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?

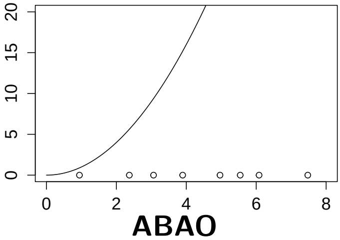
- Relapse times in medicine studies (survival analysis);
- Repair times of an industrial system (reliability);

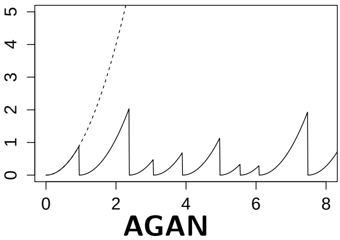
$$\Downarrow$$

How the treatment or maintenance effects can be addressed?

Maintenance effects

- 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 $N(t) = \sum_{i \geq 1} \mathbf{1}_{\{X_i \leq t\}}$;
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 $i \geq 1$;

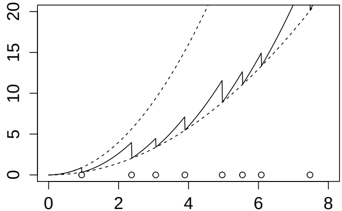
ARA: parametrization of Virtual age models ($\theta \in [0, 1]$)

Kijima type II: $\mathbf{ARA}_1$

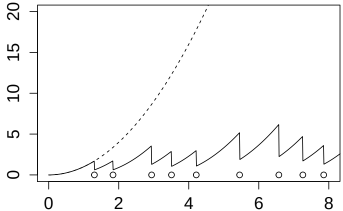
- virtual ages: $V_i = (1-\theta)(X_i - X_{i-1}) + V_{i-1}$;
- intensity: $\lambda(t - \theta X_{N(t^-)})$.

Kijima type I: $\mathbf{ARA}_\infty$

- virtual ages: $V_i = (1-\theta)(X_i - X_{i-1} + V_{i-1})$;
- intensity: $\lambda\left(t - \theta \sum_{j=0}^{N(t^-)-1} (1-\theta)^j X_{N(t^-)-j}\right)$.



$\mathbf{ARA}_1$



$\mathbf{ARA}_\infty$

A general virtual age model: Hollander and Peña (2004)

Given a vector of covariates $\boldsymbol{Z}$, the intensity of the counting process N is defined by

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

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,
- $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ 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 $\boldsymbol{\beta}$ 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_∞ 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_∞,
  - $Y(t) = \mathbf{1}_{\{t \le X_k\}}$ 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),
  - $N_i(t) = \sum_{j\geq 1} \mathbf{1}_{\{X_{i,j}\leq t\}}$ and $Y_i(t) = \mathbf{1}_{\{t\leq \mathcal{T}_i\}}$,
- Double indexed processes (Selke and Siegmund, 1983): define $Z_i^\theta(t,v) = \mathbf{1}_{\{\varepsilon_i^\theta(t)\leq v\}}$ (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

$$H_i^\theta(t,v) = \int_0^t Z_i^\theta(u,v) Y_i(u) \lambda(\varepsilon_i^\theta(u)) du.$$

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 Λ 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 $1 \leq i \leq n$

$$\begin{aligned} Y_i^\theta(s,u) &= \sum_{j=1}^{N_i(s-)} \mathbf{1}_{(\varepsilon^\theta_{i,j-1}(X_{i,j-1}+),\varepsilon^\theta_{i,j-1}(X_{i,j})]}(u) \\ &\quad + \mathbf{1}_{(\varepsilon^\theta_{i,N_i(s-)}(X_{i,N_i(s-)}+),\varepsilon^\theta_{i,N_i(s-)}(s\wedge\mathcal{T}_i)]}(u), \end{aligned}$$

with $\varepsilon^\theta_{i,j-1}$ the restriction of ε^θ_i to $(X_{i,j-1},X_{i,j}])$ for $j \geq 1$.

Profile likelihood function for ARA submodels

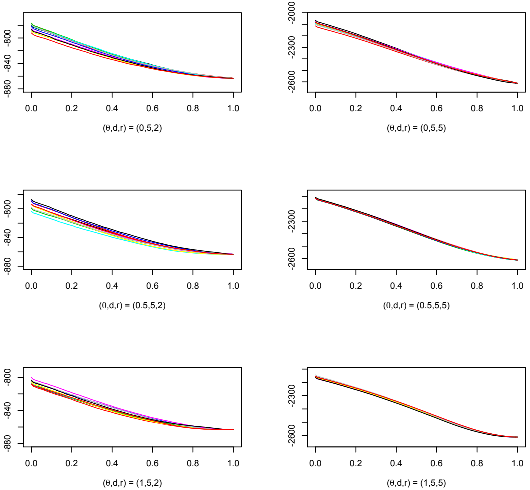
Log–likelihood function (Jacod formula, 1975):

$$\ell_{s,n}(\theta,\lambda) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}\left\{\log\lambda(u)dN_i^{\theta}(s,u) - Y_i^{\theta}(s,u)\Lambda(du)\right\}.$$

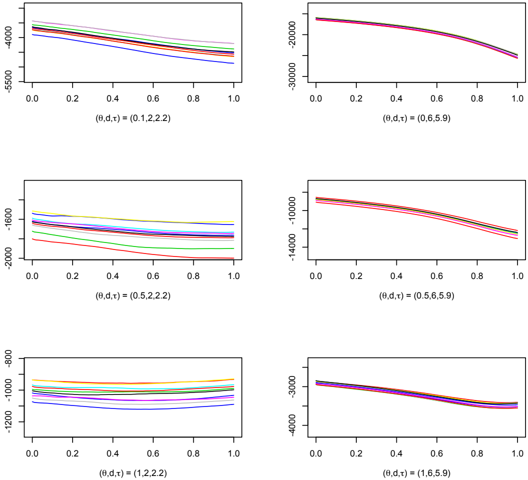
Profile log–likelihood function:

$$\begin{aligned}\ell_{s,n}(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}\left\{\log\Lambda_n^{\theta}(s,\Delta u)dN_i^{\theta}(s,u) - Y_i^{\theta}(s,u)\Lambda_n^{\theta}(s,du)\right\} \\ &= -\frac{1}{n}\sum_{i=1}^{n}\int_0^{s}\log\left(\sum_{j=1}^{n}Y_j^{\theta}(s,\varepsilon_i^{\theta}(u))\right)N_i(du) + const.\end{aligned}$$

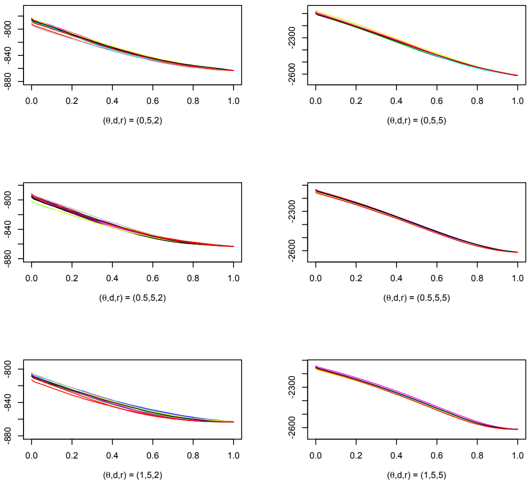
Profile likelihood function for ARA_1 and Type-II censoring



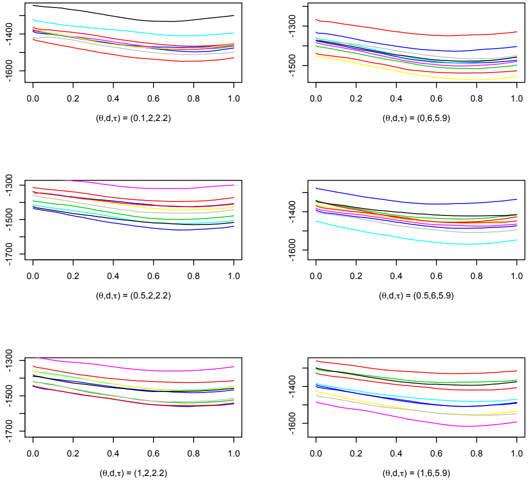
Profile likelihood function for ARA_1 and Type-I censoring



Profile likelihood function for ARA_∞ and Type-II censoring



Profile likelihood function for ARA_∞ and Type-I censoring



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 of the regression parameter.
- Zheng and Lin (2007) proposed to replace the pseudo-estimator of Λ 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 ARA_1)

The i.i.d. random elements: for $1 \le i \le n$ we have $\mathcal{T}_i$ the right censoring time and $\boldsymbol{X}_i = (X_{i,j})_{j \ge 1}$ the sequence generated by the model (we set $X_{i,0} \equiv 0$). Then we can write

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

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

- $f_{\theta,t}(\boldsymbol{z}) = \sum_{j \ge 1} \mathbf{1}_{\{x_j - \theta x_{j-1}; x_j \le s \wedge \tau\}}$,
- $g_{\theta,t}(\boldsymbol{z}) = \sum_{j \ge 1} \mathbf{1}_{\{x_j(1-\theta) < t \le x_j \wedge s \wedge \tau - \theta x_{j-1}; x_{j-1} \le s \wedge \tau\}}$.

where $\boldsymbol{z} = (\tau, \boldsymbol{x})$ and $\boldsymbol{x} \in \mathcal{X}$.

Empirical processes notations (restricted to ARA_1)

Let us define:

- the classes of functions

  $$\mathcal{F} = \{\boldsymbol{z} \mapsto f_{\theta,t}(\boldsymbol{z}); (\theta, t) \in [0,1] \times [0,s]\},$$

  and

  $$\mathcal{G} = \{\boldsymbol{z} \mapsto g_{\theta,t}(\boldsymbol{z}); (\theta, t) \in [0,1] \times [0,s]\}.$$

- the empirical measures

  $$\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\boldsymbol{Z}_i}$$

  and

  $$\mathbb{G}_n = \sqrt{n}\,(\mathbb{P}_n - P)$$

  where $\boldsymbol{Z} \sim P$.

Empirical processes notations (restricted to ARA_1)

Defining

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

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

$$\Lambda^\theta(s,t) = \int_0^t \frac{\nu^\theta(s,du)}{y^\theta(s,u)}.$$

Smoothed profile likelihood function

Coming back to the likelihood function

$$\begin{aligned}\ell_{s,n}(\theta,\lambda) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}\left\{\log\lambda(u)dN_i^{\theta}(s,u) - Y_i^{\theta}(s,u)\Lambda(du)\right\} \\ &= \int_0^{s}\left\{\log\lambda(u)\nu_n^{\theta}(s,du) - y_n^{\theta}(s,u)\Lambda(du)\right\}\end{aligned}$$

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 $b_n \searrow 0$ and $\Lambda(\cdot)$ by $\Lambda_n^{\theta}(s,\cdot)$.

Smoothed profile likelihood function

It leads to the following estimating function

$$\ell_{n,s}(\theta) = \int_0^s \log\left(\lambda_n^\theta(s,u)\right)\nu_n^\theta(s,du).$$

Then we define the estimator of θ by

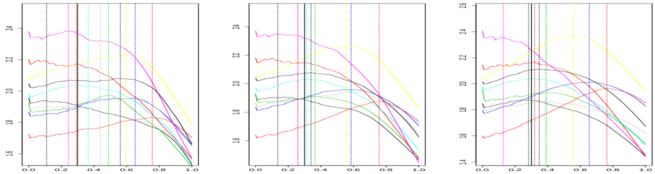
$$\theta_n = \arg\max_{\theta\in[0,1]} \ell_{n,s}(\theta)$$

and estimators of Λ and $\lambda(\cdot)$ by

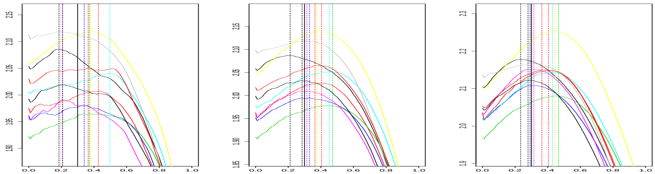
$$\Lambda_n(s,t) = \Lambda_n^{\theta_n}(s,t) \quad \text{and} \quad \lambda_n(s,t) = \lambda_n^{\theta_n}(s,t).$$

Smoothed profile likelihood: ARA_1, Type-II with $m = 2$, $\theta = 0.3$

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



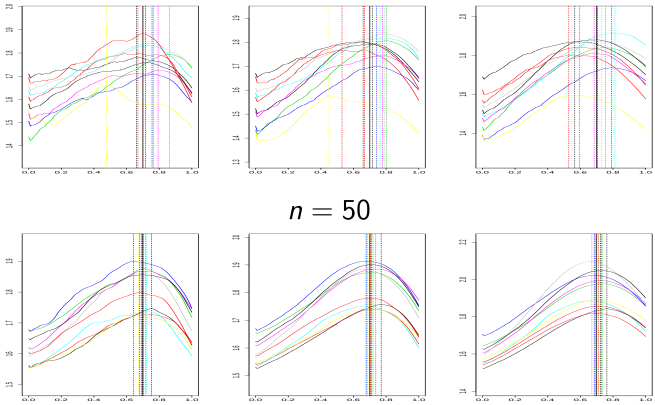
$n = 50$



$n = 200$

Smoothed profile likelihood: ARA_1, Type-II with $m = 2$, $\theta = 0.7$

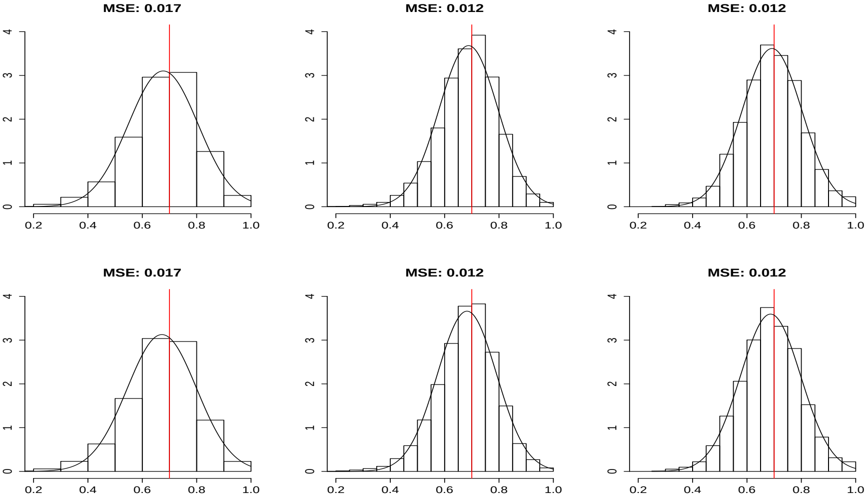
$b_n = n^{-0.4}$ $b_n = n^{-0.25}$ $b_n = n^{-0.1}$



$n = 50$

$n = 200$

Monte Carlo study: 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 λ_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 $n \to \infty$

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

2. As $n \to \infty$

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

3. $\theta \mapsto \ell_s(\theta)$ is continuous in a neighborhood of θ_0, and $\ell_s(\theta) < \ell_s(\theta_0)$ for all $\theta \in [0, 1]\backslash\{\theta_0\}$.

Consequences: $(\theta_n)_{n\geq1}$ is consistent and both λ_n and Λ_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 λ_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: $\ell_s(\theta) = \ell_s(\theta_0) \Leftrightarrow \theta = \theta_0$.

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