While a fully rigorous proof of validity for this proposed approach is still lacking, we provided here strong heuristics to support that conjecture, along with convincing numerical results across a range of settings. We demonstrate numerically that the proposed inferential model approach outperforms the more traditional maximum likelihood and Bayesian solutions in terms of coverage probability of confidence sets. And the resemblance between ours and the maximum likelihood confidence intervals in large sample size settings indicates that our coverage improvements are not at the expense of efficiency. Thus the proposed inferential model approach merits further investigation.

Aside from efforts to establish the validity property rigorously, it is of interest to explore more complicated and practical types of censored-data models, e.g., ones where censoring depends on covariates so that an assumption of random censoring might not be warranted. In principle, the approach described—with a generalized association based on the distribution of relative likelihood—would also work in more general cases, the optimization and Monte Carlo computations required to evaluate the distribution function H_θ would be much more involved. Ongoing efforts are focused on this and other more general improvements to the simple Monte Carlo computations described here.

References

M. S. Balch. Mathematical foundations for a theory of confidence structures. *International Journal of Approximate Reasoning*, 53(7):1003–1019, 2012.

M. S. Balch, R. Martin, and S. Ferson. Satellite conjunction analysis and the false confidence theorem. *arXiv:1706.08565*, 2017. Unpublished manuscript.

A. P. Dempster. Further examples of inconsistencies in the fiducial argument. *The Annals of Mathematical Statistics*, 34:884–891, 1963.

D. Dubois. Possibility theory and statistical reasoning. *Computational Statistics & Data Analysis*, 51(1):47–69, 2006.

D. Dubois and H. Prade. *Possibility Theory*. Springer, 2012.

J. Edmonson, T. R. Fleming, D. G. Decker, G. D. Malkasian, E. O. Jorgensen, J. A. Jefferies, M. J. Webb, and L. K. Kvols. Different chemotherapeutic sensitivities and host factors affecting prognosis in advanced ovarian carcinoma versus minimal residual disease. *Cancer Treatment Reports*, 63(2):241–247, 1979.

R. A. Fisher. *Statistical Methods and Scientific Inference.* Hafner Press, New York, 3rd edition, 1973.

T. R. Fleming and D. P. Harrington. *Counting Processes and Survival Analysis*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1991.

J. Hannig, H. Iyer, R. Lai, and T. Lee. Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, 111(515):1346–1361, 2016.

D. R. Helsel. *Nondetects and Data Analysis*. Wiley, 2005.

J. G. Ibrahim, M. Chen, and D. Sinha. *Bayesian Survival Analysis*. Springer, 2001.

E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

J. P. Klein and M. L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag, New York, 2003.

K. Krishnamoorthy and Z. Xu. Confidence limits for lognormal percentiles and for lognormal mean based on samples with multiple detection limits. *The Annals of Occupational Hygiene*, 55(5):495–509, 2011.

J. F. Lawless. *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, 2011.

R. Martin. Plausibility functions and exact frequentist inference. *Journal of the American Statistical Association*, 110(512):1552–1561, 2015.

R. Martin. On an inferential model construction using generalized associations. *Journal of Statistical Planning and Inference*, 195:105–115, 2018.

R. Martin. False confidence, non-additive beliefs, and valid statistical inference. *arXiv:1607.05051*, 2019. Unpublished manuscript.

R. Martin and C. Liu. Inferential models: a framework for prior-free posterior probabilistic inference. *Journal of the American Statistical Association*, 108(501):301–313, 2013.

R. Martin and C. Liu. *Inferential Models: Reasoning with Uncertainty*. CRC Press, Boca Raton, FL, 2015.

P. Meier. Nonparametric estimation from incomplete observations. 1967. Unpublished manuscript.

N. Reid and D. R. Cox. On some principles of statistical inference. *International Statistical Review*, 83(2):293–308, 2015.

G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

G. Shafer. Belief functions and parametric models. *Journal of the Royal Statistical Society Series B*, 44(3):322–352, 1982.

Terry M Therneau. *A Package for Survival Analysis in R*, 2015. URL https://CRAN.R-project.org/package=survival. Version 2.38.

M. Troffaes and S. Destercke. Probability boxes on totally preordered spaces for multivariate modelling. *International Journal of Approximate Reasoning*, 52(6): 767–791, 2011.

L. A. Wasserman. Belief functions and statistical inference. *The Canadian Journal of Statistics*, 18(3):183–196, 1990.