A central limit theorem for nonparametric regression for competing risks model with right censoring

L. Bordes, joint work with K. Gneyou

Laboratoire de Mathématiques Appliquées de Pau - CNRS UMR 5142
Université de Pau et des Pays de l'Adour

Cinquième Rencontre de Statistiques Mathématiques
BORDEAUX-SANTANDER-TOULOUSE-VALLADOLID
June 3-5, 2009

Outline

Outline

Competing risks model

Survival analysis example: event=death but several causes are possible

Flehinger et al. (Biometrika, 1998): Lung cancer data with 2 causes of death

Competing risks model

Survival analysis example: event=death but several causes are possible

Flehinger et al. (Biometrika, 1998): Lung cancer data with 2 causes of death

- Cause 1: death from cancer

Competing risks model

Survival analysis example: event=death but several causes are possible

Flehinger et al. (Biometrika, 1998): Lung cancer data with 2 causes of death

- Cause 1: death from cancer
- Cause 2: death from other causes

Competing risks model

Survival analysis example: event=death but several causes are possible

Flehinger et al. (Biometrika, 1998): Lung cancer data with 2 causes of death

- Cause 1: death from cancer
- Cause 2: death from other causes

Observations: lifetime + cause + covariates

Competing risks model

Reliability example: event=failure but several causes are possible

Craiu and Duchesne (Biometrika, 2004): hard drive data with 3 failure causes

Competing risks model

Reliability example: event=failure but several causes are possible
Craiu and Duchesne (Biometrika, 2004): hard drive data with 3 failure causes

- Cause 1: electronic hard

Competing risks model

Reliability example: event=failure but several causes are possible

Craiu and Duchesne (Biometrika, 2004): hard drive data with 3 failure causes

- Cause 1: electronic hard
- Cause 2: head flyability

Competing risks model

Reliability example: event=failure but several causes are possible

Craiu and Duchesne (Biometrika, 2004): hard drive data with 3 failure causes

- Cause 1: electronic hard
- Cause 2: head flyability
- Cause 3: head / disc magnetic

Competing risks model

Reliability example: event=failure but several causes are possible
Craiu and Duchesne (Biometrika, 2004): hard drive data with 3 failure causes

- Cause 1: electronic hard
- Cause 2: head flyability
- Cause 3: head / disc magnetic

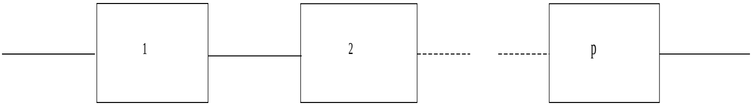

Figure: Serie system

Notations

Latent variable model:

- Lifetimes $T_1, \ldots, T_m$ rv in $\mathbb{R}^+$

Notations

Latent variable model:

- Lifetimes $T_1, \ldots, T_m$ rv in $\mathbb{R}^+$
- Set of causes $J = \{1, \ldots, m\}$

Notations

Latent variable model:

- Lifetimes $T_1, \ldots, T_m$ rv in $\mathbb{R}^+$
- Set of causes $J = \{1, \ldots, m\}$
- Censoring time C in $\mathbb{R}^+$

Notations

Latent variable model:

- Lifetimes $T_1, \ldots, T_m$ rv in $\mathbb{R}^+$
- Set of causes $J = \{1, \ldots, m\}$
- Censoring time C in $\mathbb{R}^+$
- Covariates Z in $\mathbb{R}^d$

Notations

Latent variable model:

- Lifetimes $T_1, \ldots, T_m$ rv in $\mathbb{R}^+$
- Set of causes $J = \{1, \ldots, m\}$
- Censoring time C in $\mathbb{R}^+$
- Covariates Z in $\mathbb{R}^d$

Observations without censoring:

- Duration $X = \min(T_1, \ldots, T_m)$

Notations

Latent variable model:

- Lifetimes $T_1, \ldots, T_m$ rv in $\mathbb{R}^+$
- Set of causes $J = \{1, \ldots, m\}$
- Censoring time C in $\mathbb{R}^+$
- Covariates Z in $\mathbb{R}^d$

Observations without censoring:

- Duration $X = \min(T_1, \ldots, T_m)$
- Cause $\eta = j$ if $X = T_j$

Notations

Latent variable model:

- Lifetimes $T_1, \ldots, T_m$ rv in $\mathbb{R}^+$
- Set of causes $J = \{1, \ldots, m\}$
- Censoring time C in $\mathbb{R}^+$
- Covariates Z in $\mathbb{R}^d$

Observations without censoring:

- Duration $X = \min(T_1, \ldots, T_m)$
- Cause $\eta = j$ if $X = T_j$
- Covariates Z

Notations

Latent variable model:

- Lifetimes $T_1, \ldots, T_m$ rv in $\mathbb{R}^+$
- Set of causes $J = \{1, \ldots, m\}$
- Censoring time C in $\mathbb{R}^+$
- Covariates Z in $\mathbb{R}^d$

Observations without censoring:

- Duration $X = \min(T_1, \ldots, T_m)$
- Cause $\eta = j$ if $X = T_j$
- Covariates Z

Observations with right censoring:

- Duration $Y = \min(X, C)$

Notations

Latent variable model:

- Lifetimes $T_1, \dots, T_m$ rv in $\mathbb{R}^+$
- Set of causes $J = \{1, \dots, m\}$
- Censoring time C in $\mathbb{R}^+$
- Covariates Z in $\mathbb{R}^d$

Observations without censoring:

- Duration $X = \min(T_1, \dots, T_m)$
- Cause $\eta = j$ if $X = T_j$
- Covariates Z

Observations with right censoring:

- Duration $Y = \min(X, C)$
- Censoring indicator $\delta = I(X \leq C)$

Notations

Latent variable model:

- Lifetimes $T_1, \ldots, T_m$ rv in $\mathbb{R}^+$
- Set of causes $J = \{1, \ldots, m\}$
- Censoring time C in $\mathbb{R}^+$
- Covariates Z in $\mathbb{R}^d$

Observations without censoring:

- Duration $X = \min(T_1, \ldots, T_m)$
- Cause $\eta = j$ if $X = T_j$
- Covariates Z

Observations with right censoring:

- Duration $Y = \min(X, C)$
- Censoring indicator $\delta = I(X \leq C)$
- Cause $\eta = j$ if $X = T_j$ and $\delta = 1 \Rightarrow$ set $\xi = \eta\delta$

Notations

Latent variable model:

- Lifetimes $T_1, \ldots, T_m$ rv in $\mathbb{R}^+$
- Set of causes $J = \{1, \ldots, m\}$
- Censoring time C in $\mathbb{R}^+$
- Covariates Z in $\mathbb{R}^d$

Observations without censoring:

- Duration $X = \min(T_1, \ldots, T_m)$
- Cause $\eta = j$ if $X = T_j$
- Covariates Z

Observations with right censoring:

- Duration $Y = \min(X, C)$
- Censoring indicator $\delta = I(X \leq C)$
- Cause $\eta = j$ if $X = T_j$ and $\delta = 1 \Rightarrow$ set $\xi = \eta\delta$
- Covariates Z

Aim and quantity of interest

Difficulty: generally $T_1, \ldots, T_m$ are not independent.
Tiatsis (1975): there exist independent rv $T_1^*, \ldots, T_m^*$, such that if $X^* = \min(T_1^*, \ldots, T_m^*)$ and $\eta^* = j$ if $X^* = T_j^*$ we have

$$(X, \eta) \stackrel{d}{=} (X^*, \eta^*).$$

Aim and quantity of interest

Difficulty: generally $T_1, \ldots, T_m$ are not independent.
Tiatsis (1975): there exist independent rv $T_1^*, \ldots, T_m^*$, such that if $X^* = \min(T_1^*, \ldots, T_m^*)$ and $\eta^* = j$ if $X^* = T_j^*$ we have

$$(X, \eta) \stackrel{d}{=} (X^*, \eta^*).$$

Consequence: **generally** we can not identify the joint or marginal df of $(T_1, \ldots, T_m)$ from (X, η)!

Aim and quantity of interest

Difficulty: generally $T_1, \ldots, T_m$ are not independent.
Tiatsis (1975): there exist independent rv $T_1^*, \ldots, T_m^*$, such that if $X^* = \min(T_1^*, \ldots, T_m^*)$ and $\eta^* = j$ if $X^* = T_j^*$ we have

$$(X, \eta) \stackrel{d}{=} (X^*, \eta^*).$$

Consequence: **generally** we can not identify the joint or marginal df of $(T_1, \ldots, T_m)$ from (X, η)!
Question: is it still true when in addition to (X, η) we observe Z?

Aim and quantity of interest

Difficulty: generally $T_1, \ldots, T_m$ are not independent.
Tiatsis (1975): there exist independent rv $T_1^*, \ldots, T_m^*$, such that if $X^* = \min(T_1^*, \ldots, T_m^*)$ and $\eta^* = j$ if $X^* = T_j^*$ we have

$$(X, \eta) \stackrel{d}{=} (X^*, \eta^*).$$

Consequence: **generally** we can not identify the joint or marginal df of $(T_1, \ldots, T_m)$ from (X, η)!
Question: is it still true when in addition to (X, η) we observe Z?
Heckman and Honoré (Biometrika, 1989): some general nonparametric models can be identified

Aim and quantity of interest

Difficulty: generally $T_1, \ldots, T_m$ are not independent.
Tiatsis (1975): there exist independent rv $T_1^*, \ldots, T_m^*$, such that if $X^* = \min(T_1^*, \ldots, T_m^*)$ and $\eta^* = j$ if $X^* = T_j^*$ we have

$$(X, \eta) \stackrel{d}{=} (X^*, \eta^*).$$

Consequence: **generally** we can not identify the joint or marginal df of $(T_1, \ldots, T_m)$ from (X, η)!
Question: is it still true when in addition to (X, η) we observe Z?
Heckman and Honoré (Biometrika, 1989): some general nonparametric models can be identified
Fermanian (JMVA, 2003): extends the Heckman and Honoré results but many functional parameters to estimate!

Standard functions

Cumulative incidence functions:

$$F_j(t|z) = \mathbb{P}(X \leq t, \eta = j|Z = z).$$

Standard functions

Cumulative incidence functions:

$$F_j(t|z) = \mathbb{P}(X \leq t, \eta = j | Z = z).$$

Cause specific hazard rate:

$$\lambda_j(t|z) = \lim_{s \searrow 0} \frac{1}{s} \mathbb{P}(X \in [t, t+s) | X \geq t, \eta = j) = \frac{f_j(t|z)}{\bar{F}_X(t)},$$

where $\bar{F}_X(t|z)$ is the survival function of X given $Z = z$, and is f_j the subdensity function corresponding to F_j.

Standard functions

Cumulative incidence functions:

$$F_j(t|z) = \mathbb{P}(X \le t, \eta = j | Z = z).$$

Cause specific hazard rate:

$$\lambda_j(t|z) = \lim_{s \searrow 0} \frac{1}{s} \mathbb{P}(X \in [t, t+s) | X \ge t, \eta = j) = \frac{f_j(t|z)}{\bar{F}_X(t)},$$

where $\bar{F}_X(t|z)$ is the survival function of X given $Z = z$, and is f_j the subdensity function corresponding to F_j.

Basic relations:

$$F_j(t|z) = \int_0^t f_j(s|z) ds = \int_0^t \lambda_j(s|z) \bar{F}_X(s|z) ds = \int_0^t \bar{F}_X(s|z) d\Lambda_j(s|z),$$

where $\Lambda_j(t|z) = \int_0^t \lambda_j(s|z) ds$ is the jth cumulative cause specific hazard function.

Regression

We want to estimate:

$$r_j(z) = \mathbb{E}\left[\psi(X) I(\eta = j) | Z = z\right]$$

because the quantity of interest

$$\mathbb{E}\left[\psi(T_j) | Z = z\right]$$

is generally not reachable from the distribution of (X, η).

Regression

We want to estimate:

$$r_j(z) = \mathbb{E}\left[\psi(X) I(\eta = j) | Z = z\right]$$

because the quantity of interest

$$\mathbb{E}\left[\psi(T_j) | Z = z\right]$$

is generally not reachable from the distribution of (X, η).
Examples: set $\tilde{T}_j = T_j I(\eta = j)$

- If $\psi = id$ $r_j(z) = \mathbb{E}[\tilde{T}_j | Z = z]$

Regression

We want to estimate:

$$r_j(z) = \mathbb{E}\left[\psi(X) I(\eta = j) | Z = z\right]$$

because the quantity of interest

$$\mathbb{E}\left[\psi(T_j) | Z = z\right]$$

is generally not reachable from the distribution of (X, η).
Examples: set $\tilde{T}_j = T_j I(\eta = j)$

- If $\psi = id$ $r_j(z) = \mathbb{E}[\tilde{T}_j | Z = z]$
- If $\psi(x) = x^p$ $r_j(z) = \mathbb{E}[\tilde{T}_j^p | Z = z]$

Regression

We want to estimate:

$$r_j(z) = \mathbb{E}\left[\psi(X)I(\eta = j)|Z = z\right]$$

because the quantity of interest

$$\mathbb{E}\left[\psi(T_j)|Z = z\right]$$

is generally not reachable from the distribution of (X, η).
Examples: set $\tilde{T}_j = T_j I(\eta = j)$

- If $\psi = id$ $r_j(z) = \mathbb{E}[\tilde{T}_j|Z = z]$
- If $\psi(x) = x^p$ $r_j(z) = \mathbb{E}[\tilde{T}_j^p|Z = z]$
- If $\psi_t(x) = I(x \le t)$ $r_j(z, t) = F_j(t|z)$

Assumption: C is independent of everything

$$\bar{H}(t|z) = \mathbb{P}(Y > t|Z = z) = \bar{G}(t)\bar{F}_X(t|z)$$

where $\bar{G}$ and $\bar{F}_X(\cdot|z)$ are the survival functions of C and X. We write

$$H_j(t|z) = \mathbb{P}(Y \le t, \xi = j|Z = z)$$

then for $1 \le j \le m$

$$\Lambda_j(t|z) = \int_0^t \lambda_j(s|z)ds = \int_0^t \frac{dF_j(s|z)}{\bar{F}_X(s^-|z)} = \int_0^t \frac{dH_j(s|z)}{\bar{H}(s^-|z)}$$

and

$$\begin{aligned} r_j(z) &= \int_0^{\tau_z} \psi(t) f_j(t|z)dt = \int_0^{\tau_z} \psi(t)\bar{F}_X(t|z)d\Lambda_j(t|z) \\ &= \int_0^{\tau_z} \frac{\psi(t)\bar{F}_X(t|z)}{\bar{H}(t|z)} dH_j(t|z) = \int_0^{\tau_z} \frac{\psi(t)}{\bar{G}(t)} dH_j(t|z) \end{aligned}$$

Outline

Estimating $\bar{G}$

Kaplan-Meier estimator:

$$\bar{G}_n(t) = \prod_{i \in V(t)} \left(1 - \frac{I(\xi_i = 0)}{R(Y_i)}\right)$$

where

$$V(t) = \{i; 1 \leq i \leq n, Y_i \leq t\}$$

and

$$R(t) = \#\{i; 1 \leq i \leq n, Y_i \geq t\}.$$

Estimating $H_j(\cdot|z)$

Nadaraya-Watson estimator:

$$H_{jn}(t|z) = \frac{1}{nf_n(z)} \sum_{i=1}^{n} I(Y_i \leq t, \xi_i = j) \mathbf{K}_{h_n}(z - Z_i)$$

where

- $\mathbf{K}$ is a kernel function on $\mathbb{R}^d$
- $h_n \searrow 0$ a bandwidth
- f_n a kernel type estimate of f:

$$f_n(z) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{K}_{h_n}(z - Z_i)$$

where

$$\mathbf{K}_{h_n}(z) = \frac{1}{h_n} \mathbf{K}\left(z/h_n\right).$$

Final estimator of $r_j(z)$

Plugging-in $\bar{G}_n$ and $H_{jn}(\cdot|z)$ in $r_j(z)$ we obtain with the convention $0 = 0/0$:

$$\begin{aligned}\hat{r}_{jn}(z) &= \int_0^{\tau_z} \frac{\psi(t)}{\bar{G}_n} dH_{jn}(t|z) \\ &= \frac{1}{nf_n(z)} \sum_{i=1}^{n} \frac{\psi(Y_i) I(Y_i \leq \tau_z) I(\xi_i = j) \mathbf{K}_{h_n}(z - Z_i)}{\bar{G}_n(Y_i)} \\ &= \frac{1}{f_n(z)} \int_0^{\tau_z} \frac{\psi(t)}{\bar{G}_n(Y_i)} dK_{jn}(t|z),\end{aligned}$$

with

$$K_{jn}(t|z) = \frac{1}{n} \sum_{i=1}^{n} I(t \leq \tau_z) I(\xi_i = j) \mathbf{K}_{h_n}(z - Z_i)$$

Outline

Assumptions

A. $\bar{H}(\tau_z|z) > 0$, $\bar{G}(\tau_z) > 0$ and $F_X(\tau_z|z) < 1$.

B. f continuous at z.

C. $s \mapsto H_j(t|s)$ continuous at z, uniformly in $t \in [0, \tau_z]$.

D. $\mathbf{K} = \phi \circ p$, p=polynomial and ϕ positive bounded real function with BV. supp$\mathbf{K} \subset [-1,1]^d$ and

$$(i) \quad \int_{\mathbb{R}^d} \mathbf{K}(s)ds, \qquad (ii) \quad \int_{\mathbb{R}^d} s\mathbf{K}(s)ds = 0.$$

E. $h_n = cn^{-\alpha}$ with $\alpha \in ((5d)^{-1}, d^{-1})$.

F. Functions f and $s \mapsto H_j(t|s)$ (for all $t \in [0, \tau_z]$) are twice continuously differentiable at z, and the second derivative of $s \mapsto H_j(t|s)f(s)$ is continuous at z, uniformly in $t \in [0, \tau_z]$.

Consistency and CLT

Theorem

- *Under Conditions A–E,* $\hat{r}_{jn}(z) \xrightarrow{a.s.} r_j(z)$.

Consistency and CLT

Theorem

- *Under Conditions A–E,* $\hat{r}_{jn}(z) \xrightarrow{a.s.} r_j(z)$.
- *Under Conditions A–F,* $(nh_n^d)^{1/2}(\hat{r}_{jn}(z) - r_j(z)) \rightsquigarrow \mathcal{N}(0, \sigma_j^2(z))$, *where*

$$\begin{aligned}\sigma_j^2(z) &= \frac{\|\mathbf{K}\|^2_{L^2(\mathbb{R}^d)}}{f(z)} \left(r_j(z) + 2r_j(z) \int_0^{\tau_z} \frac{\psi(t)}{\bar{G}^2(t)} H_j(t|z) dK_j(t|z) \right. \\ &+ \left. \int_0^{\tau_z} \int_0^{\tau_z} \frac{\psi(t)\psi(s)}{\bar{G}^2(t)\bar{G}^2(s)} H_j(s \wedge t|z) dK_j(s|z) dK_j(t|z) \right).\end{aligned}$$

Sketch of the proof: consistency

1. $(f_n(z), \bar{G}_n, K_{jn}(\cdot|z))$ converges uniformly a.s. to $(f(z), \bar{G}, K_j(\cdot|z))$: the empirical process part is treated by controlling the bracketing numbers (van der Vaart and Welner, 1996) and the convergence of $\bar{G}_n$ follows from Stute and Wang (1993).

Sketch of the proof: consistency

1. $(f_n(z), \bar{G}_n, K_{jn}(\cdot|z))$ converges uniformly a.s. to $(f(z), \bar{G}, K_j(\cdot|z))$: the empirical process part is treated by controlling the bracketing numbers (van der Vaart and Welner, 1996) and the convergence of $\bar{G}_n$ follows from Stute and Wang (1993).
2. Let $\phi : \mathbb{R} \times \ell[0, \tau_z] \times \ell[0, \tau_z] \to \mathbb{R}$

$$\phi(x, u, v) = \frac{1}{x} \int_{[0,\tau_z]} \frac{\psi(s)}{u(s)} dv(s)$$

is continuous at $(f(z), \bar{G}, K_j(\cdot|z))$.

Sketch of the proof: consistency

1. $(f_n(z), \bar{G}_n, K_{jn}(\cdot|z))$ converges uniformly a.s. to $(f(z), \bar{G}, K_j(\cdot|z))$: the empirical process part is treated by controlling the bracketing numbers (van der Vaart and Welner, 1996) and the convergence of $\bar{G}_n$ follows from Stute and Wang (1993).
2. Let $\phi : \mathbb{R} \times \ell[0, \tau_z] \times \ell[0, \tau_z] \to \mathbb{R}$

$$\phi(x, u, v) = \frac{1}{x} \int_{[0,\tau_z]} \frac{\psi(s)}{u(s)} dv(s)$$

is continuous at $(f(z), \bar{G}, K_j(\cdot|z))$.

3. Continuous mapping theorem.

Sketch of the proof: CLT

1. Prove that:

$$(nh_n^d)^{1/2}\left((f_n(z), \bar{G}_n, K_{jn}(\cdot|z)) - (f(z), \bar{G}, K_j(\cdot|z))\right) \rightsquigarrow (\mathcal{N}_z, 0, \mathcal{G}_z),$$

by controlling the entropy with brackets.

Sketch of the proof: CLT

1. Prove that:

$$(nh_n^d)^{1/2}\left((f_n(z),\bar{G}_n,K_{jn}(\cdot|z))-(f(z),\bar{G},K_j(\cdot|z))\right)\rightsquigarrow(\mathcal{N}_z,0,\mathcal{G}_z),$$

by controlling the entropy with brackets.

2. $\phi:\mathbb{R}\times\ell[0,\tau_z]\times\ell[0,\tau_z]\longrightarrow\mathbb{R}$ is Hadamard differentiable at $(f(z),\bar{G},K_j(\cdot|z))$, by following an example in van der Vaart (1998).

Sketch of the proof: CLT

1. Prove that:

$$(nh_n^d)^{1/2}\left((f_n(z), \bar{G}_n, K_{jn}(\cdot|z)) - (f(z), \bar{G}, K_j(\cdot|z))\right) \rightsquigarrow (\mathcal{N}_z, 0, \mathcal{G}_z),$$

   by controlling the entropy with brackets.
2. $\phi : \mathbb{R} \times \ell[0,\tau_z] \times \ell[0,\tau_z] \to \mathbb{R}$ is Hadamard differentiable at $(f(z), \bar{G}, K_j(\cdot|z))$, by following an example in van der Vaart (1998).
3. Apply the δ-méthode.

Remark

Asymptotic bias disappear because of Assumptions on $\mathbf{K}$ *and regularity conditions on* f *and* K_j.

Outline

Model

Joint conditional distribution of (T_1, T_2):

$$\bar{F}(t_1, t_2|z) = \exp(-e^z(\lambda_1 t_1 + \lambda_2 t_2 + \theta t_1 t_2))$$

for $t_1, t_2 \geq 0$ and $0 < \theta < \lambda_1 \lambda_2$. $Z \sim \mathcal{N}(0, 1)$.
Parameters of simulated data:

$$\lambda_1 = 0.1, \quad \lambda_2 = 0.15, \quad \lambda_C = 0.35, \quad \text{and} \quad \theta = 0.01.$$

Causes percentages: $\approx 42\%$ of cause 1, $\approx 38\%$ of cause 2, $\approx 20\%$ of censoring.

Estimation results

z	0	1	2
$\mathbb{E}[T_1 I(\eta=1)\|z]$	1.399	0.556	0.212
$n = 200$	1.420 (0.526)	0.575 (0.224)	0.219 (0.194)
$n = 500$	1.439 (0.363)	0.573 (0.161)	0.223 (0.131)
$n = 1000$	1.380 (0.270)	0.568 (0.121)	0.221 (0.101)

Table: Estimation of $\mathbb{E}(T_1 I(\eta = 1)|z)$ for $z \in \{0, 1, 2\}$: mean and standard deviation (within parenthesis) of $N = 1000$ estimates for various sample sizes n.

Estimation de $E[T_1 I(\eta = 1)|z]$

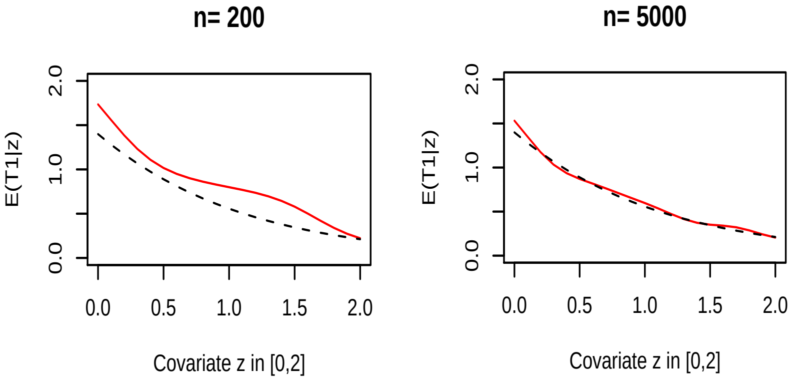

Figure: $\psi(t) = t$, $j = 1$, $n = 200$ and $n = 5000$

Estimation de $F_1(t|0)$

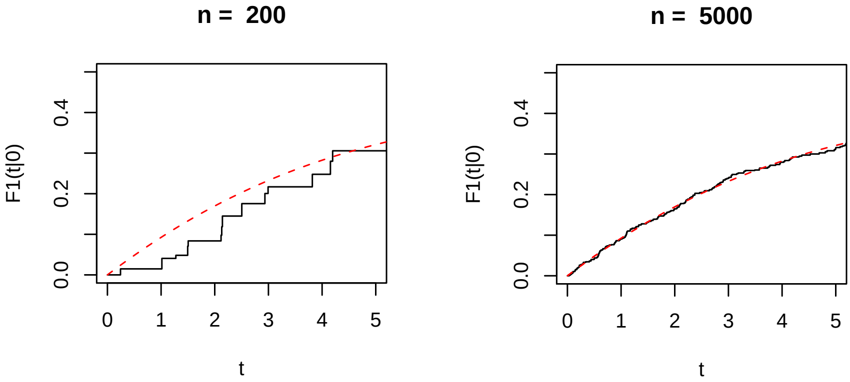

Figure: $\psi_s(t) = I(t \leq s)$, $j = 1$, $n = 200$ and $n = 5000$

Estimation de $F_1(t|1)$

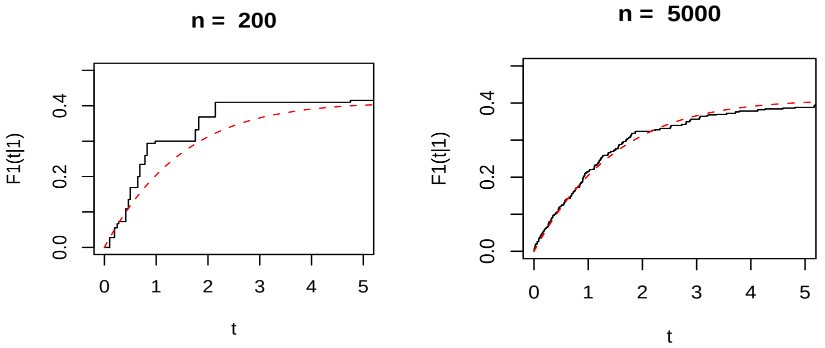

Figure: $\psi_s(t) = I(t \leq s)$, $j = 1$, $n = 200$ and $n = 5000$

Outline

Other results: convergence rates

Assumptions: (X, η) and C are independent conditionally on Z. The distribution of C depends on Z but it still holds

$$r_j(z) = \frac{1}{f(z)} \int_{[0,\tau_z]} \frac{\psi(t)}{\bar{G}(t|z)} dK_j(t|z).$$

$\bar{G}(t|z)$ is estimated by the Dabrowska (SJS, 1987) estimator. For some $\Delta \subset \text{supp}(f)$ the expected result is

$$\begin{aligned} \sup_{z\in\Delta} |\hat{r}_{jn}(z) - r_j(z)| &= O\left((nh_n^d)^{-1/2}(\log\log n + \log h_n^{-1})^{1/2}\right) \\ &+ O\left((nh_n^{2d})^{-1}\right) + O\left(h_n^{2d}\right) \quad a.s. \end{aligned}$$

by extending some results by Giné and Guillou (AIHP, 2002).

Some interesting problems: more missing data

- Uncertainty on the causes:
  there is a collection $\{S_\ell; \ell = 1, \dots, k\}$ of subsets of J, and informations are of the type

$$\xi \in \mathcal{S}_\ell \subset J \text{ with eventually } \#\mathcal{S}_\ell > 1.$$

- Competing risk including the cure assumption:
  we assume that in $X = \min(T_1, \dots, T_m)$, for some T_j we may have $\mathbb{P}(T_j = +\infty) > 0$. In this case

$$\mathbb{P}(T_j \leq t) = p_j F_j(t) + (1 - p_j),$$

  where F_j is a df and $1 - p_j = \mathbb{P}(T_j = +\infty)$.

THANK YOU!