Multiscale Bayesian survival analysis

Ismaël Castillo\textsuperscript{1} \quad Stéphanie van der Pas\textsuperscript{2}

\textsuperscript{1}Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université
\textsuperscript{2}Mathematical Institute, Leiden University

October 27, 2020
The survival function

\[ T : \text{event time}. \]

Survival function: \( P(T > t) \)
How long until...

... a patient treated in the ICU for COVID-19 dies, with or without treatment with tocilizumab?

The Lancet, August 14, 2020;
https://doi.org/10.1016/S2665-9913(20)30277-0
Estimating the survival function

100 patients, summary:

<table>
<thead>
<tr>
<th>Time</th>
<th>Still alive</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>1</td>
<td>42</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

Estimated survival
Estimating the survival function

100 patients, summary:

<table>
<thead>
<tr>
<th>Time</th>
<th>Still alive</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>1</td>
<td>42</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

Estimated survival

![Estimated survival graph](attachment:image.png)
Estimating the survival function

100 patients, summary:

<table>
<thead>
<tr>
<th>Time</th>
<th>Still alive</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>1</td>
<td>42</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

![Estimated survival graph](image)
Censoring

Source: VSNU, jaaropgaven promovendi 2014

Censoring

A survival time is **censored** if the time of event is **not** observed. It is only known that the event did not happen before the person was lost to follow-up.

Possible reasons for censoring:
- **Study ends** before the individual experiences the event.
- **Person no longer shows up** for appointments.
- **Person experiences a competing event**, e.g. death of another cause.
The Kaplan-Meier estimator

1. Protein measurement with the folin phenol agent

11. Kaplan-Meier

24. Cox model

Counted in 2014.
The Kaplan-Meier estimator with confidence intervals

Estimated survival
The Kaplan-Meier estimator with confidence intervals
Investigate Bayesian nonparametric methods.
Example of a **parametric** model:

*The survival function is given by* \( S(t) = e^{-\lambda t} \).

Example of a **nonparametric** model:

*The survival function is Hölder continuous with exponent* \( \alpha > 1/2 \).
Bayes’ theorem

For two events $A$ and $B$ such that $\mathbb{P}(B) \neq 0$:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}.$$

Example: denote by

- $HIV^+$ the event that a person is infected with HIV;
- $T^+$ the event that a person has a positive test result in an enzyme immunoassay.

The test has a sensitivity of 99.7% and a specificity of 98.5%.
Bayes’ theorem - example

In 2005, HIV prevalence in the US was 0.6% (US Preventive Services Task Force).

\[
\Pr(HIV^+ \mid T^+) = \frac{\Pr(T^+ \mid HIV^+)\Pr(HIV^+)}{\Pr(T^+ \mid HIV^+)\Pr(HIV^+) + \Pr(T^+ \mid HIV^-)\Pr(HIV^-)}
\]

\[
= \frac{0.997 \cdot 0.006}{0.997 \cdot 0.006 + 0.015 \cdot 0.994}
\]

\approx 0.29.
Bayes’ theorem - continuous version

Let $\theta \in \Theta$ be the parameter of interest.

\[
\pi(\theta \mid \text{data}) = \frac{f_\theta(\text{data})\pi(\theta)}{\int_\Theta f_\theta(\text{data})\pi(\theta)\,d\theta},
\]

where $f_\theta$ is the likelihood.
Posterior concentration

Typical goals: recovery and uncertainty quantification.

Prior

Posterior, $n = 10$, $h = 6$

Posterior, $n = 100$, $h = 60$
Posterior concentration

Typical goals: recovery and uncertainty quantification.
Typical statement for posterior concentration

The posterior distribution contracts at rate $\varepsilon_n$ at parameter $\theta_0$ if

$$\Pi(\theta : \|\theta - \theta_0\| \geq M_n\varepsilon_n \mid X^n) \to 0$$

in $P_{\theta_0}$-probability, for every $M_n \to \infty$. 
The posterior distribution contracts at rate $\varepsilon_n$ at parameter $\theta_0$ if

$$\Pi(\theta : \|\theta - \theta_0\| \geq M_n\varepsilon_n \mid X^n) \to 0$$

in $P_{\theta_0}$-probability, for every $M_n \to \infty$.

- Contraction at rate $\varepsilon_n$ implies the existence of consistent point estimators with the same convergence rate.
- Indication of usefulness of credible sets.
- For regular parametric models, the rate is $n^{-1/2}$. 
Credible sets

A **credible set** is a set $C_n(X^n)$ that contains a prescribed proportion of the posterior mass. E.g.

$$\Pi(\theta \in C_n(X^n) \mid X^n) = 0.95.$$  

Questions:

- Is it also a **confidence set**?
  That is, does
  $$\mathbb{P}_{\theta_0}(\theta_0 \in C_n(X^n)) = 0.95$$ hold?

- What is the **size** of the credible set?
Size of a credible set

TAKING A LOOK AT TOMORROW’S WEATHER...

THE HIGH TEMPERATURE WILL BE BETWEEN 40 BELOW ZERO AND 200 ABOVE!

THIS GUY’S NEVER WRONG
Posterior is approximated very well by a normal distribution centered at $\frac{h}{n}$ with variance $\frac{\theta_0(1-\theta_0)}{n}$. 
Justifying uncertainty quantification: Bernstein von Mises theorems

Posterior is approximated very well by a normal distribution centered at \( \frac{h}{n} \) with variance \( \frac{\theta_0(1-\theta_0)}{n} \).
Typical statement for Bernstein von Mises theorems

Let $\mathcal{L}(\sqrt{n}(\theta - \hat{\theta}_n) \mid X^{(n)})$ denote the law of the recentered and rescaled posterior distribution.

Let $\mathcal{L}(G_0)$ denote the law of some fixed limiting distribution.

Then the Bernstein von Mises phenomenon may be described by:

$$\beta_S(\mathcal{L}(\sqrt{n}(\theta - \hat{\theta}_n) \mid X^{(n)}) , \mathcal{L}(G_0)) \rightarrow_{P_0} 0$$

as $n \to \infty$.

Here $(S, d)$ is a separable metric space and $\beta$ is the bounded Lipschitz metric, which metrises weak convergence of probability measures on $S$:

$$\beta_S(\mu, \nu) = \sup_{u \in BL(1)} \left| \int_S u(s)(d\mu - d\nu)(s) \right| ,$$

$$BL(1) = \left\{ f : S \to \mathbb{R}, \sup_{s \in S} |f(s)| + \sup_{s \neq t, s, t \in S} |f(s) - f(t)|/d(s, t) \leq 1 \right\} .$$
Research question: for which priors is this practice justified?

Pointwise intervals

Credible band
Survival model

\[ T : \text{event time}; \ C : \text{censoring time}. \]
\[ Y = \min\{T, C\}; \ \delta = 1\{T \leq C\}. \]

Independent right censoring.

We observe \( n \) independent pairs \( (Y_1, \delta_1), \ldots, (Y_n, \delta_n) \).

Survival function: \( P(T > t) \)
Survival objects

Hazard function
Cumulative hazard
Survival

\[ \lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}. \]

\[ \Lambda(t) = \int_0^t \lambda(u) du. \]

\[ S(t) = e^{-\Lambda(t)}. \]
Hazard: CBS data of people born in 1935
The piecewise exponential model

Hazard function

Independent and dependent histogram priors

Independent

 dependent

Dependent
Rough hazard, smooth survival

Hazard function

Cumulative hazard

Survival

Hazard function

Cumulative hazard

Survival
Example question: how to set parameters?

Hazard function

Hazard function
Our approach

Our setting is non-conjugate and the techniques potentially apply to many priors.

**Key result:** establish convergence at the minimax rate in $\| \cdot \|_\infty$-norm. Using tools from Castillo (2012), Castillo (2014), Castillo & Rousseau (2015).
Major steps in the proof for the $\| \cdot \|_\infty$-rate

**Key result:** establish convergence at the minimax rate in $\| \cdot \|_\infty$-norm.

- Prove $L_1$-concentration;
- Establish a BvM result for linear functionals $\langle b, \lambda \rangle_2 = \int_0^T b(u)\lambda(u)du$; where $b$ is a bounded function.
- Relate $\| \cdot \|_\infty$ to $\{\langle \psi_k, \lambda \rangle_2\}$ (Haar wavelets) and use tight bounds from the previous step on this collection.
Our approach

Our setting is non-conjugate and the techniques potentially apply to many priors.

**Key result:** establish convergence at the minimax rate in $\| \cdot \|_\infty$-norm.

Next, obtain nonparametric BvM in appropriately weighted multiscale space, and use continuous mapping.
As in Castillo & Nickl (2014).
Assumptions

- There exists $\tau > 0$ such that $S_0(\tau) > 0$ and $P_{\lambda_0}(C \geq \tau) > 0$. We take $\tau = 1$.
- There exists $\rho > 0$ such that $\lambda_0(t) \geq \rho$ for all $t \in [0, \tau]$.
- There exist $c_1, c_2 > 0$ such that $\Lambda_0(\tau) < c_1$, $\|\lambda_0\|_\infty < c_2$.
- $C$ admits a continuous density $g$ with respect to the Lebesgue measure on $[0, \tau)$.
- There exists $\rho' > 0$ such that $g(t) \geq \rho'$ for all $t \in [0, \tau)$. 
Smoothness assumption and limiting distributions

**Smoothness assumption on the hazard**
For $\beta, D > 0$ and $l$ the largest integer smaller than $\beta$, let $\mathcal{H}(\beta, D)$ be the standard Hölder class.

$$\mathcal{H}(\beta, D) = \{ f : |f^{(l)}(x) - f^{(l)}(y)| \leq D|x - y|^{\beta - l}, x, y \in [0, 1] \}$$

We assume: $\log \lambda_0 \in \mathcal{H}(\beta, D)$ for some $0 < \beta \leq 1$.

**Limiting distributions**
For the cumulative hazard: Let $G_{\Lambda_0}(t) = W(U_0(t))$ with $W$ Brownian motion and $U_0(t) = \int_0^t \frac{\lambda_0}{M_0(u)} du$, where $M_0(u) = \mathbb{E}_{\lambda_0} 1\{ Y \geq u \}$.

For the survival function: Let $\Sigma_0(t) = -S_0(t)G_{\Lambda_0}(t)$. 
Nonparametric BvM for cumulative hazard $\Lambda$ and survival $S$

Let $\hat{\Lambda}_n$ be an efficient estimator of $\Lambda$, and $\hat{S}_n$ the corresponding estimator of the survival. Let $D[0, 1]$ be the space of càdlàg functions on $[0,1]$.

**Theorem.** Under the conditions on the previous slides, for the three classes of piecewise exponential priors studied in the paper, with the number of intervals set to $K_n = \lceil (n / \log n)^{1/2} \rceil$ with $\gamma < \beta + 1/2$:

$$\beta_{D[0,1]} \left( \mathcal{L}(\sqrt{n}(\Lambda - \hat{\Lambda}_n) | X^{(n)}), \mathcal{L}(G_{\Lambda_0}) \right) \to P_0 0,$$

as well as

$$\beta_{\mathbb{R}} \left( \mathcal{L}(\sqrt{n}\|\Lambda - \hat{\Lambda}_n\|_\infty | X^{(n)}), \mathcal{L}(\|G_{\Lambda_0}\|_\infty) \right) \to P_0 0.$$

and a similar result for the survival $S$ centered around $\hat{S}_n$, with limiting distribution $\Sigma_0$. 
Credible bands

Consequence of previous theorem: credible bands for $\Lambda$ and for $S$ are asymptotically optimal confidence bands.
Implications for practice

- Bayesian credible bands based on the studied piecewise constant priors are **reliable** and of **optimal size**.

- Several **examples of piecewise constant priors** with guaranteed good behaviour identified in paper.

- Proof technique can be applied to **other priors** to find out whether they yield useful uncertainty quantification as well.

- **Guideline** for the number of intervals: set $K = \lceil (n/ \log n)^{1/2} \rceil$.

- An open source **R software package** is freely available: BayesSurvival at [https://cran.r-project.org/web/packages/BayesSurvival/index.html](https://cran.r-project.org/web/packages/BayesSurvival/index.html).
Data example: dependent Gamma prior

Following Arjas and Gasbarra (1994):

\[
\lambda_1 \sim \text{Gamma}(\alpha_0, \beta_0)
\]

\[
\lambda_k \mid \lambda_1, \ldots, \lambda_{k-1} \sim \text{Gamma}(\alpha, \alpha/\lambda_{k-1}), \quad k = 2, \ldots, K.
\]

Then

\[
\mathbb{E}[\lambda_k \mid \lambda_1, \ldots, \lambda_{k-1}] = \lambda_{k-1}
\]

\[
\text{Var}(\lambda_k \mid \lambda_1, \ldots, \lambda_{k-1}) = \frac{\lambda_{k-1}^2}{\alpha}.
\]
Data example: dependent Gamma prior

North Central Cancer Treatment Group lung cancer data set, which contains $n = 228$ observations of which 63 are censored. $K = \lceil (n/ \log n)^{1/2} \rceil$, $\alpha_0 = 1.5$, $\beta_0 = 1$, $\alpha = 1$.

Hazard with two kernel–based estimators

- Cao et al.
- Mueller et al.
Data example: cumulative hazard and survival

![Survival function graph]

- Credible band

0.00
0.25
0.50
0.75
1.00
0 250 500 750 1000
Data example: cumulative hazard and survival

With Kaplan–Meier + CI's

Credible band
Data example: cumulative hazard and survival

With Hall–Wellner band

Credible band
Question from practitioners: non-proportional hazards

E.g. Laird and Olivier (1981).
References


