Modeling the evolution of the dependence structure between two lifetimes

Olivier Lopez, Philippe Saint Pierre

Université Pierre et Marie Curie Paris VI
Laboratoire de Statistique Théorique et Appliquée
Projet ANR Lolita

Conference Longevity 11
September 7th 2015

Outline

INTRODUCTION

- Framework :
  - Two lifetimes (T, U),
  - $T =$ lifetime of a man,
  - $U =$ lifetime of his wife.
- T and U are not independent.
- Censoring : some couples quit observation before dying.
- Truncation : observation only for people who lived long enough to enter the study.

INTRODUCTION

- Framework :
  - Two lifetimes (T, U),
  - $T =$ lifetime of a man,
  - $U =$ lifetime of his wife.
- T and U are not independent.
- Censoring : some couples quit observation before dying.
- Truncation : observation only for people who lived long enough to enter the study.

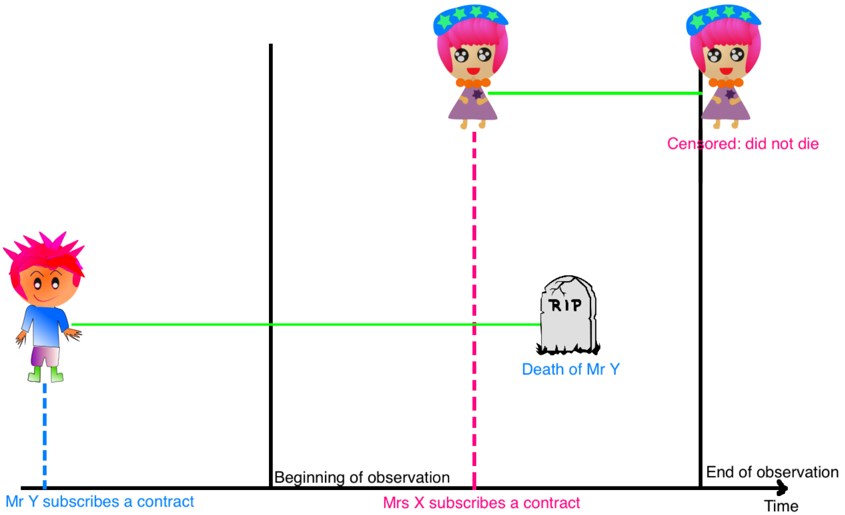
INTRODUCTION

- Framework :
  - Two lifetimes (T, U),
  - $T =$ lifetime of a man,
  - $U =$ lifetime of his wife.
- T and U are not independent.
- Censoring : some couples quit observation before dying.
- Truncation : observation only for people who lived long enough to enter the study.

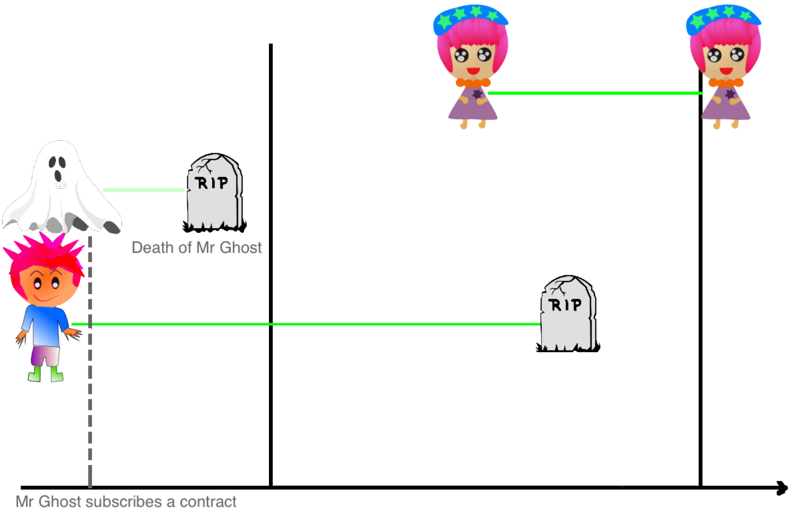
INTRODUCTION

- Framework :
  - Two lifetimes (T, U),
  - $T =$ lifetime of a man,
  - $U =$ lifetime of his wife.
- T and U are not independent.
- Censoring : some couples quit observation before dying.
- Truncation : observation only for people who lived long enough to enter the study.

Censoring and truncation



CENSORING AND TRUNCATION



Observations

Bivariate right-censored and left-truncated data

We observe n i.i.d. copies $(Y_i, Z_i, \mu_i, \nu_i, \delta_i, \gamma_i)_{1\leq i\leq n}$, with

$$\begin{cases} Y_i = \inf(T_i, C_i), \\ Z_i = \inf(U_i, D_i), \end{cases}$$

where C_i and D_i are censoring variables, and

$$\begin{cases} \delta_i = \mathbf{1}_{T_i \leq C_i}, \\ \gamma_i = \mathbf{1}_{U_i \leq D_i}, \end{cases}$$

where $Y_i \geq \mu_i$ and $Z_i \geq \nu_i$.

- We assume that :
  - (T_i, U_i) is independent from (C_i, D_i).
  - (μ_i, ν_i) is independent from (T_i, U_i, C_i, D_i).

COVARIATES

BIVARIATE RIGHT-CENSORED AND LEFT-TRUNCATED DATA

We observe n i.i.d. copies $(Y_i, Z_i, \mu_i, \nu_i, \delta_i, \gamma_i, X_i)_{1\leq i\leq n}$, with

$$\begin{cases} Y_i & = & \inf(T_i, C_i), \\ Z_i & = & \inf(U_i, D_i), \end{cases}$$

where C_i and D_i are censoring variables, and

$$\begin{cases} \delta_i & = & \mathbf{1}_{T_i\leq C_i}, \\ \gamma_i & = & \mathbf{1}_{U_i\leq D_i}, \end{cases}$$

where $Y_i \geq \mu_i$ and $Z_i \geq \nu_i$.

- We assume that :
  - (T_i, U_i) is independent from (C_i, D_i) conditionally to X_i
  - (μ_i, ν_i) is independent from (T_i, U_i, C_i, D_i) conditionally to X_i.

NATURE OF THE COVARIATES

- In the following, we focus on $X_i = (X_i^{(1)}, X_i^{(2)})$, where :
  - $X_i^{(1)}$ = birth date of the husband
  - $X_i^{(2)}$ = birth date of his wife
- May be extended to any kind of covariates :
  - salary
  - size of the family
  - ...
- One constraint : since we are using smoothing techniques, we assume the covariates to be continuous.

OUTLINE

1. INTRODUCTION
2. COPULAS
   - Definition
   - Estimation in absence of censoring and truncation
3. ESTIMATION OF THE CONDITIONAL COPULA UNDER CENSORING AND TRUNCATION
   - Estimation of the conditional distribution function
   - Estimation of the conditional copula parameter

OUTLINE

1. INTRODUCTION

2. COPULAS
   - Definition
   - Estimation in absence of censoring and truncation

3. ESTIMATION OF THE CONDITIONAL COPULA UNDER CENSORING AND TRUNCATION
   - Estimation of the conditional distribution function
   - Estimation of the conditional copula parameter

Sklar's Theorem

Sklar's Theorem - distribution functions

Let (T, U) be absolutely continuous variables with d.f. F, $F_T(t) = \mathbb{P}(T \leq t)$, $F_U(u) = \mathbb{P}(U \leq u)$. There exists a unique copula function $\mathfrak{C}$ such that

$$F(t, u) = \mathfrak{C}(F_T(t), F_U(u)).$$

Sklar's Theorem - survival functions

Let (T, U) be absolutely continuous variables with survival function S_F, $S_T(t) = \mathbb{P}(T > t)$, $S_U(u) = \mathbb{P}(U > u)$. There exists a unique copula function $\mathfrak{C}_S$ such that

$$S_F(t, u) = \mathfrak{C}_S(S_T(t), S_U(u)).$$

Moreover,

$$\mathfrak{C}_S(u, v) = u + v - 1 + \mathfrak{C}(1 - u, 1 - v).$$

Parametric copula families, notations

- We first consider that there is no covariate.
- (T_i, U_i) are i.i.d. with the same distribution as (T, U), whose dependence structure is defined by $\mathfrak{C}$.
- Assume that $\mathfrak{C} \in \{\mathfrak{C}_\theta : \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^k$.
- Let c_θ denote the copula density corresponding to $\mathfrak{C}_\theta$, that is

$$\mathfrak{c}_\theta(u, v) = \partial_u \partial_v \mathfrak{C}_\theta(u, v).$$

Estimation, margins known, no censoring nor truncation, no covariate

- Let $I = F_T(T)$ and $J = F_U(U)$.
- I and J are uniformly distributed over $[0, 1]$.
- Their joint distribution function if $\mathfrak{C}$.
- Maximum likelihood estimation :

$$\begin{aligned} \theta^* &= \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \log \mathfrak{c}_\theta(I_i, J_i) \\ &= \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \log \mathfrak{c}_\theta(F_T(T_i), F_U(U_i)). \end{aligned}$$

ESTIMATION, MARGINS UNKNOWN, NO CENSORING NOR TRUNCATION, NO COVARIATE

- Estimation of the margins by $\hat{F}_T$, $\hat{F}_U$.
- Various way of estimating the margins :
  - empirical distribution function
  - parametric models
  - semiparametric models (proportional hazard models when it comes to regression)
- Pseudo maximum likelihood estimation :

$$\begin{aligned}\theta^* &= \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \log \mathfrak{c}_\theta(\hat{I}_i, \hat{J}_i) \\ &= \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \log \mathfrak{c}_\theta(\hat{F}_T(T_i), \hat{F}_U(U_i)).\end{aligned}$$

CONDITIONAL COPULA

- The dependence structure depends on the value of the covariates X.
- We will denote $\mathfrak{C}(\cdot,\cdot|x)$ the copula of the conditional distribution of $(T,U)|X=x$.
- Nonparametric estimation : Abegaz, Gijbels, and Veraverbeke (2012), Gijbels, Omelka, Veraverbeke (2012), Acar, Craiu, and Yao (2011)...
- Here, we focus on a parametric model :

$$\mathfrak{C}(\cdot,\cdot|x)=\mathfrak{C}_{\theta(x)}(\cdot,\cdot).$$

- Our aim is then to estimate the function $\theta(x)$.

Estimation, margins UNKNOWN, no censoring nor truncation, COVARIATE

- Estimation of the conditional distributions of the margins by $\hat{F}_T(\cdot|X)$, $\hat{F}_U(\cdot|X)$.
- Pseudo maximum likelihood estimation :

$$\theta^*(x) \quad = \quad \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} w_{i,n}(x) \log \mathfrak{c}_\theta(\hat{F}_T(T_i|X_i), \hat{F}_U(U_i|X_i)),$$

where

$$w_{i,n}(x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{X_j - x}{h}\right)},$$

introducing a kernel K and a bandwidth $h > 0$.

OUTLINE

ESTIMATION OF THE DISTRIBUTION FUNCTION OF (T, U)

- Recall that
$$\frac{1}{n}\sum_{i=1}^{n}\phi(T_i, U_i) = \int \phi(t, u)d\mathbb{P}_n(t, u),$$
where $\mathbb{P}_n$ is the empirical distribution function of (T, U).
- Recall that
$$\sum_{i=1}^{n} w_{i,n}(x)\phi(T_i, U_i) = \int \phi(t, u)d\mathbb{P}_n(t, u|x),$$
where $\mathbb{P}_n$ is a kernel estimator of
$F(t, u|x) = \mathbb{P}(T \leq t, U \leq u|X = x)$.
- How to estimate (non parametrically) the distribution of (T, U) under censoring and truncation ?

PROPERTIES OF $F(t) = \mathbb{P}(T \leq t)$ (NO TRUNCATION)

- Usual definition (absolutely continuous variables) :

$$\hat{F}(t) = 1 - \prod_{Y_i \leq t} \left(1 - \frac{\delta_i}{\sum_{j=1}^{n} \mathbf{1}_{Y_j \geq Y_i}} \right).$$

- Definition as a sum (no truncation) :

$$\hat{F}(t) = \sum_{i=1}^{n} W_{in} \mathbf{1}_{Y_i \leq t},$$

where $G(t) = \mathbb{P}(C \leq t)$ and $\hat{G}$ its Kaplan-Meier estimator,

$$W_{in} = \frac{\delta_i}{n[1 - \hat{G}(Y_i -)]}.$$

Properties of the Kaplan-Meier estimator of $F(t) = \mathbb{P}(T \leq t)$ (no truncation)

- Assume that T is independent from C.
- Define

$$\begin{aligned} H_1(t) &= \mathbb{P}(Y \leq t, \delta = 1), \\ 1 - H(t) &= \mathbb{P}(Y > t) = [1 - G(t)][1 - F(t)]. \end{aligned}$$

- F is the solution of the equation :

$$dF(t) = \frac{[\int_t^\infty dF(s)]dH_1(t)}{1 - H(t)}.$$

Properties of the Kaplan-Meier estimator of $F(t) = \mathbb{P}(T \leq t)$ (no truncation)

- Assume that T is independent from C.
- Empirical version :

$$\hat{H}_1(t) = n^{-1} \sum_{i=1}^{n} \delta_i \mathbf{1}_{Y_i \leq t},$$

$$1 - \hat{H}(t) = n^{-1} \sum_{i=1}^{n} \mathbf{1}_{Y_i > t} = [1 - \hat{G}(t)][1 - \hat{F}(t)].$$

- $\hat{F}$ is the solution of the equation :

$$d\hat{F}(t) = \frac{[\int_t^\infty d\hat{F}(s)] d\hat{H}_1(t)}{1 - \hat{H}(t)}.$$

A PARTIAL DERIVATIVE EQUATION IN THE GENERAL CASE

- Assume that (T, U) is independent from (C, D), and (μ, ν) are independent from the other variables.

P.D.E. (LOPEZ (2012), *Insurance : Mathematics and Economics*)

$F(t, u) = \mathbb{P}(T \leq t, U \leq u)$ is the solution of

$$dF(t, u) = \frac{[\int_t^\infty \int_u^\infty dF(t', u')]dV(t, u)}{H(t, u)},$$

where

$$\begin{aligned} H(t, u) &= \mathbb{P}(Y \geq t, Z \geq u | Y \geq \mu, Z \geq \nu), \\ V(t, u) &= \mathbb{P}\left(\mu \leq t \leq Y, \nu \leq u \leq Z, \delta = \gamma = 1 | Y \geq \mu, Z \geq \nu\right). \end{aligned}$$

EMPIRICAL VERSION OF THE P.D.E.

- For all function ψ,

$$\frac{1}{n}\sum_{i=1}^{n}\psi(Y,Z,\delta,\gamma)\rightarrow_{n\rightarrow\infty}E[\psi(Y,Z,\delta,\gamma)|Y\geq\mu,Z\geq\nu]\ p.s.$$

- Estimate $H(t,u)=\mathbb{P}(Y\geq t,Z\geq u|Y\geq\mu,Z\geq\nu)$ par

$$\hat{H}(t,u)=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{Y_i\geq t,Z_i\geq u}.$$

- For $V(t,u)=\mathbb{P}\left(\mu\leq t\leq Y,\nu\leq u\leq Z,\delta=\gamma=1|Y\geq\mu,Z\geq\nu\right)$,

$$\hat{V}(t,u)=\frac{1}{n}\sum_{i=1}^{n}\delta_i\gamma_i\mathbf{1}_{\mu_i\leq t\leq Y_i,\nu_i\leq u\leq Z_i}.$$

- $\hat{F}$ is defined as the solution of

$$d\hat{F}(t,u)=\frac{[\int_t^\infty\int_u^\infty d\hat{F}(t',u')]d\hat{V}(t,u)}{\hat{H}(t,u)}.$$

Empirical version of the p.d.e.

- For all function ψ,

$$\frac{1}{n}\sum_{i=1}^{n}\psi(Y,Z,\delta,\gamma) \rightarrow_{n\to\infty} E[\psi(Y,Z,\delta,\gamma)|Y\geq\mu,Z\geq\nu]\ p.s.$$

- Estimate $H(t,u)=\mathbb{P}(Y\geq t,Z\geq u|Y\geq\mu,Z\geq\nu)$ par

$$\hat{H}(t,u)=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{Y_i\geq t,Z_i\geq u}.$$

- For $V(t,u)=\mathbb{P}\left(\mu\leq t\leq Y,\nu\leq u\leq Z,\delta=\gamma=1|Y\geq\mu,Z\geq\nu\right)$,

$$\hat{V}(t,u)=\frac{1}{n}\sum_{i=1}^{n}\delta_i\gamma_i\mathbf{1}_{\mu_i\leq t\leq Y_i,\nu_i\leq u\leq Z_i}.$$

- $\hat{F}$ is defined as the solution of

$$d\hat{F}(t,u)=\frac{[\int_t^\infty\int_u^\infty d\hat{F}(t',u')]d\hat{V}(t,u)}{\hat{H}(t,u)}.$$

EMPIRICAL VERSION OF THE P.D.E.

- For all function ψ,

$$\frac{1}{n}\sum_{i=1}^{n}\psi(Y,Z,\delta,\gamma)\to_{n\to\infty}E[\psi(Y,Z,\delta,\gamma)|Y\geq\mu,Z\geq\nu]\ p.s.$$

- Estimate $H(t,u)=\mathbb{P}(Y\geq t,Z\geq u|Y\geq\mu,Z\geq\nu)$ par

$$\hat{H}(t,u)=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{Y_i\geq t,Z_i\geq u}.$$

- For $V(t,u)=\mathbb{P}\left(\mu\leq t\leq Y,\nu\leq u\leq Z,\delta=\gamma=1|Y\geq\mu,Z\geq\nu\right)$,

$$\hat{V}(t,u)=\frac{1}{n}\sum_{i=1}^{n}\delta_i\gamma_i\mathbf{1}_{\mu_i\leq t\leq Y_i,\nu_i\leq u\leq Z_i}.$$

- $\hat{F}$ is defined as the solution of

$$d\hat{F}(t,u)=\frac{[\int_t^{\infty}\int_u^{\infty}d\hat{F}(t',u')]d\hat{V}(t,u)}{\hat{H}(t,u)}.$$

EMPIRICAL VERSION OF THE P.D.E.

- For all function ψ,

$$\frac{1}{n}\sum_{i=1}^{n}\psi(Y,Z,\delta,\gamma) \to_{n\to\infty} E[\psi(Y,Z,\delta,\gamma)|Y\geq\mu,Z\geq\nu]\ p.s.$$

- Estimate $H(t,u)=\mathbb{P}(Y\geq t,Z\geq u|Y\geq\mu,Z\geq\nu)$ par

$$\hat{H}(t,u)=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{Y_i\geq t,Z_i\geq u}.$$

- For $V(t,u)=\mathbb{P}\left(\mu\leq t\leq Y,\nu\leq u\leq Z,\delta=\gamma=1|Y\geq\mu,Z\geq\nu\right)$,

$$\hat{V}(t,u)=\frac{1}{n}\sum_{i=1}^{n}\delta_i\gamma_i\mathbf{1}_{\mu_i\leq t\leq Y_i,\nu_i\leq u\leq Z_i}.$$

- $\hat{F}$ is defined as the solution of

$$d\hat{F}(t,u)=\frac{[\int_t^\infty\int_u^\infty d\hat{F}(t',u')]d\hat{V}(t,u)}{\hat{H}(t,u)}.$$

EMPIRICAL VERSION OF THE P.D.E.

- We are back to a finite dimensional problem :

$$\hat{F} \leftrightarrow \mathbf{W} = (W_{1,n}, W_{2,n}, \ldots W_{n,n}, W_{\infty,n}),$$

where $W_{i,n}$ is the weight assigned to observation i where $W_{\infty,n}$ represents some residual mass.

PROPOSITION

There exists an unique solution $\mathbf{W}$ of the previous equation satisfying $\sum W_{i,n} = 1$. Moreover, all the weights are positive.

SOLVING THE EQUATION

- We need to solve

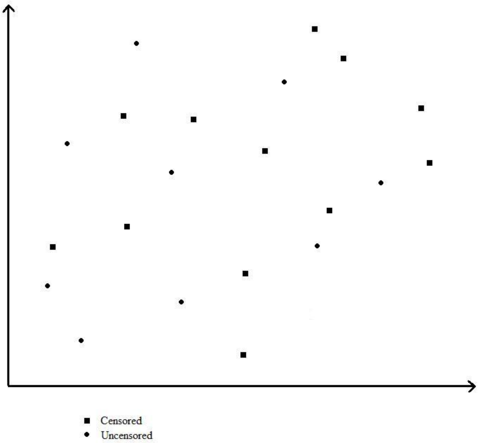
$$\mathbf{W} = A\mathbf{W}.$$

- Observe (after sorting the observations with respect to increasing values of Y)

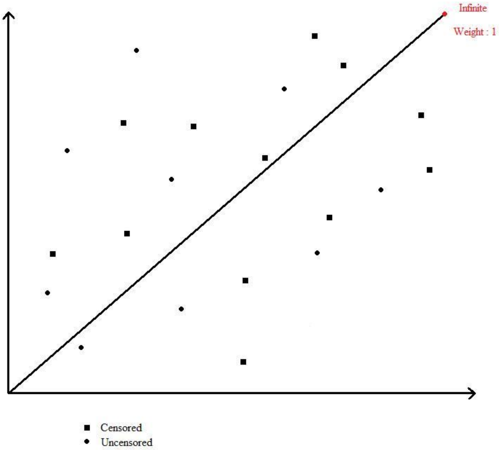
$$A = \begin{pmatrix} 0 & * & * & 0 \\ \vdots & \ddots & * & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

- From that, one deduces unicity of the solution.

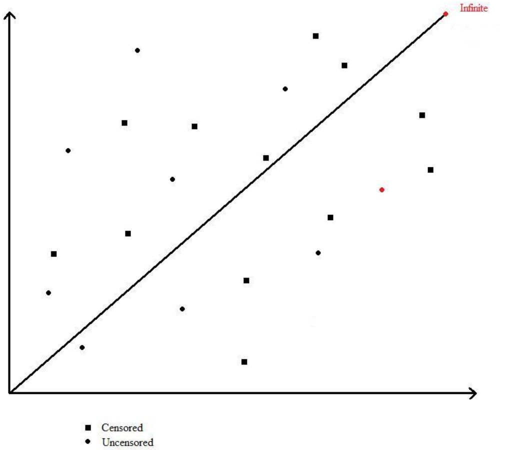
SOLVING THE EQUATION - GEOMETRICAL POINT OF VIEW



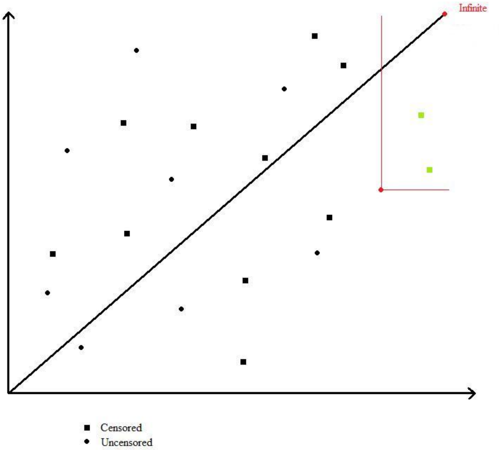
SOLVING THE EQUATION - GEOMETRICAL POINT OF VIEW



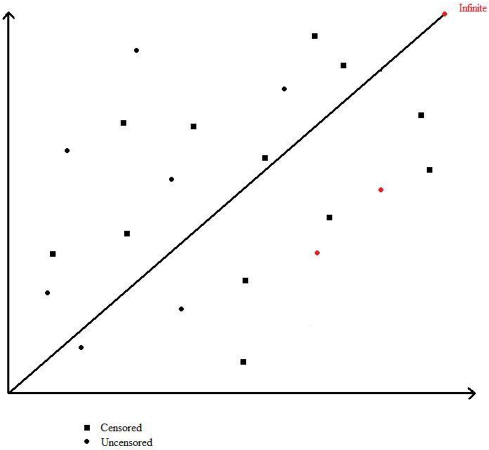
SOLVING THE EQUATION - GEOMETRICAL POINT OF VIEW



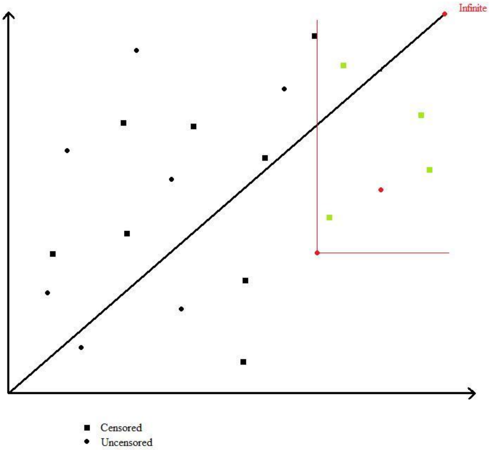
Solving the equation - geometrical point of view



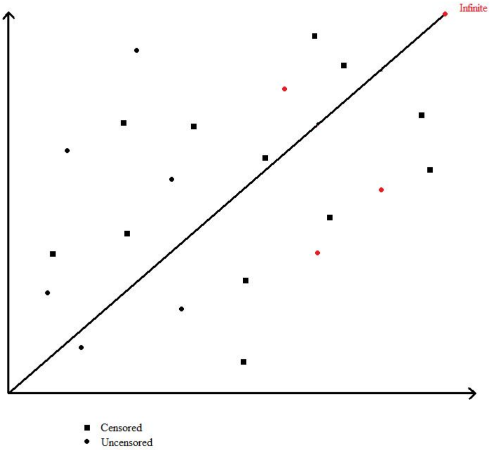
SOLVING THE EQUATION - GEOMETRICAL POINT OF VIEW



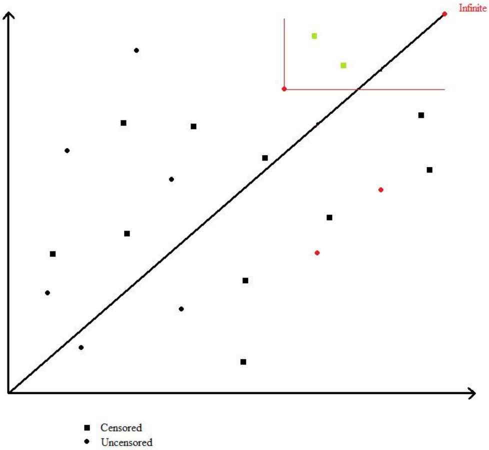
Solving the equation - geometrical point of view



SOLVING THE EQUATION - GEOMETRICAL POINT OF VIEW



SOLVING THE EQUATION - GEOMETRICAL POINT OF VIEW



EQUIVALENT EXPRESSION OF THE P.D.E.

- Define

$$W^*_{i,n} = \frac{\delta_i \gamma_i}{n} \frac{\int_{Y_i}^{\infty} \int_{Z_i}^{\infty} dF(t,u)}{H(t,u)}.$$

- Second expression : for all function ϕ,

$$E\left[nW^*_{i,n}\phi(Y_i, Z_i) | Y_i \geq \mu_i, Z_i \geq \nu_i\right] = E[\phi(T_i, U_i)].$$

- Define $F^*(t,u) = \sum_{i=1}^{n} W^*_{i,n} \mathbf{1}_{Y_i \leq t, Z_i \leq u}$, then, for all ϕ,

$$\int \phi(t,u) dF^*(t,u) = \sum_{i=1}^{n} W^*_{i,n} \phi(Y_i, Z_i) \rightarrow_{p.s.} E[\phi(T,U)].$$

- One can show that $W_{i,n}$ and $W^*_{i,n}$ are close.

EQUIVALENT EXPRESSION OF THE P.D.E.

- Define

$$W^*_{i,n} = \frac{\delta_i \gamma_i}{n} \frac{\int_{Y_i}^{\infty} \int_{Z_i}^{\infty} dF(t,u)}{H(t,u)}.$$

- Second expression : for all function ϕ,

$$E\left[nW^*_{i,n}\phi(Y_i, Z_i)|Y_i \geq \mu_i, Z_i \geq \nu_i\right] = E[\phi(T_i, U_i)].$$

- Define $F^*(t,u) = \sum_{i=1}^n W^*_{i,n} \mathbf{1}_{Y_i \leq t, Z_i \leq u}$, then, for all ϕ,

$$\int \phi(t,u) dF^*(t,u) = \sum_{i=1}^n W^*_{i,n}\phi(Y_i, Z_i) \rightarrow_{p.s.} E[\phi(T,U)].$$

- One can show that $W_{i,n}$ and $W^*_{i,n}$ are close.

EQUIVALENT EXPRESSION OF THE P.D.E.

- Define

$$W_{i,n}^* = \frac{\delta_i \gamma_i}{n} \frac{\int_{Y_i}^{\infty} \int_{Z_i}^{\infty} dF(t,u)}{H(t,u)}.$$

- Second expression : for all function ϕ,

$$E\left[nW_{i,n}^* \phi(Y_i, Z_i) | Y_i \geq \mu_i, Z_i \geq \nu_i \right] = E[\phi(T_i, U_i)].$$

- Define $F^*(t,u) = \sum_{i=1}^n W_{i,n}^* \mathbf{1}_{Y_i \leq t, Z_i \leq u}$, then, for all ϕ,

$$\int \phi(t,u) dF^*(t,u) = \sum_{i=1}^n W_{i,n}^* \phi(Y_i, Z_i) \rightarrow_{p.s.} E[\phi(T,U)].$$

- One can show that $W_{i,n}$ and $W_{i,n}^*$ are close.

Conditional version

- Assume that, conditionally on (X^1, X^2), (T, U) is independent from (C, D), and (μ, ν) is independent from (T, U, C, D). Moreover assume that all variables are continuous.

P.D.E. (similar to Lopez, 2012)

$F(t, u|x) = \mathbb{P}(T \le t, U \le u | X = x)$ is solution of

$$dF(t, u|x) = \frac{[\int_{t^+}^{\infty} \int_{u^+}^{\infty} dF(t', u'|x)] dV(t, u|x)}{H(t, u|x)},$$

where

$$H(t, u|x) = \mathbb{P}(Y \ge t, Z \ge u | Y \ge \mu, Z \ge \nu, X = x),$$
$$V(t, u|x) = \mathbb{P}\left(Y \le t, Z \le u, \delta\gamma = 1 | Y \ge \mu, Z \ge \nu, X = x\right).$$

Estimation of H (estimation of V is similar)

- Let K denote a bivariate probability density with compact support.
- Define, for $x = (x^1, x^2)$, and $X_i = (X_i^1, X_i^2)$,

$$w_{i,n}(x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{X_j - x}{h}\right)},$$

  where h is a bandwidth tending to 0 as n tends to infinity.
- Define

$$\hat{H}(t, u|x) = \sum_{i=1}^{n} w_{i,n}(x) \mathbf{1}_{Y_i \geq t, Z_i \geq u}.$$

- The bias of $\hat{H}$ is of order h^2, the variance of order $n^{-1/2}h^{-1}$.

Computation of an estimator $\hat{F}(\cdot|x)$

- The computation is similar to the case where no covariate is present.
- $\hat{F}(\cdot,\cdot|x) \leftrightarrow \mathbf{W}(x) = (W_{1,n}(x), ..., W_{n,n}(x), W_{\infty}(x))$ is the solution of an equation of the type

$$\mathbf{W} = A(x)\mathbf{W},$$

  where the matrix A now depends on x.
- The convergence rate is modified ($h^2 + n^{-1/2}h^{-p/2}$ rather than $n^{-1/2}$, where p is the dimension of X).

Estimation of $\theta(x)$

- Recall that

$$\hat{F}(t,u|x)=\sum_{i=1}^{n}W_{i,n}(x)\mathbf{1}_{Y_i\leq t, X_i\leq x}.$$

- Estimation of $\theta(x)$:

$$\hat{\theta}(x)=\arg\max_{\theta\in\Theta}\sum_{i=1}^{n}W_{i,n}(x)\log(\mathfrak{c}_\theta(\hat{F}_T(Y_i)),\hat{F}_U(U_i)).$$

ASYMPTOTIC THEORY

- Assumptions :
  - the copula function stays in the same copula family when x changes
  - $\sup_{t,u,x} |\partial_x H(t,u|x)| + |\partial_x^2 H(t,u|x)| < \infty$
  - conditions on the tail of the distributions of (T, U)
  - standard conditions on the kernel

ASYMPTOTIC THEORY

Under appropriate assumptions,

$$\sup_{x\in\mathcal{X}} \|\hat{\theta}(x) - \theta(x)\| = O_{a.s.}\left(h^2 + \frac{1}{n^{1/2}h^{p/2}}\right),$$

where p is the dimension of x, and where $\mathcal{X}$ is a compact on which the density of X is bounded away from zero.

CONCLUSION, PERSPECTIVES

- Practical choice of the bandwidth.
- Dimension of X : dimension reduction assumption to compensate the poor convergence rate when p is high.
- Goodness-of-fit, testing the assumption on the copula family...