Nonparametric Bayesian estimation of a concave distribution function with mixed interval censored data

Geurt Jongbloed, Frank van der Meulen and Lixue Pang

Delft University of Technology
October 22, 2020

Abstract
Assume we observe a finite number of inspection times together with information on whether a specific event has occurred before each of these times. Suppose replicated measurements are available on multiple event times. The set of inspection times, including the number of inspections, may be different for each event. This is known as mixed case interval censored data. We consider Bayesian estimation of the distribution function of the event time while assuming it is concave. We provide sufficient conditions on the prior such that the resulting procedure is consistent from the Bayesian point of view. We also provide computational methods for drawing from the posterior and illustrate the performance of the Bayesian method in both a simulation study and two real datasets.

Keywords: Bayesian nonparametrics, Dirichlet process, posterior consistency.

1 Introduction
In survival analysis, one is interested in the time a certain event occurs. For example, the event may be the onset of a disease. A well known complication often encountered in practice is censoring, where the precise time at which an event occurs is unknown, but partial information on it is available. In right censoring for example, one only observes the event if it occurs before a certain censoring time, otherwise one observes the censoring time accompanied by the information that the event occurred after this time. In interval censoring, one never sees the exact event time. Only an interval of positive length (possibly infinite) is observed which contains the event time of interest.

Suppose $X$ models the actual event time for one subject. Instead of observing $X$ directly, we observe a finite number of inspection times $0 < t_1 < t_2 < \cdots < t_k < \infty$, together with the information which of the intervals $[t_{j-1}, t_j]$ contains $X$. We will assume a setting in which we obtain data that are modelled as independent and identically distributed realisations of $X_1, \ldots, X_n$, each of which is distributed as $X$. For each subject, the set of inspection times, as well as the number of inspections, may be different. This type of data is known as \textit{mixed-case interval censored data}. Our model includes both the interval censoring case 1 model (also known as current status model) and interval censoring case 2 model for which $k = 1$ and $k = 2$ respectively. In many statistical models, there are reasons to impose specific assumptions on functional parameters, for example shape constraints. Incorporating such constraints into the estimation procedure often improves the accuracy of the resulting estimator. In this paper, we consider the problem of estimating the distribution function $F$ of $X$, assuming that $F$ is concave.
1.1 Related literature

In Groeneboom & Wellner (1992), the pointwise asymptotic distribution of the maximum likelihood estimator (MLE) of the distribution function in the interval censoring case 1 model is derived. For interval censoring case 2, the asymptotic pointwise distribution of the MLE is still not known. In the mixed case interval censoring model, the MLE has been studied by Schick & Yu (2000) where it is shown to be $L_1$-consistent. In Wellner & Zhang (2000) a panel count model is considered, which includes the mixed case interval censoring model as a special case, namely when the counting process has only one jump. For this panel count model, Wellner & Zhang (2000) study two estimators. In case the counting process has only one jump and there is one inspection time, their estimators coincide with the MLE for current status data ($k = 1$). If $k > 1$, this is not the case. Dümbgen, Freitag & Jongbloed (2004) consider the current status model with the additional constraint that the underlying distribution function $F_0$ is concave. It is shown that the supremum distance between the nonparametric least squares estimator and the underlying distribution function $F_0$ is of order $(\log n/n)^{2/5}$. For mixed case interval censoring, the MLE is shown to be asymptotically consistent under the assumption that $F_0$ is concave or convex-concave in Dümbgen, Freitag & Jongbloed (2006). In addition, an algorithm for computing the MLE is proposed there.

From the Bayesian perspective, Susarla & Van Ryzin (1976) derived a nonparametric Bayesian estimator for the event time distribution function based on right-censored data, using the Dirichlet process prior. A special feature in this right-censored model is that the posterior mean estimator can be constructed explicitly. For interval censored data, this explicit construction is not available. Calle & Gómez (2001) propose a nonparametric Bayesian approach in the interval censoring model and use a Markov Chain Monte Carlo algorithm to obtain estimators for the posterior mean. Doss & Huffer (2003) consider the Dirichlet Process prior in the interval censoring model. They develop and compare various Monte Carlo based algorithms for computing Bayesian estimators. A host of closely related Bayesian nonparametric models have been implemented in the DP-package in the R-language, Cf. Jara et al. (2011).

1.2 Contribution

In this paper, we define and study a Bayesian estimator of the event time distribution based on mixed-case interval censored data under the additional assumption that the distribution function is concave. An advantage of the Bayesian setup is the ease of constructing credible regions. To construct frequentist analogues of these, confidence regions, can be quite cumbersome, relying on either bootstrap simulations or asymptotic arguments. We address this problem from a theoretical perspective and provide conditions on the prior such that the resulting procedure is consistent. That is, assuming data are generated from a “true” distribution, we show that the posterior asymptotically (as the sample size increases) converges to this distribution. The proof relies on Schwartz’ method for proving posterior consistency (Cf. Section 6.4 in Ghosal & Van der Vaart (2017)). In addition, we provide computational methods for drawing from the posterior and illustrate its performance in a simulation study. Finally, we apply the Bayesian procedure on two real data sets and construct pointwise credible sets.

1.3 Outline

Section 2 sets off with introducing notation and formally describing the model. In section 3 we derive posterior consistency under a weak assumption on the prior distribution on the class of concave distribution functions. A Markov Chain Monte Carlo algorithm for obtaining draws from the posterior using the Dirichlet Mixture Process prior is detailed in section 4. In section 5 we perform a simulation study to illustrate the behaviour of the proposed Bayesian method. Furthermore, we apply it to two data sets in section 6, one concerned with Rubella and the other with breast cancer. The appendix contains proofs of some technical results.
2 Model, likelihood and prior

2.1 Model and likelihood

Suppose $X$ is a random variable in $[0, \infty)$ with concave distribution function $F_0$. Instead of observing $X$, we observe the random vector $(K, T, \Delta)$ that is constructed as follows. First, $K$ is sampled from a discrete distribution with probability mass function $p_K$ on $\{1, 2, \ldots\}$, representing the number of inspection times. Given $K = k$, $T \in \mathbb{R}^k$ is sampled from a density $g_k$ supported on the set $\{t = (t_1, \ldots, t_k) \in (0, L)^k : 0 < t_1 < \cdots < t_k < \infty\}$ for some constant $L$. This random vector contains the (ordered) inspection times. Finally, $\Delta \in \{0, 1\}^{k+1}$ is the vector indicating in which of the $k+1$ intervals generated by $T$ the event actually happened. Thus, it is defined as the vector with $j$-th component

$$\Delta_j = 1_{(t_{j-1}, t_j]}(X) \text{ for } 1 \leq j \leq k + 1$$

where $T_0 = 0$ and $T_{k+1} = \infty$ by convention.

This procedure is repeated independently, so for sample size $n$ the data is a realisation of

$$\mathcal{D}_n := \{(K_i, T^i, \Delta^i) = (K_i, T_{i,1}, \ldots, T_{i,L}, \Delta_{i,1}, \ldots, \Delta_{i,L+1}), i = 1, \ldots, n\}.$$ 

Define the sets

$$\mathcal{C}_k = \{t \in (0, L)^k : 0 < t_1 < \cdots < t_k < \infty\} \quad (2.1)$$

and $\mathcal{H}_k = \{\delta \in \{0, 1\}^{k+1} : \sum_{j=1}^{k+1} \delta_j = 1\}$, $k = 1, 2, \ldots$. Then $\mathcal{D}_n \in (\bigcup_{k=1}^\infty \{k\} \times \mathcal{C}_k \times \mathcal{H}_k)^n$.

Upon conditioning on the observed inspection times, we can define the likelihood of the distribution function $F$ by

$$L(F) = \prod_{i=1}^n \left( \frac{p_K(K_i) g_K(T^i)}{\prod_{j=1}^{K_i+1} (F(T_{i,j}) - F(T_{i,j-1}))^{\Delta_{i,j}}} \right). \quad (2.2)$$

We denote the joint distribution of $\{(K_i, T^i), 1 \leq i \leq n\}$ by $\mathbb{P}_{K,T}$. Given these $(K_i, T^i)$s the vectors $\Delta^i$ have multinomial distributions with probabilities depending on $F_0$. The distribution of $\mathcal{D}_n$ will be denoted by $\mathbb{P}_0$. Expectation with respect to measures will be denoted by $\mathbb{E}$, supplemented by a subscript referring to the measure.

2.2 Prior specification

In order to estimate the underlying concave distribution function in a Bayesian way, we construct a prior distribution on the set of all concave distribution functions. For $\theta > 0$, denote the uniform density function on $[0, \theta]$ by $\varphi(\cdot \mid \theta)$ and its distribution function by $\Psi(\cdot \mid \theta)$, i.e.

$$\varphi(x, \theta) = \frac{1}{\theta} 1\{x \leq \theta\} \text{ and } \Psi(x, \theta) = \frac{\min(x, \theta)}{\theta} \text{ respectively, } \ x \geq 0. \quad (2.3)$$

It is well known that any concave distribution function $F$ on $[0, \infty)$ allows the mixture representation (see Feller (1966))

$$F(x) = \int \Psi(x, \theta) dG(\theta), \quad (2.4)$$

where $G$ is a distribution function on $[0, \infty)$. In what follows, we sometimes stress this representation and denote the concave distribution function by $F_G$. In order to put a prior measure $\Pi$ on the set

$$\mathcal{F} = \left\{ F : F \text{ is a concave distribution on } [0, \infty) \right\},$$

we use (2.4) together with a prior distribution $\Pi^*$ on the set of all mixing distribution functions $G$ on $[0, \infty)$ (denote as $\mathcal{M}$). Having chosen such a prior measure, we denote the resulting posterior measure on $\mathcal{F}$ by $\Pi(\cdot \mid \mathcal{D}_n)$. 

3
3 Posterior consistency

In this section we establish consistency of the posterior distribution $\Pi(\cdot | \mathcal{D}_n)$ under a weak condition on the prior measure $\Pi$. Generally, the posterior is said to be consistent at $F_0$ (with respect to a semimetric $d$) if for any $\varepsilon > 0$, $\mathbb{E}_0 \Pi(d(F, F_0) > \varepsilon \mid \mathcal{D}_n) \to 0$ when $n \to \infty$.

For any distribution function $G$, denote $G_{i,j} = G(T_{i,j}) - G(T_{i,j-1})$. Given the inspection times $\{T_i, 1 \leq i \leq n\}$, we say that distribution functions $G$ and $F$ belong to the same equivalence class if the increments between the adjacent times are the same: $G_{i,j} = F_{i,j}$ for all $i = 1, \ldots, n, j = 1, \ldots, K_i + 1$. Then given data $\mathcal{D}_n$, we define a distance $d$ between two (equivalence classes of) distribution functions $G$ and $F$ by

$$d_n(G, F) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K_i+1} |G_{i,j} - F_{i,j}|. \quad (3.1)$$

Recall that $\Pi^*$ is a prior on the set $\mathcal{M}$, then $G$ is in the weak support of $\Pi^*$ if every weak neighborhood of $G$ has positive measure.

**Theorem 3.1** Fix $F_0 \in \mathcal{F}$ and $x \in [0, \infty)$. Consider the mixed-case interval censoring model described in section 1. Assume $F_0$ has a continuous density function $f_0$ on $(0, \infty)$ with $f_0(0) \leq M < \infty$ and that the weak support of the prior distribution $\Pi^*$ is $\mathcal{M}$. If $\mathbb{E}K^r < \infty$, for some $r > 1/2$, then for any $\varepsilon > 0$, we have $\mathbb{P}_0$-almost surely that

$$\Pi(F \in \mathcal{F} : d_n(F, F_0) > \varepsilon \mid \mathcal{D}_n) \to 0 \text{ as } n \to \infty.$$ 

Note that $d_n$ in Theorem 3.1 is a random semidistance since it depends on the inspection times $\{K_i, T_i, i = 1, \ldots, n\}$, also depending on $n$. Define the measure $\mu$ on the Borel $\sigma$–field $\mathcal{B}$ on $[0, \infty)$ that measures the “expected proportion of inspection times contained in a Borel set $B \in \mathcal{B}$” by

$$\mu(B) = \sum_{k=1}^{\infty} p_k(k)k^{-1} \int g_k(t) \sum_{j=1}^{k} 1_B(t_j)dt.$$ 

As a special case, assume that given $k, S_1, \ldots, S_k$ are independent and identically distributed with density function $\xi$ on $[0, \infty)$ and $\{T_1 < T_2 < \cdots < T_k\}$ are the ordered $S_j$’s. Then when $k = 1$,

$$\mu(B) = \int g_1(t_1)1_B(t_1)dt_1 = \int_B \xi(x)dx.$$ 

When $k = 2$, for any $a \in [0, \infty)$

$$\mu((0,a]) = \frac{1}{2} \int g_2(t)(1\{t_1 \leq a\} + 1\{t_2 \leq a\})dt = \frac{1}{2}(\mathbb{P}(t_1 \leq a) + \mathbb{P}(t_2 \leq a))$$

$$= \frac{1}{2} \left(1 - \left(1 - \int_0^a \xi(x)dx\right)^2 + \left(\int_0^a \xi(x)dx\right)^2\right) = \int_0^a \xi(x)dx$$

Hence, the measure $\mu$ has density $\xi$ in interval-censoring cases 1 and 2.

The follow result establishes posterior consistency with respect to $L_1(\mu)$ loss.

**Theorem 3.2** Let $F_0$, $\Pi$ and $K$ satisfy the conditions of Theorem 3.1. Then for any $\varepsilon > 0$, we have

$$\mathbb{E}_0 \Pi \left(F \in \mathcal{F} : \int |F - F_0| d\mu > \varepsilon \mid \mathcal{D}_n\right) \to 0 \text{ as } n \to \infty.$$
3.1 Proofs

For proving Theorem 3.1 we use Schwartz’ approach to derive posterior consistency. In the proof of this theorem, Lemma 3.3 is used to control the prior mass of a neighbourhood of the true distribution. Lemma 3.4 provides appropriate test functions. Both lemmas are stated below; the proofs are in appendix 7.

Lemma 3.3 Let $F_0$ and $\Pi^*$ satisfy the conditions of Theorem 3.1. Define, for $F_1, F_2 \in \mathcal{F}, k = 1, 2, \ldots$ and $t \in C_k$ as defined in (2.1):

$$h_{k,F_1,F_2}(t) = \sum_{j=1}^{k+1} (F_0(t_j) - F_0(t_{j-1})) \log \frac{F_1(t_j) - F_1(t_{j-1})}{F_2(t_j) - F_2(t_{j-1})}$$

(where $t_0 = 0$ and $t_{k+1} = \infty$ by convention). If we define,

$$S(\eta) = \left\{ F \in \mathcal{F} : \sum_{k=1}^{\infty} p_K(k) \int g_k(t) h_{k,F_0,F}(t) dt < \eta \right\},$$

then for all $\eta > 0$, $\Pi(S(\eta)) > 0$.

Note that that for the specific choice $F_1 = F_0$, by Jensen’s inequality, $h_{k,F_0,F} \geq 0$ for all $F \in \mathcal{F}$.

Lemma 3.4 For $\epsilon > 0$, define $U_\epsilon := \{ F \in \mathcal{F} : d_n(F, F_0) > \epsilon \}$. Then there exists a sequence of test functions $\Phi_n$ such that for all $n \geq 1$,

$$\mathbb{E}_0(\Phi_n) \leq C e^{-nc}$$

$$\mathbb{E}_{(K,T)} \left\{ \sup_{F \in U_\epsilon} \mathbb{E}_F[1 - \Phi_n | K, T] \right\} \leq C e^{-nc}$$

for some positive constants $c$ and $C$.

Proof of Theorem 3.1 Choose $\epsilon > 0$ and define the set $U_\epsilon$ as in Lemma 3.4. Define

$$Z_{i,j} = \frac{F(T_{i,j}) - F(T_{i,j-1})}{F_0(T_{i,j}) - F_0(T_{i,j-1})}.$$

Using expression (2.2) of the likelihood, the posterior mass of the set $U_\epsilon$ can be written as

$$\Pi(U_\epsilon | \mathcal{D}_n) = D_n^{-1} \int_{U_\epsilon} \prod_{i=1}^{n} \prod_{j=1}^{K_i+1} Z_{i,j}^{\Delta_{i,j}} d\Pi(F),$$

where

$$D_n = \int_{U_\epsilon} \prod_{i=1}^{n} \prod_{j=1}^{K_i+1} Z_{i,j}^{\Delta_{i,j}} d\Pi(F).$$

Fix $0 < \eta < c/2$, where $c$ is as it appears in Lemma 3.4. Also fix $F \in S(\eta)$.

We first show that Lemma 3.3 implies for any $\eta' > \eta$ we have $F_0$-a.s. that

$$D_n \geq \exp(-n\eta') \Pi(S(\eta'))$$

for all $n$ sufficiently large. By Lemma 3.3, we have $\Pi(S(\eta)) > 0$. Let $\Pi_{S(\eta)}$ be $\Pi$ restricted to $S(\eta)$ and normalised to a probability measure. For $i \geq 1$ define

$$Y_{i,j} = -\int \Delta_{i,j} \log Z_{i,j} d\Pi_{S(\eta)}(F) 1_{\{1, 2, \ldots, K_i+1\}}(j).$$
Note that,
\[
\mathbb{E}_0 \left[ \sum_{j=1}^{K_i+1} Y_{i,j} \right] = \mathbb{E}_{K_i,T_i} \left[ \mathbb{E}_0 \left[ \sum_{j=1}^{K_i+1} Y_{1,j} \mid T^{K_i}, K_i \right] \right] \\
= \mathbb{E}_{K_i,T_i} \left[ \sum_{j=1}^{K_i+1} \int -\left( F_0(T_{1,j}) - F_0(T_{1,j-1}) \right) \log Z_{i,j} d\Pi_S(\eta)(F) \right] \\
= \sum_{k=1}^{\infty} p_K(k) \int \int g_k(t) h_{k,F_0,F}(t) dt d\Pi_S(\eta)(F) \leq \eta < \infty.
\]

Therefore, the law of large numbers yields
\[
\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K_i+1} Y_{i,j} \to \mathbb{E}_0 \left[ \sum_{j=1}^{K_i+1} Y_{1,j} \right] \leq \eta, \quad \mathbb{P}_0 \text{-a.s.}
\]

Hence, \( \mathbb{P}_0 \)-a.s. for any \( \eta' > \eta \),
\[
D_n \geq \int \prod_{i=1}^{n} \prod_{j=1}^{K_i+1} Z_{i,j}^{\Delta_{i,j}} d\Pi(F) = \Pi(S(\eta)) \int \prod_{i=1}^{n} \prod_{j=1}^{K_i+1} Z_{i,j}^{\Delta_{i,j}} d\Pi_S(\eta)(F) \\
= \Pi(S(\eta)) \int \exp \left( \sum_{i=1}^{n} \sum_{j=1}^{K_i+1} \Delta_{i,j} \log Z_{i,j} \right) d\Pi_S(\eta)(F) \\
\geq \Pi(S(\eta)) \exp \left( -n \cdot \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K_i+1} Y_{i,j} \right) \\
\geq \exp(-n\eta') \Pi(S(\eta)) \quad (3.5)
\]

for \( n \) sufficiently large, where we used Jensen’s inequality in the second inequality.

Now we can finish the proof by combining this result with the test functions \( \Phi_n \) satisfying (3.4) (by Lemma 3.4).

By inequality (3.5), we can bound \( \mathbb{E}_0 \Pi(U_\epsilon \mid D_n) \) as follows,
\[
\mathbb{E}_0 \Pi(U_\epsilon \mid D_n) = \mathbb{E}_0 \Pi(U_\epsilon \mid D_n) \Phi_n + \mathbb{E}_0 \Pi(U_\epsilon \mid D_n)(1 - \Phi_n) \\
\leq \mathbb{E}_0 \Phi_n + \Pi(S(\eta))^{-1} e^{n\eta'} \mathbb{E}_0 \int_{U_\epsilon} \prod_{i=1}^{n} \prod_{j=1}^{K_i+1} Z_{i,j}^{\Delta_{i,j}} (1 - \Phi_n) d\Pi(F) \\
= \mathbb{E}_0 \Phi_n + \Pi(S(\eta))^{-1} e^{n\eta'} \mathbb{E}_0 \int_{U_\epsilon} \mathbb{E}_F(1 - \Phi_n)d\Pi(F) \\
\leq C e^{-cn} + \Pi(S(\eta))^{-1} \cdot C e^{-(c-\eta')n} = o(1) \quad \text{as } n \to \infty.
\]

The final step follows by choosing \( \eta' < c \). Since \( \sum_{n=1}^{\infty} e^{-bn} < \infty \) for any constant \( b \), almost sure convergence follows by the Borel-Cantelli lemma.

\[\square\]

**Proof of Theorem 3.2** First note that the proof of (16) in Dümbgen, Freitag & Jongbloed (2006) shows for all distribution functions \( F, F_0 \in \mathcal{F} \),
\[
d'_n(F,F_0) = \frac{1}{n} \sum_{i=1}^{n} K_i^{-1} \sum_{j=1}^{K_i} |F(T_{i,j}) - F_0(T_{i,j})| \leq d_n(F,F_0). \quad (3.6)
\]
For any $\epsilon > 0$, denote set
\[
A_n = \left\{ \sup_{F \in \mathcal{F}} \left| d_n^r(F, F_0) - \int |F - F_0| d\mu \right| > \epsilon / 2 \right\},
\]
Now we prove that $\mathbb{P}_{(K,T)}(A_n) \to 0$ as $n \to \infty$. Fix $F_0 \in \mathcal{F}$ and denote
\[
\psi_i(F) = n^{-1} K_i^{-1} \sum_{j=1}^{K_i} |F(T_{i,j}) - F_0(T_{i,j})|.
\]
Then $d_n^r(F, F_0) = \sum_{i=1}^n \psi_i(F)$. Note that $\mathbb{E}_{(K,T)} d_n^r(F, F_0) = \int |F - F_0| d\mu$. It is sufficient to show that
\[
\mathbb{E}_{(K,T)} \sup_{F \in \mathcal{F}} |d_n^r(F, F_0) - \mathbb{E}_{(K,T)} d_n^r(F, F_0)| \to 0. \tag{3.7}
\]
By theorem 7.2, it is implied by the existence of a sequence $\delta_n \to 0$ such that
\[
\mathbb{E}_{(K,T)} \sum_{i=1}^n \sup_{F \in \mathcal{F}} |\psi_i(F)| = O(1), \tag{3.8}
\]
\[
\mathbb{E}_{(K,T)} \sum_{i=1}^n \mathbf{1}\{ \sup_{F \in \mathcal{F}} |\psi_i(F)| > \delta_n \} \sup_{F \in \mathcal{F}} |\psi_i(F)| = o(1), \tag{3.9}
\]
for any $u > 0$, $\log \mathcal{N}(u, \mathcal{F}, \rho_n) = c(u)$. \tag{3.10}

Here
\[
\mathcal{N}(u, \mathcal{F}, \rho_n) = \min \left\{ \# \mathcal{G} : \mathcal{G} \subset \mathcal{F}, \inf_{G \in \mathcal{G}} \rho_n(F, G) \leq u \text{ for all } F \in \mathcal{F} \right\},
\]
and
\[
\rho_n(F, F') = \sum_{i=1}^n |\psi_i(F) - \psi_i(F')|.
\]
For (3.8) and (3.9), note that $\sup_{F \in \mathcal{F}} |\psi_i(F)| \leq n^{-1}$, hence $\mathbb{E}_{(K,T)} \sum_{i=1}^n \sup_{F \in \mathcal{F}} |\psi_i(F)| \leq 1$. By taking $n \delta_n \to \infty$, e.g. $\delta = \frac{1}{\sqrt{n}}$,
\[
\mathbb{E}_{(K,T)} \sum_{i=1}^n \mathbf{1}\{ \sup_{F \in \mathcal{F}} |\psi_i(F)| > \delta_n \} \sup_{F \in \mathcal{F}} |\psi_i(F)| \leq n^{-1} \mathbb{E} \sum_{i=1}^n \mathbf{1}\{ n^{-1} > \delta_n \} = n^{-1} \to 0
\]
For (3.10), note that
\[
\rho_n(F, F') = \sum_{i=1}^n |\psi_i(F) - \psi_i(F')| \leq n^{-1} \sum_{i=1}^n K_i^{-1} \sum_{j=1}^{K_i} |F(T_{i,j}) - F_0(T_{i,j})| - |F'(T_{i,j}) - F_0(T_{i,j})| + |F(T_{i,j}) - F'(T_{i,j})| = \int |F - F' - d\mu| \leq \left( \int |F - F'|^2 d\mu \right)^{1/2}
\]
where the measure $\nu$ is defined by $\nu(\cdot) = n^{-1} \sum_{i=1}^n K_i^{-1} \sum_{j=1}^{K_i} \delta_{T_{i,j}}(\cdot)$. In the final step, we use Hölder’s inequality and that $\nu$ has total mass 1. Further, using Lemma 2.1 and equation (25) in van de Geer (2000) we obtain
\[
\log \mathcal{N}(u, \mathcal{F}, \rho_n) \leq \log \mathcal{N}(u, \mathcal{F}, L_2(\nu)) \leq C u^{-1}
\]
for some constant $C$ and any $u > 0$. 

7
Therefore, denote \( B_\epsilon = \{ F \in \mathcal{F} : |F - F_0| d\mu > \epsilon \} \), by \( \mathbb{P}_{(K,T)}(A_n) \to 0 \) as \( n \to \infty \) and inequality (3.6), we have

\[
\mathbb{E}_0 \Pi (B_\epsilon \mid D_n) = \mathbb{E}_0 \Pi (B_\epsilon \mid D_n) 1_{A_n} + \mathbb{E}_0 \Pi (B_\epsilon \mid D_n) 1_{A_n^c} \\
\leq \mathbb{E}_0 (1_{A_n}) + \mathbb{E}_0 \Pi (F \in \mathcal{F} : d'_n(F, F_0) > \epsilon / 2 \mid D_n) \\
\leq \mathbb{P}_{(K,T)}(A_n) + \mathbb{E}_0 \Pi (F \in \mathcal{F} : d(F, F_0) > \epsilon / 2 \mid D_n) \to 0
\]

as \( n \to \infty \). \hfill \Box

4 Computational methods

Assume the mixing measure \( G \) is a Dirichlet process with base measure \( G_0 \) (with density \( g_0 \)) and concentration rate \( \alpha \). The prior distribution this induces on \( \mathcal{F} \) through (2.4) is called a Dirichlet Mixture Process (DMP). Denoting by \( \#(x) \) the number of distinct values in a vector \( x \), a sample \( X_1, \ldots, X_n \) from the DMP can be generated using the following steps:

\[
Z := (Z_1, \ldots, Z_n) \sim \text{CRP}(\alpha) \\
\Theta_1, \ldots, \Theta_{\#(Z)} \text{ iid } \sim G_0 \\
X_1, \ldots, X_n \mid \Theta_1, \ldots, \Theta_{\#(Z)}, Z_1, \ldots, Z_n \text{ iid } \sim \text{Unif}(0, \Theta_{Z_i}).
\] (4.1)

Here CRP(\( \alpha \)) denotes the “Chinese Restaurant Process” that can be viewed as follows. Assume in a Chinese restaurant, the first customer sits at the first table. Then, given a number of occupied tables, the next customer joins one of these tables with a probability proportional to the number of customers already there, or starts a new table with probability proportional to \( \alpha \). Interpreting \( Z_i \) as the number of customers sitting at table \( i \) after \( n \) customer arrivals, this leads to a distribution on the space of partitions of the integers \( \{1, 2, \ldots, n\} \).

In the interval censoring model, we do not observe the \( X_i \)'s, but for each \( i \) the interval \( (L_i, R_i] = (T_{i,J-1}, T_{i,J}] \) that contains \( X_i \). We are then interested in the conditional distribution of \( (Z, \Theta) \) given the data \( D_n \). In case we would have complete observations \( X_1, \ldots, X_n \), there are algorithms to sample from this conditional distribution (see Neal (2000)). Having only the interval censored data, we can adapt such algorithms, treating the unobserved event times \( X_i \) as latent variables in the same fashion as this is done in the case of right censoring by Hansen & Lauritzen (2002). Given the exact values \( X_i \), we can use existing algorithms to generate samples from the posterior. Subsequently, we update the \( X_i \)'s in each iteration by sampling conditionally on the time intervals \( (L_i, R_i] \) where the event happened.

We initialise a Gibbs sampler by specifying values of \( (Z, \Theta, X) \) that satisfy the constraints in the model. This means that \( \Theta_{Z_i} \geq X_i \) and \( X_i \in (L_i, R_i] \) for \( i = 1, \ldots, n \). For ease of notation let \( \Theta = (\Theta_1, \ldots, \Theta_{\#(Z)}) \) and \( X = (X_1, \ldots, X_n) \) for \( i = 1, \ldots, n \). Then the following steps are iterated:

1. sample \( Z \mid (X, \Theta, D_n) \);  
2. sample \( \Theta \mid (X, Z, D_n) \);  
3. sample \( X \mid (D_n, \Theta, Z) \).

Given \( X, D_n \) does not play any role when sampling \( Z \) and \( \Theta \). Hence the first two steps are the same as in the case of precise observations. More details on this step in that setting can be found in Jongbloed, van der Meulen & Pang (2018) and Neal (2000). The final step is to sample the latent variables \( X \) given \( D_n \), \( Z \) and \( \Theta \). For this, note that

\[
f_{X_i | D_n, \Theta, Z}(x \mid D_n, \theta, z) \propto f(x \mid \theta_{z_i}) 1_{(L_i, R_i]}(x) = \varphi(x \mid \theta_{z_i}) 1_{(L_i, R_i]}(x).
\]
This is the density of the uniform distribution on interval \((L_i, R_i] \cap [0, \Theta_{Z_i}]\). Note that with the initialisation described above, \((L_i, R_i] \cap [0, \Theta_{Z_i}]\) is non-empty.

Using the conjugacy property of Dirichlet process (see e.g. Ferguson (1973)), the conditional expectation of the posterior of \(F\) is given by

\[
\mathbb{E} \left[ \int \Psi(x, \theta) dG(\theta) \mid \Theta, Z, D_n \right] = \frac{1}{\alpha + n} \left( \alpha \int \Psi(x, \theta) dG_0(\theta) + \sum_{i=1}^{n} \Psi(x, \Theta_{Z_i}) \right). 
\]

Hence, the posterior mean of \(F\) can be obtained using a Markov Chain Monte Carlo approximation of the posterior of \((\Theta, Z)\) given \(D_n\). Having the algorithms to generate from the distribution of \((\Theta, Z) \mid D_n\), assume in the \(j\)-th iteration we obtained \(\left(\Theta_{z_i}^{(j)}, \ldots, \Theta_{z_n}^{(j)}\right)\). At iteration \(j\), a sample from the posterior is given by

\[
\hat{F}^{(j)}(x) := \frac{\alpha}{\alpha + n} \int \Psi(x, \theta) dG_0(\theta) + \frac{1}{\alpha + n} \sum_{i=1}^{n} \Psi(x, (\Theta_{z_i}^{(j)})). 
\]

After \(J\) iterations, an estimator for the posterior mean is given by \(J^{-1} \sum_{j=1}^{J} \hat{F}^{(j)}(x)\).

**Remark 4.1** In case the Dirichlet process is truncated, the target density is of fixed dimension. One of the referees raised the question whether probabilistic programming languages such as JAGS, BUGS, Stan or Turing can be used. First of all, we do not consider truncation here as, strictly speaking, it is not necessary. However, we fully agree that from a practical point of view the proposed approach may be implemented using one of the suggested Bayesian computational packages in case of truncation. What might be tricky here is that the workhorse algorithm in for example Stan (Hamiltonian Monte Carlo) uses automatic differentiation for computing gradients. However, the density of the uniform distribution on \([0, \theta]\), viewed as a function of \(\theta\) is not differentiable.

A host of related Bayesian nonparametric models have been implemented in the DP-package (Cf. Jara et al. (2011)).

## 5 Simulation results

In this section, we first study the posterior mean estimators of a concave distribution function based on simulated interval censored data. Next, we compare the Bayesian and the frequentist methods in this setting.

We simulate data by repeating independently \(n\) times the following scheme:

1. sample \(K\) from the discrete uniform distribution on the integers \(\{1, \ldots, 20\}\);
2. sample \(K\) inspection times \(T_1 < \cdots < T_K\) by sorting \(K\) independent and identically distributed random variables (we choose the Gamma distribution with shape parameter equal to 2 and rate parameter equal to 1);
3. sample \(X\) from the standard Exponential distribution;
4. set \(L := \sup \{T_j : T_j < X\}\) and \(R := \inf \{T_j : T_j \geq X\}\) (where \(T_0 = 0, T_{k+1} = \infty\)).

This leads to the dataset \(D_n\) containing the observation intervals \((L_i, R_i] \mid 1 \leq i \leq n\).

The prior is specified by a Dirichlet Process for the mixture measure. As seen in the formula (4.2), the concentration parameter \(\alpha\) expresses our confidence on the prior. As such the choice of \(\alpha\) can be interpreted as a prior sample size. In the following we choose \(\alpha = 1\), expressing a small prior sample size.
There is no obvious “optimal” choice for the base measure. The approach we take is motivated by the numerical study in (Jongbloed, van der Meulen & Pang (2018), section 5.1), where it is shown to yield a reasonable balance between computational tractability and performance. Write $Y \sim \text{Par}(s, \xi)$ with $s > 0$ and $\xi > 0$ if $f_Y(y) = \xi y^{-\xi-1}1\{y \geq s\}$. We choose the base measure to be a mixture of $\text{Par}(s, 1)$-distributions, where $s \sim \text{Gamma}(2, 1)$. This implies that the density of the base measure, $g_0$, satisfies $g_0(\theta) \sim \theta$ for $\theta \approx 0$. The Pareto-distribution is a conjugate prior to the Uniform distribution, alleviating computations in a Gibbs sampler. Mixing over $s$ is a practical way to robustify the prior.

For updating $S$, it follows from the short computation in section 5.1 in (Jongbloed, van der Meulen & Pang (2018)) that

$$f_{S|\Theta, Z}(s \mid \theta, z) \propto s^{1(z)}e^{-s}1\{s \leq \wedge(\theta)\}.$$

We take sample size $n = 100$. To show the algorithm’s performance, we show a traceplot and autocorrelation function of $\hat{F}^i(1)$ over 30,000 iterations in Figure 1.

![Figure 1](image-url)

Figure 1: Traceplot (left) and autocorrelation plot (right) for the posterior distribution function evaluated at 1 using the algorithm detailed in Section 4. The horizontal line in the left-hand figure depicts the true value $F_0(1) = 1 - e^{-1}$.

We compute the posterior mean estimator for the function $F_0$ using equation (4.2) for two samples from the standard exponential distribution: one with sample size 50 and the other with sample size 500. Figure 2 shows the results. The total number of MCMC iterations was chosen to be 30,000, with 15,000 burn-in iterations.

We now compare different estimation methods:

- the posterior mean for a concave distribution function;
- the maximum likelihood estimator under concavity;
- the maximum likelihood estimator without shape constraints.

We took $n = 500$ and considered $K_i = 1$, $K_i = 2$ for $i = 1, \ldots, n$ (interval censoring case 1 and 2) and $K_i$ independently sampled from the discrete uniform distribution on the integers $\{1, 2, \ldots, 20\}$, which we denote by $K \sim \text{Unif}(1, 20)$.
Figure 2: Posterior mean in case the data are sampled from the standard exponential distribution. Two solid lines depict the posterior mean and the dashed line is true distribution function (standard exponential). The concentration parameter is $\alpha = 1$ and the base measure is a mixture of Pareto distributions. The total number of MCMC iterations was chosen to be 30,000, with 15,000 burn-in iterations.
We use the same prior specification as before. Figure 3 depicts the estimators $\hat{F}$ (here we have three estimators: the NPMLE using the algorithm in Wellner & Zhan (1997), the concave MLE studied in Dümbgen, Freitag & Jongbloed (2006) and the Bayesian posterior mean estimator) and error curves $\hat{F} - F_0$, where $F_0$ is the true underlying distribution function. As the true distribution is smooth it is not surprising that NPMLE performs worst, as it is a step function. With an increasing number of inspection times, the procedure of generating the inspection time and event time gives a narrow inspection interval for each event. Although the NPMLE does not consider the concavity assumption on $F_0$, it suggests a concave shape. As can be seen in all cases, the concave MLE and the posterior mean estimator behave similarly.

![Figure 3: Three cumulative distribution function estimators of $F_0$ (left) and error curves (right) $\hat{F} - F_0$, $\hat{F}$: posterior mean, NPMLE and concave MLE in case $n = 500$ sample size data are sampled from the standard exponential distribution. From top to bottom corresponding different inspection time $K = 1, 2$ and $K \sim \text{Unif}(1, 20)$.](image)

Using the setting of mixed interval censoring ($K \sim \text{Unif}(1, 20)$), we generated 50 data sets of sizes $n = 50, 100, 200, 400, 800$ from the standard exponential and half-normal distribution and computed the NPMLE, the concave MLE, the posterior mean for each of the cases. Fix grid points $t_j = j/100, j = 1, \ldots, m$, where we took $m = 800$. Figures 4 and 5 show the log of the mean square error of $\hat{F}$ evaluated at $t = t_j, j = 1, \ldots, m$ for each sample size $n$, that is

$$\log R(\hat{F}, F)(t) = \log \frac{1}{50} \sum_{k=1}^{50} (\hat{F}^{(k)}(t) - F(t))^2$$

where $\hat{F}^{(k)}$ represent estimator based on the $k$–th data set. We see that all three estimators give small error. As seen from figure 3, it can be explained by the setting of how to generate mixed interval censoring data. We see that the posterior mean gives smallest error when $t$ is small, whereas all three estimators are comparable when $t \in [1, 4]$ of case $n = 800$. Finally, the NPMLE performs best when $t$ is big based on the data sets sample from the half-normal distribution.
Figure 4: The log mean square error (log $R(\hat{F}, F)$) evaluated at grid points $\{0.01, 0.02, \ldots, 8.0\}$ for NPMLE, the concave MLE and posterior mean in 50 data sets of different sample size $n = 50, 200, 800$ case sampled from the standard exponential distribution.
Figure 5: The log mean square error (log R(\hat{F}, F)) evaluated at grid points \{0.01, 0.02, \ldots, 8.0\} for NPMLE, the concave MLE and posterior mean in 50 data sets of different sample size \(n = 50, 200, 800\) case sampled from the standard halfnormal distribution.
We also consider a global value, the integrated square errors:

\[ ISE^{(k)}(\hat{F}, F) = \frac{1}{m} \sum_{j=1}^{m} (\hat{F}^{(k)}(t_j) - F(t_j))^2 \]

for each sample size \( n \), where \( \hat{F}^{(k)} \) represent estimator based on the \( k \)-th data set, \( k = 1, \ldots, 50 \). Figure 6 shows the mean of integrated square errors. In most of the cases, we see that the concave MLE has the smallest mean integrated square error, The posterior mean laying between NPMLE and the concave MLE and close to the concave MLE in case of half-normal distribution.

![Figure 6](image_url)

Figure 6: Mean of \( ISE^{(k)}(\hat{F}, F) \) for NPMLE, the concave MLE and posterior mean in 50 data sets of different sample size \( n = 50, 100, 200, 400, 800 \) case sampled from the standard exponential and halfnorm distribution.

## 6 Case study

In this section we illustrate the applicability of our method in real data examples. Using a nonparametric frequentist approach, producing confidence bands for the underlying distribution usually needs quite some fine tuning (see e.g. Groeneboom & Jongbloed). Contrary to the frequentist approach, within the
Bayesian approach it is simple to construct pointwise credible regions from MCMC output. We applied the Bayesian approach and two frequentist estimators to the Rubella data and Breast cancer data sets.

**Example 6.1** Rubella is a highly contagious childhood disease. The Rubella data concerns the prevalence of rubella in \( n = 230 \) Austrian males (see for more information Keiding et al. (1996)). The male individuals included in the data set represent an unvaccinated population. The data records whether a person got infected or not before a certain time. Here the upper limit of a person's life span is set equal to 100. Because there is only one inspection time per person, the data are actually case 1 interval censored. Figure 7 visualises the data, showing that the time intervals either start at 0 or end at 100.

The settings for computing the posterior mean are as described in the previous section (DP as the prior, with concentration parameter \( \alpha = 1 \) and the mixture of Pareto as the base measure. The total number of iterations was set to 30,000 where the initial 15,000 iterations have been treated as burn. Figure 8 shows the three estimators and 95\% pointwise credible sets for the underlying distribution function. The mle (assuming the distribution function to be concave) is comparable with the posterior mean. However, the posterior mean provides a smoother estimator as it is obtained by averaging and not as a maximizer of a likelihood (both the mle and mle under concavity assumption only change slope at censoring times).

![Figure 7: Visualisation of Rubella data. The x-axis is the range of event time. The horizontal lines displays the time intervals.](image)

**Example 6.2** In the Breast cancer study discussed in Finkelstein & Wolfe (1986), 94 early breast cancer patients were given radiation therapy with (RCT, 48) or without (RT, 46) adjuvant chemotherapy between 1976 and 1980. They were supposed to be seen at clinic visits every 4 to 6 months. However, actual visit times differ from patient to patient, and times between visits also vary. In each visit, physicians evaluated the appearance breast retraction. The data contain information about the time to breast retraction, hence, interval censored. Figure 9 visualises the data, we use the right end point 100 for the right censoring case.
Figure 8: NPMLE, concave MLE and posterior mean estimators (A) and 95% pointwise credible sets of the estimated posterior mean (B) for the underlying distribution function based on the Rubella data.

Figure 9: Visualisation of the Breast cancer data (left: RT, right: RCT). The x-axis is the range of event times. The horizontal lines display the time intervals.
The settings for computing the posterior mean are as in example 6.1. Figure 10 shows the three estimators under two treatments (RT and RCT) and 95% credible sets for the underlying survival function.

Figure 10: NPMLE, concave MLE and posterior mean estimators (top) and 95% pointwise credible sets of the estimated posterior mean (bottom) for the underlying survival function $1 - F(x)$ ($F(x)$ denotes the distribution function) based on the Breast cancer data.

7 Appendix: proofs of technical results

In the proof of lemma 3.3, we use the following lemma, it constructs a sequence of approximations for $F_0$.

**Lemma 7.1** Let $F_0$ satisfy the conditions stated in theorem 3.1. Then there exists a sequence of piece-
wise linear concave distribution functions \((F_m)\) such that

\[
\sum_{k=1}^{\infty} p_k(k) \int g_k(t) h_{k,F_0,F_m}(t) dt \to 0 \quad \text{as} \quad m \to \infty.
\]

**Proof** Since \(F_0\) is a concave distribution function, its density \(f_0\) is decreasing on \([0, \infty)\). We start off with the construction of functions \(f_m\) that approximate \(f_0\) (Cf. Theorem 18 in Wu & Ghosal (2008)). Choose \(m \in \mathbb{N}\) and let \(\hat{f}_{0,m} = \frac{f_0}{f_0(m)}\), then \(\hat{f}_{0,m} \to f_0\) pointwise as \(m \to \infty\). Let \(a_1\) and \(a_2\) be real numbers such that \(f_0(0) > a_1 > a_2 > 0\). By the continuity of \(f_0\), there exists \(x_2 > x_1\) satisfying \(f_0(x_1) = a_1\) and \(\hat{f}_{0,m}(x_2) = a_2\). See also Figure 11. Let \(m_1 \in \mathbb{N}\) and \(m_2 \in \mathbb{N}\) satisfy \(\frac{m_1}{m} < x_1 \leq \frac{m_1+1}{m}\) and \(\frac{m_2}{m} < x_2 \leq \frac{m_2+1}{m}\). Then define

\[
\hat{f}_m(x) = \begin{cases} 
\hat{f}_{0,m}(\frac{i}{m}), & \frac{i-1}{m} < x \leq \frac{i}{m}, 1 \leq i \leq m_1 \\
\hat{a}_1, & \frac{m_1}{m} < x \leq \frac{m_1+1}{m} \\
\hat{f}_{0,m}(\frac{i-1}{m}), & \frac{i-1}{m} < x \leq \frac{i}{m}, m_1 + 1 < i \leq m^2.
\end{cases}
\]

and \(\hat{f}_m(0) = \hat{f}_{0,m}(m^{-1})\). Because \(f_0\) is continuous on \([0, m]\), \(\hat{f}_m\) converges pointwise to \(f_0\) as \(m \to \infty\). Note \(\hat{f}_m\) is not a probability density function, as it will not integrate to one. We now normalize \(\hat{f}_m\) to a density function \(f_m\). First we can rewrite \(\hat{f}_m\) as

\[
\hat{f}_m(x) = \sum_{i=1}^{m^2} \hat{w}_i \varphi(x, i/m),
\]

where \(\varphi\) is defined as (2.3) and

\[
\hat{w}_i = \begin{cases} 
\frac{i}{m} (\hat{f}_0(\frac{i}{m}) - \hat{f}_0(\frac{i-1}{m})), & 1 \leq i < m_1 \\
\frac{m_1+1}{m} (\hat{f}_0(\frac{m_1+1}{m}) - \hat{a}_1), & i = m_1 \\
\frac{m_1+1}{m} (\hat{a}_1 - \hat{f}_0(\frac{m_1+1}{m})), & i = m_1 + 1 \\
\frac{i}{m} (\hat{f}_0(\frac{i-1}{m}) - \hat{f}_0(\frac{i}{m})), & m_1 + 1 < i < m^2 \\
\frac{m_1}{m} \hat{f}_0(\frac{m^2-1}{m}), & i = m^2.
\end{cases}
\]

Let

\[
w_i = \begin{cases} 
\hat{w}_i \frac{1}{\sum_{j=1}^{m_1-1} \hat{w}_j - \sum_{j=m_1+2}^{m^2} \hat{w}_j}, & m_1 \leq i \leq m_2, \\
\hat{w}_i, & \text{otherwise}.
\end{cases}
\]

Then \(\sum_{i=1}^{m^2} w_i = 1\) and \(w_i \geq 0\) (for \(m\) sufficiently large). Finally, define a sequence of probability density functions

\[
f_m(x) = \sum_{i=1}^{m^2} w_i \varphi(x, i/m). \quad (7.1)
\]
Note that for \( x \geq x_2 \), \( f_m(x) = \hat{f}_m(x) \). For each \( x \in [0, m] \),

\[
|f_m(x) - \hat{f}_m(x)| = \left| \sum_{i=1}^{m^2} w_i \varphi(x, \frac{i}{m}) - \sum_{i=1}^{m^2} \hat{w}_i \varphi(x, i/m) \right|
\]

\[
= \sum_{i=m_1}^{m^2} (w_i - \hat{w}_i) \varphi(x, i/m)
\]

\[
= \left| \left( 1 - \frac{\sum_{j=2}^{m_1} \hat{w}_j - \sum_{j=m_2+1}^{m} \hat{w}_j}{\sum_{j=m_1}^{m_2} \hat{w}_j} \right) - 1 \right| \sum_{i=m_1}^{m_2} \hat{w}_i \varphi(x, i/m)
\]

\[
\leq \left| 1 - \frac{\sum_{j=2}^{m_1} \hat{w}_j - \sum_{j=m_2+1}^{m} \hat{w}_j}{\sum_{j=m_1}^{m_2} \hat{w}_j} \right| - 1 \left( \sum_{i=m_1}^{m_2} \hat{w}_i \right) \frac{m}{m_1}
\]

\[
= 1 - 1 - \frac{1}{m} \sum_{i=2}^{m_2} \hat{f}_{0,m}(i/m) - \frac{m_1}{m_1} \to 0
\]

Here we use that \( m/m_1 \to x_1^{-1} \) and that the expression within the modular signs converges to 0 as difference between \( \int_0^m \hat{f}_{0,m}(x)dx \) and its Riemann sum approximate. Then we have \( |f_m - \hat{f}_{0,m}| \to 0 \) pointwise and \( \hat{f}_{0,m} \to f_0 \) pointwise. Hence \( f_m \) is a decreasing density and converges to \( f_0 \) pointwise. See an example in figure 11 for visualize \( f_0, \hat{f}_m \) and \( f_m \).

![Figure 11: Approximation for a decreasing function \( f_0 \). First we construct a step function \( \hat{f}_m \), then we normalize the weights \( \hat{w}_i \) to \( w_i \) such that \( f_m \) defined by (7.1) is a decreasing density function.](image)

Define \( F_m(x) = \int_0^x f_m(t)dt \), then using dominated convergence, we have \( F_m \to F_0 \) pointwise. As
\[ m \to \infty \ (m > L), \] then it follows that for all \( k \) and \( t \)

\[
h_{k, F_0, F_m}(t) = \sum_{j=1}^{k+1} (F_0(t_j) - F_0(t_{j-1})) \log \frac{F_0(t_j) - F_0(t_{j-1})}{F_m(t_j) - F_m(t_{j-1})} \to 0.
\]

The next step is to find an integrable upper bound for \( |h_{k, F_0, F_m}| \). Denote \( p_j = F_0(t_j) - F_0(t_{j-1}) \) for \( j = 1, \ldots, k + 1 \) and note that \( \sum_{j=1}^{k+1} p_j = 1 \). Then

\[
|h_{k, F_0, F_m}(t)| \leq \sum_{j=1}^{k+1} p_j \log p_j + \sum_{j=1}^{k+1} p_j \log(F_m(t_j) - F_m(t_{j-1})).
\]

Using Lagrange multipliers, the first sum achieves its maximal value over all probability vectors when all \( p_j \)'s would be equal. Hence it can be bounded by \( \log(k+1) \). For the second sum, by the construction of \( f_m \) we know that when \( x < x_2 \), \( f_m(x) \geq f_m(x_2) = f_0(x_2) = f_0(m/m_2/m) \geq a_2 \); when \( x_2 \leq x \leq m \), \( f_m(x) = f_0(x/m) = \frac{f_0(x)}{f_0(m)} \geq f_0(x) \). Since there exists \( j_0 \in \{1, \ldots, k+1\} \) such that \( t_{j_0-1} < x_2 \leq t_{j_0} \), the second sum can be bounded by \( I_1 + I_2 + I_3 \), where

\[
I_1 = -\sum_{j=1}^{j_0-1} p_j \log(a_2(t_j - t_{j-1}))
\]

\[
I_2 = -p_{j_0} \log(a_2(x_2 - t_{j_0-1}) + F_0(t_{j_0} \land m) - F_0(x_2))
\]

\[
I_3 = -\sum_{j=j_0+1}^{k} p_j \log p_j + p_{k+1} |\log(F_0(m) - F_0(t_k))|
\]

Again using the Lagrange multipliers, we have

\[
I_1 = -\sum_{j=1}^{j_0-1} p_j \frac{1}{a_2(t_j - t_{j-1})} (a_2(t_j - t_{j-1})) \log(a_2(t_j - t_{j-1}))
\]

\[
\leq -\frac{M}{a_2} \sum_{j=1}^{j_0-1} (a_2(t_j - t_{j-1})) \log(a_2(t_j - t_{j-1})) \leq \frac{M}{a_2} \log k
\]

In the second step we use \( p_j \leq M(t_j - t_{j-1}) \). In the final step, we use that \( \sum_{j=1}^{j_0-1} a_2(t_j - t_{j-1}) \leq 1 \). To bound \( I_2 \), we know that \( -x \log x \leq \frac{1}{e} \) when \( x \in (0, 1] \). Splitting \( I_2 \) into two parts, we have

\[
I_2 \leq -(F_0(t_{j_0}) - F_0(x_2)) \log(F_0(t_{j_0} \land m) - F_0(x_2)) - (F_0(x_2) - F_0(t_{j_0-1})) \log(a_2(x_2 - t_{j_0-1}))
\]

\[
\leq \frac{1}{e} \left( \frac{F_0(t_{j_0}) - F_0(x_2)}{F_0(t_{j_0} \land m) - F_0(x_2)} + \frac{F_0(x_2) - F_0(t_{j_0-1})}{a_2(x_2 - t_{j_0-1})} \right)
\]

\[
\leq \frac{1}{e} \left( \frac{F_0(t_{j_0})}{F_0(t_{j_0} \land m)} + \frac{M}{a_2} \right) \leq \frac{1}{e} \left( \frac{1}{F_0(L)} + \frac{M}{a_2} \right)
\]

In the last step, we used that \( \frac{F_0(t_{j_0})}{F_0(t_{j_0} \land m)} \leq \max(1, 1/F_0(m)) \leq 1/F_0(L) \). Similarly, we can bound \( I_3 \) by

\[
I_3 \leq \log k + p_{k+1} |\log(F_0(m) - F_0(t_k))|
\]

\[
\leq \log k + \frac{1}{e} \frac{1 - F_0(t_k)}{F_0(m) - F_0(t_k)} \leq \log k + \frac{1}{e} \frac{1}{F_0(m)} \leq \log k + \frac{1}{e} \frac{1}{F_0(L)}
\]
Therefore, having these bounds we obtain
\[ \left| h_{k,F_0,F_m}(t) \right| \leq \left( \frac{M}{a_2} + 2 \right) \log(k + 1) + \frac{1}{e} \left( \frac{2}{F_0(L)} + \frac{M}{a_2} \right). \]

By the assumption in theorem 3.1, we have \( \mathbb{E} \log(K + 1) \leq C(r)K^r \infty < \infty \) for some constant \( C(r) \)
depend on \( r \), hence
\[ \sum_{k=1}^{\infty} p_k(k) \int g_k(t) \mid h_{k,F_0,F_m}(t) \mid dt < \infty. \]

Therefore, by the dominated convergence theorem,
\[ \sum_{k=1}^{\infty} p_k(k) \int g_k(t)h_{k,F_0,F_m}(t)dt \to 0. \]

### 7.1 Proof of lemma 3.3

**Proof** By lemma 7.1, for any \( \eta > 0 \) there exists a sequence of piecewise linear concave distribution functions \( (F_m) \) such that
\[ \sum_{k=1}^{\infty} p_k(k) \int g_k(t)h_{k,F_0,F_m}(t)dt < \eta/2 \quad (7.2) \]
for all \( m \) big enough. Recall definition (7.1), \( f_m(x) = \sum_{i=1}^{m^2} w_i \varphi(x, \frac{i}{m}) = \int \varphi(x, \theta) dP_m(\theta) \), where \( P_m(\cdot) = \sum_{i=1}^{m^2} w_i \delta_{i/m}(\cdot) \). Without loss of generality, assume \( w_i > 0 \) for all \( i = 1, \ldots, m^2 \). Given \( m \) fixed, for some \( 0 < \epsilon < \min(1, e^{\eta/4} - 1) \), define a discrete probability measure \( P_{m,\epsilon}(\cdot) = \sum_{i=1}^{m^2} w_i \delta_{(i+\epsilon)/m}(\cdot) \). Moreover, define the bounded Lipschitz distance on the set of probability measure on \([0, \infty)\) by
\[ d_{BL}(P,Q) = \sup_{\psi \in C_1} \left| \int \psi dP - \int \psi dQ \right|, \]
where \( C_1 \) denotes the set of Lipschitz continuous functions on \([0, \infty)\) with Lipschitz constant 1. Then \( d_{BL} \) induces the weak topology (See Appendix A.2 in Ghosal & Van der Vaart (2017)). Choose \( 0 < \delta \leq \frac{\epsilon}{4m} \min_{1 \leq i \leq m^2} w_i \) and define the open set
\[ \Omega_m = \{ P \in \mathcal{M} : d_{BL}(P, P_{m,\epsilon}) < \delta \}. \]

Choose Lipschitz continuous functions \( \psi_j, j = 1, \ldots, m \) with compact support \([\frac{j}{m}, \frac{j+\epsilon}{m}]\), satisfying
\[ \psi_j(\theta) = \frac{\epsilon}{4m} \text{ if } \theta \in \left( \frac{j+\epsilon}{m}, \frac{j+\epsilon}{m} \right) \text{ and } 0 \leq \psi_j \leq \frac{\epsilon}{4m}. \]
Denote \( U_j = \left[ \frac{j}{m}, \frac{j+\epsilon}{m} \right], j = 1, \ldots, m^2 \). Then for any \( P \in \Omega_m, j = 1, \ldots, m^2 \), we have
\[ \left| \int \psi_j dP - \int \psi_j dP_{m,\epsilon} \right| \leq d_{BL}(P, P_{m,\epsilon}) < \delta. \]

It also follows that for \( j = 1, \ldots, m^2 \),
\[ \frac{\epsilon}{4m} P(U_j) \geq \int \psi_j dP \geq \int \psi_j dP_{m,\epsilon} - \delta \]
\[ \geq \frac{\epsilon}{4m} \int_{(j+\epsilon)/m}^{(j+\epsilon+1)/m} 1dP_{m,\epsilon} - \delta = \frac{\epsilon}{4m} \frac{w_j - \delta}{4m} \geq \frac{\epsilon}{4m} e^{-\eta/4} w_j. \]
That is \( P(U_j) \geq e^{-n/4} w_j \), for \( j = 1, \ldots, m^2 \). Using this lower bound and the mixture representation (2.4), we have for any \( x \geq 0 \), \( P \in \Omega_m, \\

\frac{f_m(x)}{f_P(x)} \leq \frac{\sum_{i=1}^{m^2} w_i \varphi(x, \frac{x}{m})}{\sum_{i=1}^{m^2} \int U_i \varphi(x, \theta) dP(\theta)} \leq \frac{\sum_{i=1}^{m^2} w_i \mathbb{1}_{\{x \leq \frac{x}{m}\}}}{\sum_{i=1}^{m^2} m^2 \mathbb{1}_{\{x \leq \frac{x}{m}\}} P(U_j)} \leq (1 + \epsilon) e^{n/4} \leq e^{n/2}.

As this implies

\[
F_m(t_j) - F_m(t_{j-1}) = \int_{t_{j-1}}^{t_j} f_m(x) dx \leq e^{n/2} \int_{t_{j-1}}^{t_j} f_P(x) dx = e^{n/2} (F_P(t_j) - F_P(t_{j-1})),
\]

we have that

\[
h_{k,F_m,F_P}(t) = \sum_{j=1}^{k+1} (F_0(t_j) - F_0(t_{j-1})) \log \frac{F_m(t_j) - F_m(t_{j-1})}{F_P(t_j) - F_P(t_{j-1})} \\
\leq \frac{n}{2} \sum_{j=1}^{k+1} (F_0(t_j) - F_0(t_{j-1})) \leq \eta/2. \tag{7.3}
\]

Note that \( h_{k,F_0,F_P}(t) = h_{k,F_0,F_m}(t) + h_{k,F_m,F_P}(t) \). Combining inequalities (7.2) and (7.3), we have

\[
\sum_{k=1}^{\infty} p_k(k) \int g_k(t) h_{k,F_0,F_P}(t) dt < \eta.
\]

That means \( \{F_P \in \mathcal{F} : P \in \Omega_m\} \subset S(\eta) \). Since \( \Omega_m \) is an open weak neighborhood of \( P_m' \) in the neighborhood \( a \) and support(\( \Pi^* \)) = \( \mathcal{M} \), we have \( \Pi^*(\Omega_m) > 0 \).

Recall that the prior \( \Pi \) on \( \mathcal{F} \) is induced by the prior \( \Pi^* \) on \( \mathcal{M} \) and the mixture representation (2.4), therefore \( \Pi(S(\eta)) \geq \Pi^*(\Omega_m) > 0 \).

\[\square\]

### 7.2 Proof of lemma 3.4

**Proof** We construct a test function depending on data \( D_n \). For any \( \epsilon > 0 \), define the event \( A_n = \{d_n(\hat{F}_n, F_0) \geq \epsilon/2\} \), where \( \hat{F}_n \) is the maximum likelihood estimator of the underlying distribution based on observations \( D_n \) (see Theorem 3 in Dümbgen, Freitag & Jongbloed (2006)) and \( d_n \) is defined as (3.1). Define \( \Phi_n = 1\{A_n\} \), then as \( n \to \infty \),

\[
\mathbb{E}_0 \Phi_n = \mathbb{E}_{K,T} \{ \mathbb{E}_{F_0}[\Phi_n|K,T]\} = \mathbb{E}_{K,T} \{ \mathbb{P}_{F_0}[d_n(\hat{F}_n, F_0) \geq \epsilon/2|K,T]\} \to 0 \tag{7.4}
\]

The final step holds because the consistency of \( \hat{F}_n \), \( \mathbb{P}_{F_0}[d_n(\hat{F}_n, F_0) \geq \epsilon/2|K,T] \to 0 \) and this probability is bounded by 1. Similarly, given \( (K,T) \), for all \( F \in U_\epsilon \\

\[
\mathbb{E}_F[1 - \Phi_n|(K,T)] = \mathbb{P}_F[\{d_n(\hat{F}_n, F_0) \leq \epsilon/2\} \cap \{d_n(F, F_0) > \epsilon\}](K,T)] \\
\leq \mathbb{P}_F[d_n(F, F_0) - d_n(\hat{F}_n, F_0) \geq \epsilon/2](K,T)] \\
\leq \mathbb{P}_F[d_n(F, \hat{F}_n) \geq \epsilon/2](K,T)]
\]

Then it is sufficient to prove for any \( \epsilon > 0 \),

\[
\mathbb{E}_{(K,T)} \left\{ \sup_{F \in U_\epsilon} \mathbb{P}_F[d_n(F, \hat{F}_n) > \epsilon|(K,T)] \right\} \to 0.
\]

23
We state that
\[ \sup_{F \in U} \mathbb{P}_F[\epsilon_n(F, \hat{F}_n) > \epsilon] \rightarrow 0. \] (7.5)

Then (7.4) and (7.5) are equivalent to the existence of a uniformly exponentially consistent test for testing \( H_0 : F = F_0 \) versus \( H_1 : F \in U \), (see Proposition 4.4.1 in Ghosh & Ramamoorthi (2003)).

Now we show the inequality (7.5) holds. For a fixed \( F \in \mathcal{F} \), the consistency result in Dümbgen, Freitag & Jongbloed (2006) claims that \( d_n(F, \hat{F}_n) \rightarrow 0 \). Actually, they proved that \( \mathbb{P}_F[d_n(F, \hat{F}_n) > \epsilon] \rightarrow 0 \) given the censoring times \((K, T)\). We checking all steps of the proof in Dümbgen, Freitag & Jongbloed (2006), the consistency is follows from the finite expectation of \( K \) and the bound \( F \leq 1 \).

Define
\[ H^2(F, G) = (2n)^{-1} \sum_{i,j}(F_{i,j} - G_{i,j})^2. \]

The consistency result is follows from the following steps:

1. \( d_n(F, \hat{F}_n) \leq 8^{1/2}H(F, \hat{F}_n) \);
2. \( H(F, \hat{F}_n)^2 \leq n^{-1} \sum_{i,j}(\Delta_{i,j} - F_{i,j})(\hat{F}_{n,i,j} - F_{i,j})^{1/2} \);
3. \( n^{-1} \sum_{i,j}(\Delta_{i,j} - F_{i,j})(\hat{F}_{n,i,j} - F_{i,j})^{1/2} \leq \sup_{G \in \mathcal{F}}[\sum_{i}(\psi_i(G) - \mathbb{E}_F\psi_i(G))]; \)

where \( \psi_i(G) = n^{-1} \sum_j \Delta_{i,j}(G_{i,j} - F_{i,j})^{1/2} \). Hence, it is sufficient to show
\[ \mathbb{P}_F \left( \sup_{G \in \mathcal{F}} \left| \sum_i(\psi_i(G) - \mathbb{E}_F\psi_i(G)) \right| > \epsilon \right) \rightarrow 0. \]

By theorem 7.2, this is a consequence of the following conditions: for some sequences \( \delta_n \rightarrow 0, b_n \rightarrow 0 \),

\[ \mathbb{E}_F \sum_{i=1}^n \sup_{G \in \mathcal{F}} |\psi_i(G)| = O(1), \] (7.6)
\[ \mathbb{E}_F \sum_{i=1}^n 1 \{ \sup_{G \in \mathcal{F}} |\psi_i(G)| > \delta_n \} \sup_{G \in \mathcal{F}} |\psi_i(G)| \rightarrow b_n, \] (7.7)

for any \( u > 0 \), \( \log \mathcal{N}(u, \mathcal{F}, \rho_n) \leq c(u) \). (7.8)

where
\[ \mathcal{N}(u, \mathcal{F}, \rho_n) = \min \left\{ \# \mathcal{G} : \mathcal{G} \subset \mathcal{F}, \inf_{G \in \mathcal{G}} \rho_n(G, G') \leq u \text{ for all } G \in \mathcal{F} \right\}, \]

and
\[ \rho_n(G, G') = \sum_{i=1}^n |\psi_i(G) - \psi_i(G')|. \]

We first give the main inequalities to derive these conditions. For (7.6),
\[ \mathbb{E}_F \sum_{i=1}^n \sup_{G \in \mathcal{F}} |\psi_i(G)| \leq n^{-1} \sum_{i}(K_i + 1)^{1/2}. \]

For (7.7),
\[ \mathbb{E}_F \sum_{i=1}^n 1 \{ \sup_{G \in \mathcal{F}} |\psi_i(G)| > \delta_n \} \sup_{G \in \mathcal{F}} |\psi_i(G)| \leq n^{-1} \sum_{i}(K_i + 1)^{\kappa}(n\delta_n)^{-2\kappa} \rightarrow 0, \]

where \( \kappa \in (0, 1/2) \), recall that \( EK^\kappa < \infty \) and choosing \( n\delta_n \rightarrow \infty \). As for (7.8), \( \rho_n \) can be bounded by a finite measure, hence
\[ \log \mathcal{N}(u, \mathcal{F}, \rho_n) \leq Cu^{-1} \]
for some constant $C$. (For more details see the proof of Theorem 3 in Dümbgen, Freitag & Jongbloed (2006)). Hence,

$$b_n = n^{-1} \sum_i (K_i + 1)^r(n\delta_n)^{-2r}, c(u) = Cu^{-1}.$$

By equation (7.12), we have

$$\mathbb{P} \left\{ \sup_{G \in \mathcal{F}} \left| \sum_i \left( \psi_i(G) - \mathbb{E}_F \psi_i(G) \right) \right| > \epsilon \right\} < 4\epsilon^{-1}b_n + 128Ce^{-1}e^{-1}\exp \left( -\frac{\epsilon^2}{512n\delta_n^2} \right)$$

Note that the right side do not depend on $F$, hence the inequality (7.5) holds.

\[\square\]

### 7.3 A technical result for proving uniform convergence

The following theorem follows from theorem 8.2 in Pollard (1990).

**Theorem 7.2** Let $f_1(w, t), f_2(w, t), \ldots, f_n(w, t)$ be independent processes with integrable envelopes $F_1(w), F_2(w), \ldots, F_n(w)$. If for each $\epsilon > 0$,

1. there is a sequence $\delta_n \to 0$ such that

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}F_i 1\{F_i > \delta_n\} < \epsilon, \text{ for all } n,$$

2. log $N(u, \mathcal{F}_{nw}, \rho_n) = c(u)$,

then

$$\sup_t \left| \sum_{i=1}^n (f_i(w, t) - \mathbb{E}f_i(w, t)) \right| \to 0 \text{ in probability.}$$

Here $N(u, \mathcal{F}_{nw}, \rho_n)$ is the covering number of $\mathcal{F}_{nw}$ with distance

$$\rho_n = \rho_n(t, t') = \sum_{i=1}^n |f_i(w, t) - f_i(w, t')|.$$

**Proof** Define event $A_{n, i} := \{F_i > \delta_n\}$, then we split the expectation into two parts:

$$\sup_t \left| \sum_i (f_i(w, t) - \mathbb{E}f_i(w, t)) \right| \leq \sup_t \left| \sum_i (f_i(w, t)1\{A_{n, i}\} - \mathbb{E}f_i(w, t)1\{A_{n, i}\}) \right|$$

$$+ \sup_t \left| \sum_i (f_i(w, t)1\{A_{n, i}^c\} - \mathbb{E}f_i(w, t)1\{A_{n, i}^c\}) \right|$$

For the first item in the right side, by the condition 1, we have

$$\mathbb{P} \left\{ \sup_t \left| \sum_i (f_i(w, t)1\{A_{n, i}\} - \mathbb{E}f_i(w, t)1\{A_{n, i}\}) \right| > \epsilon/2 \right\}$$

$$\leq 2\epsilon^{-1}\mathbb{E} \left\{ \sup_t \left| \sum_i (f_i(w, t)1\{A_{n, i}\} - \mathbb{E}f_i(w, t)1\{A_{n, i}\}) \right| \right\}$$

$$\leq 4\epsilon^{-1}\mathbb{E} \left\{ \sum_t \sup_i f_i(w, t)1\{A_{n, i}\} \right\} = 4\epsilon^{-1}b_n$$

25
For the second item, denote \( f_i^* = f_i 1\{A_{n,i}^c\} \). Using symmetrization, we have

\[
P \left\{ \sup_t \left| \sum_i (f_i^* (w, t) - \mathbb{E} f_i^* (w, t)) \right| > \epsilon / 2 \right\} \leq 4 \mathbb{E} \left\{ \sup_t \left| \sum_i f_i^* (w, t) \right| > \epsilon / 8 \right\},
\]

where \( \sigma_i = 1 \) or \(-1 \) with probability \( 1/2 \) independently. By the definition of covering number \( \mathcal{N}(\epsilon/16, \mathcal{F}_n, \rho_n) \), given \( w \), for each \( t \) in \( \mathcal{F}_n \), there exists \( t' \) such that the distance \( \rho_n (t, t') \leq \epsilon/16 \). Then we have

\[
P \left\{ \sup_t \left| \sum_i \sigma_i f_i^* (w, t) \right| > \epsilon / 8 \right\} \leq \mathbb{P} \left\{ \max_{t'} \left| \sum_i \sigma_i f_i^* (w, t') \right| + \rho_n (t, t') > \epsilon / 8 \right\}
\leq \mathbb{P} \left\{ \max_{t'} \left| \sum_i \sigma_i f_i^* (w, t') \right| > \epsilon / 16 \right\}
\leq N(\epsilon/16, \mathcal{F}_n, \rho_n) \max_{t'} \mathbb{P} \left\{ \left| \sum_i \sigma_i f_i^* (w, t') \right| > \epsilon / 16 \right\} \tag{7.10}
\]

By the Hoeffding’s inequality and \( f_i^* (w, t') \leq \delta_n \), we further have

\[
P \left\{ \left| \sum_i \sigma_i f_i^* (w, t') \right| > \epsilon / 16 \right\} \leq 2 \exp \left( -\frac{2\epsilon^2 / 16^2}{\sum_i (2f_i^* (w, t'))^2} \right) \leq 2 \exp \left( -\frac{\epsilon^2}{512n\delta_n^2} \right) \tag{7.11}
\]

Therefore, combining inequalities (7.9), (7.10) and (7.11), we have

\[
P \left\{ \sup_t \left| \sum_i (f_i (w, t) - \mathbb{E} f_i (w, t)) \right| > \epsilon \right\} \leq 4\epsilon^{-1} + 8C(\epsilon/16)\epsilon^{-1} \exp \left( -\frac{\epsilon^2}{512n\delta_n^2} \right). \tag{7.12}
\]

By choosing \( n\delta_n^2 \to 0 \), we have the right side tend to 0.

\[\square\]

References

Calle, M.L. and Gómez, G. (2001). Nonparametric Bayesian estimation from interval-censored data using Monte Carlo methods. *Journal of Statistical Planning and Inference* **98**, p. 73–87.

Doss, H., and Huffer F.W. (2003). Monte Carlo Methods for Bayesian Analysis of Survival Data Using Mixtures of Dirichlet Process Priors. *Journal of Computational and Graphical Statistics* **12**, p. 282-307.

Dümbgen, L., Freitag, S., and Jongbloed, G. (2004). Consistency of Concave Regression, With an Application to Current Status Data. *Mathematical Methods of Statistics* **13**, p. 69-81.

Dümbgen, L., Freitag, S., and Jongbloed, G. (2006). Estimating a Unimodal Distribution From Interval-Censored Data. *Journal of the American Statistical Association* **101**, p. 1094–1106.

Feller, W. (1966). An Introduction to Probability Theory and Its Applications. *Vol. II, John Wiley and Sons, New York."

Ferguson, T.S. (1973). A Bayesian Analysis of Some Nonparametric Problem. *The Annals of Statistics* **1**, p. 209–230.

Finkelstein, D. M. and Wolfe, R. A. (1986). Isotonic Regression for interval censored survival data using an E-M algorithm. *Communications in Statistics: Theory and Methods* **15**, p. 2493–2505.

van de Geer, S. (2000) *Empirical processes in M-estimation*. Cambridge University Press.

26
Ghosh, J.K. and Ramamoorthi, R.V. (2003). Bayesian Nonparametrics. *Springer Series in Statistics.*

Ghosh, J.K., Ramamoorthi, R.V., and Srikant, K.R. (1999). Bayesian analysis of censored data. *Statistics and Probability Letters* 41, p. 255-265.

Ghosal, S., and Van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference.* Cambridge University Press.

Gómez, G., Calle, M.L., and Oller, R. (2004). Frequentist and Bayesian approaches for interval-censored data. *Statistical Papers* 45, p. 139–173.

Gómez, G., Calle, M.L., Oller, R. and Langohr, K. (2009). Tutorial on methods for interval-censored data and their implementation in R. *Statistical Modelling* 9, p. 259-297.

Groeneboom, P. and Jongbloed, G. (2015). Nonparametric isotonic confidence intervals for monotone functions. *The Annals of Statistics* 43, p. 2019–2054.

Groeneboom, P., and Wellner, J.A. (1992). Information Bounds and Nonparametric Maximum Likelihood Estimation. Basel: Birkhäuser.

Hansen, M.B., and Lauritzen, S.L. (2002). Nonparametric Bayes inference for concave distribution functions. *Statistica Neerlandica* 56, p. 110-127.

Alejandro Jara and Timothy Hanson and Fernando Quintana and Peter Miller and Gary Rosner (2011). DPpackage: Bayesian Semi- and Nonparametric Modeling in R. *Journal of Statistical Software, Articles* 40(5), p. 1–30.

Jongbloed, G., van der Meulen, F., and Pang, L. (2020). Bayesian estimation of a decreasing density. arXiv:1801.02539, to appear in Brazilian Journal of Statistics.

Keiding, N., Begtrup, K., Scheike, T. H. and Hasibeder, G. (1996). Estimation from Current Status Data in Continuous Time. *Lifetime Data Anal.* 2, p. 119–129.

Krachey, E.C. (2009) Variations on the Accelerated Failure Time Model: Mixture Distributions, Cure Rates, and Different Censoring Scenarios. PhD in Statistics, North Carolina State University.

Neal, R.M. (2000). Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics* 9, p. 249–265.

Pollard, D. (1990). *Empirical Processes: Theory and Applications.* Haywood, CA: IMS.

Schick, A., and Yu, Q. (2000). Consistency of the GMLE With Mixed Case Interval-Censored Data. *Scandinavian Journal of Statistics* 27, p. 45-55.

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

Wellner, J.A., and Zhan, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association* 92, p. 945–959.

Wellner, J.A., and Zhang, Y. (2000). Two Estimators of the Mean of a Counting Process With Panel Count Data. *The Annals of Statistics* 28, p. 779-814.

Wu, Y. Ghosal, S. (2008). Kullback Leibler property of kernel mixture priors in Bayesian density estimation. *Electron. J. Statist.* 28, p. 298–331.