Habilitation Thesis :
On Adaptive Nonparametric Estimation for Survival Data

"Contribution à l'estimation non-paramétrique adaptative dans les modèles de durées"

Elodie BRUNEL
I3M, UMR 5149 CNRS, Université Montpellier 2 , FRANCE

Contents

Introduction

This report is about my research works from the beginning of my career at the Paris Descartes University where I had my first assistant professor position during year 2000 to now at the Montpellier 2 University, since my transfer in 2007.

These works deals with nonparametric estimation of different functions of interest and the crucial question of model selection in the particular application field of survival data.

I give some perspectives on how to take advantage of model selection tools for adaptive estimation in survival models. Censoring stands for the distinctive feature of such data and different schemes can appear. In all cases, it requires specific procedures to take into account the presence of censoring. Covariables are sometimes of great interest in the context of survival data and statisticians need to consider conditional models.

The report will address the following three issues: first, we consider estimators of the hazard rate, density or cumulative distribution functions in various sampling situations such as right-censoring, interval censoring or selection bias; secondly, our study includes the presence of covariables *via* the estimation of conditional mean and conditional distributions; thirdly, we explore different regression functions such as conditional mean residual life or conditional hazard rate. Finally, the concluding part of the document presents recent works, some of them are more or less successfully completed and give pathes that have not yet been explored.

Here, I don't take again the results of my Phd thesis supervised by Pr. Alain Berlinet which dealt with optimal choice of bins for modified histograms introduced by Barron in the sense of the Kullbak-Leibler divergence (publications [2], [3], [4]). Nevertheless, I have envisaged applications of Barron density estimators for estimating hazard rate and it seems to me the beginning of my interest in survival analysis.

Survival data

In many biomedical applications, the variable of interest is often a time which cannot be directly observed. *Right-censoring* constitutes the common illustration of such incomplete data. In fact, let us denote by X the time of interest, namely the time which goes by from the start of the observation to the realisation of a terminal event of interest. If for some individuals, we observe a time C before the terminal event happens (such that $X > C$), we say the data are *right-censored*. More formally, let us denote the pair (Z, δ) where $Z = \min(X, C)$ and $\delta = \mathbb{1}_{(X \leq C)}$ stands for the non-censoring indicator. The mechanism of censoring can depend on observations or not, can be random or not... In general, the random times X and C are supposed to be independent. In addition, we consider only random variables admitting a probability density with respect to the Lebesgue mesure on $\mathbb{R}$.

Interval censoring is another important feature. Let us introduce the following example: we have to follow children until the aquisition of fluent reading. Some of them are already able to read at the beginning of the shool year, others are going to learn during the period, whereas a few of them will not read at the end year. We say the data are *interval censored* because the event of interest is never observed. We just know that the event has happened or not. Interval-censoring is frequent in epidemic studies, where the contamination time cannot be exactly known.

In the presence of censoring, estimation problems need specific statistical tools. Since the famous product-limit estimator of Kaplan & Meier (1958) for the survival function $\mathbb{P}(X \geq t) =: 1 - F(t)$, many nonparametric methods have been developed for censored data.

Following Aalen (1978)'s works, statistical inference in survival models expand rapidly, especially thanks to the point processes theory. As an example, the Nelson-Aalen estimator of the cumulative hazard rate $H(t) = \int_0^t h(u)du$ has been introduced by involving the point process $N(t) = \mathbb{1}_{(Z \leq t, \delta = 1)}$ where $h(t)$ is the intensity or hazard rate function. For details, we refer to Andersen *et al.* (1993). The distribution of the time of interest can be indifferently defined from the survival function, density probability function or hazard rate, but the last one is often prefered by praticians who aim to interpret its form.

The first part of the report is devoted to different proposals of nonparametric hazard rate estimators under right-censoring. Our contribution is to provide completely data-driven and optimal procedures for estimating hazard curves.

Next, we envisage sampling situations where, in addition to the right-censoring, the data suffer from selection bias. This is often the case in epidemiology studies. We propose different nonparametric density and hazard rate estimators for data with both censoring and bias.

In the second part, we aim to take into account one or more covariables. We are interested in various regression functions such as conditional mean, conditional (cumulative) distribution, conditional density, hazard rate or mean residual life. We provide new and flexible alternatives to the proportional hazard model introduced by Cox (1972), which states that the hazard rate h has the following form:

$$h(t, x) = h_0(t) \exp(\beta_0' x), \;\; t \geq 0, \, x \in \mathbb{R}^d$$

where h_0 is a determinist function, called the baseline hazard, and the vector $\beta_0 \in \mathbb{R}^d$ stands for the unknown multivariate parameter. In this set-up, the coefficient coefficient $e^{\beta_0^j}$ quantifies the relative risk associated to the prognostic covariate x^j.

Adaptive nonparametric estimation

The common thread of my works can be found in the development of nonparametric methods for estimating adaptively various functionals of interest appearing in survival models. We adopt the minimax risk point of view, as usual in functional estimation, namely the one that is the most pessimistic over a given class of functions $\mathcal{F}$. Optimal estimation in the minimax sens has known a booming in the 90's with the works by Donoho *et al.* (1995) or Birgé & Massart (1997). Most often, we just give convergence rates by providing upper bound of the ($\mathbb{L}^2$)-risk, seeing that the minimax rate have already been established in the literature. Unfortunately, optimal estimators do depend on the regularity of the function to be estimated, and this regularity is described by the considered class $\mathcal{F}$ of functions (Hölder, Sobolev, Besov,...). For being sure, we are able to build an estimator which achieves the best possible rate of convergence over the class $\mathcal{F}$, one should know the regularity of the function. Such "estimators" are often referred as *oracles* but they cannot be computed from the data since they depend on unknown quantities. Adaptive estimation solve this problem by providing *data-driven* estimators, which achieve automatically the optimal rate of convergence with respect to the unknown regularity of the function to be estimated. Oracle inequalities constitute the main tool to obtain such adaptive results and are very attractive in a non-asymptotic point of view. Our contribution falls within the *model selection* paradigm developped by Barron *et al.* (1999), using penalised contrast estimators.

Let $s \in \mathcal{S}$ an unknown function to be estimated from the observations $X_1, \ldots, X_n$. An empirical $\gamma(X_1, \cdots, X_n; t) := \gamma_n(t)$ is called a *contrast* for s, if $t \mapsto \mathbb{E}[\gamma_n(t)]$ is minimum at the point $t = s$ over the set $\mathcal{S}$. The minimisation of a contrast over a collection of finite-dimensional approximation spaces called *models* (cf. Appendice A), provides a collections of estimators. Appendice A is devoted to the description of model collections and their approximation properties. Then, we have to choose one estimator in the collection and *model selection* allows to make this choice in an optimal way. To realize the usual trade-off between the squared bias term and the variance, we have to determine a penalty term, which has to become greater the more complex the model is. We define estimation strategies for the fonctions of interest presented above, following this line. Our works show how to take advantage of model selection tools, in addition with specific technicalities to deal with survival data. Nuisance effects due to censoring or other sampling situations do not affect convergence rates, compared to the i.i.d case. But there is a price to pay for fixed sample sizes and we illustrate in our publications, as often as possible the deterioration due to the sampling schemes.

Chapter 1

Adaptive Estimation in various sampling situations

Publications [5] , [8], [10], [11], [12]

In this part, I present the results obtained for estimation of the probability density function f, the hazard rate h and the cumulative distribution function F of a random time X in different sampling situations: right-censoring, bias selection (with known and unknown bias) and interval-censoring of type I (also referred as current status data). Here are also the first published papers. We do not consider in these works the presence of covariates.

The first section 1.1 is devoted to different poroposals for estimating the hazard rate. This also constitutes the first works with Fabienne Comte when we met in Paris Descartes University in 2001-2002. Next, with Agathe Guilloux, who had submitted us the problem of data suffering from selection bias, we interested ourselves to this problem and the section 1.2 presents these works. The last section 1.3 introduces the estimator of the cumulative distribution function for type I interval-censored data.

1.1 Hazard rate Estimators

Publications [5], [8]

Many density or hazard rate estimators have been proposed in the literature of survival data taking into account right-censoring; among others, we can cite: Marron & Padgett (1987), Dabrowska (1987), Stute (1995), Dabrowska *et al.* (1999), Efromovich (2001) or Antoniadis & Grégoire (1990), Patil (1997), Antoniadis *et al.* (1999). In these papers, the proposed procedures are not adaptive, in the sense that a regularity assumption on the unknown distribution of the sample is required. However, the estimators (kernel or wavelet-type) achieve the optimal convergence rate associated to the Mean Integrated Squared Error.

More recently, the general study done by Reynaud-Bouret (2006) includes the adaptive hazard rate estimation as a particular case of the intensity of a Poisson process one. Our estimators are close to this work (in the same spirit), by the fact they use, as well as ours, penalized contrast estimators with histogram bases.

Our goal is to propose *data-driven* estimators that automatically realize the trade-off between bias and variance without any assumption on the unknown regularity of the hazard rate. Note that, the same procedure failed to exist for kernel approaches, which are often applied for twice differentiable functions, until the recent works by Goldenshluger & Lepski (2011).

1.1.1 Right-censoring model

Here are the main notations for describing the right-censoring model. Let $X_1, \cdots, X_n$ be a n-sample of positive random variables, which are the times of interest, independent with common cumulative distribution function (c.d.f) F and let $C_1, \cdots, C_n$, be the censoring random variables, independent with c.d.f G. In addition, suppose that the C_i's are independent of the X_i's. The hazard rate h is defined as follows:

$$h(x) = \lim_{u \to 0} \frac{\mathbb{P}(x \leq X \leq X + u | X \geq x)}{u}$$

Moreover, if F has a probability density f, then we can write:

$$h(x) = \frac{d}{dx} H(x) = \frac{f(x)}{\bar{F}(x)}, \quad \text{if } F(x) < 1.$$

where $H = -\log(\bar{F})$ is the cumulative hazard function and $\bar{F} = 1 - F$ is the survival function. In the right-censoring scheme, one observes the pairs $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$ where

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

with the notation $a \wedge b = \min(a, b)$. Roughly speaking, the indicator δ_i equals 1 if the survival time of individual i is observed. Otherwise $\delta_i = 0$ and the observation is said to be censored. Let L denote the common c.d.f of $Z_1, \cdots, Z_n$ and $\bar{L} := 1 - L = (1 - F)(1 - G)$ their associated survival function.

Our estimation procedure is only valid on an interval $[0, \tau]$ with $\tau < \tau_L = \min\{\tau_F, \tau_G\}$ and $\tau_F = \sup\{x : F(x) < 1\}$ for any c.d.f F. Without loss of generality, we set $\tau = 1$ in the sequel, up to a scaling change. From Antoniadis *et al.* (1999), since $Z_{(n)} \to \tau_L \, p.s$ as $n \to \infty$ where $Z_{(n)}$ stands for the n-th order statistic, one can choose the value of τ greater than $Z_{(n)}$.

The Kaplan and Meier (1958) estimator of the survival function $\bar{F}$, is defined as follows:

$$KM_n(x) = \begin{cases} \displaystyle\prod_{i=1, Z_{(i)} \leq x}^{n} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}} & \text{if } x \leq Z_{(n)} \\ 0 & \text{if } x > Z_{(n)}. \end{cases}$$

With a slight abuse of notation, the indicator $\delta_{(i)}$ is the one associated to the observation $Z_{(i)}$. But, this definition needs the largest $Z_{(n)}$ to be non-censored. Moreover, we have to assume the estimator does not go to zero and we use a slight modification of the Kaplan-Meier estimator proposed by Lo *et al.* (1989):

$$\bar{F}_n(x) = \begin{cases} \displaystyle\prod_{i=1,\, Z_{(i)} \leq x}^{n} \left(\frac{n-i+1}{n-i+2}\right)^{\delta_{(i)}} & \text{si } x \leq Z_{(n)} \\ \bar{F}_n(Z_{(n)}) & \text{if } x > Z_{(n)}. \end{cases} \tag{1.1}$$

This estimator has good properties, in particular it satisfies $\bar{F}_n(x) \geq (n+1)^{-1}$ for all x and $\sup_{0 \leq x \leq T} |KM_n(x) - \bar{F}_n(x)| = O(n^{-1})$ a.s for all $0 < T < \min\{\tau_F, \tau_G\}$, see Lo *et al.* (1989).

1.1.2 Hazard rate estimator based on strong representation of Kaplan-Meier estimator

Here, we give the presentation of the estimator proposed and studied in [8]. Let $(\varphi_\lambda)_{\lambda \in \Lambda_m}$ be an orthonormal basis of $\mathbb{L}^2(A)$ (with $A = [0,1]$), described in Appendix A. The hazard rate estimator $h = h\mathbf{1}_A$ is defined by:

$$\hat{h}_m = \sum_{\lambda \in \Lambda_m} \hat{a}_\lambda \varphi_\lambda \; ; \; \hat{a}_\lambda = \int_0^1 \varphi_\lambda(x) dH_n(x), \tag{1.2}$$

where $H_n = -\ln(\bar{F}_n)$ and $\bar{F}_n$ defined by (1.1). We can write for $x \in A$:

$$\hat{h}_m(x) = - \sum_{i / Z_{(i)} < 1} \delta_{(i)}\, \mathcal{K}_m(Z_{(i)}, x) \ln\left(1 - \frac{1}{n-i+2}\right)$$

with $\mathcal{K}_m(Z_{(i)}, x) = \sum_{\lambda \in \Lambda_m} \varphi_\lambda(Z_{(i)}) \varphi_\lambda(x)$. This last expression underlines the similarity with the kernel estimator replacing $\mathcal{K}_m(\,.\,, x)$ with a standard kernel. We define our estimator as a projection estimator on different bases such as Fourier basis [T], piecewise polynomial basis [DP] or wavelet [W]. In model selection, we call *model* the finite-dimensional linear subspace S_m of $\mathbb{L}^2([0,1])$ generated by the basis $(\varphi_\lambda)_{\lambda \in \Lambda_m}$,

$$S_m := \text{span}(\varphi_\lambda, \lambda \in \Lambda_m)$$

Its dimension is denoted by $D_m := |\Lambda_m|$. Let us also introduce the orthogonal projection h_m on S_m of h restricted to $[0,1]$. For one estimator $\hat{h}_m$, we obtain the following upper bound of the $\mathbb{L}^2$-risk:

Proposition 1 *Let $\hat{h}_m \in S_m$ defined by (1.2), with collection* [T], [DP] *or* [W], *see Appendix A, then,*

$$\mathbb{E}(\|h - \hat{h}_m\|^2) \leq \|h - h_m\|^2 + \frac{2\Phi_0^2 D_m}{n} \int_0^1 \frac{h(x)}{1 - L(x)} dx + \kappa \frac{D_m^2 \ln^2(n)}{n^2}, \tag{1.3}$$

where κ is a numerical constant that does not depend on the basis choice.

The key argument of Proposition 1 relies on the decomposition of the variance term based on the representation of the Kaplan-Meier estimator as a sum of independent variables plus a remainder term. Lemma 3.1 in [8] gives such a decomposition and uses influence curves obtained by Reid (1981), and also Lo *et al.* (1989). Note that we also consider spline basis in [8]. We deduce easily from Proposition 1, convergence rates for the estimator provided that the function h satisfies a regularity assumption, namely h belongs to a Besov space $B_{\alpha,2,\infty}([0,1])$ and for some $\alpha > 0$. In fact, Lemma 12 in Barron *et al.* (1999) implies that the approximation term $\|h - h_m\|$ is of order $D_m^{-\alpha}$ for collection [T], [DP] and [W], pour tout réel $\alpha > 0$. The minimum in (1.3) is achieved for a model S_{m_n} such that $D_{m_n} = O([n^{1/(1+2\alpha)}])$. Then, if $h \in \mathcal{B}_{\alpha,2,\infty}([0,1])$, with $\alpha > 0$, we recover the usual nonparametric convergence rate $n^{-2\alpha/(1+2\alpha)}$.

The next step consists in finding an automatic procedure which allows an optimal choice of the dimension D_m without any knowledge of the regularity α. For this purpose, we write our estimator defined by (1.2) as the minimizer of the following contrast:

$$\gamma_n(t) = \|t\|^2 - 2\int_0^1 t(x)dH_n(x) \tag{1.4}$$

for a function $t \in \mathbb{L}^2([0,1])$, and with $\|t\|^2 = \int_0^1 t^2(x)dx$. Then,

$$\int_0^1 t(x)dH_n(x) = -\sum_{i/Z_{(i)}<1} \delta_{(i)}\, t(Z_{(i)}) \ln\left(1 - \frac{1}{n-i+2}\right),$$

and we have

$$\hat{h}_m = \arg\min_{t\in S_m} \gamma_n(t).$$

Finally, we proceed by penalization of the contrast as follows:

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n} [\gamma_n(\hat{h}_m) + \text{pen}(m)] \tag{1.5}$$

where pen(m) is a penalty term to be defined. Our result gives the form of an adequate (but theoretical) penalty:

$$\text{pen}(m) = \kappa\Phi_0^2 \left(\int_0^1 \frac{h(x)}{1-L(x)}dx\right)\frac{D_m}{n},$$

with κ a numerical constant satisfying the following oracle inequality:

$$\mathbb{E}(\|\hat{h}_{\hat{m}} - h\|^2) \leq \inf_{m\in\mathcal{M}_n} \left(3\|h - h_m\|^2 + 4\text{pen}(m)\right) + \frac{K\ln^2(n)}{n}. \tag{1.6}$$

Since the penalty term depends on unkown quantity $\mathbb{E}\left(\frac{\mathbf{1}_{(\delta_1=1,Z_1\leq 1)}}{(1-L(Z_1))^2}\right) = \int_0^1 \frac{h(x)}{1-L(x)}dx$ we have to replace it by:

$$\hat{s}_2 = \frac{1}{n}\sum_{i=1}^n \frac{\mathbf{1}_{(Z_i\leq 1)}\mathbf{1}_{(\delta_i=1)}}{(1-\hat{L}_n(Z_i))^2}, \quad \hat{L}_n(x) = \frac{1}{n+1}\sum_{i=1}^n \mathbf{1}_{(Z_i\leq x)}.$$

We can also prove an oracle inequality for the random penalty, see Theorem 4.1 and 4.2 in [8] for details.

1.1.3 Hazard rate estimator as a *ratio*

Another quantity of interest in the right-censoring model is the subdensity ψ of the uncensored data $\delta_i Z_i$. Let us define

$$\psi(x) = f(x)(1 - G(x))$$

As a consequence, we have:

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{\psi(x)}{1 - L(x)} \quad \text{if } L(x) < 1.$$

The estimation of the subdensity ψ gives another way for estimating the hazard rate as a ratio and without the use of the Kaplan-Meier estimator. We assume that:

$$\exists c_L > 0, \ \ \forall x \in [0,1], \ \ c_L \leq 1 - L(x) < 1, \tag{1.7}$$

with $c_L = \inf_{x\in[0,1]}(1 - L(x)) = 1 - L(1)$.
Then, we are in position to define the following contrast, which is just the application of the usual density contrast:

$$\gamma_n^\psi(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^n \mathbf{I}_{(Z_i \leq 1)}\delta_i t(Z_i) \tag{1.8}$$

for a function $t \in \mathbb{L}^2([0,1])$ and $\hat{\psi}_m = \arg\min_{t\in S_m} \gamma_n^\psi(t)$.

Let us define the penalized estimator $\hat{\psi}_{\hat{m}}$ with $\hat{m}$ such that:

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n} [\gamma_n^\psi(\hat{\psi}_m) + \widehat{\text{pen}}^\psi(m)] \tag{1.9}$$

$$\widehat{\text{pen}}^\psi(m) = \kappa\Phi_0^2 \left(\frac{1}{n}\sum_{i=1}^n \mathbf{I}_{(X_i\leq 1)}\mathbf{I}_{(\delta_i=1)}\right)\frac{D_m}{n}$$

for κ a numerical constant. We can prove an adaptive result for the function ψ belonging to some Besov space $\mathbb{B}_{\alpha_\psi,2,\infty}(R)$, see Theorem 3.2 in [5].

Next, we make use of the adaptive estimator of the subdensity ψ to define an estimator of the hazard rate h:

$$\tilde{h}_\psi = \frac{\hat{\psi}_{\hat{m}}}{1 - \hat{L}_n} \quad \text{with } \hat{L}_n(x) = \frac{1}{n+1}\sum_{i=1}^n \mathbf{I}_{(Z_i\leq x)}. \tag{1.10}$$

With the following decomposition of the risk:

$$\tilde{h}_\psi - h = \left(\frac{\hat{\psi}_{\hat{m}} - \psi}{1 - \hat{L}_n} + \psi(\frac{1}{1-\hat{L}_n} - \frac{1}{1-L})\right)$$

we obtain the bound:

$$\mathbb{E}\|\tilde{h}_\psi - h\|^2 \leq \frac{2^4}{c_L^2}\,\mathbb{E}\|\hat{\psi}_{\hat{m}} - \psi\|^2 + \frac{C(c_L, \|\psi\|)}{n}, \tag{1.11}$$

where $C(c_L, \|\psi\|)$ is a constant depending on c_L and $\|\psi\|$. From (1.28), we deduce that $\tilde{h}_\psi$ is an adaptive estimator of the function h provides h and ψ have the same regularity $\alpha = \alpha_h = \alpha_\psi$. We can compare our result with the one of Antoniadis *et al.* (1999) (without adaptation). It is well-known that methodologies using ratio are not optimal if the regularity α_h is greater that α_ψ; in this case the resulting rate is the worst one, namely $\tilde{h}_\psi = n^{-\alpha_\psi/(1+2\alpha_\psi)}$, instead of the optimal one $n^{-\alpha_h/(1+2\alpha_h)}$. Nevertheless, in the "good" case, we recover the minimax rate proved by Huber & MacGibbon (2004). In practice, simulation study (see [5]) shows that this strategy is not always satisfactory even if the estimation of ψ is nearly perfect! For this reason, we aim to explore a direct mean-square strategy.

1.1.4 Direct mean-square contrast estimator

In [5], we also propose the following contrast:

$$\hat{h}_m = \arg\min_{t\in S_m} \gamma_n^h(t) \quad \text{où } \gamma_n^h(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^n \mathbf{1}_{(Z_i\leq 1)} \frac{\delta_i t(Z_i)}{1-\hat{L}_n(Z_i)} \tag{1.12}$$

with $\hat{L}_n(x)$ defined above by (1.10). For a model S_m generated by a basis $(\varphi_\lambda)_{\lambda\in\Lambda_m}$, we define:

$$\hat{h}_m = \sum_{\lambda\in\Lambda_m} \hat{a}_\lambda \varphi_\lambda \quad \text{où} \quad \hat{a}_\lambda = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{(Z_i\leq 1)} \frac{\delta_i \varphi_\lambda(Z_i)}{1-\hat{L}_n(X_i)}. \tag{1.13}$$

We propose the to select the dimension as:

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n} \left(\gamma_n^h(\hat{h}_m) + \text{pen}^h(m)\right). \tag{1.14}$$

with

$$\widehat{\text{pen}}^h(m) = \kappa\Phi_0^2 \left(\frac{1}{n}\sum_{i=1}^n \frac{\mathbf{1}_{(Z_i\leq 1)}\delta_i}{(1-\hat{L}_n(Z_i))^2}\right)\frac{D_m}{n}, \tag{1.15}$$

and κ a numerical constant. Under some assumptions (see Theorem 4.1 and Corollary 4.1 in [5]), we prove adaptive results for the mean-square estimator. The proof is based on an adequate decomposition of the contrast, leading to a centred linear empirical process whose supremum on balls is controlled thanks to Talagrand (1996) Inequality. In addition a non-linear remainder term is also proved to be negligeable by using arguments of Baraud (2002). Then, we prove that the penalized estimator $\tilde{h} = \hat{h}_{\hat{m}}$ is adaptive without any restriction, contrary to the quotient estimator $\tilde{h}_\psi$: the estimator $\tilde{h}$ achieves automatically the minimax rate in all the cases. Surprisingly, simulations give a slight advantage to the quotient strategy. This is perhaps due to our choice of hazard functions that are quite smooth. But, the deterioration of the estimation at the end of the interval is the same whatever the strategy. Note that in [5], we also deal with non-regular collection of piecewise polynomial, see Section 4.4 in [5].

1.2 Estimation under the presence of both bias and censoring

1.2.1 Sampling with known bias

In publication [11], we consider data suffering from bias selection. The relevance of selection bias in statistical inference has been first pointed out by Fisher (1934). Since then, many authors noticed its presence in data from a wide range of fields. We refer to Cox (1969) for industrial applications, Chakraborty and Rao (2000) for biomedical applications, and Heckman (1985) in Economics, among many others. The review by Patil and Rao (1977) gives numerous practical examples of weighted distributions. Instead of observing the time of interest X, on observes the random variable X_w whose density probability is given by:

$$f_w(x) = \frac{w(x)f(x)}{\mu}, \qquad \text{and } \mu = \int w(u)f(u)du\,, \tag{1.16}$$

The weight function $w(.)$ is determinist and supposed to be known in this set-up. We shall say that X is suffering from selection bias. In addition, the data can also be right-censored, in this case we just observe

$$X_w \wedge C \text{ and } \delta_w = \mathbf{1}_{(X_w \leq C)}.$$

The C_i's are i.i.d. random variables and are independent of the $X_{w,i}$'s. The special case where $w(x) = x$ for all $x > 0$, called "length-biased sampling", has received a particular attention, see Vardi (1982), de Uña-Àlvarez (2002), Marron and de Uña-Àlvarez (2004) and Asgharian *et al.* (2002).

1.2.2 Adaptive estimation of the density and of the hazard rate functions with known bias

Many papers are devoted to the estimation of the density or cumulative distribution functions in this context but without considering censoring. We mention Gill *et al.* (1988) and Efromovich (2004a)for the estimation of biased data and also Vardi (1982) in the particular case of length-bias $w(x) = x$. In Efromovich (2004a), the distribution function is estimated by projection on trigonometric polynomial spaces. Optimal results are given for class of functions admitting a trigonometric serie development with rate of order $O(\log(n)/n)$.

In Efromovich (2004b), the conditions are less restrictive and the results can be compared with ours. In [11], we also have the problem of the estimation on the whole support due to the presence of censoring. Again, we restrict our attention on an interval $A = [0, \tau]$ with $\tau = \sup\{x \in \mathbb{R}^+ : (1-F_w)(1-G)(x) > 0\}$, and F_w stands for the cumulative distribution function of X_w. We set $A = [0, 1]$ for simplicity. The estimation contrast is defined for $t \in \mathbb{L}^2([0, 1])$), par

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n}\hat{\mu}\sum_{i=1}^{n} \frac{\delta_{w,i} t(X_{w,i})}{w(X_{w,i}\hat{\hat{G}}(X_{w,i})} \qquad \text{avec } \hat{\mu}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_{w,i}}{w(X_{w,i})\hat{\hat{G}}(X_{w,i})}\right)^{-1}. \tag{1.17}$$

Note that we use the modified version (1.1) of the Kaplan-Meier estimator of $\bar{G}$.

Let $\{Sm : m \in \mathcal{M}_n\}$ be the collection of projection spaces. For each m, the space S_m, of dimension D_m is generated by an orthonormal basis on $[0,1]$, described in Appendix A.

Our projection estimator of the density f is defined by:

$$\hat{f}_m = \arg\min_{t \in S_m} \gamma_n(t) \tag{1.18}$$

and the penalized criterion is:

$$\hat{m} = \arg\min_{m \in \mathcal{M}_n} [\gamma_n(\hat{f}_m) + \widehat{\text{pen}}(m)] \text{ with } \widehat{\text{pen}}(m) = \kappa \Phi_0^2 \hat{\mu}^2 \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\delta_{w,i}}{w^2(X_{w,i}) \hat{\bar{G}}^2 (X_{w,i})} \right) \frac{D_m}{n}$$

κ a numerical constant and Φ_0 a basis dependent constant. We need the following assumption on the weight function w:

$$\exists w_2 > 0, \quad 0 < w(x) \leq w_2 < +\infty, \ \forall x \in A.$$

This technical assumption excludes the case of length-bias. We prove in [11] that the penalized estimator $\hat{f}_{\hat{m}}$ achieves the optimal minimax rate of convergence for densities belonging to Besov classes.

1.2.3 Sampling in a Lexis diagram

Consider, in a population of individuals I, the random variables of their birth dates $(\sigma_i)_{i \in I}$, and the non-negative random variables of their lifetimes $(X_i)_{i \in I}$. In the Lexis (1875) diagram, an individual can be represented by his life-line, $\mathcal{L}(\sigma, X)$,

$$\mathcal{L}(\sigma, X) = \{(\sigma + y, y), \ 0 \leq y \leq X\},$$

which is a unit-slope line whose points have as coordinates the calendar time $(\sigma + y)$ and the age (y), see Figure 1.1. The individual i with birth date σ_i and lifetime X_i is included in the sample if:

$$\mathcal{L}(\sigma_i, X_i) \cap \mathcal{S} \neq \emptyset \Leftrightarrow a_{\mathcal{S}}(\sigma_i) < \infty \text{ and } X_i \geq a_{\mathcal{S}}(\sigma_i). \tag{1.19}$$

Lexis diagrams have been considered for such modelization purpose by Keiding (1990), Lund (2000) and Guilloux (2007). For example, time-window studies or cohort studies can be described by the following sampling patterns. If $\mathcal{S}$ is a deterministic Borel set in the Lexis diagram, then only individuals with life-lines intersecting $\mathcal{S}$ can be included in the study, i.e. only pairs (σ, X) such that $\mathcal{L}(\sigma, X) \cap \mathcal{S} \neq \emptyset$.

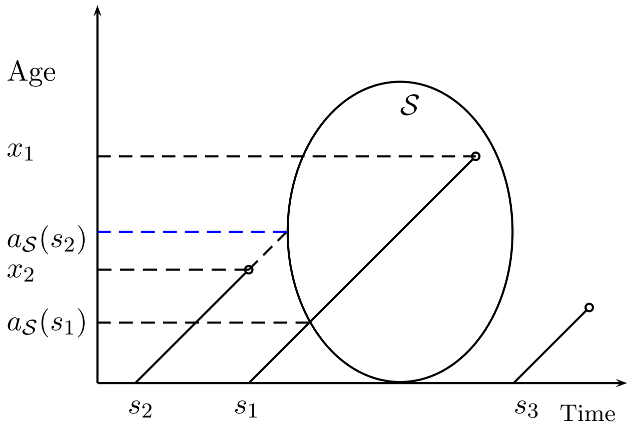


Figure 1.1: Diagramme de Lexis

Let $\sigma_{\mathcal{S}}$ denotes the birth time and $X_{\mathcal{S}}$ the lifetime for the included individuals. From now on, the pair $(\sigma_{\mathcal{S}}, X_{\mathcal{S}})$ will be referred to the observation opposed to the unobservable pair (σ, X).

Let the point process $\eta = \sum_{i \in I} \varepsilon_{\sigma_i}$, with the birth times as occurrence times, be a nonhomogeneous Poisson process on $\mathbb{R}$ with intensity φ (where ε_a is the Dirac measure at point a). Assume furthermore, that the lifetimes X_i, for $i \in I$, are i.i.d. with common probability density (p.d.f.) function f. The marking theorem ensures that the point process $\mu = \sum_{i \in I} \varepsilon_{(\sigma_i, X_i)}$ is a nonhomogenous Poisson process with intensity φf.

Here, we are interested in the individuals whose life-lines intersect the Borel set $\mathcal{S}$. In other words, we are interested in the restriction $\mu_{|\mathcal{S}}$ of the process μ to the Borel set $\mathcal{S}$. The restriction theorem ensures that the restriction $\mu_{|\mathcal{S}}$ is a Poisson process with mean measure $\int_{B \cap \mathcal{S}} \varphi f / \int_{\mathcal{S}} \varphi f$, for any Borel set B in $\mathcal{B}_{\mathbb{R} \times \mathbb{R}_+}$. Given the number $\mu(\mathcal{S})$ of points in the Borel set $\mathcal{S}$, the points of Poisson process $\mu_{|\mathcal{S}}$ look exactly like independent random variables, with common probability measure $\mathbb{P}(\cdot) = \int_{\cdot \cap \mathcal{S}} \varphi f / \int_{\mathcal{S}} \varphi f$ on Borel subsets of $\mathbb{R} \times \mathbb{R}_+$.
As a consequence, using Equation (1.19), we have, for all $s \in \mathbb{R}$ and $x \in \mathbb{R}_+$:

$$\begin{aligned} \mathbb{P}\left(\sigma_{\mathcal{S}} \leq s, X_{\mathcal{S}} \leq x\right) &= \frac{\iint_{]-\infty, s] \times [0,x]} \mathbf{1}_{\{(u,v) \in \mathcal{S}\}} \varphi(u) f(v) du dv}{\iint_{\mathbb{R} \times \mathbb{R}_+} \mathbf{1}_{\{(u,v) \in \mathcal{S}\}} \varphi(u) f(v) du dv} \\ &= \frac{1}{\mu_{\mathcal{S}}} \iint_{]-\infty, s] \times [0,x]} \varphi(u) f(v) \mathbf{1}_{\{a_{\mathcal{S}}(u) < \infty\}} \mathbf{1}_{\{a_{\mathcal{S}}(u) \leq v\}} du dv, \end{aligned} \quad (1.20)$$

where $\mu_{\mathcal{S}} = \iint_{\mathbb{R} \times \mathbb{R}_+} \mathbf{1}_{\{a_{\mathcal{S}}(u) < \infty\}} \mathbf{1}_{\{a_{\mathcal{S}}(u) \leq v\}} \varphi(u) f(v) du dv$. Hence the marginal distribution of the $X_{\mathcal{S}}$ is given, for all $x \in \mathbb{R}_+$, by :

$$F_{\mathcal{S}}(x) = \mathbb{P}\left(X_{\mathcal{S}} \leq x\right) = \frac{1}{\mu_{\mathcal{S}}} \int_0^x w(v) f(s) ds, \quad (1.21)$$

with

$$w(x) \quad = \quad \int_{-\infty}^{\infty} \mathbf{1}_{\{a_{\mathcal{S}}(u) \leq x\}} \varphi(u) du. \tag{1.22}$$

For example, in the time-window study, the weight function w is given, for $x \geq 0$, by:

$$w(x) = \int_{t_1 - x}^{t_2} \varphi(u) du.$$

In the particular case where $t_1 = t_2$ and φ is a constant, such a sample is called a "length-biased sample", see Asgharian *et al.* (2002) and de Uña-Àlvarez (2002).

In the example of the cohort study, the weight function w is constant and given, for $x \geq 0$, by

$$w(x) = \int_{t_1}^{t_2} \varphi(u) du.$$

Censoring

The lifetimes can also be subject to right-censoring. In this model, we can thus address the question of estimating the density f or the hazard rate λ of the underlying X, *without knowing the bias function.*

Now only the individuals whose life-lines intersect the Borel set $\mathcal{S}$ are included in the study. For an included individual i, with birth date $\sigma_{\mathcal{S},i}$ and lifetime $X_{\mathcal{S},i}$, we assume that its age at inclusion $a_{\mathcal{S}}(\sigma_{\mathcal{S},i})$ is observable. The lifetime $X_{\mathcal{S},i}$ can straightforwardly be written as follows:

$$X_{\mathcal{S},i} \quad = \quad \underbrace{a_{\mathcal{S}}(\sigma_{\mathcal{S},i})}_{\text{age at inclusion}} \quad + \quad \underbrace{(X_{\mathcal{S},i} - a_{\mathcal{S}}(\sigma_{\mathcal{S},i}))}_{\text{time spent in the study}}.$$

As the time spent in the study is given by $X_{\mathcal{S},i} - a_{\mathcal{S}}(\sigma_{\mathcal{S},i})$, we shall assume that this time can be censored. It would be the case, for example, if an individual i leaves the study before his death.

For that matter, we introduce a non-negative random variable C with density function h and c.d.f. H, independent of $X_{\mathcal{S}}$ and $a_{\mathcal{S}}(\sigma_{\mathcal{S}})$, such that the observable time for individual i is

$$Z_i = a_{\mathcal{S}}(\sigma_{\mathcal{S},i}) \quad + \quad (X_{\mathcal{S},i} - a_{\mathcal{S}}(\sigma_{\mathcal{S},i})) \wedge C_i.$$

As usual, we assume furthermore that the r.v. $\mathbf{1}_{\{X_{\mathcal{S},i} - a_{\mathcal{S}}(\sigma_{\mathcal{S},i}) \leq C\}}$ is observable. As a consequence, the available data are i.i.d. replications of:

$$\begin{cases} \sigma_{\mathcal{S},i} \\ Z_i = a_{\mathcal{S}}(\sigma_{\mathcal{S},i}) + (X_{\mathcal{S},i} - a_{\mathcal{S}}(\sigma_{\mathcal{S},i})) \wedge C_i, \quad \text{for i=1,\ldots,n.} \\ \mathbf{1}_{\{X_{\mathcal{S},i} - a_{\mathcal{S}}(\sigma_{\mathcal{S},i}) \leq C_i\}} \end{cases} \tag{1.23}$$

Counting processes for the estimation

In this context, Guilloux (2007) introduces the following counting processes. For all $x \geq 0$, let

$$D(x) = \sum_{i=1}^{n} \mathbf{I}_{\{Z_i \leq x, X_{\mathcal{S},i} - a_{\mathcal{S}}(\zeta_i) \leq C_i\}}. \tag{1.24}$$

For $x \geq 0$, the random variable $D(x)$ is the "number of observed deaths before age x" in the sample. Let furthermore the process O be defined, for all $x \geq 0$, by:

$$O(x) = \sum_{i=1}^{n} \mathbf{I}_{\{a_{\mathcal{S}}(\sigma_{\mathcal{S},i}) \leq x \leq Z_i\}} = \sum_{i=1}^{n} \mathbf{I}_{\{a_{\mathcal{S}}(\sigma_{\mathcal{S},i}) \leq x \leq X_{\mathcal{S},i}, x \leq a_{\mathcal{S}}(\sigma_{\mathcal{S},i}) + C_i\}}$$

The random variable $O(x)$ represents the "number of individuals at risk at age x". In the sampling situation considered here, to be at risk at age x for an individual means that it was included in the study at an age less than x and is neither dead nor censored before age x.

Let Λ denote the cumulative hazard function of X and be defined as:

$$\Lambda(x) = \int_0^x \frac{f(s)ds}{1 - F(s)},$$

for all $x \geq 0$. As usual in survival analysis, it seems natural to define its estimator $\widehat{\Lambda}$ by:

$$\widehat{\Lambda}_n(x) = \int_0^x \frac{dD(s)}{O(s) + n\epsilon_n}, \tag{1.25}$$

for all $x \geq 0$, where $(\epsilon_n)_{n \geq 1}$ is a sequence of positive numbers such that $\epsilon_n \to 0$ and $\sqrt{n}\varepsilon_n \to 0$ as $n \to \infty$, see Guilloux (2007).

We can also define the estimator $\widehat{F}_n$ for the cumulative distribution function F of X. Mimicking the construction of the Kaplan-Meier estimator, we define, for all $x \geq 0$:

$$\hat{F}_n(x) \quad = \quad 1 - \prod_{i: Z_{(i)} \leq x} \left(1 - \frac{\mathbf{I}_{\{X_{\mathcal{S},i} - a_{\mathcal{S}}(\sigma_{\mathcal{S},i}) \leq C_i\}}}{O(Z_{(i)}) + n\epsilon_n}\right),$$

where $\left(Z_{(i)}\right)_{i=1,\ldots,n}$ are the ordered statistics of the sample $(Z_i)_{i=1,\ldots,n}$.

Here, it is useful to mention the following result proved by Guilloux (2007):

Theorem 1 *Assume that there exists w_1 such that, for all $x \geq 0$, $w_1 \leq w(x)$. For all $u > 0$:*

$$\mathbb{P}\left(\sqrt{n} \sup_{x \geq 0} \left| \left(\hat{F}_n(x) - F(x)\right)(1 - H)(x) w_1 \right| > u\right) \leq 2.5 \exp(-2u^2 + Cu),$$

where C is an universal constant.

The following consequence of Theorem 1, is also useful in the sequel.

Lemma 1 *Assume that there exists w_1 such that, for all $x \geq 0$, $w_1 \leq w(x)$. For all $k \in \mathbb{N}^*$, there exists a constant $C_F(k)$ depending on k, w, $\mu_{\mathcal{S}}$ and c_G such that*

$$\mathbb{E}\left(\sup_{x\in A}|\hat{F}_n(x)-F(x)|^{2k}\right) \leq C_F(k)n^{-k}.$$

Remark: The condition $w_1 \leq w(x)$ holds for the window study case and for the cohort study as soon as the interior of $\mathcal{S} \cap \{(x,0), x \in \mathbb{R}\}$ is non empty. It means that a death cannot occur immediatly at the inclusion of an individual.

1.2.4 Adaptative estimation of the density and the hazard rate functions with unknown bias

In publication [10], we propose estimators of both density and hazard rate functions of X. We consider models S_m with dimension $D_m = m$ generated by trigonometric bases $[T]$ (see Appendix A). To avoid repetition, we present here only our methodology to estimate the hazard aret and we refer to [10] for the density estimation. Let us introduce the following contrast:

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^{n}\frac{\delta_i t(Z_i)}{O(Z_i)/n}. \tag{1.26}$$

The criterion $\gamma_n(t)$ is the empirical counterpart of $\|t\|^2 - 2\langle t,\lambda\rangle = \|t-\lambda\|^2 - \|\lambda\|^2$, with $\|t\|^2 = \int_0^1 t^2(x)dx$. By writing

$$\gamma_n(t) = \|t\|^2 - 2\int_0^1 t(x)d\widehat{\Lambda}_n(x)$$

with $\widehat{\Lambda}_n$ defined by (1.25) and taking $\epsilon_n = 0$. This contrast can be compared with the one proposed in [5] and with the works of Reynaud-Bouret (2006) for the Aalen's multiplicative intensity model.

Thus, the estimator built for one model S_m is defined by $\hat{\lambda}_m = \arg\min_{t\in S_m}\gamma_n(t)$, and the penalized estimator $\hat{\lambda}_{\hat{m}}$ by choosing $\hat{m}$ as follows:

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n}\left\{\gamma_n(\hat{\lambda}_m) + \text{pen}(m)\right\} \text{ where } \text{pen}(m) = \kappa\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{O^2(Z_i)/n^2}\right)\frac{m}{n}.$$

We give in Theorem 3.1 in [10] an oracle inequality for the $\mathbb{L}^2$-risk, from which we can deduce the minimax convergence rate. We refer to [10] for details about the theoretical results.

1.3 Estimation of the cumulative distribution function for current status data

Let X be a survival time with unknown cumulative distribution function (c.d.f) F. In the interval censoring case 1 model, we are not able to observe the survival time X. Instead, an observation consists of the pair (U, δ) where U is an examination time and δ is the indicator function of the event $(X \leq U)$. Roughly speaking, the only knowledge about the variable of interest X is wether it has occurred before U or not. Early examples of such interval censoring can be found in demography studies. In epidemiology, these censoring schemes also arise for instance in AIDS studies or more generally in the study of infectious diseases when the infection time is an unobservable event. We assume that U is independent of X, that F has density f and that the c.d.f G of U has density g. Such data, also known as current status data, may remind us right-censored data where the observed data is the pair $(\min(X, C), \mathbb{1}(X \leq C))$ where C is a censoring variable. However, the estimation procedure in these two censoring models is substantially different. Current status data have been studied by many authors in the last two decades, see Jewell & van der Laan (2004) for a state of the art. In the interval censoring model, the nonparametric maximum likelihood estimator (NPMLE) of the survival function is proved to be uniformly consistent, pointwise convergent to a nonnormal asymptotic distribution at the rate $n^{-1/3}$ in Groeneboom & Wellner (1992). In van de Geer (1993), it is also established that the NPMLE converges at rate $n^{-1/3}$ in $\mathbb{L}^2$-norm. Recent extensions take two directions. First, more general contexts are considered. For example, van der Vaart & van der Laan (2006) build nonparametric estimates of the survival function for current status data in presence of time dependent and high dimensional covariates: they provide limit central theorems with rate $n^{-1/3}$ and nonstandard limiting processes. The second direction aims at proposing smooth estimates that may take into account the possible smoothness of the survival function. Indeed, the NPMLE estimator is a piecewise constant function. The locally linear smoother proposed by Yang (2000), contrary to the NPMLE may be non monotone, but it has a better convergence rate than the NPMLE when the density f is smooth and the kernel function and the bandwidth are properly chosen. In the same spirit, Ma & Kosorok (2006) introduce an adaptive modified penalized least square estimator built with smoothing splines but their main objective is the study of semiparametric models. They have in mind the same type of penalization device that we present here, but their penalty functions contain many complicated terms that would be difficult to estimate. Here, we also pursue the search for smooth (or piecewise smooth) adaptive estimators. We present two different penalized minimum contrast estimators built on trigonometric, polynomial or wavelet spaces whose associated penalty terms are really simple; the minimization of the penalized contrast function allows to choose a space that leads to both a non asymptotic automatic squared bias/variance compromise and to an asymptotic optimal convergence rate according to the regularity of the function F in term of Besov spaces. An interesting feature of the procedure is that the estimators and their study is made straightforward by the most powerful Talagrand (1996) inequality for empirical centered processes. We also use technical properties proved in a regression framework by Baraud et al. (2001) and Baraud. (2002) for the mean-square estimator. Globally, the available tools and algorithms for adaptive density and regression estimation make our solution easy to study and to implement.

We have to mention extensions with multivariate or time-dependent covariates in van der Vaart & van der Laan (2006) or recently Plancade (2011) who proposes a generalization of our work.

Quotient strategy

We define the density ψ of the U_i's such that $\delta_i = 1$. As X and U are independent, we get:

$$\mathbb{P}(U \leq X, \delta = 1) = \mathbb{P}(U \leq X, X \leq U) = \int_0^x F(u)g(u)du$$

and thus,

$$\psi(x) = F(x)\, g(x). \tag{1.27}$$

This expression suggests a quotient estimator of ψ as follows:

Step 1 : Build an adaptive estimator of g:

$$\hat{g}_m = \arg\min_{t \in S_m} \gamma_n^g(t) \quad \text{with } \gamma_n^g(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^n t(U_i).$$

and $\tilde{g} = \hat{g}_{\hat{m}_g}$ with $\hat{m}_g = \arg\min_{m \in \mathcal{M}_n}[\gamma_n^g(\hat{g}_m) + \kappa\Phi_0^2 D_m/n]$.

Step 2 : Build an adaptive estimator $\tilde{\psi}$ of the density ψ :

$$\hat{\psi}_m = \arg\min_{t \in S_m} \gamma_n^\psi(t) \quad \text{with} \gamma_n^\psi(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^n \delta_i t(U_i),$$

and take $\tilde{\psi} = \hat{\psi}_{\hat{m}}$ by choosing $\hat{m} = \arg\min_{m \in \mathcal{M}_n}[\gamma_n^\psi(\hat{\psi}_m) + \widehat{\text{pen}}^\psi(m)]$, with

$$\widehat{\text{pen}}^\psi(m) = \kappa\Phi_0^2 \left(\frac{1}{n}\sum_{i=1}^n \delta_i\right)\frac{D_m}{n}.$$

Step 3 : Define the estimator of the c.d.f F by:

$$\tilde{F}(x) = \begin{cases} 0 & \text{if } \tilde{\psi}(x)/\tilde{g}(x) < 0 \\ \dfrac{\tilde{\psi}(x)}{\tilde{g}(x)} & \text{if } 0 \leq \tilde{\psi}(x)/\tilde{g}(x) \leq 1 \\ 1 & \text{if } \tilde{\psi}(x)/\tilde{g}(x) > 1 \end{cases}$$

Proposition 2 *If the density g of U satisfies $g(u) \geq g_0 > 0$, for all $u \in [0,1]$ and under some assumptions described in Lemma 3.1 in [12], then*

$$\mathbb{E}\|\tilde{F} - F\|^2 \leq \frac{2^4}{g_0^2}\left(\mathbb{E}\|\tilde{\psi} - \psi\|^2 + \mathbb{E}\|\tilde{g} - g\|^2\right) + \frac{C(g_0, \|\psi\|)}{n}, \tag{1.28}$$

with $C(g_0, \|\psi\|)$ a constant depending on g_0 and $\|\psi\|$.

Next with regularity assumptions $g \in \mathcal{B}_{\alpha_g,2,\infty}([0,1])$ and $\psi \in \mathcal{B}_{\alpha_\psi,2,\infty}([0,1])$, we deduce from Proposition 2, the rate of convergence of the estimator $\tilde{F}$. As a quotient strategy, if the regularity α_F is greater than the one of $\psi = Fg$, namely α_ψ,then we do not recover the optimal rate.

Direct strategy

The censoring mechanism is such that $\delta = \mathbb{I}_{(X \leq U)}$ given $U = u$ is a Bernoulli random variable with parameter $F(u)$:

$$\mathbb{E}(\delta | U = u) = F(u) \tag{1.29}$$

We can consider a standard mean-square regression contrast:

$$\gamma_n^{\text{MS}}(t) = \frac{1}{n} \sum_{i=1}^{n} [\delta_i - t(U_i)]^2$$

to define an estimator $\hat{F}_m = \arg\min_{t \in S_m} \gamma_n^{\text{MS}}(t)$. Then, we obtain the penalized esimator $\hat{F}_{\hat{m}_0}$ with:

$$\hat{m}_0 = \arg \min_{m \in \mathcal{M}_n} \{\gamma_n^{MS}(\hat{F}_m) + \kappa_0 \frac{D_m}{n}\}.$$

and κ_0 a numerical constant. We prove an adaptive result for $\hat{F}_{\hat{m}_0}$ and globally it appears that the regression estimation is better than the quotient estimator, from both theoretical and empirical points of view, as shown in the simulation study in [12].

Chapter 2

Conditional distribution for right-censored data

Publications [6], [7], [9] and [13]

In the previous chapter, we consider only curves associated with the marginal probability law of the time of interest. Herafter, we investigate conditional models to take into account the presence of covariables. Let us denote the time of interest Y, possibly right-censored, and $\vec{X}$ a covariate which is completely observed.

2.1 Regression in presence of one or more covariates

The model

However, for modeling the relationship between a response and a multivariate regressor, new methodologies have to be found especially to solve the problem of practical implementation in higher dimension. The main objective of the article is to propose a multivariate method of model selection for an additive regression function of a low-dimensional covariate vector. In fact, the particular case of additive models seems to be more realistic in practice and may constitute a way to make the dimension of the covariate greater than one. Suppose that $\vec{X}_i$ is a d-dimensional covariate in a compact set, without loss of generality we assume that $\vec{X}_i$ is taking value into $[0,1]^d$. Let $(\vec{X}_1, Y_1)$, $(\vec{X}_2, Y_2)$, $\ldots$, $(\vec{X}_n, Y_n)$ be independent identically distributed random variables. Let $T > 0$ be a fixed time for collecting the data. Therefore, the response variables *before censoring* are denoted by $Y_{i,T} = Y_i \wedge T$, where $a \wedge b$ denotes the infimum of a and b. Then, the model is defined for $i = 1, \ldots, n$, by :

$$\mathbf{E}(Y_{i,T}|\vec{X}_i) = r_T(\vec{X}_i) = r_{T,1}(X_i^{(1)}) + \cdots + r_{T,d}(X_i^{(d)}). \tag{2.1}$$

with $\vec{X}_i = (X_i^{(1)}, \ldots, X_i^{(d)})$. For identifiability we suppose that $\mathbf{E}(r_{T,j}(X_1^{(j)})) = 0$ for $j = 2, \ldots, d$. A comment on the model is required. The setting is analogous to Kohler *et al.* (2003)

and the constant T time is due to the fact that functionals of the survival function under censoring cannot be estimated on the complete support as mentioned by Gross and Lai (1996). Note that the fixed time T is not considered in the empirical setting and the procedure does work without any truncation. This fixed bound is introduced for technical and theoretical purposes. Of course, it is often mentioned that the function of interest would be r in the regression model $\mathbf{E}(Y_i|\vec{X}_i) = r(\vec{X}_i)$, instead of its biased version r_T. In addition, we suppose the C_i's to be independent of the $(\vec{X}_i, Y_{i,T})$'s for $i = 1, \cdots, n$ and that the c.d.f F and G are supported on the whole half-line $\mathbb{R}^+$. This assumption is not too restrictive since most of parametric survival models share it. Consequently,

$$\mathbb{P}(Y_i \geq T) = \mathbb{P}(Y_{i,T} = T) > 0 \text{ and } \mathbb{P}(C_i > T) > 0.$$

Thus, we state

$$\begin{cases} \forall i = 1, \dots, n,\ 1 - G(Y_{i,T}) \geq 1 - G(T) := c_G, \\ \\ \forall t \in [0, T],\ 1 - F(t) \geq 1 - F(T) := c_F > 0. \end{cases}$$

The method consists in building projection estimators of the d components $r_{T,1}, \dots, r_{T,d}$ on different projection spaces. The strategy is based on a standard mean-square contrast as in Baraud (2002) together with an optimized version of the data transformation proposed by Fan & Gijbels (1996).

The multivariate setting of additive models

For estimating additive regression function, the approximation spaces can be described as

$$S_m = \left\{ t(x^{(1)}, \dots, x^{(d)}) = a + \sum_{i=1}^{d} t_i(x^{(i)}),\ (a, t_1, \dots, t_d) \in \mathbf{R} \times \Pi_{i=1}^{d} S_{m_i} \right\}$$

where S_{m_i} is chosen as a piecewise polynomial space with dimension D_{m_i}.

The method of estimation

As usual in regression problems, a mean-square contrast can lead to an estimator of r_T. But we need first to transform the data to take the censoring mechanism into account.

We consider the following transformation of the censored data

$$\varphi_\alpha(Z) = (1 + \alpha) \int_0^Z \frac{dt}{1 - G(t)} - \alpha \frac{\delta Z}{1 - G(Z)}. \tag{2.2}$$

The main interest of the transformation is the following property: $\mathbb{E}(\varphi_\alpha(Z_1)|\vec{X}_1) = \mathbb{E}(Y_{1,T}|\vec{X}_1)$. Indeed,

$$\mathbb{E}\left[\frac{\delta_1 Z_1}{\bar{G}(Z_1)}|\vec{X}_1\right] = \mathbb{E}\left[\left(\frac{\delta_1 Y_{1,T}}{\bar{G}(Y_{1,T})}|\vec{X}_1, \varepsilon_1\right) | \vec{X}_1\right] = \mathbb{E}\left[\mathbb{E}\left(\delta_1|\vec{X}_1, \varepsilon_1\right) \frac{Y_{1,T}}{\bar{G}(Y_{1,T})}|\vec{X}_1\right] = \mathbb{E}\left(Y_{1,T}|\vec{X}_1\right),$$

$$
\begin{aligned}
\mathbb{E}\left[\int_0^{Z_1} \frac{dt}{1-G(t)} | \vec{X}_1\right] &= \mathbb{E}\left[\int_0^{+\infty} \frac{\mathbb{E}(\mathbf{I}_{Y_{1,T}\wedge C_1 \geq t} | \vec{X}_1, \varepsilon_1)}{1-G(t)} dt | \vec{X}_1\right] \\
&= \mathbb{E}\left[\int_0^{+\infty} \frac{\mathbf{I}_{Y_{1,T}\geq t}\mathbb{E}(\mathbf{I}_{C_1 \geq t} | \vec{X}_1, \varepsilon_1)}{1-G(t)} dt | \vec{X}_1\right] \\
&= \mathbb{E}\left[\int_0^{+\infty} \mathbf{I}_{Y_{1,T}\geq t} dt | \vec{X}_1\right] = \mathbb{E}(Y_{1,T} | \vec{X}_1).
\end{aligned}
$$

The transformation φ_α was considered by Koul *et al.* (1981) for $\alpha = -1$ and this case is often the only one studied in most of theoretical results. Leurgans (1987) proposed the transformation corresponding to $\alpha = 0$ and the general form (2.2) is described in Fan & Gijbels (1996). The main problem then lies in the choice of the parameter α. We experimented the proposition of Fan & Gijbels (1996) for this choice, but we did not find it really satisfactory. Therefore, we performed a choice of α in order to minimize the variance $\mathrm{Var}(\varphi_\alpha(\mathrm{Z}))$ of the resulting transformed data and took the empirical version of

$$
\hat{\alpha} = -\frac{\mathrm{cov}(\varphi_1(Z), \varphi_1(Z) - \varphi_2(Z))}{\mathrm{Var}(\varphi_1(Z) - \varphi_2(Z))}, \tag{2.3}
$$

with

$$
\varphi_1(Z) = \int_0^Z \frac{dt}{\bar{G}(t)}, \quad \varphi_2(Z) = \varphi_1(Z) - \frac{\delta Z}{\bar{G}(Z)}. \tag{2.4}
$$

In all cases, the transormed data are unobservable since we need to define $\hat{\bar{G}}$, a relevant estimator of $\bar{G}$. We propose to take the Kaplan-Meier (1958) product-limit estimator $\hat{\bar{G}}$, modified in the way suggested by Lo et al. (1989). Finally, by substituting $\bar{G}$ by its estimator $\hat{\bar{G}}$, we obtain the empirical version of the transformed data :

$$
\hat{\varphi}_{\hat{\alpha}}(Z) = (1+\hat{\alpha}) \int_0^Z \frac{dt}{\hat{\bar{G}}(t)} - \hat{\alpha} \frac{\delta Z}{\hat{\bar{G}}(Z)}. \tag{2.5}
$$

The mean-square contrast

The mean-square strategy leads to study the following contrast:

$$
\gamma_n(t) = \frac{1}{n} \sum_{i=1}^n [\hat{\varphi}_{\hat{\alpha}}(Z_i) - t(\vec{X}_i)]^2. \tag{2.6}
$$

In this context, it is useful to consider the empirical norm associated with the design

$$
\|t\|_n^2 = \frac{1}{n} \sum_{i=1}^n t^2(\vec{X}_i).
$$

Here we define

$$
\hat{r}_m = \arg \min_{t \in S_m} \gamma_n(t). \tag{2.7}
$$

The function $\hat{r}_m$ may be uneasy to define but the vector $(\hat{r}_m(\vec{X}_1), \ldots, \hat{r}_m(\vec{X}_n))'$ is always well defined since it is the orthogonal projection in $\mathbb{R}^n$ of the vector $(\hat{\varphi}_{\hat{\alpha}}(Z_1), \ldots, \hat{\varphi}_{\hat{\alpha}}(Z_n))'$ onto the subspace of $\mathbf{R}^n$ defined by $\{(t(\vec{X}_1), \ldots, t(\vec{X}_n))',\ t \in S_m\}$. This explains why the empirical norms are particularly suitable for the mean-square contrast.

Next, model selection is performed by selecting the model $\hat{m}$ such that:

$$\hat{m} = \arg\min_{m \in \mathcal{M}_n} \{\gamma_n(\hat{r}_m) + \text{pen}(m)\}, \tag{2.8}$$

where we have to determine the relevant form of pen(.) for $\hat{r}_{\hat{m}}$ to be an adaptive estimator of r.

Main result for the adaptive mean-square estimator

The automatic selection of the projection space can be performed via penalization and the following theoretical result is proved in [6], for the particular choice of $\hat{\alpha} = \alpha = -1$ i.e. for a contrast γ_n defined by (2.6) with variables $\hat{\varphi}_{-1}(Z_i) = \delta_i Z_i/(1 - \hat{G}(Z_i))$.

Assume that the common density f of the covariate vector $\vec{X}_i$ is such that $\forall x \in [0,1]^d, 0 < f_0 \leq f(x) < f_1 < +\infty$ and that the Y_i's admit moments of order 8. Consider the collection of models [DP] (defined in Appendix A) with $N_n \leq n/(16 f_1 K_\varphi)$ for [DP] where K_φ is a (known) constant depending on the basis. Let $\hat{r}_m$ be the adaptive estimator defined by (2.6) with $\hat{\alpha} = -1$ and (2.8) with

$$\text{pen}(m) = \kappa \Phi_0^2 \mathbf{E}\left[\left(\frac{\delta_1 Z_1}{\bar{G}(Z_1)}\right)^2\right] \frac{D_m}{n},$$

where κ is a numerical constant. Then

$$\mathbf{E}(\|\hat{r}_{\hat{m}} - r_T\|_n^2) \leq C \inf_{m \in \mathcal{M}_n} \left(\|r_m - r_T\|^2 + \text{pen}(m)\right) + C' \frac{\sqrt{\ln(n)}}{n}, \tag{2.9}$$

where r_m is the orthogonal projection of r_T onto S_m and C and C' are constants depending on Φ_0, $\|f\|$ and c_G.

The unknown expectation therein has to be replaced by

$$\widehat{\text{pen}}(m) = \kappa \hat{\sigma}^2 \Phi_0^2 \frac{D_m}{n}, \text{ with } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \left(\frac{\delta_i Z_i}{\hat{\bar{G}}(Z_i)}\right)^2 \tag{2.10}$$

where the constant κ is a universal numerical constant, determined by simulations experiments. It can be proved that the estimator obtained by random penalization still satisfies Inequality (2.9).

The left-hand side term of Inequality (2.9) shows that an automatic and non asymptotic trade-off is automatically performed between an unavoidable squared bias term $\|r_m - r_T\|^2$ and a term having a variance order pen(m). The nonasymptotic properties of the estimation algorithm can be appreciated when the selected model has small dimension but allows a good adequation between the true function and the estimate.

The asymptotic rates can be deduced, if a regularity assumption is set on the function to estimate

r_T. It is worth emphasizing here that the rates recovered in the addivitive d-dimensional model correspond to the one-dimensional rates ($d = 1$), whatever the number d of covariates. Indeed, when r_T has regularity β, the resulting minimax rates should be $n^{-2\beta/(2\beta+d)}$, see [6], with the standard loss due to the high dimension of the problem (Stone (1982)). The additive model has therefore undeniable virtues from a theoretical point of view and is illustrated through some simulations, see [9] .

Recently, *single-index* models have been explored by Lopez *et al.* (2013) for the estimation of the distribution of the pair $(Y, \vec{X})$. This appoach is an alternative to additive models for dimension reduction.

2.2 Conditional density and cumulative distribution functions

Publications [7] and [13].
Now, I explain how we get estimators for both conditional density and cumulative distribution functions. These papers are not specific to the context of censored data. However, we have those applications in mind and we always propose corrections to take censoring into account, but this does not constitutes the main contribution of publications [7] and [13].

First, we aim to estimate the conditional density $\pi(x, y)$ on a compact set $A = A_1 \times A_2$ from observations of $(X_1, Y_1), \ldots, (X_n, Y_n)$, i.i.d. pairs of random variables. The estimation of the conditional density of Y given $X = x$ brings more information than the regression model studied in the previous section, that only deals with the conditional expectation of the probability law. We can write:

$$\pi(x, y) = \frac{f_{X,Y}(x, y)}{f_X(x)} \quad \text{when } f_X(x) > 0$$

with the usual notations $f_{X,Y}$ and f_X standing respectively for the density of (X, Y) and the marginal density of X. Few references in the literature contribute to that subject, even in a non adaptive setting. We can cite: Györfi & Kohler (2007) for histogram estimators, Faugeras (2009) for kernel methods using copula, De Gooijer & Zerom (2003) for dependent data and we refer to the review in Lacour (2007) for the particular case of Markov chain with $Y_i = X_{i+1}$. Adaptive estimators have been proposed by Efromovich (2007) and Efromovich (2008). These works, as ours, deal with a possible anisotropy of the conditional density, that is different regularity in each direction.

2.2.1 Contrast stories

Density contrast

If our goal wa to estimate a univariate density f_Y of Y (on a bounded interval A_1) from observations $Y_1, \ldots, Y_n$, the usual contrast, associated to the loss function $\ell(f_Y, t) = \|f_Y - t\|^2$ would be

$$\gamma_n^{(1)}(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^{n} t(Y_i).$$

In fact, the function $\hat{f}_{m_1}$ minimizing $\gamma_n^{(1)}$ on the model $S_{m_1}^{(1)}$, described below or in Appendix A, is directly obtained as:

$$\hat{f}_{m_1} = \sum_{j\in J_{m_1}} \hat{a}_j \varphi_j^{m_1}, \quad \text{où } \hat{a}_j = \frac{1}{n}\sum_{i=1}^{n} \varphi_j^{m_1}(Y_i).$$

Now, let us denote $f_{m_1} = \sum_{j\in J_{m_1}} a_j \varphi_j^{m_1}$ the orthogonal projection of the function f_Y on $S_{m_1}^{(1)}$ (for the scalar product in $\mathbb{L}^2$). Pythagora's theorem allows to write:

$$\begin{aligned} \|f_Y - \hat{f}_{m_1}\|^2 &= \|f_Y - f_{m_1}\|^2 + \|f_{m_1} - \hat{f}_{m_1}\|^2 \\ &= \|f_Y - f_{m_1}\|^2 + \sum_{j\in J_{m_1}} (\hat{a}_j - a_j)^2, \end{aligned}$$

and then

$$\begin{aligned} \mathbb{E}(\|f_Y - \hat{f}_{m_1}\|^2) &= \|f_Y - f_{m_1}\|^2 + \frac{1}{n}\sum_{j\in J_{m_1}} \mathrm{Var}(\varphi_j^{m_1}(Y_1)) \\ &\leq \|f_Y - f_{m_1}\|^2 + \frac{1}{n}\sum_{j\in J_{m_1}} \mathbb{E}[(\varphi_j^{m_1}(Y_1))^2] \\ &\leq \|f_Y - f_{m_1}\|^2 + \phi_1 \frac{D_{m_1}^{(1)}}{n}. \end{aligned}$$

where the constant ϕ_1 depends on the basis $(\varphi_j)_{j\in J_{m_1}}$ and $D_{m_1}^{(1)}$ is the dimension of the model $S_{m_1}^{(1)}$.

Regression Contrast

Now, let us examine the case of a regression function r given by $\mathbb{E}(Y|X = x) =: r(x)$. From i.i.d. observations $(X_i, Y_i)_{1\leq i\leq n}$, one can define a mean-square contrast associated to the loss function $\|b - t\|_{f_X}^2$ with $\|t\|_{f_X}^2 = \int t^2(x) f_X(x) dx$. This $\mathbb{L}^2$-norm weighted by f_X the marginal density of the design is the natural reference norm appearing in regression problems. The mean-square contrast is:

$$\gamma_n^{(2)}(t) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - t(X_i))^2$$

or equivalently

$$\gamma_n^{(3)}(t) = \frac{1}{n}\sum_{i=1}^{n}(t^2(X_i) - 2Y_i t(X_i)).$$

As in the density case, the decomposition of the risk gives a bias term and a variance term of order $\phi_1 D_{m_1}^{(1)}/n$.

Density/regression mixed contrast

For the estimation of the conditional density $\pi(x, y)$ of Y given $X = x$, a *mixed* strategy is setting up. Before going further, we need to define model collection $\{S_m, m \in \mathcal{M}_n\}$ for functions defined in $\mathbb{R}^2$ with $\mathcal{M}_n$ a multiple index set (see Appendix A). For each pair of index $m = (m_1, m_2)$, S_m is a subspace of functions with support in $A = A_1 \times A_2$ defined as a product space $S^{(1)}_{m_1} \otimes S^{(2)}_{m_2}$ with $S^{(i)}_{m_i} \subset (L^2 \cap L^\infty)(\mathbb{R})$, for $i = 1, 2$, each one generated by two different (or not) orthonomal bases $(\varphi^{m_1}_j)_{j\in J_{m_1}}$ with dimension $|J_{m_1}| = D^{(1)}_{m_1}$ and $(\psi^{m_2}_k)_{k\in K_{m_2}}$ with $|K_{m_2}| = D^{(2)}_{m_2}$. Thus, we have

$$S^{(1)}_{m_1} = \{t \;/\; t(x) = \sum_{j\in J_{m_1}} a^{m_1}_j \varphi^{m_1}_j(x)\},$$

$$S^{(2)}_{m_2} = \{t \;/\; t(y) = \sum_{k\in K_{m_2}} a^{m_2}_k \psi^{m_2}_k(y)\}$$

and

$$S_m = S^{(1)}_{m_1} \otimes S^{(2)}_{m_2} = \{T \;/\; T(x, y) = \sum_{j\in J_{m_1}} \sum_{k\in K_{m_2}} A^m_{j,k} \varphi^{m_1}_j(x)\psi^{m_2}_k(y),\; A^m_{j,k} \in \mathbb{R}\}.$$

From i.i.d. observations $(X_i, Y_i)_{1\le i\le n}$, we define the contrast

$$\gamma^{(4)}_n(T) = \frac{1}{n}\sum_{i=1}^{n}\left(\int T^2(X_i, y)dy - 2T(X_i, Y_i)\right).$$

Comparing this contrast to $\gamma^{(1)}_n$ and $\gamma^{(2)}_n$ or $\gamma^{(3)}_n$, it can be interpreted as a density contrast with respect to the y-direction and as a regression contrast with respect to the x-direction.

2.2.2 Contrast for the estimation of the conditional density in presence of censoring

If the time Y id right-censored, we have to modify the contrast by using a similar transformation as in section 2.1, and this gives:

$$\gamma^{(4)}_n(T) = \frac{1}{n}\sum_{i=1}^{n}\left(\int T^2(X_i, y)dy - 2\hat{w}_i T(X_i, Z_i)\right) \text{ with } \hat{w}_i = \begin{cases} 1 & \text{if no censoring occurs} \\ \dfrac{\delta_i}{\hat{\bar{G}}(Z_i)} & \text{otherwise} \end{cases}$$

The existence of the minimizer of this constrast on a model S_m is not always guaranteed. But, we refer to proposition 2.1 in [7] which ensures that the function $\hat{\pi}_m(., .)$ at point (X_i, y):

$$(\hat{\pi}_m(X_i, y))_{1\le i\le n} = P_{\mathcal{W}}\left(\left(\sum_k \psi_k(Z_i)\psi_k(y)\right)_{1\le i\le n}\right)$$

is well-defined, where $P_{\mathcal{W}}$ is the orthogonal projector on $\mathcal{W} = \{(t(X_i, y))_{1\le i\le n}, t \in S_m\}$ for the euclidean scalar product in $\mathbb{R}^n$. Moreover, the natural empirical norm appearing in this problem is:

$$\|t\|_n = \left(\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} t^2(X_i, y)dy\right)^{1/2}$$

We prove in [7] that the empirical risk gives a bias term plus a variance term upper bounded by $\phi_1\phi_2 D^{(1)}_{m_1} D^{(2)}_{m_2}/n$ with ϕ_1 (resp. ϕ_2) basis-dependent constants. Then, the dimensions $D^{(1)}_{m_1}$ et $D^{(2)}_{m_2}$ have to be chosen in an optimal way and the resulting rate of convergence depends on both regularities of the function $\pi(x,y)$ in the x- and y-directions. The penalized estimator $\tilde{\pi} = \hat{\pi}_{\hat{m}}$ is defined by $\hat{\pi}_m = \mathrm{argmin}_{t\in S_m}\, \gamma^{(4)}_n(t)$, with $\hat{m} = \mathrm{argmin}_{m\in\mathcal{M}_n}\, \gamma^{(4)}_n(\hat{\pi}_m) + \mathrm{pen}(m)$ and the penalty term is

$$\mathrm{pen}(m) = \kappa\frac{\|\pi\|_\infty}{c_G} D_{m_1}D_{m_2}/n \text{ or } \ \mathrm{pen}(m) = \kappa\frac{\Phi_0}{f_0}\mathbb{E}\left(\frac{\delta_1}{\bar{G}^2(Z_1)}\right) D_{m_1}D_{m_2}/n$$

where we set

$$f_0 = \inf_{x\in A_1} f_X(x) \quad \text{and} \quad c_G = \inf_{y\in A_2} \bar{G}(y). \tag{2.11}$$

For both choices of the penalty, we prove an oracle inequality with assumptions slightly restrictive for the second one. Nevertheless, the latter is preferable in practice since it involves the empirical moment of the squared $\delta_1/\bar{G}(Z_1)$ instead of the constant c_G. Here, I refer to [7] for the detailed statement of the results. Note that our penalized estimator achieves the minimax rate thanks to a lower bound of the risk obtained by mimicking arguments developed in Lacour (2007).

Endly, Akakpo & Lacour (2011) aim at considering inhomogenous functions, namely $\pi \in \mathcal{B}^{\alpha}_{p,\infty}$ and $0 < p < 2$), whereas we have only considered $p = 2$ in [7]. Cohen & Lepennec (2011) explore model selection for maximum likelihood criterion associated with Kullback-Leibler loss. These recent contributions show the interest of the statistical community for the conditional density estimation.

2.2.3 Adaptive estimation of the conditional cumulative distribution function

Nonparametric methods for estimating the conditional c.d.f. $F(y|x)$ are most of the time not adaptive. Stute (1986), in the setting of completely observed data, or Dabrowska (1989), for censored data have studied the properties of the Beran's estimator, which is a generalisation of the Kaplan-Meier estimator in presence of covariates. These works consider kernel approaches, under the usual assumption that the unknown c.d.f is twice differentiable with respect to the variable x and the choice of the optimal bandwidth is not adaptive. On the contrary, we propose in [13], adaptive estimation procedure for the unknown c.d.f. by using a penalized contrast.

For the understanding of the contrast, let us examine first the case of the estimation of the c.d.f without any covariate. We just modify the density contrast $\gamma^{(1)}_n$ as follows :

$$\gamma^{(5)}_n(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^{n}\int t(y)\mathbf{1}_{(Y_i\le y)}dy,$$

which gives by minimization over $S_{m_2}^{(2)}$ the estimator

$$\hat{F}_{Y,m_2}(y) = \sum_{j\in J_{m_2}} \hat{a}_j \psi_j^{m_2}(y) \text{ with } \hat{a}_j = \frac{1}{n}\sum_{i=1}^n \int \psi_j^{m_2}(y)\mathbf{1}_{(Y_i\le y)}dy = \int \psi_j^{m_2}(y)F_n(y)dy.$$

This expression makes appear the estimator $\hat{F}_{Y,m_2}(y)$ as the orthogonal projection on $S_{m_2}^{(2)}$ of the standard empirical function c.d.f. $F_n(y) = (1/n)\sum_{i=1}^n \mathbf{1}_{(Y_i\le y)}$. This remark explains why we do obtain a smoother estimator than the empirical c.d.f provided the projection basis is smooth. However, there is no gain to hope regarding the rate of convergence since the empirical c.d.f. has already the parametric rate $n^{-1/2}$.

This behavior is still true in the y-direction when we add a covariable. We porpose the following contrast defined for a function $T \in S_m = S_{m_1}^{(1)} \otimes S_{m_2}^{(2)}$:

$$\Gamma_n^0(T) = \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}} (T^2(X_i,y) - 2T(X_i,y)\mathbf{1}_{(Y_i\le y)})dy.$$

We can remark that the contrast $\Gamma_n^0(T)$ behaves like a mean-square contrast in the x-direction. As in the univariate setting, the variance term does not depend on the dimension $D_{m_2}^{(2)}$ (the one associated with the y-direction) but only on the dimension $D_{m_1}^{(1)}$ of the covariate. This is why the compromise between the (squared) bias and variance term only involves one dimension, namely $D_{m_1}^{(1)}$, while dimension $D_{m_2}^{(2)}$ has to be chosen the largest as possible, exactly as in the univariate setting described above.

By computing the expectation of the contrast, we get

$$\mathbb{E}(\Gamma_n^0(T)) \quad = \quad \|T-F\|_{f_X}^2 - \|F\|_{f_X}^2$$

where $\|T\|_{f_X}^2 = \iint_A T^2(x,y)f_X(x)dxdy$. By the strong law of large numbers, $\Gamma_n^0(T)$ is the empirical counterpart of $\|T-F\|_{f_X}^2 - \|F\|_{f_X}^2$ and thus minimizing it leads to minimize $\|T-F\|_{f_X}^2$ in mean.

As before in the density case, we need to take right-censoring effect into account and we define, in the same way the following weighting of the contrast:

$$\hat{w}_i = \frac{\delta_i}{\hat{\bar{G}}(Z_i)}$$

where the estimator $\hat{\bar{G}}$ of the survival function $\bar{G} = 1 - G$ of the censoring variable C, has already been defined by (1.1). The contrast becomes in presence of censoring:

$$\Gamma_n(T) = \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}} (T^2(X_i,y) - 2\hat{w}_i T(X_i,y)\mathbf{1}_{(Z_i\le y)})dy.$$

Then, the penalized estimator is given by $\tilde{F} = \hat{F}_{\hat{m}}$ with

$$\hat{m} = \arg\min_{m_1\in\mathcal{M}_n} \{\Gamma_n(\hat{F}_{m_1}) + \text{pen}(m_1)\} \tag{2.12}$$

$$\text{pen}(m) \geq K_0 \ell(A_2) \xi \frac{D^{(1)}_{m_1}}{n}, \text{ with } \xi = \frac{\phi_1}{f_0} \mathbb{E}\left(\frac{\delta_1}{\bar{G}^2(Z_1)} \right) \text{ or } \xi = \frac{1}{c_G},$$

K_0 is a numerical constant and the constants f_0 and c_G are defined in the following assumptions [$\mathcal{A}1$] and [$\mathcal{A}2$], under which we are able to establish our Theorem 2:

[$\mathcal{A}1$] The density f_X satisfies $\|f_X\|_\infty := \sup_{x \in A_1} |f_X(x)| < \infty$ and there exists a constant $f_0 > 0$ such that, $\forall x$ in A_1, $f_X(x) \geq f_0$,

[$\mathcal{A}2$] $\forall y \in A_2$, $1 - G(y) \geq c_G > 0$.

For any function h and a given subspace S of $\mathbb{L}^2(A)$, we set

$$d(h, S) = \inf_{g \in S} \|h - g\| = \inf_{g \in S} \left(\iint |h(x, y) - g(x, y)|^2 dx dy \right)^{1/2}.$$

Theorem 2 *(see Theorem 4.1 in [13]) Under assumptions [$\mathcal{A}1$] and [$\mathcal{A}2$], consider the penalized estimator $\tilde{F}$ of the conditional c.d.f $F\mathbf{1}_A$ restricted to a compact A, with model collection described in Appendix A satisfying $\mathcal{D}^{(1)}_n \leq \sqrt{n}$, then, we have:*

$$\mathbb{E}\|F\mathbf{1}_A - \tilde{F}\|_n^2 \leq C \inf_{m_1 \in \mathcal{M}_n} \{ d^2(F\mathbf{1}_A, S^{(1)}_{m_1} \otimes \mathcal{S}^{(2)}_n) + \text{pen}(m_1) \} + \frac{C'}{n}$$

with C and C' constants depending on the problem.

Our method does not guarantee that we obtain a strict estimator (that is a c.d.f), in particular, the penalized estimator $\tilde{F}$ is not increasing. To rectify this drawback, we use the *a posteriori* rearrangement method proposed by Chernozhukov et al. (2009). For X_i, $i = 1, \ldots, n$,

$$\tilde{F}^*(X_i, y) = \inf \left\{ z \in \mathbb{R}, \int \mathbf{1}_{\{\tilde{F}(X_i, u) \leq z\}} du \geq y \right\}.$$

This gives an estimator that is a c.d.f with respect to the variable y. Then, the estimator $\tilde{F}^\star$ is a c.d.f:

$$\tilde{F}^\star(x, y) = \begin{cases} 0 & \text{if } \tilde{F}^*(x, y) < 0 \\ \tilde{F}^*(x, y) & \text{if } 0 \leq \tilde{F}^*(x, y) \leq 1 \\ 1 & \text{if } \tilde{F}^*(x, y) > 1 \end{cases}$$

Chapter 3

Estimation of regression functions by mean-square-type contrasts

Publications [14], [15].

After the work on conditional density and c.d.f functions, we overviewed in Chapter 2, new perspectives rise to other regression-type functions such the Mean Residual Life (MRL) and the hazard rate in presence of a covariable.

3.1 Conditional Mean Residual Life

In randomized clinical trials, survival times are often measured from randomization or treatment implementations. But studying survival functions or hazard rates may be inadequate to answer a patient asking during the trial, how much more time he still has or whether the new treatment improves his life expectancy. To correctly address these questions, life expectancy must be studied as a function of time, via the so-called mean residual life (MRL) function: We use the same notations as in Section 2.2 of Chapter 2 for the conditional c.d.f. and we introduce the MRL as:

$$e(y) = \mathbb{E}(Y - y | Y > y) \ , \ y > 0,$$

with $\mathbb{E}(Y) < +\infty$.
The MRL $e(y)$ can also be written:

$$e(y) = \begin{cases} \int_y^{+\infty} \bar{F}(u)du/\bar{F}(y) & \text{if } \bar{F}(y) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

This equality leads to several proposals of nonparametric estimators, built by plug-in of Kaplan-Meier survival estimators, see Hall & Wellner (1981), Csörgo & Zitikis (1996) and the references therein. Under adequate assumptions, these estimators inherit the parametric rates of the Kaplan-Meier estimator, but unfortunately they are not smooth. To circumvent this drawback, regularized estimators based on kernel smoothing have been proposed by Chaubey & Sen (1999) or Abdous & Berred (2005).

To measure the combined effect of a covariate X on the MRL, we shall rather define and study the conditional MRL:

$$e(y|x) = \mathbb{E}(Y - y|Y > y, X = x) = \begin{cases} \int_y^{+\infty} \overline{F}(u|x)du/\overline{F}(y|x) & \text{if } \overline{F}(y|x) > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

where $\bar{F}(y|x)$ is the conditional survival function of Y given $X = x$:

$$\bar{F}(y|x) = \mathbb{P}(Y > y|X = x) = \frac{\int_y^{+\infty} f_{(X,Y)}(x, u)du}{f_X(x)}.$$

Here $f_{(X,Y)}$ denotes the joint probability density of (X, Y) and f_X denotes the marginal density of X. In semi-parametric regression analysis, Oakes & Dasu (1990) propose a proportional mean residual life model to study the association with related covariates, when the response is completely observed: $e(y|X) = e_0(y) \exp(\beta^t X)$ where $e_0(y)$ is a baseline MRL and β is the regression parameter to be estimated. This model is studied in Maguluri & Zhang (1994). Then, Chen & Cheng (2005) and Chen *et al.* (2005) have developed strategies in this model for censored response.

Our goal is rather to develop purely nonparametric method to estimate the conditional MRL $e(y|x)$ on a compact set $A = A_1 \times A_2$. For this purpose, we have to find a mean-square-type contrast. Let $T : (x, y) \mapsto T(x, y)$ be a bivariate mesurable function with compact support $A = A_1 \times A_2$ and define:

$$\Gamma_n(T) = \frac{1}{n} \sum_{i=1}^{n} \left(\int T^2(X_i, y) \mathbf{1}_{(Y_i \geq y)} dy - 2\Psi_T(X_i, Y_i) \right)$$

with

$$\Psi_T(x, y) = \int_0^y (y - u)T(x, u)du.$$

We also introduce the two functions

$$\bar{F}_1(x, y) = \int_y^{+\infty} f_{(X,Y)}(x, u)du \quad \text{and} \quad \bar{F}_2(x, y) = \int_y^{+\infty} \bar{F}_1(x, u)du$$

supposed to be mesurable. Then, we can write:

$$e(y|x) = \frac{\bar{F}_2(x, y)}{\bar{F}_1(x, y)} \text{ if } \bar{F}_1(x, y) > 0.$$

Let the functions S and T be such that

$$\iint S^2(x, y)\bar{F}_1(x, y)dxdy < +\infty \quad \text{and} \quad \iint T^2(x, y)\bar{F}_1(x, y)dxdy < +\infty.$$

We define a μ-scalar product between S and T by:

$$\langle S,T\rangle_\mu = \iint S(x,y)T(x,y)d\mu(x,y) \text{ where } d\mu(x,y) = \bar{F}_1(x,y)dxdy$$

and the associated norm is denoted by $\|.\|_\mu$.

$$\mathbb{E}(\Gamma_n(T)) = \mathbb{E}\left(\int T^2(X_1,y)\mathbf{1}_{(Y_1\geq y)}dy - 2\Psi_T(X_1,Y_1)\right). \tag{3.2}$$

By computing separately the expectations involved in (3.2), we get on the one hand:

$$\begin{aligned}\mathbb{E}\left[\int T^2(X_1,y)\mathbf{1}_{(Y_1\geq y)}dy\right] &= \iint \left(\int T^2(x,y)\mathbf{1}_{(u\geq y)}dy\right) f_{(X,Y)}(x,u)dxdu \\ &= \iint T^2(x,y)\bar{F}_1(x,y)dxdy = \|T\|_\mu^2,\end{aligned}$$

on the other hand, since $\mathbb{E}\left[\Psi_T(X_1,Y_1)\right] < +\infty$

$$\begin{aligned}\mathbb{E}\left[\Psi_T(X_1,Y_1)\right] &= \iint \int_0^y (y-u)T(x,u)du\, f_{(X,Y)}(x,y)dxdy \\ &= \iint \left(\int \mathbf{1}_{(u\leq y)}(y-u)f_{(X,Y)}(x,y)dy\right) T(x,u)dxdu\end{aligned}$$

Here we need the assumption $\lim_{y\to+\infty} y\bar{F}_1(x,y) = 0$, for all $x\in A_1$ to make an integration by part:

$$\mathbb{E}\left[\Psi_T(X_1,Y_1)\right] = \iint T(x,u)\bar{F}_2(x,u)dxdu = \iint T(x,u)e(u|x)\bar{F}_1(x,u)dxdu = \langle T,e\rangle_\mu.$$

We have proved that: $\mathbb{E}(\Gamma_n(T)) = \|T\|_\mu^2 - 2\langle T,e\rangle_\mu$ and thus

$$\mathbb{E}(\Gamma_n(T)) = \|T-e\|_\mu^2 - \|e\|_\mu^2.$$

Note that the assumption $\lim_{y\to+\infty} y\bar{F}_1(x,y) = 0$ is not too restrictive.

We define a collection of models $\{S_m : m\in\mathcal{M}_n\}$ such that for each m, the subspace S_m is compactly supported on $A = A_1\times A_2$, more precisely:

$$S_m = S_{m_1}\otimes\mathcal{H}_n = \left\{T, \quad T(x,z) = \sum_{j\in J_m}\sum_{k\in\mathcal{K}_n} a_{j,k}\varphi_j^m(x)\psi_k(z),\ a_{j,k}\in\mathbb{R}\right\},$$

where S_{m_1} and $\mathcal{H}_n\subset(\mathbb{L}^2\cap\mathbb{L}^\infty)(\mathbb{R})$ are spanned by orthonormal bases $(\varphi_j^m)_{j\in J_m}$ and $(\psi_k)_{k\in\mathcal{K}_n}$. The dimension $|J_m| = D_m$ has to be properly chosen while $|\mathcal{K}_n| = \mathcal{D}_n^{(2)}$ is fixed the largest as possible. The behavior is the same as in the conditional c.d.f: no model selection is required in the y-direction because the conditional MRL behaves like a conditional c.d.f. As a consequence, the dimension $\mathcal{D}_n^{(2)}$ is fixed and corresponds to the dimension of the greatest subspace $\mathcal{H}_n$. On the contrary, the dimension D_m of S_{m_1} has to be chosen in an optimal way. In [14], we proved results for piecewise polynomial collections but we could consider other bases as well.

We have the following assumptions:

(A0) $\forall x \in A_1, \quad \lim_{y\to+\infty} y\bar{F}_1(x,y) = 0$,

(A1) $\exists \bar{F}_0, f_1 > 0$ such that $\forall (x,y) \in A_1 \times A_2,\ \bar{F}_1(x,y) \geq \bar{F}_0$ and $f_X(x) \leq f_1$.

(A2) $\forall (x,y) \in A_1 \times A_2,\ e(y|x) \leq \|e\|_{\infty,A} < +\infty$.

The procedure for computing the estimator has to be described.
Step 1 : We define one model for each dimension. By equating the gradient of the contrast to zero, $\hat{e}_m = \arg\min_{T\in S_m} \Gamma_n(T)$ leads to:

$$\forall j_0 \in J_m, \forall k_0 \in \mathcal{K}_n, \quad \frac{\partial \Gamma_n(T)}{\partial a_{j_0,k_0}} = 0$$

for any function T in S_m, we have $T(x,y) = \sum_{j\in J_m}\sum_{k\in\mathcal{K}_n} a_{j,k}\varphi_j^m(x)\psi_k(y)$. We can write equivalently:

$$G_m\hat{A}_m = \Upsilon_m,$$

where

- $\hat{A}_m$ stands for $vec((\hat{a}_{j,k})_{j\in J_m,k\in\mathcal{K}_n})$ the coefficients of the estimator in the basis,
- $G_m := \Big(\frac{1}{n}\sum_{i=1}^n \varphi_j^m(X_i)\varphi_\ell^m(X_i)\int \psi_k(z)\psi_p(z)\mathbf{1}_{\{Y_i\geq z\}}dz\Big)_{(j,k),(\ell,p)\in(J_m\times\mathcal{K}_n)^2}$
- $\Upsilon_m := vec\left(\Big(\frac{1}{n}\sum_{i=1}^n \varphi_j^m(X_i)\int_0^{Y_i}(Y_i-u)\psi_k(u)du\Big)_{j\in J_m,k\in\mathcal{K}_n}\right)$.

$vec(.)$ is the operator that stacks the columns of a matrix into a vector. *Step 2:* To solve the minimization problem, the matrix G_m has to be inversible. We easily verify that its eigenvalues are non negative. We refer to [14] for the details on algebra. We modify the definition of our estimator as follows:

$$\hat{e}_m := \begin{cases} \arg\min_{T\in S_m}\Gamma_n(T) & \text{on } \hat{H}_m \\ 0 & \text{on } \hat{H}_m^c \end{cases},$$

with the random set $\hat{H}_m := \Big\{\min \mathrm{Sp}(G_m) \geq \max(\hat{\bar{F}}_0/3, n^{-1/2})\Big\}$ and with $\mathrm{Sp}(G_m)$ the spectrum of G_m. For each m, on the set $\hat{H}_m$, the matrix G_m is inversible. The estimator $\hat{\bar{F}}_0$ of the bound $\bar{F}_0$ defined by (A1) satisfy the condition:

(A3) For all integer $k \geq 1$, $\mathbb{P}(|\hat{\bar{F}}_0 - \bar{F}_0| > \bar{F}_0/2) \leq C_k/n^k$.

Assumption (A3) allows to substitute $\hat{\bar{F}}_0$ to $\bar{F}_0$ and to study the risk on the random set on which the eigenvalues of G_m are positive. The precise definition of $\hat{\bar{F}}_0$ is given in [14]. This procedure is rather technical than the one we used in Section 2.2 for the condtional density or c.d.f. but our estimator of the conditional MRL e is well-defined on the whole compact A, instead to be defined only at the points (X_i, y). As a consequence, we can consider the $\mathbb{L}^2(A)$-risk (and not only the empirical risk as for the conditional c.d.f).

Step 3: Model selection has to be performed to obtain an optimal procedure: We chose $\hat{m}$ such that

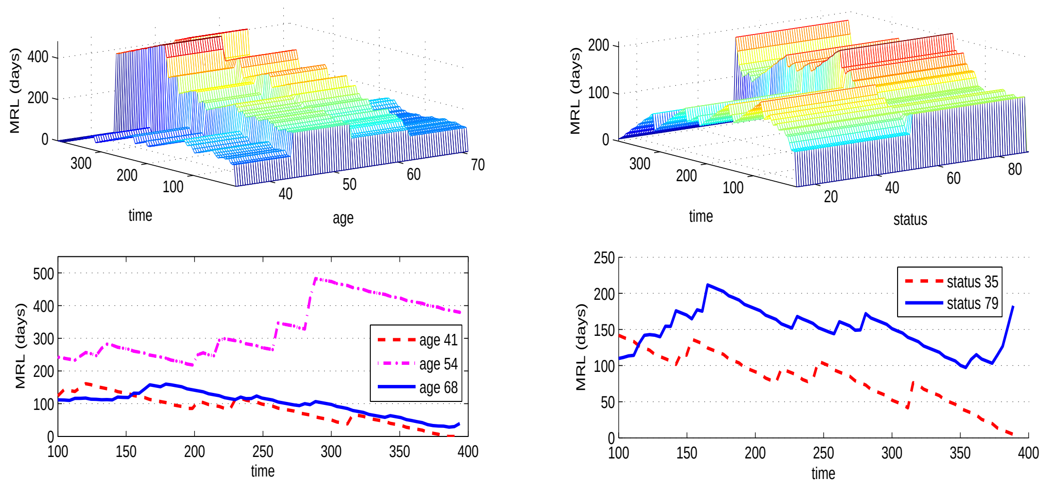
$$\hat{m} = \arg\min_{m\in\mathcal{M}_n} \Big(\Gamma_n(\hat{e}_m) + \mathrm{pen}(m)\Big),$$

where

$$\mathrm{pen}(m) = \kappa\phi_1 \frac{\mathbb{E}(Y_1^3) + \ell(A_2)\mathbb{E}(Y_1^2)}{\bar{F}_0}\frac{D_m}{n},$$

Our estimator of e on the compact A is defined by $\tilde{e} := \hat{e}_{\hat{m}}$. With additional assumptions on the model collections (and in particular their maximal dimension), we can prove that our estimator is adaptive and achieves automatically the optimal rate of convergence, see Theorem 3.1 and Corollary 3.1 in [14].

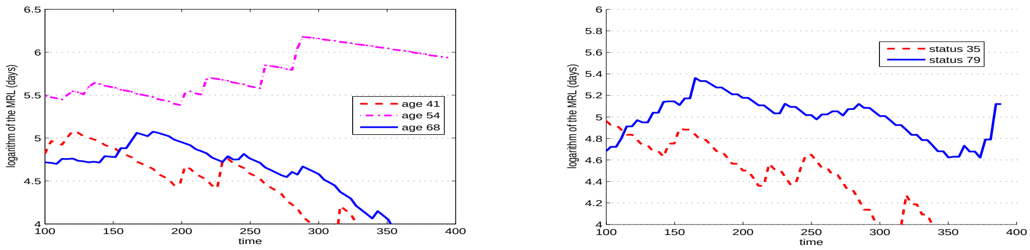
Figure 3.1: MRL Estimator as a function of the time and the age (left-top) and time and autonomy score (right-top). Estimator of the MRL given for different fixed values of age 41, 54 and 68 (left-bottom) and for autonomy score 35 and 79 (right-bottom).



Application

Let us conclude this section with an example. As an illustration, we consider data from the Veterans Administration Lung Cancer Trial presented by Prentice (1973), in which males with advanced inoperable lung cancer received chemotherapy. Several covariates were observed for each patient. Hereafter, we study the survival times of 128 uncensored patients which range from 1 to 587 and we focus on the patients age as a covariate. These data were also studied in

Figure 3.2: MRL Estimator in logarithmic scale, as a function of the age (left) and autonomy score (right).



Chen & Cheng (2005) but only for a subgroup of 97 patients with no prior therapy, and for two other covariates: a categorical covariate which gives the tumor type, which we cannot handle with our method and the performance status which is a therapeutic score ranging from 0 to 100 and evaluating the autonomy of the patient. The curve shows three parts with respect to the age, with change points corresponding somehow to early, middle and old ages. This behaviour also appears on Figure 3.1 where the estimated MRL is shown for three fixed values of ages 41, 54 and 68. While the MRL is quite the same for ages 41 and 68, it is of interest that the MRL is twice better for middle age 54. We could think that there is an optimal age (the middle age) to receive the treatment. We also estimate the MRL conditionally to the performance status as a continuous covariate with the subgroup of 97 patients with no prior therapy. The estimated MRL is shown in Figure 3.1. For the performance status, we detect as for the age two parts in the curve corresponding to low and high status. Moreover, our purely nonparametric estimator gives a new graphical method to check whether the proportionality assumption of the Oakes-Dasu model is true or not. Then, in log-scale, the curves in Figure 3.2 should be parallel to each other for any fixed value of the covariate. But, it seems that for the covariate age, the Oakes-Dasu model cannot be reasonably considered. Besides, for the covariate performance status, as mentioned in the work by Chen & Cheng (2005), after an initial period from 0 to 150 days, the curves corresponding, respectively, to statuses 35 and 79 appear quite parallel, which may suggest the adequacy of the proportionality assumption.

3.2 Conditional hazard rate

Publication [15]

We consider in [15] the problem of estimation from right-censored data in presence of covariates, when the censoring indicator is missing. Let T be a random variable representing the time to death from the cause of interest. Let C denote a right-censoring random time. Under usual random censorship, the observation is $Y = T \wedge C$ and $\delta = \mathbf{1}(T \leq C)$. Let X denote a real covariate. In what follows, it is assumed that T, C and X admit densities respectively denoted by f_T, g and f_X. In addition, C is assumed to be independent of T conditionally to X. When the cause of death is not recorded, the censoring indicator is missing: this is the missing censoring indicator (MCI) model, see Subramanian (2004), which is defined as follows. Let ξ be the missingness indicator, that is $\xi = 1$ if δ is observed and $\xi = 0$ otherwise. The observed data are then given for individual $i \in \{1, \dots, n\}$:

$$(Y_i, X_i, \delta_i, \xi_i = 1) \quad \text{or} \quad (Y_i, X_i, \xi_i = 0).$$

We shall say that the model is:

- MCAR under the assumption that the indicator are Missing Completely At Random, i.e. ξ is independent of T, C and X.
- MAR under the assumption that the indicator is Missing At Random i.e. ξ and δ are independent conditionally to Y, X.

In [15], we mainly concentrate on the MAR model. This model has been considered by several authors in the last decade. Most papers are interested in survival function and cumulative hazard rate estimation. In particular, van der Laan and McKeague (1998) improve Lo (1991)'s paper and build a sieved nonparametric maximum likelihood estimator of the survival function in the MAR case. Their estimator is a generalization of the Kaplan & Meier (1958) estimator to this context and is the first proposal reaching the efficiency bound. Subramanian (2004) also proposes an efficient estimator of the survival function in the MAR case; he proves his estimate to be efficient as well. Gijbels *el al.* (2007) study semi-parametric and nonparametric Cox regression analysis in several contexts.

Kernel methods have also been used to build different estimators in the MAR context. Subramanian (2006) estimates the cumulative hazard rate with a ratio of kernel estimators. Recently, Wang *et al.* (2009) proposed density estimator based on kernels and Kaplan Meier-type corrections of censoring. They prove a CLT and suggest a bandwidth selection strategy. Extensions of these works to conditional functions (both cumulative hazard and survival functions) in the presence of covariates is developed in Wang & Shen (2008).

Note that several authors (Dikta (1998), and more recently Subramanian (2009), Subramanian (2011)) study semiparametric models for the missing process and imputation methods – but for Kaplan-Meier estimator – while we remain in pure nonparametric setting.

Both our method and our aim are rather different. We indeed consider the estimation of the conditional hazard rate given a covariate. Moreover, we provide a nonparametric mean square strategy by considering approximations of the target function on finite dimensional linear spaces spanned by convenient and simple orthonormal (functional) bases. A collection of estimators is thus defined, indexed by the dimension of the multidimensional projection space, and a penalization device allows us to select a "good" space among all the proposals.

Contrary to standard kernel methodology, our estimator has the advantage of being defined as a contrast minimizer and not a ratio of two estimators, see Wang & Shen (2008), Subramanian (2006). It depends on an unknown function, in its definition, which has to be replaced by an estimator; this step is shared by the kernel approach. However, our precise study of the plug-in estimator allows us to non-asymptotically control the mean square risk. From an asymptotic point of view, we provide anisotropic rates corresponding to the regularity of the function under estimation, plus the rate of the intermediate plug-in estimator.

3.2.1 Construction of the estimator in the MAR case

We aim at estimating the conditional hazard function on a compact $A = A_1 \times [0, \tau]$:

$$\lambda(x, y) = \lambda_{Y|X}(x, y) = \frac{f_{Y|X}(x, y)}{1 - F_{Y|X}(x, y)},$$

with $f_{Y|X}$ and $F_{Y|X}$ the conditional density and c.d.f. of Y given X. We define two functions ξ and δ playing the role of "nuisance" parameters:

$$\pi(x, y) = \mathbb{E}(\xi | X = x, Z = y)$$
$$\zeta(x, y) = \mathbb{E}(\delta | X = x, Z = y).$$

Construction of the contrast with no covariable

For simplicity, we explain the construction of the contrast without any covariable. The hazard rate is simply

$$\lambda(y) = \frac{f_Y(y)}{1 - F_Y(y)}.$$

If $\zeta(y) = \mathbb{E}[\delta | Z = y]$ was known, a contrast to estimate λ would be:

$$\Gamma_n^{th}(h) = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau h^2(y) \mathbf{1}_{(Z_i \geq y)} dy - \frac{2}{n} \sum_{i=1}^{n} \left(\xi_i \delta_i + (1 - \xi_i)\zeta(Z_i)\right) h(Z_i).$$

In fact, under the MAR assumption,

$$\begin{aligned} \mathbb{E}(\delta_i \xi_i + (1 - \xi_i)\zeta(Z_i) | Z_i) &= \mathbb{E}(\delta_i | Z_i)\mathbb{E}(\xi_i | Z_i) + \mathbb{E}[(1 - \xi_i)\mathbb{E}(\delta_i | Z_i) | Z_i] \\ &= \mathbb{E}(\mathbb{E}(\delta_i | Z_i)(\xi_i + (1 - \xi_i)) | Z_i) \\ &= \zeta(Z_i) \end{aligned}$$

and the expectation of the contrast is

$$\begin{aligned} \mathbb{E}\big(\Gamma_n^{th}(h)\big) &= \int h^2(y) d\mu(y) - 2 \int h(y)\lambda(y) d\mu(y) \\ &= \|h\|_\mu^2 - 2\langle h, \lambda \rangle_\mu = \|h - \lambda\|_\mu^2 - \|\lambda\|_\mu^2. \end{aligned}$$

where we set $d\mu(y) = (1-L)(y)dy$ and $(1-L)(y) = (1-F_Y)(1-G)(y) = \mathbb{P}(Z \geq y)$.

But, as ζ is unknown, we have to substitute an estimator $\tilde{\zeta}$ and the contrast Γ_n^{th} becomes:

$$\Gamma_n^0(h) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau h^2(y)\mathbf{1}_{(Z_i \geq y)}dy - \frac{2}{n}\sum_{i=1}^n \left(\xi_i\delta_i + (1-\xi_i)\tilde{\zeta}(Z_i)\right) h(Z_i). \tag{3.3}$$

Extension of the contrast with a covariable

The contrast (3.3) can be extend to the presence of a covariable:

$$\Gamma_n(h) \quad = \quad \frac{1}{n}\sum_{i=1}^n \int_0^\tau h^2(X_i,y)\mathbf{1}_{(Z_i \geq y)}dy - \frac{2}{n}\sum_{i=1}^n \left(\xi_i\delta_i + (1-\xi_i)\tilde{\zeta}(X_i,Z_i)\right) h(X_i,Z_i), \tag{3.4}$$

for the "reference" norm $\iint h^2(x,y)d\mu(x,y) = \iint h^2(x,y)(1-L_{Z|X}(y,x))f_X(x)dxdy$ and

$$1 - L_{Z|X}(y,x) := \mathbb{P}(Z \geq y|X=x) = (1-F_{Y|X}(x,y))(1-G_{C|X}(x,y)).$$

Estimation of ζ

Now, we come back to the estimation of the auxiliary function ζ. That is the key of our estimation strategy. In [15], we propose a mean-square criterion for the estimation of $\zeta(x,y) = \mathbb{E}(\delta|X=x, Z=y)$ on A:

$$\hat{\zeta}_m = \arg\min_{T \in S_m} \tilde{\gamma}_n(T) = \frac{1}{n}\sum_{i=1}^n [\xi_i T^2(X_i,Z_i) - 2\xi_i\delta_i T(X_i,Z_i)].$$

Then, we obtain the penalized estimator $\tilde{\zeta} := \hat{\zeta}_{\tilde{m}}$ with:

$$\tilde{m} = \arg\min_{m \in \mathcal{M}_n} \tilde{\gamma}_n(\hat{\zeta}_m) + \tilde{\kappa}\frac{dim(S_m)}{n}.$$

The estimator $\tilde{\zeta}$ can be substituted to ζ in Γ_n. We control the $\mathbb{L}^2$-risk (or equivalent norm) of $\tilde{\zeta}$.

The lack of this procedure is to apply in two steps, with a *plug-in* of the estimator of ζ in Γ_n. As a consequence, we prove an oracle inequality where both errors of λ and ζ add up. The resulting rate of convergence is thus, see Corollary 2 in [15]:

$$\mathbb{E}\left(\|\lambda - \hat{\lambda}_{\hat{m}}\|_A^2\right) = O(n^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+2}}) + O(n^{-\frac{2\bar{\beta}}{2\bar{\beta}+2}}),$$

that is the estimator $\hat{\lambda}_{\hat{m}}$ would achieve the optimal rate provided the function ζ is most regular than λ.

MCAR case

Under the MCAR assumption, that is ξ independent of T, C (and possibly the covariable X if there is any), we can consider the following contrast:

$$\gamma_n^{(1)}(h) = \frac{1}{n}\sum_{i=1}^{n}\int_0^1 h^2(y)\xi_i \mathbb{1}(Y_i \geq y)dy - \frac{2}{n}\sum_{i=1}^{n}\delta_i\xi_i h(Y_i),$$

here the "reference" measure is $d\mu(y) = \mathbb{E}(\xi)(1 - L(y))dy$. This contrast doesn't need the estimation step of ζ but it uses only the observations such that $\xi_i = 1$.

3.2.2 Extensions

We envisage other proposals of contrasts based on the function $\pi(x, y) = \mathbb{E}(\xi|X = x, Z = y)$:

$$\Gamma_n^{\text{EST}}(h) = \frac{1}{n}\sum_{i=1}^{n}\int_0^1 h^2(X_i, y)\tilde{\pi}(X_i, y)\mathbb{1}(Z_i \geq y)dy - \frac{2}{n}\sum_{i=1}^{n}\delta_i\xi_i h(X_i, Y_i). \tag{3.5}$$

Here, we also have to estimate the function π. The function π is just a regression function $\mathbb{E}(\xi|X, Z)$ and an estimator can be obtained directly from the complete sample.

We are interested in comparison of *imputation versus regression* approaches. Let γ_i be a Bernoulli random variable with parameter $\zeta(Y_i)$ given $\xi_i = 0$. The quantity $\zeta(Y_i)$ has to be estimated as well as in the estimation approach.
Then, we define a contrast where the missing value are substituted by imputation:

$$\Gamma_n^{\text{IMP}}(h) = \frac{1}{n}\sum_{i=1}^{n}\int_0^1 h^2(y)\mathbb{1}(Z_i \geq y)dy - \frac{2}{n}\sum_{i=1}^{n}\left(\xi_i\delta_i + (1 - \xi_i)\gamma_i\right)h(Z_i). \tag{3.6}$$

We study this strategy in [18].

Chapter 4

Works in progress, perspectives

In this last chapter, I expose recent works and perspectives. Functional data and more generally high-dimension methods have initiated the working group "ADONF" (Analyse des Données Fonctionnelles in french) in Montpellier. This group aim at encouraging interactions between the researchers of the Institute of Mathematics at University and our colleagues at INRA/SupAgro. From now on, I turn my works towards functional data. Since 2011, I supervise with André Mas, the Phd thesis of Angelina Roche on *Model selection for functional data*. There are many connexions with regression models I have studied before and future works could be devoted to fonctional covariables in survival analysis; I give a brief overview in the sequel. I'm going on with other themes like inference for recurrent events or confidence bands.

Functional data

Prépublications [16], [17]

Functional data analysis have known recent advances in the past two decades. In many practical situations, we aim to predict values of a scalar response by using functional predictors, or roughly speaking, curves. Many fields of applications are concerned with this kind of data, such as medicine, chemometrics or econometrics. This is especially the case when people have to predict electric consumption from a daily temperature curve, or in medicine when spectrometric signals are used to detect abnormality. We refer to Ferraty & Vieu (2006) and Ramsay & Silverman (2005) for detailed examples. In [16], we focuse on the functional linear model, where the dependence between a scalar response Y and the functional random predictor X is given by:

$$Y = \int_0^1 \beta(t)X(t)dt + \varepsilon, \tag{4.1}$$

where the centred random variable ε stands for a noise term with variance σ^2 and is independent of X. Our aim is to estimate the unknown slope function β from an independent and identically distributed (i.i.d.) sample (X_i, Y_i), for $i = 1, \cdots, n$. In the sequel, we suppose that the random function X takes value in $\mathbf{L}^2(A)$ with A a compact set and for sake of simplicity, we fix $A = [0, 1]$. We recall that the usual inner product $\langle ., . \rangle$ in $\mathbf{L}^2[0, 1]$ is defined by $\langle f, g \rangle = \int_0^1 f(u)g(u)du$ for all $f, g \in \mathbf{L}^2[0, 1]$. The random curve X will be supposed to be centred and periodic, that is to say the function $s \mapsto \mathbf{E}[X(s)]$ is identically equal to zero and $X(0) = X(1)$. This context matches the description of circular data considered in Comte & Johannes (2010).

By multiplying both sides of Equation (4.1) by $X(s)$ and by taking expectation, we easily obtain:

$$\mathbf{E}[Y X(s)] = \int_0^1 \beta(t)\mathbf{E}[X(t)X(s)]dt =: \Gamma\beta(s), \text{ for all } s \in [0, 1], \tag{4.2}$$

where Γ is the covariance operator associated to the random function X. Then, the problem of the estimation of β is related to the inversion of the covariance operator Γ or of its empirical version:

$$\Gamma_n := \frac{1}{n}\sum_{i=1}^{n} \langle X_i, . \rangle X_i.$$

Many authors have studied the functional linear model. Strategies using regression on functional principal components have been proposed by Cardot *et al.* (2003) or Cardot *et al.* (2007) among others. The estimator of the slope function is usually obtained on the linear space spanned by the first eigenfunctions associated to the greatest eigenvalues of the empirical covariance operator Γ_n. Although the resulting estimator is shown to be convergent, its behaviour is often erratic in simulation studies. Smoothing splines estimator minimizing a standard least squares criterion has been improved by Crambes *et al.* (2009) with a slight modification of the usual penalty. The authors have shown that rates of convergence for the risk defined by the mean squared prediction error depend on both the smoothness of the slope function and the structure of the covariance operator (in particular, the decreasing rate of the eigenvalues). They also prove that the obtained rates are minimax over large classes of slope functions. In a different way, Cardot & Johannes (2010) propose a thresholded projection estimator to circumvent instability problems, which can reach optimal convergence rate for the risk associated with the mean squared prediction error. Their techniques based on dimension reduction follow inverse problems ideas starting from Efromovich & Koltchinskii (2001) and covered more recently by Hoffmann & Reiss (2008). But all theses procedures depend on one or more tuning parameters, which is a difficult problem to solve in practice.

Recently, Verzelen (2010) has proposed an adaptive estimation procedure for the slope coefficient, say $\theta \in \mathbf{R}^p$, in the context of high dimensional regression models. He has obtained an oracle-inequality for the risk associated with the prediction error on $\mathbf{R}^p$, which remains true when $p >> n$. Any knowledge on the covariance matrix of the design is required but his results are obtained under the assumption that both the design and the noise are gaussian. We can also mention the different but related works of the inverse problem community such as Cavalier

& Hengartner (2005) among others.

Cai & Hall (2006) addressed the problem of prediction from an estimator of the slope function. In this work, the choice of the tuning parameter plays an important role in the performance of the estimators. The usual practical strategy of empirically choosing the smoothing parameter value is performed through the generalized cross validation.

But nonasymptotic results providing adaptive data-driven estimators were missing up to the recent paper by Comte & Johannes (2012). They have proposed a model selection procedure for the orthogonal series estimator introduced by Cardot & Johannes (2010). The resulting estimator is completely data-driven and it is shown to achieve optimal minimax rates for general weighted $\mathbf{L}^2$-risks. The dimension is selected by minimization of a penalized contrast which requires the knowledge of the sequence of weights defining the risk, the aim being to estimate accurately the slope function and its derivatives. In this sense their work is more general than ours. Nevertheless, we explain hereafter why their data-driven estimation procedure cannot encompass the prediction error, which can be seen as a particular weighted risk whose weights are the unknown eigenvalues of the covariance operator.

At a first sight, our penalized estimator may look like a special case of the Comte and Johannes's one in the case of the prediction error, but it is definitely not since their penalty term involves the weights defining the risk, that is to say the unknown eigenvalues. In our paper, even though a less general risk is handled, we propose a very simple data-driven procedure to select the adequate dimension of the functional space over which the standard mean square contrast is minimized. We want to emphasize that the prediction error is of particular interest in applications and thus takes an important place in most of the papers related to functional linear models. Though, our goal differs mainly from Comte & Johannes (2010) but the tools are those of model selection as well. Our penalized estimator is proved to satisfy an oracle-type inequality for the risk associated to the prediction error and to reach optimal rates for slope functions belonging to Sobolev classes.

A second part of the work, more ambitious, consists in using the eigen elements of the empirical covariance operator. In [17], we do not deal with fixed bases anymore, but wit random bases and the control of the bias term of the estimator becomes difficult to handle. Thanks to perturbation theory, we obtain in this context an oracle inequality for the functional principal component estimator. Our proposal gives an adaptive estimator which achieves the optimal and minimax rates over Sobolev classes.

There are numerous perspectives around functional data. In particular, regression functions in survival analysis can possibly involve functional covariables (mass spectrometry for the renal impairment, medical imaging, blood sugar or blood pressure curves,...). I think that the *single index* model, could have interesting developments for functional covariable X: $Y = g(\langle \beta, X\rangle)+\varepsilon$.

A recent paper by Chen *et al.* (2011) consider this context, but up to my knowledge, adaptive procedures do not exist for estimating simultaneaously both functions g and β.

Recurrent events (work in progress)

With Segolen Geffray, we are working on nonparametric estimation for cumulative distribution functions for recurrent events in presence of right-censoring. In epidemiology, the patients can suffer from recurrent events, like tumoral recidives or opportunist diseases, asthma crisis or heart failure. We refer to Schaubel & Cai (2004) or Cook & Lawless (2008) for examples of such data. We consider situations where the death of a patient occur with high probability. Roughly speaking, the observation of a recurrent event or death can be observed or not if censoring occurs.

An illustration is given in figure 4.1 for 6 patients S_1, ..., S_6. The observation of a patient is represented by a line and our aim is to study the second recurrent event.

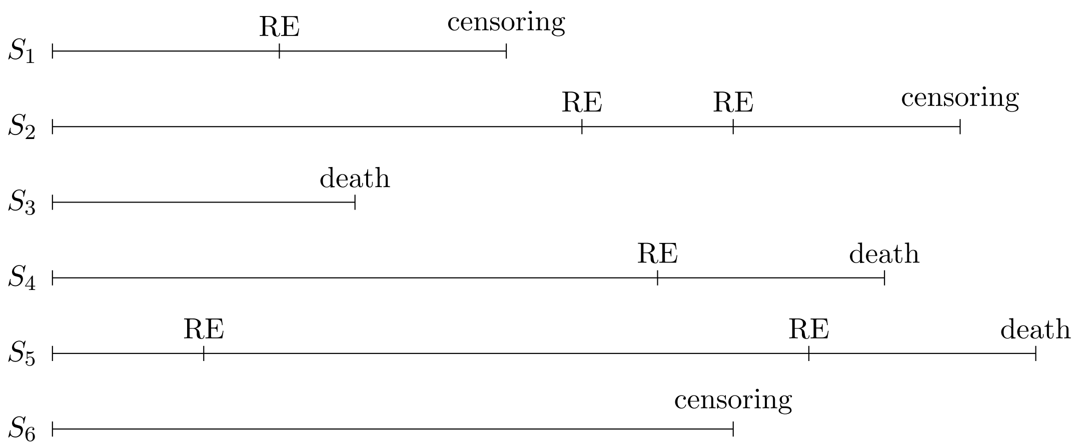


Figure 4.1: Recurrent Events (RE = event of interest).

In addition, we assume that recurrent events of interest, terminal events (death) and censoring events cannot occur at the same time. The statistical scheme is as follows:

- Let K be the total number of observed events for one individual.
- For $k = 1, \ldots, K$, $Y^{[k]}$ is the time between the $(k-1)$th and the kth event.
- For $k = 1, \ldots, N$, $\mathcal{C}^{[k]}$ is the indicator of the type of the kth event:

$$\mathcal{C}^{[k]} = \begin{cases} 1 & \text{for a recurrent event,} \\ 2 & \text{for a death.} \end{cases}$$

- For $k = 1, \ldots, K$, $Z^{[k]} = \min(Y^{[k]}, C - \sum_{l=0}^{k-1} Y^{[l]})$ stands for the time between the $(k-1)$th and the kth observed event (with $Y^{[0]} = 0$).
- For $k = 1, \ldots, K$, $J^{[k]} = \mathcal{C}^{[k]} I(\sum_{l=1}^{k} Y^{[l]} \leq C)$ is the indicator for the kth event.

Our goal is to provide an adaptive estimation, inspired from Section 2.2, of the cumulative distribution function $\mathbb{P}\left[Y^{[2]} \leq y_2 | Y^{[1]} = y_1\right]$ in this context.

Confidence intervals (work in progress)

With O. Bouaziz and F. Comte, we would look into confidence bands for survival curves. A drawback of the methods described in all the previous works is that they are built to be compatible with $\mathbb{L}^2$-risk (or $\mathbb{L}^p$). In fact, model selection approaches make "global" estimation of functions (namely estimation of the coefficients of the unknown function on a given basis). In the contrary, kernel methods give "local" estimation and one can often dispose of ponctual or uniform CLT from which it is easy to derive confidence bounds. Recent and very general works by Giné and Nickl (2010) deal with the construction of confidence bands for the density. The key tool is to provide control of uniform risk (instead of $\mathbb{L}^p$ ones). In the same spirit as Picard & Tribouley (2000) or Tribouley (2004) for wavelet bases, we propose to define confidence bands for the estimator of the hazard rate described in [5], with piecewise polynomial bases.

Appendix A

Choix des bases et propriétés d'approximation

Les bases $(\varphi_\lambda)_{\lambda\in\Lambda_m}$ orthonormales utilisées pour définir les collections de modèles $(S_m)_{m\in\mathcal{M}_n}$ sont construites sur $A = [0,1]$. On peut facilement se ramener à n'importe quel intervalle $A = [a,b]$ par un changement d'échelle, puisque si $\{\varphi_\lambda\}_{\lambda\in\Lambda_m}$ est une base orthonormale de $\mathbb{L}^2([0,1])$ alors $\left\{\frac{1}{\sqrt{b-a}}\varphi_\lambda\left(\frac{\cdot-a}{b-a}\right)\right\}$ est une base orthonormale de $\mathbb{L}^2([a,b])$.

La notation Λ_m de l'ensemble d'indices, possiblement multiples, est une notation générique qui permet d'uniformiser l'écriture des bases. On explicite dans les exemples ci-dessous la signification de cette notation et on donne la dimension $D_m = |\Lambda_m|$ de chacun des espaces S_m engendré par la base $(\varphi_\lambda)_{\lambda\in\Lambda_m}$.

$[T]$ Base Trigonométrique : $\{\varphi_0, ..., \varphi_{m-1}\}$ avec $\varphi_0(x) = \mathbb{I}_{[0,1]}(x)$,

$$\varphi_{2j}(x) = \sqrt{2}\cos(2\pi jx)\mathbb{I}([0,1])(x) \ ; \ \varphi_{2j-1}(x) = \sqrt{2}\sin(2\pi jx)\mathbb{I}([0,1])(x)$$

pour $j \geq 1$. Cette base engendre un modèle S_m de dimension $D_m = m$.

$[DP]$ Base de Polynômes par morceaux dyadiques : l'ensemble d'indices $\Lambda_m = \{\lambda = (k,d), 1 \leq k \leq 2^m, 0 \leq d \leq r\}$ est un ensemble de couples indexant des polynômes de degré $0,\cdots,r$ (le degré maximal r est fixé), définis sur les intervalles $[(k-1)/2^m, k/2^m]$ avec $k = 1,\cdots,2^m$ d'une partition (dyadique) de $[0,1]$. L'espace S_m engendré est alors de dimension $D_m = (r+1)2^m$.

$[W]$ Base d'ondelettes sur $[0,1]$: $\{\Psi_{j,k}, j = -1,\cdots,m \text{ et } k \in \mathcal{K}(j)\}$, avec $\Psi_{j,k}(x) = 2^{j/2}\psi(2^jx-k)$, ψ est l'ondelette *mère*. L'ensemble $\mathcal{K}(j) = \{0,\cdots,K2^j-1\}$, pour $j \geq 0$, où K est une constante qui ne dépend que de ψ. Par exemple, pour la base de Haar, construite avec $\psi(x) = \mathbb{I}_{[0,1/2[}(x) - \mathbb{I}_{[1/2,1[}(x)$, on a $K = 1$. L'espace S_m engendré par la base d'ondelettes est de dimension $D_m = \sum_{j=1}^m |\mathcal{K}(j)| = |\mathcal{K}(-1)| + K(2^{m+1}-1)$.

Par construction, les éléments de cette base ont leur support inclus dans $[0,1]$.

$[H]$ Base d'histogramme : $\varphi_1, \cdots, \varphi_{2^m}$ avec $\varphi_j = 2^{m/2}\mathbb{1}([(j-1)/2^m, j/2^m[)$ pour $j = 1, \cdots, 2^m$. Dans ce cas, la dimension est $D_m = 2^m$. La base d'histogramme est un exemple particulier de bases [DP] et [W] qui sont des bases localisées.

Toutes ces bases engendrent, pour chaque $m \in \mathcal{M}_n$, un sous-espace vectoriel de $\mathbb{L}^2(A) \cap L^\infty(A)$ (ou $\mathbb{L}^2([0,1]) \cap L^\infty([0,1])$), qui possède les propriétés suivantes :

$(\mathcal{M}_1)$ $\forall m \in \mathcal{M}_n$, $\dim(S_m) = D_m \leq n$.

$(\mathcal{M}_2)$ Il existe une constante $\Phi_0 > 0$ telle que $\forall t \in S_m$, $\|t\|_\infty \leq \Phi_0\sqrt{D_m}\|t\|$.

C'est une propriété importante qui permet de relier la norme infinie à la norme dans $\mathbb{L}^2(A)$, introduite sous cette forme par Barron *et al.* (1999).

Enfin, pour les résultats d'adaptation, nous avons besoin que les modèles soient emboîtés :

$(\mathcal{M}_3)$ $\forall m, m' \in \mathcal{M}_n$, $D_m \leq D_{m'} \implies S_m \subset S_{m'}$.

Extension des modèles en dimension 2

Dans les chapitres 2 et 3, nous estimons des fonctions de deux variables sur $A = A_1 \times A_2$ un compact de $\mathbb{R}^2$. Une base orthonormale de $\mathbb{L}^2(A_1 \times A_2)$ peut être construite en prenant simplement le produit tensoriel de deux bases de $\mathbb{L}^2(A_1)$ et de $\mathbb{L}^2(A_2)$.

Pour chaque couple d'indices $m = (m_1, m_2)$, S_m est un espace de fonctions à support dans $A = A_1 \times A_2$ défini comme un espace produit de $S^{(1)}_{m_1} \otimes S^{(2)}_{m_2}$ avec $S^{(i)}_{m_i} \subset (L^2 \cap L^\infty)(\mathbb{R})$, pour $i = 1, 2$, engendrés par deux bases orthonormales $(\varphi_j^{m_1})_{j \in J_{m_1}}$ avec $|J_{m_1}| = D^{(1)}_{m_1}$ et $(\psi_k^{m_2})_{k \in K_{m_2}}$ avec $|K_{m_2}| = D^{(2)}_{m_2}$. Les indices j et k peuvent désigner des couples d'entiers comme nous l'avons décrit pour la base de polynômes par morceaux. Ainsi, on a :

$$S^{(1)}_{m_1} = \{t \;/\; t(x) = \sum_{j \in J_{m_1}} a_j^{m_1}\varphi_j^{m_1}(x)\},$$

$$S^{(2)}_{m_2} = \{t \;/\; t(y) = \sum_{k \in K_{m_2}} a_k^{m_2}\psi_k^{m_2}(y)\}$$

et

$$S_m = S^{(1)}_{m_1} \otimes S^{(2)}_{m_2} = \{T \;/\; T(x,y) = \sum_{j \in J_{m_1}} \sum_{k \in K_{m_2}} A^m_{j,k}\varphi_j^{m_1}(x)\psi_k^{m_2}(y),\; A^m_{j,k} \in \mathbb{R}\}.$$

On peut utiliser la même base dans les deux directions ou au contraire autoriser des bases différentes. La dimension de chaque modèle $S_m = S_{m_1} \otimes S_{m_2}$ est alors égale à $D_{m_1}D_{m_2}$.

Propriétés d'approximation

Les méthodes d'estimation par projection sur des bases orthonormales, consistent à estimer une fonction s (ou $s\mathbf{I}_A$ sa restriction sur un compact A) par approximation de sa projection orthogonale S_m sur un espace de dimension finie, le modèle S_m. Le résultat suivant nous fournit l'ordre de l'erreur d'approximation et se déduit de Hochmuth (2002) pour les bases localisées (polynômes ou ondelettes) et de Nikol'skii (1975) pour les bases trigonométriques.

Proposition 3 *Soit s une fonction de $\mathcal{B}^{\alpha}_{2,\infty}(A)$, avec $A \subset \mathbb{R}$, et $\alpha > 0$. Considérons un modèle S_m décrit ci-dessus de dimension D_m. Si s_m est la projection orthogonale de $s\mathbf{I}_A$ sur S_m, alors il existe une constante $C > 0$ telle que*

$$\|s\mathbf{I}_A - s_m\|_A = \left(\int_A (s - s_m)^2\right)^{1/2} \leq C D_m^{-\alpha}$$

où la constante C dépend uniquement de la base et de la norme de s dans l'espace de Besov.

En dimension 2, le résultat de la proposition précédente devient :

Proposition 4 *Soit s une fonction de $\mathcal{B}^{\alpha}_{2,\infty}(A)$, avec $A = A_1 \times A_2 \subset \mathbb{R}^2$ et $\alpha = (\alpha_1, \alpha_2)$. Considérons un modèle $S_m = S_{m_1} \otimes S_{m_2}$ décrit ci-dessus de dimension $D_{m_1} D_{m_2}$. Si s_m est la projection orthogonale de $s\mathbf{I}_A$ sur S_m, alors il existe une constante $C' > 0$ telle que*

$$\|s\mathbf{I}_A - s_m\|_A = \left(\int_A (s - s_m)^2\right)^{1/2} \leq C'[D_{m_1}^{-\alpha_1} + D_{m_2}^{-\alpha_2}]$$

où la constante C dépend uniquement de la base et de la norme de s dans l'espace de Besov.

Publications

The articles are referenced by a number in brackets. The publications [1]–[4] correspond to the works of my Phd thesis. Chapter 1 presents papers [5], [8], [10], [11] et [12], Chapter 2: [6], [7], [9] and [13], Chapter 3: [14] and [15].

[1] Brunel E., (1998) Applications d'Estimateurs de la Densité à la Simulation d'Episodes Pluvieux Extrêmes en Languedoc-Roussillon, Rev. Statistique Appliquée, XLVI(4), 45-58.

[2] Berlinet A., Brunel E., (2000) Choix optimal du nombre de classes pour l'estimateur de Barron. Comptes Rendus de l'Académie des Sciences de Paris, 331, no 9, 713-716.

[3] Berlinet A., Brunel E., (2002) Kullback-Leibler optimization of density estimates. Recent Advances in Statistical Methods (Montreal, QC, 2001), 31-41, Imp. Coll. Press, London.

[4] Berlinet A., Brunel E., (2004) Cross-validated density estimates based on Kullback-Leibler information, Journal of Nonparametric Statistics, 16, 3-4, 493-513.

[5] Brunel E., Comte F., (2005) Penalized contrast estimation of density and hazard rate with censored data, Sankhya, 67, Part 3, 441-475.

[6] Brunel E., Comte F., (2006) Adaptive nonparametric regression estimation in presence of right censoring, Math. Methods Statist. , 15, 3, 233–255.

[7] Brunel E., Comte F., Lacour C., (2007) Adaptive estimation of the conditional density in presence of censoring, Sankhya 69, 4, 734-763.

[8] Brunel E., Comte F., (2008) Adaptive estimation of hazard rate with censored data, Comm. Statist. Theory Methods 37, no. 8, 1284-1305.

[9] Brunel E., Comte F., (2008) Model selection for additive regression models in the presence of censoring, Chapitre 1, Mathematical Methods in Survival Analysis, Reliability and Quality of Life. Edité par C. Huber, N. Limnios, M. Mesbah and M. Nikulin, ISTE and Wiley, p. 17-31.

[10] Brunel E., Comte F., Guilloux A., (2008) Estimation strategies for censored lifetimes with a Lexis-diagram type model, Scand. J. Statist. 35, 3, 557-576.

[11] Brunel E., Comte F., Guilloux A., (2009) Nonparametric density estimation in presence of bias and censoring, Test , 18, no. 1, 166-194.

[12] Brunel E., Comte F., (2009) Cumulative distribution function estimation under interval censoring case 1, Electron. J. Stat. 3, 1-24.

[13] Brunel E., Comte F., Lacour C., (2010) Minimax estimation of the conditional cumulative distribution function, Sankhya. 72 , no. 2, 293-330.

[14] Brunel E., Comte F., (2011) Conditional mean residual life estimation. J. Nonparam. Statist., 23, no. 2, 471-495.

[15] Brunel, E., Comte F., Guilloux A., (2012) Nonparametric estimation for survival data with censoring indicators missing at random. To appear in Journal of Statistical Planning and Inference. HAL http://hal.archives-ouvertes.fr/hal-00679799.

Submitted papers:

[16] Brunel, E., Roche, A., (2012) Penalized contrast estimation in functional linear models with circular data. *submitted.* HAL http://hal.archives-ouvertes.fr/hal-00651399

[17] Brunel, E., Mas, A., Roche, A. (2013) Non-asymptotic Adaptive Prediction in Functional Linear Models. *submitted.* HAL http://hal.archives-ouvertes.fr/hal-00763924

[18] Brunel, E., Comte, F., Guilloux, A., (2013) Estimation/Imputation strategies for missing data in survival analysis. *submitted.* HAL http://hal.archives-ouvertes.fr/hal-00818859

Bibliography

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.*, **6**, 701-726.

Abdous, B. and Berred, A. (2005). Mean residual life estimation. *J. of Statist. Planning and Inference* **132**, 3-19.

Allignol, A., Latouche, A., Yan, J., Fine, J.P. (2011). A regression model for the conditional probability of a competing event: application to monoclonal gammopathy of unknown significance, *Applied Statistics* **60**, 135-142.

Andersen, P.K. , Borgan, O., Gill, R. D., Keiding, N. (1993). Statistical models based on counting processes. Springer Series in Statistics. Springer-Verlag, New York.

Akakpo, N., Lacour, C. (2011). Inhomogeneous and anisotropic conditional density estimation from dependent data. *Electron. J. Stat.* **5**, 1618–1653.

Antoniadis, A., Grégoire, G. (1990). Penalized likelihood estimation for rates with censored survival data. *Scand. J. Statist.* **17**, 43-63.

Antoniadis, A., Grégoire, G., Nason, G. (1999). Density and hazard rate estimation for right-censored data by using wavelet methods. *J. R. Stat. Soc.*, Ser. B Stat. Methodol. **61**, 63-84.

Asgharian, M., M'Lan, C.E., Wolfson, D.B. (2002). Length-biased sampling with right censoring: an unconditional approach. *J. Amer. Statist. Assoc.* **97**, 201-209.

Asgharian, M. (2003). Biased sampling with right censoring: a note on Sun, Cui & Tiwari (2002). *Canad. J. Statist.* **30**, 475-490.

Azzalini, A., Bowman, A.W. (1990). A look at some data on the Old Faithfull geyser. *Applied Statistics* **39**, 357-365.

Baraud, Y. (2002) Model selection for regression on a random design. *ESAIM P&S* **6**, 127-146.

Baraud, Y., Comte, F., Viennet, G. (2001). Adaptive estimation in autoregression or β-mixing regression via model selection. *Ann. Statist* **29**, 839-875.

Barron, A.R., Cover, T. M. (1991). Minimum complexity density estimation. *IEEE Trans. Inform. Theory* **37**, 1034-1054.

Barron, A.R., Birgé, L., Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields* **113**, 301-413.

Bhattacharya, P.K., Gangopadhyay, A.K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Stat.* **18**, 1400-1415.

Birgé, L. (1999). Interval censoring: a nonasymptotic point of view. *Math. Methods Statist.* **8**, 285–298.

Birgé, L., Massart, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Relat. Fields* **97**, 113-150.

Birgé, L., Massart, P. (1997). From model selection to adaptive estimation, in *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics* (D. Pollard, E. Torgersen and G. Yang, eds), 55-87, Springer-Verlag, New-York.

Birgé, L., Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4**, 329-375.

Bitouzé, D., Laurent, B., Massart, P. (1999). A Dvoretsky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator. *Ann. Inst. Henri Poincaré* **32**, 735-763.

Cai, T., Hall, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34**, 2159–2179.

Cai, T., Yuan, M. (2012). Minimax and Adaptive Prediction for Functional Linear Regression, *J. Amer. Statist. Assoc.* **107**, 1201-1216.

Cardot, H. , Ferraty, F., Sarda, P. (2003). Spline estimator for the functional linear model, *Statistica Sinica*, 571-591.

Cardot, H., Mas, A., Sarda, P. (2007). CLT in functional linear regression models. *Probab. Theory Relat. Fields*, **138**, 325–361.

Cardot, H., Johannes, J. (2010). Thresholding projection estimators in functional linear models. *J. Multivariate Anal.* **101**, 395–408.

Comte, F., Johannes, J. (2010). Adaptive estimation in circular functional linear models. *Math. Methods Statist.* **19** 42–63.

Comte, F., Johannes, J. (2012). Adaptive functional linear regression, *Ann. Statist.* **40**, 2765-2797.

Cook, R.J., Lawless, J.F., (2008) The statistical analysis of recurrent events, Springer, New-York.

Chagny, G. (2011). Warped bases for conditional density estimation. Preprint. hal-00641560, version 2.

Chen, Dong., Hall, P., Müller, H-G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* **39**, 1720-1747.

Chen, Y. Q., Cheng, S. (2005). Semiparametric regression analysis of mean residual life with censored survival data, *Biometrika* **92**, 19-29.

Chernozhukov, V., Fernández-Val, I., Galichon, A. (2009). Improving point and interval estimates of monotone functions by rearrangement. *Biometrika* **96**, 559-575.

Comte, F., Gaiffas, S., Guilloux, A. (2011). Adaptive estimation of the conditional intensity of marker-dependent counting processes, *Ann. Inst. H. Poincaré Probab. Statist.* **47**, 4, 1171-1196.

Cohen, A., Daubechies, I., Vial, P.B. (1993). Wavelets on the interval and fast wavelet transforms.*Appl. Comput. Harmon. Anal.* **1**, 54-81.

Cohen, S., Lepennec, E. (2011). Conditional density estimation by penalized likelihood model selection and applications. Preprint. arXiv:1103.2021v5.

Crambes, C., Kneip, A., Sarda, P. (2009). Smoothing Splines Estimators for Functional Linear Regression, *Ann. Statist.*, **1**, 35–72.

Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B*, **34**, 187-220.

Dabrowska, D.M. (1987). Nonparametric regression with censored survival time data. *Scand. J. Statist.* **14**, 181-197.

Dabrowska, D.M., Duffy, D.L., Zhang, Z.D. (1999). Hazard and density estimation from bivariate censored data. *J. Nonparametr. Statist.* **10**, 67-93.

de Boor, C. (1978). *A practical guide to splines.* Applied Mathematical Science, Springer, New York.

De Gooijer, J. G., Zerom, D. (2003). On conditional density estimation. *Statist. Neerlandica*, **57**, 159-176.

Cavalier, L. & Hengartner, N.W. (2005). Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, **21**, 1345-1361.

Chakraborty, R., and Rao, C.R., (2000). Selection biases of samples and their resolutions. In Handbook of Statistics, Vol. 18 (P.K. Sen and C.R. Rao, eds), 675-712. Elsevier Science.

Chaubey, Y.P. and Sen, P.K. (1999). On smooth estimation of mean residual life. *J. Statist. Plann. Inference* **75**, 223Ű236.

Chen, Y. Q. and Cheng (2005), S. Semiparametric regression analysis of mean residual life with censored survival data. *Biometrika* **92**, 19-29.

Chen, Y.Q., Jewell, N. P., Lei, X., Cheng, S.C. (2005). Semiparametric estimation of proportional mean residual life model in presence of censoring. *Biometrics* **61**, 170-178.

Cox, D.R. (1969). In *New development in survey sampling.* Eds Johnson and Smith, Wiley.

Csorgo, M. and Zikitis, R. (1996). Mean Residual Life Processes. *Ann. Statist.* **24**, 1717-1739.

Delecroix, M., Yazourh, Y. (1991). Nonparametric inference for failure rates with random right censorship: the method of orthogonal functions [French]. *Statist. Anal. Données* **16**, 39-62.

DeVore, R.A., Lorentz, G.G. (1993). *Constructive approximation.* Springer-Verlag.

Diamond, I.D., McDonald, J.W. (1991). The analysis of current status data. *Demographic applications of event history analysis.*(eds J. Trussell, R. Hankinson & J. Tilton). Oxford University Press, Oxford.

Dikta, M.K., (1998). On semiparametric random censorship models. *J. Stat. Plann. Inference* **66**, 253-279.

Donoho, D.L., Johnstone, I.M. (1998). Minimax estimation with wavelet shrinkage. *Ann. Statist.*, **26**, 879-921.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., Picard, D. (1995). Wavelet shrinkage: asymptopia? *J. Roy. Statist. Soc. Ser. B* **57**, 301-369.

Efromovich, S. (2001). Density estimation under random censorship and order restrictions: from asymptotic to small sample sizes. *J. Amer. Statist. Assoc.* **96**, 667-685.

Efromovich, S. (2004). Density estimation for biased data. *Ann. Statist.* **32**, 1137-1161.

Efromovich, S. (2004). Distribution estimation for biased data. *J. Stat. Plann. Inference* **124**, 1-43.

Efromovich, S. (2007). Conditional density estimation in a regression setting. *Ann. Statist.***35**, 2504-2535.

Efromovich, S. (2008). Oracle inequality for conditional density estimation and an actuarial example. *Ann. Inst. Statist. Math.*, 1-27.

Efromovich, S., Koltchinskii, V. (2001). On inverse problems with unknown operators. *IEEE Transactions on Information Theory*, **47**, 2876–2894.

Engl, H. W., Hanke, M., Neubauer, A. (1996). Regularization of inverse problems, Mathematics and its Applications, 375 Kluwer Academic Publishers Group.

Faugeras, O.P. (2009). A quantile-copula approach to conditional density estimation. *J. Multivariate Anal.* **100**, 2083-2099.

Fan, J., Gijbels, I. (1994). Censored regression: local linear approximations and their applications. *J. Amer. Statist. Assoc.* **89**, 560-570.

Fan, J., Hu, T.-C., Truong, Y. (1994) Robust nonparametric function estimation. *Scand. J. Statist.*, **21**, 433-446.

Ferraty, F., Vieu, P. (2006). Nonparametric functional data analysis, Springer Series in Statistics, Theory and practice, New York, Springer.

Fisher, R.A. (1934). The effect of methods of ascertainment upon the estimation of frequencies. *Ann. Eugen.* **6**, 13-25.

Gijbels, I., Lin, D., Ying, Z., (2007). Non- and semi-parametric analysis of failure time data with missing failure indicators. Complex datasets and inverse problems, 203-223, IMS Lecture Notes Monogr. Ser., 54, Inst. Math. Statist., Beachwood, OH.

Gill, R.D., Vardi, Y., Wellner, J. A. (1988). Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.* **16**, 1069–1112.

Giné, E. and Nickl, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38**, 1122-1170.

Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39**, 1608-1632.

Groeneboom, P., Wellner, J.A. (1992). Information bounds and nonparametric maximum likelihood estimation. Boston, Birkhäuser Verlag.

Guilloux, A. (2007). Non-parametric estimation for censored lifetimes suffering from an unknown selection bias. *Math. Methods Statist.* **16**, 202-216.

Györfi, L., Kohler, M., Krzyzak, A., Walk, H. (2002). A distribution-free theory of nonparametric regression. Springer Series in Statistics, Springer 2002.

Györfi, L., Kohler, M. (2007). Nonparametric estimation of conditional distributions. *IEEE Trans. Inform. Theory* **53**, 1872-1879.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35**, 70–91.

Hall, W. J. and Wellner, J. (1981) Mean residual life. Statistics and related topics (Ottawa, Ont., 1980), 169–184, North-Holland, Amsterdam-New York.

He, X., Shi, P. (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *J. Nonparam. Statist.*, **3**, 299-308.

Heckman, J. (1985). Selection bias and self-selection. In The new Palgrave : A Dictionary of Economics, 287-296. MacMillan Press, Stockton, N.Y.

Heuchenne, C., Van Keilegom, I. (2007). Nonlinear regression with censored data. *Technometrics* **49**, 34-44.

Hochmuth, R. (2002) Wavelet characterizations for anisotropic Besov spaces. *Appl. Comput. Harmon. Anal.* **2**, 179-208.

Hoffmann, M., Reiss, M. (2008). Nonlinear estimation for linear inverse problems with error in the operator. *Ann. Statist.*, **36**, 310–336.

Huber, C., MacGibbon, B. (2004). Lower bounds for estimating a hazard. *Advances in survival analysis,* 209–226, Handbook of Statist., 23 Elsevier, Amsterdam.

Hudgens, M.G., Maathuis, M.H., Gilbert, P.B. (2007). Nonparametric estimation of the joint distribution of a survival time subject to interval censoring and a continuous mark variable. *Biometrics* **63**, 372–380.

Jewell, N. P., van der Laan, M. (2004). *Current status data: review, recent developments and open problems.* Advances in survival analysis, 625–642, Handbook of Statist., 23, Elsevier, Amsterdam.

Kaplan, E.L., Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

Keiding, N. (1990). Statistical inference in the Lexis diagram. *Philos. Trans. Roy. Soc. London Ser. A* **332**(1627):487–509.

Koenker, R. (2005). *Quantile regression.* Econometric Society Monographs, Cambridge.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33-50.

Koenker, R., Ng, P., Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, **81**, 673-680.

Koul, H., Susarla, V., Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9**, 1276–1288.

Lacour, C. (2007). Adaptive estimation of the transition density of a Markov chain. *Ann. Inst. Henri Poincaré Probab. Statist.* **43**, 571-597.

Lejeune, M., Sarda, P. (1988). Quantile regression: a nonparametric approach. *Comput. Statist. Data Anal.* , **6**, 229-239.

Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74**, 301–309.

Lexis, W. (1875). Einleitung in die Theorie der Bevölkerung-Statistik. In Mathematical demography (édition D. Smith and N. Keyfitz), Biomathematics 6, 39-41, 1977. Springer-Verlag, Berlin.

Li, Y., Hsing, T. (2007). On rates of convergence in functional linear regression, *J. Multivariate Anal.* **98**, 1782–1804.

Little, R.J.A Rubin, D.B. (2002). Statistical analysis with missing data. Second edition. Wiley Series in Probability and Statistics.

Lo, S.-H., (1991). Estimating a survival function with incomplete cause-of-death data. *J. Multivariate Anal.* **39**, 217-235.

Lo, S.H., Mack, Y.P., Wang, J.L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Probab. Theory Relat. Fields* **80**, 461-473.

Lopez, O., Patilea, V. , Van Keilegom, I.(2013). Single index regression models in the presence of censoring depending on the covariates. à paraître *Bernoulli.*

Lund, J. (2000). Sampling bias in population studies - how to use the Lexis diagram. *Scand. J. Statist.* **27**, 589-604.

Ma, S., Kosorok, M.R. (2006). Adaptive penalized M-estimation with current status data. *Ann. Inst. Statist. Math.* **58** , 511–526.

Maguluri, G. and Zhang, C.-H. (1994). Estimation in the mean residual life regression model. *J. Roy. Statist. Soc. Ser. B* **56**, 477-489.

Marron, J.S. and de Uña-Àlvarez, J. (2004). SiZer for length biased, censored density and hazard estimation. *J. Statist. Plann. Inference* **121**, 149-161.

Marron, J., Padgett, W.J. (1987). Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples. *Ann. Statist.* **15**, 1520-1535.

Massart, P. (2007). Concentration inequalities and model selection. Ecole d'été de Probabilités de Saint-Flour 2003. Lecture Notes in Mathematics 1896, Springer Berlin Heidelberg.

Mielniczuk, J. (1985). Properties of some kernel estimators and of the adapted Loftsgarden-Quesenberry estimator of a density function for censored data. *Period. Math. Hungar.* **16**, 69-81.

Nikol'skii, S.M. (1975). Approximations of functions of several variables and imbedding theorems, Springer-Verlag, New York, translated from the Russian by John M. Danskin, Jr., Die Grundlehren der Mathematischen Wissenschaften, Band 205.

Oakes, D., Dasu, T. (1996). A note on residual life. *Biometrika* **77**, 409-410.

Patil, P. (1997). Nonparametric hazard rate estimation by orthogonal wavelet method. *J. Statist. Plann. Inference* **60**, 153-168.

Park, J. (2004).Optimal global rate of convergence in nonparametric regression with left-truncated and right-censored date. *J. Multivariate Anal.* **89**, 70–86.

Patil, G.P. and Rao, C.R., (1977). The weighted distributions : A survey of their applications. In *Applications of Statistics*, (P.R. Krishnaiah, ed.), 383-405. North-Holland Publishing Co., Amsterdam.

Picard, D., Tribouley, K.(2000) Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28**, 298-335.

Plancade, S. (2011). Estimation of conditional cumulative distribution function from current status data. Preprint. arXiv:1110.5927v1.

Poiraud-Casanova, S., Thomas-Agnan, C. (1998). Quantiles conditionnels. *J. Soc. Française Stat.*, **139**, 31-44.

Prentice, R.I. (1973). Exponential survivals with censoring and explanatory variables, *Biometrika* **60**, 279-288.

Preda, C., Saporta, G. (2005). PLS regression on a stochastic process, *Comput. Statist. Data Anal.*, **48**, 149–158.

Ramsay, J.O., Silverman, B.W., (2005). Functional Data Analysis, Springer Series in Statistics, 2nd edition Springer.

Ramsay, J.O., Dalzell, C.J., (1991). Some Tools for Functional Data Analysis, *J. R. Statist. Soc.* **53**, 539–572.

Reid, N. (1981) Influence functions for censored data. *Ann. Statist.* **9**, 78–92.

Reynaud-Bouret, P. (2006). Penalized projection estimators of the Aalen multiplicative intensity. *Bernoulli*, **12**, 633-661.

Ruppert, D., Caroll, J. (1988). Transformation and weighting in regression. Chapman and Hall, New York.

Schaubel, D.E., Cai, J. (2004). Regression Methods for Gap Time Hazard Functions of Sequentially Ordered Multivariate Failure Time Data , *Biometrika* **91**, 291–303.

Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.* **23**, 422-439.

Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.*, **126**, 505-563.

Subramanian, S., (2004). Asymptotically efficient estimation of a survival function in the missing censoring indicator model. *J. Nonparam. Statist.* **16**, 797-817.

Subramanian, S., (2006). Survival analysis for the missing censoring indicator model using kernel density estimation techniques. *Stat. Methodol.* **3**, 125-136.

Subramanian, S., (2009). The multiple imputations based Kaplan-Meier estimator. *Statist. Probab. Lett.* **79**, 1906-1914.

Subramanian, S., (2011). Multiple imputations and the missing censoring indicator model. *J. Multivariate Anal.* **102**, 105-117.

Triboulay, K. (2004). Adaptive simultaneous confidence interval in nonparametric estimation. *Statist. Probab. Letters* **69** , 37-51.

de Uña-Àlvarez, J. (2002). Product-limit estimation for length-biased censored data. *Test* **11**, 109-125.

Vardi, Y. (1982). Nonparametric estimation in renewal processes. *Ann. Statist.* **10**, 772–785.

van de Geer, S. (1993). Hellinger-consistency of certain nonparametric likelihood estimators. *Ann. Statist.* **21**, 14-44.

van der Laan, M. J., McKeague, I. W., 1998. Efficient estimation from right-censored data when failure indicators are missing at random. *Ann. Statist.* **26**, 164-182.

van der Vaart, A., van der Laan, M. J. (2006). Estimating a survival distribution with current status data and high-dimensional covariates. *Int. J. Biostat.* **2**, Art 9, 42pp.

Van Keilegom, I., de Una-Álvarez, J., Meira-Machado, L. (2011). Nonparametric location-scale models for successive survival times under dependent censoring, *J. Statist. Plann. Infer.* **141**, 1118-1131.

Verzelen, N. (2010). High-dimensional Gaussian model selection on a Gaussian design. *Ann. Inst. Henri Poincaré Probab. Stat.*, **46**, 480–524.

Wang, Q., Shen, J., (2008). Estimation and confidence bands of a conditional survival function with censoring indicators missing at random. *J. Multivariate Anal.* **99**, 928-948.

Wang, Q., Ng, K. W., (2008). Asymptotically efficient product-limit estimators with censoring indicators missing at random. *Statist. Sinica* **18**, 749-768.

Wang, Q. , Liu, W., Liu, C., (2009). Probability density estimation for survival data with censoring indicators missing at random. *J. Multivariate Anal.* **100**, 835-850.

Yang, S. (2000). Functional estimation under interval censoring case 1. *J. Statist. Plann. Inference* **89**, 135-144.

Zheng, Z. K. (1987). A class of estimators of the parameters in linear regression with censored data, *Acta Math. Appl. Sinica* **3**, 231-341.

Zheng, Z. K. (1988). Strong consistency of nonparametric regression estimates with censored data, *J. Math. Res. Exposition* **8**, 307-313.