Generalized inferential models for censored data

Joyce Cahoon* and Ryan Martin†

November 27, 2021

Abstract

Inferential challenges that arise when data are censored have been extensively studied under the classical frameworks. In this paper, we provide an alternative generalized inferential model approach whose output is a data-dependent plausibility function. This construction is driven by an association between the distribution of the relative likelihood function at the interest parameter and an unobserved auxiliary variable. The plausibility function emerges from the distribution of a suitably calibrated random set designed to predict that unobserved auxiliary variable. The evaluation of this plausibility function requires a novel use of the classical Kaplan–Meier estimator to estimate the censoring rather than the event distribution. We prove that the proposed method provides valid inference, at least approximately, and our real- and simulated-data examples demonstrate its superior performance compared to existing methods.

Keywords and phrases: Kaplan–Meier estimator; plausibility; random set; relative likelihood; survival analysis.

1 Introduction

A class of challenging and practically relevant problems are those where the data are corrupted in some way. Examples of such corruption include missingness, measurement error, coarsening, etc. One special type of corruption, common in time-to-event studies, is *censoring*, where at least one observation is incomplete in the sense that only an interval that contains the actual value is available. For example, in a clinical trial, it may happen that only lower bounds on some patients' remission times are 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 here 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.

*Gray Systems Lab, Microsoft; joyce.cahoon@microsoft.com

†Department of Statistics, North Carolina State University; rgmarti3@ncsu.edu

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

$$T_i = \min(X_i, C_i), \quad D_i = 1(X_i \leq C_i), \quad i = 1, \ldots, n, \tag{1}$$

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 denote an amount of time (or some other quantity), the statistical models, F_θ, considered here and in the literature more generally 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) give the likelihood function

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

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; for example, 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 non-additive, data-dependent belief/plausibility function. 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 derived from (2). The belief function 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 belief function output inherits a calibration or *validity* property. A precise statement is given in Section 2, but an important practical consequence of the validity property is that the confidence, or plausibility, regions derived from the inferential model achieve the nominal frequentist coverage probability.

Unfortunately, the presence of censoring complicates the basic inferential model construction and validity properties described in Martin (2015, 2018). In particular, the distribution G of the censoring times is an infinite-dimensional nuisance parameter whose influence is difficult to overcome. Here we propose an extension of the basic approach above, one that makes novel use of the Kaplan–Meier estimator (e.g., Kaplan and Meier 1958) for the censoring distribution G. We then develop a Monte Carlo algorithm to evaluate the belief and plausibility of any hypothesis about θ, and we show—both theoretically and empirically—that inference drawn from the generalized inferential model output 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 Background

2.1 Basic 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 a running example for this section, consider a binomial problem $Y \sim \mathsf{P}_{Y|\theta} = \mathsf{Bin}(n, \theta)$, where the number of trials, n, is known but θ is unknown and to be inferred. More complicated examples, involving censored data, will be considered in the sections that follow.

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, $\underline{\Pi}_y : 2^\Theta \to [0, 1]$, where $\underline{\Pi}_y(A)$ denotes the data analyst's degree of belief about the hypothesis $A \subseteq \Theta$ based on the observed data $Y = y$. As the notation suggests, the interpretation of the inferential model's degree of belief is similar to that of a lower bound on a collection of posterior probabilities for A, given y. So, naturally, inferences would be drawn based on the magnitudes of $\underline{\Pi}_y(A)$ for various assertions of interest A. This definition of an 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.

What properties should $\underline{\Pi}_y$ have? In the scientific applications we have in mind here, if it is desired that large $\underline{\Pi}_y(A)$ be interpreted as support for the claim that A is true, then it becomes essential that the degrees of belief be calibrated so that we know what a "large" $\underline{\Pi}_y$ 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:

$$\sup_{\theta \notin A} \mathsf{P}_{Y|\theta}\{\underline{\Pi}_Y(A) > 1 - \alpha\} \leq \alpha, \quad \forall\, \alpha \in [0, 1], \quad \forall\, A \subseteq \Theta. \tag{3}$$

That is, if the hypothesis A is false, so that $A \not\ni \theta$, the degree of belief $\underline{\Pi}_Y(A)$, as a function of $Y \sim \mathsf{P}_{Y|\theta}$, will be stochastically no larger than $\mathsf{Unif}(0, 1)$. This property ensures that relatively low degree of belief values will be assigned to false hypotheses, thus protecting the data analyst from systematic errors.

The desire for belief/plausibility functions to satisfy some sort of calibration property relative to a statistical model is not unique to us here. Indeed, Walley (2002) considered a validity condition very similar to—and slightly weaker than—that in (3). His goal was to show that a reconciliation of frequentist calibration properties with generalized Bayes was possible. However, his proposed method to achieve this reconciliation is rather inefficient from a statistical point of view, and the inferential model approach taken here is generally more efficient. Balch (2012) develops a theory of confidence structures with similar motivation and properties to us here but with somewhat less flexibility. Finally, Denœux and Li (2018) provided a review of some of these developments and proposed a new notion of calibration in the context of parameter estimation but this requires fixing an α value on which the degrees of belief will depend.

The validity property (3) can be equivalently expressed in terms of the plausibility function, $\overline{\Pi}_y(A) = 1 - \underline{\Pi}_y(A^c)$, the belief function's dual (Shafer 1976). In terms of this dual inferential model output, validity is satisfied if

$$\sup_{\theta \in A} \mathsf{P}_{Y|\theta}\{\overline{\Pi}_Y(A) \leq \alpha\} \leq \alpha, \quad \forall\, \alpha \in [0,1], \quad \forall\, A \subseteq \Theta. \tag{4}$$

Following this constraint, the plausibility values can be compared to a $\mathsf{Unif}(0,1)$ scale, and decisions based on such comparisons will control frequentist error rates (Martin 2018).

Based on the *false confidence theorem* in Balch et al. (2019), 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, e.g., the belief/plausibility functions in Shafer (1976) or the special case of necessity/possibility functions in Dubois and Prade (1988), Dubois (2006), Destercke and Dubois (2014), and Hose and Hanss (2020).

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 in terms of 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., Dempster 1963; Fisher 1973; Hannig et al. 2016), which generally leads to additive beliefs that are not valid. Thanks to the intuition provided by the auxiliary variable formulation, non-additivity can be introduced by using random sets targeting the unobserved value of U in (5) that corresponds to the observed $Y = y$ and true θ. The following three steps summarize this construction.

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

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

Here and in what follows, we write ϑ for a generic value of the parameter θ.

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). Roughly, this random set is designed to contain those "typical" values of U with certain frequencies; more precise details are provided below.

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

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

The intuition is that $\Theta_y(\mathcal{S})$ contains the unknown θ if and only if $\mathcal{S}$ contains the unobserved value of U. Consequently, if $\mathcal{S}$ can reliably contain the unobserved value of U, then $\Theta_y(\mathcal{S})$ can contain the true θ. This implies $\Theta_y(\mathcal{S})$ carries some information about the location of θ, so it makes sense to define the belief function,

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

as a measure of the degree of belief in the truthfulness of the assertion A about θ. Then its dual, the plausibility function, is given by $\overline{\Pi}_y(A) = 1 - \underline{\Pi}_y(A^c)$. Another important summary is the plausibility contour function

$$\pi_y(\vartheta) = \mathsf{P}_{\mathcal{S}}\{\Theta_y(\mathcal{S}) \ni \vartheta\}, \quad \vartheta \in \Theta. \tag{7}$$

For illustration, consider the binomial data problem mentioned above. For the A-step, n association that links the binomial Y and the success probability θ with an unobservable auxiliary variable $U \sim \mathsf{P}_U = \mathsf{Unif}(0,1)$ is

$$Y = F_\theta^{-1}(U) \iff F_\theta(Y-1) \le U < F_\theta(Y),$$

where F_θ is the $\mathsf{Bin}(n,\theta)$ distribution function. This leads to the set-valued map

$$(y,u) \mapsto \Theta_y(u) = \{\vartheta : F_\vartheta(y-1) \le u < F_\vartheta(y)\}.$$

The P-step proceeds by introducing a random set $\mathcal{S}$ targeting the unobserved value of U. In this case, a reasonable choice is

$$\mathcal{S} = \{u \in [0,1] : |u - 0.5| \le |\tilde{U} - 0.5|\}, \quad \tilde{U} \sim \mathsf{Unif}(0,1). \tag{8}$$

Then the distribution $\mathsf{P}_{\mathcal{S}}$ is completely determined by that of $\tilde{U}$, which is known and easy to compute. For the C-step, we combine the above results to get a new, data-dependent random set

$$\Theta_y(\mathcal{S}) = \bigcup_{u \in \mathcal{S}} \Theta_y(u) = \bigcup_{u \in \mathcal{S}} \{\vartheta : F_\vartheta(y-1) \le u < F_\vartheta(y)\}.$$

If we define $B_{a,b}$ to be the $\mathsf{Beta}(a,b)$ distribution function, and apply the well-known identity $F_\vartheta(y) = B_{n-y,y+1}(1-\vartheta)$, then the above can be simplified:

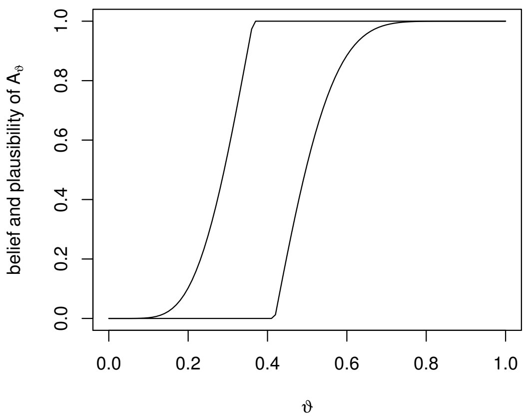
$$\Theta_y(\mathcal{S}) = \bigl[1 - B_{n-y+1,y}^{-1}(\tfrac{1}{2} + |\tilde{U} - \tfrac{1}{2}|), 1 - B_{n-y,y+1}^{-1}(\tfrac{1}{2} - |\tilde{U} - \tfrac{1}{2}|)\bigr].$$

Since the right-hand side is simply a function of $\tilde{U} \sim \mathsf{Unif}(0,1)$, formulas for the belief and plausibility functions are readily available. For example, if $A_\vartheta = [0,\vartheta]$, for $\vartheta \in [0,1]$, is a one-sided assertion about θ, then the belief function is

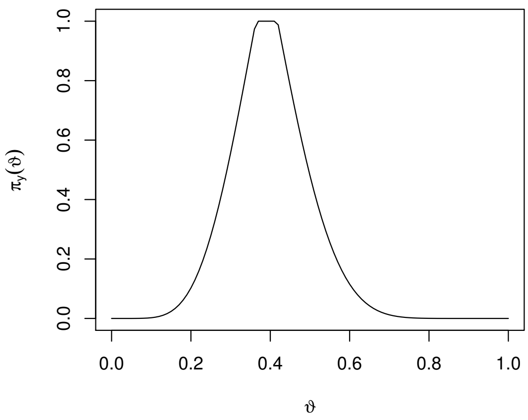
$$\begin{aligned}\underline{\Pi}_y(A_\vartheta) &= \mathsf{P}_{\mathcal{S}}\{\Theta_y(\mathcal{S}) \subseteq A_\vartheta\} \\ &= \mathsf{P}_U\{1 - B_{n-y,y+1}^{-1}(\tfrac{1}{2} - |U - \tfrac{1}{2}|) \le \vartheta\} \\ &= \max\{0, 2B_{y+1,n-y}(\vartheta) - 1\},\end{aligned}$$

where the last inequality uses the fact that $B_{a,b}(1-\vartheta) = B_{b,a}(\vartheta)$. The plausibility function $\overline{\Pi}_y(A_\vartheta)$ at A_ϑ can be found similarly. The plausibility contour function, too, can be written in closed form, as

$$\begin{aligned}\pi_y(\vartheta) &= \mathsf{P}_{\mathcal{S}}\{\Theta_y(\mathcal{S}) \ni \vartheta\} \\ &= 1 - \max\{0, 2B_{y+1,n-y}(\vartheta) - 1\} - \max\{0, 1 - 2B_{n-y+1,y}(\vartheta)\}.\end{aligned}$$



(a) Belief and plausibility functions at A_ϑ



(b) Plausibility contour at ϑ

Figure 1: Inferential model output for the binomial example with $(n, y) = (18, 7)$.

For illustration, suppose that $n = 18$ and $y = 7$. Panels (a) and (b) of Figure 1 show the c-box determined by the belief and plausibility functions at A_ϑ and the plausibility contour function, respectively.

Under 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 to predict 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 $\gamma(u) = \mathsf{P}_\mathcal{S}(\mathcal{S} \ni u)$. Then validity in the sense of (3) is implied by a stochastic dominance property in γ, namely, that is stochastically no smaller than $\mathsf{Unif}(0, 1)$, i.e.,

$$\mathsf{P}_U\{\gamma(U) \leq \alpha\} \leq \alpha, \quad \forall\, \alpha \in [0, 1]. \tag{9}$$

See Definition 1 in Martin and Liu (2013). For example, with the random set $\mathcal{S}$ in (8) targeting a $\mathsf{Unif}(0, 1)$ auxiliary variable U, we have

$$\gamma(u) = \mathsf{P}_\mathcal{S}(\mathcal{S} \ni u) = 1 - |2u - 1|, \quad u \in \mathbb{U} = [0, 1].$$

Then it is easy to see that $U \sim \mathsf{Unif}(0, 1)$ implies $\gamma(U) \sim \mathsf{Unif}(0, 1)$, thus verifying the above sufficient condition for validity as in (3).

Though not strictly necessary for validity, efficiency considerations suggest that $\mathcal{S}$ be nested—like $\mathcal{S}$ in (8)—which makes the belief function consonant. Moreover, when $\mathcal{S}$ is nested, the inferential model's output has the mathematical form of a consonant belief/plausibility function or, equivalently, a necessity/possibility measure (e.g., Liu and Martin 2020). The importance of possibility theory in the broader context of statistical inference was discussed recently in Martin (2021).

2.2 Generalized inferential models

As Martin (2018) argued, the above formulation can be rather rigid; more flexibility can be achieved by working with a so-called *generalized association*, one that does not fully characterize the posited statistical model. As before, suppose we have data $Y \sim \mathsf{P}_{Y|\theta}$, and let $(y, \vartheta) \mapsto R_y(\vartheta)$ be a real-valued function of generic data and parameter values.

Knowing the distribution of $R_Y(\theta)$ does not provide sufficient information to simulate copies of Y from $\mathsf{P}_{Y|\theta}$, but it does provide an avenue for making inference on θ.

As in the previous section, our goal is to forge a relationship, or association, between observable data Y, unknown parameter θ, and an unobservable auxiliary variable with known distribution. For this, we proposed to use the summary $R_Y(\theta)$ and write

$$R_Y(\theta) = H_\theta^{-1}(U), \quad U \sim \mathsf{P}_U = \mathsf{Unif}(0,1), \tag{10}$$

where H_θ is the distribution function of $R_Y(\theta)$,

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

The advantage of this generalized association is, as explained in Martin (2018), 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. What is an appropriate choice of $R_y(\theta)$? The options are virtually unlimited, but 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*

$$R_y(\theta) = L_y(\theta)/L_y(\hat{\theta}), \tag{11}$$

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 $\Theta_y(u) = \{\vartheta : R_y(\vartheta) = H_\vartheta^{-1}(u)\}$ for $u \in [0,1]$.

P-step. For the relative likelihood-based construction in the A-step, since values of $R_y(\vartheta)$ indicate ϑ is a plausible value of the parameter, it makes sense to use a one-sided random set $\mathcal{S}$. Here we work with

$$\mathcal{S} = [\tilde{U}, 1], \quad \tilde{U} \sim \mathsf{P}_U := \mathsf{Unif}(0,1). \tag{12}$$

Two properties of $\mathcal{S}$ deserve mentioning: first, $\gamma(u) = u$ and, hence, $\gamma(U) \sim \mathsf{Unif}(0,1)$, which implies (9) is satisfied; second, $\mathcal{S}$ is nested just like that in Section 2.1.

C-step. Combine the two sets above to get

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

Then the plausibility contour π_y in (7) is

$$\pi_y(\vartheta) = H_\vartheta(R_y(\vartheta)), \quad \vartheta \in \Theta,$$

which determines the full belief and plausibility functions via consonance, e.g.,

$$\overline{\Pi}_y(A) = \sup_{\vartheta \in A} \pi_y(\vartheta), \quad A \subseteq \Theta.$$

It follows from Theorem 1 in Martin (2018) that the generalized inferential model with plausibility function determined by (7) achieves the validity property in (4).

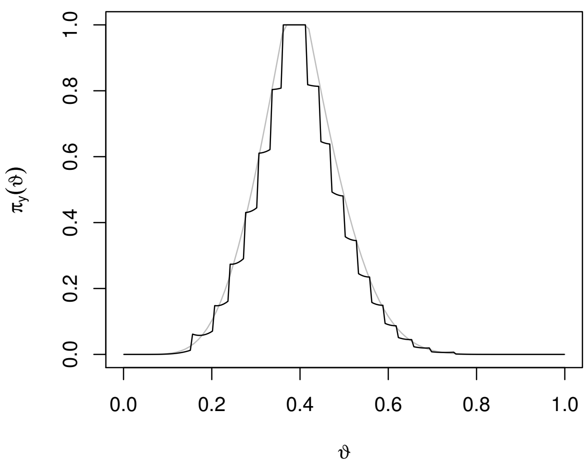


Figure 2: Plot of the plausibility contour $\pi_y(\vartheta)$ based on the inferential model construction here (black) and that in the previous section (gray).

This generalized inferential model can also be applied to the binomial example presented above. In this case, the likelihood function is

$$L_y(\vartheta) \propto \vartheta^y (1-\vartheta)^{n-y}, \quad \vartheta \in [0,1]$$

and, since the maximum likelihood estimator is $\hat{\theta} = y/n$, the relatively likelihood is

$$R_y(\vartheta) = \frac{L_y(\vartheta)}{L_y(\hat{\theta})} = \Bigl(\frac{n\vartheta}{y}\Bigr)^y \Bigl(\frac{n-n\vartheta}{n-y}\Bigr)^{n-y}.$$

Then the generalized inferential model proceeds by finding the distribution function $H_\theta(r)$ of $R_Y(\theta)$. This has no convenient, closed-form expression but the relevant inferential model output can easily and exactly be evaluated numerically as

$$H_\theta(r) = \sum_{y:R_y(\theta)\leq r} f_\theta(y),$$

where f_θ is the $\mathsf{Bin}(n,\theta)$ probability mass function. From this computation, all of the relevant inferential model output can be evaluated. For illustration, we revisit that same example presented in the previous section, where $(n,y) = (18,7)$. Figure 2 plots the plausibility contour of this generalized inferential model, compared that from the basic inferential model presented in Figure 1. The stair-step pattern is standard in discrete-data problems; see, e.g., Martin (2015) and Balch (2020). Note, however, that the generalized inferential model is slightly more efficient that that developed in the previous section, as indicated by its narrower plausibility contour.

Finally, it is often the case that the full parameter of the statistical model is of the form (θ,η), i.e., $Y \sim \mathsf{P}_{Y|\theta,\eta}$, where θ is the quantity of interest and η is a so-called *nuisance parameter*. The censored data application considered here is of this form—with the censoring distribution G being the nuisance parameter—as is the meta-analysis application in Cahoon and Martin (2020). A natural way to proceed with marginal inference on θ, which we describe in more detail in Section 3, is to define a function $R_Y(\theta)$ that does not directly depend on the value of the nuisance parameter η. This does not immediately resolve the η-dependence, however, because the distribution function

$$r \mapsto H_{\theta,\eta}(r) := \mathsf{P}_{Y|\theta,\eta}\{R_Y(\theta) \leq r\} \tag{13}$$

will generally depend on the unknown η. To overcome this dependence on the unknown nuisance parameter, one might try plugging in an estimator $\hat{\eta}$ based on the available data, which amounts to constructing a generalized inferential model based on the approximate distribution function for $R_Y(\theta)$, namely, $H_{\theta,\hat{\eta}}$. Of course, plugging in an estimate affects the exact validity of the generalized inferential model but, at least intuitively, if $\hat{\eta}$ is a reasonably accurate estimate of η, then the corresponding plug-in generalized inferential model ought to be approximately valid. This is precisely the situation encountered in censored-data applications, and Theorem 1 below confirms the above intuition.

3 Generalized inferential models under censoring

3.1 Construction

For random right-censored data, the full likelihood for Y under distribution $\mathsf{P}_{Y|\theta,G}$, where Y_i are independently generated from (1), is given by

$$L_y(\theta, G) = \prod_{i=1}^{n} \bar{G}(t_i)^{d_i} g(t_i)^{1-d_i} \prod_{i=1}^{n} f_\theta(t_i)^{d_i} \bar{F}_\theta(t_i)^{1-d_i}, \quad \theta \in \Theta. \tag{14}$$

Censoring is a special case of *data coarsening* (Heitjan and Rubin 1991; Jacobsen and Keiding 1995), which has attracted considerable attention in both the statistics and imprecise probability community. For the latter, interest in such models is clear since coarsening of the data is one common way that imprecision can be introduced into a statistical model; see, e.g., Denœux (2014) and Couso and Dubois (2018). Here we are considering random censoring which, according to Jacobsen and Keiding (1995, Example 4) corresponds to *coarsening at random*, or CAR, in the terminology of Couso and Dubois (2018).

Since our interest is only in θ and the censoring distribution G does not depend on θ, it is easy to see that the maximizer, $\hat{\theta}$, of $\vartheta \mapsto L_y(\vartheta, G)$ does not depend on G. Therefore, we can adopt a modified relative likelihood

$$R_y(\vartheta) = L_y(\vartheta, G)/L_y(\hat{\theta}, G), \quad \vartheta \in \Theta.$$

With this choice, note that the nuisance parameter G gets canceled out, leaving only a function of data and interest parameter. Of course, the *distribution* of $R_Y(\theta)$, as a function of $Y \sim \mathsf{P}_{Y|\theta,G}$, does depend on G; see below.

Our generalized inferential model for censored data proceeds as outlined in Section 2.2. The modified relative likelihood above is the connection between data, interest parameter θ, and a scalar auxiliary variable U. That is, for the A-step, we have

$$R_Y(\theta) = H_{\theta,G}^{-1}(U), \quad \text{where } U \sim \mathsf{P}_U = \mathsf{Unif}(0,1),$$

with $H_{\theta,G}$ the distribution function of the modified relative likelihood as in (13). Just like in Section 2.2, values of the modified relative likelihood closer to 1 suggest values of the parameter that are more plausible so, for the P-step, we choose random sets in the form of nested intervals as in (12), i.e., $\mathcal{S} = [\tilde{U}, 1]$, where $\tilde{U} \sim \mathsf{Unif}(0,1)$. Finally, the C-step proceeds exactly as in Section 2.2, with all of the relevant output being based on the plausibility contour

$$\pi_y(\vartheta) = H_{\vartheta,G}\big(R_y(\vartheta)\big), \quad \vartheta \in \Theta.$$

If G were known, then it would be relatively simple to evaluate the distribution function $H_{\theta,G}$ and, hence, the plausibility contour in (7). Moreover, validity of the generalized inferential model would follow immediately from the general theory.

Of course, G is never known in applications, so we need to suitably modify the above strategy in order to overcome this challenge. As we indicated above, it makes sense to plug in an estimator of G. Unfortunately, the construction of an estimator and justification of the corresponding plug-in method are non-trivial when G is an infinite-dimensional object. The next two subsections address these two challenges in turn.

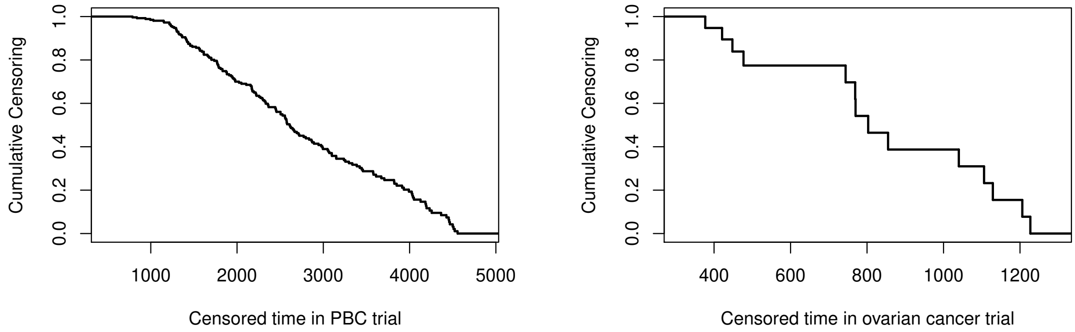
3.2 Implementation

Putting the above inferential model construction into practice requires that the distribution function of the relative likelihood be evaluated, at least approximately, for every θ. This is straightforward to do 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:

$$H_{\theta,G}(r) \approx \frac{1}{M}\sum_{m=1}^{M} 1\{R_{Y^{(m)}}(\theta) \leq r\}, \tag{15}$$

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, $H_{\theta,G}$ 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 $\widehat{G}$.

The Kaplan–Meier estimator (Kaplan and Meier 1958), summarized in Chapter 4 of Klein and Moeschberger (2003), is a powerful tool in the basic censored data analysis toolbox. It is designed to estimate the survival/distribution function of the event times when censoring is present, and its implementation is included in all statistical software programs, e.g., the `survival` package in R (Therneau 2014). Our goal here is slightly different, however; we want to estimate the censoring distribution function G rather than the event time distribution. Fortunately, there is a relatively straightforward way to overcome this. Our proposal is simply to reverse the event/censored classifications. That is, recall that the observations are $T_i = \min(X_i, C_i)$. But now we treat C_i as the "event time" and X_i as the "censoring time," which amounts to simply defining a new censoring indicator $d_i' = 1 - d_i$. From here, we can immediately construct the Kaplan–Meier estimator $\widehat{G}$ based on this modified interpretation of the data. An R code implementation that returns the plausibility contour and corresponding plausibility



(a) Data from Section 4.1

(b) Data from Section 4.2

Figure 3: Plots of the Kaplan–Meier estimator of the survival function, $1 - G$.

regions based on this generalized inferential model formulation, with the Kaplan–Meier plug-in estimator, is available from the second author's website.

Many different presentations of the Kaplan–Meier estimator can be given, but the one that we find the easiest to understand is through the lens of maximum likelihood. Just as the likelihood function $L_y(\theta, G)$ in (14) factored nicely for dealing with θ independent of G, the same factorization can be used here to deal with G. That is, to maximize this function over G, independent of θ, we can ignore the second term in (14) and focus exclusively on the first, which only depends on G. Being a function of an infinite-dimensional G complicates the details, but it turns out that the Kaplan–Meier is the (nonparametric) maximum likelihood estimator of G under this event-to-censoring time renaming device. There is a closed-form expression for the Kaplan–Meier estimator (see Klein and Moeschberger 2003, Chap. 4), but the formula provides no additional insights so we opt not to give it here. Instead, we briefly mention two relevant properties. First, being a maximum likelihood estimator, the Kaplan–Meier estimator inherits many of the nice properties, including consistency and statistical efficiency, properties that we make use of in our theoretical analysis presented in Appendix A. Second, like many other nonparametric maximum likelihood estimators, $\widehat{G}$ is a discrete distribution, so it has a step-function form. For illustration, Panels (a) and (b) of Figure 3 plot the Kaplan–Meier estimator of the censoring survival function $1 - G$ for the two real-data examples in Sections 4.1 and 4.2, respectively.

While this Kaplan–Meier plug-in estimator strategy is conceptually straightforward, there is one technical point worth making about the estimation process. In typical applications of the Kaplan–Meier estimator, where the goal is nonparametric estimation of the event time distribution, if the largest observation corresponds to a censored outcome, the the Kaplan–Meier estimator of the survival function at time t does not vanish as $t \to \infty$. This amounts to the estimated event time distribution putting some mass at ∞. In our context, since we interpret the original event times as censoring times, our estimate $\widehat{G}$ will put positive mass at ∞ when the largest observation is an event, under which C_i^*'s drawn from $\widehat{G}$ will equal ∞ with positive probability and, consequently, T_i^*'s drawn will correspond to an event time as $X_i^* < C_i^*$. The question is whether this simulation process produces Y^* data sets that reflect the censoring pattern seen in the original data. Our numerical simulations presented below suggested that it does.

3.3 Validity properties

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 plausibility contours are stochastically no larger than uniform when $Y \sim \mathsf{P}_{Y|\theta,G}$, it follows that

$$\sup_{\theta \in A} \mathsf{P}_{Y|\theta,G}\{\overline{\Pi}_Y(A) \leq \alpha\} \leq \alpha, \quad \forall\, \alpha \in [0,1], \quad \forall\, A \subseteq \Theta. \tag{16}$$

A desirable consequence of validity is that confidence regions having the nominal frequentist coverage probability can be constructed immediately based on the plausibility function output. Indeed, if π_y is the plausibility contour function corresponding to $\overline{\Pi}_y$, then the set

$$\{\vartheta \in \Theta : \pi_y(\vartheta) > \alpha\} \tag{17}$$

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 $\pi_Y(\theta) > \alpha$, which equals $1-\alpha$.

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

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

at least approximately? Here $\pi_y(\theta;\widehat{G})$ denotes the plausibility function obtained by applying the above algorithm with $\widehat{G}$ plugged in for the unknown G, i.e., simulating C_i^*'s iid from $\widehat{G}$. The dependence of $\pi_y(\theta;\widehat{G})$ on the Kaplan–Meier estimator, an infinite-dimensional quantity, is quite complicated, but at the very least, under mild assumptions, our proposed generalized inferential model should be valid for large n. The following theorem confirms this. Since we are considering asymptotic properties as $n \to \infty$, we embellish on our previous notation to emphasize the dependence on n.

Theorem 1. *Let $Y^n = (Y_1,\ldots,Y_n)$ be a sample obtained under random censoring, $\mathsf{P}_{Y^n|\theta,G}$, as described in (1), where both θ and G are unknown. The proposed plausibility function for inference on θ, defined by*

$$\pi_{Y^n}(\theta;\widehat{G}_n) = H^n_{\theta,\widehat{G}_n}(R_{Y^n}(\theta)),$$

with $\widehat{G}_n$ the Kaplan–Meier estimator of G described above based on Y^n, satisfies

$$\pi_{Y^n}(\theta;\widehat{G}_n) \to \mathsf{Unif}(0,1) \quad \text{in distribution as } n \to \infty.$$

Therefore, the proposed generalized inferential model is approximately valid for large n.

Proof. See Appendix A. □

Theorem 1 establishes approximate validity, but the theoretical support for our approach might actually be stronger than the theorem suggests. The proof amounts to comparing the distribution of $R_{Y^n}(\theta)$ under two different distributions for Y^n, one based on (θ, G) and the other based on $(\theta, \widehat{G}_n)$. The relative likelihood is well-known to be an approximate pivot, that is, when n is large, the distribution $R_{Y^n}(\theta)$, under $\mathsf{P}_{Y^n|\theta,G}$, does not depend on θ or G. Since $\vartheta \mapsto R_{y^n}(\vartheta)$ itself does not depend on G, the distribution of $R_{Y^n}(\theta)$ is relatively insensitive to small changes in G. That is, if G' and G are two censoring distributions that are "reasonably close," then the distribution of $R_{Y^n}(\theta)$ is roughly the same under $\mathsf{P}_{Y^n|\theta,G}$ and $\mathsf{P}_{Y^n|\theta,G'}$. So all we need is for $\widehat{G}_n$ to be in a neighborhood of G to conclude that $\pi_{Y^n}(\theta; \widehat{G}_n)$ is approximately uniform. Fortunately, we know that $\widehat{G}_n$ gets close to G very quickly as $n \to \infty$, so we expect that $\pi_{Y^n}(\theta; \widehat{G}_n) \to \mathsf{Unif}(0, 1)$ very quickly too. Details behind these heuristics are made precise in the proof in Appendix A, but it remains to determine if there really some kind of higher-order accuracy here.

To see this unexpected accuracy in action, we take 10,000 small samples of size $n = 15$ 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 $\alpha \mapsto \mathsf{P}_{Y|\theta,G}\{\pi_Y(\theta; \widehat{G}) \le \alpha\}$ shown in Figure 4 is approximately uniform, hence approximate validity. Moreover, by starting with the relative likelihood $R_y(\theta)$ in (11), we removed almost all dependence on the nuisance parameter G; that is, the exact distribution of our relative likelihood is roughly constant in G and thus the plug-in estimator we used to get $\widehat{G}$ apparently does not need to be especially accurate. As a result, the plausibility output using our plug-in method as described in Section 3.2 is close to the exact distribution. Simulated- and real-data examples in Section 4 further demonstrate the proposed method's strong performance compared to others.

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 results suggest, plausibility functions consistently outperform 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 $y = \{(t_i, d_i) : i = 1, \ldots, n\}$ as above, the maximum likelihood estimate is $\hat{\theta} = \sum_{i=1}^n d_i / \sum_{i=1}^n t_i$. From its classical asymptotic normality, a 95% confidence interval for θ is easily obtained as $\hat{\theta} \pm 1.96 I(\hat{\theta})^{-1/2}$, where $I(\hat{\theta})$ is the observed Fisher information, i.e., the negative second derivative of the log-

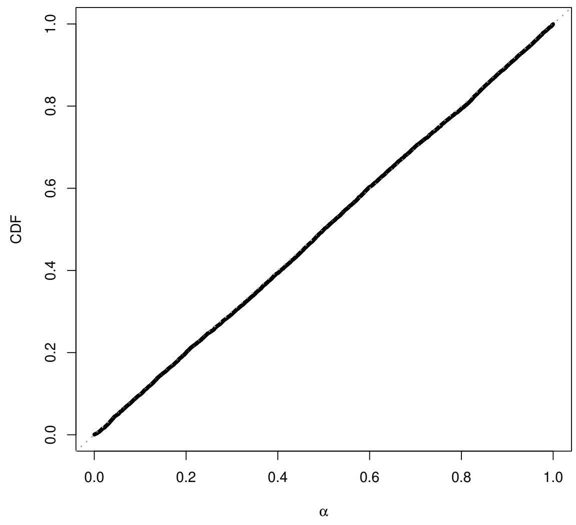


Figure 4: Distribution of $\alpha \mapsto \mathsf{P}_{Y|\theta,G}\{p_Y(\theta;\widehat{G}) \leq \alpha\}$ (solid) compared with that of $\mathsf{Unif}(0,1)$ (dotted) based on Monte Carlo samples from a standard exponential distribution subject to random right censoring. The average censoring level among all 10,000 replications at this setting is 19.9%.

likelihood evaluated at $\hat{\theta}$. 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. Experiments with various values of (α_0, β_0) revealed that $\alpha_0 = 2$ and $\beta_0 = 1$ had the best overall performance across our settings with respect to coverage probability of the credible intervals.

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

$$\frac{\theta^{\sum_i D_i} e^{-\theta \sum_i T_i}}{\hat{\theta}^{\sum_i D_i} e^{-\hat{\theta} \sum_i T_i}} = H_{\theta,G}^{-1}(U), \quad U \sim \mathsf{P}_U = \mathsf{Unif}(0,1). \tag{18}$$

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

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

Next, the P-step requires introducing a predictive random set $\mathcal{S}$ in (12) for U. We then combine our A- and P-steps

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

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

$$\pi_y(\vartheta) = H_{\vartheta,G}(R_y(\vartheta)), \quad \theta > 0.$$

A $100(1-\alpha)\%$ confidence interval can be obtained as the upper level set of the plausibility contour as in (17). Evaluating this plausibility contour requires the Monte Carlo procedure discussed in Section 3.

For comparison, we simulate 10,000 replications of lifetimes arising from nine different θ values—between 0.5 and 15—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 5 demonstrate that the nominal $100(1-\alpha)\%$ coverage is achieved by our proposed method. Note that this problem is particularly challenging in the $n = 15$ and large θ case, since large θ implies more censoring. The maximum likelihood and Bayes approaches appear to be substantially affected by this extreme censoring, while our approach is not.

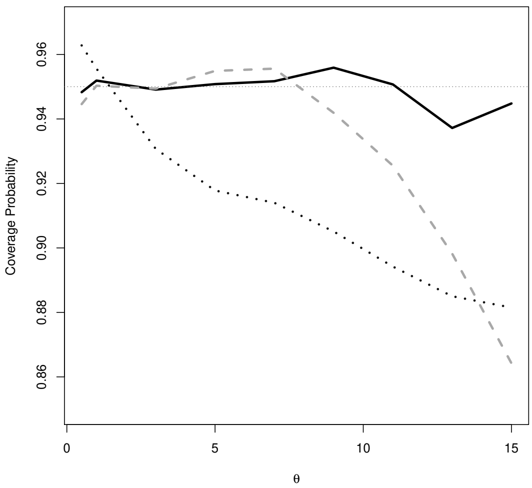
For a real-data illustration, we consider the primary biliary cirrhosis (PBC) data from a clinical trial at the Mayo Clinic from 1974 to 1984. The data consists of $n = 312$ recorded survival times for patients involved in the randomized trial, along with a corresponding right censoring indicator; there are 168 censored cases, more than 50% of total observations. Figure 6 shows the point plausibility function $\pi_y(\theta)$ for a range of parameter values, along with the corresponding 95% plausibility interval (17). For comparison, 95% confidence intervals based on asymptotic normality of the maximum likelihood estimate are also displayed. The intervals derived from the plausibility function are almost indistinguishable from the likelihood-based intervals, which is a sign of our proposed approach's efficiency, since the latter are asymptotically "best."

4.2 Weibull

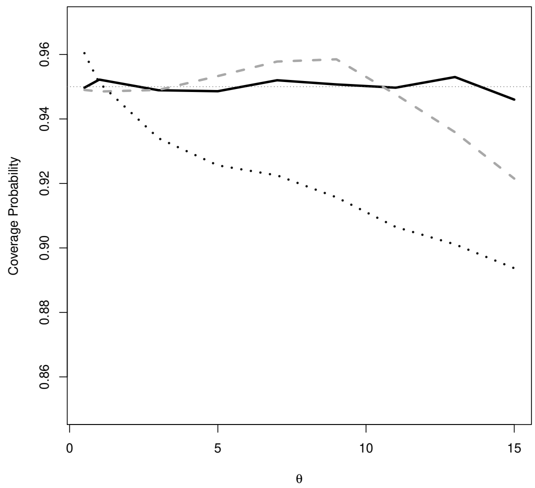
One of the most widely used time-to-event distributions is the Weibull, with applications in manufacturing, health, etc., as it has sufficient flexibility to capture changes in the hazard rate (Lawless 2011). Exponential is a special case of the Weibull when the shape parameter $\beta = 1$. The density and survival functions, indexed by $\theta = (\beta, \lambda)$, are

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

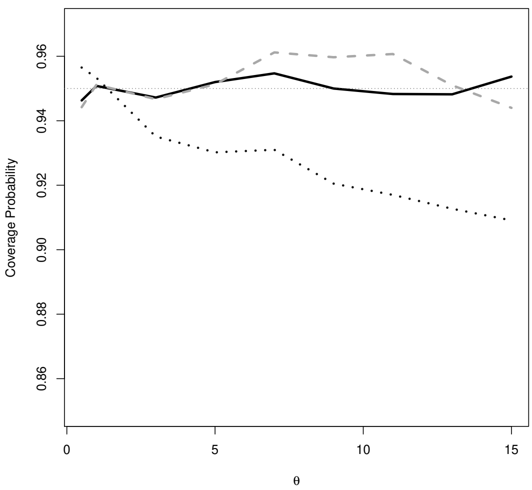
Similar to the setup as described for the exponential example, we compare the performance of our proposed approach against that of a more traditional frequentist or objective Bayesian approach. An inferential model requires that we simulate the distribution of $R_Y(\theta)$; so for a finite grid of $\theta = (\beta, \lambda)$ 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 $C^* \sim \widehat{G}$, the modified Kaplan–Meier estimate. We implement this procedure for 10,000 replications of lifetimes arising from six different settings, with β ranging between 0.5 and 3.0, of the Weibull distribution. These 60,000 replications were each subject to random right censoring from $G \sim \mathsf{Unif}(0, 4)$. To compare against a Bayes approach, multiple non-informative and weakly informative priors were used, from which the $\mathsf{Gamma}(0.1, 1)$ prior on the shape and $\mathsf{N}(0, 10)$ prior on the log transformed scale were selected, as they resulted in credible regions with the highest coverage. Surprisingly, as shown in Figure 7, despite the careful specification of these priors, the generalized inferential model still remains the only method that achieves nominal coverage across these censored data settings. The joint confidence sets from maximum likelihood under-cover, while the joint



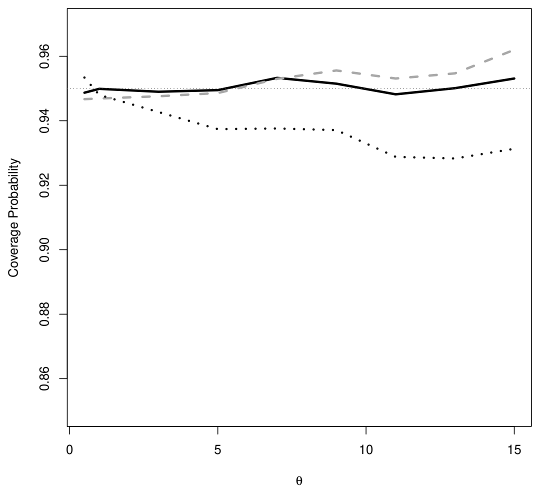
(a) $n = 15$



(b) $n = 20$



(c) $n = 25$



(d) $n = 50$

Figure 5: Coverage probability of the 95% plausibility region for θ in the exponential model (black). Results compared to those based on maximum likelihood (dashed) and Bayesian with a $\mathsf{Gamma}(2, 1)$ prior (dotted).

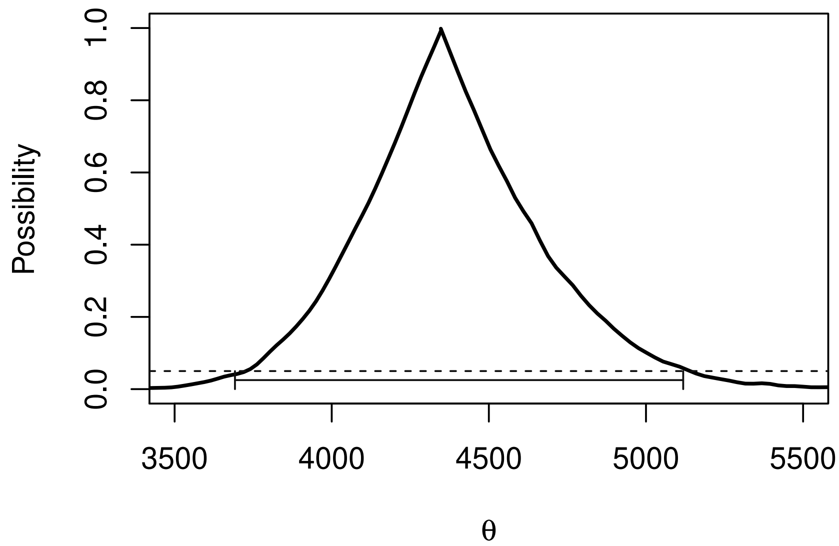


Figure 6: Point plausibility function for the mean in the PBC example under an exponential model (black). Reference line at $\alpha = 0.05$ (dotted) and approximate 95% confidence intervals based on maximum likelihood (grey).

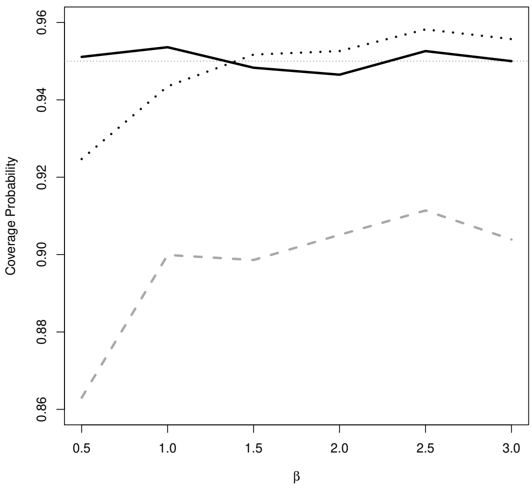
credible sets from our Bayes approach achieves nominal coverage only past a specific β value as, under our simulation settings, the average censoring levels tend be much larger with small values of β. Further investigations into interval lengths (not shown) also demonstrate our plausibility intervals are on average shorter than the Bayesian intervals.

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 sample size and high censoring level, our plausibility contours capture the non-elliptical shape as shown by the Bayesian posterior in Figure 8.

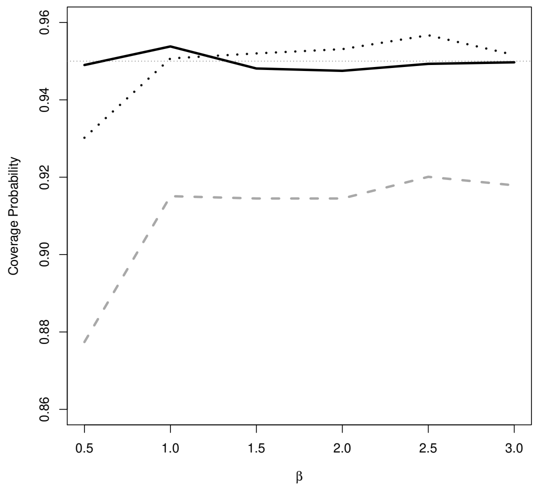
4.3 Log-normal

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

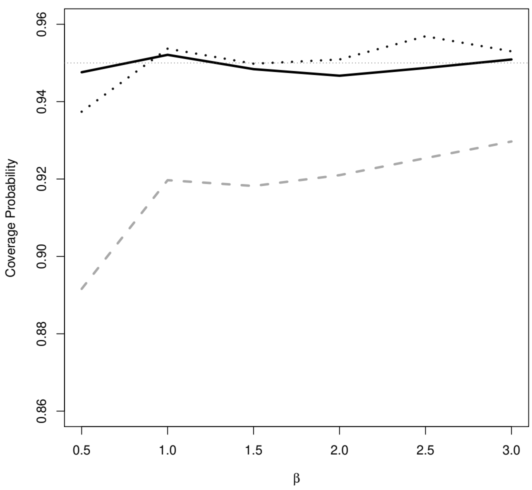
$$f_\theta(t) = \frac{1}{(2\pi)^{1/2}\sigma t} \exp\Big\{-\frac{1}{2}\Big(\frac{\log t - \mu}{\sigma}\Big)^2\Big\}.$$



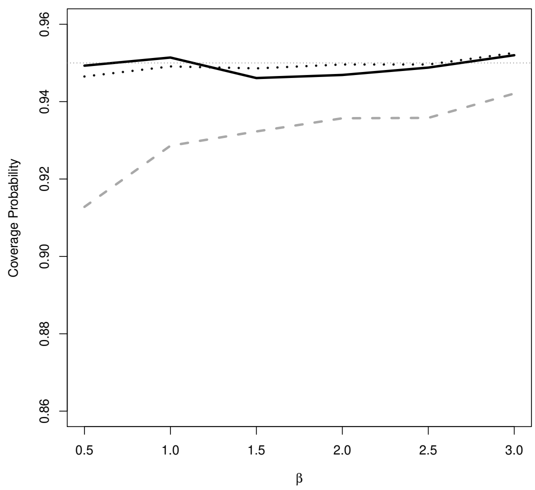
(a) $n = 15$



(b) $n = 20$



(c) $n = 25$



(d) $n = 50$

Figure 7: Coverage probability of the 95% plausibility region for $\theta = (\beta, \lambda)$ in the Weibull model (black). Results compared to maximum likelihood (dashed) 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 (dotted).

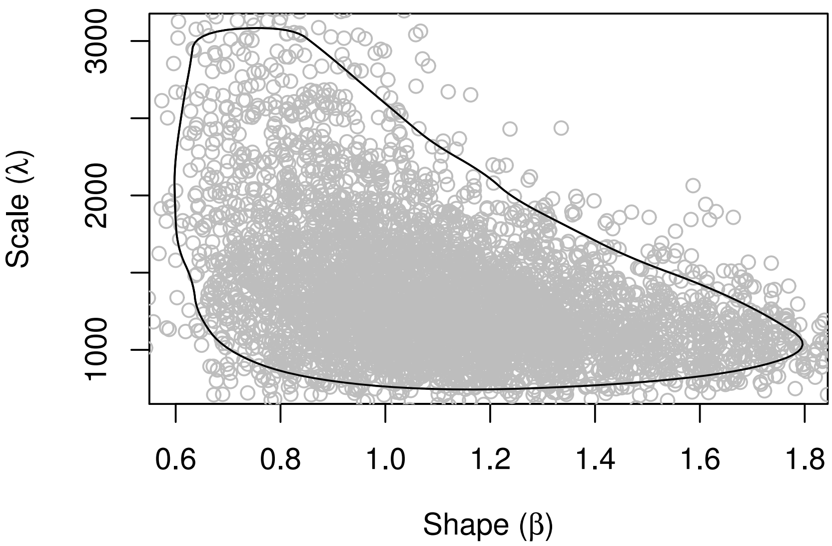
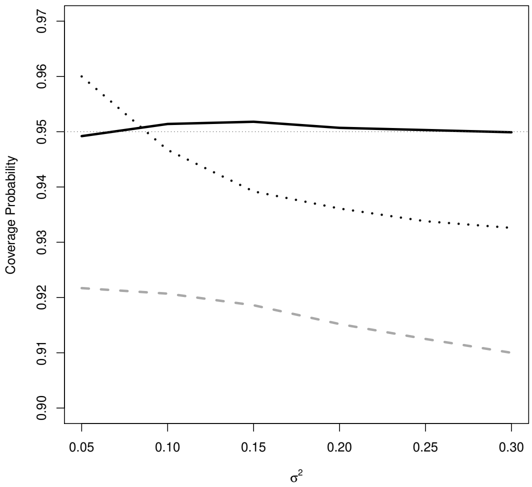


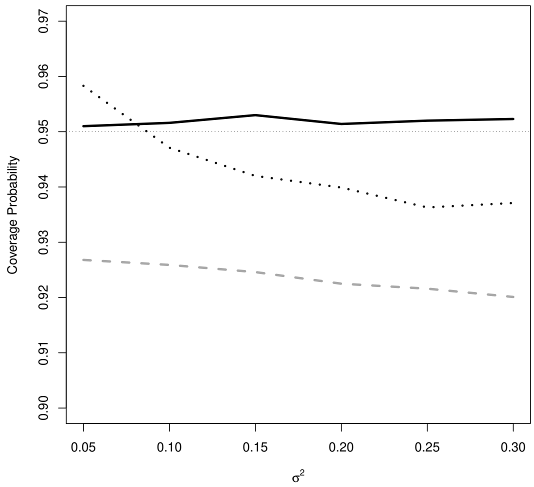
Figure 8: Plausibility contour (black) for $\theta = (\beta, \lambda)$, the shape and scale parameter pair, in the ovarian cancer data under a Weibull model subject to Type I right censoring. Bayesian posterior samples based on a $\mathsf{Gamma}(1, 0.1)$ prior for the shape and $\mathsf{N}(0, 10)$ prior for the log transformed scale parameter (gray).

Similar to our examples above, we compare the coverage performance of our plausibility 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 6 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 $\widehat{G}$ 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. A relevant quantity of interest in log-normal model applications is the mean, $\psi = \exp(\mu + \sigma^2/2)$, a non-linear function of (μ, σ). Figure 9 shows that, under various censoring levels, our proposed method gives marginal plausibility intervals for ψ that achieve the nominal $100(1-\alpha)\%$ coverage while, again, the other methods drastically under-cover.

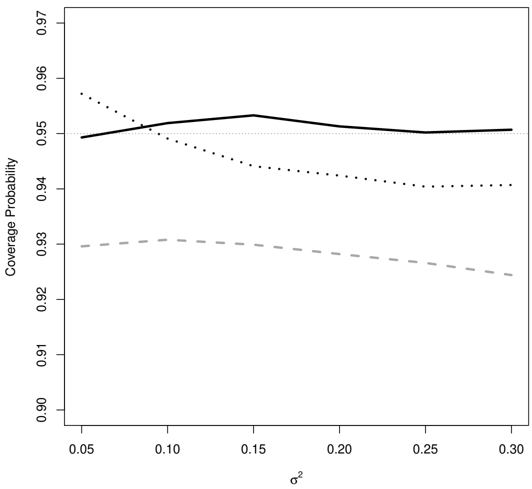
We use Atrazine concentration data collected from a well in Nebraska as an example. This set of 24 observations were randomly subject to two lower detection limits of 0.01 and 0.05 μg/l of which 11 observations were censored. Despite this censoring level of 45.8%, previous studies indicate the log-normality assumption holds (Helsel 2005). We apply our Monte Carlo approach to determine the joint plausibility contours for $\theta = (\mu, \sigma^2)$ in Figure 10, along with the marginal plausibility function for the log-normal mean, ψ, in Figure 11. The point at which we assign the highest plausibility aligns with the maximum



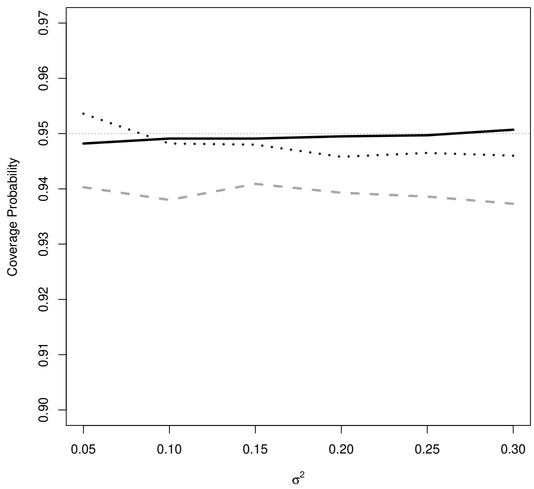
(a) $n = 15$



(b) $n = 20$



(c) $n = 25$



(d) $n = 50$

Figure 9: Coverage probability of the 95% plausibility interval for ψ in the log-normal model (black). Results are compared to maximum likelihood (dashed) 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 (dotted).

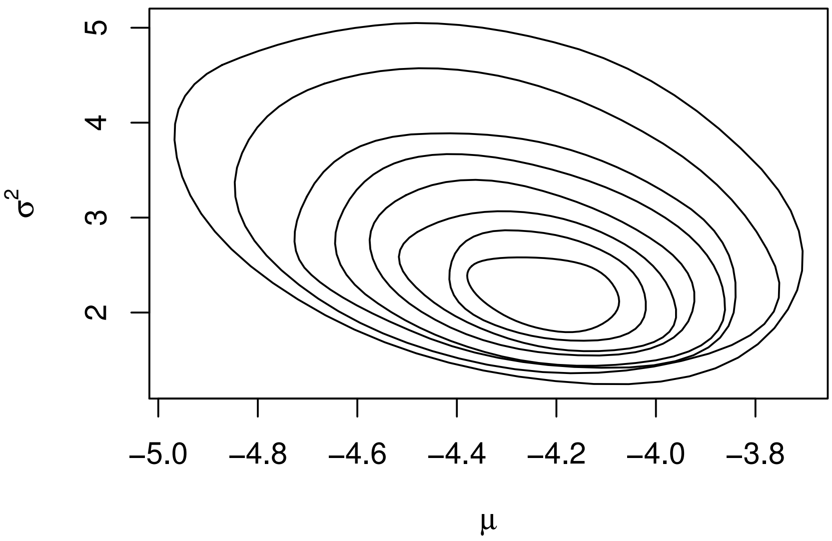


Figure 10: Plausibility contours $\alpha = \{0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90\}$ for the Atrazine example under a log-normal model with Type I left censoring.

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 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. 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. Approximate validity is established in Theorem 1, but we argued that the validity result might actually be stronger than the theorem suggests. We demonstrate numerically that the proposed inferential model approach outperforms traditional maximum likelihood and Bayesian solutions in terms of coverage probability.

Aside from efforts to establish the validity property more rigorously for small n, it is of interest to explore complicated and practical types of censored-data models. First, there are interesting problems where censoring depends on covariates, so that an assumption of random censoring might not be justified. 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

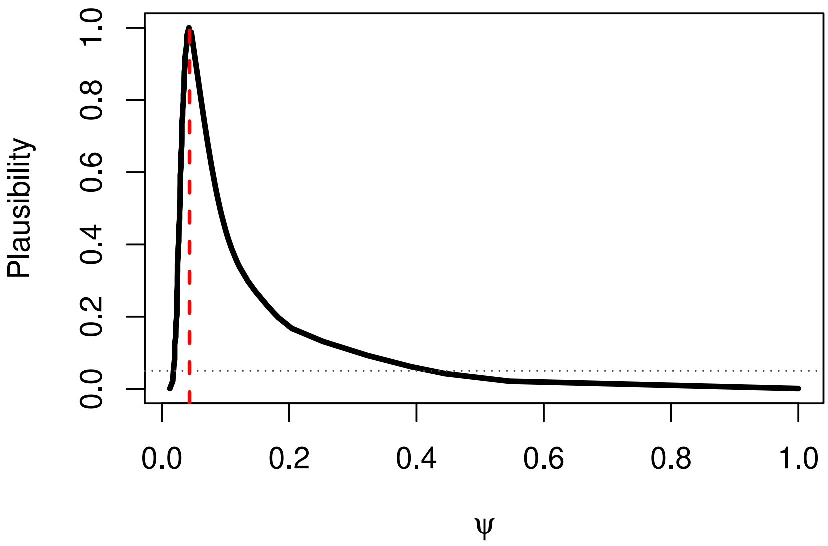


Figure 11: Marginal plausibility function for the mean ψ in the Atrazine example. Reference lines at $\alpha = 0.05$ and at the maximum likelihood estimate (dashed).

evaluate the distribution function $H_{\theta,G}$ would be much more involved. Ongoing efforts are focused on this and other general improvements to the simple Monte Carlo computations described here. Second, so-called current status examples are those where the event times are unobservable, the only data is an examination time and a status indicator which simply indicates if the event has happened by the examination time or not. This data is corrupted even more so than with censored data, but it would still be possible to extend the methodology developed here. Third, we could also consider interval-censored data, where the event time is censored unless it falls within a certain (random) interval. Here, like with the addition of covariates, extension of the proposed methodology is at conceptually straightforward, but computationally more challenging. More generally, there is an entire class of examples involving corrupted or coarsened data in Couso and Dubois (2018) and it would be interesting to explore the application of the inferential model machinery to those problems.

Acknowledgments

The authors thank the anonymous reviewers for their helpful feedback on a earlier version of this manuscript. This work is partially supported by the U.S. National Science Foundation, DMS–1811802.

A Proof of Theorem 1

Start by writing $\pi_{Y^n}(\theta; \widehat{G}_n) = \pi_{Y^n}(\theta; G) + \Delta_n$, where

$$\Delta_n = H^n_{\theta,\widehat{G}_n}(R_{Y^n}(\theta)) - H^n_{\theta,G}(R_{Y^n}(\theta)),$$

with $\widehat{G}_n$ the Kaplan–Meier estimate and G the true censoring distribution. The key insight is that $\pi_{Y^n}(\theta; G)$ is exactly uniformly distributed under $\mathsf{P}_{Y^n|\theta,G}$, so if we can show that $\Delta_n \to 0$ in probability, the claim will follow from Slutsky's theorem.

A first observation is that

$$|\Delta_n| \leq \sup_{r\in[0,1]} \bigl|H^n_{\theta,\widehat{G}_n}(r) - H^n_{\theta,G}(r)\bigr|,$$

so we can prove the claim by showing that the above difference vanishes uniformly. But since these are distribution functions, it is enough to show that the difference vanishes pointwise, at each fixed r. To prove pointwise convergence, we refer to Banerjee (2005) who shows that the usual large sample properties for the relative likelihood $R_{Y^n}(\theta)$ hold under $\mathsf{P}_{Y^n|\theta,G}$ and under $\mathsf{P}_{Y^n|\theta,G_n}$, as long as G and G_n are "close." In particular, he shows that, for any G_n that satisfies $G_n = G + n^{-1/2} Z_n$ for Z_n bounded in probability, the two distributions $\mathsf{P}_{Y^n|\theta,G}$ and $\mathsf{P}_{Y^n|\theta,G_n}$ are mutually contiguous and, therefore,

$$-2 \log R_{Y^n}(\theta) \to \mathsf{ChiSq}(\dim(\theta)) \quad \text{in distribution as } n \to \infty, \tag{19}$$

under both $\mathsf{P}_{Y^n|\theta,G}$ and $\mathsf{P}_{Y^n|\theta,G_n}$; see also, Murphy and van der Vaart (1997, 2000). Theorem 5 in Breslow and Crowley (1974) establishes that the Kaplan–Meier estimator satisfies

$$n^{1/2}\|\widehat{G}_n - G\| = O(1) \quad \text{in probability as } n \to \infty,$$

where $\|G - G'\| = \sup_{t\leq\tau} |G(t) - G'(t)|$ and τ is any value such that $\{1 - F_\theta(\tau)\}\{1 - G(\tau)\} > 0$. Therefore, we have

$$H^n_{\theta,G}(r) \to H^\infty(r) \quad \text{and} \quad H^n_{\theta,\widehat{G}_n}(r) \to H^\infty(r), \quad n \to \infty \tag{20}$$

where H^∞ is the limiting distribution function of $R_{Y^n}(\theta)$ from (19). If we write

$$\bigl|H^n_{\theta,\widehat{G}_n}(r) - H^n_{\theta,G}(r)\bigr| \leq \bigl|H^n_{\theta,\widehat{G}_n}(r) - H^\infty(r)\bigr| + \bigl|H^n_{\theta,G}(r) - H^\infty(r)\bigr|,$$

then we immediately see that the right-hand converges to 0 in $\mathsf{P}_{Y^n|\theta,G}$-probability as $n \to \infty$. This, in turn, implies the same for Δ_n and, applying Slutsky's theorem as discussed above, we can conclude that $p_{Y^n}(\theta; \widehat{G}_n) \to \mathsf{Unif}(0,1)$ in distribution.

References

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

Balch, M. S. (2020). New two-sided confidence intervals for binomial inference derived using Walley's imprecise posterior likelihood as a test statistic. *International Journal of Approximate Reasoning*, 123:77–98.

Balch, M. S., Martin, R., and Ferson, S. (2019). Satellite conjunction analysis and the false confidence theorem. *Proceedings of the Royal Society A*, 475(2227):20180565.

Banerjee, M. (2005). Likelihood ratio tests under local alternatives in regular semiparametric models. *Statistica Sinica*, 15(3):635.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics*, 2(3):437–453.

Cahoon, J. and Martin, R. (2020). Generalized inferential models for meta-analyses based on few studies. *Statistics and Applications*, 18(2):299–316.

Couso, I. and Dubois, D. (2018). A general framework for maximizing likelihood under incomplete data. *International Journal of Approximate Reasoning*, 93:238–260.

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

Denœux, T. (2014). Likelihood-based belief function: Justification and some extensions to low-quality data. *International Journal of Approximate Reasoning*, 55(7):1535 – 1547.

Denœux, T. and Li, S. (2018). Frequency-calibrated belief functions: review and new insights. *International Journal of Approximate Reasoning*, 92:232–254.

Destercke, S. and Dubois, D. (2014). Special cases. In *Introduction to Imprecise Probabilities*, Wiley Ser. Probab. Stat., pages 79–92. Wiley, Chichester.

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

Dubois, D. and Prade, H. (1988). *Possibility Theory.* Plenum Press, New York.

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

Fisher, R. A. (1973). *Statistical Methods and Scientific Inference.* Hafner Press.

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

Heitjan, D. F. and Rubin, D. B. (1991). Ignorability and coarse data. *The Annals of Statistics*, 19(4):2244–2253.

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

Hose, D. and Hanss, M. (2020). On data-based estimation of possibility distributions. *Fuzzy Sets and Systems*, 399:77–94.

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

Jacobsen, M. and Keiding, N. (1995). Coarsening at random in general sample spaces and random censoring in continuous time. *The Annals of Statistics*, 23(3):774–786.

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

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

Krishnamoorthy, K. and Xu, Z. (2011). 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.

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

Liu, C. and Martin, R. (2020). Inferential models and possibility measures. `arXiv:2008.06874`.

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

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

Martin, R. (2019). False confidence, non-additive beliefs, and valid statistical inference. *International Journal of Approximate Reasoning*, 113:39–73.

Martin, R. (2021). An imprecise-probabilistic characterization of frequentist statistical inference. `https://researchers.one/articles/21.01.00002`.

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

Martin, R. and Liu, C. (2015). *Inferential Models: Reasoning with Uncertainty*. CRC Press.

Murphy, S. A. and van der Vaart, A. W. (1997). Semiparametric likelihood ratio inference. *The Annals of Statistics*, 25(4):1471–1509.

Murphy, S. A. and van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95(450):449–465.

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

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

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

Therneau, T. M. (2014). A package for survival analysis in S. `https://cran.r-project.org/package=survival`. R package version 2.37-7.

Walley, P. (2002). Reconciling frequentist properties with the likelihood principle. *Journal of Statistical Planning and Inference*, 105(1):35–65.

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