Reliable survival analysis based on the Dirichlet Process

Francesca Mangili[*,1]**, Alessio Benavoli** [1]**, Cassio P. de Campos** [1]**,** and **Marco Zaffalon** [1]

[1] IPG-IDSIA, Galleria 2, 6928 Manno-Lugano, Switzerland

We present a robust Dirichlet process for estimating survival functions from samples with right-censored data. It adopts a prior near-ignorance approach to avoid almost any assumption about the distribution of the population lifetimes, as well as the need of eliciting an infinite dimensional parameter (in case of lack of prior information), as it happens with the usual Dirichlet process prior. We show how such model can be used to derive robust inferences from right-censored lifetime data. Robustness is due to the identification of the decisions that are prior-dependent, and can be interpreted as an analysis of sensitivity with respect to the hypothetical inclusion of fictitious new samples in the data. In particular, we derive a nonparametric estimator of the survival probability and a hypothesis test about the probability that the lifetime of an individual from one population is shorter than the lifetime of an individual from another. We evaluate these ideas on simulated data and on the Australian AIDS survival dataset. The methods are publicly available through an easy-to-use `R package`.

Key words: Prior near-ignorance; Dirichlet Process; Survival analysis; Censored data; Bayesian nonparametrics; Log-rank test.

1 Introduction

Studies are conducted daily to compare the survival time of individuals of different medical, demographic, environmental, and behavioral characteristics. For instance, consider the Australian AIDS dataset (Ripley and Solomon, 1994; Venables and Ripley, 2002), where analyses suggested that a difference in survival time existed when discriminating individuals with AIDS by the use (or not) of drugs (for whom a different survival was arguably expected), but also suggested that individuals with AIDS from the Queensland region in Australia have significantly worse survival time than those from the New South Wales region. Even if the latter conclusion has been questioned in those studies because survival difference associated with Australian regions was at first not expected, no formal analysis was used to assess the reliability of the result. This is a common situation in practice, partially related to the lack of methods with such purpose. Reliability of conclusions often come at a later stage by comparing the obtained results with those of other studies. The overall goal of this work is to provide interpretable and easy-to-use reliable methods for survival estimation and for testing differences in survival time. We address this problem from a Bayesian nonparametric perspective. In particular, we consider a Bayesian nonparametric approach based on a robust Dirichlet prior process.

Bayesian nonparametric procedures have appeared after Ferguson (1973), who introduced the Dirichlet process (DP). DP priors are behind the most popular Bayesian nonparametric models, and a number of authors have proposed nonparametric approaches based on them to estimate survival functions with censored data (Susarla and Van Ryzin, 1976; Blum and Susarla, 1977; Zhou, 2004). In particular, Susarla and Van Ryzin (1976) developed an estimator for survival functions that converges to the Kaplan-Meier estimator as the prior strength of the DP goes to zero. An extension to the so-called "neutral to the right"

---

*Corresponding author: e-mail: francesca@idsia.ch

priors was considered by Ferguson and Phadia (1979). Other authors focus on the hazard function; for instance, Hjort (1990) uses Beta processes to develop a Bayesian estimator for the cumulative hazard.

A DP is completely characterized by its prior strength (or precision), which is a positive scalar number, and a prior base probability measure, which is an infinite-dimensional "parameter". Without detailed prior information, the subjective choice of the infinite-dimensional prior base measure may be difficult and may considerably affect inferences, leading to non-robust decisions. Generalizing earlier ideas developed in Bayesian parametric robustness (Berger, 1994; Berger et al., 2000) and near-ignorance models (Berger, 1994; Pericchi and Walley, 1991; Walley, 1991), our first contribution is to adopt a viewpoint that models incomplete prior information by considering, instead of a single prior probability measure, the family of all probability measures that are compatible with the available information (Benavoli et al., 2014b; Benavoli et al., 2014a). In this view, one models the lack of prior information by considering the set of all DP priors with fixed prior strength, and with normalized base measure free to vary within the set of all probability measures. The only required parameter is the prior strength (a positive scalar), and thus we refer to this prior as near-ignorance DP (or IDP) (Benavoli et al., 2014b; Benavoli et al., 2014a). On the one hand, this viewpoint simplifies prior elicitation; on the other hand, it provides an extra level of robustness, because for all inferences it is able to estimate bounds for the inferences that are independent of the prior base probability measure.

Based on the IDP, we propose a simple and robust estimator for the survival probability under a scenario with right-censored data. This can be seen as a generalization to continuous spaces of the work of Coolen (1997), who takes a different approach for survival analysis and considers a finite partition of the continuous timeline (which depends on the observed events) in order to apply the (discrete and parametric) Imprecise Dirichlet Model (Walley, 1996). Our IDP approach naturally reduces to that model when only a finite partition of the space is considered, yet it allows us to model survival data in continuous-time domains in a natural way. The probabilistic interpretation of posterior distributions provided by the IDP approach gives a straightforward way to quantify the uncertainty of the predictions. When it comes to estimating the prediction performance, we show on artificial data that despite the actual curves differ little from those estimated with Kaplan-Meier, the credible intervals of IDP are more reliable than Kaplan-Meier's linear (and log-transformed) confidence intervals obtained from Greenwood's formula.

Our second contribution is to use the IDP survival function estimator to develop a hypothesis test for the probability $P(X < X')$ that an individual from a population survives for a time X shorter than the survival time X' of an individual from another population. We call this test *generalized IDP rank-sum test* (or *IDP test* for short), since it extends the procedure proposed in (Benavoli et al., 2014b) to right-censored data. Many authors have considered the problem of comparing two survival curves and proposed several tests. Letòn and Zuluaga (2005) provide a review of the different tests and compare them on a number of artificial datasets. The most commonly used tests are the log-rank test (Mantel, 1966) and some generalizations of the Mann-Witney-Wilcoxon rank-sum test (Mann and Whitney, 1947; Wilcoxon, 1945), which are more powerful in case of early differences in the hazard function, since they weight more the first part of the survival curve. However, there are no clear recommendations as to which test is the best with respect to differences in the hazard function and in the setting of samples (with respect to number of observations and censoring distributions). Moreover, those tests are designed to detect generic differences between the two curves, whereas often we are interested in verifying whether the probability of survival is significantly larger in one population when compared to another. Those tests might be less reliable for such purpose, which is for instance prominent in case of crossing hazards. IDP allows deriving the posterior probability distribution of $P(X < X')$ and thus it is well suited to test directional hypothesis such as $P(X < X') > 1/2$, or equivalently $P(X < X') > P(X \geq X')$ (meaning that it is more probable that an individual of the first population has shorter life than an individual of the second population rather than longer or equal). When it comes to decision making, we can interpret the IDP model in two ways. A first way, more strictly related to prior near-ignorance models, verifies whether decisions are prior-dependent or not given a predefined value of the prior strength. Decisions are taken only when they do not depend on the choice of the prior. A second, perhaps more intuitive, interpretation considers the upper and lower bounds given by the IDP

model as obtained by adding in the most favorable/unfavorable positions a number of *fictitious* data equal to the strength of the DP prior. This is possible because we demonstrate that upper and lower posterior bounds are obtained by considering discrete priors made of a finite number of Dirac delta functions. This interpretation of the IDP shares some similarity with the method of influence curves used in robust analysis (Hampel, 1974; Huber, 2010) to study how an estimator varies when a new observation is added. In light of this interpretation, we consider as an indicator of robustness of the IDP test decision the minimum number of individuals that, if added to the available samples, could make the test contradict its decision (that is, the minimum prior strength that would make the decision prior-dependent).

The IDP test is widely applicable in clinical trials, and can be of interest also in reliability analysis of industrial components or social sciences, where X and X' represent the time to an event of interest. We will show through numerical simulations that the test is more consistent than traditional ones, especially in the case of non-proportional hazards, and that the IDP test can identify when decisions taken by the log-rank and Wilcoxon-type tests are not robust. We will also show that the results of the IDP test cannot be easily reproduced by the simple decrease of the significance level of the traditional tests (typically used to increase robustness); indeed, the IDP test provides a more sensible way to achieve reliable results.

Besides experiments using simulated data, we analyze the Australian AIDS dataset (Ripley and Solomon, 1994; Venables and Ripley, 2002). Previous analyses suggested that a difference in survival time of AIDS patients existed when considering use (or not) of drugs, disease transmission, as well as Australian regions (Ripley and Solomon, 1994) (even though it is acknowledged that this last finding required further investigation), but not when considering gender. Our IDP test confirms those findings and outputs how reliable they are. In particular, the IDP test confirms that the survival time difference identified by traditional tests between individuals of distinct Australian regions is not reliable and that further information is needed, thus providing a mathematical model and reasoning that clarifies previous results.

2 Robust estimation of survival probabilities from right censored data

Let $X_1, \ldots, X_n$ be independent random variables describing the lifetimes of n individuals which are censored on the right by n independent follow up times $Y_1, \ldots, Y_n$ which are also independent of $X_1, \ldots, X_n$. The X_i are identically distributed (as X) with cumulative distribution function (CDF) $F(t)$. It will be slightly more convenient in what follows to deal with survival functions $S(t) = 1 - F(t)$. In practice, we can only observe the event (either death or censoring) that occurs first, and thus each observation of the sample $\mathbf{Z}_n = \{(Z_1, d_1), \ldots, (Z_n, d_n)\}$ includes the time of the observation, $Z_i = \min(X_i, Y_i)$, and the indicator d_i which is 1 if $X_i < Y_i$ and 0 otherwise. In this work we only consider non-informative censoring, which occurs if individuals leave the trial independently of their actual state. Our main goal is to estimate the survival function $S(t)$. Let us define $\tilde{X}_1, \ldots, \tilde{X}_{n^d}$ as the subset of n^d uncensored observations (for which $d_i = 1$) and $\tilde{Y}_1, \ldots, \tilde{Y}_{n^c}$ the subset of $n^c = n - n^d$ censored observations (for which $d_i = 0$). Without loss of generality, we assume that Z_i, $\tilde{X}_i$ and $\tilde{Y}_i$ are ordered. For ease of presentation, it is assumed that X and Y are continuous and there are no ties; however, if ties are present, one can still use the results of this paper by introducing between tied observations a fictitious distance going down to zero (Coolen and Yan, 2003).

A traditional nonparametric estimator for the survival function from lifetime data is the Kaplan-Meier (KM) estimator (Kaplan and Meier, 1958):

$$\hat{S}(t) = \prod_{\tilde{X}_i \leq t} \frac{n_i^+}{n_i}, \tag{1}$$

where $n_i = n - i + 1$ and $n_i^+ = n - i$ are the numbers of individuals at risk at time $\tilde{X}_i$ and just after $\tilde{X}_i$, respectively (that is, the individuals that have not yet been removed from the sample by censoring or

death). The variance of this estimator is (usually) approximated by Greenwood's formula:

$$E[(\hat{S}(t) - E[\hat{S}(t)])^2] = \sum_{\tilde{X}_i \leq t} \frac{\hat{S}^2(t)}{n_i n_i^+}, \tag{2}$$

and pointwise confidence intervals are based on the assumption of normality for the distribution of $\hat{S}(t)$. However, this is a rather rough assumption if n is small and $\hat{S}(t)$ is close to 1 or 0. A better approximation assumes normality for the logarithmic transformation of the hazard function (Moeschberger and Klein, 2003).

For the purpose of Bayesian nonparametric estimation, the Dirichlet process (Ferguson, 1973) can be used. Let $\mathbb{X}$ be a standard Borel space with Borel σ-field $\mathscr{B}_{\mathbb{X}}$ and $\mathbb{P}$ be the space of probability measures on $(\mathbb{X}, \mathscr{B}_{\mathbb{X}})$ equipped with the weak topology and the corresponding Borel σ-field $\mathscr{B}_{\mathbb{P}}$. Let $\mathbb{M}$ be the class of all probability measures on $(\mathbb{P}, \mathscr{B}_{\mathbb{P}})$. We call the elements $\mu \in \mathbb{M}$ nonparametric priors. An element of $\mathbb{M}$ is called a DP distribution $\mathrm{Dp}(\alpha)$ with base measure α if for every finite measurable partition $B_1, \ldots, B_m$ of $\mathbb{X}$, the vector $(P(B_1), \ldots, P(B_m))$ has a Dirichlet distribution with parameters $(\alpha(B_1), \ldots, \alpha(B_m))$, where $\alpha(\cdot)$ is a finite positive Borel measure on $\mathbb{X}$. Said $s = \alpha(\mathbb{X})$ the prior strength of the DP and $\alpha^* = \alpha/s$ the normalized base measure, we will use $\mathrm{Dp}(s, \alpha^*)$ as an alternative notation equivalent to $\mathrm{Dp}(\alpha)$; moreover, if $\mathbb{X} = \mathbb{R}$, we shall also describe $P \sim \mathrm{Dp}(s, \alpha^*)$ by saying $P \sim \mathrm{Dp}(s, G)$, where G stands for the cumulative distribution function (CDF) of α^*. As $F(t) = P(0, t]$ and $S(t) = 1 - P(0, t]$, hereafter we will refer indifferently to $\mathrm{Dp}(s, \alpha^*)$ as the prior of P, $F(t)$ and $S(t)$.

Consider the partition B_1 and $B_1^c = \mathbb{X} \backslash B_1$; then, if $P \sim \mathrm{Dp}(s, \alpha^*)$, from the definition of DP we have that $(P(B_1), P(B_1^c)) \sim Dir(s\alpha^*(B_1), s(1 - \alpha^*(B_1^c)))$, which is a Beta distribution. From the moments of the Beta distribution, we can thus derive that:

$$\mathscr{E}[P(B_1)] = \alpha^*(B_1), \quad \mathscr{E}[(P(B_1) - \mathscr{E}[P(B_1)])^2] = \frac{\alpha^*(B_1)(1 - \alpha^*(B_1))}{(s+1)}, \tag{3}$$

where we have used the calligraphic letter $\mathscr{E}$ to denote expectation with respect to the Dirichlet process. This shows that the normalized measure α^* of DP reflects the prior expectation of P, while the scaling parameter s controls how much P is allowed to deviate from its mean.

Let f be a real-valued bounded function on $\mathbb{X}$. We call $E[f] = \int f dP$ a predictive inference about X, where P is a probability measure on $(\mathbb{X}, \mathscr{B}_{\mathbb{X}})$. If $P \sim \mathrm{Dp}(s, \alpha^*)$, then the expectation with respect to the Dirichlet process of $E[f]$ is

$$\mathscr{E}\left[E(f)\right] = \mathscr{E}\left[\int f dP\right] = \int f d\mathscr{E}[P] = \int f d\alpha^*. \tag{4}$$

One of the most remarkable properties of DP priors is that the posterior distribution of P is again a DP. Let $X_1, \ldots, X_n$ be independent and identically distributed samples from P and $P \sim \mathrm{Dp}(s, \alpha^*)$. Then the posterior distribution of P given the observations is

$$P|X_1, \ldots, X_n \sim \mathrm{Dp}\left(s+n, \frac{s}{s+n}\alpha^* + \frac{1}{s+n}\sum_{i=1}^{n}\delta_{X_i}\right), \tag{5}$$

where δ_{X_i} is an atomic probability measure centered at X_i. This means that the Dirichlet process satisfies a property of conjugacy, in the sense that the posterior for P is again a Dirichlet process with updated unnormalized base measure $\alpha + \sum_{i=1}^{n}\delta_{X_i}$. From (3) and (5) we can easily derive the posterior mean and variance of P.

If we use $\mathrm{Dp}(s, \alpha^*)$ as prior for the distribution of $S(t)$, then the posterior distribution given the sample $\mathbf{Z}_n$ of randomly censored observations is a mixture of Dirichlet processes (Blum and Susarla, 1977) and

thus the conjugacy property is not satisfied anymore when data are censored. The p-th order moment of the posterior distribution of $S(t)$ is (Susarla and Van Ryzin, 1976)

$$\mathscr{E}[S(t)^p|\mathbf{Z}_n] = \prod_{j=0}^{p-1} \frac{\alpha(t,\infty) + n_t^+ + j}{s + n + j} \prod_{\tilde{Y}_i \leq t} \frac{\alpha[\tilde{Y}_i,\infty) + n_i + j}{\alpha[\tilde{Y}_i,\infty) + n_i^+ + j}, \tag{6}$$

where n_t and n_t^+ are the numbers of individuals at risk at time t and just after t. Here, as in the rest of the paper, the product is taken to be one if the number of factors is zero (that is, if $t < \tilde{Y}_1$). All prior inferences about $S(t)$ are fully determined by the choice of α^*, which reflects our prior guess about the distribution of X, and s, which reflects the strength of our belief in such guess. In the absence of prior information about the distribution of X, choosing the infinite dimensional parameter α^* may be hard and arbitrary, and may affect the reliability of posterior inferences. The only solution to this problem that has been proposed so far is the limiting DP obtained when the prior strength goes to zero. This model has been subject to some controversy, since it is not actually non-informative and assigns zero posterior probability to any set that does not include the observations (see Rubin (1981) for a detailed discussion). This situation is aggravated in survival analysis with censored data. In this case, the posterior survival distribution may assign positive probability to survival times later than that of the latest observed lifetime; since there are no observed deaths there, the posterior distribution will depend on the prior base measure even if the prior strength goes to zero. As a consequence, this choice of prior does not always remove the dependence of the posterior inferences from the prior base measure. Therefore, we propose a different choice, which calls back to the ideas of sets of prior probabilities and near-ignorance models (Berger, 1994; Pericchi and Walley, 1991; Walley, 1991).

Definition 1. *Let $\mu \in \mathbb{M}$ be a nonparametric prior on P and $\mathscr{E}_\mu[E(f)]$ be the expectation of $E[f]$ with respect to μ. A class of nonparametric priors $\mathscr{T} \subset \mathbb{M}$ is called a prior ignorance model for predictive inferences about X if, for any real-valued bounded function f on $\mathbb{X}$, it satisfies:*

$$\underline{\mathscr{E}}[E(f)] = \inf_{\mu \in \mathscr{T}} \mathscr{E}_\mu[E(f)] = \inf_{x \in \mathbb{X}} f(x), \ \overline{\mathscr{E}}[E(f)] = \sup_{\mu \in \mathscr{T}} \mathscr{E}_\mu[E(f)] = \sup_{x \in \mathbb{X}} f(x), \tag{7}$$

where $\underline{\mathscr{E}}[E(f)]$ and $\overline{\mathscr{E}}[E(f)]$ are the lower and upper bounds of $\mathscr{E}_\mu[E(f)]$, respectively. ∎

From (7) it can be observed that the range of $\mathscr{E}_\mu[E(f)]$ under the class $\mathscr{T}$ is the same as the original range of f. In other words, by specifying the class $\mathscr{T}$, we are not giving any information on the value of the expectation of f. Here, we have focused on the expectation of f as most of the nonparametric predictive inferences about a variable X can be expressed in terms of expectations of f, i.e., $\int f(X)dP_X = E[f]$. For instance, in this paper we are interested in $S(t) = \int_t^\infty dF(x) = E[f_t(x)]$ with $f_t(x) = I_{(t,\infty)}(x)$. Therefore, being prior ignorant on this kind of inferences is important in nonparametric statistics. Note that, if $\mathscr{T}$ is a prior ignorance model, then all sets $\mathscr{T}^*$ including $\mathscr{T}$ as a subset are prior ignorance models. However, a set of priors should not be too large otherwise the posterior inferences would become too little informative to be of any practical use. For instance, if we take $\mathscr{T} = \mathbb{M}$, that is, the largest possible set of priors, we have a prior ignorance model whose posterior inferences remain vacuous, or, in other words, a model that cannot learn from data (Walley, 1991, Sec. 7.3.7). Such model would be useless in practice.

We are now ready to define the IDP.

Definition 2. *We call prior near-ignorance DP (or IDP) the class $\mathscr{T}$ of DPs such that*

$$\mathscr{T} = \{\mathrm{Dp}(s,\alpha^*) : \ \forall \alpha^* \in \mathbb{P}\}.$$

∎

Thus, the IDP set of priors is obtained by fixing the prior strength s and letting the normalized base measure α^* vary in the set of all probability distributions $\mathbb{P}$. We say that the IDP is a model of *near*-ignorance because to define it the modeler has to choose s. By considering the DP priors with base measures concentrated on the values of X that give the inf and sup of f it is easy to see from (4) that $\mathscr{T}$ verifies (7).

Given independent and identically distributed $X_1, \ldots, X_n$ from $P \sim \mathrm{Dp}(s, \alpha^*)$, the posterior lower and upper bounds of $\mathscr{E}_\mu[E(f)|X_1, \ldots, X_n]$ are (Benavoli et al., 2014b):

$$\begin{aligned}
\underline{\mathscr{E}}\big[E(f)|X_1, \ldots, X_n\big] &= \frac{s}{s+n} \inf_{x \in \mathbb{X}} f(x) + \frac{\sum_{i=1}^n f(X_i)}{s+n}, \\
\overline{\mathscr{E}}\big[E(f)|X_1, \ldots, X_n\big] &= \frac{s}{s+n} \sup_{x \in \mathbb{X}} f(x) + \frac{\sum_{i=1}^n f(X_i)}{s+n},
\end{aligned} \tag{8}$$

and the effect of prior near-ignorance vanishes asymptotically, that is, it holds (for any finite s)

$$\overline{\mathscr{E}}\big[E(f)|X_1, \ldots, X_n\big] - \underline{\mathscr{E}}\big[E(f)|X_1, \ldots, X_n\big] = \frac{s}{s+n}\big(\sup_{x \in \mathbb{X}} f(x) - \inf_{x \in \mathbb{X}} f(x)\big) \to 0, \quad \text{for } n \to \infty.$$

Moreover, from (8), it also follows that if we collect n i.i.d. random variable $X_1, \ldots, X_n$ the IDP estimates converge to Σ as $n \to \infty$, where $\Sigma = \lim_{n \to \infty} \frac{\sum_{i=1}^n f(X_i)}{s+n}$. Therefore, IDP provides an asymptotically consistent estimator of Σ.

2.1 Survival curve estimator

We apply the IDP to obtain a robust estimator of the survival function $S(t)$ and provide credible intervals for its value. As the IDP is a prior near-ignorance model, we have that prior inferences for the survival function $S(t) = E[I_{(t,\infty)}(x)]$ at a given time t are vacuous. Theorem 1, whose proof is given in the supplementary material, gives the posterior moments of the survival function $S(t)$ when using the IDP approach.

Theorem 1. *Given the IDP prior, the posterior lower and upper p-th order moments of $S(t)$ are*

$$\underline{\mathscr{E}}[S(t)^p|\mathbf{Z}_n] = \prod_{j=0}^{p-1} \frac{n_t^+ + j}{n_t^+ + s + j} \prod_{\tilde{X}_i \leq t} \frac{n_i^+ + s + j}{n_i + s + j}, \quad \overline{\mathscr{E}}[S(t)^p|\mathbf{Z}_n] = \prod_{j=0}^{p-1} \prod_{\tilde{X}_i \leq t} \frac{n_i^+ + s + j}{n_i + s + j}, \tag{9}$$

and are obtained for the prior DP with base probability measure $\alpha^* = \delta_{X_0}$, where $X_0 = t$ for the lower expectation, and $X_0 > t$ for the upper expectation. ∎

Hereafter, we will denote by $\underline{\mu}_t$ and $\overline{\mu}_t$ the prior measures that give, respectively, the lower and the upper moments of $S(t)$. In the interpretation of the prior δ_{X_0} as fictitious data (deaths in this case), we can see that the upper bound for $S(t)$ is found when all fictitious deaths happen after the time of interest, i.e. $X_0 > t$, whereas the lower is obtained when they happen at $X_0 = t$. From (9) we can see that the imprecision $\overline{\mathscr{E}}[S(t)^p|\mathbf{Z}_n] - \underline{\mathscr{E}}[S(t)^p|\mathbf{Z}_n]$ increases at the increase of s and that if $s \to \infty$ then $\underline{\mathscr{E}}[S(t)^p|\mathbf{Z}_n] \to 0$ and $\overline{\mathscr{E}}[S(t)^p|\mathbf{Z}_n] \to 1$. When $s \to \infty$ we are in the situation discussed in Section 2 where the set of priors $\mathscr{T} = \mathbb{M}$ is too large and thus the model cannot learn from data.

From (9) we can also derive upper and lower bounds for the expectation of $S(t)$,

$$\underline{\mathscr{E}}[S(t)|\mathbf{Z}_n] = \frac{n_t^+}{s + n_t^+} \prod_{\tilde{X}_i \leq t} \frac{s + n_i^+}{s + n_i}, \quad \overline{\mathscr{E}}[S(x)|\mathbf{Z}_n] = \prod_{\tilde{X}_i \leq t} \frac{s + n_i^+}{s + n_i}, \tag{10}$$

and state the following result:

Corollary 1. *The upper and lower bounds in (10) for the expectation of $S(t)$ always encompass the KM*

estimate $\hat{S}(t)$. ∎

We can see that the difference between the bounds of $\mathscr{E}_\mu[S(t)|\mathbf{Z}_n]$, which is

$$\overline{\mathscr{E}}[S(t)|\mathbf{Z}_n] - \underline{\mathscr{E}}[S(t)|\mathbf{Z}_n] = \frac{s}{s+n_t^+} \prod_{\tilde{X}_i \leq t} \frac{s+n_i^+}{s+n_i},$$

goes to zero for $n_t^+ \to \infty$. From this and from the result in Corollary 1, it follows that the IDP estimator converges to the Kaplan-Meier one, which, for $t \leq \max_i\{Z_i\}$, is a uniformly consistent estimator of $S(t)$ (Wang, 1987).

The posterior distribution of $S(t)$ derived from the IDP is used to build credible intervals for the survival function without resorting to the Gaussian approximation. We define the pointwise two-sided $100(1-\gamma)\%$ credible interval $[S_*(t), S^*(t)]$ for $S(t)$, as the symmetric interval having lower probability $1-\gamma$ of including the true value; that is, we want to ensure probability at least $1-\gamma$ that $S_*(t) \leq S(t) \leq S^*(t)$. Such interval is obtained by solving the equations:

$$\underline{\mathscr{E}}\big[I_{S(t)\leq S^*(t)}\big|\mathbf{Z}_n\big] = 1-\gamma/2, \ \ \text{and} \ \ \underline{\mathscr{E}}\big[I_{S(t)\geq S_*(t)}\big|\mathbf{Z}_n\big] = 1-\gamma/2, \tag{11}$$

with I the indicator function. This implies that

$$\begin{aligned} \mathscr{E}_\mu\big[I_{S_*(t)\leq S(t)\leq S^*(t)}\big|\mathbf{Z}_n\big] &= 1 - \big(1 - \mathscr{E}_\mu\big[I_{S(t)\geq S_*(t)}\big|\mathbf{Z}_n\big]\big) - \big(1 - \mathscr{E}_\mu\big[I_{S(t)\leq S^*(t)}\big|\mathbf{Z}_n\big]\big) \\ &= \mathscr{E}_\mu\big[I_{S(t)\geq S_*(t)}\big|\mathbf{Z}_n\big] + \mathscr{E}_\mu\big[I_{S(t)\leq S^*(t)}\big|\mathbf{Z}_n\big] - 1 \\ &\geq 1-\gamma. \end{aligned}$$

Theorem 2. *Given the IDP prior, the posterior lower bounds of $\mathscr{E}_\mu\big[I_{S(t)\leq a}\big|\mathbf{Z}_n\big]$ and $\mathscr{E}_\mu\big[I_{S(t)\geq a}\big|\mathbf{Z}_n\big]$ are found for the same priors that give, respectively, the upper and the lower moments of $S(t)$. That is,*

$$\underline{\mathscr{E}}\big[I_{S(t)\leq a}\big|\mathbf{Z}_n\big] = \mathscr{E}_{\overline{\mu}_t}\big[I_{S(t)\leq a}\big|\mathbf{Z}_n\big], \ \ \text{and} \ \ \underline{\mathscr{E}}\big[I_{S(t)\geq a}\big|\mathbf{Z}_n\big] = \mathscr{E}_{\underline{\mu}_t}\big[I_{S(t)\geq a}\big|\mathbf{Z}_n\big].$$

∎

Deriving analytical expressions for these expectations is usually a difficult task, especially as the number of observations increases, since conjugacy does not hold in case of censored data. Nevertheless, we have developed an efficient method for the numerical computation of posterior upper and lower distributions of $S(t)$ by Monte Carlo sampling, based on the fact that they correspond to the atomic priors in Theorem 1 (theoretical derivations and additional technical details are given in the supplementary material).

2.2 Comparison of survival curves

Let X and X' be lifetimes of individuals from two populations, F_1 and F_2 their distribution functions and S_1 and S_2 their survival functions; we want to test the hypothesis $P(X < X') \leq 1/2$ against $P(X < X') > 1/2$, which resembles a standard Wilcoxon sum-rank test. For this, we need to estimate

$$P(X < X') = \int F_1(t)dF_2(t) = \int \big[1 - S_1(t)\big]d[-S_2(t)]$$

based on samples $\mathbf{Z}_n$ and $\mathbf{Z}'_{n'}$ of n and n' observations from S_1 and S_2, respectively, which are assumed to be independent. We take as prior distributions for S_1 and S_2 the Dirichlet Processes $\mu_1 = \mathrm{Dp}(s, G_1)$ and $\mu_2 = \mathrm{Dp}(s, G_2)$, respectively (we use the same s for ease of presentation, but that is not a limitation). Then, by the properties of the IDP, it follows that a-priori: $\mathscr{E}_{\mu_1,\mu_2}[P(X < X')] = \int G_1(t)dG_2(t)$. The IDP satisfies the condition of prior near-ignorance, hence a-priori we are fully ignorant about the hypothesis

$P(X < X') \leq 1/2$. Using IDP, the hypothesis test can be performed in two steps. First, we define a loss function (Berger, 1993, Ch 4.4)

$$L(P,a) = \begin{cases} k_0 \cdot I_{\{P(X<X')>\frac{1}{2}\}}, & \text{if } a = 0, \\ k_1 \cdot I_{\{P(X<X')\leq\frac{1}{2}\}}, & \text{if } a = 1, \end{cases} \tag{12}$$

where a is our action and k_0 is the loss we incur by taking the action $a = 0$ (that is, declaring that $P(X < X') \leq \frac{1}{2}$) when actually $P(X < X') > \frac{1}{2}$, while k_1 gives the loss we incur by taking the action $a = 1$ (that is, declaring that $P(X < X') > \frac{1}{2}$) when actually $P(X < X') \leq \frac{1}{2}$. Second, we compute the expected value of this loss and we choose $a = 1$ if $\mathscr{E}_{\mu_1,\mu_2}[L(P,0)|\mathbf{Z}_n,\mathbf{Z}'_{n'}] \geq \mathscr{E}_{\mu_1,\mu_2}[L(P,1)|\mathbf{Z}_n,\mathbf{Z}'_{n'}]$, i.e.,

$$k_0 \mathscr{E}_{\mu_1,\mu_2}\left[I_{\{P(X<X')>\frac{1}{2}\}}|\mathbf{Z}_n,\mathbf{Z}'_{n'}\right] \geq k_1 \mathscr{E}_{\mu_1,\mu_2}\left[I_{\{P(X<X')\leq\frac{1}{2}\}}|\mathbf{Z}_n,\mathbf{Z}'_{n'}\right]$$
$$\Rightarrow \mathscr{E}_{\mu_1,\mu_2}\left[I_{\{P(X<X')>\frac{1}{2}\}}|\mathbf{Z}_n,\mathbf{Z}'_{n'}\right] \geq \frac{k_1}{k_0+k_1}, \tag{13}$$

and $a = 0$ otherwise. When the above inequality is satisfied, we can declare that $P(X < X') > \frac{1}{2}$ with probability at least $\frac{k_1}{k_0+k_1}$. For comparison with the traditional test we will take $\frac{k_1}{k_0+k_1} = 1 - \gamma$, where γ is the significance level of the test (usually the notation α is used, but α has another meaning here); notice however that, while a principled way of choosing γ is lacking in the traditional tests, we can set it in a more informed way based on the losses k_0 and k_1 expected in case of error. Finally, according to the decision rule in (13), we verify whether $\underline{\mathscr{E}}\left[I_{\{P(X<X')>\frac{1}{2}\}}|\mathbf{Z}_n,\mathbf{Z}'_{n'}\right] > 1 - \gamma$, and whether $\overline{\mathscr{E}}\left[I_{\{P(X<X')>\frac{1}{2}\}}|\mathbf{Z}_n,\mathbf{Z}'_{n'}\right] > 1 - \gamma$, and then:

1. If both inequalities are true, we declare that the probability of longer survival is higher for the first group than for the second group with probability greater than $1 - \gamma$.

2. If both inequalities are false, we declare that the probability of longer survival is higher for the first group than for the second group with probability lower than the desired probability of $1 - \gamma$.

3. If the left-hand inequality is false but the right-hand inequality is true, we are in an *indeterminate* situation.

Thus, to perform the test we need to compute lower and upper bounds for the probability that $P(X < X') > \frac{1}{2}$. Lower bounds for the expectation of $P(X < X')$ and the probability of $P(X < X') > \frac{1}{2}$ in the case of censored data are given in the two following theorems. Similar results can be found for the upper bounds.

Theorem 3. *The lower bound of the expectation of $P(X < X')$ is found using the DP priors $\overline{\mu}_1 = \mathrm{Dp}(s, \alpha_{10}^*)$ and $\underline{\mu}_2 = \mathrm{Dp}(s, \alpha_{20}^*)$ with base probability measure*

$$d\alpha_{10} = s\delta_{Z_0}, \quad d\alpha_{20} = s\sum_{i=0}^{n'}\pi_i\delta_{Z_i'^+}, \tag{14}$$

where $Z_0 > \max\{Z_1,\ldots,Z_n,Z_1',\ldots,Z_{n'}'\}$, $Z_i' < Z_i'^+ < Z_{i+1}$, with $i = 0,\ldots,n'$, $Z_0' = 0$ and $Z_{n'+1}' = \infty$, and where $\pi = (\pi_0, \pi_1, \ldots, \pi_{n'})$ is a vector of weights verifying $\pi_i = 0$ if $d_i' = 1$ and $\sum_{i=0}^{n'}\pi_i = 1$. The vector π is obtained by minimizing the expectation

$$\mathscr{E}_{\overline{\mu}_1,\underline{\mu}_2}[P(X<X')|\mathbf{Z}_n,\mathbf{Z}'_{n'}] = \sum_{i=1}^{n^d}\left(\overline{\mathscr{E}}[S_1(\tilde{X}_{i-1})|\mathbf{Z}_n] - \overline{\mathscr{E}}[S_1(\tilde{X}_i)|\mathbf{Z}_n]\right)\mathscr{E}[S_2(\tilde{X}_i)|\mathbf{Z}'_{n'}], \tag{15}$$

where $\tilde{X}_0 = 0$, $\overline{\mathcal{E}}[S_1(t)|\mathbf{Z}_n]$ is given by (10), and the posterior expectation of $S_2(t)$ is derived from (6) and is equal to:

$$\mathcal{E}_{\underline{\mu}_2}[S_2(t)|\mathbf{Z}'_{n'}] = \prod_{Z'_i \leq t} \frac{\alpha_{20}(Z'_i, +\infty) + n'_i - d'_i}{\alpha_{20}[Z'_i, +\infty) + n'_i}.$$

$\blacksquare$

Theorem 4. *The infimum of the probability that $P(X < X') > \frac{1}{2}$ is found for the DP priors $\overline{\mu}_1 = \mathrm{Dp}(s, \alpha_{10}^*)$ and $\underline{\mu}_2^* = \mathrm{Dp}(s, \alpha_{20}^*)$ with base probability measure*

$$d\alpha_{10} = s\delta_{Z_0}, \quad d\alpha_{20} = s\sum_{i=0}^{n'} \pi'_i \delta_{Z_i'^+}, \tag{16}$$

where $Z_0 > \max\{Z_1, \ldots, Z_n, Z'_1, \ldots, Z'_{n'}\}$, $Z'_i < Z_i'^+ < Z_{i+1}$, with $i = 0, \ldots, n'$, $Z'_0 = 0$ and $Z'_{n'+1} = \infty$, and where $\pi' = (\pi'_0, \pi'_1, \ldots, \pi'_{n'})$ is a vector of weights that verifies $\pi'_i = 0$ if $d'_i = 1$ and $\sum_{i=0}^{n'} \pi'_i = 1$. The vector π' is obtained by minimizing $\mathcal{E}_{\pi'}[I_{P(X<X')>\frac{1}{2}}|\mathbf{Z}_n, \mathbf{Z}'_{n'}]$. $\blacksquare$

From these theorems we see that the extreme values of the expectation of $P(X < X')$ and of the probability of $P(X < X') > \frac{1}{2}$ are found for discrete priors. Hence, the priors can still be easily interpreted as additional deaths and the posterior distribution of $P(X < X')$ at these priors efficiently computed by Monte Carlo sampling (see proofs and results in the supplementary material). However, while minimizing (15) can be reduced to a convex optimization problem, so that the prior $\underline{\mu}_2$ can be easily found, the minimization problem in Theorem 4 is much more computationally expensive since for each vector π' the objective function $\mathcal{E}_{\pi'}[I_{P(X<X')>\frac{1}{2}}|\mathbf{Z}_n, \mathbf{Z}'_{n'}]$ has to be computed by Monte Carlo sampling. Thus, finding the exact extrema in this case is not tractable. Conservative bounds for the value of $\mathcal{E}_{\mu_1,\mu_2}[I_{P(X<X')>\frac{1}{2}}|\mathbf{Z}_n, \mathbf{Z}'_{n'}]$ can be efficiently computed based on the following result.

Theorem 5. *Let $S_{\underline{\mu}_t,i}$, $i = 1, 2$, be a process such that at each time t the distribution of $S_{\underline{\mu}_t,i}(t)$ is the posterior distribution of $S_i(t)$ when the prior base measure is δ_t (that is, the one that gives the lower expectation of $S_i(t)$) and $S_{\overline{\mu},i}$ the posterior distribution of $S_i(t)$ when the prior base measure is δ_{X_0}, $X_0 > t$. Then,*

$$\begin{aligned}
\mathcal{E}_{\mu_1,\mu_2}\left[I_{\{P(X<X')>\frac{1}{2}\}}\big|\mathbf{Z}_n, \mathbf{Z}'_{n'}\right] &\geq \mathcal{E}\left[I_{\int S_{\underline{\mu}_t,2}(t)d[-S_{\overline{\mu},1}(t)]>\frac{1}{2}}\big|\mathbf{Z}_n, \mathbf{Z}'_{n'}\right] \\
\mathcal{E}_{\mu_1,\mu_2}\left[I_{\{P(X<X')>\frac{1}{2}\}}\big|\mathbf{Z}_n, \mathbf{Z}'_{n'}\right] &\leq 1 - \mathcal{E}\left[I_{\int S_{\underline{\mu}_t,1}(t)d[-S_{\overline{\mu},2}(t)]>\frac{1}{2}}\big|\mathbf{Z}_n, \mathbf{Z}'_{n'}\right].
\end{aligned} \tag{17}$$

$\blacksquare$

We have verified in several case studies that, when s is small, the bounds in (17) are close to the exact ones (see the supplementary material). Thus, for small s, they do not increase too much the fraction of indeterminate instances. Instead, for large s these bounds are too rough and in most cases do not lead to determinate decisions. For this reason, we propose to use a less conservative approach that approximates the probability obtained using $\underline{\mu}_2^*$ with that obtained for $\underline{\mu}_2$. In the absence of censoring this result is exact as we have the equality between these two probabilities.

Theorem 6. *In the absence of censoring, the upper and lower bounds of $\mathcal{E}_{\mu_1,\mu_2}\left[I_{\{P(X<X')>\frac{1}{2}\}}\big|\mathbf{Z}_n, \mathbf{Z}'_{n'}\right]$ are found for the priors that give, respectively, the upper and lower expectation of $P(X < X')$.* $\blacksquare$

Unfortunately, this result does not hold in case of censored data (see the supplementary material for a counterexample). Then, the lower and upper probabilities obtained using $\underline{\mu}_2$ are inner approximations of the correct ones, but we have verified empirically that the difference between them is very small (see the supplementary material for more details). Moreover, results of the simulations in Section 3 show that a test performed using this approximation is calibrated and more reliable than the traditional tests. Therefore, the use of the priors $\overline{\mu}_1, \underline{\mu}_2$ allows an efficient computation of the lower and upper probabilities and, at the same time, provides a good approximation of the true lower and upper probabilities. For this reason we suggest to use this approximation.

Although the IDP test shares several similarities with a standard Bayesian approach, it embodies a change of paradigm when it comes to taking decisions. The IDP test has the advantage of producing an *indeterminate* outcome when the decision about the hypothesis being tested is *prior-dependent*, that is, the IDP test is able to tell whether the decision which minimizes the expected loss would change depending on the DP base measure one chooses. Therefore, the IDP test is robust in providing a determinate decision only when all DP priors in the class represented by the IDP agree. Indeterminate outcomes indicate that collecting additional data could provide valuable further evidence to the decision maker.

To perform a sensitivity analysis of the test decision with respect to the possible values of the prior strength s, we propose to evaluate the maximum value (s_{max}) for which the decision remains determinate. Then, we are guaranteed that by repeating the test with s_{max} more observations in each group we will never contradict the decision taken with the available data. In this view, s_{max} can be seen as a measure of robustness of the test decision. Notice that values of s_{max} greater than 1 indicate already a quite robust decision: for example, if $s_{max} = 2$ we are implicitly considering the very unlikely event that all the 4 fictitious observations (2 for each group) fall in the most unfavorable position for the hypothesis under test.

3   Simulations

Using simulated data, we compare the IDP estimator for survival functions with the traditional KM estimator, where we employ Greenwood's formula plus the Gaussian approximation for the estimation of confidence intervals regarding the survival probability. For obtaining the KM estimator, we use the R package survival (Therneau, 2014) and we consider both the linear *plain* confidence interval (which is routinely constructed by most statistical packages), and the interval built using a *log-log* transformation of the cumulative hazard rate (Moeschberger and Klein, 2003). We consider simulated cases where lifetime data and follow up times are generated from an exponential distribution, such that $X_1, \ldots, X_n, Y_1, \ldots, Y_n \sim Exp(\lambda = 5)$. As an illustration, Figure 1 shows examples of survival curves estimated for $n = 25$ and $n = 100$ with 95% confidence intervals for the log-log KM and the 95% credible interval defined in (11) for the IDP with $s = 0.25$. While a (somewhat small) difference between the two estimators can be noticed for $n = 25$, the same is unperceivable with $n = 100$. Notice however that the lower bound for the IDP credible interval accounts better for the observed censored data. The lower bound of the IDP decreases at censoring times, thus reflecting the fact that, after censoring, we have less individuals under observation and thus the estimate we can (robustly) provide has wider uncertainty and, thus, wider credible intervals.

Figure 2 shows the coverage of the 95% intervals obtained by the IDP and the KM estimators (both plain and log-log) over 1000 repetitions (and varying time instants). The plain KM confidence intervals provide a coverage lower than the target value of 0.95, as it is known that Greenwood's estimator tends to underestimate the true variance of KM for small to moderate samples (Moeschberger and Klein, 2003). The log-log KM estimator provides a better coverage than plain KM, but still below the target. Note that coverage decreases at later time instants, which is exactly when fewer data are available for the estimation (fewer individuals at risk). On the other hand, IDP is more robust and, in general, achieves a coverage slightly larger than 0.95. If one chooses a very low value of s, for instance $s = 0.01$, then IDP cannot reach the target coverage either. Table 1 presents coverages (averaged over all time instants) for different values of s. The greater the choice of s the more conservative is IDP and the width of its credible intervals

(a) $n = 25$

(b) $n = 100$
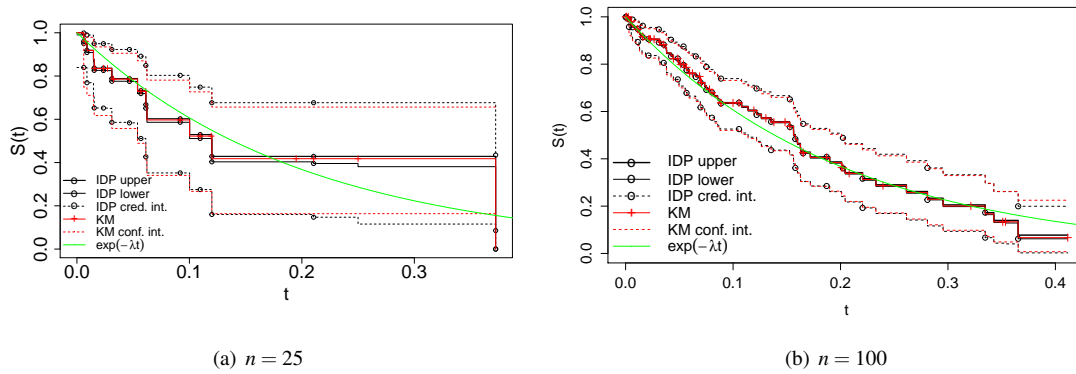
Figure 1 Survival curve estimates from the IDP ($s = 0.25$) and the log-log KM approaches after n observations. The continuous lines represent the KM estimator (red line) and the lower (thin black line) and upper (thick black line) IDP estimates; the dotted lines represent the confidence/credible intervals estimated by the KM (red) and the IDP (black).

Table 1 Coverage of pointwise KM's confidence intervals and IDP's credible intervals.

	KM (plain)	KM (log-log)	IDP(s=0.01)	IDP(s=0.25)	IDP(s=0.5)
n=25	0.878	0.932	0.924	0.961	0.974
n=100	0.932	0.953	0.945	0.961	0.968

increases. In these results, as it can be seen in Table 1, values for s above 0.5 are already too conservative. We take $s = 0.25$ for most experiments from now on, because in general it already achieves the desired coverage.

(a) $n = 25$

(b) $n = 100$
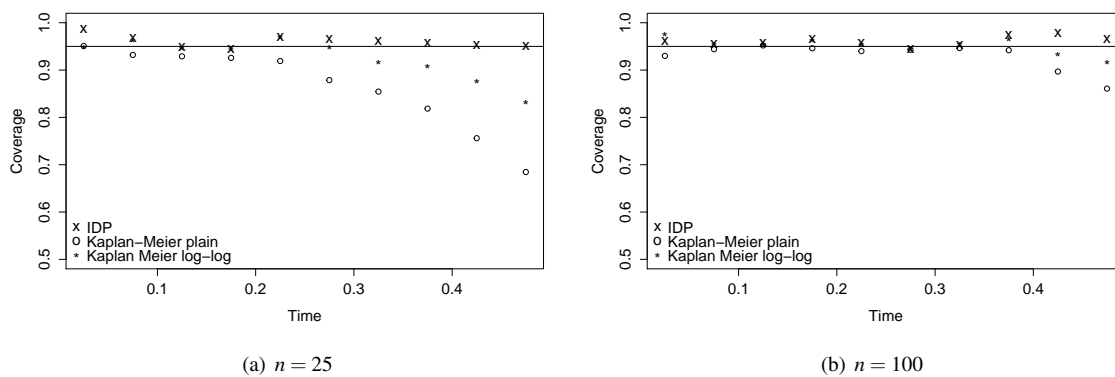
Figure 2 Coverage of the Kaplan-Meier and IDP credible intervals with $s = 0.25$.

For the purpose of comparing the survival of individuals of two independent populations, we have applied the IDP test to a number of artificial datasets simulating the different scenarios proposed by Letòn and Zuluaga (2005), that is, same hazard (SH), proportional hazards (PH), early hazard difference (EHD) and late hazard difference (LHD), and we have compared it against the log-rank test and the (Peto-Peto

Table 2 Fraction of H1 decisions for the different tests in the cases where the IDP is determinate or indeterminate (round brackets). Into square brackets in the IDP column, we report the fraction of indeterminate outcomes.

	Log-rank	Peto-Peto	IDP($s = 0.25$)
SH	0.0261 (0.405)	0.0282 (0.392)	0.0250 [0.079]
PH	0.7201 (0.678)	0.6948 (0.433)	0.7069 [0.171]
EHD	0.7563 (0.202)	0.8578 (0.896)	0.8268 [0.163]
LHD	0.4411 (0.970)	0.2122 (0.090)	0.2333 [0.100]

modification of the) Gehan-Wilcoxon test, both implemented in the `R package survival`. Tests are one-sided, always analyzing whether $P(X < X') > \frac{1}{2}$ (which is considered as the alternative hypothesis), with X and X' denoting the corresponding variables of interest in two independent populations. Table 2 gives the fraction of rejection of the null hypothesis over the cases where IDP yields a determinate decision and over the cases where it cannot take a decision (into round brackets); the fraction of indeterminate cases for the IDP test is also shown (square brackets). The SH line regards same hazards, so the null hypothesis is true and the fraction of rejections represents the type-I error (which should be lower than $\gamma = 0.05$). In the other lines, the fraction of rejections measures the power of the test. Further details about the simulation setting are presented in the supplementary material.

Let us first consider the determinate cases. We see that all tests perform similar for SH. As expected, the Peto-Peto test outperforms the log-rank test only in the EHD scenario. Overall, IDP presents similar power to Peto-Peto, and is outperformed by the log-rank for LHD. Even though the IDP test is never the most powerful, it outperforms the log-rank in the EHD scenario and the Peto-Peto test in the LHD scenario (it ties with Peto-Peto for PH), thus it can be seen as a compromise between the other two tests when the scenario under study is unknown (which is often the case). The IDP test also gives us an extra information: it suggests that part of the cases (indicated between square brackets) are *indeterminate* and could be better decided if more data were available (with few more data, some of those decisions could easily become a rejection of the null hypothesis). While traditional tests always issue a determinate decision, the IDP test acknowledges that a reliable decision cannot be made without collecting more data. In the cases where the IDP test is indeterminate we can see from Table 2 that the type-I error in the SH scenario is much larger than γ both for the log-rank and Peto-Peto tests. In the PH scenario the power of the traditional tests reduces with respect to that obtained when the IDP test is determinate. In the EHD scenario the power of the Peto-Peto test is much larger than that of the Log-rank test since the former is biased toward H1 in such scenario, whereas the log-rank test is biased toward H0. The opposite situation happens for the LHD scenario. Indeed those two tests carry some assumptions that are usually hard to verify a priori. The IDP test is capable of detecting the situations where those assumptions mostly affect the test decisions; the fact that the two tests usually disagree in those cases shows that the decision is difficult and may not be reliable. Moreover, we will see in Example 3.2 that the assumptions underlying the log-rank and Peto-Peto tests may affect the reliability of their decisions under the null hypothesis. Therefore, for the robustness of the decision it is important to isolate those instances where the decision strongly relies on those uncertain assumptions.

We believe that robustness is an important feature of a test, as it may contribute to avoid misleading decisions and false discoveries which affect the replicability of results. In this context, we show an example where log-rank and Peto-Peto tests are not reliable (a reliable test should not reject the null hypothesis with probability higher than $\gamma = 0.05$ when it is true).

Example 3.1 Consider two groups with different sizes ($n = 20$ and $n' = 100$) from the same population. Lifetime data are taken from the same exponential distribution $Exp(\lambda = 1)$; censoring times are sampled from the uniform with support $[0, 2]$. Due to the very different sample size, the type-I error, evaluated over 1000 repetitions, is greater than the desired value $\gamma = 0.05$ both for log-rank and Peto-Peto tests (0.065

and 0.061, respectively). If we limit to consider only the samples where the IDP test is determinate (89% of the cases), the Type-I error is 0.032 both for the log-rank and the Peto-Peto tests and 0.030 for the IDP test. Considering instead only the samples where the IDP is indeterminate, the error become 0.300 for the log-rank and 0.336 for the Peto-Peto test.

Besides the robustness that is provided by the IDP test in general, Example 3.2 shows that the IDP test is particularly useful in situations where hazards cross (as well as in situations where one is unsure about such crossing, if he wants to take reliable decisions), as most classical tests become unreliable.

Example 3.2 Consider two scenarios A and B, each of which with two groups of $n = 50$ and $n' = 25$ samples. Lifetime data are sampled from two populations with survival functions given in Figure 3.

(a) Scenario A   (b) Scenario B
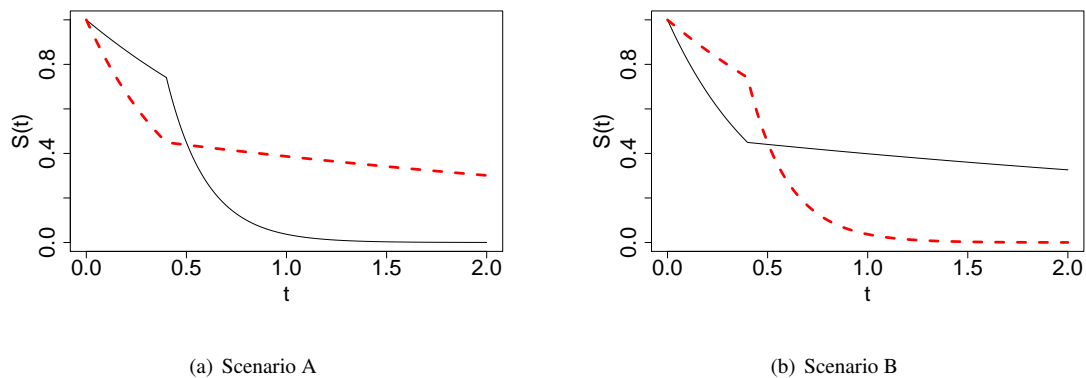
Figure 3 Survival functions for control (continuous thin line) and treated (dashed thick line) populations.

The survival curves in the two scenarios are very similar, but curves of control and treatment have their shape interchanged. The hazard function is designed so that in both scenarios $P(X < X')$ is slightly smaller than $\frac{1}{2}$ (with X representing the control group); hence the null hypothesis is true in both scenarios. The type-I error of the three tests, evaluated on 1000 different repetitions, is shown in Table 3. The log-rank has a large Type-I error in scenario A, while Peto-Peto has very large Type-I error in scenario B, and both are way greater than the target of 0.05. Indeed, these tests are designed to respectively weight more late and early differences in the curves and, thus, cannot account correctly for the overall behavior of the functions. The IDP test does not assume any characteristic for the survival function; its Type-I error has been always smaller than 0.05. When the IDP is indeterminate the error is 1 or almost 1 for the cases where the tests are biased toward the alternative hypothesis (that is, scenario A for log-rank and scenario B for Peto-Peto test). In the opposite scenario, tests are biased toward favoring the null hypothesis and thus the error is much smaller; this is particularly evident for the log-rank, which is more strongly biased and thus never rejects the null hypothesis apart for only 2 extreme cases where also the IDP and the Peto-Peto do reject it. The log-rank does not achieve error below 0.05 in Scenario A even if we set the significance level of the test at 0.01. Peto-Peto test, instead, obtains an error below 0.05 if we set $\gamma = 0.01$ (but not below the target γ itself), but such choice is completely arbitrary, as one cannot know (or rely on) which would be the γ to achieve type-I error of 5% (besides the sharp drop in power that the test would face). Into round brackets we report the error over the indeterminate cases identified by an IDP test with $\gamma = 0.05$. We expect that, in those cases which are critical for a decision criteria with $\gamma = 0.05$, a more conservative test using $\gamma = 0.01$ would not reject the null hypothesis thus achieving a very small Type-I error error. This is always true except for the log-rank test in scenario A (with an error of 1) thus confirming once more the strong bias of this test in case of crossing hazards.

Table 3 Type-I error of different tests for significance level $\gamma = 0.05$ and $\gamma = 0.01$ over the determinate cases. Into round brackets we report the Type-I error for Log-rank and Peto-Peto in the IDP indeterminate instances, and into square brackets the fraction of indeterminate outcomes for IDP.

	Log-rank		Peto-Peto		IDP ($s = 0.25$)
	$\gamma = 0.05$	$\gamma = 0.01$	$\gamma = 0.05$	$\gamma = 0.01$	$\gamma = 0.05$
Scenario A	0.320 (1.000)	0.140 (0.929)	0.044 (0.071)	0.007 (0.000)	0.048 [0.028]
Scenario B	0.014 (0.000)	0.006 (0.000)	0.173 (1.000)	0.063 (0.800)	0.045 [0.020]

Table 4 p-values (Log-rank and Peto-Peto tests) and posterior probabilities (IDP test) for the AIDS dataset. (H0)/(H1) indicates the decision taken at the level $\gamma = 0.05$, while (I) indicates an indeterminate outcome.

	Group 1	Group 2	N. Data	Log-rank	Peto-Peto	IDP ($s = 0.25$)	s_{max}
1	Male	Female	2754/89	0.182 (H0)	0.525 (H0)	0.575/0.738 (H0)	6.2
2	NSW	VIC	1780/588	0.228 (H0)	0.024 (H1)	0.907/0.921 (H0)	1.5
3	QLD	NSW	226/1780	0.046 (H1)	0.027 (H1)	0.929/0.969 (I)	-
4	No Drug	Drug	2506/120	0.011 (H1)	0.019 (H1)	0.986/0.990 (H1)	3.2
5	Blood	Haemoph.	94/46	0.046 (H1)	0.006 (H1)	0.991/0.995 (H1)	2.1

4 Australian AIDS survival data

In this section, we consider the Australian AIDS survival dataset (Ripley and Solomon, 1994; Venables and Ripley, 2002) of cases reported to the Australian National Centre in HIV Epidemiology and Clinical Research. The dataset contains records of 2843 individuals diagnosed from 1982 to 1991, of which 1787 died prior to the end of the study.

We consider the five pairs of groups listed in Table 4. The first considers gender, the next two consider populations from distinct regions of Australia (*NSW*, *VIC*, and *QLD* mean, respectively, New South Wales including the Australian Capital Territory, Victoria, and Queensland, the three most populated regions in the country), the fourth considers users of drugs (or not), and the fifth considers the reported transmission between *Blood* (that is, receipt of blood, blood components or tissue) and *Haemophilia* (coagulation disorder). These tests have been chosen to confirm (or to find evidence against) the results reported long ago by the responsible for the dataset himself (Ripley and Solomon, 1994). For each pair, we test the hypothesis that the survival time is shorter for the first group of individuals than for the second group. More precisely, let X be the survival time of individuals from Group 1 (for example, *Male* in the first line of Table 4) and X' that of individuals from Group 2. We test $P(X < X') \leq \frac{1}{2}$ (null hypothesis, or H0) against $P(X < X') > \frac{1}{2}$ (alternative hypothesis, or H1) using the log-rank and the Peto-Peto modification of the Gehan-Wilcoxon test, as well as the IDP test. In Table 4, we present p-values for the log-rank and Peto-Peto tests and posterior probabilities for the IDP test. We use the significance level $\gamma = 0.05$; then the null hypothesis is rejected (denoted H1) if the p-value is smaller than γ (log-rank and Peto-Peto tests) or if the lower probability of the alternative hypothesis is greater than $1 - \gamma$ (IDP test); if, instead, the upper probability is greater than $1 - \gamma$ but the lower probability is not, then we declare it *indeterminate*. Figure 4 shows the survival curves related to the tests under consideration.

According to the previous study (Ripley and Solomon, 1994), no difference in survival has been verified between *Male* and *Female*, and between *NSW* and *VIC*, while a difference in survival time was identified for *QLD* versus *NSW*, *No-drug* versus *drug* usage, and *Blood* versus *Haemophilia* (the last comparison was not directly performed, but can be inferred from their conclusions). Table 4 shows that our results are consistent with those findings. The IDP test confirms that gender is not significant to discriminate survival time even if the strength of its prior were as large as 6, that is, even if 6 fictitious additional data

(a) *Male* and *Female*.

(b) *NSW* and *VIC*.

(c) *NSW* and *QLD*.

(d) *No-drug* and *Drug*.

(e) *Blood* and *Haemophilia*.
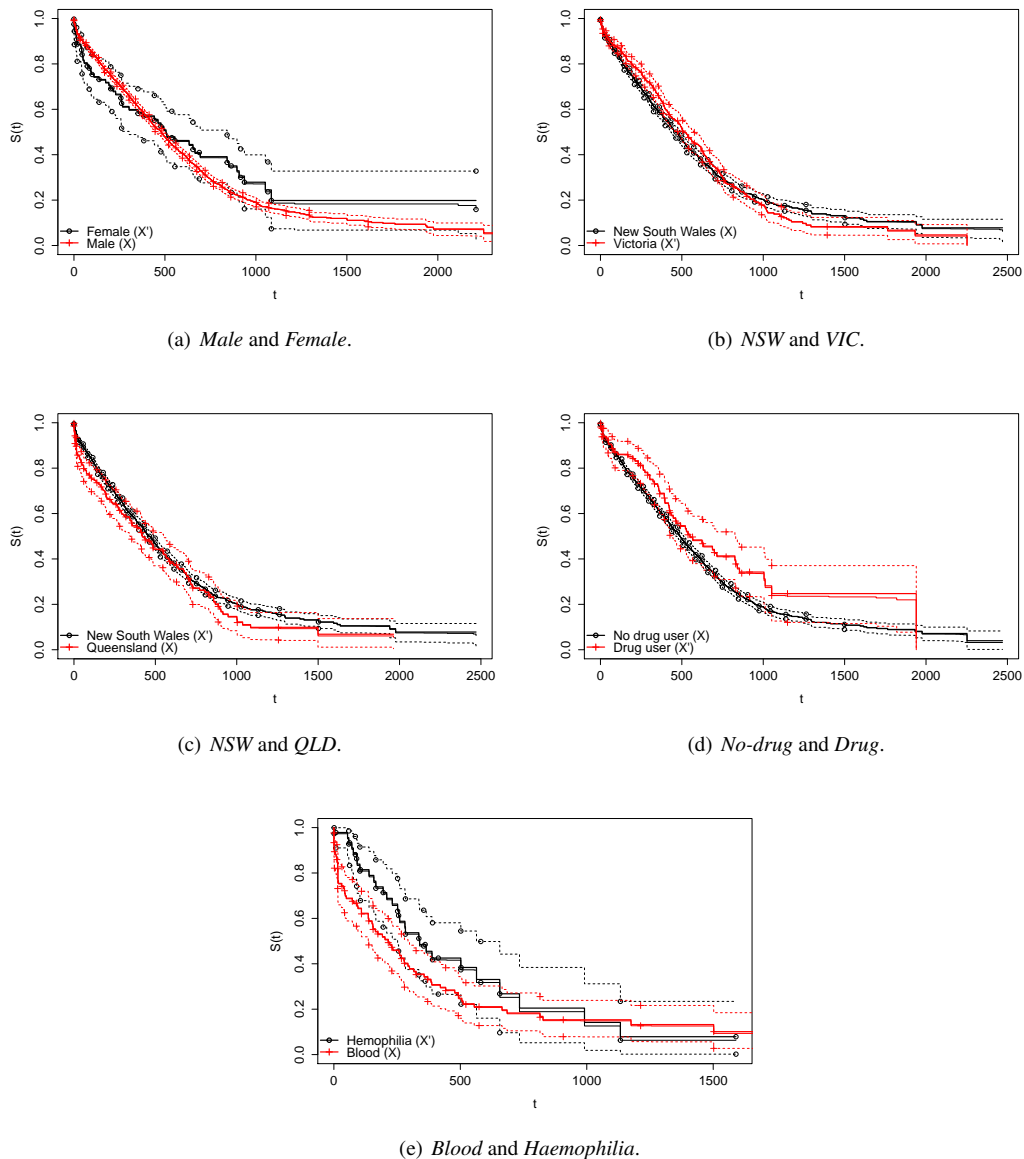
Figure 4 Survival curves estimated with IDP ($s = 0.25$) for the groups in Table 4. Continuous lines represent the IDP lower (thin line) and upper (thick line) expectations of $S(t)$; the dotted lines represent the upper and lower bounds of the credible intervals.

samples were placed in each group in the most adversarial positions in time, the conclusion would remain unchanged. Log-rank and Peto-Peto do not reject H0 either.

For *NSW* against *VIC*, log-rank and Peto-Peto tests provide very different p-values. The Peto-Peto test weights more the first part of the survival curves, where the curve of *NSW* lies below that of *VIC*. In the second part, the curves appear interchanged, but this is somehow neglected by the test. This is a clear case where the proportional hazards assumption fails, and traditional tests are less reliable. IDP indicates that

up to $s = 1.5$ we can reliably *not* reject H0. These results are supported by data and discussions presented in Australian technical reports (National Centre in HIV Epidemiology and Clinical Research, 2013; Kirby Institute, 2013).

Interestingly, both log-rank and Peto-Peto reject H0 for the comparison *QLD* versus *NSW*, in accordance with previous results (Ripley and Solomon, 1994). However, Ripley and Solomon (1994) clearly acknowledge that the poor survival of *QLD* with respect to *NSW* is *"obscure and require further investigation"*. Indeed recent Queensland reports (Queensland Health, 2011, 2012) (as well as the previously mentioned Australian reports) show no reason to believe that survival of AIDS patients in that region is shorter than in New South Wales. The IDP test identifies such doubtful situation and responds an *indeterminate* outcome (with an indeterminacy that exists even if $s \to 0$), suggesting that further data should be collected for reaching a better decision.

Finally, both *No-drug* versus *Drug* and *Blood* versus *Haemophilia* tests have H0 rejected with all the three tests, even though p-values of log-rank and Peto-Peto are quite different in the latter. This might be explained by the possible late cross in the curves, as shown in Figure 4. In both comparisons, IDP provides a robust decision of rejecting H0 together with a measure of its robustness, given by the maximum strength s such that the decision of rejecting H0 would yet take place. For *Blood* versus *Haemophilia*, even if $s_{max} = 2.1$ additional fictitious data samples were included in each of the groups ($s_{max} = 3.2$ fictitious data in the *No-drug* versus *Drug*), we would still reject H0, so the results are indeed quite reliable.

From these results we can see that the IDP test does not simply issue an indeterminate outcome when the p-value of the traditional tests is close to the significance level γ: for instance, the p-value of the log-rank test is very close to the confidence level γ both in comparison *QLD* versus *NSW* and *Blood* versus *Haemophilia* (the p-value is 0.046 in both cases), but only in the first comparison the IDP test is indeterminate. The same situation can be seen for the Peto-Peto test with comparisons *NSW* versus *VIC* (p-value= 0.024) and *QLD* versus *NSW* (p-value= 0.027). Neither can the IDP test be mimicked by issuing an indeterminate outcome when the decision of the other tests differ: they both rejected H0 in the comparison *QLD* versus *NSW*, while IDP test is indeterminate and they take different decisions in the comparison *NSW* versus *VIC* while IDP issues H0.

5   Conclusions

In this paper we have employed a model of prior near-ignorance for nonparametric inferences based on the Dirichlet process to develop robust methods for survival analysis with right-censored data. The approach has the benefits of a Bayesian inference while avoiding completely the need of specifying the infinite-dimensional parameter of the Dirichlet Process. The only free parameter is the strength of the prior, which has a clear interpretation as the number of additional fictitious data (placed in the most adversarial way) that we impose to the model while checking whether the decision would remain unchanged. This makes the elicitation of the prior very easy, and allows us to compute posterior inferences for which no closed form expression exists by a simple Monte Carlo sampling from the Dirichlet distribution, thus avoiding more demanding sampling approaches typically used for the Dirichlet process (for example, stick breaking). Based on this prior near-ignorance model, we have developed an estimator for the survival curves which can provide reliable credible intervals for the probability of survival. We have also proposed a general, simple and conservative approach for testing the difference in survival of individuals from two independent populations called IDP test, which is a robust alternative to log-rank and other weighted Wilcoxon rank-sum tests. The IDP test is able to identify whether the decision is prior-dependent, and gives the possibility of evaluating the size of the difference between survival times through the posterior distribution of $P(X < X')$. Moreover, the IDP test allows us to perform an analysis of robustness with respect to the only parameter that has to be elicited in the IDP test, the prior strength, by computing its maximum value for which the IDP test remains determinate. This can be interpreted as a measure of robustness of the decision. Results have shown that this test has similar power than that of classical tests

and yet is more reliable. The study of the Australian AIDS dataset has demonstrated that the IDP test was able to identify the low reliability of a previous conclusion that the survival of individuals with AIDS from the Queensland region in Australia is significantly worst than for individuals from New South Wales, a result that had been questioned by the curator of the dataset himself.

6 Supplementary Material and Software

Supplementary material is available online at the journal website. It contains details about the numerical approximation of the posterior distribution of $S(t)$ and $P(X < X')$ and about the simulations and proofs of all results presented in this paper.

The `R package IDPsurvival` available from the CRAN Repository provides the software in the form of R code, together with a sample input data set and complete documentation. Source code to reproduce the results in this paper is available as Supporting Information on the journal's web page (`http://onlinelibrary.wiley.com/doi/xxx/suppinfo`).

Acknowledgments

Conflict of Interest
The authors have declared no conflict of interest.

References

Benavoli, A., Corani, G., Mangili, F., Zaffalon, M., and Ruggeri, F. (2014a). A Bayesian Wilcoxon signed-rank test based on the Dirichlet process. In *Proceedings of the 31st International Conference on Machine Learning, Beijing*.

Benavoli, A., Mangili, F., Ruggeri, F., and Zaffalon, M. (2014b). Imprecise Dirichlet process with application to the hypothesis test on the probability that $X < Y$. *ArXiv:1402.2755*.

Berger, J. (1994). An overview of robust Bayesian analysis with discussion. *Test*, 3(1):5–124.

Berger, J., Rios Insua, D., and Ruggeri, F. (2000). Bayesian robustness. In Rios Insua, D. and Ruggeri, F., editors, *Robust Bayesian Analysis*, volume 152 of *Lecture Notes in Statistics*, pages 1–32. Springer New York.

Berger, J. O. (1993). *Statistical decision theory and Bayesian analysis*. Springer-Verlag.

Blum, J. and Susarla, V. (1977). On the posterior distribution of a Dirichlet process given randomly right censored observations. *Stochastic Processes and their Applications*, 5(3):207–211.

Coolen, F. (1997). An imprecise Dirichlet model for Bayesian analysis of failure data including right-censored observations. *Reliability Engineering & System Safety*, 56(1):61–68.

Coolen, F. and Yan, K. (2003). Nonparametric predictive comparison of two groups of lifetime data. In *International Symposium on Imprecise Probabilities and Their Applications. (ISIPTA)*, volume 3, pages 148–161.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230.

Ferguson, T. and Phadia, E. (1979). Bayesian nonparametric estimation based on censored data. *The Annals of Statistics*, 7(1):163–186.

Hampel, F. (1974). The influence curve and its role in robust estimation. *J. of the American Statistical Association*, 69(346):383–393.

Hjort, N. (1990). Nonparametric Bayes estimators based on Beta processes in models for life history data. *Annals of statistics*, 18(3):985–1500.

Huber, C. (2010). Robust versus nonparametric approaches and survival data analysis. In *Advances in Degradation Modeling*, pages 323–337. Springer.

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

Kirby Institute (2013). HIV, viral hepatitis and sexually transmissible infections in Australia Annual Surveillance Report 2013. Technical report, University of New South Wales.

Letòn, E. and Zuluaga, P. (2005). Relationships among tests for censored data. *Biometrical J.*, 47(3):377–387.

Mann, H. and Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematics and Statistics*, 18(1):50–60.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports. Part 1*, 50(3):163–170.

Moeschberger, M. and Klein, J. (2003). *Survival analysis: Techniques for censored and truncated data: Statistics for Biology and Health*. Springer.

National Centre in HIV Epidemiology and Clinical Research (2013). Mapping HIV outcomes: geographical and clinical forecasts of numbers of people living with HIV in Australia. Technical report, University of New South Wales.

Pericchi, L. and Walley, P. (1991). Robust Bayesian credible intervals and prior ignorance. *Int. Statistical Review*, 59:1–23.

Queensland Health (2011). 2011 HIV/AIDS Report: Epidemiology and Surveillance. Technical report, Queensland Government.

Queensland Health (2012). HIV in Queensland 2012. Technical report, Queensland Government.

Ripley, B. and Solomon, P. (1994). A note on Australian AIDS survival. Technical report, University of Adelaide. Department of Statistics Research Report 94/3.

Rubin, D. (1981). The Bayesian bootstrap. *The Annals of Statistics*, 9(1):130–134.

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *J. of the American Statistical Association*, 71(356):897–902.

Therneau, T. (2014). *A Package for Survival Analysis in S*. R package version 2.37-7.

Venables, W. and Ripley, B. (2002). *Modern applied statistics with S*. Springer.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.

Walley, P. (1996). Inferences from multinomial data: learning about a bag of marbles. *J. of the Royal Statistical Society. Series B (Methodological)*, 58(1):3–57.

Wang, J.-G. (1987). A note on the uniform consistency of the kaplan-meier estimator. *The Annals of Statistics*, 15(3):1313–1316.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.

Zhou, M. (2004). Nonparametric Bayes estimator of survival functions for doubly/interval censored data. *Statistica Sinica*, 14(2):533–546.

Supplementary material to "Reliable survival analysis based on the Dirichlet Process"

Francesca Mangili[*,1], **Alessio Benavoli** [1], **Cassio P. de Campos** [1], and **Marco Zaffalon** [1]

[1] IPG-IDSIA, Galleria 2, 6928 Manno-Lugano, Switzerland

This supplementary material contains an initial Section 1 with a brief description of the properties of the Dirichlet Process (DP), omitted in the paper, that will be used in the sequel. In Sections 2 and 3 we derive the procedures for the numerical approximation of the posterior distribution of, respectively, the survival probability and the probability $P(X < X')$ for survival curve comparison and propose a solution to manage the presence of ties in the IDP test. We also provide a discussion about the choice of the prior strength of the DP (Section 4), further details about the simulations (Section 5) and proofs of the theorems appearing in the paper (Section 6). Finally, in Section 7 we derive an analytical expression for the posterior variance of $P(X < X')$ that can be used for the Gaussian approximation of the distribution of $P(X < X')$ when the number of observations is large.

1 Some properties of the Dirichlet process

Let $P \sim \mathrm{Dp}(s, \alpha^*)$, α^* be a finite positive Borel measure on $\mathbb{X}$, with $\alpha^*(\mathbb{X}) = 1$, and G stand for the cumulative distribution function (CDF) of α^*. Hereafter we list some useful properties of DP that will be used in the sequel.

(a) In case $\mathbb{X} = \mathbb{R}$, since P is completely defined by its CDF F, a-priori we say $F \sim \mathrm{Dp}(s, G)$ and a posteriori we can write as follows:

$$F|X_1, \ldots, X_n \sim \mathrm{Dp}\left(s+n, \frac{s}{s+n}G + \frac{1}{s+n}\sum_{i=1}^{n} I_{[X_i, \infty)}\right), \tag{S.1.1}$$

where I is the indicator function.

(b) Consider an element $\mu \in \mathbb{M}$ which puts all its mass at the probability measure $P = \delta_x$ for some $x \in \mathbb{X}$. This can also be modeled as $\mathrm{Dp}(s, \delta_x)$ for each $s > 0$.

(c) Assume that $P_1 \sim \mathrm{Dp}(s_1, \alpha_1^*)$, $P_2 \sim \mathrm{Dp}(s_2, \alpha_2^*)$, $(w_1, w_2) \sim Dir(s_1, s_2)$ and P_1, P_2, (w_1, w_2) are independent, then (Ghosh and Ramamoorthi, 2003, Sec. 3.1.1):

$$w_1 P_1 + w_2 P_2 \sim \mathrm{Dp}\left(s_1 + s_2, \frac{s_1}{s_1 + s_2}\alpha_1^* + \frac{s_2}{s_1 + s_2}\alpha_2^*\right). \tag{S.1.2}$$

(d) Let P_x have distribution $\mathrm{Dp}(s+n, \frac{s}{s+n}\alpha^* + \frac{1}{s+n}\sum_{i=1}^{n}\delta_{X_i})$. We can write

$$P_x = w_0 P + \sum_{i=1}^{n} w_i \delta_{X_i}, \tag{S.1.3}$$

where $(w_0, w_1, \ldots, w_n) \sim Dir(s, 1, \ldots, 1)$ and $P \sim \mathrm{Dp}(s, \alpha^*)$ (it follows from (b)–(c)).

---

[*]Corresponding author: e-mail: francesca@idsia.ch

2 Numerical approximation of the posterior distributions of $S(t)$

In Blum and Susarla (1977) an analytical expression of the posterior distribution of the survival function $S(t)$ given a Dirichlet process prior and randomly censored data is derived and proven to be a mixture of Dirichlet processes. To deal with such mixture may be hard in practice. A solution is to approximate the posterior distribution of $S(t)$ by Monte Carlo sampling. For this, it is convenient to use the formalism introduced by Peterson (1977): let $L(t,d)$ be the distribution of $(Z_i, d_i) \in \mathbb{X} = \mathbb{R} \times \{0,1\}$, Peterson expresses the Kaplan-Meier estimator based on the sub-survival functions $S^*(t) = \int_t^\infty dL(s,1)$ and $Q^*(t) = \int_t^\infty dL(s,0)$ from, respectively, deaths and censoring. The following results hold (Peterson, 1977, Properties 1 and 3):

$$S^*(t) = \int_t^\infty Q(s)d[-S(s)],$$

$$S^*(t) + Q^*(t) = P(Z > t) = P(X > t)P(Y > t) = S(t)Q(t),$$

(S.2.4)

where $Q(t) = 1 - C(t)$ and $C(t)$ is the CDF of the i.i.d. censoring times $Y_1, \ldots, Y_n$.

From (S.2.4) it follows:

$$\frac{dS(t)}{S(t)} = \frac{dS^*(t)}{S^*(t) + Q^*(t)}.$$

(S.2.5)

Let $P_1 = \int_0^\infty dS^*(t)$ and $P_0 = \int_0^\infty dQ^*(t)$ be the probabilities of death and censoring, we assume that (P_1, P_0), $S^*(t)$ and $Q^*(t)$ are independent and take the priors

$$dL(t|d=1)\delta_1(d) = \frac{dS^*(t)\delta_1(d)}{P_1} \quad \sim \quad \mathrm{Dp}(s, \alpha_s^*(t)\delta_1(d)),$$

$$dL(t|d=0)\delta_0(d) = \frac{dQ^*(t)\delta_0(d)}{P_0} \quad \sim \quad \mathrm{Dp}(q, \alpha_q^*(t)\delta_0(d)),$$

$$(P_1, P_0) \quad \sim \quad Dir(s, q),$$

where $\delta_i(d)$, $i = 0, 1$, is the Dirac delta function which takes value 1 if $d = i$ and 0 otherwise. Then by (S.1.2)

$$dL(t,d) = dS^*(t)\delta_1(d) + dQ^*(t)\delta_0(d) = P_1 dL(t|d=1)\delta_1(d) + P_0 dL(t|d=0)\delta_0(d) \qquad \text{(S.2.6)}$$

$$\sim \mathrm{Dp}\left(s+q, \frac{s}{s+q}\alpha_s^*(t)\delta_1(d) + \frac{q}{s+q}\alpha_q^*(t)\delta_0(d)\right) = \mathrm{Dp}(s+q, \alpha_L^*(t,d)). \qquad \text{(S.2.7)}$$

A posteriori it holds

$$dL(t,d)|\mathbf{Z}_n \sim \mathrm{Dp}(s+q+n, \tfrac{s+q}{s+q+n}\alpha_L^*(t,d) + \tfrac{n}{s+q+n}\sum_{i=1}^n \delta_{(Z_i, d_i)}(t,d)). \qquad \text{(S.2.8)}$$

We see from (S.2.8) that the conjugacy property is satisfied for $L(t,d)$, and for this reason it is easier to sample from its posterior rather than from the posterior of $S(t)$. Moreover, the following Lemma S.1 shows that for $q \to 0$ and $\alpha_L^* = \alpha_s^*$ the prior in (S.2.6) for $L(t,d)$ is equivalent to the prior $S(t) \sim \mathrm{Dp}(s, \alpha^*)$ used in the IDP model. Since, from Theorem 1 in the paper, the prior base measure $\alpha_s(t)$ is discrete for upper and lower expectations of $S(t)$, the posterior distribution of $L(t,d)$ at the extrema is also discrete, and thus we can easily draw from it samples $L^{(k)}$, $k = 1, 2, \ldots$, based on the result of Lemma S.2 below. Hence, we propose to sample from the posterior distribution of $L(t,d)$, and then derive for each sample $L^{(k)}(t,d)$ the corresponding values $S^{(k)}(t)$ based on the results of Lemma S.3. Having chosen, based on Lemma S.1, a prior for $L(t,d)$ which is equivalent to the prior $\mathrm{Dp}(s, \alpha^*)$ used in the IDP model for $S(t)$, the samples $S^{(k)}$ derived from $L^{(k)}$, $k = 1, 2, \ldots$, will be distributed according to the posterior distribution of $S(t)$ given the prior $S(t) \sim \mathrm{Dp}(s, \alpha^*)$ and thus provide a numerical approximation of it.

Lemma S.1. *For $q \to 0$ and $\alpha_s^*(t) = \alpha^*(t)$, the prior in (S.2.6) for $L(t,d)$ implies the prior $S(t) \sim \mathrm{Dp}(s, \alpha^*)$ of the IDP model for $S(t)$.* ∎

Proof. *First notice that if $q \to 0$ the probability with respect to the DP $\mathrm{Dp}(s+q, \alpha_L^*)$ that $P_0 = 0$ goes to 1 (recall that P_0 is the probability of sampling a censored data) and so does the probability that the sub-survival function $Q^*(t) = P_0 \int_t^\infty dL(s|d=0)$ is equal to 0 (notice that $Q^*(t)$ is not a probability measure and thus it is not normalized). The relation between the survival function $S(t)$ and the sub-survival functions $S^*(t), Q^*(t)$ is given by (S.2.5); from it one can see that if $Q^*(t) = 0$, then $S^*(t) = S(t)$. It also holds that if $q \to 0$ then the probability that $P_1 = 1$ goes to 1, and thus $\frac{S^*(t)}{P_1} = S^*(t) = S(t)$ almost surely. Then, $S(t)$ will be distributed as $\frac{S^*(t)}{P_1}$, that is as the Dirichlet process $\mathrm{Dp}(s, \alpha_s^*) = \mathrm{Dp}(s, \alpha^*)$.* ∎

The following result is used for sampling from $L(t,d)$ when the base measure is discrete.

Lemma S.2. *Let $\alpha_L = \sum_{i=1}^m g_i \delta_{(Z_i, d_i)}$ be the base measure of the Dirichlet process $L(t,d)$. A sample $L^{(k)}(z,d)$ from $\mathrm{Dp}(\alpha_L)$ has the form*

$$dL^{(k)}(t,d) = \sum_{i=1}^m w_i \delta_{(Z_i, d_i)}(t,d), \tag{S.2.9}$$

where the vector of weights $(w_1, \ldots, w_m)$ has a Dirichlet distribution with parameters $(g_1, \ldots, g_m)$. ∎

Proof. *First, notice that the probability of sampling outside the support of α_L is 0. Consider now the partition $\{(Z_1, d_1)\}, \ldots, \{(Z_m, d_m)\}$ of the support of α_L; it follows from the definition of the DP that $(P(Z_1, d_1), \ldots, P(Z_m, d_m)) \sim Dir(g_1, \ldots, g_m)$.* ∎

The following result allows deriving samples $S^{(k)}(t)$ from the collection of samples $L^{(k)}(t,d)$, $k = 1, 2, \ldots$.

Lemma S.3. *Let $S^{*(k)}$ and $Q^{*(k)}$ be the sub-survival functions of the sample distribution $L^{(k)}(t,d)$ in (S.2.9). If $S^{*(k)}$ and $Q^{*(k)}$ have no discontinuities in common, the corresponding survival function $S^{(k)}(t)$ is:*

$$S^{(k)}(t) = \prod_{\{Z_i \leq t\}} \frac{S^{*(k)}(Z_{i+1}) + Q^{*(k)}(Z_i)}{S^{*(k)}(Z_i) + Q^{*(k)}(Z_i)} = \prod_{\{i: d_i = 1, Z_i \leq t\}} \frac{\sum_{j=1}^n w_j I_{(Z_i, \infty)}(Z_j)}{\sum_{j=1}^n w_j I_{[Z_i, \infty)}(Z_j)}. \tag{S.2.10}$$

∎

Proof. *The first equality follows from Theorem 2.1 in Peterson (1977) by considering that $dS^{*(k)}(t)$ is discrete. To obtain the second equality, we consider that the fractions*

$$\frac{S^{*(k)}(Z_{i+1}) + Q^{*(k)}(Z_i)}{S^{*(k)}(Z_i) + Q^{*(k)}(Z_i)}$$

in (S.2.10) simplify if $d_i = 0$, since in this case $S^{*(k)}(Z_{i+1}) = S^{*(k)}(Z_i)$. If, instead, $d_i = 1$ we have that

$$S^{*(k)}(Z_i) = \int_{Z_i}^{\infty} d[L^{(k)}(s, d = 1)] = \int_{Z_i}^{\infty} \sum_{j=1}^{m} w_j \delta_{(Z_j, d_j)}(s, 1) ds$$

$$= \sum_{j=1}^{m} w_j I_{[Z_i, \infty)}(Z_j) \delta_1(d_j),$$

and, similarly

$$Q^{*(k)}(Z_i) = \int_{Z_i}^{\infty} d[L^{(k)}(s, d = 0)] = \int_{Z_i}^{\infty} \sum_{j=1}^{m} w_j \delta_{(Z_j, d_j)}(s, 0) ds$$

$$= \sum_{j=1}^{m} w_j I_{[Z_i, \infty)}(Z_j) \delta_0(d_j).$$

From this and from the fact that $I_{[Z_{i+1}, \infty)}(Z_j) = I_{(Z_i, \infty)}(Z_j)$ one can easily derive (S.2.10). ∎

To sum up, the lower and upper posterior distributions of $S(t)$ can be numerically approximated by drawing a large number of samples $L^{(k)}(t, d)$ from a DP $L(t, d)$ with base measure $\alpha_L = s\delta_{(X_0, 1)} + \sum_{i=1}^{n} \delta_{(Z_i, d_i)}$ where, from Theorem 1, one has $X_0 = t$ for the lower and $X_0 > t$ for the upper. From Lemma S.2 this corresponds to sampling the vector of weights $(w_0, \ldots, w_n)$ from the Dirichlet distribution with parameters $(s, \overbrace{1, \ldots, 1}^{n})$ and computing from (S.2.10) the value $S^{(k)}(t)$ corresponding to each sample $L^{(k)}(t)$. The empirical distribution of the samples $S^{(k)}(t)$ when $X_0 = t$ (respectively, $X_0 > t$) numerically approximates the lower (respectively, the upper) distribution of $S(t)$.

3   Posterior upper and lower distributions of $P(X < X')$

In this section we complete the results about the posterior upper and lower expectations of $P(X < X')$ and $I_{P(X < X') > \frac{1}{2}}$ given in the paper and derive a procedure for the numerical approximation by Monte Carlo sampling of the posterior distribution of $P(X < X')$ at the priors $\overline{\mu}_1 \underline{\mu}_2$.

Theorem 4 of the paper states that the prior base measure $\alpha_{20} = s \sum_{j=0}^{n'} \pi'_j \delta_{Z'^+_j}$ that gives the lower probability for $P(X < X') > \frac{1}{2}$ depends on the relative position of the data in $\mathbf{Z}_n$ and $\mathbf{Z}'_{n'}$ and has to be found each time by minimizing $\mathcal{E}_{\pi'} \left[I_{\{P(X < X') > \frac{1}{2}\}} \middle| \mathbf{Z}_n, \mathbf{Z}'_{n'} \right]$. This minimization is computationally expensive since for every vector of weights π' the value of the target function $\mathcal{E}_{\mu} \left[I_{\{P(X < X') > \frac{1}{2}\}} \middle| \mathbf{Z}_n, \mathbf{Z}'_{n'} \right]$ has to be computed by Monte Carlo sampling (details about how this computation can be carried out are given below) – notice that the number of Monte Carlo samples needed for a precision of about 10^{-3} is 10^6. An alternative solution is to use the bounds given in Theorem 5, which are very fast to compute and robust, as they always encompass the true extreme values of $\mathcal{E}_{\mu} \left[I_{\{P(X < X') > \frac{1}{2}\}} \middle| \mathbf{Z}_n, \mathbf{Z}'_{n'} \right]$. A different solution is to approximate the true lower value of the posterior probability of $P(X < X') > \frac{1}{2}$ by that obtained for the prior $\overline{\mu}_1, \underline{\mu}_2$ which can be computed quite efficiently. We will show in this section that, while the bounds given in Theorem 5 can be too conservative for large s, the probability obtained using $\overline{\mu}_1, \underline{\mu}_2$ always provides a good approximation. Before presenting these results, we provide an artificial example which shows that in case of censoring the probabilities obtained using the priors $\underline{\mu}_2$ and $\underline{\mu}_2^*$ can be different.

Example 3.1 Let us consider a sample with 20 deaths and no censored observations and a second sample with only two observations $(Z'_1, 0)$ and $(Z'_2, 1)$, with $Z_i < Z'_1 < Z_j$, $i = 1, \ldots, 5$, $j = 6, \ldots, 20$ and
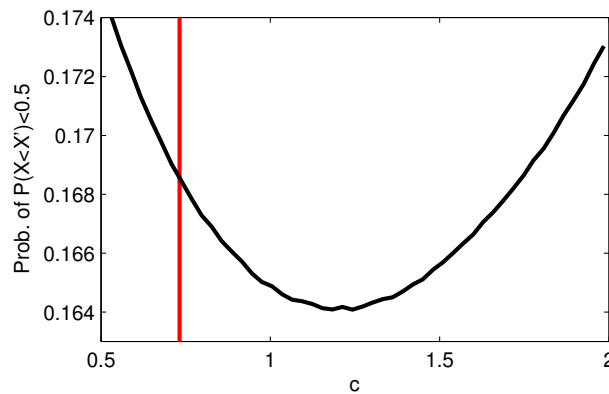
Figure 1   Probability that $P(X < X') > \frac{1}{2}$ for different values of c.

$Z'_2 > Z_i$, $i = 1, \ldots, 20$. Then, π and π' are vectors of two elements: $(s - c, c)$. The posterior expectation of $P(X < X')$ given $\overline{\mu}_1$ and π is

$$\frac{5}{20 + s} \frac{2 + c}{2 + s} + \frac{15}{20 + s} \frac{2 + c}{2 + s} \frac{1}{1 + c}.$$

If $s > \sqrt{3} - 1$ the lower is found for $c = \sqrt{3} - 1$. Samples of $P(X < X')$ are obtained from

$$w_{11}(w_{21} + w_{22} + w_{2c}) + w_{12}(w_{21} + w_{22} + w_{2c})\frac{w_{22}}{w_{22} + w_{2c}},$$

where $(w_{10}, w_{11}, w_{12}) \sim Dir(s, 5, 15)$ and $(w_{20}, w_{2c}, w_{21}, w_{22}) \sim Dir(s - c, c, 1, 1)$. We have evaluated the probability that $P(X < X') > \frac{1}{2}$ by Monte Carlo sampling for $s = 3$ and for different values of c. Results are shown in Figure 1. Notice that the lower value of the probability (minimum of the black curve) is found for c greater than $\sqrt{3} - 1$ (the vertical line in the figure), which is the value that minimizes the expectation of $P(X < X')$. However, the posterior probability of $P(X < X') > \frac{1}{2}$ obtained using $\underline{\mu}_2$ is 0.168, that obtained using $\underline{\mu}_2^*$ is 0.164 and so the difference is small.

To compare the two solutions proposed in different scenarios, we have simulated several datasets by sampling the survival and censoring times for the first populations from an exponential distribution with parameter $\lambda_1 = 1$. We have sampled the survival times for the second population from an exponential distribution with parameter $\lambda_2 = 1 + \Delta$ and included a single censored data so to be able to solve the minimization problem of Theorem 4 in a reasonable time. We have considered $\Delta = 0, 10$, $s = 0.25, 1, 5$ and $n = n' = 5, 10, 20$. For each case we have generated 100 samples. When the number of censored data is large, the minimization cannot be carried out in a reasonable time. Then, to evaluate the goodness of the approximation in more general cases, that is with more than 1 censored data, we have simulated large datasets and used the Gaussian approximation presented in Section 7 below, which avoids Monte Carlo sampling, to solve the minimization problem more efficiently. We have simulated 100 datasets for each case considered, sampling the survival and censoring times for both populations from exponential distributions with $\lambda_1 = 1$ and $\lambda_2 = 1 + \Delta$. We have considered $\Delta = 0, 0.5$ $s = 1, 5, 10$ and $n = n' = 50, 100, 200$. Results are shown in Tables 1 and 2. The lower bound given by Theorem 5 can considerably underestimate the true lower value of $\mathscr{E}_{\mu_1,\mu_2}\left[I_{\{P(X<X')>\frac{1}{2}\}} \big| \mathbf{Z}_n, \mathbf{Z}'_{n'} \right]$ for $s \geq 1$, while for $s = 0.25$ it is close to the true value. Moreover, if used to perform the hypothesis tests in Table 2 of the paper, the fraction of undecided cases never increases of more than 0.065 (the fraction of undecided cases in the four situation is SH-0.092, PH-0.248, EHD-0.206, LDH-0.178). On the other side, in all the situations considered the maximum distance between the probability that $P(X > X') > \frac{1}{2}$ obtained for $\underline{\mu}_2$ and for $\underline{\mu}_2^*$ is smaller than 0.003, which is

Table 1　Maximum value of the difference $\mathscr{E}_{\overline{\mu}_1,\underline{\mu}_2}\left[I_{\{P(X<X')>\frac{1}{2}\}}\middle|\mathbf{Z}_n,\mathbf{Z}'_{n'}\right] - \mathscr{E}_{\overline{\mu}_1,\underline{\mu}_2^*}\left[I_{\{P(X<X')>\frac{1}{2}\}}\middle|\mathbf{Z}_n,\mathbf{Z}'_{n'}\right]$. GA stands for Gaussian approximation.

$n=n'=$	5	10	20	$n=n'=$	50 (GA)	100 (GA)	200 (GA)
$s=0.25,\ \Delta=0$	0.002	0.001	0.002	$s=1,\ \Delta=0$	$0.43\ 10^{-4}$	$1.33\ 10^{-4}$	$3.68\ 10^{-4}$
$s=1,\ \Delta=0$	0.002	0.002	0.002	$s=5,\ \Delta=0$	$0.47\ 10^{-4}$	$1.34\ 10^{-4}$	$2.96\ 10^{-4}$
$s=5,\ \Delta=0$	0.001	0.001	0.001	$s=10,\ \Delta=0$	$0.182\ 10^{-4}$	$0.810\ 10^{-4}$	$3.369\ 10^{-4}$
$s=0.25,\ \Delta=10$	0.002	0.001	0.001	$s=1,\ \Delta=0.5$	$0.40\ 10^{-4}$	$1.35\ 10^{-4}$	$2.98\ 10^{-4}$
$s=1,\ \Delta=10$	0.002	0.001	0.001	$s=5,\ \Delta=0.5$	$0.50\ 10^{-4}$	$1.22\ 10^{-4}$	$2.81\ 10^{-4}$
$s=5,\ \Delta=10$	0.001	0.001	0.003	$s=10,\ \Delta=0.5$	$0.67\ 10^{-4}$	$6.10\ 10^{-4}$	$3.34\ 10^{-4}$

Table 2　Maximum difference between $\mathscr{E}_{\overline{\mu}_1,\underline{\mu}_2^*}\left[I_{\{P(X<X')>\frac{1}{2}\}}\middle|\mathbf{Z}_n,\mathbf{Z}'_{n'}\right]$ and the lower bound given in Theorem 5. GA stands for Gaussian approximation.

$n=n'=$	5	10	20	$n=n'=$	50 (GA)	100 (GA)	200 (GA)
$s=0.25,\ \Delta=0$	0.052	0.038	0.027	$s=1,\ \Delta=0$	0.126	0.109	0.082
$s=1,\ \Delta=0$	0.154	0.138	0.096	$s=5,\ \Delta=0$	0.364	0.324	0.249
$s=5,\ \Delta=0$	0.013	0.061	0.135	$s=10,\ \Delta=0$	0.040	0.243	0.377
$s=0.25,\ \Delta=10$	0.044	0.030	0.018	$s=1,\ \Delta=0.5$	0.143	0.097	0.050
$s=1,\ \Delta=10$	0.029	0.250	0.357	$s=5,\ \Delta=0.5$	0.373	0.338	0.228
$s=5,\ \Delta=10$	0.151	0.110	0.074	$s=10,\ \Delta=0.5$	0.427	0.473	0.422

very close to the precision with which this probability has been estimated (a slightly larger difference of 0.004 was found only in the unrealistic example 3.1). These results and those presented in the simulation of Section 3 of the paper, suggest that the probability obtained for the prior $\underline{\mu}_2$ is a good approximation of the true lower value of $\mathscr{E}_{\mu_1,\mu_2}\left[I_{\{P(X<X')>\frac{1}{2}\}}\middle|\mathbf{Z}_n,\mathbf{Z}'_{n'}\right]$ and that its use does not affect the robustness of the test. In the rest of this section we explain how to compute the posterior probability of $P(X<X')>\frac{1}{2}$ for the priors $\overline{\mu}_1,\underline{\mu}_2$.

Theorem 3 of the paper states that the prior distribution $\alpha_{20}=s\sum_{j=0}^{n'}\pi_j\delta_{Z'_j+}$ that gives the lower expectations for $P(X<X')$ depends on the relative position of the data in $\mathbf{Z}_n$ and $\mathbf{Z}'_{n'}$ and has to be found each time by minimizing (15) which can be written as

$$\mathscr{E}_{\overline{\mu}_1,\underline{\mu}_2}[P(X<X')|\mathbf{Z}_n,\mathbf{Z}'_{n'}] = \sum_{i=1}^{n^d}\left(\Delta S_1(\tilde{X}_i|\mathbf{Z}_n)\prod_{Z'_j\leq\tilde{X}_i}\frac{\Pi_j+n'_j-\tilde{d}'_j}{\Pi_j+n'_j}\right), \qquad (S.3.11)$$

where $\Delta S_1(\tilde{X}_i|\mathbf{Z}_n) = \overline{\mathscr{E}}[S_1(\tilde{X}_{i-1})|\mathbf{Z}_n] - \overline{\mathscr{E}}[S_1(\tilde{X}_i)|\mathbf{Z}_n]$, $\Pi_j = s\sum_{k=j}^{n'}\pi_k$, $\tilde{d}'_j = d'_j+s\pi_j$. By writing $\Pi_j+n'_j-\tilde{d}'_j = \Pi_{j+1}+n'_{j+1}+(1-d'_j)$ we obtain

$$\prod_{Z'_j\leq\tilde{X}_i}\frac{\Pi_j+n'_j-\tilde{d}'_j}{\Pi_j+n'_j} = \frac{\Pi_1+n'_1+1-d'_0}{\Pi_0+n'_0}\frac{\Pi_2+n'_2+1-d'_1}{\Pi_1+n'_1}\cdots\frac{\Pi_{l+1}+n'_{l+1}+1-d'_l}{\Pi_l+n'_l}$$

$$= \frac{\Pi_{l+1}+n'_{l+1}+1-d'_l}{s+n'}\prod_{j=1}^{l}\frac{\Pi_j+n'_j+1-d'_{j-1}}{\Pi_j+n'_j},$$

where l is the index of the largest observation such that $Z'_l \leq \tilde{X}_i$ and where we have used the fact that $\Pi_0 = s$ and $n'_0 = n'$. Then, by the transformation $\Pi_i + n'_i = e^{u_i}$ we obtain the convex function

$$\frac{e^{u_{l+1}} + 1 - d'_l}{s + n'} \prod_{j=1}^{l} \left(1 + (1 - d'_{j-1})e^{u_j}\right),$$

and the problem of minimizing (S.3.11) can be transformed into an equivalent convex minimization problem which makes the traditional tools of convex optimization applicable.

Once the vector of weights π is obtained, the posterior distributions $S_1(t)|\mathbf{Z}_n$ and $S_2(t)|\mathbf{Z}'_{n'}$ corresponding to the DP priors with (discrete) measures α_{10} and α_{20} given by Theorem 3 can be numerically computed by Monte Carlo sampling, based on the results of Section 2 of this supplementary material. In particular,

- samples for $S_1(t)$ are obtained by sampling the vectors of weights $W_1 = (w_{1,0}, w_{1,1}, \ldots, w_{1,n})$ from $Dir(s, \overbrace{1, \ldots, 1}^{n})$; $S_1^{(k)}(t)$ is computed based on (S.2.10) as

$$S_1^{(k)}(t) = \prod_{\tilde{X}_i \leq t} \frac{\sum_{j=i+1}^{n} w_{1,j} + w_{1,0}}{\sum_{j=i}^{n} w_{1,j} + w_{1,0}}, \tag{S.3.12}$$

where the weight $w_{1,0}$ is always included in the summation since $Z_0 > \max\{Z_j\}$;

- samples for $S_2(t)$ are obtained by sampling the vectors of weights $W_2 = (w_{2,-n'}, \ldots, w_{2,-1}, w_{2,0}, w_{2,1}, \ldots, w_{2,n'})$ from $Dir(s\pi_{n'}, \ldots, s\pi_1, s\pi_0, \overbrace{1, \ldots, 1}^{n'})$; the sample $S_2^{(k)}(t)$ is computed as

$$S_2^{(k)}(t) = \prod_{Z'_i \leq t} \frac{\sum_{j=i}^{n'}(w_{2,-j} + w_{2,j}) - w_{2,-i} - w_{2,i}\delta_1(d'_i)}{\sum_{j=i}^{n'}(w_{2,-j} + w_{2,j})}. \tag{S.3.13}$$

Equation (S.3.13) can be better understood by interpreting the prior α_{20} as $n' + 1$ fictitious deaths falling just after each observation and just after $Z'_0 = 0$. For each fictitious death we sample the weights $w_{2,0}, w_{2,-1}, \ldots, w_{2,n'}$ that play the same role as the weights associated to observed deaths in the computation of $S_2^{(k)}(t)$.

The samples $S_1^{(k)}(t)$ and $S_2^{(k)}(t)$ are step functions; $S_1^{(k)}(t)$ has jumps at the n^d deaths $\tilde{X}_1 \ldots, \tilde{X}_{n^d}$, whereas $S_2^{(k)}(t)$ has jumps at all observations $Z_1, \ldots, Z_{n'}$ and $Z_0 = 0$. Then, we can write:

$$S_1^{(k)}(t) = 1 - \sum_{i=1}^{n^d} \overline{w}_i I_{[0,t]}(\tilde{X}_i), \quad \overline{w}_i = S_1^{(k)}(\tilde{X}_{i-1}) - S_1^{(k)}(\tilde{X}_i),$$

$$S_2^{(k)}(t) = 1 - \sum_{i=0}^{n'} \underline{w}_i I_{[0,t]}(Z'_i), \quad \underline{w}_i = S_2^{(k)}(Z'_{i-1}) - S_2^{(k)}(Z'_i).$$

Finally, the posterior of $P(X < X')|\mathbf{Z}_n, \mathbf{Z}'_{n'}$ for the priors $\overline{\mu}_1, \underline{\mu}_2$ can be approximated based on the following theorem.

Theorem S.1. *For any $a \in [0,1]$ it holds that*

$$\mathscr{E}_{\overline{\mu}_1, \underline{\mu}_2}\left[I_{P(X<X')>a}|\mathbf{Z}_n, \mathbf{Z}'_{n'}\right] = P\left[g(W_1, W_2, \mathbf{Z}_n, \mathbf{Z}'_{n'}) > a\right], \tag{S.3.14}$$

where P is the probability with respect to the (Dirichlet) distribution of $W_1 = (w_{1,0}, w_{1,1}, \ldots, w_{1,n})$ and $W_2 = (w_{2,-n'}, \ldots, w_{2,-1}, w_{2,0}, w_{2,1}, \ldots, w_{2,n'})$ and

$$g(W_1, W_2, \mathbf{Z}_n, \mathbf{Z}'_{n'}) = \sum_{i=1}^{n^d} \sum_{j=0}^{n'} \overline{w}_i \underline{w}_j I_{(Z'_j, \infty)}(\tilde{X}_i), \tag{S.3.15}$$

■

Proof. *The posterior expectation of $I_{\{P(X<X')>a\}}$ is equal to the probability that $\int(1-S_1^{(k)}(t))d[-S_2^{(k)}(t)] > a$, and thus we have:*

$$\mathscr{E}_{\overline{\mu}_1,\underline{\mu}_2}\big[I_{P(X<X')>a}|\mathbf{Z}_n,\mathbf{Z}'_{n'}\big] = P\big[\int(1-S_1^{(k)}(t))d[-S_2^{(k)}(t)] > a|\mathbf{Z}_n,\mathbf{Z}'_{n'}\big] =$$

$$P\left[\int\sum_{i=1}^{n^d}\overline{w}_iI_{[0,t]}(\tilde{X}_i)\left(\sum_{j=0}^{n'}\underline{w}_j\delta_{Z'_j}(t)\right)dt > a\right] = P\left[\sum_{i=1}^{n^d}\sum_{j=0}^{n'}\overline{w}_i\underline{w}_jI_{[0,Z'_j]}(\tilde{X}_i) > a\right].$$

■

Similar results can be derived for the upper distribution of $P(X < X')$ by inverting the roles of X and X' based on the fact that the upper bound for $\mathscr{E}_{\mu_1,\mu_2}[I_{P(X<X')>0.5}|\mathbf{Z}_n,\mathbf{Z}'_{n'}]$ corresponds to the lower bound of $\mathscr{E}_{\mu_1,\mu_2}[I_{P(X'\leq X)>0.5}|\mathbf{Z}_n,\mathbf{Z}'_{n'}]$ since

$$\overline{\mathscr{E}}[I_{P(X<X')>a}|\mathbf{Z}_n,\mathbf{Z}'_{n'}] = \overline{\mathscr{E}}[1-I_{P(X\geq X')>a}|\mathbf{Z}_n,\mathbf{Z}'_{n'}] = 1-\underline{\mathscr{E}}[I_{P(X'\leq X)>a}|\mathbf{Z}_n,\mathbf{Z}'_{n'}].$$

In practice the inequality $X' < X$ can be considered instead of $X' \leq X$ having assumed that there are no ties. Thus, by Monte Carlo sampling the vectors $(w_{1,0},w_{1,1},\ldots,w_{1,n})$ and $(w'_{2,-n'},\ldots,w'_{2,-1},w'_{2,0},w'_{2,1},\ldots,w'_{2,n'})$, it is possible to approximate the distributions of $P(X < X')$ at $\overline{\mu}_1,\underline{\mu}_2$ and $\underline{\mu}_1,\overline{\mu}_2$ and from that, approximate $\underline{\mathscr{E}}[I_{P(X<X')>0.5}|\mathbf{Z}_n,\mathbf{Z}'_{n'}]$ and $\overline{\mathscr{E}}[I_{P(X<X')>0.5}|\mathbf{Z}_n,\mathbf{Z}'_{n'}]$.

4   Choice of the prior strength *s*

The value of *s* determines how quickly lower and upper posterior expectations converge at the increase of the number of observations. A way to select a value of *s* is by imposing that the degree of imprecision, measured by the difference between the upper and lower expectations of the functions of interest, is reduced to a fraction of its prior value after one observation. Imposing a degree of imprecision close to 1 after the first observation increases the width of the credible intervals in the survival curves estimation or the probability of an indeterminate outcome in the IDP test, whereas a value close to 0 makes the outcomes less reliable (in fact, in Section 3 of the paper the limiting value of 0 has produced unreliable estimates of credible intervals for the survival probability). The intermediate value of $1/2$ is a frequent choice in prior-ignorance modeling (Pericchi and Walley, 1991; Walley, 1996). In figure 2 we show the imprecision obtained for the expectation of $S(t)$ (the function of interest in survival curve estimation) and $I_{P(X<X')>0.5}$ (the function of interest in the IDP rank sum test) after 1 observed death as a function of *s*. For values of *s* smaller than 1 the two curves are very close, but the imprecision for the IDP test increases faster with *s*. A reduction of the imprecision to 1/2 of its initial value is obtained for $s \approx 1$. Notice that, in case of censoring, the imprecision is augmented by the presence of censored data. Thus, to limit the number of indeterminate decision in the IDP rank sum test we have adopted a less conservative choice by empirically searching for a value of *s* which assured always a Type I error lower that γ while keeping the fraction of undecided cases satisfactorily low. We adopted the value $s = 0.25$ (see Section 3 in the paper) which corresponds to a reduction of the indeterminacy to about 1/5 of its initial value after 1 observed death. Although these are subjective ways of choosing the degree of conservativeness (indeterminacy), they have a clear interpretation as the number of fictitious additional deaths that we are considering. For example, in the IDP test we are implicitly considering the event that *s* fictitious deaths are added to each group in the most unfavorable position for the hypothesis under test. In this view, a value $s > 1$ would be already a very
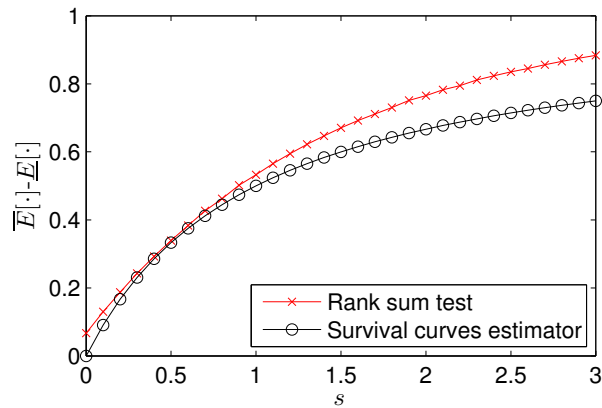
Figure 2 Difference between the upper and lower posterior expectation of the function of interest for the IDP rank sum test ($I_{P(X<X')>0.5}$) and the survival curve estimation ($S(t)$).

conservative choice, whereas the choice $s = 0.25$ is less robust, since one single additional observation could change the decision of the test.

A different approach that can be used when a sharp decision is not required is to perform the test for $s \to 0$ and compute the value s_{max} of s after which the decision becomes indeterminate, as shown for the AIDS dataset. When the IDP test with $s \to 0$ is not indeterminate, this provides the decision that would be taken at the lowest degree of conservativeness and at the same time a measure of its robustness.

4.1 Managing ties $X_i = X_j'$

In the case of ties between samples from the two populations ($X_i = X_j'$), the common approach is to test the hypothesis $[P(X < X') + \frac{1}{2}P(X = X')] \leq 0.5$ against $[P(X < X') + \frac{1}{2}P(X = X')] > 0.5$. Since

$$P(X < X') + \frac{1}{2}P(X = X') = E\left[I_{(X,\infty)}(X') + \frac{1}{2}I_{\{X\}}(X')\right] = E[\mathrm{H}(X' - X)],$$

where $\mathrm{H}(\cdot)$ denotes the Heaviside step function, i.e., $\mathrm{H}(t) = 1$ for $t > 0$, $\mathrm{H}(t) = 0.5$ for $t = 0$ and $\mathrm{H}(t) = 0$ for $t < 0$, in case of ties the results presented in this Section are still valid if we substitute $I_{(X_j,\infty)}(X_k')$ with $\mathrm{H}(X_k' - X_j)$ in function g of Theorem S.1 and in matrix A of Theorem S.2 in Section 7.

5 Details of the simulations

We have applied the IDP test to a number of artificial datasets simulating four scenarios (same hazard (SH), proportional hazard (PH), early hazard difference (EHD) and late hazard difference (LHD)) considered by Letòn and Zuluaga (2005) to study the performance of several rank tests for censored lifetime data. For this, the lifetime and follow up data of a large number of samples are generated. The four situations (SH, PH, EHD and LHD) are simulated by sampling $n = n' = 30$ lifetimes X_i and X_i' with hazard function $\lambda_1(t)$ and $\lambda_2(t)$ defined by the step functions

- SH: $\lambda_1(t) = \lambda_2(t) = 1$;

- PH: $\lambda_1(t) = 2, \lambda_2(t) = 1$;

- EHD: $\lambda_1(t) = 3I_{[0,0.4)}(t) + 0.75I_{[0.4,0.6)}(t) + 1I_{[0.6,\infty)}(t),$

  $\lambda_2(t) = 0.75I_{[0,0.4)}(t) + 3I_{[0.4,0.6)}(t) + 1I_{[0.6,\infty)}(t);$

- LHD: $\lambda_1(t) = 2I_{[0,0.4)}(t) + 4I_{[0.4,\infty)}(t),$ $\lambda_2(t) = 2I_{[0,0.4)}(t) + 0.4I_{[0.4,\infty)}(t).$

The censoring times are sampled from a uniform distribution between 0 and 2 ($U[0,2]$). The type-I error of the test for SH and the power of the test for PH, EHD and LHD are evaluated over 1000 random repetitions. All tests are performed at the level $\gamma = 0.05$.

When simulating the scenarios A and B of Example 3.2, lifetime data are sampled from two populations with hazard functions:

- Case A: $\lambda_1(t) = 0.75I_{[0,0.4)}(t) + 5I_{[0.4,\infty)}(t),$ $\lambda_2(t) = 2I_{[0,0.4)}(t) + 0.25I_{[0.4,\infty)}(t);$

- Case B: $\lambda_1(t) = 2I_{[0,0.4)}(t) + 0.2I_{[0.4,\infty)}(t),$ $\lambda_2(t) = 0.75I_{[0,0.4)}(t) + 5I_{[0.4,\infty)}(t).$

The censoring times are sampled from $U[0,2]$. The type-I error is evaluated for $\gamma = 0.05$ and $\gamma = 0.01$ over 1000 random repetitions.

6   Proofs

Proof. of Theorem 1 *The first term in (6),*

$$\frac{\alpha(t,+\infty) + n_t^+ + j}{s + n + j},$$

is minimized when $\alpha(t,+\infty) = 0$. Moreover, since the derivative with respect to $\alpha[Z_i,+\infty)$ of the other terms in (6),

$$\frac{\alpha[Z_i,+\infty) + n_i + j}{\alpha[Z_i,+\infty) + n_i^+ + j}, \tag{S.6.16}$$

is non-positive, their minimum is found by maximizing $\alpha[Z_i,+\infty)$, i.e., at $\alpha[Z_i,+\infty) = s$. These conditions are satisfied by a prior which concentrates all the probability in t, i.e., $\alpha^* = \delta_t$. The lower moments in (9) are derived by using $\alpha = s\delta_t$ in (6).

Similarly, the maximum value of the terms in (S.6.16) is found by minimizing $\alpha[Z_i,+\infty)$. Since $\alpha[Z_1,+\infty) \geq \alpha[Z_2,+\infty) \geq \cdots \geq \alpha(t,+\infty)$, the minimum of all terms $\alpha[Z_i,+\infty)$ is $\alpha[Z_i,+\infty) = \alpha(t,+\infty) = \alpha_t$, for any given value of α_t. Then, we can write the product in (6) as:

$$\prod_{j=0}^{p-1} \frac{\alpha_t + n_{i_1} + j}{s + n + j} \frac{\alpha_t + n_t^+ + j}{\alpha_t + n_{i_c}^+ + j} \prod_{k=2}^{c} \frac{\alpha_t + n_{i_k} + j}{\alpha_t + n_{i_{k-1}} + j}, \tag{S.6.17}$$

where $i_1,\ldots,i_c < l$ are the indexes of the observations censored before t. Each term of the product in (S.6.17) has non-negative derivative with respect to α_t, so the maximum is found at $\alpha_t = s$. These conditions are satisfied by a prior which concentrates all the probability after t, as for example $G = \delta_{X_0}$, with $X_0 > t$. The upper moments in (9) are derived by using $\alpha = s\delta_{X_0}$ in (6). ∎

Proof. of Corollary 1 *We need to prove that $\underline{\mathscr{E}}[S(t)|\mathbf{Z}_n] \leq \hat{S}(t) \leq \overline{\mathscr{E}}[S(t)|\mathbf{Z}_n]$, where $\hat{S}(t)$ is the Kaplan-Meier estimator. The inequality for the upper bound follows easily from the fact that all the terms $\frac{s + n_i^+}{s + n_i}$*

of the upper are larger than the terms $\frac{n_i^+}{n_i}$ of the KM estimator. To prove the inequality for the lower bound we need to rewrite $\hat{S}(t)$ and $\underline{\mathscr{E}}[S(t)|\mathbf{Z}_n]$ as follows:

$$\hat{S}(t) = \prod_{\tilde{X}_i \leq t} \frac{n_i^+}{n_i} = \prod_{Z_i \leq t} \frac{n_i^+}{n_i} \prod_{\tilde{Y}_i \leq t} \frac{n_i}{n_i^+} = \frac{n_t^+}{n} \prod_{\tilde{Y}_i \leq t} \frac{n_i}{n_i^+};$$

$$\underline{\mathscr{E}}[S(t)|\mathbf{Z}_n] = \frac{n_t^+}{s+n_t} \prod_{\tilde{X}_i \leq t} \frac{s+n_i^+}{s+n_i} = \frac{n_t^+}{s+n_t} \prod_{Z_i \leq t} \frac{s+n_i^+}{s+n_i} \prod_{\tilde{Y}_i \leq t} \frac{s+n_i}{s+n_i^+}$$

$$= \frac{n_t^+}{s+n_t} \frac{s+n_t}{s+n} \prod_{\tilde{Y}_i \leq t} \frac{s+n_i}{s+n_i^+} = \frac{n_t^+}{s+n} \prod_{\tilde{Y}_i \leq t} \frac{s+n_i}{s+n_i^+}.$$

Now we can see that all terms of $\underline{\mathscr{E}}[S(t)|\mathbf{Z}_n]$ are smaller than the corresponding terms of $\hat{S}(t)$, i.e. $\frac{n_t^+}{s+n} <$ $\frac{n_t^+}{n}$ and $\frac{s+n_i}{s+n_i^+} < \frac{n_i}{n_i^+}$, for all $i = 1, \ldots, n$. ∎

Proof. of Theorem 2 *A sample from the posterior distribution of $L(t,d)$ can be written as*

$$L^{(k)}(t,d) = w_0 P^{(k)}(t) \delta_1(d) + \sum_{i=1}^n w_i \delta_{Z_i, d_i}(t,d),$$

where $(w_0, w_1, \ldots, w_n) \sim Dir(s, 1, \ldots, 1)$ and $P^{(k)}(t) \sim \mathrm{Dp}(s, \alpha_s^*(t))$. Notice that any sample $P^{(k)}$ from a Dirichlet process has a discrete distribution. Then, from Lemma S.3 the sample $S^{(k)}(t)$ corresponding to $L^{(k)}(t,d)$ can be written as the product

$$S^{(k)}(t) = \prod_{\{Z_i^* \leq t\}} \frac{S^{*(k)}(Z_{i+1}^*) + Q^{*(k)}(Z_i^*)}{S^{*(k)}(Z_i^*) + Q^{*(k)}(Z_i^*)} = \prod_{\{i:d_i=1, Z_i^* \leq t\}} \frac{w_0 P(Z_i^*, \infty) + \sum_{j=1}^n w_j I_{(Z_i^*, \infty)}(Z_j)}{w_0 P[Z_i^*, \infty) + \sum_{j=1}^n w_j I_{[Z_i^*, \infty)}(Z_j)}. \quad \text{(S.6.18)}$$

where $S^{*(k)}$ and $Q^{*(k)}$ are the subsurvival functions of the sample distribution $L^{(k)}$ and $Z_i^*(t)$ are all the points of discontinuity of $L^{(k)}$. By the same reasoning used in the proof of Corollary 1 we can rewrite (S.6.18) as

$$S^{(k)}(t) = \left[w_0 P^{(k)}(t, \infty) + \sum_{j=1}^n w_j I_{(t, \infty)}(Z_j) \right] \prod_{\{\tilde{Y}_i \leq t\}} \frac{w_0 P^{(k)}[\tilde{Y}_i, \infty) + \sum_{j=1}^n w_j I_{[\tilde{Y}_i, \infty)}(Z_j)}{w_0 P^{(k)}[\tilde{Y}_i, \infty) + \sum_{j=1}^n w_j I_{(\tilde{Y}_i, \infty)}(Z_j)}.$$

This equation has the same form as (6) with $p = 1$ and with $w_0 P^{(k)}$, $\sum_{j=1}^n w_j I_{[t, \infty)}(Z_j)$ and $\sum_{j=1}^n w_j I_{(t, \infty)}(Z_j)$ instead of α, n_t and n_t^+. Then, the same reasoning used in the proof of Theorem 1 to prove that $\alpha^* = \delta_t$ and $\alpha^* = \delta_{X_0}$ (with $X_0 > t$) give the lower and upper bounds of (6) can be used here to show that $P^{(k)} = \delta_t$ and $P^{(k)} = \delta_{X_0}$ give lower and upper bounds for $S^{(k)}(t)$. Thus, we have

$$\begin{aligned} \mathscr{E}_\mu[I_{S(t) \geq a}|\mathbf{Z}_n] \quad &= \mathscr{E}_{(w_0, w_1, \ldots, w_n), P^{(k)}}[I_{S^{(k)}(t) \geq a}|\mathbf{Z}_n] \\ &\geq \mathscr{E}_{(w_0, w_1, \ldots, w_n)}[I_{S^{(k)}(t) \geq a}|\mathbf{Z}_n, P^{(k)} = \delta_t] \\ &= \mathscr{E}_{\underline{\mu}}[I_{S(t) \geq a}|\mathbf{Z}_n]. \end{aligned}$$

and

$$\begin{aligned} \mathscr{E}_\mu[I_{S(t) \leq a}|\mathbf{Z}_n] \quad &= \mathscr{E}_{(w_0, w_1, \ldots, w_n), P^{(k)}}[I_{S^{(k)}(t) \leq a}|\mathbf{Z}_n] \\ &\geq \mathscr{E}_{(w_0, w_1, \ldots, w_n)}[I_{S^{(k)}(t) \leq a}|\mathbf{Z}_n, P^{(k)} = \delta_{X_0}] \\ &= \mathscr{E}_{\overline{\mu}}[I_{S(t) \leq a}|\mathbf{Z}_n]. \end{aligned}$$

■

Proof. of Theorem 3 *We need to prove that the lower expectation of $\int [1 - S_1(t)] d[-S_2(t)]$ is found using the priors in (14). The other results follow straightforwardly.*

First, notice that both terms $[1 - S_1(t)]$ and $d[-S_2(t)]$ are positive and independent and that the first term is minimized at the maximum of $S_1(t)$ which, from Theorem 1, is obtained at all observations in $\mathbf{Z}_n$ and $\mathbf{Z}'_{n'}$ using the prior $\overline{\mu}_1 = \mathrm{Dp}(s, \alpha^*_{10})$ in (14). Since $P(X < X') = 1 - P(X' \le X)$ the integral of interest can be rewritten as

$$\int [1 - S_1(t)] d[-S_2(t)] = \int S_2(t) d[-S_1(t)]$$

and thus

$$\mathscr{E}_{\overline{\mu}_1,\mu_2}\left[P(X < X')|\mathbf{Z}_n, \mathbf{Z}'_{n'}\right] = \mathscr{E}_{\overline{\mu}_1,\mu_2}\left[\int S_2(t) d[-S_1(t)] \,\bigg|\, \mathbf{Z}_n, \mathbf{Z}'_{n'}\right] = \int \mathscr{E}_{\mu_2}\left[S_2(t)|\mathbf{Z}'_{n'}\right] d\mathscr{E}_{\overline{\mu}_1}[-S_1(t)|\mathbf{Z}_n].$$

Using the fact that $\mathscr{E}_{\overline{\mu}_1}[S_1(t)|\mathbf{Z}_n]$ is as step function with jumps at the death times $\tilde{X}_1, \ldots, \tilde{X}_{n^d}$, we can write

$$d\mathscr{E}_{\overline{\mu}_1}[-S_1(t)|\mathbf{Z}_n] = \sum_{i=1}^{n^d} \left(\mathscr{E}_{\overline{\mu}_1}[S_1(\tilde{X}_{i-1})|\mathbf{Z}_n] - \mathscr{E}_{\overline{\mu}_1}[S_1(\tilde{X}_i)|\mathbf{Z}_n]\right) \delta_{\tilde{X}_i}(t) dt = \sum_{i=1}^{n^d} \Delta S_1(\tilde{X}_i|\mathbf{Z}_n) \delta_{\tilde{X}_i}(t) dt$$

where $\Delta S_1(\tilde{X}_i|\mathbf{Z}_n)$ is the size of each jump. Then, we obtain

$$\mathscr{E}_{\overline{\mu}_1,\mu_2}\left[P(X < X')|\mathbf{Z}_n, \mathbf{Z}'_{n'}\right] = \sum_{i=1}^{n^d} \Delta S_1(\tilde{X}_i|\mathbf{Z}_n) \mathscr{E}_{\mu_2}[S_2(\tilde{X}_i)|\mathbf{Z}'_{n'}]. \tag{S.6.19}$$

From (6) we have that

$$\mathscr{E}_{\mu_2}[S_2(\tilde{X}_i)|\mathbf{Z}'_{n'}] = \frac{\alpha_{20}(\tilde{X}_i, +\infty) + n'^+_{\tilde{X}_i}}{n' + s} \prod_{\tilde{Y}'_i \le \tilde{X}_i} \frac{\alpha_{20}[\tilde{Y}'_i, +\infty) + n'_i}{\alpha_{20}[\tilde{Y}'_i, +\infty) + n'^+_i}.$$

By defining $\pi'_j = \alpha_{20}[\tilde{Y}'_j, \tilde{Y}'_{j+1})$, with $j = 0, \ldots, n'^c - 1$ and $\tilde{Y}'_0 = 0$, and $\pi'_{n'^c} = \alpha_{20}[\tilde{Y}'_{n'^c}, \infty)$, we can write $\alpha_{20}[\tilde{Y}'_j, +\infty) = \sum_{j=k}^{n'^c} \pi'_j$. Then we have

$$\prod_{\tilde{Y}'_j \le t} \frac{\alpha_{20}[\tilde{Y}'_j, +\infty) + n'_j}{\alpha_{20}[\tilde{Y}'_j, +\infty) + n'^+_j} = f(\pi', t)$$

and

$$\mathscr{E}_{\overline{\mu}_1,\mu_2}\left[P(X < X')|\mathbf{Z}_n, \mathbf{Z}'_{n'}\right] = \sum_{i=1}^{n^d} \Delta S_1(\tilde{X}_i|\mathbf{Z}_n) \frac{\alpha_{20}(\tilde{X}_i, +\infty) + n'^+_{\tilde{X}_i}}{n' + s} f(\pi', \tilde{X}_i). \tag{S.6.20}$$

The vector π fixes the mass assigned by the prior α_{20} to the intervals $[\tilde{Y}'_j, \tilde{Y}'_{j+1})$; we are left free to decide how this mass is distributed inside each interval. Given a vector π, the minimum of (S.6.20) is found by minimizing all terms

$$\alpha_{20}(\tilde{X}_i, +\infty) = \alpha_{20}(\tilde{X}_i, \tilde{Y}'_{h_i}) + \alpha_{20}[\tilde{Y}'_{h_i}, \infty),$$

where h_i is the index of the smallest censored observation $\tilde{Y}'_j > \tilde{X}_i$ (so that $\tilde{X}_i \in [\tilde{Y}'_{h_i-1}, \tilde{Y}'_{h_i})$). Then, the minimum of $\alpha_{20}(\tilde{X}_i, +\infty)$ is obtained for $\alpha_{20}(\tilde{X}_i, \tilde{Y}'_{h_i}) = 0$. This condition can be satisfied by concentrating the mass π'_{h_i-1}, assigned by α_{20} to the interval $[\tilde{Y}'_{h_i-1}, \tilde{Y}'_{h_i})$, in the sub-interval $(\tilde{Y}'_{h_i-1} \tilde{X}_i]$. This means that the mass assigned to any interval $[\tilde{Y}'_j, \tilde{Y}'_{j+1})$, $j = 0, \ldots, n'^c$ has to lie before any possible death $\tilde{X}_i$ falling in that interval. A solution is to choose a prior that concentrates all the mass π_j at $\tilde{Y}'_j$. However, since we have assumed in Lemma S.3 that there are no common discontinuities in $S_2^*(t)$ and $Q_2^*(t)$, we make the mass concentrate at $\tilde{Y}'^+_j$, with $\tilde{Y}'_j < \tilde{Y}'^+_j < \min\{\tilde{X}_i : \tilde{Y}'_j < \tilde{X}_i\}$. This yields the DP prior with base measure α_{20} in (14). ∎

Proof. of Theorem 4 *A sample from the posterior distribution of $L_i(t,d)$, $i = 1, 2$ can be written as*

$$L_i^{(k)}(t,d) = w_{i0}P_i^{(k)}(t)\delta_1(d) + \sum_{j=1}^{n} w_{ij}\delta_{Z_j, d_j}(t,d),$$

where $(w_{i0}, w_{i1}, \ldots, w_{in}) \sim Dir(s, 1, \ldots, 1)$, $P_i^{(k)}(t) \sim Dp(s, \alpha_s^*(t))$. The corresponding sample $S_i^{(k)}(t)$ is

$$S_i^{(k)}(t) = \left[w_{i0}P_i^{(k)}(t, \infty) + \sum_{j=1}^{n} w_{ij}I_{(t,\infty)}(Z_j) \right] \prod_{\{\tilde{Y}_j \leq t\}} \frac{w_{i0}P_i^{(k)}[\tilde{Y}_j, \infty) + \sum_{l=1}^{n} w_{il}I_{[\tilde{Y}_j, \infty)}(Z_l)}{w_{i0}P_i^{(k)}[\tilde{Y}_j, \infty) + \sum_{k=1}^{n} w_{il}I_{(\tilde{Y}_j, \infty)}(Z_l)}$$

and the corresponding sample for the probability $P(X < X')$ can be written as

$$P^{(k)}(X < X') = 1 - \int S_1^{(k)}(t)d[-S_2^{(k)}(t)] = 1 - \sum_{j=1}^{n'^*} \left[S_2^{(k)}(Z'^*_{j-1}) - S_2^{(k)}(Z'^*_j) \right] S_1^{(k)}(Z'^*_j), \quad \text{(S.6.21)}$$

where Z'^*_j are all the points of discontinuity of L_2 and n'^* its (possibly infinite) number. For any sample $S_2^{(k)}$, the same argument used in Theorem 1 to show that the lower expectation of $S_1(t)$ is found concentrating all prior mass at $X_0 > t$ can be used now to prove that $S_1^{(k)}(Z'^*_j)$, $j = 0, \ldots, n'^d$, is upper bounded by the value obtained for $P_1^{(k)} = \delta_{X_0}$, with $X_0 > Z'^*_j$ for all $j = 0, 1 \ldots$. Then $1 - \sum_{j=1}^{n'^*} \left[S_2^{(k)}(Z'^*_{j-1}) - S_2^{(k)}(Z'^*_j) \right] S_1^{(k)}(Z'^*_j)$ is lower bounded by the value obtained when $P_1^{(k)} = \delta_{X_0}$ and thus

$$\begin{aligned} \mathcal{E}_{\mu_1, \mu_2}[I_{P(X<X') \geq \frac{1}{2}}|\mathbf{Z}_n, \mathbf{Z}'_{n'}] &= \mathcal{E}_{w_1, P_1^{(k)}, \mu_2}[I_{P^{(k)}(X<X') \geq \frac{1}{2}}|\mathbf{Z}_n, \mathbf{Z}'_{n'}] \\ &\geq \mathcal{E}_{w_1, \mu_2}[I_{P^{(k)}(X<X') \geq \frac{1}{2}}|\mathbf{Z}_n, \mathbf{Z}'_{n'}, dP_1^{(k)} = \delta_{X_0}]. \end{aligned}$$

Samples $dP_1^{(k)} = \delta_{X_0}$ are obtained when $\alpha_{10}^* = \delta_{X_0}$, with $X_0 > Z'^*_j$.
For $\alpha_{10}^* = \delta_{X_0}$, the sample for the probability $P(X < X')$ can be written as

$$P^{(k)}(X < X') = \sum_{j=1}^{n^d} \left[S_1^{(k)}(Z_{j-1}) - S_1^{(k)}(Z_j) \right] S_2^{(k)}(Z_j). \quad \text{(S.6.22)}$$

This equation has the same form as (S.6.19) with $w_{20}P_2^{(k)}$, $\sum_{j=1}^{n} w_{ij}I_{(t,\infty)}(Z_j)$ and $\sum_{j=1}^{n} w_{ij}I_{[t,\infty)}(Z_j)$ instead of α_{20}, n_t^+ and n_t. Let $\pi'^{(k)}_j$ be the mass assigned by $P_2^{(k)}$ to the intervals $[\tilde{Y}'_j, \tilde{Y}'_{j+1})$, $j = 0, \ldots, n'^c$, the same reasoning used in Theorem 3 to show that the lower of $\mathcal{E}_{\overline{\mu}_1, \mu_2}[P(X < X')|\mathbf{Z}_n, \mathbf{Z}'_{n'}]$ is found when all the mass of μ_2 lies before any possible death $\tilde{X}_i$ falling in that interval, can be used here to prove that the lower

of $P^{(k)}(X < X')$ is found when $\pi_j'^{(k)}$ concentrates at $\tilde{Y}_j'^+$, with $\tilde{Y}_j' < \tilde{Y}_j'^+ < \min\{\tilde{X}_i : \tilde{Y}_j' < \tilde{X}_i\}$. Then,

$$
\begin{aligned}
\mathscr{E}_{\overline{\mu}_1, \mu_2}[I_{P(X < X') \geq \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}'_{n'}] & = \mathscr{E}_{\overline{\mu}_1, w_2, P_2^{(k)}}[I_{P^{(k)}(X < X') \geq \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}'_{n'}] \\
& \geq \mathscr{E}_{\overline{\mu}_1, w_2, (\pi_0'^{(k)}, \pi_1'^{(k)}, \dots, \pi_{n'^c}'^{(k)})}[I_{P^{(k)}(X < X') \geq \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}'_{n'}, dP_2^{(k)} = \sum_{i=0}^{n'^c} \pi_i'^{(k)} \delta_{\tilde{Y}_j'^+}].
\end{aligned}
$$

Samples $dP_2^{(k)} = \sum_{i=0}^{n'} \pi_i'^{(k)} \delta_{\tilde{Y}_j'^+}$ are obtained for a prior of the form $\alpha_{20}^{*(k)} = \sum_{i=0}^{n'} \pi_i' \delta_{\tilde{Y}_j'^+}$. ∎

Proof. of Theorem 5 *The fact that the lower bound for the posterior probability that $P(X > X') > \frac{1}{2}$ is found for $\overline{\mu}_1 = \mathrm{Dp}(s, \delta_{X_0})$ follows from Theorem 4. In the proof of Theorem 2 we have shown that, for a fixed t, a lower bound for $S_2^{(k)}(t)$, hereafter denoted by $\underline{S}_2^{(k)}(t)$, is found when $P_2^{(k)} = \delta_t$. It follows that*

$$
P^{(k)}(X < X') \geq \sum_{j=1}^{n^*} \left[S_1^{(k)}(Z_{j-1}^*) - S_1^{(k)}(Z_j^*) \right] \underline{S}_2^{(k)}(Z_j^*).
$$

Then

$$
\begin{aligned}
\mathscr{E}_{\mu_1, \mu_2}[I_{P(X < X') \geq \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}'_{n'}] & = \mathscr{E}_{w_1, P_1^{(k)}, w_2, P_2^{(k)}}[I_{P^{(k)}(X < X') \geq \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}'_{n'}] \\
& \geq E_{w_1, w_2}[I_{\sum_{j=1}^{n^*} \left[S_1^{(k)}(Z_{j-1}^*) - S_1^{(k)}(Z_j^*) \right] \underline{S}_2^{(k)}(Z_j^*) \geq \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}'_{n'}, dP_1^{(k)} = \delta_{X_0}].
\end{aligned}
$$

The proof for the upper bound follows from the fact that $\mathscr{E}_{\mu_1, \mu_2}[I_{P(X < X') \geq \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}'_{n'}] = 1 - \mathscr{E}_{\mu_1, \mu_2}[I_{P(X' < X) \geq \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}'_{n'}]$. Then,

$$
\mathscr{E}_{\mu_1, \mu_2}[I_{P(X' < X) \geq \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}'_{n'}] \geq \mathscr{E}\left[I_{\int S_{\underline{\mu}_t, 1}(t) dS_{\overline{\mu}_t, 2}(t) > \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}'_{n'} \right]
$$

implies

$$
\mathscr{E}_{\mu_1, \mu_2}[I_{P(X < X') \geq \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}'_{n'}] \leq 1 - \mathscr{E}\left[I_{\int S_{\underline{\mu}_t, 1}(t) dS_{\overline{\mu}_t, 2}(t) > \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}'_{n'} \right].
$$

∎

Proof. of Theorem 6 *In the absence of censoring, the samples for S_i, $i = 1, 2$ can be written as*

$$
S_i^{(k)}(t) = w_{i0} P_i^{(k)}(t, \infty) + \sum_{j=1}^{n_1} w_{ij} I_{(t, \infty)}(X_j).
$$

Given the two samples $S_i^{(k)}$ and $S_2^{(k)}$ the probability $P^{(k)}(X < X')$ is

$$
\begin{aligned}
P^{(k)}(X < X') & = \int S_2^{(k)} d[-S_1^{(k)}] \\
& = w_{10} w_{20} \int P_2^{(k)}(t, \infty) dP_1^{(k)} + w_{10} \sum_{j=1}^{n_2} w_{2j} P_1[0, X_j') \\
& \quad + w_{20} \sum_{j=1}^{n_1} w_{1j} P_2(X_j, \infty) + \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{1i} w_{2j} I_{(X_i, \infty)}(X_j') \\
& \geq \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{1i} w_{2j} I_{(X_i, \infty)}(X_j').
\end{aligned}
$$

As a consequence

$$
\begin{aligned}
\mathscr{E}_{\mu_1, \mu_2}[I_{P^{(k)}(X < X') > \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}_n'] & = \mathscr{E}_{P_1^{(k)}, w_1, P_2^{(k)}, w_2}[I_{\int S_2^{(k)} d[-S_1^{(k)}] > \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}_n'] \\
& \geq E_{w_1, w_2}[I_{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{1i} w_{2j} I_{(X_i, \infty)}(X_j') > \frac{1}{2}} | \mathbf{Z}_n, \mathbf{Z}_n'].
\end{aligned}
$$

The bound $\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{1i} w_{2j} I_{(X_i, \infty)}(X_j')$ is reached for the priors $\alpha_1^* = \delta_{X_0}$ and $\alpha_2^* = \delta_{X_0'}$ with $X_0 > X_i'$, $i = 0, \dots, n_2$ and $X_0' < X_i$, $i = 0, \dots, n_1$ which are the same priors that give the lower expectation of $P(X < X')$. ∎

7   Gaussian approximation of $P(X < X')$

In this section, and more precisely in Theorem S.2, we derive an analytical expression of the posterior variance σ^2 of the lower distribution of $P(X < X')$. This results can be used to approximate the distribution of $P(X < X')$ by a Gaussian distribution with mean given by (S.3.11) and variance equal to σ^2. The Gaussian approximation can be used to reduce the computational costs of the Monte Carlo sampling when the size of the available samples $\mathbf{Z}_n$ and $\mathbf{Z}'_{n'}$ is large.

We start by the following lemma which is necessary to prove Theorem S.2.

Lemma S.4. *Given the Dirichlet process $L(t,d)$ with base measure $\alpha_L(t,d) = \alpha_s(t)\delta_1(d) + \alpha_q(t)\delta_0(d)$ with $\alpha_s(t) = \sum_{i=1}^{n^d}\delta_{\tilde{X}_i}(t)$ and $\alpha_q(t) = \sum_{i=1}^{n^c}\delta_{\tilde{Y}_i}(t)$ it holds*

$$\mathscr{E}\left[\prod_{i=1}^{m}\left(\frac{S^*(Z_i) + Q^*(Z_i)}{S^*(Z_{i-1}) + Q^*(Z_i)}\right)^{p_i}\right] = \prod_{i=1}^{m}\prod_{j=0}^{p_i-1}\frac{\alpha_s(Z_i,\infty) + \alpha_q[Z_i,\infty) + j}{\alpha_s[Z_i,\infty) + \alpha_q[Z_i,\infty) + j}. \tag{S.7.23}$$

∎

Proof. *Partition $\mathbb{X} = \mathbb{R} \times \{0,1\}$ into the $2m+1$ subsets $A_1 = [0,Z_1) \times \{0,1\}, A_2 = \{(Z_1,1)\}, A_3 = \{(Z_1,0)\} \cup (Z_1,Z_2) \times \{0,1\},\ldots, A_{2m} = \{(Z_m,1)\}, A_{2m+1} = \{(Z_m,0)\} \cup (Z_m,\infty) \times \{0,1\}$. Then, the random vector of probabilities $P(A_i)$ has a Dirichlet distribution with parameter vector $(\beta_1,\ldots,\beta_{2m+1})$, where $\beta_i = \alpha_L(A_i)$, and hence*

$$\mathscr{E}\left[\prod_{i=1}^{m}\left(\frac{S^{*(k)}(Z_i^+) + Q^{*(k)}(Z_i)}{S^{*(k)}(Z_i^-) + Q^{*(k)}(Z_i)}\right)^{p_i}\right]$$

$$= K\int\ldots\int\prod_{i=1}^{m}\left(1 - \sum_{j=1}^{2i}\theta_j\right)^{p_i}\left(1 - \sum_{j=1}^{2i-1}\theta_j\right)^{-p_i}\prod_{j=1}^{2m}\theta_j^{\beta_j-1}\left(1 - \sum_{i=1}^{2m}\theta_i\right)^{\beta_{2m+1}-1}\prod_{i=1}^{2m}d\theta_i$$

$$= K\int\ldots\int\prod_{i=1}^{m}\theta_{2i-1}^{\beta_{2i-1}-1}\theta_{2i}^{\beta_{2i}-1}\left(1 - \sum_{j=1}^{2i}\theta_j\right)^{p_i}\left(1 - \sum_{j=1}^{2i-1}\theta_j\right)^{-p_i}\left(1 - \sum_{i=1}^{2m}\theta_i\right)^{\beta_{2m+1}-1}\prod_{i=1}^{2m}d\theta_i,$$

where the $2m$-dimensional integral is carried over the volume $\{(\theta_1,\ldots,\theta_{2m}) : \theta_i \geq 0, \sum_{i=1}^{m}\theta_i \leq 1\}$ and

$$K = \frac{\Gamma(s+n)}{\prod\limits_{i=1}^{2m+1}\Gamma(\beta_i)}.$$

Having defined $\theta_i^c = (1 - \sum_{j=1}^{i}\theta_j)$ and $B(\eta,\nu) = \frac{\Gamma(\eta)\Gamma(\nu)}{\Gamma(\eta+\nu)}$, and using the equality

$$\int_0^c \theta^{\eta-1}(c-\theta)^{\nu-1}d\theta = c^{\eta+\nu-1}B(\eta,\nu) \text{ for } 0 \leq c \leq 1 \text{ and } \eta,\nu > 0,$$

we obtain by integrating first with respect to $d\theta_{2m}d\theta_{2m-1}$

$$\int_0^{\theta_{2m-2}^c}\theta_{2m-1}^{\beta_{2m-1}-1}(\theta_{2m-2}^c - \theta_{2m-1})^{-p_m}d\theta_{2m-1}\int_0^{\theta_{2m-1}^c}\theta_{2m}^{\beta_{2m}-1}(\theta_{2m-1}^c - \theta_{2m})^{p_m+\beta_{2m+1}-1}d\theta_{2m}$$

$$= \int_0^{\theta_{2m-2}^c}\theta_{2m-1}^{\beta_{2m-1}-1}(\theta_{2m-2}^c - \theta_{2m-1})^{\beta_{2m}+\beta_{2m+1}-1}d\theta_{2m-1}B(\beta_{2m},p_m+\beta_{2m+1}) \tag{S.7.24}$$

$$= (\theta_{2m-3}^c - \theta_{2m-2})^{\beta_{2m-1}+\beta_{2m}+\beta_{2m+1}-1}B(\beta_{2m-1},\beta_{2m}+\beta_{2m+1})B(\beta_{2m},p_m+\beta_{2m+1}).$$

By repeating this double integration for m times, we obtain

$$\mathscr{E}\left[\prod_{i=1}^{m}\left(\frac{S^*(Z_i^+)+Q^*(Z_i)}{S^*(Z_i^-)+Q^*(Z_i)}\right)^{p_i}\right] = K\prod_{i=1}^{m}B\left(\beta_{2i-1},\sum_{j=2i}^{2m+1}\beta_j\right)B\left(\beta_{2i},p_i+\sum_{j=2i+1}^{2m+1}\beta_j\right)$$

Finally, using the fact that $B(\eta,\nu) = \frac{\Gamma(\eta)\Gamma(\nu)}{\Gamma(\eta+\nu)}$, $\Gamma(\eta+1) = \eta\Gamma(\eta)$,

$$\sum_{j=2i+1}^{2m+1}\beta_j = \alpha_s(Z_i,+\infty)+\alpha_q[Z_i,+\infty)$$

and

$$\sum_{j=2i}^{2m+1}\beta_j = \alpha_s[Z_i,+\infty)+\alpha_q[Z_i,+\infty)$$

after some calculations, one obtains

$$\prod_{i=1}^{m}\prod_{j=0}^{p_i-1}\frac{\alpha_s(Z_i,+\infty)+\alpha_q[Z_i,+\infty)+j}{\alpha_s[Z_i,+\infty)+\alpha_q[Z_i,+\infty)+j}.$$

∎

Theorem S.2. *The posterior variance of $P(X < X')$ at the priors $\overline{\mu}_1$ and $\underline{\mu}_2$ with base measures given in (14) is*

$$\sigma^2 = trace[A^T E_{W_1}[\overline{W}\,\overline{W}^T]AE_{W_2}[\underline{W}\,\underline{W}^T]] - \mathscr{E}_{\overline{\mu}_1,\underline{\mu}_2}[P(X<X')|\mathbf{Z}_n,\mathbf{Z}'_{n'}]^2, \tag{S.7.25}$$

where the expectations E_{W_1} and E_{W_2} are taken w.r.t. the (Dirichlet) distributions of W_1 and W_2, $\overline{W} = [\overline{w}_1,\ldots,\overline{w}_{n^d}]^T$, $\underline{W} = [\underline{w}_0,\ldots,\underline{w}_{n'}]^T$, A is an $n^d \times n'$ matrix with elements $a_{jk} = I_{(\tilde{X}_j,\infty)}(Z'_k)$, and $E_{W_1}[\overline{W}\,\overline{W}^T]$ and $E_{W_2}[\underline{W}\,\underline{W}^T]$ are, respectively, $n^d \times n^d$ and $n' \times n'$ symmetric matrices of elements, respectively, $\overline{e}_{ij} = E[\overline{w}_i\overline{w}_j]$ and $\underline{e}_{ij} = E[\underline{w}_i\underline{w}_j]$, which, for $i < j$ are given by

$$E[\overline{w}_i\overline{w}_j] = \frac{d_id_j}{(n_i+s+1)(n_j+s)}\prod_{h=1}^{i-1}\frac{n_h^+ +s+1}{n_h+s+1}\prod_{h=1}^{j-1}\frac{n_h^+ +s}{n_h+s},$$

$$E[\overline{w}_i^2] = \frac{(1+d_i)d_i}{(n_i+s+1)(n_i+s)}\prod_{h=1}^{i-1}\frac{(n_h^+ +s+1)(n_h^+ +s)}{(n_h+s+1)(n_h+s)}$$

$$E[\underline{w}_i\underline{w}_j] = \frac{\tilde{d}'_i\tilde{d}'_j}{(n'_i+\Pi_i+1)(n'_j+\Pi_j)}\prod_{h=1}^{i-1}\frac{n'_h+\Pi_h-\tilde{d}'_h+1}{n'_h+\Pi_h+1}\prod_{h=1}^{j-1}\frac{n'_h+\Pi_h-\tilde{d}'_h}{n_h+\Pi_h}$$

$$E[\underline{w}_i^2] = \frac{(1+\tilde{d}'_i)\tilde{d}'_i}{(n'_i+\Pi_i+1)(n'_i+\Pi_i)}\prod_{h=1}^{i-1}\frac{(n'_h+\Pi_h-\tilde{d}'_h+1)(n'_h+\Pi_h-\tilde{d}'_h)}{(n'_h+\Pi_h+1)(n'_h+\Pi_h)}.$$

∎

Proof.

From (S.3.15), the variance of $P(X < X')$ can be written as

$$\sigma^2 = E\left[\left(\sum_{i=1}^{n^d}\sum_{j=0}^{n'} \overline{w}_i \underline{w}_j I_{(\tilde{X}_i, \infty)}(Z'_j)\right)^2\right] - \mathscr{E}_{\overline{\mu}_1, \underline{\mu}_2}[P(X < X')|\mathbf{Z}_n, \mathbf{Z}'_{n'}]^2.$$

Thus, by exploiting the equality

$$\sum_{i=1}^{n^d}\sum_{j=0}^{n'} \overline{w}_{1i} \underline{w}_{2j} I_{(\tilde{X}_i, \infty)}(Z'_j) = \overline{W}^T A \, \underline{W},$$

the linearity of expectations and the independence of $\overline{W}$ and $\underline{W}$, one obtains

$$\begin{aligned} E[\overline{W}^T A \underline{W} \overline{W}^T A \underline{W}] &= E[\underline{W}^T A^T \overline{W} \, \overline{W}^T A \underline{W}] \\ &= E_{W_2}[\underline{W}^T A^T E_{W_1}[\overline{W} \, \overline{W}^T] A \underline{W}]. \end{aligned}$$

Since the result of this product is a scalar, it is equal to its trace and thus we can use the cyclic property

$$trace[E_{W_2}[\underline{W}^T A^T E_{W_1}[\overline{W} \, \overline{W}^T] A \underline{W}]] = trace[A^T E_{W_1}[\overline{W} \, \overline{W}^T] A E_{W_2}[\underline{W} \, \underline{W}^T]],$$

and finally obtain

$$\sigma^2 = trace[A^T E_{W_1}[\overline{W} \, \overline{W}^T] A E_{W_2}[\underline{W} \, \underline{W}^T]] - \mathscr{E}_{\overline{\mu}_1, \underline{\mu}_2}[P(X < X')|\mathbf{Z}_n, \mathbf{Z}'_{n'}]^2.$$

The proof is completed by deriving $E_{W_1}[\overline{W} \, \overline{W}^T]$ (and $E_{W_2}[\underline{W} \, \underline{W}^T]$) from the fact that

$$E[\overline{w}_i \overline{w}_j] = \mathscr{E}_{\overline{\mu}_1}[(S_1(\tilde{X}_{i-1}) - S_1(\tilde{X}_i))(S_1(\tilde{X}_{j-1}) - S_1(\tilde{X}_j))|\mathbf{Z}_n] \tag{S.7.26}$$

and a similarly

$$E[\underline{w}_i \underline{w}_j] = \mathscr{E}_{\underline{\mu}_2}[(S_2(Z'_{i-1}) - S_2(Z'_i))(S_2(Z'_{j-1}) - S_2(Z'_j))|\mathbf{Z}'_{n'}]. \tag{S.7.27}$$

Assuming (w.l.o.g.) that $Z_i < Z_j$ each term of (S.7.26) and (S.7.27) can be written as

$$\begin{aligned} \mathscr{E}_{\overline{\mu}_l}[S_l(Z_i)S_l(Z_j)|\mathbf{Z}_n] &= \mathscr{E}_{\mu_l}\left[\prod_{Z_h \le Z_i} \frac{S_l^*(Z_h) + Q_l^*(Z_h)}{S_l^*(Z_{h-1}) + Q_l^*(Z_h)} \prod_{Z_h \le Z_j} \frac{S_l^*(Z_h) + Q_l^*(Z_h)}{S_l^*(Z_{h-1}) + Q_l^*(Z_h)} \Big| \mathbf{Z}_n\right] \\ &= \mathscr{E}_{\overline{\mu}_l}\left[\prod_{Z_h \le Z_i} \left(\frac{S_l^*(Z_h) + Q_l^*(Z_h)}{S_l^*(Z_{h-1}) + Q_l^*(Z_h)}\right)^2 \prod_{Z_i < Z_h \le Z_j} \frac{S_l^*(Z_h) + Q_l^*(Z_h)}{S_l^*(Z_{h-1}) + Q_l^*(Z_h)} \Big| \mathbf{Z}_n\right], \end{aligned}$$

where $l = 1, 2$, and where one should use Z'_i instead of Z_i and $\underline{\mu}_l$ instead of $\overline{\mu}_l$ for $l = 2$. Then, by Lemma S.4, we have

$$\begin{aligned} \mathscr{E}_{\overline{\mu}_l}[S_l(Z_i)S_l(Z_j)|\mathbf{Z}_n] &= \prod_{Z_h \le Z_i} \frac{(\alpha_s(Z_h, \infty) + \alpha_q[Z_h, \infty) + 1)(\alpha_s(Z_h, \infty) + \alpha_q[Z_h, \infty))}{(\alpha_s[Z_h, \infty) + \alpha_q[Z_h, \infty) + 1)(\alpha_s(Z_h, \infty) + \alpha_q[Z_h, \infty))} \cdot \\ &\quad \cdot \prod_{Z_i < Z_h \le Z_j} \frac{\alpha_s(Z_h, \infty) + \alpha_q[Z_h, \infty)}{\alpha_s[Z_h, \infty) + \alpha_q[Z_h, \infty)}, \end{aligned} \tag{S.7.28}$$

where $\alpha_s = s\delta_{Z_0} + \sum_{i=1}^{n^d} \delta_{\tilde{X}_i}$ and $\alpha_q = \sum_{i=1}^{n^c} \delta_{\tilde{Y}_i}$ for $l = 1$ and $\alpha_s = \sum_{i=1}^{n'}(\pi_i + d'_i)\delta_{Z'_i}$ and $\alpha_q = \sum_{i=1}^{n'^c} \delta_{\tilde{Y}'_i}$ for $l = 2$. Then, if $i < j$, by substitution of (S.7.28) into (S.7.26) or (S.7.27) we obtain

$$\mathscr{E}_{\overline{\mu}_l}[(S_l(Z_i^-) - S_l(Z_i^+))(S_l(Z_j^-) - S_l(Z_j^+))|\mathbf{Z}_n] = \tag{S.7.29}$$
$$\prod_{Z_h < Z_i} \frac{(\alpha_s(Z_h,\infty) + \alpha_q[Z_h,\infty) + 1)(\alpha_s(Z_h,\infty) + \alpha_q[Z_h,\infty))}{(\alpha_s[Z_h,\infty) + \alpha_q[Z_h,\infty) + 1)(\alpha_s[Z_h,\infty) + \alpha_q[Z_h,\infty))} \cdot$$
$$\cdot \prod_{Z_i \le Z_h < Z_j} \frac{\alpha_s(Z_h,\infty) + \alpha_q[Z_h,\infty)}{\alpha_s[Z_h,\infty) + \alpha_q[Z_h,\infty)} \left(\frac{\alpha_s[Z_i,Z_i]\alpha_s[Z_j,Z_j]}{(\alpha_s[Z_i,\infty) + \alpha_q[Z_i,\infty) + 1)(\alpha_s[Z_j,\infty) + \alpha_q[Z_j,\infty))} \right),$$

whereas if $i = j$ we have

$$\mathscr{E}_{\overline{\mu}_l}[(S_l(Z_i^-) - S_l(Z_i^+))^2|\mathbf{Z}_n] = \tag{S.7.30}$$
$$\prod_{Z_h < Z_i} \frac{(\alpha_s(Z_h,\infty) + \alpha_q[Z_h,\infty) + 1)(\alpha_s(Z_h,\infty) + \alpha_q[Z_h,\infty))}{(\alpha_s[Z_h,\infty) + \alpha_q[Z_h,\infty) + 1)(\alpha_s[Z_h,\infty) + \alpha_q[Z_h,\infty))} \cdot$$
$$\cdot \left(\frac{\alpha_s[Z_i,Z_i](\alpha_s[Z_i,Z_i] + 1)}{(\alpha_s[Z_i,\infty) + \alpha_q[Z_i,\infty) + 1)(\alpha_s[Z_i,\infty) + \alpha_q[Z_i,\infty))} \right).$$

The matrices $E_{W_1}[\overline{W}\ \overline{W}^T]$ and $E_{W_2}[\underline{W}\ \underline{W}^T]$ can finally be derived by substituting the expressions for α_s and α_q into (S.7.29) and (S.7.30).　■

References

Blum, J. and Susarla, V. (1977). On the posterior distribution of a Dirichlet process given randomly right censored observations. *Stochastic Processes and their Applications*, 5(3):207–211.

Ghosh, J. and Ramamoorthi, R. (2003). *Bayesian nonparametrics*. Springer.

Letòn, E. and Zuluaga, P. (2005). Relationships among tests for censored data. *Biometrical J.*, 47(3):377–387.

Pericchi, L. and Walley, P. (1991). Robust Bayesian credible intervals and prior ignorance. *Int. Statistical Review*, 59:1–23.

Peterson, A. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *J. of the American Statistical Association*, 72(360):854–858.

Walley, P. (1996). Inferences from multinomial data: learning about a bag of marbles. *J. of the Royal Statistical Society. Series B (Methodological)*, 58(1):3–57.