Semiparametric consistent estimators for ARA models under right censoring

Laurent Bordes
Université of Pau and Pays de l’Adour
joint work with Eric Beutner and Laurent Doyen

AMMSI final workshop – Grenoble
Why recurrent events?

- Relapse times in medicine studies (survival analysis);
- Repair times of an industrial system (reliability);

How the treatment or maintenance effects can be addressed?
The basic assumptions on maintenance are:

- "As Bad As Old": minimal maintenance action ⇒ Poisson process;
- "As Good As New": perfect maintenance action ⇒ Renewal process;

The reality is between these two extreme cases: imperfect maintenance models.
Virtual age models: Kijima (1989)

Maintenance times: \( 0 = X_0 < X_1 < X_2 < \cdots \);
Counting process: \( N(t) = \sum_{i \geq 1} 1\{X_i \leq t\} \);
Virtual ages: \( 0 = V_0 < V_1 < V_2 < \cdots \);

- \( \lambda(\cdot) \) is a deterministic failure rate function;
- for \( i \geq 1 \):

\[
P(X_{i+1} - X_i > t|X_1, \ldots, X_i, V_1, \ldots, V_i) = \exp \left( - \int_{V_i}^{V_i+t} \lambda(u)du \right).
\]

- It leads to an intensity function for \( N \) defined by:

\[
\lambda \left( t - (X_{N(t^-)} - V_{N(t^-)}) \right)
\]

\( \Rightarrow \) AGAN: \( V_i = 0 \) for \( i \geq 1 \);
\( \Rightarrow \) ABAO: \( V_i = X_i \) for \( i \geq 1 \);
ARA: parametrization of Virtual age models ($\theta \in [0, 1]$)

Kijima type II: $\text{ARA}_1$
- virtual ages: $V_i = (1 - \theta)(X_i - X_{i-1}) + V_{i-1}$;
- intensity: $\lambda(t - \theta X_{N(t^-)})$.

Kijima type I: $\text{ARA}_\infty$
- virtual ages: $V_i = (1 - \theta)(X_i - X_{i-1} + V_{i-1})$;
- intensity: $\lambda\left(t - \theta \sum_{j=0}^{N(t^-)-1}(1 - \theta)^j X_{N(t^-)-j}\right)$.
Given a vector of covariates $\mathbf{Z}$, the intensity of the counting process $N$ is defined by

$$Y(t) U \lambda(\varepsilon(t, \omega)) \psi(\beta^T \mathbf{Z}) \rho(N(t^-), \alpha),$$

where

- $Y(t) = 1_{\{t \leq T\}}$ is the at $\{0, 1\}$–risk process,
- $U$ is a positive random effect,
- $t \mapsto \varepsilon(t, \omega)$ is the virtual age function, possibly random, that gives the virtual age at the calendar time $t$,
- $\lambda(\cdot)$ is an unknown hazard rate function,
- $\alpha$ and $\beta$ are unknown Euclidean regression parameters,
- $\psi(\cdot)$ and $\rho(\cdot)$ are known link functions.

**Exemple:** $U \equiv 1$, $\varepsilon(t, \omega) = t - X_{N(t^-)}$, $\rho(\cdot) \equiv 1$ and $\psi(x) = \exp(x) \Rightarrow$ Cox model.
Two recent semi-/non-parametric inference result

- Peña (2014): for $U \equiv 1$, $\varepsilon(t, \omega)$ observable, $\rho(\cdot) \equiv 1$ and $\psi(x) = \exp(x) \Rightarrow$ obtain the asymptotic behavior of estimators of $\beta$ and $\Lambda(\cdot) = \int_{0}^{\cdot} \lambda(t) dt$.

- BBD (2015): the profile likelihood approach fails to give consistent estimators for the ARA$_1$ and ARA$_\infty$ submodels, that is intensities of the form

$$Y(t)\lambda(\varepsilon^\theta(t, \omega))$$

where

- $\varepsilon^\theta(t, \omega) = t - \theta X_{N(t^-;\omega)}(\omega)$ i.e. ARA$_1$,
- $\varepsilon^\theta(t, \omega) = t - \theta \sum_{j=0}^{N(t^-;\omega)-1} (1 - \theta)^j X_{N(t^-;\omega)-j}(\omega)$ i.e. ARA$_\infty$,
- $Y(t) = 1\{t \leq X_k\}$ i.e. Type-II censoring.
Profile likelihood function for ARA submodels

- Data:
  - \((N_i, Y_i)_{1 \leq i \leq n}\) be \(n\) i.i.d. copies of \((N, Y)\),
  - \(N_i(t) = \sum_{j \geq 1} 1_{\{X_{i,j} \leq t\}}\) and \(Y_i(t) = 1_{\{t \leq T_i\}}\).

- Double indexed processes (Selke and Siegmund, 1983): define \(Z_i^\theta(t, v) = 1_{\{\epsilon_i^\theta(t) \leq v\}}\) (i.e. 1 if the virtual age at calendar time \(t\) is less than \(v\)) and introduce

\[
N_i^\theta(t, v) = \int_0^t Z_i^\theta(u, v) N_i(du),
\]
and
\[
H_i^\theta(t, v) = \int_0^t Z_i^\theta(u, v) Y_i(u) \lambda(\epsilon_i^\theta(u)) du.
\]
Profile likelihood function for ARA submodels

Assume that $s$ is the total duration of the study. With a change of variables (possible under ARA submodels) we obtain a nonparametric estimator (Peña et al. from 2001) of $\Lambda$ which is a NPMLE (BBD, 2015):

$$
\Lambda_n^\theta(s, t) = \sum_{i=1}^{n} \int_{0}^{t} \frac{N_i^\theta(s, du)}{\sum_{j=1}^{n} Y_j^\theta(s, u)}
$$

where for $1 \leq i \leq n$

$$
Y_i^\theta(s, u) = \sum_{j=1}^{N_i(s-)} 1_{(\varepsilon_{i,j-1}^{\theta},\varepsilon_{i,j-1}^{\theta},\varepsilon_{i,j}^{\theta}(X_{i,j})]}(u) + 1_{(\varepsilon_{i,N_i(s-)}^{\theta},\varepsilon_{i,N_i(s-)}^{\theta},\varepsilon_{i,N_i(s-)}^{\theta}(s \wedge T_i)]}(u),
$$

with $\varepsilon_{i,j-1}^{\theta}$ the restriction of $\varepsilon_{i}^{\theta}$ to $(X_{i,j-1}, X_{i,j}]$ for $j \geq 1$. 
Profile likelihood function for ARA submodels

Log–likelihood function (Jacod formula, 1975):

\[ \ell_{s,n}(\theta, \lambda) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} \left\{ \log \lambda(u) dN_{i}^{\theta}(s, u) - Y_{i}^{\theta}(s, u) \Lambda(du) \right\}. \]

Profile log–likelihood function:

\[
\ell_{s,n}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} \left\{ \log \Lambda_{n}^{\theta}(s, \Delta u) dN_{i}^{\theta}(s, u) - Y_{i}^{\theta}(s, u) \Lambda_{n}^{\theta}(s, du) \right\}
\]

\[
= - \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{s} \log \left( \sum_{j=1}^{n} Y_{j}^{\theta}(s, \varepsilon_{i}(u)) \right) N_{i}(du) + \text{const.}
\]
Profile likelihood function for $\text{ARA}_1$ and Type-II censoring.
Profile likelihood function for ARA$_1$ and Type-I censoring

$(\theta, d, \tau) = (0.1, 2, 2.2)$

$(\theta, d, \tau) = (0.5, 2, 2.2)$

$(\theta, d, \tau) = (1, 2, 2.2)$

$(\theta, d, \tau) = (0.6, 5.9)$

$(\theta, d, \tau) = (0.5, 6.5, 9)$

$(\theta, d, \tau) = (1, 6, 5.9)$
Profile likelihood function for $\text{ARA}_\infty$ and Type-II censoring
Profile likelihood function for $\text{ARA}_\infty$ and Type-I censoring

\[
(\theta,d,\tau) = (0.1,2,2.2)
\]

\[
(\theta,d,\tau) = (0.5,2,2.2)
\]

\[
(\theta,d,\tau) = (1,2,2.2)
\]

\[
(\theta,d,\tau) = (0.6,5.9)
\]

\[
(\theta,d,\tau) = (0.5,6,5.9)
\]

\[
(\theta,d,\tau) = (1,6,5.9)
\]
An alternative approach: smoothing $\Lambda_n(s, \cdot)$

- A similar phenomenon can be observed for the accelerated failure time model for which the profile likelihood function does not even depend on the regression parameter.
- Zheng and Lin (2007) proposed to replace the pseudo-estimator of $\Lambda$ by a regularized version in the profile likelihood function. They proved that the resulting estimators are consistent and asymptotically normal.
Empirical processes notations (restricted to $\text{ARA}_1$)

The i.i.d. random elements: for $1 \leq i \leq n$ we have $T_i$ the right censoring time and $X_i = (X_{i,j})_{j \geq 1}$ the sequence generated by the model (we set $X_{i,0} \equiv 0$). Then we can write

$$N_i^\theta(s, t) = f_{\theta,t}(Z_i) \text{ and } Y_i^\theta(s, t) = g_{\theta,t}(Z_i)$$

where $Z_i = (T_i, X_i) \in \mathcal{Z} = \mathbb{R} \times \mathcal{X}$ ($\mathcal{X}$ is the set of unbounded non decreasing sequences (that is without accumulation points) on $\mathbb{R}^+$):

- $f_{\theta,t}(z) = \sum_{j \geq 1} 1\{x_j - \theta x_{j-1}; x_j \leq s \land \tau\}$,

- $g_{\theta,t}(z) = \sum_{j \geq 1} 1\{x_j(1-\theta) < t \leq x_j \land \tau - \theta x_{j-1}; x_{j-1} \leq s \land \tau\}$.

where $z = (\tau, x)$ and $x \in \mathcal{X}$. 
Empirical processes notations (restricted to ARA$_1$)

Let us define:

- the classes of functions
  \[ \mathcal{F} = \{ z \mapsto f_{\theta,t}(z); (\theta, t) \in [0, 1] \times [0, s] \}, \]
  and \[ \mathcal{G} = \{ z \mapsto g_{\theta,t}(z); (\theta, t) \in [0, 1] \times [0, s] \}. \]
- the empirical measures
  \[ \mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Z_i}, \]
  and \[ \mathcal{G}_n = \sqrt{n} (\mathbb{P}_n - P) \]
  where $Z \sim P$. 
Defining

- \( \nu_n^\theta(s, t) = \mathbb{P}_n f_{\theta, t} = \frac{1}{n} \sum_{i=1}^n f_{\theta, t}(Z_i) = \frac{1}{n} \sum_{i=1}^n N_i^\theta(s, t) \),
- \( \nu^\theta(s, t) = Pf_{\theta, t} = \mathbb{E}(f_{\theta, t}(Z)) \),
- \( y_n^\theta(s, t) = \mathbb{P}_n g_{\theta, t} = \frac{1}{n} \sum_{i=1}^n g_{\theta, t}(Z_i) = \frac{1}{n} \sum_{i=1}^n Y_i^\theta(s, t) \),
- \( y^\theta(s, t) = Pg_{\theta, t} = \mathbb{E}(g_{\theta, t}(Z)) \),

we have for \( t \in [0, s] \)

\[
\Lambda_n^\theta(s, t) = \int_0^t \frac{\nu_n^\theta(s, du)}{y_n^\theta(s, u)},
\]

and

\[
\Lambda^\theta(s, t) = \int_0^t \frac{\nu^\theta(s, du)}{y^\theta(s, u)}.
\]
Coming back to the likelihood function

\[ \ell_{s,n}(\theta, \lambda) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} \left\{ \log \lambda(u) dN_{i}^{\theta}(s, u) - Y_{i}^{\theta}(s, u) \Lambda(du) \right\} \]

\[ = \int_{0}^{s} \left\{ \log \lambda(u) \nu_{n}^{\theta}(s, du) - y_{n}^{\theta}(s, u) \Lambda(du) \right\} \]

we replace \( \lambda(\cdot) \) by

\[ \lambda_{n}^{\theta}(s, u) = \frac{1}{b_{n}} \int_{\mathbb{R}} \kappa \left( \frac{u - v}{b_{n}} \right) \Lambda_{n}^{\theta}(s, dv) = \frac{1}{b_{n}} \int_{\mathbb{R}} \kappa \left( \frac{u - v}{b_{n}} \right) \frac{\nu_{n}^{\theta}(s, dv)}{y_{n}^{\theta}(s, v)} , \]

where \( b_{n} \downarrow 0 \) and \( \Lambda(\cdot) \) by \( \Lambda_{n}^{\theta}(s, \cdot) \).
Smoothed profile likelihood function

It leads to the following estimating function

\[ \ell_n(s) = \int_0^s \log \left( \lambda^s_n(s, u) \right) \nu^s_n(s, du). \]

Then we define the estimator of \( \theta \) by

\[ \theta_n = \arg \max_{\theta \in [0,1]} \ell_n(s, \theta) \]

and estimators of \( \Lambda \) and \( \lambda(\cdot) \) by

\[ \Lambda_n(s, t) = \Lambda^\theta_n(s, t) \quad \text{and} \quad \lambda_n(s, t) = \lambda^\theta_n(s, t). \]
Smoothed profile likelihood: $\text{ARA}_1$, Type-II with $m = 2$, $\theta = 0.3$

\[ b_n = n^{-0.4} \quad b_n = n^{-0.25} \quad b_n = n^{-0.1} \]

$n = 50$

$n = 200$
Smoothed profile likelihood: $\text{ARA}_1$, Type-II with $m = 2$, $\theta = 0.7$

$$b_n = n^{-0.4}$$

$$b_n = n^{-0.25}$$

$$b_n = n^{-0.1}$$

$n = 50$

$n = 200$
Monte Carlo study: $\text{ARA}_1$, Type-II with $m = 2, \theta = 0.7$

$n = 50$
Main theoretical results for the ARA$_1$ model

Under some technical assumptions including the fact that $\lambda_0$ is continuous and strictly increasing on $[0, s]$, that $b_n = cn^{-d}$ with $d \in (0, 1/2)$, that the support of pdf of $\mathcal{T}$ is $\mathbb{R}^+$ we have:

1. As $n \rightarrow \infty$

   $$\sup_{(\theta, t) \in [0,1] \times [0, s]} \left| \frac{\lambda_n^\theta(s, t)}{\lambda^\theta(s, t)} - 1 \right| \rightarrow 0 \quad \text{a.s.}$$

2. As $n \rightarrow \infty$

   $$\sup_{\theta \in [0,1]} |\ell_{n,s}(\theta) - \ell_s(\theta)| \rightarrow 0 \quad \text{a.s.}$$

3. $\theta \mapsto \ell_s(\theta)$ is continuous in a neighborhood of $\theta_0$, and $\ell_s(\theta) < \ell_s(\theta_0)$ for all $\theta \in [0, 1] \backslash \{\theta_0\}$.

Consequences: $(\theta_n)_{n \geq 1}$ is consistent and both $\lambda_n$ and $\Lambda_n$ are uniformly consistent.
Key results: most important intermediate results

1. $\mathcal{F}$ and $\mathcal{G}$ are $P$–Donsker classes of functions using bracketing numbers: under ARA$_1$ and $\lambda_0$ non decreasing.

2. Convergence rates:

$$\sup_{(\theta,t)} \left| \nu_{n}^{\theta}(s, t) - \nu^{\theta}(s, t) \right| = o_{a.s.}(b_n),$$

and

$$\sup_{(\theta,t)} \left| y_{n}^{\theta}(s, t) - y^{\theta}(s, t) \right| = o_{a.s.}(b_n).$$

3. Identifiability: $\ell_s(\theta) = \ell_s(\theta_0) \iff \theta = \theta_0$.

Other challenges: central limit theorem and bandwidth selection, a new way to obtain asymptotic results for recurrent event models.