# Estimation à support non-compact de la fonction de risque instantané en analyse des durées de vie.

Elodie BRUNEL
IMAG, CNRS
Université de Montpellier

based on works with F. Comte

Onzième Biennale Française des Mathématiques Appliquées et Industrielles

# Hazard rate in survival analysis

- Survival analysis is a collection of statistical procedures employed on time-to-event data.
- The major limitation of time-to event-data is the possibility of incomplete data : censored observations. (Kaplan, 1958 ; Meier, 1958 ; Cox, 1972).
- The main objectives of survival analysis include :
  - analysis of patterns of time-to-event data
  - comparing survival curves (between different groups)
  - assessing the relationship between explanatory variables and survival time.

# Outline

1. Hazard rate in survival analysis
2. Definition of the estimator
   - The statistical model
   - Least squares projection estimator
3. Bounds for the empirical and integrated risk
   - Specific Compact case
   - Non-compact bases
4. Model selection and simulations

# Outline

# Hazard rate in survival analysis

Let $X$ be a lifetime of interest, a nonnegative random variable with density $f$ w.r.t the Lebesgue measure, and with cumulative distribution function $F$ and survival function $S = 1 - F$ .

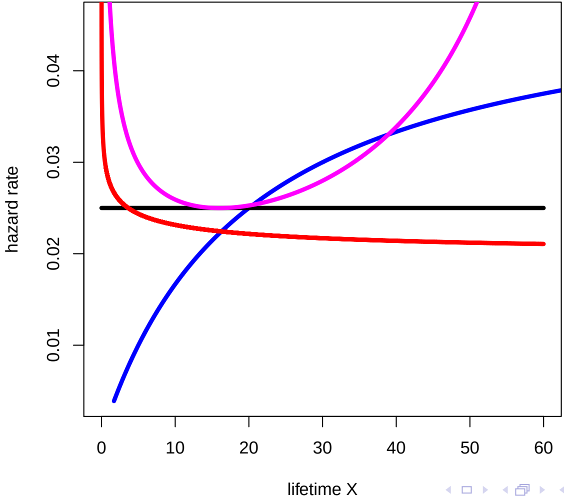
- The hazard rate function (also called instantaneous failure rate) is a good way to model data distribution in survival analysis.

The hazard rate function $\lambda$ is defined by :

$$\begin{aligned}\lambda(x) &= \lim_{\Delta\to 0} \frac{1}{\Delta}\mathbb{P}\left(x < X \leq x + \Delta | X > x\right) \\ &= \lim_{\Delta\to 0} \frac{1}{\Delta}\frac{F(x+\Delta) - F(x)}{S(x)} = \frac{f(x)}{S(x)}\end{aligned}$$

# Different shapes for the hazard rate

$$\lambda(x) = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{P}(x < X \leq x + \Delta | X > x) = \frac{f(x)}{S(x)}$$



# Outline

Aim : nonparametric estimation of the hazard rate function $\lambda = f/S$ in the presence of right-censored observations.

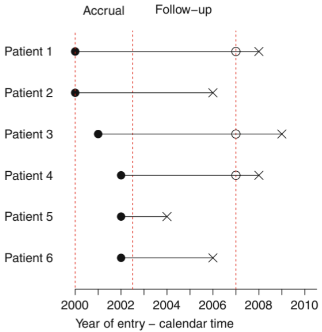
Consider the model where, instead of $X_1, \cdots, X_n$, the observations are

$Z_i = X_i \wedge C_i, \quad \delta_i = \mathbf{1}_{\{X_i \leq C_i\}},$

where the sequences $(X_i)_i$ and $(C_i)_i$ are two independent sequences of i.i.d. nonnegative absolutely continuous random variables. The $Z_i$'s are called right-censored observations, and the $\delta_i$'s are non-censoring indicators.

$\times$ indicates the event occurs before the end-study $Z_2 = X_2$ and $\delta_2 = 1$

$\circ$ the event has not been observed : it is right censored $Z_3 = C_3$ and $\delta_3 = 0$



Aim : nonparametric estimation of the hazard rate function $\lambda = f/S$ in the presence of right-censored observations.

The hazard rate function $\lambda = f/S$ of the lifetime $X$ is the function of interest.
But $X_1, \cdots, X_n$ are not completely observed
Instead, we observe $(Z_1, \delta_1), \cdots, (Z_n, \delta_n)$ with $Z_i = X_i \wedge C_i$ and $\delta_i = \mathbf{1}_{\{X_i \leq C_i\}}$ :

- $(X_i)$ and $(C_i)$ are independent sequences of i.i.d. nonnegative and absolutely continuous random variables.
- $(Z_i)$ i.i.d. with survival function $S_Z$
- $(C_i)$ i.i.d. with survival function $S_C$

We need to find a strategy to recover $\lambda$ from the observations $(Z_i, \delta_i)_{1 \leq i \leq n}$

We need to find a strategy to recover $\lambda = \frac{f}{S}$ from the observations $(Z_i, \delta_i)_{1\leq i\leq n}$.

Classical nonparametric strategies are of two types :

- Quotient estimator : very popular for kernel methods in general regression estimation. (Nadaraya, 1964 ; Watson, 1964 and Müller & al. 1994, Bouezmarni & al. , 2011, Barbeito & Cao, 2018 ; Brunel & Comte, 2005).
- Direct estimator : Least squares contrast minimization. (Barron, Birgé, Massart, 1999 ; Baraud, 2002 for general regression setting, Plancade, 2011 ; Comte & al, 2011 ; Brunel & Comte, 2021 for specific hazard rate setting)

We focus on hazard rate estimation by the least squares projection method.

How is it possible to handle the case of non compact supported bases ?

What is the interest of non compactly supported bases ?

- The estimation set and the support of the bases in projection methods are usually considered as fixed in the theoretical study.
- But, in practice you have to adjust the support to the data : so that the support becomes random !
- With a non compact support, you don't need to fix the estimation support in advance.

Let $s, t : A \mapsto \mathbb{R}$ be two square integrable functions from $A \subseteq \mathbb{R}^+$ into $\mathbb{R}$
The following criterion is called a contrast and

$$\boxed{\gamma_n(t) := \|t\|_n^2 - \frac{2}{n}\sum_{i=1}^n \delta_i t(Z_i),}$$

where the empirical inner product and its associated empirical squared norm are defined by

$$\langle s, t\rangle_n := \frac{1}{n}\sum_{i=1}^n \int s(x)t(x)\mathbf{1}_{\{Z_i>x\}}dx, \quad \|t\|_n^2 := \frac{1}{n}\sum_{i=1}^n \int t^2(x)\mathbf{1}_{\{Z_i>x\}}dx\,.$$

$$= \int s(x)t(x)\widehat{S}_{n,Z}(x)dx \qquad = \int t^2(x)\widehat{S}_{n,Z}(x)dx$$

with $\widehat{S}_{n,Z}(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{Z_i>x\}}$ the empirical survival function of $Z$

Let $s, t : A \mapsto \mathbb{R}$ be two square integrable functions from $A \subseteq \mathbb{R}^+$ into $\mathbb{R}$
The following criterion is called a contrast and

$$\boxed{\gamma_n(t) := \|t\|_n^2 - \frac{2}{n}\sum_{i=1}^n \delta_i t(Z_i),}$$

where the empirical inner product and its associated empirical squared norm are defined by

$$\begin{aligned}\langle s, t\rangle_n &:= \frac{1}{n}\sum_{i=1}^n \int s(x)t(x)\mathbf{1}_{\{Z_i>x\}}dx, \quad \|t\|_n^2 := \frac{1}{n}\sum_{i=1}^n \int t^2(x)\mathbf{1}_{\{Z_i>x\}}dx\,.\\ &= \int s(x)t(x)\widehat{S}_{n,Z}(x)dx \qquad\qquad = \int t^2(x)\widehat{S}_{n,Z}(x)dx\end{aligned}$$

with $\widehat{S}_{n,Z}(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{Z_i>x\}}$ the empirical survival function of $Z$

Why is the contrast $\gamma_n(t)$ related to our hazard rate estimation problem ?

$$\mathbb{E}(\gamma_n(t)) = \mathbb{E}\left(\|t\|_n^2 - \frac{2}{n}\sum_{i=1}^n \delta_i t(Z_i)\right)$$
$$= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\int t^2(x)\mathbf{1}_{\{Z_i>x\}}dx\right) - \frac{2}{n}\sum_{i=1}^n \mathbb{E}(\delta_i t(Z_i))$$

$\mathbb{E}\left(\int t^2(x)\mathbf{1}_{\{Z_i>x\}}dx\right) = \int t^2(x)S_Z(x)dx = \|t\|_{S_Z}^2$

$\mathbb{E}(\delta_i t(Z_i)) = \mathbb{E}(\mathbf{1}_{\{X_i\le C_i\}}t(X_i)) = \int t(x)S_C(x)f(x)dx = \int t(x)S_Z(x)\frac{f(x)}{S(x)}dx.$
(as $S_Z = S_C S$)

Therefore, we find that

$$\boxed{\mathbb{E}(\gamma_n(t)) = \|t\|_{S_Z}^2 - 2\int t(x)\lambda(x)S_Z(x)dx}$$

Why is the contrast $\gamma_n(t)$ related to our hazard rate estimation problem ?

$$\begin{aligned}\mathbb{E}(\gamma_n(t)) &= \mathbb{E}\left(\|t\|_n^2 - \frac{2}{n}\sum_{i=1}^n \delta_i t(Z_i)\right)\\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\int t^2(x)\mathbf{1}_{\{Z_i>x\}}dx\right) - \frac{2}{n}\sum_{i=1}^n \mathbb{E}(\delta_i t(Z_i))\end{aligned}$$

$\mathbb{E}\left(\int t^2(x)\mathbf{1}_{\{Z_i>x\}}dx\right) = \int t^2(x)S_Z(x)dx = \|t\|_{S_Z}^2$

$\mathbb{E}(\delta_i t(Z_i)) = \mathbb{E}(\mathbf{1}_{\{X_i\leq C_i\}}t(X_i)) = \int t(x)S_C(x)f(x)dx = \int t(x)S_Z(x)\frac{f(x)}{S(x)}dx.$
(as $S_Z = S_C S$)

Therefore, we find that

$$\boxed{\mathbb{E}(\gamma_n(t)) = \|t\|_{S_Z}^2 - 2\int t(x)\lambda(x)S_Z(x)dx}$$

Why is the contrast $\gamma_n(t)$ related to our hazard rate estimation problem ?

$$\begin{aligned}\mathbb{E}(\gamma_n(t)) &= \mathbb{E}\left(\|t\|_n^2 - \frac{2}{n}\sum_{i=1}^n \delta_i t(Z_i)\right)\\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\int t^2(x)\mathbf{1}_{\{Z_i>x\}}dx\right) - \frac{2}{n}\sum_{i=1}^n \mathbb{E}(\delta_i t(Z_i))\end{aligned}$$

$$\mathbb{E}\left(\int t^2(x)\mathbf{1}_{\{Z_i>x\}}dx\right) = \int t^2(x)S_Z(x)dx = \|t\|_{S_Z}^2$$

$$\mathbb{E}(\delta_i t(Z_i)) = \mathbb{E}(\mathbf{1}_{\{X_i\le C_i\}}t(X_i)) = \int t(x)S_C(x)f(x)dx = \int t(x)S_Z(x)\frac{f(x)}{S(x)}dx.$$

(as $S_Z = S_C S$)

Therefore, we find that

$$\boxed{\mathbb{E}(\gamma_n(t)) = \|t\|_{S_Z}^2 - 2\int t(x)\lambda(x)S_Z(x)dx}$$

Why is the contrast $\gamma_n(t)$ related to our hazard rate estimation problem ?

$$\begin{aligned}\mathbb{E}(\gamma_n(t)) &= \mathbb{E}\left(\|t\|_n^2 - \frac{2}{n}\sum_{i=1}^n \delta_i t(Z_i)\right)\\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\int t^2(x)\mathbf{1}_{\{Z_i>x\}}dx\right) - \frac{2}{n}\sum_{i=1}^n \mathbb{E}(\delta_i t(Z_i))\end{aligned}$$

$\mathbb{E}\left(\int t^2(x)\mathbf{1}_{\{Z_i>x\}}dx\right) = \int t^2(x)S_Z(x)dx = \|t\|_{S_Z}^2$

$\mathbb{E}(\delta_i t(Z_i)) = \mathbb{E}(\mathbf{1}_{\{X_i\leq C_i\}}t(X_i)) = \int t(x)S_C(x)f(x)dx = \int t(x)S_Z(x)\frac{f(x)}{S(x)}dx.$

(as $S_Z = S_C S$)

Therefore, we find that

$$\boxed{\mathbb{E}(\gamma_n(t)) = \|t\|_{S_Z}^2 - 2\int t(x)\lambda(x)S_Z(x)dx}$$

Why is the contrast $\gamma_n(t)$ related to our hazard rate estimation problem ?

$$\mathbb{E}(\gamma_n(t)) = \|t\|^2_{S_Z} - 2\int t(x)\lambda(x)S_Z(x)dx$$

Thus, we can write,

$$\mathbb{E}(\gamma_n(t)) \;=\; \|t\|^2_{S_Z} - 2\langle t, \lambda\rangle_{S_Z} = \|t-\lambda\|^2_{S_Z} - \|\lambda\|^2_{S_Z}$$

Thus, minimizing $\gamma_n$ for large $n$ (by the Law of Large Numbers), should provide a function $t$ minimizing $\int (t(x) - \lambda(x))^2 S_Z(x)dx$.

Therefore, we should estimate the $\mathbb{L}^2$ orthogonal projection of $\lambda$ with respect to the $S_Z$-weighted inner product on a subspace $S_m$ of functions over which the minimization is performed.

Why is the contrast $\gamma_n(t)$ related to our hazard rate estimation problem ?

$$\mathbb{E}(\gamma_n(t)) = \|t\|^2_{S_Z} - 2\int t(x)\lambda(x)S_Z(x)dx$$

Thus, we can write,

$$\mathbb{E}(\gamma_n(t)) \;=\; \|t\|^2_{S_Z} - 2\langle t, \lambda\rangle_{S_Z} = \|t-\lambda\|^2_{S_Z} - \|\lambda\|^2_{S_Z}$$

Thus, minimizing $\gamma_n$ for large $n$ (by the Law of Large Numbers), should provide a function $t$ minimizing $\int (t(x)-\lambda(x))^2 S_Z(x)dx$.

Therefore, we should estimate the $\mathbb{L}^2$ orthogonal projection of $\lambda$ with respect to the $S_Z$-weighted inner product on a subspace $S_m$ of functions over which the minimization is performed.

Three norms are involved in the estimation problem and must be compared.

For a given square integrable function $t$ defined on $A$,

- the integral $\mathbb{L}^2(A, dx)$ norm $\|t\|^2 = \int t^2(x)dx$ associated to the basis.
- the empirical norm $\|t\|_n^2 = \frac{1}{n}\sum_{i=1}^n \int t^2(x)\mathbf{1}_{\{Z_i>x\}}dx$ involved in the definition of the least squares contrast
- and its expectation corresponding to a weighted $\mathbb{L}^2(A, S_Z(x)dx)$ norm

$$\|t\|_{S_Z}^2 = \int t^2(x)S_Z(x)dx$$

In previous works : Only compactly supported bases have been considered : easier to deal with !

↪ New insight with a work by Cohen, Davenport, Leviatan (2013, 2019) on the stability of Least Squares approximations.

↪ New way of dealing with regression estimators (Comte & Genon-Catalot, 2020) and hazard rate estimation (Brunel & Comte, 2021).

Three norms are involved in the estimation problem and must be compared.

For a given square integrable function $t$ defined on $A$,

- the integral $\mathbb{L}^2(A, dx)$ norm $\|t\|^2 = \int t^2(x)dx$ associated to the basis.
- the empirical norm $\|t\|_n^2 = \frac{1}{n}\sum_{i=1}^n \int t^2(x)\mathbf{1}_{\{Z_i > x\}}dx$ involved in the definition of the least squares contrast
- and its expectation corresponding to a weighted $\mathbb{L}^2(A, S_Z(x)dx)$ norm

$$\|t\|_{S_Z}^2 = \int t^2(x)S_Z(x)dx$$

In previous works : Only compactly supported bases have been considered : easier to deal with !

↪ New insight with a work by Cohen, Davenport, Leviatan (2013, 2019) on the stability of Least Squares approximations.

↪ New way of dealing with regression estimators (Comte & Genon-Catalot, 2020) and hazard rate estimation (Brunel & Comte, 2021).

Let $A \subseteq \mathbb{R}^+$ and let $(\varphi_j, j = 0, \cdots, m-1)$ be an orthonormal system of functions supported on $A$ belonging to $\mathbb{L}^2(A, dx)$, i.e. such that

$$\langle \varphi_j, \varphi_k \rangle = \delta_{j,k},\ 0 \leq j, k \leq m-1.$$

We define $S_m = \mathrm{span}(\varphi_0, \cdots, \varphi_{m-1})$. Thus, $\dim(S_m) = m$.

Now, our projection estimator $\widehat{\lambda}_m$ of the hazard rate $\lambda$ is defined by :

$$\boxed{\widehat{\lambda}_m := \arg\min_{t \in S_m} \gamma_n(t).}$$

with

$$\gamma_n(t) = \|t\|_n^2 - \frac{2}{n} \sum_{i=1}^{n} \delta_i t(Z_i),$$

Standard algebra gives, for $t = \sum_{j=0}^{m-1} a_j \varphi_j \in S_m$,

$$\nabla \gamma_n(a_0, \cdots, a_{m-1}) = 2\widehat{\Psi}_{m,Z}\vec{a}^{(m)} - \frac{2}{n}\widehat{\Phi}_m^\top \vec{\delta}, \text{ with } \vec{a}^{(m)} = \begin{pmatrix} a_0 \\ \vdots \\ a_{m-1} \end{pmatrix}$$

with

$$\begin{cases} \widehat{\Psi}_{m,Z} := (\langle \varphi_j, \varphi_k \rangle_n)_{0 \le j,k \le m-1} = \left( \int \varphi_j(x)\varphi_k(x)\widehat{S}_{Z,n}(x)dx \right)_{0 \le j,k \le m-1} \\ \widehat{\Phi}_m = (\varphi_j(Z_i))_{1 \le i \le n, 0 \le j \le m-1} \text{ and } \vec{\delta} = (\delta_1, \cdots \delta_n)^\top \end{cases}$$

Provided that $\widehat{\Psi}_{m,Z}$ is *a. s.* invertible,

$$\boxed{\widehat{\lambda}_m = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j \text{ with } \vec{\hat{a}}^{(m)} = \begin{pmatrix} \hat{a}_0 \\ \vdots \\ \hat{a}_{m-1} \end{pmatrix} = \frac{1}{n}\widehat{\Psi}_{m,Z}^{-1}\widehat{\Phi}_m^\top \vec{\delta}}$$

Provided that $\widehat{\Psi}_{m,Z}$ is *a. s.* invertible,

$$\widehat{\lambda}_m = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j \text{ with } \vec{\hat{a}}^{(m)} = \begin{pmatrix} \hat{a}_0 \\ \vdots \\ \hat{a}_{m-1} \end{pmatrix} = \frac{1}{n} \widehat{\Psi}_{m,Z}^{-1} \widehat{\Phi}_m^\top \vec{\delta}$$

and we need to define the trimmed estimator :

$$\boxed{\tilde{\lambda}_m = \begin{cases} \widehat{\lambda}_m & \text{if } \|\widehat{\Psi}_{m,Z}^{-1}\|_{\mathrm{op}} \leq \mathfrak{c} \dfrac{n}{\ln n} \\ 0 & \text{otherwise} \end{cases}}$$

where $\|\widehat{\Psi}_{m,Z}^{-1}\|_{\mathrm{op}} = +\infty$ if $\widehat{\Psi}_{m,Z}$ is not invertible and $\mathfrak{c}$ is a constant defined further.

# Outline

Consider a general context where the estimation support $A$ is such that : $A \subseteq \mathbb{R}^+$ and

$$\int_A \lambda^2(x) S_Z(x) dx < +\infty.$$

This condition is fulfilled for most classical models in survival analysis. Indeed as $S_Z \leq S$, the condition holds if the distribution of $X$ is such that $\int_A \lambda^2 S < +\infty$.

**Examples of survival models satisfying $\int_{\mathbb{R}^+} \lambda^2 S < +\infty$.**

(E.1) Exponential model, $\lambda(x) = \theta \mathbf{1}_{\{x \geq 0\}}$, $S(x) = \exp(-\theta x) \mathbf{1}_{\{x \geq 0\}}$, $\theta > 0$.

(E.2) Weibull model, $\lambda(x) = \alpha \theta^\alpha x^{\alpha-1} \mathbf{1}_{\{x \geq 0\}}$, $S(x) = \exp(-(\theta x)^\alpha) \mathbf{1}_{\{x \geq 0\}}$, $\alpha > \frac{1}{2}$, $\theta > 0$

(E.3) Gamma model, $f(x) = \theta^\nu x^{\nu-1} e^{-\theta x} / \Gamma(\nu) \mathbf{1}_{\{x \geq 0\}}$, $\nu > \frac{1}{2}$, $\theta > 0$,

(E.4) Gompertz–Makeham, $\lambda(x) = \gamma_0 + \gamma_1 e^{\gamma_2 x}$, $S(x) = e^{-\gamma_0 x - (\gamma_2/\gamma_1)(e^{\gamma_2 x} - 1)} \mathbf{1}_{\{x \geq 0\}}$, for real numbers $\gamma_0, \gamma_1, \gamma_2 > 0$,

(E.5) Log-logistic, $\lambda(x) = \theta \nu x^{\nu-1} (1 + \theta x^\nu)^{-1} \mathbf{1}_{\{x \geq 0\}}$, $\nu > \frac{1}{2}$, $\theta > 0$, $S(x) = 1/(1 + \theta x^\nu) \mathbf{1}_{\{x \geq 0\}}$,

(E.6) Log-normal $\lambda(x) = (x\sigma)^{-1} \phi\left(\sigma^{-1}(\ln x - \mu)\right) \left[1 - \Phi\left(\sigma^{-1}(\ln x - \mu)\right)\right]^{-1} \mathbf{1}_{\{x \geq 0\}}$, where $\phi(x)$ and $\Phi(x)$ are respectively the density and the cumulative distribution function of a standard Gaussian, $\mu \in \mathbb{R}$, $\sigma > 0$.

Consider a general context where the estimation support $A$ is such that : $A \subseteq \mathbb{R}^+$ and

$$\int_A \lambda^2(x) S_Z(x) dx < +\infty.$$

This condition is fulfilled for most classical models in survival analysis. Indeed as $S_Z \leq S$, the condition holds if the distribution of $X$ is such that $\int_A \lambda^2 S < +\infty$.

**Examples of survival models satisfying $\int_{\mathbb{R}^+} \lambda^2 S < +\infty$.**

(E.1) Exponential model, $\lambda(x) = \theta \mathbf{1}_{\{x \geq 0\}}$, $S(x) = \exp(-\theta x)\mathbf{1}_{\{x \geq 0\}}$, $\theta > 0$.

(E.2) Weibull model, $\lambda(x) = \alpha\theta^\alpha x^{\alpha-1}\mathbf{1}_{\{x \geq 0\}}$, $S(x) = \exp(-(\theta x)^\alpha)\mathbf{1}_{\{x \geq 0\}}$, $\alpha > \frac{1}{2}$, $\theta > 0$

(E.3) Gamma model, $f(x) = \theta^\nu x^{\nu-1} e^{-\theta x}/\Gamma(\nu)\mathbf{1}_{\{x \geq 0\}}$, $\nu > \frac{1}{2}$, $\theta > 0$,

(E.4) Gompertz–Makeham, $\lambda(x) = \gamma_0 + \gamma_1 e^{\gamma_2 x}$, $S(x) = e^{-\gamma_0 x - (\gamma_2/\gamma_1)(e^{\gamma_2 x} - 1)}\mathbf{1}_{\{x \geq 0\}}$, for real numbers $\gamma_0, \gamma_1, \gamma_2 > 0$,

(E.5) Log-logistic, $\lambda(x) = \theta\nu x^{\nu-1}(1 + \theta x^\nu)^{-1}\mathbf{1}_{\{x \geq 0\}}$, $\nu > \frac{1}{2}$, $\theta > 0$, $S(x) = 1/(1 + \theta x^\nu)\mathbf{1}_{\{x \geq 0\}}$,

(E.6) Log-normal $\lambda(x) = (x\sigma)^{-1}\phi\left(\sigma^{-1}(\ln x - \mu)\right)\left[1 - \Phi\left(\sigma^{-1}(\ln x - \mu)\right)\right]^{-1}\mathbf{1}_{\{x \geq 0\}}$, where $\phi(x)$ and $\Phi(x)$ are respectively the density and the cumulative distribution function of a standard Gaussian, $\mu \in \mathbb{R}$, $\sigma > 0$.

First, We can prove a bound for the empirical risk : let us denote $\lambda_A := \lambda \mathbf{1}_A$

## Proposition (Empirical Risk Bound)

$\mathcal{H}1.$ $\displaystyle\int_A \lambda^2(z)\sqrt{S_Z(z)}dz < +\infty.$

$\mathcal{H}2.$ In addition, we assume that the basis $(\varphi_j)_{0\leq j\leq m-1}$ is such that $L(m) := \sup_{x\in A}\sum_{j=0}^{m-1}\varphi_j^2(x) < +\infty.$

$\mathcal{H}3.$ The matrix $\Psi_{m,Z} := (\langle \varphi_j, \varphi_k\rangle_{S_Z})_{0\leq j,k\leq m-1}$ is invertible and

$$\|\Psi_{m,Z}^{-1}\|_{\mathrm{op}} \leq \frac{\mathfrak{c}}{2}\frac{n}{\ln n}, \quad \mathfrak{c} = \frac{3\ln\frac{3}{2}-1}{10},$$ "stability condition" (Cohen & al, 2013, 2019)

Then, for any $m$ such that $L(m) \leq n$,

$$\mathbb{E}\left[\|\tilde{\lambda}_m - \lambda_A\|_n^2\right] \quad \leq \quad \inf_{t\in S_m}\|t-\lambda_A\|_{S_Z}^2 + 2\frac{\mathrm{Tr}(\Psi_{m,Z}^{-1}\Psi_{m,\lambda S_Z})}{n} + \frac{\mathcal{C}_1}{n}.$$

where $\mathcal{C}_1$ is a positive constant and $\Psi_{m,\lambda S_Z} = \left(\int \varphi_j(x)\varphi_k(x)\lambda(x)S_Z(x)dx\right)_{0\leq j,k\leq m-1}$.

First, We can prove a bound for the empirical risk : let us denote $\lambda_A := \lambda \mathbf{1}_A$

## Proposition (Empirical Risk Bound)

$\mathcal{H}1.$ $\displaystyle\int_A \lambda^2(z)\sqrt{S_Z(z)}dz < +\infty.$

$\mathcal{H}2.$ In addition, we assume that the basis $(\varphi_j)_{0\leq j\leq m-1}$ is such that $L(m) := \sup_{x\in A}\sum_{j=0}^{m-1}\varphi_j^2(x) < +\infty$.

$\mathcal{H}3.$ The matrix $\Psi_{m,Z} := (\langle \varphi_j, \varphi_k\rangle_{S_Z})_{0\leq j,k\leq m-1}$ is invertible and

$$\|\Psi_{m,Z}^{-1}\|_{\mathrm{op}} \leq \frac{\mathfrak{c}}{2}\frac{n}{\ln n}, \quad \mathfrak{c} = \frac{3\ln\frac{3}{2} - 1}{10},$$ "stability condition" (Cohen & al, 2013, 2019)

Then, for any $m$ such that $L(m) \leq n$,

$$\mathbb{E}\left[\|\tilde{\lambda}_m - \lambda_A\|_n^2\right] \leq \inf_{t\in S_m}\|t - \lambda_A\|_{S_Z}^2 + 2\frac{\mathrm{Tr}(\Psi_{m,Z}^{-1}\Psi_{m,\lambda S_Z})}{n} + \frac{\mathcal{C}_1}{n}.$$

Bias $\searrow$ with $m$

where $\mathcal{C}_1$ is a positive constant and $\Psi_{m,\lambda S_Z} = \left(\int \varphi_j(x)\varphi_k(x)\lambda(x)S_Z(x)dx\right)_{0\leq j,k\leq m-1}$.

First, We can prove a bound for the empirical risk : let us denote $\lambda_A := \lambda \mathbf{1}_A$

## Proposition (Empirical Risk Bound)

$\mathcal{H}1.$ $\displaystyle\int_A \lambda^2(z)\sqrt{S_Z(z)}dz < +\infty.$

$\mathcal{H}2.$ In addition, we assume that the basis $(\varphi_j)_{0\le j\le m-1}$ is such that $L(m) := \sup_{x\in A}\sum_{j=0}^{m-1}\varphi_j^2(x) < +\infty.$

$\mathcal{H}3.$ The matrix $\Psi_{m,Z} := (\langle \varphi_j, \varphi_k\rangle_{S_Z})_{0\le j,k\le m-1}$ is invertible and

$$\|\Psi_{m,Z}^{-1}\|_{\rm op} \le \frac{\mathfrak{c}}{2}\frac{n}{\ln n}, \quad \mathfrak{c} = \frac{3\ln\frac{3}{2}-1}{10},$$ "stability condition" (Cohen & al, 2013, 2019)

Then, for any $m$ such that $L(m) \le n$,

$$\mathbb{E}\left[\|\tilde\lambda_m - \lambda_A\|_n^2\right] \le \inf_{t\in S_m}\|t-\lambda_A\|_{S_Z}^2 + 2\frac{\mathrm{Tr}(\Psi_{m,Z}^{-1}\Psi_{m,\lambda S_Z})}{n} + \frac{\mathcal{C}_1}{n}.$$

Bias $\searrow$ with $m$ Variance ???

where $\mathcal{C}_1$ is a positive constant and $\Psi_{m,\lambda S_Z} = \left(\int \varphi_j(x)\varphi_k(x)\lambda(x)S_Z(x)dx\right)_{0\le j,k\le m-1}.$

Second, We can prove a bound for the integrated risk :

## Proposition (Integrated Risk Bound)

Under Assumptions $\mathcal{H}1$, $\mathcal{H}2$ , $\mathcal{H}3$ For any $m$ such that $L(m) \leq n$,

$$\mathbb{E}\left[\|\tilde{\lambda}_m - \lambda_A\|^2_{S_Z}\right] \leq \left(1 + 8\frac{\mathfrak{c}}{\log n}\right) \inf_{t \in S_m} \|t - \lambda_A\|^2_{S_Z} + 2\frac{\mathrm{Tr}(\Psi^{-1}_{m,Z}\Psi_{m,\lambda S_Z})}{n} + \frac{\mathcal{C}_1}{n}.$$

where $\mathcal{C}_1$ is a positive constant and $\Psi_{m,\lambda S_Z} = \left(\int \varphi_j(x)\varphi_k(x)\lambda(x)S_Z(x)dx\right)_{0 \leq j,k \leq m-1}$.

Now, we can prove the following (not obvious !) lemma :

**Lemma**

*Let the collection $(S_m)$ be nested ($m \leq m' \implies S_m \subset S_{m'}$), then $m \mapsto \mathrm{Tr}(\Psi_{m,Z}^{-1}\Psi_{m,\lambda S_Z})$ is increasing.*

Therefore, both bounds lead to the same conclusion that a compromise has to be found for the choice of $m$, making a trade-off between bias and variance.

$$\mathbb{E}\left[\|\tilde{\lambda}_m - \lambda_A\|_n^2\right] \leq \underbrace{\inf_{t \in S_m} \|t - \lambda_A\|_{S_Z}^2}_{\text{Bias} \searrow \text{ with } m} + \underbrace{2\frac{\mathrm{Tr}(\Psi_{m,Z}^{-1}\Psi_{m,\lambda S_Z})}{n}}_{\text{Variance} \nearrow \text{ with } m} + \frac{\mathcal{C}_1}{n}.$$

$$\mathbb{E}\left[\|\tilde{\lambda}_m - \lambda_A\|_{S_Z}^2\right] \leq \left(1 + 8\frac{\mathfrak{c}}{\log n}\right) \underbrace{\inf_{t \in S_m} \|t - \lambda_A\|_{S_Z}^2}_{\text{Bias} \searrow \text{ with } m} + \underbrace{2\frac{\mathrm{Tr}(\Psi_{m,Z}^{-1}\Psi_{m,\lambda S_Z})}{n}}_{\text{Variance} \nearrow \text{ with } m} + \frac{\mathcal{C}_1}{n}.$$

Any condition on the estimation set $A$ is required.

## Main ingredient for the proof of the empirical and integrated risk bounds :

Two sets are of interest :

$$\Omega_m = \left\{ \forall t \in S_m, \left| \frac{\|t\|_n^2}{\|t\|_{S_Z}^2} - 1 \right| \leq \frac{1}{2} \right\}$$

$$\Lambda_m = \left\{ \|\widehat{\Psi}_{m,Z}^{-1}\|_{\mathrm{op}} \leq \mathfrak{c} \frac{n}{\log(n)} \right\}$$

The following Lemma provides preliminary results which are the main ingredients to bound the empirical risk and the integrated risk of one estimator.

### Lemma

Under the assumptions $\mathcal{H}1.$, $\mathcal{H}2.$, $\mathcal{H}3.$ ,

$$\mathbb{P}(\Omega_m^c) \leq 2/n^4 \quad \text{and} \quad \mathbb{P}(\Lambda_m^c) \leq 2/n^4$$

## Previous works with compactly supported bases

Assume that we want to estimate $\lambda_A$ with $A$ compact (Brunel & Comte, 2005 ; Reynaud-Bouret, 2006 ; Plancade, 2011 ; Comte & al. 2011).

- trigonometric basis on $A = [0, a]$ : $\varphi_0(x) = a^{-1/2}\mathbf{1}_{[0,a]}(x)$, $\varphi_{2j-1}(x) = \sqrt{2/a}\cos(2\pi jx/a)\mathbf{1}_{[0,a]}(x)$, $\varphi_{2j}(x) = \sqrt{2/a}\sin(2\pi jx/a)\mathbf{1}_{[0,a]}(x)$, $j \geq m$.
- histogram basis on $A = [0, a]$ , we set $\varphi_j(x) = \sqrt{ma}\mathbf{1}_{[ja/m,(j+1)a/m)}$ for $j = 0, \dots, m-1$.
- general piecewise polynomials with given degree $r$, by rescaling $Q_0, \dots, Q_r$, the Legendre basis on each sub-interval $[ja/m, (j+1)a/m)$, $j = 0, \dots, m-1$.

Condition $\mathcal{H}2.L(m) = \sup_{x\in A}\sum_{j=0}^{m-1}\varphi_j^2(x) < +\infty$ is satisfied for these bases, and $L(m) \leq c_\varphi^2 m$, where $c_\varphi^2$ is a known constant depending on the basis and not on $m$.

## Previous works with compactly supported bases

Assume that we want to estimate $\lambda_A$ with $A$ compact (Brunel & Comte, 2005 ; Reynaud-Bouret, 2006 ; Plancade, 2011 ; Comte & al. 2011).

Instead of Assumption $\mathcal{H}1$. $\int_A \lambda^2(z)\sqrt{S_Z(z)}dz < +\infty$., it is assumed :

$$\mathcal{H}1'.\ \forall x \in A,\ S_Z(x) \geq S_0 > 0 \text{ and } \lambda(x) \leq \|\lambda_A\|_\infty < +\infty$$

In the compact setting, $\mathcal{H}1'. \implies \mathcal{H}1$.

For $A = \mathbb{R}^+$, $\mathcal{H}1'$. does not hold anymore.

Our results encompass previous ones :

**Lemma**

*Let $A$ be a compact set and consider a basis such that $L(m) \leq c_\varphi^2 m$. Under $\mathcal{H}1'$, condition $\mathcal{H}1$, is fulfilled. Moreover,*

(i) $\|\Psi_{m,Z}^{-1}\|_{\mathrm{op}} \leq 1/S_0$, *"stability condition" automatically fulfilled for sufficiently large $n$*

(ii) $0 \leq \mathrm{Tr}(\Psi_{m,Z}^{-1}\Psi_{m,\lambda S_Z}) \leq m\|\lambda_A\|_\infty$ *explicit upper bound for the variance term*

We can recover the previous variance bound :

**Proposition (Plancade, 2011)**

$$\mathbb{E}[\|\tilde{\lambda}_m - \lambda_A\|_n^2] \leq \|\lambda_m - \lambda_A\|^2 + 2\|\lambda_A\|_\infty \frac{m}{n} + \frac{\mathcal{C}_1}{n},$$

where $\lambda_m$ is the $\mathbb{L}^2$-orthogonal projection of $\lambda_A$ on $S_m$.

Our results encompass previous ones :

**Lemma**

*Let $A$ be a compact set and consider a basis such that $L(m) \leq c_\varphi^2 m$. Under $\mathcal{H}1'$, condition $\mathcal{H}1$, is fulfilled. Moreover,*
(i) $\|\Psi_{m,Z}^{-1}\|_{\mathrm{op}} \leq 1/S_0$, *"stability condition" automatically fulfilled for sufficiently large $n$*
(ii) $0 \leq \mathrm{Tr}(\Psi_{m,Z}^{-1}\Psi_{m,\lambda S_Z}) \leq m\|\lambda_A\|_\infty$ *explicit upper bound for the variance term*

We can recover the previous variance bound :

**Proposition (Plancade, 2011)**

$$\mathbb{E}[\|\tilde{\lambda}_m - \lambda_A\|_n^2] \leq \|\lambda_m - \lambda_A\|^2 + 2\|\lambda_A\|_\infty \frac{m}{n} + \frac{\mathcal{C}_1}{n},$$

where $\lambda_m$ is the $\mathbb{L}^2$-orthogonal projection of $\lambda_A$ on $S_m$.

Consequently, We can recover the standard optimal nonparametric proved previously :

For a function $\lambda_A$ in a class of functions of regularity $\alpha$ (Besov or Sobolev classes), the choice $m^* = n^{1/(2\alpha+1)}$ will lead to the risk order $\mathbb{E}[\|\tilde{\lambda}_{m^*} - \lambda_A\|_n^2] \lesssim n^{-2\alpha/(2\alpha+1)}$ usual optimal nonparametric rate (Hubber & MacGibbon, 2004 ; Comte & al. 2011).

## Non-compact bases

Example of the Laguerre basis on $A = \mathbb{R}^+$. Let us define :

$$P_j(x) = \sum_{k=0}^{j} (-1)^k \binom{j}{k} \frac{x^k}{k!}, \qquad \varphi_j(x) = \sqrt{2} P_j(2x) e^{-x} \mathbf{1}_{x \geq 0}, \quad j \geq 0.$$

The collection $(\varphi_j)_{j \geq 0}$ is a complete orthonormal system on $\mathbb{L}^2(\mathbb{R}^+)$, such that $\forall j \geq 0, \forall x \in \mathbb{R}^+, |\varphi_j(x)| \leq \sqrt{2}$ (Abramowitz and Stegun ,1964).

$\hookrightarrow$ Therefore $L(m) = \sum_{j=0}^{m-1} \varphi_j(x)^2 \leq 2m$ and condition $\mathcal{H}2.$ is satisfied.

Other bases could be considered but with Laguerre basis the estimators are general combination of Gamma-type distributions, particularly well-suited in the context of survival models.

## What is the order of the variance term in the non-compact setting ?

- Compact bases : explicit bound $0 \leq \mathrm{Tr}(\Psi_{m,Z}^{-1}\Psi_{m,\lambda S_Z}) \leq m\|\lambda_A\|_\infty$
- Non-compact bases : the order of $\mathrm{Tr}(\Psi_{m,Z}^{-1}\Psi_{m,\lambda S_Z})$ is not obvious !

We can prove :

**Lemma**

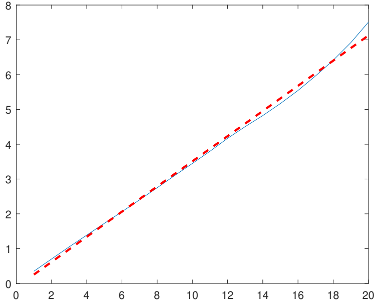
*If $\mu(\mathbb{R}^+ \cap \mathrm{Supp}(S_Z)) > 0$ where $\mu$ is Lebesgue measure and $\mathrm{Supp}(S_Z) = \{x \in \mathbb{R}^+, S_Z(x) > 0\}$ is the support of $S_Z$, then $\Psi_{m,Z}$ is invertible. Moreover, there exists $c^\star > 0$ such that, for sufficiently large $m$,*

$$\|\Psi_{m,Z}^{-1}\|_{\mathrm{op}} \geq c^\star\sqrt{m}.$$

## Evaluating the variance order on numerical examples

Note that if $X \sim \mathcal{E}(\beta)$ i.e. $f(x) = \beta e^{-\beta x}\mathbf{1}_{\mathbb{R}^+}(x)$ and $S(x) = e^{-\beta x}\mathbf{1}_{\mathbb{R}^+}(x)$, then $\lambda(x) = \beta$. Therefore $\Psi_{m,Z} = \beta^{-1}\Psi_{m,\lambda S_Z}$ and

$$\mathrm{Tr}(\Psi_{m,Z}^{-1}\Psi_{m,\lambda S_Z}) = \beta\mathrm{Tr}(\mathrm{Id}_m) = \beta\, m.$$




Plots of $m \mapsto \mathrm{Tr}(\widehat{\Psi}_{m,Z}^{-1}\widehat{\Psi}_{m,\lambda S_Z})$ for $m = 1, \dots, 20$, from $n = 10000$ observations of $X \sim \mathcal{E}(1/3)$ with no censoring, in blue.

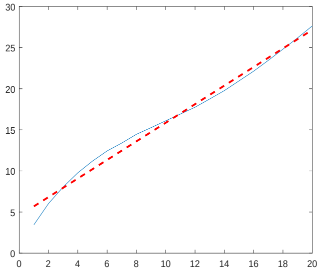
In bold dotted red, the best approximating line $y = \hat{a} + \hat{b}x$, with $\hat{a} = -0.11$, $\hat{b} = 0.36$ .

## Evaluating the variance order on numerical examples

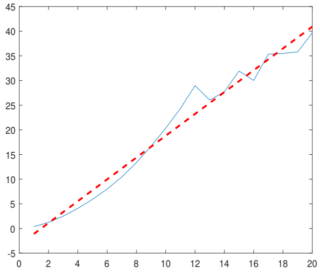
Conjecture : the variance can remain of order $m/n$ in the non-compact setting as well !

$$\boxed{\mathrm{Tr}(\Psi_{m,Z}^{-1}\Psi_{m,\lambda S_Z}) \text{ is of order } bm.}$$

$f(x) = 3/(1+x)^4 \mathbf{1}_{x\geq 0}$




$f(x) = xe^{-x^2/2}\mathbf{1}_{x\geq 0}$




Plots of $m \mapsto \mathrm{Tr}(\widehat{\Psi}_{m,Z}^{-1}\widehat{\Psi}_{m,\lambda S_Z})$ for $m = 1, \ldots, 20$, from $n = 10000$ observations with no censoring, in blue. In bold dotted red, the best approximating line $y = \hat{a} + \hat{b}x$, with $\hat{a} = -0.11$, $\hat{b} = 0.36$ (left) and $\hat{a} = -3.31$, $\hat{b} = 2.21$ (right).

# Outline

Now, let $(S_m)_{m\in\mathcal{M}_n}$ be the theoretical collection of models with $\mathcal{M}_n$ defined by

$$\mathcal{M}_n = \left\{ m \in \{1, \ldots, n\}, \|\Psi_{m,Z}^{-1}\|_{\mathrm{op}} \leq \frac{\mathfrak{c}}{2} \frac{n}{\ln n} \right\}$$

and its empirical version

$$\widehat{\mathcal{M}}_n = \left\{ m \in \{1, \ldots, n\}, \|\widehat{\Psi}_{m,Z}^{-1}\|_{\mathrm{op}} \leq \mathfrak{c} \frac{n}{\ln n} \right\}.$$

Then we select

$$\hat{m} = \arg \min_{m\in\widehat{\mathcal{M}}_n} \left( \gamma_n(\hat{\lambda}_m) + \widehat{\mathrm{pen}}(m) \right), \quad \widehat{\mathrm{pen}}(m) = \kappa \frac{\mathrm{Tr}(\widehat{\Psi}_{m,Z}^{-1} \widehat{\Psi}_{m,\lambda S_Z})}{n},$$

- The penalty $\widehat{\mathrm{pen}}(m)$ is the empirical version of the variance order. The penalized criterion is thus an empirical version of the squared bias / variance decomposition.
- The constant $\kappa$ is numerical and depends on neither $\lambda$ nor $n$; We took $\kappa = 2$ in numerical experiments.

## Estimation of $\Psi_{m,\lambda S_Z}$ in the penalty term

An important preliminary remark is that, as $\lambda S_Z = f S_C$, the matrix $\Psi_{m,\lambda S_Z}$ can easily be estimated by

$$\widehat{\Psi}_{m,\lambda S_Z} = \left( \frac{1}{n} \sum_{i=1}^{n} \delta_i \varphi_j(Z_i) \varphi_k(Z_i) \right)_{0 \le j,k \le m-1}.$$

$$\begin{aligned}
\mathbb{E}\left(\delta_1 \varphi_j(Z_1)\varphi_k(Z_1)\right) &= \mathbb{E}\left(\mathbf{1}_{(X_1 \le C_1)} \varphi_j(X_1)\varphi_k(X_1)\right) \\
&= \mathbb{E}\left(\mathbb{E}(\mathbf{1}_{(X_1 \le C_1)} \varphi_j(X_1)\varphi_k(X_1) | X_1)\right) \\
&= \mathbb{E}\left(\varphi_j(X_1)\varphi_k(X_1) S_C(X_1)\right) \\
&= \int \varphi_j(x)\varphi_k(x) f(x) S_C(x) dx = \langle \varphi_j, \varphi_k \rangle_{\lambda S_Z}
\end{aligned}$$

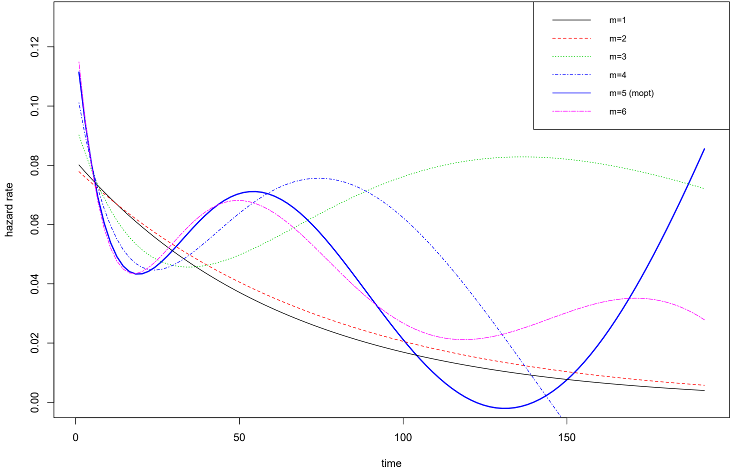
## Estimation of $\Psi_{m,\lambda S_Z}$ in the penalty term

An important preliminary remark is that, as $\lambda S_Z = f S_C$, the matrix $\Psi_{m,\lambda S_Z}$ can easily be estimated by

$$\widehat{\Psi}_{m,\lambda S_Z} = \left( \frac{1}{n} \sum_{i=1}^{n} \delta_i \varphi_j(Z_i) \varphi_k(Z_i) \right)_{0 \le j,k \le m-1} .$$

$$\begin{aligned}
\mathbb{E}\left(\delta_1 \varphi_j(Z_1) \varphi_k(Z_1)\right) &= \mathbb{E}\left(\mathbf{1}_{(X_1 \le C_1)} \varphi_j(X_1) \varphi_k(X_1)\right) \\
&= \mathbb{E}\left(\mathbb{E}(\mathbf{1}_{(X_1 \le C_1)} \varphi_j(X_1) \varphi_k(X_1) | X_1)\right) \\
&= \mathbb{E}\left(\varphi_j(X_1) \varphi_k(X_1) S_C(X_1)\right) \\
&= \int \varphi_j(x) \varphi_k(x) f(x) S_C(x) dx = \langle \varphi_j, \varphi_k \rangle_{\lambda S_Z}
\end{aligned}$$

# Model selection : a real data example for breast-feeding lifetime ($n = 927$ women)




Collection of estimators for $n = 927$ with 96% of non-censoring. cohort from the National Longitudinal Survey of Youth of the U.S. Bureau of Labor Statistics

## Comparison of the Mean Squared Error (MSE) over 200 replications with previous Projection methods

(a) Gamma case. $X_i \sim$Gamma$(\nu, \theta)$ with $\nu = 5$ (shape) and $\theta = 1$ (scale) and $C_i \sim \mathcal{E}xp(1/6)$.

(b) Bimodal case. The $X_i$'s have a bimodal density defined by $f = 0.8u + 0.2v$ where $u$ is the density of $\exp(Y/2)$ with $Y \sim \mathcal{N}(0,1)$ and $v = 0.17Y + 2$ and $C_i \sim \mathcal{E}xp(2/5)$.

| Method | Gamma | | Bimodal | |
|---|---|---|---|---|
| $n$ | 200 | 500 | 200 | 500 |
| Wavelet<br>Antoniadis & et al (1999) | 0.112 | 0.099 | 2.080 | 0.197 |
| Histogram<br>Reynaud-Bouret (2004) | 0.055 | 0.057 | 1.259 | 1.122 |
| Fourier<br>Brunel & Comte (2005) | 0.086 | 0.090 | 0.902 | 0.706 |
| Laguerre basis | 0.0275 | 0.0084 | 0.629 | 0.487 |

## Comparison of the MSE with a Kernel estimator (non-compact setting)

| Method | Weibull | $\chi^2(2)$ | $\chi^2(3)$ |
|---|---|---|---|
| $n$ | 100 | 100 | 100 |
| Kernel estimator<br>Barbeito & Cao (2019) | 0.029 | 0.056 | 0.020 |
| Laguerre basis | 0.028 | 0.040 | 0.029 |

Merci de votre attention !