An Improved nonparametric estimator of distribution function for bivariate competing risks model

Advisor: Takeshi Emura（江村剛志）教授
Presenter: Fan-Hsuan Kao
NCU

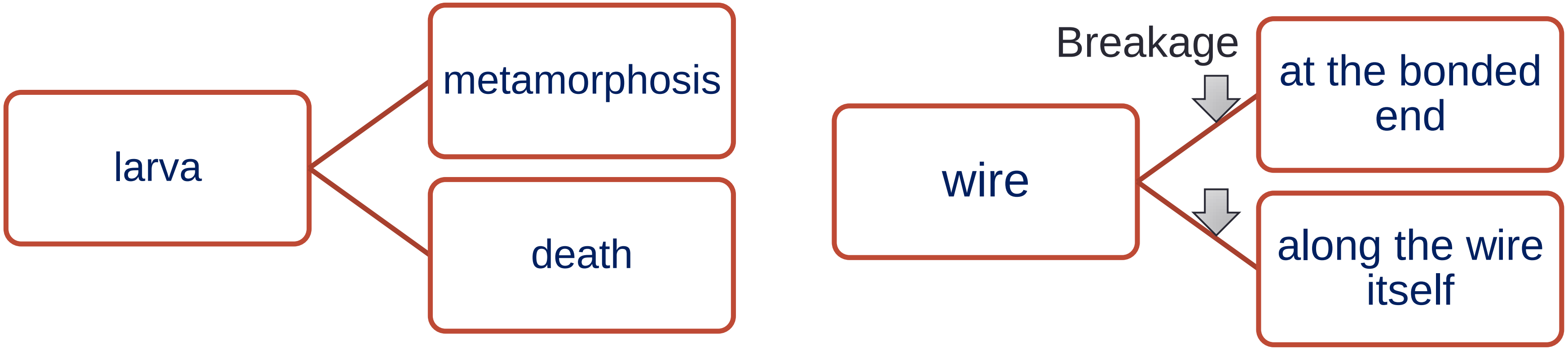
2013/05/30

Outline

- Introduction
- Method
- Asymptotic theory
- Simulation
- Data Analysis
- Conclusion

Introduction

- In the lifetime experiments, the failure causes of individual may be classified into more than one disjoint class.



- So-called competing risks model are used to analyze such data.

- Two frameworks are used to deal with standard competing risks settings.

- Let continuous failure time $T > 0$ and a cause of failure $C \in \{1, 2, ..., \gamma\}$ can be observed for an individual:

- (1) Cause-specific hazard function

$$\lambda_i(t) = \lim_{\Delta t \to 0} \frac{\Pr(T < t + \Delta t, C = i \mid T \geq t)}{\Delta t}, i = 1, 2, ..., \gamma.$$

- (2) Cause-specific distribution function

$$F_i(t) = \Pr(T \leq t, \ C = i), i = 1, 2, ..., \gamma.$$

- In biological application, bivariate competing risks arise naturally.

- For example, the occurrence of blindness in the left and right eyes, a genetic study examining the age at death of parents and children.

- In the reality, it is hard to get bivariate failure times and failure causes for all individuals. One usually encounter truncation and censoring.

- Cohort follow-up studies : Dementia →Doubly censored case (i.e., both left censoring and right censoring together) **(Shen, 2011).**

- Sankaran et al. (2006) proposed an estimator under random right censoring.

Object:

- We consider a new nonparametric estimator which improves upon the estimator of Sankaran et al (2006) under the same situation.

Method-Estimator of Sankaran et al. (2006)

- We describe a nonparametric method which is proposed by Sankaran et al. (2006).

- Let $T = (T_1, T_2)$ be a pair of non-negative random variables.
- Let $Z = (Z_1, Z_2)$ be a pair of random censoring times.

- Under the bivariate right censoring, observable variables are $Y = (Y_1, Y_2)$ and $\delta = (\delta_1, \delta_2)$.

$$Y_k = \min(T_k, Z_k)\begin{cases} T_k \Rightarrow \delta_k = 1 \\ Z_k \Rightarrow \delta_k = 0 \end{cases}, \quad k = 1, 2 \quad .$$

- Let $S(t_1, t_2) = \Pr(T_1 > t_1, T_2 > t_2)$ be the joint survivor functions of T.
- Let $G(t_1, t_2) = \Pr(Z_1 > t_1, Z_2 > t_2)$ be the survivor function of censoring times Z
- $H(t_1, t_2) = \Pr(Y_1 > t_1, Y_2 > t_2)$ be the survivor function of observable variables Y.
- Assume that T and Z are independent, then we have

$$H(t_1, t_2) = G(t_1, t_2) S(t_1, t_2).$$

- In the bivariate competing risks model, the cause-specific distribution function is

$$F_{ij}(t_1, t_2) = \Pr(T_1 \le t_1, T_2 \le t_2, C_1 = i, C_2 = j), \quad i = 1, 2, \ldots, \gamma_1, \; j = 1, 2, \ldots, \gamma_2,$$

where $C = (C_1, C_2)$ is a pair of causes corresponding to $T = (T_1, T_2)$, and suppose that are $\gamma_1 \times \gamma_2$ different failure for (T_1, T_2).

.

- Let

$$F_{ij}^{*}(t_1, t_2) = \Pr(T_1 \le t_1, T_2 \le t_2, \delta_1 = 1, \delta_2 = 1, C_1 = i, C_2 = j)$$

- Sankaran et al. (2006) derive a formula which is

$$F_{ij}(t_1, t_2) = \int_0^{t_1}\int_0^{t_2} \frac{S(u^-, v^-) F_{ij}^{*}(du, dv)}{H(u^-, v^-)}, \quad i = 1, 2, \ldots, \gamma_1, \; j = 1, 2, \ldots, \gamma_2 \, .$$

- The observations are $(Y_{1u}, Y_{2u}, \delta_{1u}, \delta_{2u}, C_{1u}, C_{2u})$ for $u = 1, 2, \ldots, n$.

- Define

$$\hat{H}(t_1, t_2) = \frac{1}{n}\sum_{u=1}^{n}\mathbf{I}(Y_{1u} > t_1, Y_{2u} > t_2).$$

- An unbiased estimate of $F_{ij}^{*}(t_1, t_2)$ is given by

$$\hat{F}_{ij}^{*}(t_1, t_2) = \frac{1}{n}\sum_{u=1}^{n}\mathbf{I}(Y_{1u} \le t_1, Y_{2u} \le t_2, \delta_{1u} = 1, \delta_{2u} = 1, C_{1u} = i, C_{2u} = j).$$

- Since $\hat{H}(t_1, t_2) > 0$, we can get the estimator

$$\hat{F}_{ij}(t_1, t_2) = \int_0^{t_1}\int_0^{t_2} \frac{\hat{S}(u^-, v^-)\hat{F}^*_{ij}(du, dv)}{\hat{H}(u^-, v^-)}, \quad i = 1, 2, \ldots, \gamma_1, \; j = 1, 2, \ldots, \gamma_2.$$

- Sankaran et al. (2006) propose to apply the Dabrowska's nonparametric estimator (1988) of the $\hat{S}(t_1, t_2)$.

- $\hat{F}_{ij}(t_1, t_2) \xrightarrow{a.s} F(t_1, t_2)$.

- The estimator $\hat{F}_{ij}(t_1, t_2)$ is weak convergence.
(Sankaran et al, 2006).

Method-Independence Estimator

- If (T_1, C_1, Z_1) and (T_2, C_2, Z_2) is independent, it is easy to show

$$F_{ij}(t_1, t_2) = \int_0^{t_1} \frac{S_1(u-)}{H_1(u-)} F_{1i}^*(du) \int_0^{t_2} \frac{S_2(v-)}{H_2(v-)} F_{2j}^*(dv), \quad i = 1, 2, \ldots, \gamma_1,\ j = 1, 2, \ldots, \gamma_2 .$$

where $S_k(u) = \Pr(T_k > u)$, $H_k(u) = \Pr(Z_k > u)$ for $k = 1, 2$.

- The "independence estimator" of $F_{ij}(t_1, t_2)$ defined as

$$\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2),$$

where $\hat{F}_{1i}(t_1)$ and $\hat{F}_{2j}(t_2)$ are the univariate estimators of the form

$$\hat{F}_{1i}(t_1) = \int_0^{t_1} \frac{\hat{S}_1(u^-)}{\hat{H}_1(u^-)} F_{1i}^*(du), \quad i = 1, 2, \ldots, \gamma_1 .$$

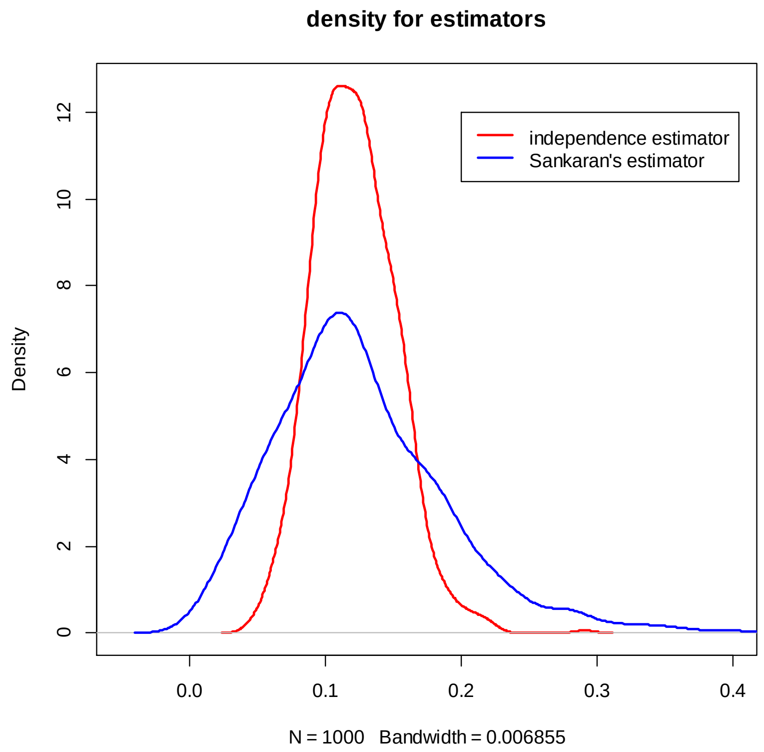
$$\hat{F}_{2j}(t_2) = \int_0^{t_2} \frac{\hat{S}_2(v^-)}{\hat{H}_2(v^-)} F_{2j}^*(dv), \quad j = 1, 2, \ldots, \gamma_2 .$$

and

$$\hat{F}_{1i}^*(t_1) = \frac{1}{n}\sum_{u=1}^{n} \mathbf{I}(Y_{1u} \le t_1, \delta_{1u} = 1, C_{1u} = i), \quad \hat{H}_k(t^-) = \frac{1}{n}\sum_{u=1}^{n} \mathbf{I}(Y_{ku} \ge t), \; k = 1, 2,$$

$$\hat{F}_{2j}^*(t_2) = \frac{1}{n}\sum_{u=1}^{n} \mathbf{I}(Y_{2u} \le t_2, \delta_{2u} = 1, C_{2u} = j).$$

- This estimator is consistent only under the independent assumption.
- The independence estimator reduce the variance and it is much easier to compute.
- We take advantage of the small variance to develop a new estimator.



Method-Proposed estimator

- The idea of improves the estimator is combined the existing estimator and independence estimator as follow:

$$\hat{F}_{ij}^{a}(t_1,t_2)=a\hat{F}_{ij}(t_1,t_2)+(1-a)\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2),\quad a\in[0,1].$$

$$\hat{F}_{ij}^{a}(t_1,t_2)\begin{cases}a=0 & \Longrightarrow \quad \hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2)\\ a=1 & \Longrightarrow \quad \hat{F}_{ij}(t_1,t_2)\end{cases}$$

- How to choose the optimal value of a ?

- We usually use MSE to evaluate different estimators, which considers bias and variance.

- The approach method is similar with Akritas and Keilegom (2003).

- Let $F_{ij}(t_1, t_2)$ be the true value of distribution function. Define

$$\begin{aligned}&\mathrm{MSE}[\hat{F}_{ij}^{a}(t_1,t_2)]\\&=E[\hat{F}_{ij}^{a}(t_1,t_2)-F_{ij}(t_1,t_2)]^2\\&=E\{a[\hat{F}_{ij}(t_1,t_2)-F_{ij}(t_1,t_2)]+(1-a)[\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2)-F_{ij}(t_1,t_2)]\}^2.\end{aligned}$$

- It is easy to see that

$$\begin{aligned}&\frac{d}{da}\mathrm{MSE}[\hat{F}_{ij}^{a}(t_1,t_2)]\\&=2a\times E[\hat{F}_{ij}(t_1,t_2)-F_{ij}(t_1,t_2)]^2-2(1-a)E[\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2)-F_{ij}(t_1,t_2)]^2\\&\quad+2(1-2a)E\{[\hat{F}_{ij}(t_1,t_2)-F_{ij}(t_1,t_2)][\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2)-F_{ij}(t_1,t_2)]\},\end{aligned}$$

$$\frac{d^2}{da^2}\mathrm{MSE}[\hat{F}_{ij}^{a}(t_1,t_2)]>0.$$

- Define

$$x(t_1, t_2) = E[\hat{F}_{ij}(t_1, t_2) - F_{ij}(t_1, t_2)]^2,$$

$$y(t_1, t_2) = E[\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) - F_{ij}(t_1, t_2)]^2,$$

$$z(t_1, t_2) = E\{[\hat{F}_{ij}(t_1, t_2) - F_{ij}(t_1, t_2)][\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) - F_{ij}(t_1, t_2)]\}.$$

- Then,

$$a^*(t_1, t_2) = \arg\min_a \mathrm{MSE}[\hat{F}_{ij}^a(t_1, t_2)] = \frac{y(t_1, t_2) - z(t_1, t_2)}{x(t_1, t_2) + y(t_1, t_2) - 2z(t_1, t_2)}.$$

Theorem 1

- *We assume the following conditions:*

$$A.\ x(t_1, t_2) \neq z(t_1, t_2),$$
$$B.\ y(t_1, t_2) \neq z(t_1, t_2),$$
$$C.\ x(t_1, t_2) + y(t_1, t_2) - 2z(t_1, t_2) > 0.$$

Then

$$MSE[\hat{F}_{ij}^{a^*}(t_1, t_2)] < MSE[\hat{F}_{ij}(t_1, t_2)],$$
$$MSE[\hat{F}_{ij}^{a^*}(t_1, t_2)] < MSE[\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2)].$$

- ***Proof***

Take $a^*(t_1, t_2)$ into the formula of $\text{MSE}[\hat{F}_{ij}^{a}(t_1, t_2)]$, we get

$$
\begin{aligned}
&\text{MSE}[\hat{F}_{ij}^{a^*}(t_1, t_2)] \\
&= E\{a^*[\hat{F}_{ij}(t_1, t_2) - F_{ij}(t_1, t_2)] + (1-a^*)[\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) - F_{ij}(t_1, t_2)]\}^2 \\
&= a^{*2} E[\hat{F}_{ij}(t_1, t_2) - F_{ij}(t_1, t_2)]^2 + (1-a^*)^2 E[\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) - F_{ij}(t_1, t_2)]^2 \\
&\quad + 2a^*(1-a^*)E\{[\hat{F}_{ij}(t_1, t_2) - F_{ij}(t_1, t_2)][\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) - F_{ij}(t_1, t_2)]\} \\
&= \left(\frac{y-z}{x+y-2z}\right)^2 x + \left(1 - \frac{y-z}{x+y-2z}\right)^2 y + 2\left(\frac{y-z}{x+y-2z}\right)\left(1 - \frac{y-z}{x+y-2z}\right) z \\
&= \frac{xy - z^2}{x+y-2z}.
\end{aligned}
$$

- Under conditions A and B, C, we can get the desired results

$$\mathrm{MSE}[\hat{F}_{ij}^{a^*}(t_1,t_2)] < \mathrm{MSE}[\hat{F}_{ij}(t_1,t_2)] \Leftrightarrow \frac{xy-z^2}{x+y-2z} < y \Leftrightarrow (y-z)^2 > 0,$$

$$\mathrm{MSE}[\hat{F}_{ij}^{a^*}(t_1,t_2)] < \mathrm{MSE}[\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2)] \Leftrightarrow \frac{xy-z^2}{x+y-2z} < x \Leftrightarrow (x-z)^2 > 0. \quad \square$$

- Since the values $x(t_1,t_2)$, $y(t_1,t_2)$ and $z(t_1,t_2)$ are the expectations of three very different variable, the conditions are usually holds.

- Real data $\xrightarrow{?}$ $a^*(t_1, t_2)$

- We use the nonparametric bootstrap method to get $\hat{x}(t_1, t_2)$, $\hat{y}(t_1, t_2)$ and $\hat{z}(t_1, t_2)$.

- ***The bootstrap method for estimating*** $a^*(t_1, t_2)$
- **1.** Calculate $\hat{F}_{ij}(t_1, t_2)$ from the real data.
- **2.** Let

$$\{(Y_{1,u}^{*(b)}, Y_{2,u}^{*(b)}, \delta_{1,u}^{*(b)}, \delta_{2,u}^{*(b)}, C_{1,u}^{*(b)}, C_{2,u}^{*(b)}) : u = 1, 2, \ldots, n)\}$$

be a random sample with replacement from the data

$$\{(Y_{1,u}, Y_{2,u}, \delta_{1,u}, \delta_{2,u}, C_{1,u}, C_{2,u}) : u = 1, 2, \ldots, n\}$$

for $b = 1, 2, \ldots, B$, where B is the bootstrap number.

- **3.** Calculate the distribution function $\hat{F}_{ij}^{*(b)}(t_1, t_2)$ and $\hat{F}_{1i}^{*(b)}(t_1)\hat{F}_{2j}^{*(b)}(t_2)$ based on the resampled data , then we compute

$$\hat{x}(t_1, t_2) = \frac{1}{B}\sum_{b=1}^{B}[\hat{F}_{ij}^{*(b)}(t_1, t_2) - \hat{F}_{ij}(t_1, t_2)]^2,$$

$$\hat{y}(t_1, t_2) = \frac{1}{B}\sum_{b=1}^{B}[\hat{F}_{1i}^{*(b)}(t_1)\hat{F}_{2j}^{*(b)}(t_2) - \hat{F}_{ij}(t_1, t_2)]^2,$$

$$\hat{z}(t_1, t_2) = \frac{1}{B}\sum_{b=1}^{B}[\hat{F}_{1i}^{*(b)}(t_1)\hat{F}_{2j}^{*(b)}(t_2) - \hat{F}_{ij}(t_1, t_2)][\hat{F}_{ij}^{*(b)}(t_1, t_2) - \hat{F}_{ij}(t_1, t_2)].$$

Then, we obtain the estimator

$$\hat{a}(t_1, t_2) = \frac{\hat{y}(t_1, t_2) - \hat{z}(t_1, t_2)}{\hat{x}(t_1, t_2) + \hat{y}(t_1, t_2) - 2\hat{z}(t_1, t_2)}.$$

Asymptotic theory

- We prove the consistency of proposed estimator,

$$\hat{F}_{ij}^{\hat{a}}(t_1,t_2)=\hat{a}\,\hat{F}_{ij}(t_1,t_2)+(1-\hat{a})\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2).$$

- We consider two cases,

(i) T_1 and T_2 are independent,

$$\Longrightarrow \quad \hat{F}_{1i}(t_1)\hat{F}_{2i}(t_2)\xrightarrow{P}F_{1i}(t_1)F_{2j}(t_2)=F_{ij}(t_1,t_2)$$

$$\Longrightarrow \quad \hat{F}_{ij}(t_1,t_2)\xrightarrow{P}F(t_1,t_2)$$

(ii) T_1 and T_2 are not independent,

$$\Longrightarrow \quad n\to\infty,\quad \hat{a}(t_1,t_2)\xrightarrow{P}a^*(t_1,t_2)\to 1$$

$$\Longrightarrow \quad \hat{F}_{ij}^{\hat{a}}(t_1,t_2)\to\hat{F}_{ij}(t_1,t_2)\xrightarrow{P}F_{ij}(t_1,t_2)$$

- **Lemma 1** **(Theorem Ⅳ.4.1, Andersen et al., 1993)**

  For fixed t_k *with* $G_k(t_k) > 0$, *where* $G_k(t) = \Pr(Z_k > t)$. *Then, as* $n \to \infty$,

$$\hat{F}_{k,j}(t_k) \xrightarrow{P} F_{k,j}(t_k) \quad \textit{for} \ k = 1, 2 \ .$$

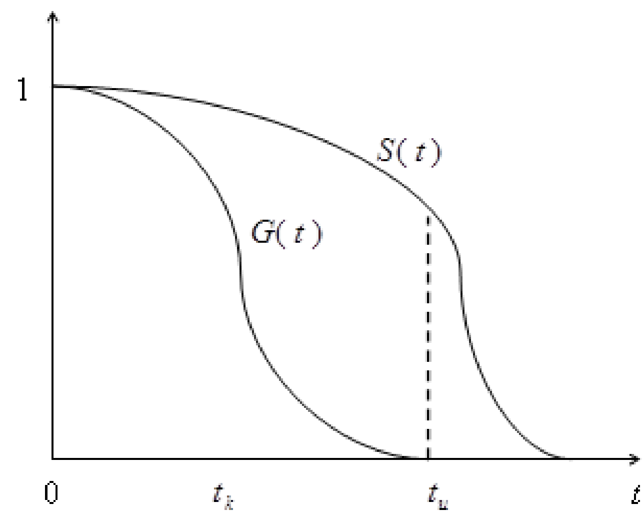


Figure 1 A survival function plot for the censored time and the failure time

- **Lemma 2**

For fixed (t_1, t_2) *with* $G(t_1, t_2) > 0$ *, as* $n \to \infty$,

$$\hat{F}_{1i}(t_1)\hat{F}_{2i}(t_2) \xrightarrow{P} F_{1i}(t_1)F_{2i}(t_2).$$

Theorem 2

Suppose that (t_1, t_2) *satisfies* $G(t_1, t_2) > 0$, T_1 *and* T_2 *are not independent, Then, as* $n \to \infty$,

$$a^*(t_1, t_2) \to 1 .$$

- ***Proof***

We know

$$\hat{F}_{ij}(t_1, t_2) \xrightarrow{P} F_{ij}(t_1, t_2),$$

$$\hat{F}_{1i}(t_1)\hat{F}_{2i}(t_2) \xrightarrow{P} F_{1i}(t_1)F_{2j}(t_2).$$

Since these two estimators are uniformly bounded sequence,

$$\hat{F}_{ij}(t_1, t_2) \xrightarrow{L_P} F(t_1, t_2),$$

$$\hat{F}_{1i}(t_1)\hat{F}_{2i}(t_2) \xrightarrow{L_P} F_{1i}(t_1)F_{2j}(t_2),\ 0 < p < \infty$$ (Chung, 2001).

To prove the Theorem 2, we need three claims.

- **Claim 1**

$$\lim_{n\to\infty} E[\hat{F}_{ij}(t_1, t_2) - F_{ij}(t_1, t_2)]^2 = 0$$

Proof

The proof follows by the definition of convergence in L_P at $p = 2.$ □

- **Claim 2**

$$\lim_{n\to\infty} E[\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) - F_{ij}(t_1,t_2)]^2 = [F_{1i}(t_1)F_{2j}(t_2) - F_{ij}(t_1,t_2)]^2$$

Proof

Since $\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) \xrightarrow{L_P} F_i(t_1)F_j(t_2)$, $0 < p < \infty$, it follows

$p = 2$ ⇒ $\lim_{n\to\infty} E[\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) - F_{1i}(t_1)F_{2j}(t_2)]^2 = 0,$

$p = 1$ ⇒ $\lim_{n\to\infty} E[\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) - F_{1i}(t_1)F_{2j}(t_2)] = 0,$

then we can get

⇒ $\lim_{n\to\infty} E[\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) - F_{ij}(t_1,t_2)]^2 = [F_{1i}(t_1)F_{2j}(t_2) - F_{ij}(t_1,t_2)]^2$

- **Claim 3**

$$\lim_{n\to\infty} E\{[\hat{F}_{ij}(t_1,t_2)-F_{ij}(t_1,t_2)][\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2)-F_{ij}(t_1,t_2)]\}=0$$

Proof

- Since $\hat{F}_{ij}(t_1,t_2)\xrightarrow{P}F_{ij}(t_1,t_2)$ and $\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2)\xrightarrow{P}F_i(t_1)F_j(t_2)$,

$\Rightarrow$ $\hat{F}_{ij}(t_1,t_2)\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2)\xrightarrow{P}F_{ij}(t_1,t_2)F_{1i}(t_1)F_{2j}(t_2)$ (Slutsky's theorem).

Since $\{\hat{F}_{ij}(t_1,t_2)\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2)\}$ is also uniformly bounded

$\Rightarrow$ $\hat{F}_{ij}(t_1,t_2)\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2)\xrightarrow{L_P}F_{ij}(t_1,t_2)F_{1i}(t_1)F_{2j}(t_2)$, $0<p<\infty$.

$\Rightarrow$ $\lim_{n\to\infty} E\{[\hat{F}_{ij}(t_1,t_2)-F_{ij}(t_1,t_2)][\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2)-F_{ij}(t_1,t_2)]\}=0.$

- Finally, we use Claims 1~3 and the definition of $a^*(t_1, t_2)$ to conclude.

$$\lim_{n\to\infty} x(t_1,t_2) = \lim_{n\to\infty} E[\hat{F}_{ij}(t_1,t_2) - F_{ij}(t_1,t_2)]^2 = 0,$$

$$\lim_{n\to\infty} y(t_1,t_2) = \lim_{n\to\infty} E[\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) - F_{ij}(t_1,t_2)]^2 = [F_{1i}(t_1)F_{2j}(t_2) - F_{ij}(t_1,t_2)]^2,$$

$$\lim_{n\to\infty} z(t_1,t_2) = \lim_{n\to\infty} E[\hat{F}_{ij}(t_1,t_2) - F_{ij}(t_1,t_2)][\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) - F_{ij}(t_1,t_2)] = 0,$$

$$\begin{aligned}\lim_{n\to\infty} a^*(t_1,t_2) &= \lim_{n\to\infty} \frac{y(t_1,t_2) - z(t_1,t_2)}{x(t_1,t_2) + y(t_1,t_2) - 2z(t_1,t_2)} \\ &= \frac{\lim_{n\to\infty} y(t_1,t_2) - \lim_{n\to\infty} z(t_1,t_2)}{\lim_{n\to\infty} x(t_1,t_2) + \lim_{n\to\infty} y(t_1,t_2) - \lim_{n\to\infty} 2z(t_1,t_2)} \\ &= 1.\end{aligned}$$

The proof of Theorem 2 is completed. □

- **Conjecture 1**
- *Suppose that* (t_1, t_2) *satisfies* $G(t_1, t_2) > 0$. *Then, as* $n \to \infty$,

$$\hat{a}(t_1, t_2) \xrightarrow{P} a^*(t_1, t_2).$$

- The results should follow to Dabrowska's bootstrap convergence results (Corollary 2.2 of Dabrowska, 1989).

Theorem 3

Suppose that (t_1, t_2) *satisfies* $G(t_1, t_2) > 0$. *Then, as* $n \to \infty$

$$\hat{F}_{ij}^{\hat{a}}(t_1, t_2) = \hat{a}\,\hat{F}_{ij}(t_1, t_2) + (1-\hat{a})\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) \xrightarrow{P} F_{ij}(t_1, t_2) \; .$$

- ***Proof***
- To prove the theorem we consider two cases,

A. T_1 and T_2 are independent,

We know

$$\hat{F}_{ij}(t_1,t_2) \xrightarrow{P} F_{ij}(t_1,t_2),$$

$$\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) \xrightarrow{P} F_i(t_1)F_j(t_2) = F_{ij}(t_1,t_2),$$

then,

$$\hat{F}_{ij}^{\hat{a}}(t_1,t_2) = \hat{a}\,\hat{F}_{ij}(t_1,t_2) + (1-\hat{a})\hat{F}_{1i}(t_1)\hat{F}_{2j}(t_2) \xrightarrow{P} F_{ij}(t_1,t_2).$$

- B. T_1 and T_2 are not independent.

  We know

  $a^*(t_1, t_2) \to 1$ (Theorem 2),

  $\hat{a}(t_1, t_2) \xrightarrow{P} a^*(t_1, t_2)$ (Conjecture 1),

  $\Rightarrow$ $\hat{a}(t_1, t_2) \xrightarrow{P} 1$ $\Rightarrow$ $\hat{F}_{ij}^{\hat{a}}(t_1, t_2) \to \hat{F}_{ij}(t_1, t_2)$

Since $\hat{F}_{ij}(t_1, t_2) \xrightarrow{P} F_{ij}(t_1, t_2)$,

$$\hat{F}_{ij}^{\hat{a}}(t_1, t_2) \xrightarrow{P} F_{ij}(t_1, t_2).$$

Simulation

- 500 simulation times.
- For the dependent data, we consider the Clayton model (1978):

$$\Pr(T_1 \le t_1, T_2 \le t_2) = \max\left[\{ F_1(t_1)^{-(\alpha-1)} + F_2(t_2)^{-(\alpha-1)} - 1 \}^{\frac{-1}{\alpha-1}}, 0 \right], \quad \alpha \in [0, \infty) \setminus \{1\}.$$

- When $\alpha \in (1, \infty]$, T_1 and T_2 have positive correlation.

- We use "Kendall's tau (τ)" to calculate the correlation coefficient. For the Clayton model (1978) the formula is

$$\tau = \frac{\alpha - 1}{\alpha + 1}.$$

- We consider two distributions for T_1 and T_2 :
  (i) $T_1, T_2, Z_1, Z_2 \sim \exp(1)$,
  (ii) $T_1, T_2, Z_1, Z_2 \sim LN(0, 1)$.
- For these two distributions, we consider two cases,
  (i) T_1 and T_2 are independent,
  (ii) T_1 and T_2 are dependent by the Clayton model.
- Z_1 and Z_2 are independent.

- $C_k = \begin{cases} 1, \text{ with prob.} = 0.5 \\ 2, \text{ with prob.} = 0.5 \end{cases}, \quad k = 1, 2.$
- C_1 and C_2 are independent.
- The number of nonparametric bootstrap resample is taken to be B=500.

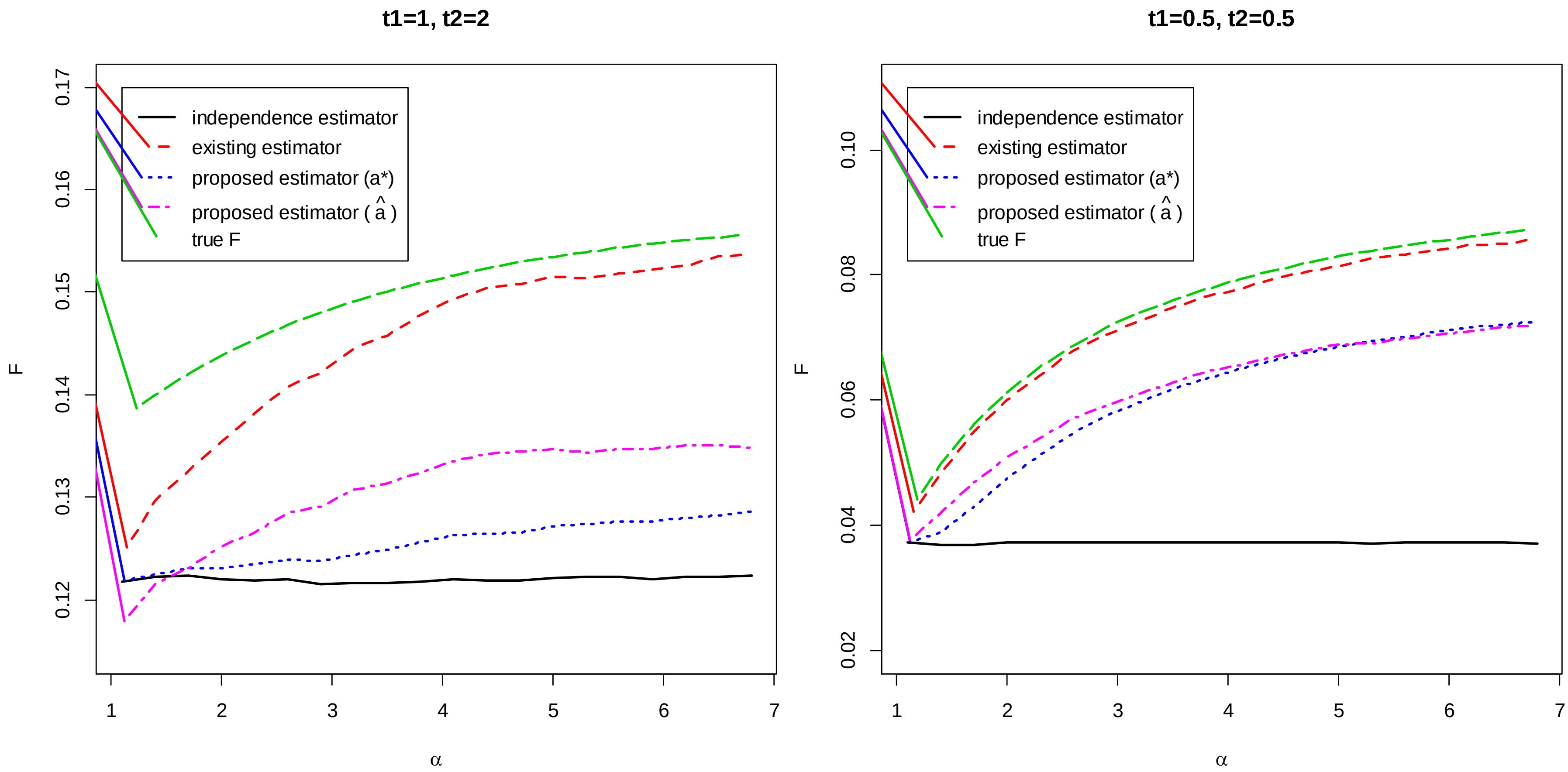
- **Table** Simulation results for four different estimators with n = 100. Data follow $T_1, T_2, Z_1, Z_2 \sim \exp(1)$ with the Clayton model. The corresponding pair of causes are (i, j) = (1, 1).

Independent case						
(t_1, t_2)	True value $F_{11}(t_1, t_2)$	Estimator	Estimated value F	MSE	a^*	$\hat{a}$
(1, 2)	0.1366	$\hat{F}_{11}(t_1, t_2)$	0.1290	0.004378		
		$\hat{F}_1(t_1)\hat{F}_1(t_2)$	0.1219	0.001132		
		$\hat{F}_{11}^{a^*}(t_1, t_2)$	0.1221	0.001134	0.0226	
		$\hat{F}_{11}^{\hat{a}}(t_1, t_2)$	0.1213	0.002766		0.3595
(0.5, 0.5)	0.0387	$\hat{F}_{11}(t_1, t_2)$	0.0386	0.000601		
		$\hat{F}_1(t_1)\hat{F}_1(t_2)$	0.0376	0.000164		
		$\hat{F}_{11}^{a^*}(t_1, t_2)$	0.0376	0.000164	0	
		$\hat{F}_{11}^{\hat{a}}(t_1, t_2)$	0.0360	0.000376		0.3880

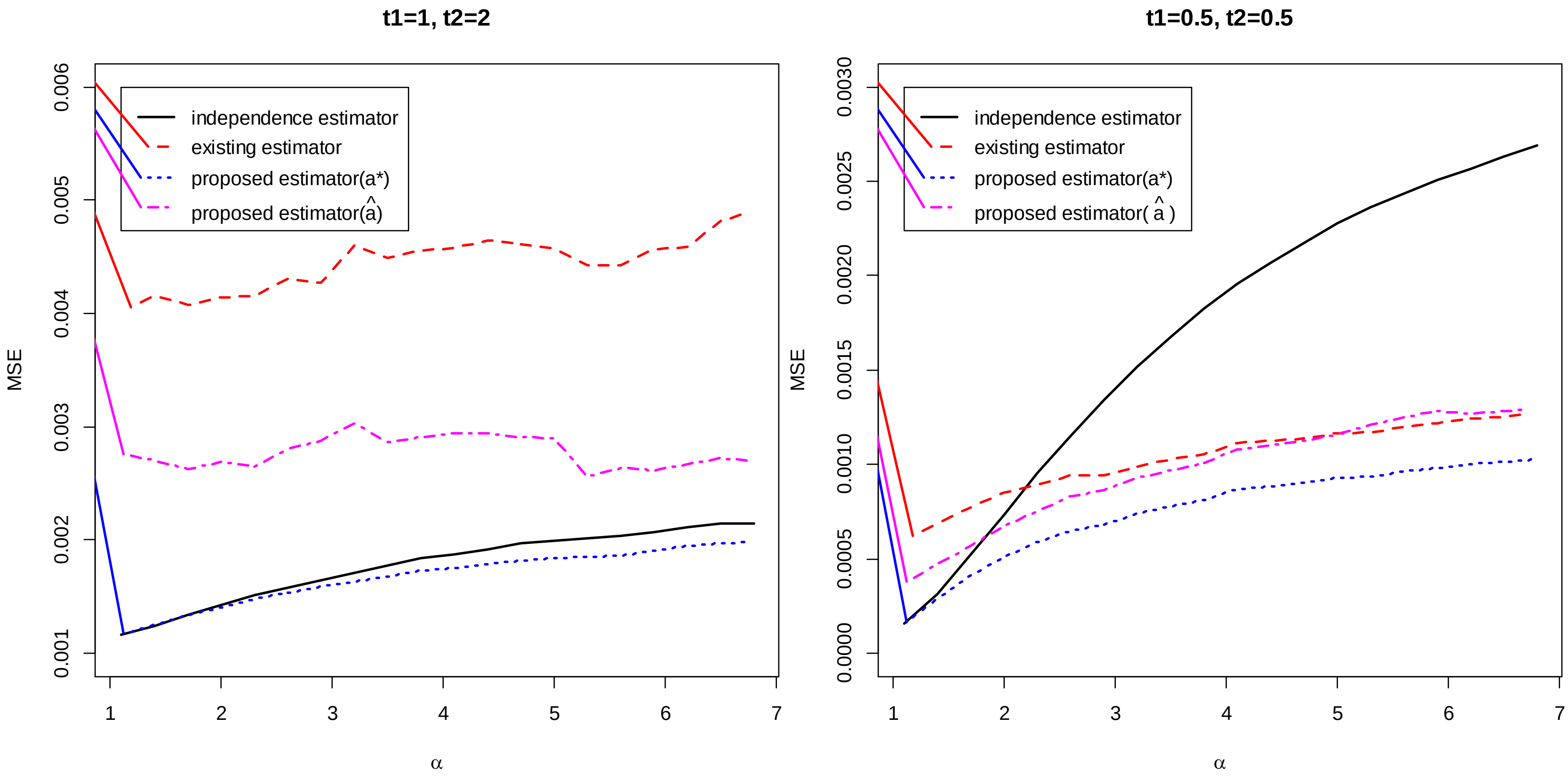
- **Table** Simulation results for four different estimators with n = 100. Data follow $T_1, T_2, Z_1, Z_2 \sim \exp(1)$ with the Clayton model. The corresponding pair of causes are (i, j) = (1, 1).

			Dependent case with $\alpha = 5$ $(\tau = 0.667)$			
(t_1, t_2)	True value $F_{11}(t_1, t_2)$	Estimator	Estimated value F	MSE	a^*	$\hat{a}$
(1, 2)	0.1534	$\hat{F}_{11}(t_1, t_2)$	0.1514	0.004574		
		$\hat{F}_1(t_1)\hat{F}_1(t_2)$	0.1221	0.001999		
		$\hat{F}_{11}^{a^*}(t_1, t_2)$	0.1274	0.001836	0.1805	
		$\hat{F}_{11}^{\hat{a}}(t_1, t_2)$	0.1348	0.002897		0.3717
(0.5, 0.5)	0.0830	$\hat{F}_{11}(t_1, t_2)$	0.0816	0.001166		
		$\hat{F}_1(t_1)\hat{F}_1(t_2)$	0.0372	0.002275		
		$\hat{F}_{11}^{a^*}(t_1, t_2)$	0.0675	0.000931	0.6829	
		$\hat{F}_{11}^{\hat{a}}(t_1, t_2)$	0.0688	0.001150		0.6037

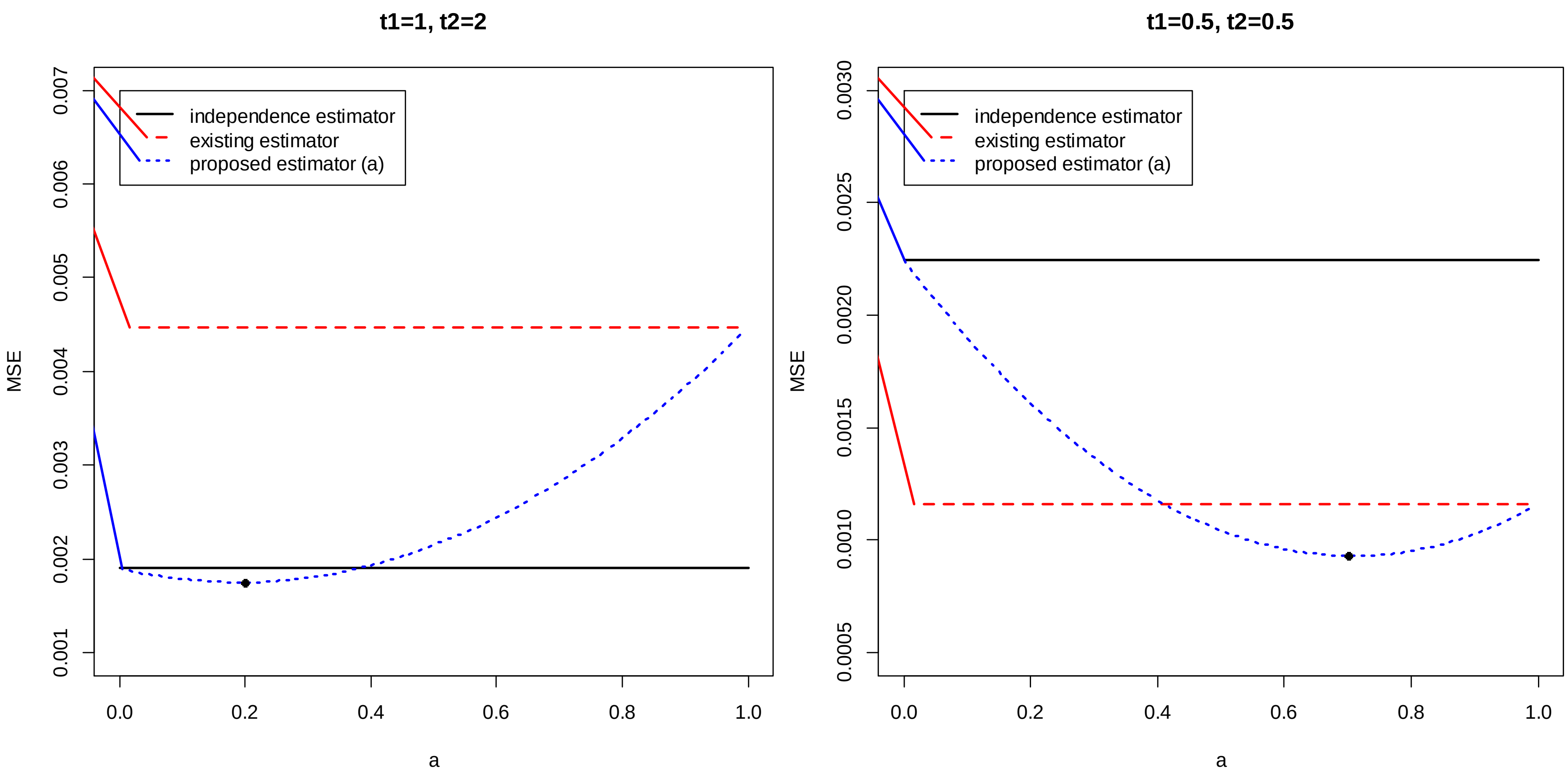
- **Figure** Simulation results for estimated values of four different estimators with n = 100. Data follow $T_1, T_2, Z_1, Z_2 \sim \exp(1)$ under the Clayton model with $\alpha = 1.1 \sim 7$ $(\tau = 0.05 \sim 0.75)$ The corresponding pair of causes are $(i, j) = (1, 1)$.



- **Figure** Simulation results for MSE of four different estimators with n = 100. Data follow $T_1, T_2, Z_1, Z_2 \sim \exp(1)$ under the Clayton model with $\alpha = 1.1 \sim 7$ $(\tau = 0.05 \sim 0.75)$. The corresponding pair of causes are $(i, j) = (1,1)$



- **Figure** Simulation results for MSE of three different estimators with n = 100. Data follow $T_1, T_2, Z_1, Z_2 \sim \exp(1)$ under the Clayton copula with $\alpha = 5$ $(\tau = 0.667)$ The corresponding pair of causes are $(i, j) = (1, 1)$, "•" indicates the optimal value at a^*.



Data analysis

- **Mouse data analysis**

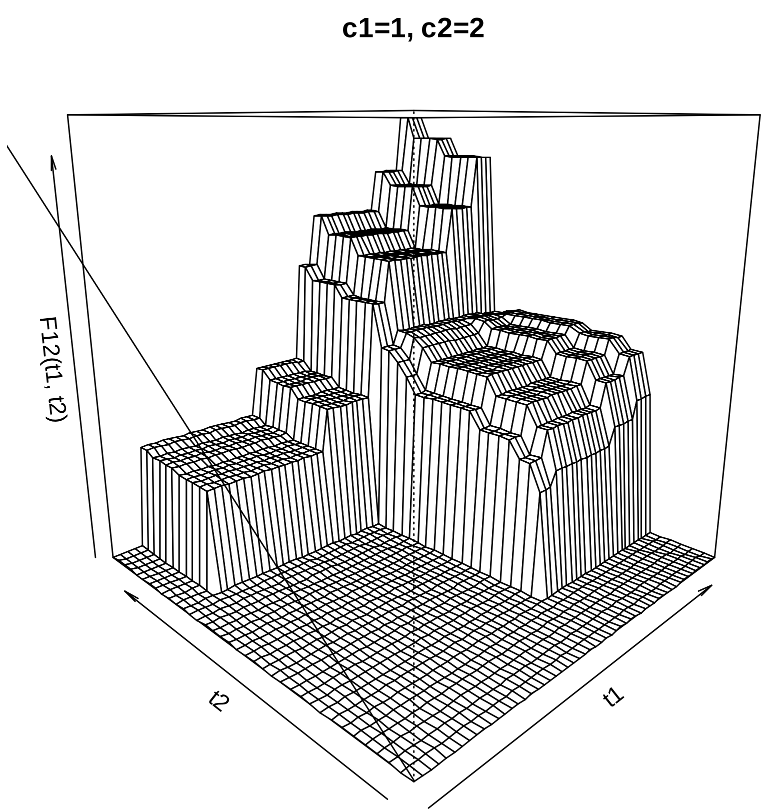
- The mouse data concerning the times to tumor appearance or death for 50 pairs of mice from the same litter in a tumor genesis experiment (Mantel and Ciminera, 1979).

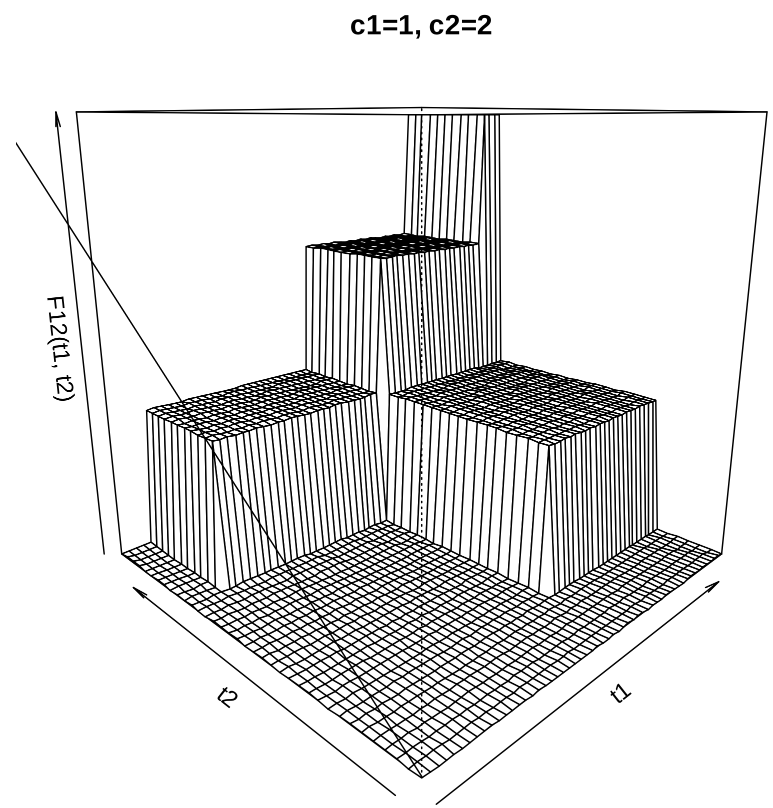
- In this data, T_1 and T_2 is failure times (in weeks) for a pair of mice.

$C_k \; (k = 1, 2)$

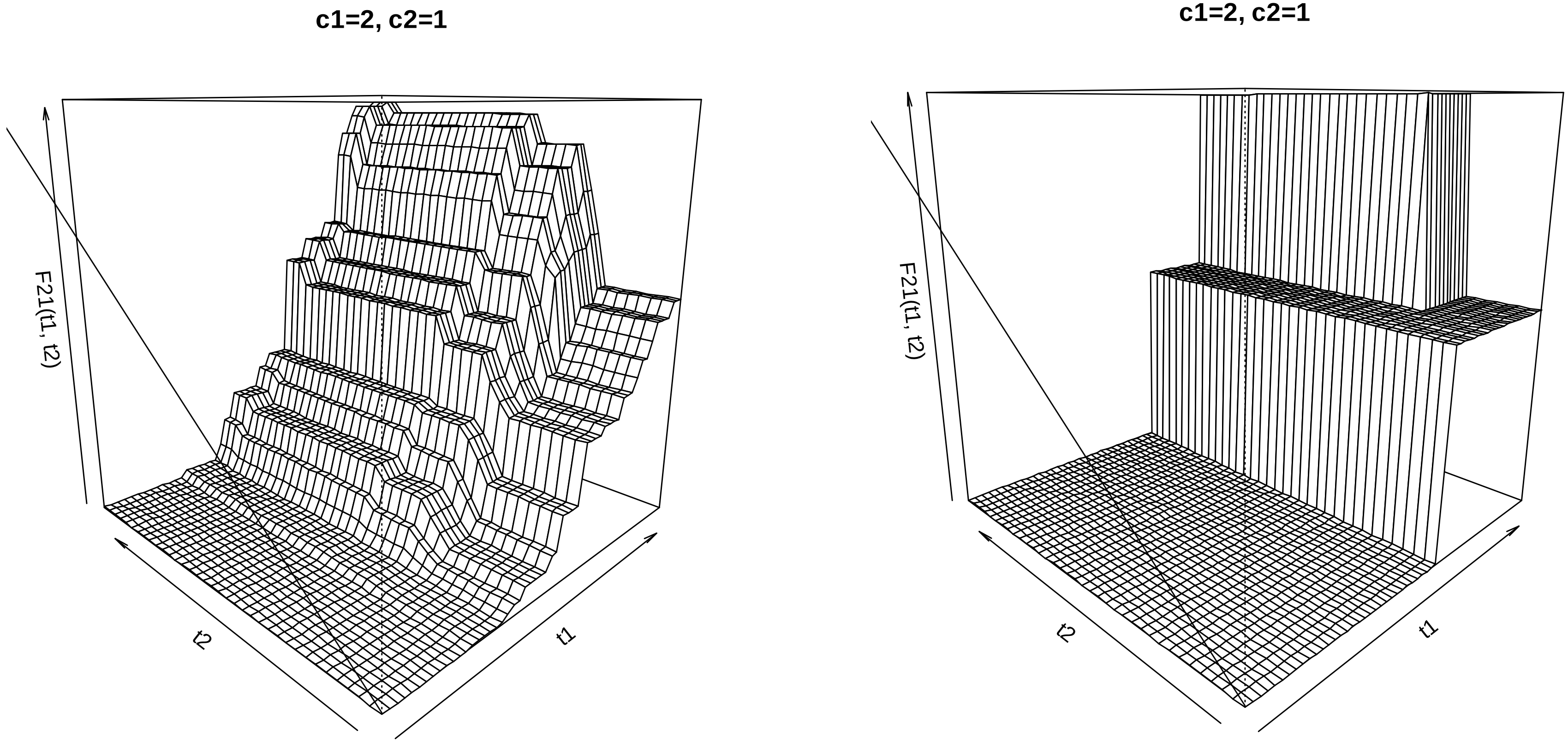
- 1 → Failure was appearance of a tumour
- 2 → Death happened before tumour appearance
- 0 → Censored observation (104 weeks)

- **Figure** The proposed estimator $\hat{F}_{ij}^{\hat{a}}(t_1, t_2)$ and the existing estimator $\hat{F}_{ij}(t_1, t_2)$ of cause-specific distribution function with the mouse data. Two plots correspond to pairs of cause $(i, j) = (1, 2)$.





- **Figure** The proposed estimator $\hat{F}_{ij}^{\hat{a}}(t_1, t_2)$ and the existing estimator $\hat{F}_{ij}(t_1, t_2)$ of cause-specific distribution function with the mouse data. Two plots correspond to pairs of cause $(i, j) = (2, 1)$.

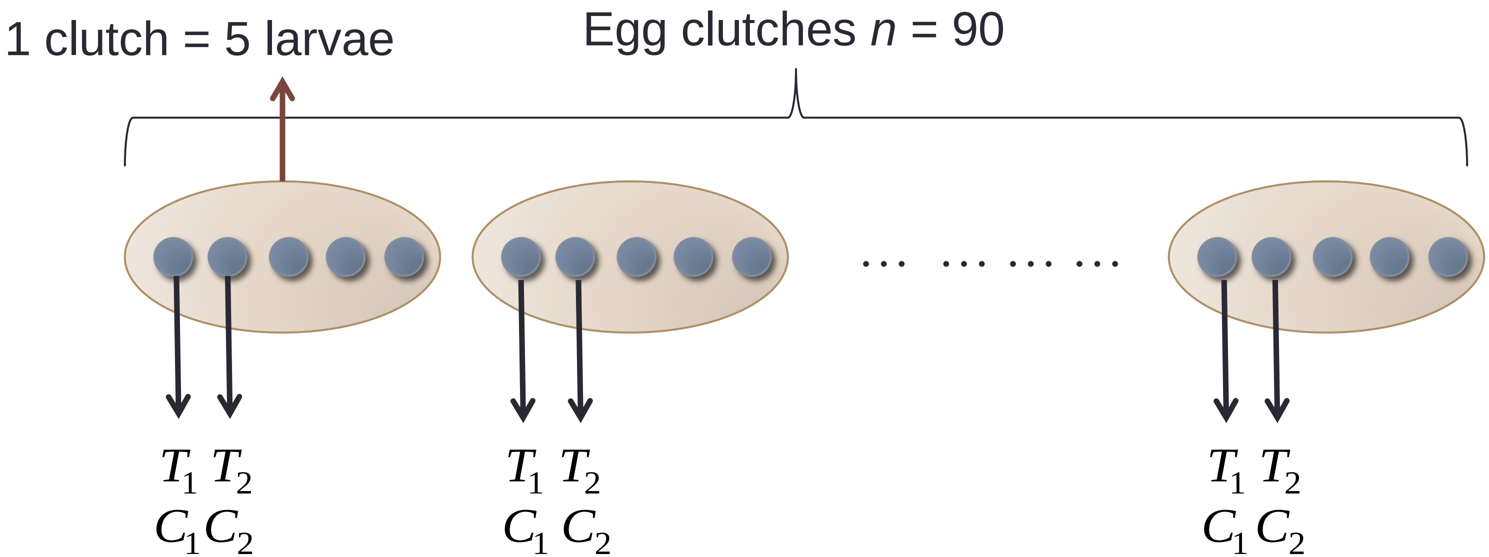


- **Salamander data analysis**

- For the salamander data, we consider the time that completion of metamorphosis on the salamander living in Hokkaido Japan (Michimae and Emura, 2012).

- They conduct an experimental design where two experimental factors (water levels and food types) are randomly assigned.

- We focus on the experiment factor water levels in this paper.

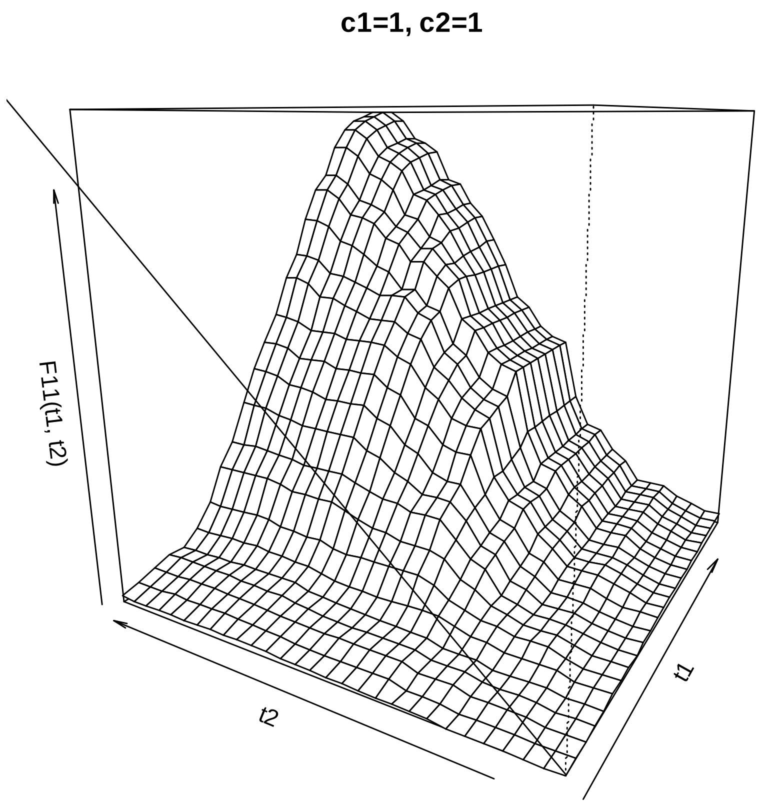


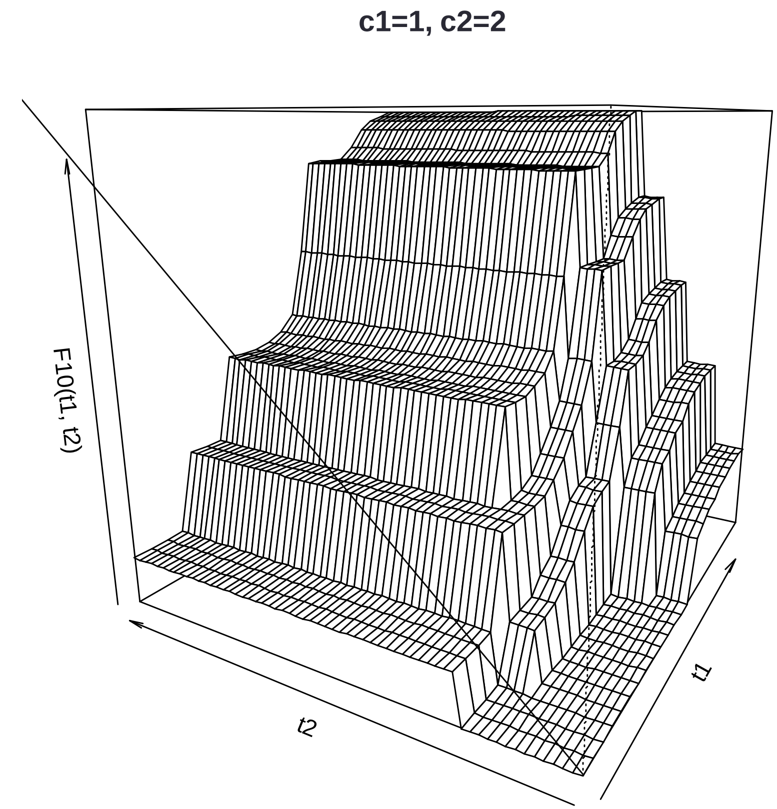
$C_k \ (k = 1, 2) \begin{cases} 1 \\ 2 \end{cases}$

1 → The event was metamorphosis

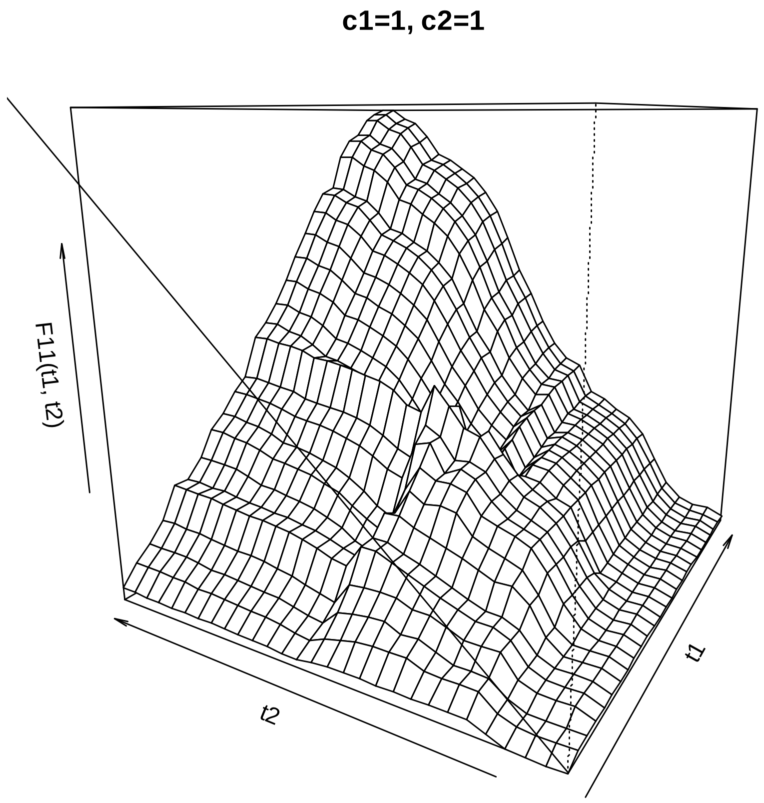
2 → The larvae was death prior to metamorphosis

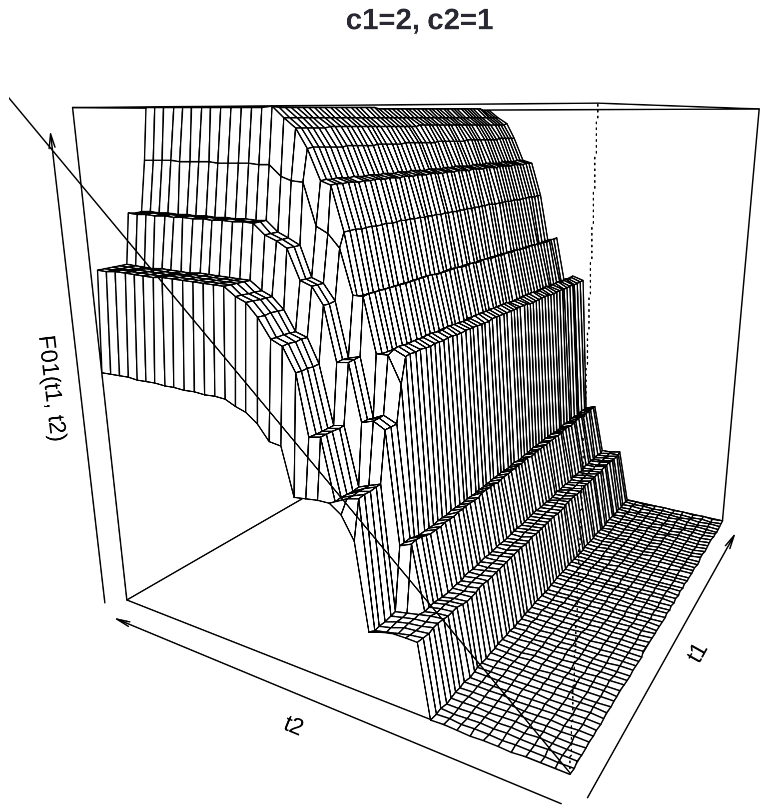
- **Figure 12** The proposed estimator $\hat{F}_{ij}^{\hat{a}}(t_1, t_2)$ of cause-specific distribution function with the salamander data for low water level. Two plots correspond to pairs of two different causes (i, j) = (1, 1), (1, 2).





- **Figure 13** The proposed estimator $\hat{F}_{ij}^{\hat{a}}(t_1, t_2)$ of cause-specific distribution function with the salamander data for high water level. Three plots correspond to pairs of three different causes (i, j) = (1, 1), (1, 2) and (2, 1).





Conclusion

- We have developed a new nonparametric estimator which improved the Sankaran et al. (2006) estimator.
- The idea of combining two estimators to improve the MSE is similar with Akritas and Keilegom (2003).
- Chen et al. (2009) also proposed an estimator which combines the “model-free” estimator and “model-based” estimator that imposes a strong independence assumption.
- We show that the MSE of the proposed method is smaller than the existing estimator in theoretical and numerical.
- In future work, the uniform consistency, weak convergence and the bootstrap convergence of the proposed estimator need to be established.

Thank You for Your Attention