Nonparametric copula estimation for censored data

Svetlana Gribkova, Olivier Lopez

Laboratoire de Statistique Théorique et Appliquée, Paris 6

27 August 2013

Outline

Copula function

A copula is a multivariate distribution function on $[0,1]^d$ with uniform marginal distributions.

Theorem (Sklar, 1959)

If $F_1, \ldots, F_d$ are continuous marginal distribution functions of $F(t_1, \ldots, t_d)$, then there exists a unique distribution function $C(u_1, \ldots, u_d)$ on $[0,1]^d$ with uniform marginals (copula) such that

$$F(t_1, \ldots, t_d) = C(F_1(t_1), \ldots, F_d(t_d))$$

Copulas permit to model separately the marginal distributions and the dependence structure

Problem formulation

- Nonparametric copula estimation for bivariate censored data
- Extension of nonparametric goodness-of-fit test to censored data

Bivariate right censored data

Instead of $T^{(1)}$ and $T^{(2)}$ one observes

$$\begin{cases} Y^{(1)} = T^{(1)} \wedge C^{(1)} \quad \textit{and} \quad \delta^{(1)} = \mathbb{1}_{T^{(1)} \leq C^{(1)}} \\ Y^{(2)} = T^{(2)} \wedge C^{(2)} \quad \textit{and} \quad \delta^{(2)} = \mathbb{1}_{T^{(2)} \leq C^{(2)}} \end{cases}$$

where $C^{(1)}$ and $C^{(2)}$ are censoring random variables.

Problem formulation

- Nonparametric copula estimation for bivariate censored data
- Extension of nonparametric goodness-of-fit test to censored data

Bivariate right censored data

Instead of $T^{(1)}$ and $T^{(2)}$ one observes

$$\begin{cases} Y^{(1)} = T^{(1)} \wedge C^{(1)} \quad \textit{and} \quad \delta^{(1)} = \mathbb{1}_{T^{(1)} \leq C^{(1)}} \\ Y^{(2)} = T^{(2)} \wedge C^{(2)} \quad \textit{and} \quad \delta^{(2)} = \mathbb{1}_{T^{(2)} \leq C^{(2)}} \end{cases}$$

where $C^{(1)}$ and $C^{(2)}$ are censoring random variables.

The variables of interest are not directly observed and the classical empirical copula estimator is not available

Examples

Loss and allocated loss adjustment expenses

Bivariate insurance data with

- $T^{(1)}$ – loss amount, censored by the policy limit $C^{(1)}$
- $T^{(2)}$ – allocated loss adjustment expense, observed

Observations are composed of $(Y_i^{(1)}, Y_i^{(2)}, \delta_i^{(1)}, \delta_i^{(2)})_{1\le i\le n}$ with $\delta_i^{(2)} = 1$ and $Y_i^{(2)} = T_i^{(2)}$ a.s.

Joint life insurance policies

Joint survival data on married couples, subscribed a pension contract

At the end of the observation period most couples are still alive $\Rightarrow$ censoring

- $T^{(1)}, T^{(2)}$ – lifetimes of spouses
- $C^{(1)}, C^{(2)}$ – their ages at the exit from the study, $\varepsilon = C^{(2)} - C^{(1)}$

Observations are composed of $(Y_i^{(1)}, Y_i^{(2)}, \delta_i^{(1)}, \delta_i^{(2)}, \varepsilon_i)_{1\le i\le n}$

Estimation: i.i.d. observations vs censored observations

Uncensored data

$(T_i^{(1)}, T_i^{(2)})_{1 \leq i \leq n}$

- Distribution function estimator:

$$F_n(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{T_i^{(1)} \leq t_1, T_i^{(2)} \leq t_2}$$

- Parametric copula estimator: maximum likelihood
- Nonparametric copula estimator (Deheuvels, 1979):

$$\mathfrak{C}_n(u, v) = F_n(F_{1n}^{-1}(u), F_{2n}^{-1}(v))$$

- Goodness-of-fit

Censored data

$(Y_i^{(1)}, Y_i^{(2)}, \delta_i^{(1)}, \delta_i^{(2)})_{1 \leq i \leq n}$

Estimation: i.i.d. observations vs censored observations

Uncensored data

$(T_i^{(1)}, T_i^{(2)})_{1 \le i \le n}$

- Distribution function estimator:

$$F_n(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{T_i^{(1)} \le t_1, T_i^{(2)} \le t_2}$$

- Parametric copula estimator: maximum likelihood
- Nonparametric copula estimator (Deheuvels, 1979):

$$\mathfrak{C}_n(u, v) = F_n(F_{1n}^{-1}(u), F_{2n}^{-1}(v))$$

- Goodness-of-fit

Censored data

$(Y_i^{(1)}, Y_i^{(2)}, \delta_i^{(1)}, \delta_i^{(2)})_{1 \le i \le n}$

- Distribution function estimator:

  $d = 1$: Kaplan-Meier estimator

  $d = 2$: several bivariate generalizations

Estimation: i.i.d. observations vs censored observations

Uncensored data

$(T_i^{(1)}, T_i^{(2)})_{1 \le i \le n}$

- Distribution function estimator:

$$F_n(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{T_i^{(1)} \le t_1, T_i^{(2)} \le t_2}$$

- Parametric copula estimator: maximum likelihood
- Nonparametric copula estimator (Deheuvels, 1979):

$$\mathfrak{C}_n(u, v) = F_n(F_{1n}^{-1}(u), F_{2n}^{-1}(v))$$

- Goodness-of-fit

Censored data

$(Y_i^{(1)}, Y_i^{(2)}, \delta_i^{(1)}, \delta_i^{(2)})_{1 \le i \le n}$

- Distribution function estimator:

  $d = 1$: Kaplan-Meier estimator

  $d = 2$: several bivariate generalizations

- Parametric copula estimator: censored maximum likelihood

Estimation: i.i.d. observations vs censored observations

Uncensored data

$(T_i^{(1)}, T_i^{(2)})_{1 \leq i \leq n}$

- Distribution function estimator:

$$F_n(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{T_i^{(1)} \leq t_1, T_i^{(2)} \leq t_2}$$

- Parametric copula estimator: maximum likelihood
- Nonparametric copula estimator (Deheuvels, 1979):

$$\mathfrak{C}_n(u, v) = F_n(F_{1n}^{-1}(u), F_{2n}^{-1}(v))$$

- Goodness-of-fit

Censored data

$(Y_i^{(1)}, Y_i^{(2)}, \delta_i^{(1)}, \delta_i^{(2)})_{1 \leq i \leq n}$

- Distribution function estimator:

  $d = 1$: Kaplan-Meier estimator

  $d = 2$: several bivariate generalizations
- Parametric copula estimator: censored maximum likelihood
- Nonparametric copula estimator valid for a large scope of model configurations and censoring schemes: ?
- Goodness-of-fit: ?

Inference from censored sample: $d = 1$

- The variable of interest T (df. F_T) is censored by C (df. G_C).
- Observations are composed of $(Y_i, \delta_i)_{1 \leq i \leq n}$ with $Y_i = \min(T_i, C_i)$.
- Kaplan-Meier estimator of the distribution function F_T:

$$F_n(t) = \sum_{i=1}^{n} \delta_i W_{in}(Y_i) \mathbb{1}_{Y_i \leq t}, \quad W_{in}(y) = \frac{1}{n(1 - \hat{G}_C(y-))}.$$

  - $\hat{G}_C$ – Kaplan-Meier estimator of G_C
  - $nW_{in}(y)$ can be seen as estimator of $W^*(y) = (1 - G_C(y-))^{-1}$
  - For any function ϕ we have $\mathbb{E}(\delta W^*(Y)\phi(Y)) = \mathbb{E}(\phi(T))$
  - $\sqrt{n}(F_n(t) - F_T(t)) \rightsquigarrow G_F(t)$

Inference from censored sample: $d = 2$

- To estimate copula one needs a distribution function estimator
- General case: estimators suffer from various drawbacks
- Several bivariate $\sqrt{n}$-consistent generalizations of Kaplan-Meier estimator in particular models:
  - Stute (JMVA, 1993), estimation when only one variable is censored
  - Lopez, Saint-Pierre (JSPI, 2012), estimation when censoring variables are linked through a copula
  - Lopez (IME, 2012), estimation of the bivariate distribution function under censoring and truncation
  - Gribkova, Lopez, Saint-Pierre (JMVA, 2013), estimation of the bivariate distribution function when the difference between censoring variables is observed

Discrete copula estimator from censored sample

- For these and some other censoring configurations, the bivariate distribution function estimator takes a generic form

$$F_n(t_1, t_2) = \sum_{i=1}^{n} \delta_i^{(1)} \delta_i^{(2)} W_{in}(Y_i^{(1)}, Y_i^{(2)}) \mathbb{1}_{Y_i^{(1)} \leq t_1, Y_i^{(2)} \leq t_2}$$

- Copula estimator relies on this generic form without imposing any condition on the form of the weights

Discrete copula estimator

$$\begin{aligned} C_n(u, v) &= F_n(F_{1n}^{-1}(u), F_{2n}^{-1}(v)) \\ &= \sum_{i=1}^{n} \delta_i^{(1)} \delta_i^{(2)} W_{in}(Y_i^{(1)}, Y_i^{(2)}) \mathbb{1}_{Y_i^{(1)} \leq F_{1n}^{-1}(u), Y_i^{(2)} \leq F_{2n}^{-1}(v)}. \end{aligned}$$

Examples of censoring configurations

1. One variable $T^{(1)}$ is censored, and $T^{(2)}$ is observed

$$W_{in} = \frac{1}{n(1 - \hat{G}_{C^{(1)}}(Y_i^{(1)}-))}$$

2. Censoring variables linked through a copula function $\tilde{C}$

$$W_{in} = \frac{1}{n\tilde{C}(1 - \hat{G}_{C^{(1)}}(Y_i^{(1)}-), 1 - \hat{G}_{C^{(2)}}(Y_i^{(2)}-))}$$

3. Difference between censoring variables $\varepsilon_i = C_i^{(2)} - C_i^{(1)}$ is observed

$$W_{in} = \frac{1}{n(1 - \hat{G}_{C^{(1)}}(max(Y_i^{(1)}, Y_i^{(2)} - \varepsilon_i)-))}$$

Censored data: smooth copula estimator

Smooth bivariate estimator of the distribution function,

$$\widehat{F}_n^1(t_1, t_2) = \sum_{i=1}^{n} \delta_i^{(1)} \delta_i^{(2)} W_{in}(Y_i^{(1)}, Y_i^{(2)}) K\left(\frac{t_1 - Y_i^{(1)}}{h}\right) K\left(\frac{t_2 - Y_i^{(2)}}{h}\right),$$

permits to define a smooth estimator of copula by

$$\widehat{C}_n^1(u, v) = \widehat{F}_n^1((\widehat{F}_{1n}^1)^{-1}(u), (\widehat{F}_{2n}^1)^{-1}(v)).$$

Copula density estimator is given by

$$c_n(u, v) = \frac{\partial^2}{\partial u \partial v} C_n(u, v) = \frac{1}{h^2} \sum_{i=1}^{n} \frac{\delta_i^{(1)} \delta_i^{(2)} W_{in} k\left(\frac{F_{1n}^{-1}(u) - Y_i^{(1)}}{h}\right) k\left(\frac{F_{2n}^{-1}(v) - Y_i^{(2)}}{h}\right)}{f_{1n}(F_{1n}^{-1}(u)) f_{2n}(F_{2n}^{-1}(v))},$$

with $f_{in}(x) = \frac{\partial}{\partial x} F_{in}(x)$.

Kernel estimator based on the transformed variables

As in Omelka, Gijbels, Veraverbeke, 2009, for some df. $\Phi(x)$ consider a random vector

$$(\tilde{T}_1, \tilde{T}_2) = (\Phi^{-1}[F_1(T^{(1)})], \Phi^{-1}[F_2(T^{(2)})])$$

and pseudo-observations $(\Phi^{-1}[F_{1n}(Y_i^{(1)})], \Phi^{-1}[F_{2n}(Y_i^{(2)})])_{1 \le i \le n}$.

$\tilde{T}_1$ and $\tilde{T}_2$ are linked by the same copula as $T^{(1)}$ and $T^{(2)}$.

$$\widehat{\mathbb{F}}_n^2(t_1, t_2) = \sum_{i=1}^{n} \delta_i^{(1)} \delta_i^{(2)} W_{in} K\left(\frac{t_1 - \Phi^{-1}[F_{1n}(Y_i^{(1)})]}{h}, \frac{t_2 - \Phi^{-1}[F_{2n}(Y_i^{(2)})]}{h}\right)$$

$$\widehat{C}_n(u, v) = \widehat{F}_n^2(\Phi^{-1}(u), \Phi^{-1}(v))$$

- Less affected by the marginal distributions
- Allows copula density to be unbounded in the corners of the unit square

Asymptotic results - discrete estimator

Theorem (Consistency)

Let $\mathcal{T}_1 = [-\infty, A_1]$, *and* $\mathcal{T}_2 = [-\infty, A_2]$, *such that*

$$\sup_{t_1 \in \mathcal{T}_1, t_2 \in \mathcal{T}_2} |F_n(t_1, t_2) - F(t_1, t_2)| = O_P(n^{-1/2}).$$

Let $F_1(\mathcal{T}_1) = [0, u_1]$, *and* $F_2(\mathcal{T}_2) = [0, u_2]$. *Then, for any* $\eta > 0$,

$$\sup_{u \leq u_1 - \eta, v \leq u_2 - \eta} |\mathfrak{C}_n(u, v) - C(u, v)| = O_P(n^{-1/2}).$$

Theorem (Weak convergence)

Suppose that F *has continuous marginal distribution functions and partial derivatives of its copula function exist and are continuous. Suppose that* $\sqrt{n}(F_n(t_1, t_2) - F(t_1, t_2)) \rightsquigarrow G_F(t_1, t_2)$. *Then the empirical process*

$$\sqrt{n}(C_n(u, v) - C(u, v)) \rightsquigarrow G_C(u, v) \text{ in } l^\infty([0, 1]^2)$$

Smooth estimation: conditions on weights

Let $\tau_j = \inf\{t : F_j(t) = 1\}$, for all $\mathcal{Y} \in (-\infty, \tau_1) \times (-\infty, \tau_2)$, assume that

1. $n\delta_i^{(1)}\delta_i^{(2)} W_{in} = \delta_i^{(1)}\delta_i^{(2)}\hat{g}(Y_i^{(1)}, Y_i^{(2)})$, where $\hat{g}(y_1, y_2) \underset{P}{\rightarrow} g(y_1, y_2)$ and g is bounded and twice continuously differentiable on $\mathcal{Y}$ with bounded second order partial derivatives.
2. $\sup_{(t_1,t_2)\in\mathcal{Y}} |\hat{g}(t_1, t_2) - g(t_1, t_2)| = O_P(n^{-1/2})$
3. for all $\psi \in \mathcal{F}$, where $\mathcal{F}$ denotes a Donsker class of bounded functions,

$$\sum_{i=1}^{n}[nW_{in} - W_i]\psi(Y_{1i}, Y_{2i}) = \frac{1}{n}\sum_{i=1}^{n}\eta^{\psi}(Y_i^{(1)}, Y_i^{(2)}, \delta_i^{(1)}, \delta_i^{(2)}) + R_n(\psi),$$

   with $W_i = \delta_i^{(1)}\delta_i^{(2)} g(Y_i^{(1)}, Y_i^{(2)})$ and $\sup_{\psi\in\mathcal{F}} |R_n(\psi)| = o_P(n^{-1/2})$,
4. $E(\delta^{(1)}\delta^{(2)} g(T_1, T_2)^2) < +\infty$ (permits to have convergence on whole $[0, 1]^2$)

Asymptotic results - smooth estimators

Theorem

Under the conditions on the weights W_{in} and with $h^2\sqrt{n} \to 0$,

1. *If $F(t_1, t_2)$ is twice differentiable with bounded second order derivatives on R^2, then*
$$\sqrt{n}(\widehat{C}_n^1 - C)(u, v) \rightsquigarrow G_C(u, v) \text{ in } l^\infty([0, 1]^2).$$

2. *Suppose that,*
   - *Φ is a distribution function such that $\Phi'(x)$ and $\frac{\Phi'(x)^2}{\Phi(x)}$ are bounded*
   - *C has bounded second order partial derivatives on $(0, 1)^2$ and*
   $$C_{uu} = O\left(\frac{1}{u(1-u)}\right), C_{vv} = O\left(\frac{1}{v(1-v)}\right), C_{uv} = O\left(\frac{1}{\sqrt{uv(1-u)(1-v)}}\right)$$
   - *C_u,C_v are continuous in $[0, 1]^2 \setminus \{(0, 0), (1, 0), (0, 1), (1, 1)\}$*

   Then,
$$\sqrt{n}(\widehat{C}_n^2 - C)(u, v) \rightsquigarrow G_C(u, v) \text{ in } l^\infty([0, 1]^2)$$

Application to a particular model

- Motivated by data from a large Canadian insurer containing joint lifetimes of members of couples subscribed an insurance contract.
- Instead of observing lifetimes $(T^{(1)}, T^{(2)})$, one observes

$$\begin{cases} Y^{(1)} = \inf(T^{(1)}, C^{(1)}) \ \ et \ \ \delta^{(1)} = \mathbb{1}_{T^{(1)} \leq C^{(1)}} \\ Y^{(2)} = \inf(T^{(2)}, C^{(2)}) \ \ et \ \ \delta^{(2)} = \mathbb{1}_{T^{(2)} \leq C^{(2)}} \end{cases}$$

  Model assumptions:

  - $C^{(2)} = C^{(1)} + \varepsilon$, with ε observed (e.g. age difference)
  - $(T^{(1)}, T^{(2)})$ independant from ε and from $C^{(1)}$
  - $P(T^{(1)} = C^{(1)}) = P(T^{(2)} = C^{(1)} + \varepsilon) = 0$
  - $C^{(1)}$ is independent from ε

- Observations are composed of $(Y_i^{(1)}, Y_i^{(2)}, \delta_i^{(1)}, \delta_i^{(2)}, \varepsilon_i)_{1 \leq i \leq n}$.

The bivariate distribution function estimator

- This model falls in the general setup (G., Lopez, Saint-Pierre (JMVA, 2013)):

$$F_n(t_1, t_2) = \sum_{i=1}^{n} \delta_i^{(1)} \delta_i^{(2)} W_{in} \mathbb{1}_{Y_i^{(1)} \leq t_1, Y_i^{(2)} \leq t_2},$$

  with the weights

$$W_{in}(Y_i^{(1)}, Y_i^{(2)}, \varepsilon_i) = \frac{1}{n\hat{S}_G(\max(Y_i^{(1)}, Y_i^{(2)} - \varepsilon_i)-)},$$

  where $\hat{S}_G$ is the Kaplan-Meier estimator of the sf. of the variable $C^{(1)}$.

- Idea: $nW_{in}(y_1, y_2, e)$ estimate $W^*(y_1, y_2, e) = (S_G(\max(y_1, y_2 - e)-))^{-1}$, with

$$E(\delta^{(1)}\delta^{(2)} W^*(Y^{(1)}, Y^{(2)}, \epsilon)\psi(Y^{(1)}, Y^{(2)})) = E(\psi(T^{(1)}, T^{(2)}))$$

Asymptotic property of the estimator

Asymptotic representation uniform over a class of function $\mathcal{F}$

- Let $\mathcal{G}$ be the class of positive monotonic functions bounded by 1 and $\xi(T^{(1)}, T^{(2)}, C^{(1)}, \varepsilon) = \delta^{(1)}\delta^{(2)} S_G(\max(T^{(1)}, T^{(2)} - \varepsilon)-)^{-2}$. For any (t_0, u_0) in $\mathbb{R}^2$ such that $S_F(t_0, u_0) > 0$, assume that the class $\{(T^{(1)}, T^{(2)}, C^{(1)}, \varepsilon) \to \xi(T^{(1)}, T^{(2)}, C^{(1)}, \varepsilon) f(T^{(1)}, T^{(2)}) g(\max(T^{(1)}, T^{(2)} - \varepsilon)-) \mathbb{1}_{T^{(1)} \leq t_0, T^{(2)} \leq u_0}, f \in \mathcal{F}, g \in \mathcal{G}\}$ is Donsker.
- Uniformly over $\phi \in \mathcal{F}$,

$$\int \phi(t_1, t_2) d(F_n - F^*)(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} \psi_\phi(Y_i^{(1)}, Y_i^{(2)}, \varepsilon_i) + R_n(\phi),$$

where $E(\psi_\phi(Y^{(1)}, Y^{(2)}, \varepsilon)) = 0$ and $|R_n(\phi)| = o_P(n^{-1/2})$.

Corollary:

- $\sqrt{n}(F_n(t_1, t_2) - F(t_1, t_2)) \rightsquigarrow G_F(t_1, t_2)$
- $C_n(u, v) = F_n(F_{1n}^{-1}(u), F_{2n}^{-1}(v))$.

Application to goodness-of-fit testing

A bivariate parametric copula model

The distribution function of random vector $(T^{(1)}, T^{(2)})$ is modeled by

$$F_\theta(t_1, t_2) = C_\theta(F_1(t_1), F_2(t_2)), \quad C_\theta \in \mathcal{C}_0 = \{C_\theta, \theta \in \Theta\}.$$

Given a sample of data, the model is estimated by

$$\hat{F}(t_1, t_2) = C_{\hat{\theta}_n}(\hat{F}_1(t_1), \hat{F}_2(t_2))$$

Application to goodness-of-fit testing

A bivariate parametric copula model

The distribution function of random vector $(T^{(1)}, T^{(2)})$ is modeled by

$$F_\theta(t_1, t_2) = C_\theta(F_1(t_1), F_2(t_2)), \quad C_\theta \in \mathcal{C}_0 = \{C_\theta, \theta \in \Theta\}.$$

Given a sample of data, the model is estimated by

$$\hat{F}(t_1, t_2) = C_{\hat{\theta}_n}(\hat{F}_1(t_1), \hat{F}_2(t_2))$$

- Copula misspecification may lead to an important error of estimation
- Goodness-of-fit test of $H_0 : C \in \mathcal{C}_0 \quad \text{vs} \quad H_1 : C \notin \mathcal{C}_0$

Goodness-of-fit test

Goodness-of-fit procedure:

- Estimate copula parameter under H_0 and obtain $C_{\theta_n}(u, v)$ (maximum likelihood under censoring)

Goodness-of-fit test

Goodness-of-fit procedure:

- Estimate copula parameter under H_0 and obtain $C_{\theta_n}(u, v)$ (maximum likelihood under censoring)
- Take a nonparametric weakly convergent copula estimator $C_n(u, v)$ (under censoring!)

Goodness-of-fit test

Goodness-of-fit procedure:

- Estimate copula parameter under H_0 and obtain $C_{\theta_n}(u, v)$ (maximum likelihood under censoring)
- Take a nonparametric weakly convergent copula estimator $C_n(u, v)$ (under censoring!)
- Consider the empirical process

$$\gamma_n(u, v) = \sqrt{n}(C_n - C_{\theta_n})(u, v), \quad 0 \leq u, v \leq 1$$

and test statistics

$$d_n = \int_{[0,1]^d} \gamma_n(u, v)^2 dC_n(u, v).$$

Goodness-of-fit test

Goodness-of-fit procedure:

- Estimate copula parameter under H_0 and obtain $C_{\theta_n}(u,v)$ (maximum likelihood under censoring)
- Take a nonparametric weakly convergent copula estimator $C_n(u,v)$ (under censoring!)
- Consider the empirical process

$$\gamma_n(u,v) = \sqrt{n}(C_n - C_{\theta_n})(u,v), \quad 0 \leq u, v \leq 1$$

  and test statistics

$$d_n = \int_{[0,1]^d} \gamma_n(u,v)^2 dC_n(u,v).$$

- Critical region: $\mathcal{R} = \{d_n > c\}$

Calculation of p-value

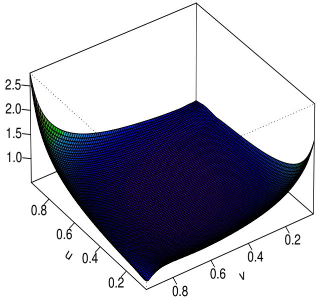
Bootstrap procedure for calculating p-value: for $b = 1, ..., B$

- simulate $(T_{1i}^b, T_{2i}^b)_{1 \le i \le n}$ according to the distribution defined by C_{θ_n} and the marginal distributions F_{1n} and F_{2n}
- simulate $(C_{1i}^b)_{1 \le i \le n}$ according to G_n (K-M estimator of censoring)
- simulate $(\varepsilon_{1i}^b)_{1 \le i \le n}$ according to the empirical d.f. of $(\varepsilon_i)_{1 \le i \le n}$
- let $C_{2i}^b = C_{1i}^b + \varepsilon_i$ and compute the $b-$th bootstrap sample $(Y_{1i}^b, Y_{2i}^b, \varepsilon_i^b, \delta_{1i}^b, \delta_{2i}^b)_{1 \le i \le n}$
- using the $b-$th bootstrap sample, compute estimators θ_n^b and C_n^b and the corresponding distance d_n^b
- using the bootstrap sample for d_n, calculate the p-value

Application to Canadian dataset

- Database from a large Canadian insurer, 11454 contracts, more than 90% of censoring. Test for the survival copula linking two lifetimes:

Model	Parameter estimation	p-value
Clayton	4.89	0.391
Frank	11.41	0.416
Nelsen 4.2.20	1.33	0.103



Conclusion

- Nonparametric estimation of copula and its density under bivariate censoring
- The estimators are valid for a range of censoring configurations
- Central limit theorems for the copula estimators
- An application to a particular censoring configuration
- Extension of nonparametric goodness-of-fit test to the case of censored data
- New bootstrap procedure to approximate the p-value