Seminar in
Taida Inst. of Mathematical Sciences

Copula modeling for dependent truncation

Takeshi Emura
Inst. Stat. Science, Academia Sinica

Joint work with Weijing Wang
Inst. Stat., National Chiao Tung U.

Outlines

Part I: Copula: Review

- Copula - definition
- Copula - examples

Part II: Truncation data

- Truncation data
- Semi-survival copula
- Existing procedures   - moment method -

Part III: Proposed method

- Reverse-time hazard model
- Proposed method - nonparametric likelihood method –
- Simulation & data analysis
- Comparison with existing methods
- Conclusion & future work

Part I

Copula: Review

Copula

- *Definition*

The function $C: [0,1]\times[0,1]\mapsto[0,1]$ is said to be Copula when it is a bivariate distribution function having the uniform [0, 1] marginals

$$C[u,1]=u, \quad C[1,v]=v$$

- *Any bivariate distribution function* $F(x,y)$ *has a representation*

$$F(x,y)=C[F_X(x),F_Y(y)], \quad \text{where} \begin{cases} F_X(x)=F(x,\infty) \\ F_Y(y)=F(\infty,y) \end{cases}$$

Sklar's theorem (Sklar, 1959)

Copula

$$\Pr(X \le x, Y \le y) = C[\Pr(X \le x), \Pr(Y \le y)]$$

- Example 1: Independence copula

$$C[u,v] = uv$$

- Example 2: Frank copula (Genest, 1986; Frank, 1979)

$$C_\alpha[u,v] = \log_{\alpha^{-1}}\left\{1 + \frac{(\alpha^{-u}-1)(\alpha^{-v}-1)}{(\alpha^{-1}-1)}\right\}, \quad \alpha > 0$$

$$\lim_{\alpha \to 1} C_\alpha[u,v] = uv$$

- Example 3: Normal copula

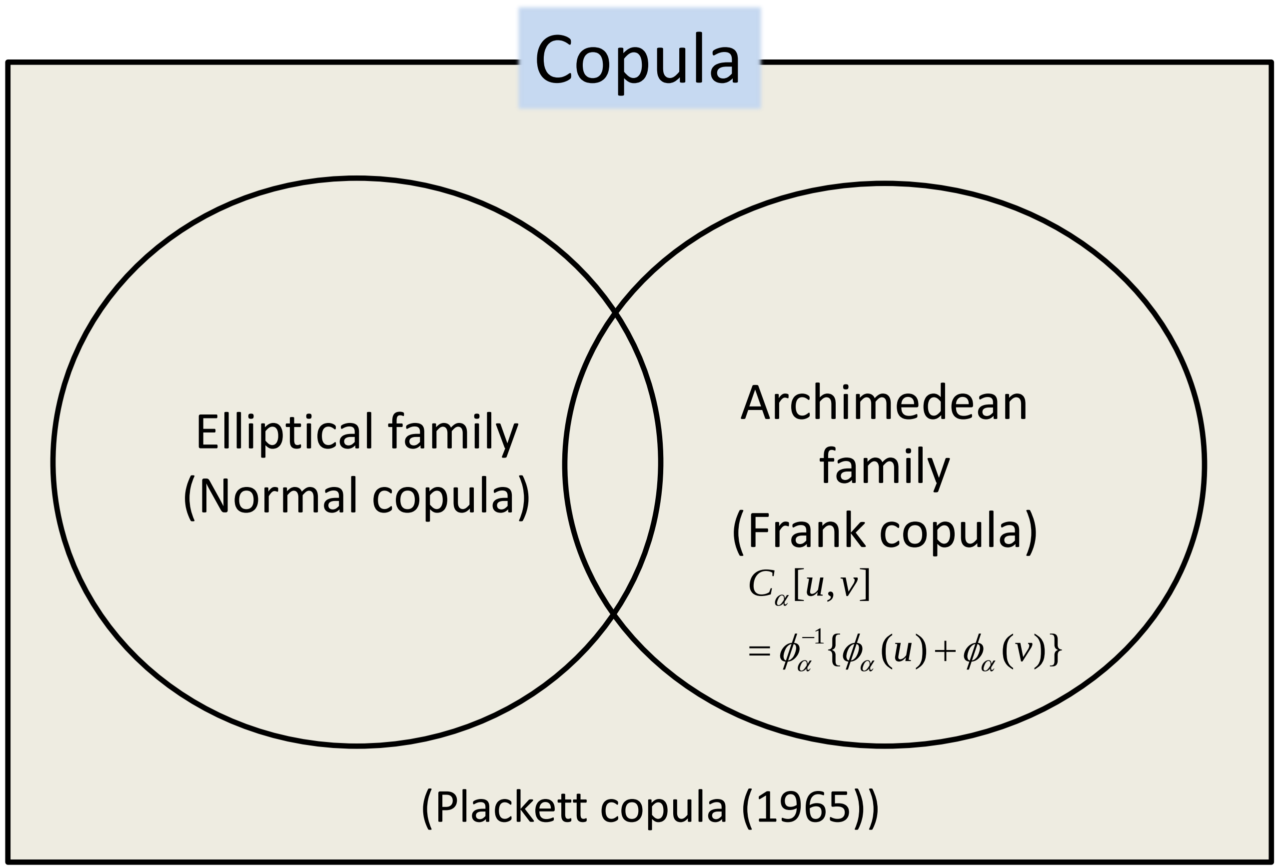
$$C_\rho[u,v] = \Phi_\rho[\Phi^{-1}(u), \Phi^{-1}(v)], \quad -1 < \rho < 1$$

Φ_ρ : Joint CDF of standard bivariate normal

$$\lim_{\rho \to 0} C_\rho[u,v] = uv$$

Copula

$$C[u, v]$$


Copula
Elliptical family
(Normal copula)
Archimedean
family
(Frank copula)
$C_\alpha[u,v]$
$= \phi_\alpha^{-1}\{\phi_\alpha(u) + \phi_\alpha(v)\}$
(Plackett copula (1965))

Copula

Copula in parametric setting

$$\Pr(X \le x, Y \le y) = C[F_X(x), F_Y(y)]$$

- Example 1: Election in UK (Smith, 2004)

  X : election time : Weibull
  Y : the number of votes: Normal
  } Joint?

  * Ali-Mikhail-Haq copula (Fukumoto, 2009) based on AIC

- Example 2: Insurance payment

  X : Indemnity payment: Parete
  Y : expenses (termed ALAE): Parete
  } Joint?

  * Frees and Valdez (1998) fit Gumbel copula based on AIC

Copula

Copula in semi-parametric setting

$$\Pr(X \le x, Y \le y) = C[F_X(x), F_Y(y)]$$

- Example 3: Australian Twin study (Prentice & Hsu, 1997)

  X : Time to disease for child 1 : Non-parametric
  Y : Time to disease for child 2: Non-parametric
  } Joint?

  * Prentice and Hsu fit Clayton copula
  * Gumbel copula may be the best one (Emura et al., 2010)

- Example 4: Transfusion-related AIDS (Lagakos et al., 1988)

  X : Incubation time of AIDS: Non-parametric
  Y : Infection time of AIDS: Non-parametric
  } Joint?

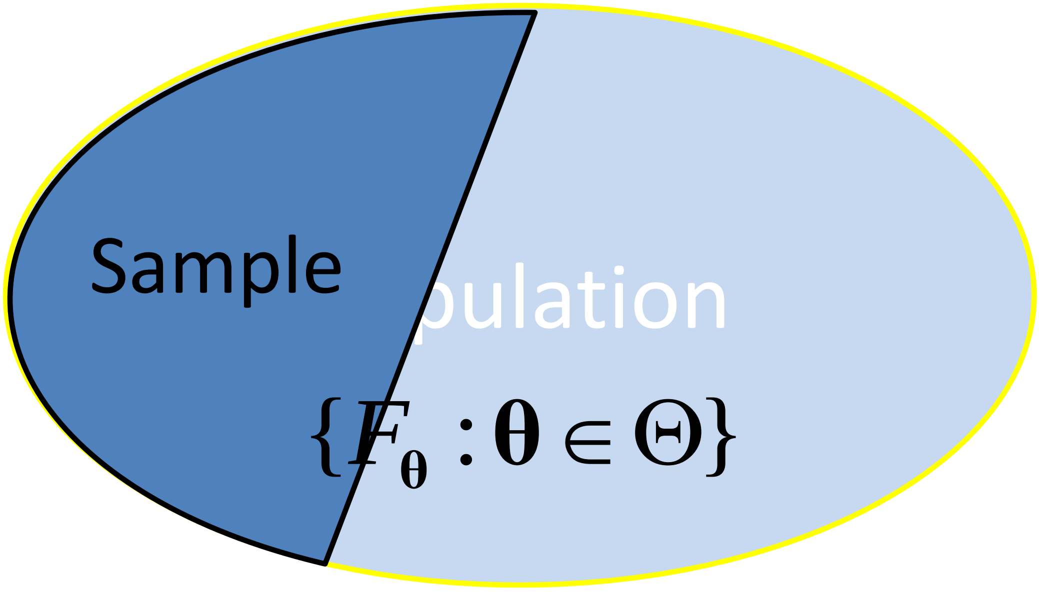
  * Chaieb et al. (2006) fit Frank copula
  * Beaudoin & Lakhal-Chaieb (2008) shows Clayton is better
  * We will argue that Clayton copula is the best one

Part II

Truncation data: Review

Truncation data

- *Truncated samples are those from which certain population values are entirely excluded* (Truncated and censored sample by Cohen, 1991)



Industry & Reliability (Book of Cohen, 1991; Navaro & Ruiz, 1996)

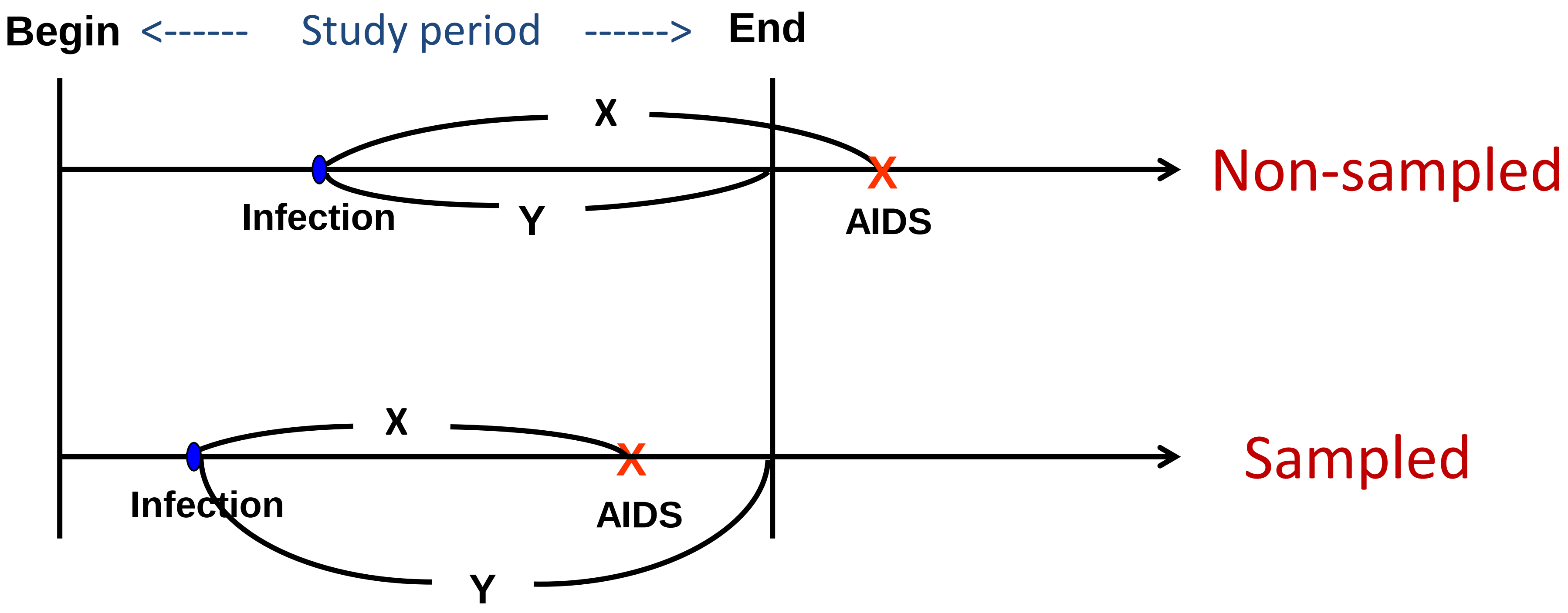
Biomedical studies (Book of Klein & Moeschberger, 2003)

Econometrics (Book of Amemiya, 1994, Chap 13)

Truncation data

- Transfusion-related AIDS

  (Lagakos et al., 1988; Kalbfleisch & Lawless, 1989)



Truncation criteria : $X \leq Y$

Truncation data

- Truncation data:

$\{(X_j, Y_j);\ j = 1, \ldots, n\}$

subject to $X_j \le Y_j$

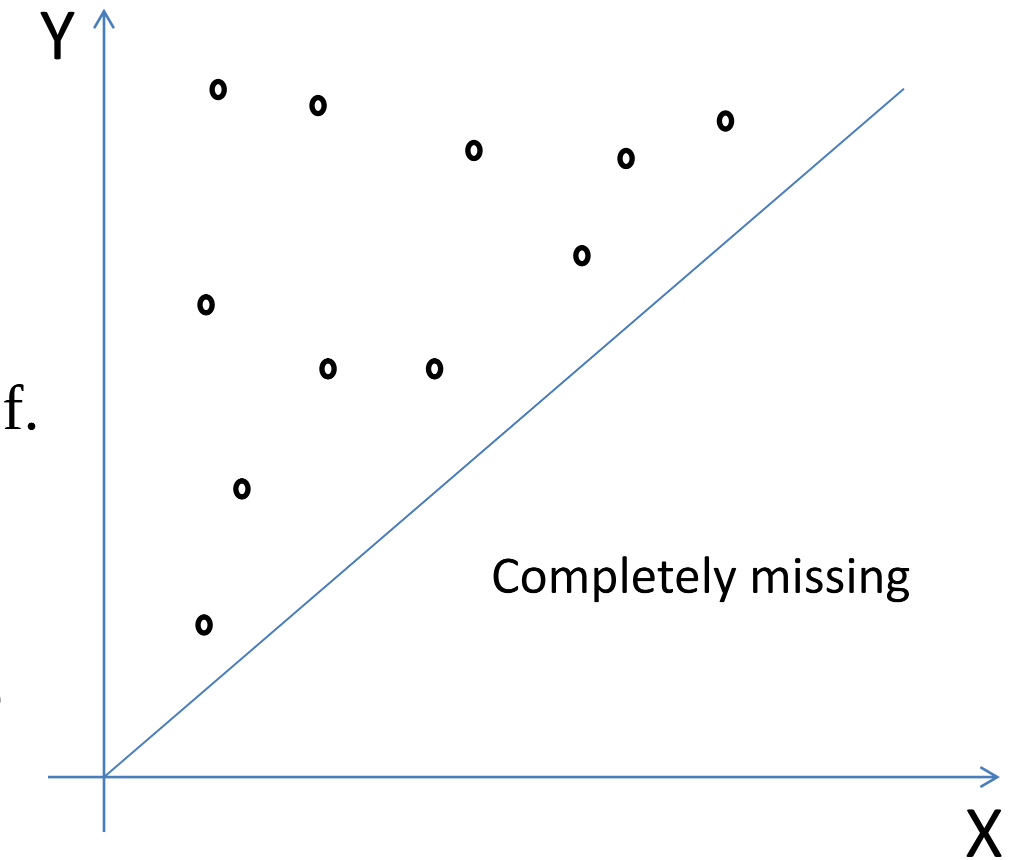
$\Downarrow$

i.i.d. from the conditional c.d.f.

$\Pr(X \le x, Y \le y \mid X \le Y)$,

where (X, Y) is

the population random variable



$$\frac{1}{n}\sum_{j=1}^{n} I(X_j \le x) \rightarrow_{Bias} \Pr(X \le x)$$

Truncation data

Traditional analysis

- Estimation of $F_X(x) = \Pr(X \le x)$

$$\hat{F}_X(x) = \prod_{u>x} \left\{ 1 - \frac{\sum_{j=1}^{n} I(X_j = u)}{\sum_{j=1}^{n} I(X_j \le u, Y_j \ge u)} \right\}$$

(Lynden-Bell, 1971; Lagakos et al., 1988)

- Quasi-independence assumption (Tsai, 1991):

$$\Pr(X \le x, Y \le y \mid X \le Y) \propto \iint_{\substack{u \le x, v \le y \\ u \le v}} dF_X(u) dF_Y(v)$$

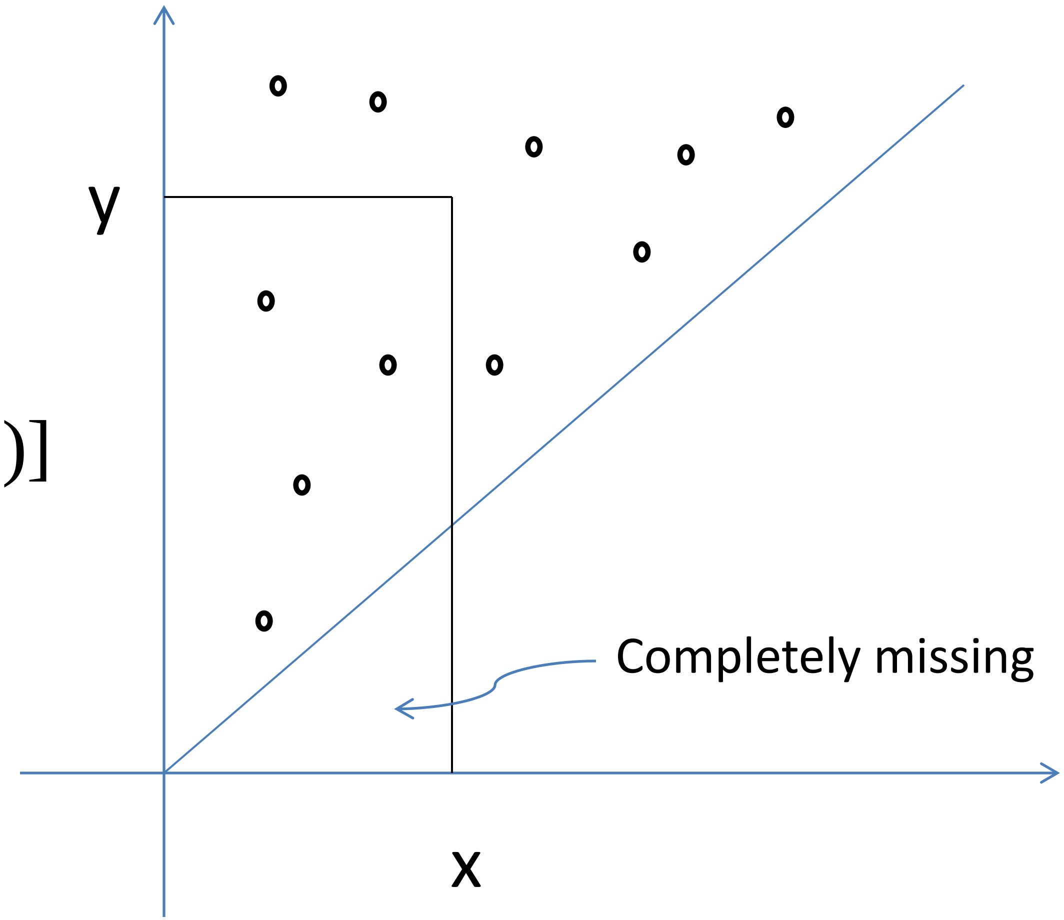
*Quasi-independence assumption is testable

(Chen et al., 1996; Martin & Betensky, 2005; Emura & Wang; 2010)

Truncation data

$$\Pr(X \le x, Y \le y)$$
$$= C[\Pr(X \le x), \Pr(Y \le y)]$$

The model is unidentifiable

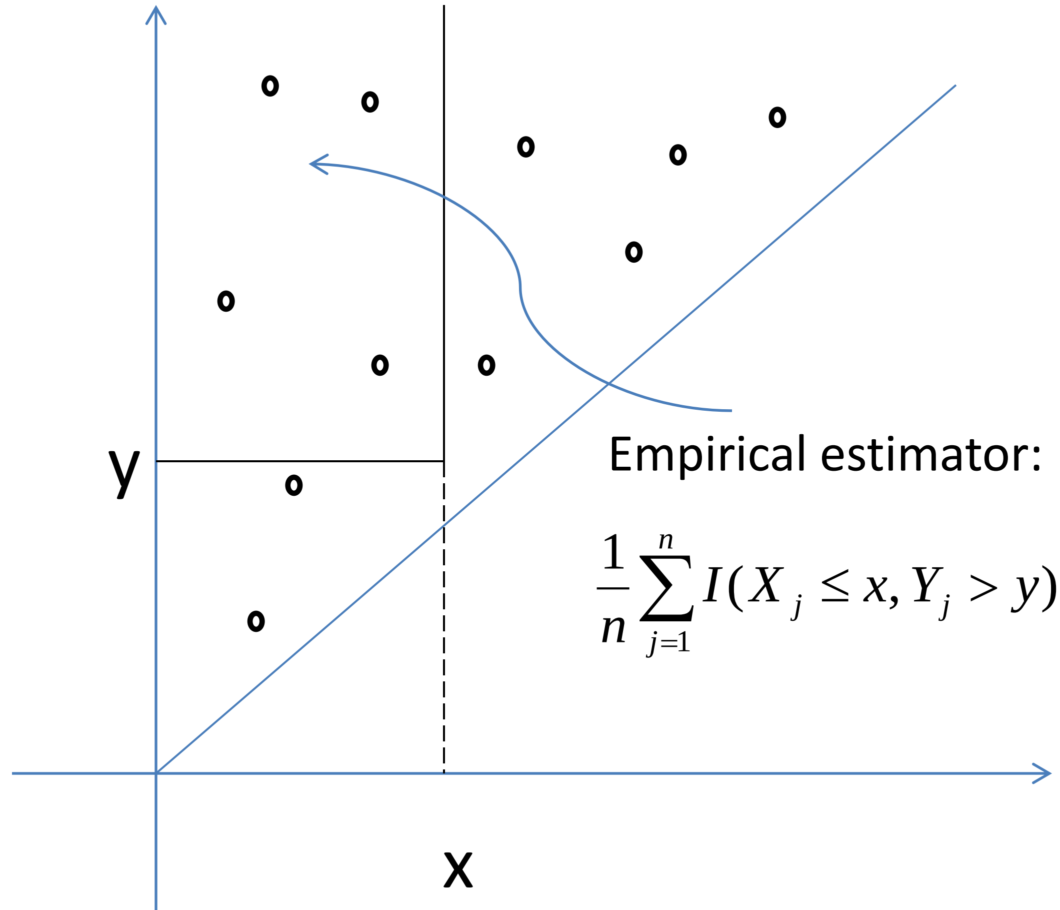


Truncation data

$$\Pr(X \le x, Y > y \mid X \le Y) = \frac{C_\alpha[F_X(x), S_Y(y)]}{c(\alpha, F_X, S_Y)}$$

where

$$c(\alpha, F_X, S_Y) = \iint_{x \le y} \frac{\partial^2}{\partial x \partial y} C_\alpha[F_X(x), S_Y(y)]dxdy$$



Empirical estimator:

$$\frac{1}{n}\sum_{j=1}^{n} I(X_j \le x, Y_j > y)$$

- Semi-survival copula

(Chaieb et al., 2006, *Biometrika*)

- Quasi-independence: $C_\alpha[u, v] = uv$

Truncation data

- Estimator for (F_X, S_Y)

$$\frac{1}{n}\sum_{j=1}^{n} I(X_j \le t, Y_j > t) = \frac{C_\alpha[F_X(t), S_Y(t)]}{c(\alpha, F_X, S_Y)},$$

where $t \in (X_1, \ldots X_n, Y_1, \ldots, Y_n)$

(Chaieb et al., 2006)

- Estimator for α

1. Conditional Kendall's tau (Chaieb et al, 2006)
2. Conditional likelihood (Emura, Wang & Hung, 2011, Sinica)

* Conditional likelihood achieves higher efficiency

Drawbacks of existing procedures

- Chaieb et al. (2006) and Emura et al. (2011) are restricted to Archimedean family

$$C_\alpha[u,v]=\phi_\alpha^{-1}\{\phi_\alpha(u)+\phi_\alpha(v)\}$$

$$\because)\ \frac{1}{n}\sum_{j=1}^{n} I(X_j \le t, Y_j > t) = \frac{C_\alpha[F_X(t), S_Y(t)]}{c(\alpha, F_X, S_Y)},$$

$$\Leftrightarrow \phi_\alpha\left(\frac{c(\alpha, F_X, S_Y)}{n}\sum_{j=1}^{n} I(X_j \le t, Y_j > t)\right) = \phi_\alpha(F_X(t)) + \phi_\alpha(S_Y(t))$$

$$\Leftrightarrow F_X(t) = \phi_\alpha^{-1}\left\{\phi_\alpha\left(\frac{c(\alpha, F_X, S_Y)}{n}\sum_{j=1}^{n} I(X_j \le t, Y_j > t)\right) - \phi_\alpha(S_Y(t))\right\}$$

- The assumption of no ties: $t \in (X_1, \ldots X_n, Y_1, \ldots, Y_n)$
- Efficiency concern

- Part III: Proposed method

Proposed method

- The preceding two methods use moment-based estimating equations for (F_X, S_Y)

- In this talk, we propose to get $(\hat{\alpha}, \hat{F}_X, \hat{S}_Y)$ by the nonparametric maximum likelihood estimator (NPMLE)

* Motivation: Higher efficiency of the NPMLE

Proposed method

- Re-parameterize (F_X, S_Y)

$$F_X(x) = e^{-H_X(x)}, \quad S_Y(y) = e^{-\Lambda_Y(y)}$$

$* H_X(x)$: Reverse - time cumulative hazard

(Lagakos et al., 1988; Navaro & Ruiz, 1996)

$* \Lambda_Y(y)$: Cumulative hazard

- Copula model:

$$\Pr(X \le x, Y > y \mid X \le Y) = \frac{C_\alpha[e^{-H_X(x)}, e^{-\Lambda_Y(y-)}]}{c(\alpha, H_X, \Lambda_Y)},$$

$$\text{where} \quad c(\alpha, H_X, \Lambda_Y) = \iint_{x \le y} -\frac{\partial^2}{\partial x \partial y} C_\alpha[e^{-H_X(x)}, e^{-\Lambda_Y(y-)}]dxdy$$

Proposed method

- Density

$$\Pr(X = x, Y = y \mid X \le Y) = \frac{\eta_\alpha[H_X(x), \Lambda_Y(y-)]}{c(\alpha, H_X, \Lambda_Y)}\{-dH_X(x)\}d\Lambda_Y(y),$$

where $\eta_\alpha[x, y] = e^{-x}e^{-y}\frac{\partial^2}{\partial u \partial u}C_\alpha[u, u]\Big|_{u=e^{-x}, v=e^{-y}}$

- Log-likelihood

$$l_n(\alpha, H_X, \Lambda_Y) =$$

$$\sum_{j=1}^{n}\log\eta_\alpha[H_X(X_j), \Lambda_Y(Y_j-)] + \log\{-dH_X(X_j)\} + \log d\Lambda_Y(Y_j) - \log c(\alpha, H_X, \Lambda_Y)$$

- Maximization for (*2n+1*) parameters

$$(\alpha, -dH_X(X_1), \dots, -dH_X(X_n), d\Lambda_Y(Y_1), \dots, d\Lambda_Y(Y_n))$$

Proposed method

- Identifiability problem

  We found that the maximum of $l_n(\alpha, H_X, \Lambda_Y)$ is not unique

  (# of parameters =2n+1 > # of observed points = 2n)

- Reduces to 2n-1 parameters

$$(\alpha, -dH_X(X_1), \ldots, -dH_X(X_n), d\Lambda_Y(Y_1), \ldots, d\Lambda_Y(Y_n))$$

$$\Downarrow$$

$$(\alpha, \underbrace{-dH_X(X_{(1)})}_{\equiv 1}, \ldots, -dH_X(X_{(n)}), d\Lambda_Y(Y_{(1)}), \ldots, \underbrace{d\Lambda_Y(Y_{(n)})}_{\equiv 1})$$

Proposed method

Geometrical understanding of $-dH_X(X_{(1)})=1$

$$\Pr(X \le x_{(1)}, Y > x_{(1)} \mid X \le Y)$$

$$= \frac{C_\alpha[F_X(x_{(1)}), S_Y(x_{(1)})]}{c(\alpha, F_X, S_Y)}$$

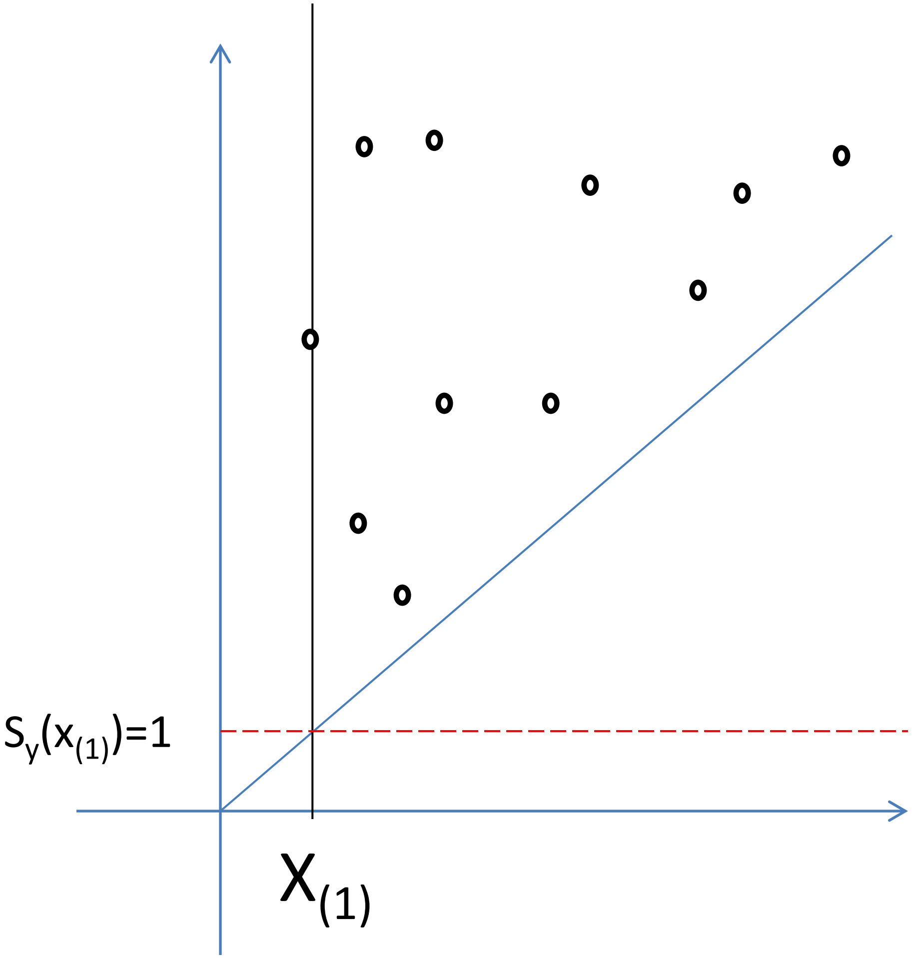
$$= \frac{F_X(x_{(1)})}{c(\alpha, F_X, S_Y)} \qquad (1)$$

$$\therefore) -dH_X(x_{(1)})$$

$$= \frac{dF_X(x_{(1)})}{F_X(x_{(1)})} \qquad \because F_X = e^{-H_X}$$

$$= \frac{\Pr(X = x_{(1)} \mid X \le Y)}{\Pr(X \le x_{(1)} \mid X \le Y)} \qquad \text{by (1)}$$

$$\leftarrow \frac{1/n}{1/n} = 1 \qquad \text{see the plot}$$



Proposed method

- *2n-1* score equations

$$0 = \partial l_n(\alpha, H_X, \Lambda_Y)/\partial\alpha$$

$$0 = \partial l_n(\alpha, H_X, \Lambda_Y)/\partial\{-dH_X(X_{(j)})\}, \quad j = 2,..,n$$

$$0 = \partial l_n(\alpha, H_X, \Lambda_Y)/\partial d\Lambda_Y(Y_{(j)}), \quad j = 1,..,n-1$$

- Breslow-Aalen type expression

$$H_X(x) = \int_x^\infty \frac{\sum_{j=1}^n I(X_j = u)}{\sum_{j=1}^n \Psi_j^{(1,0)}(u;\alpha, H_X, \Lambda_Y)}$$

$$\Lambda_Y(x) = \int_0^y \frac{\sum_{j=1}^n I(Y_j = u)}{\sum_{j=1}^n \Psi_j^{(0,1)}(u;\alpha, H_X, \Lambda_Y)}$$

Proposed method

- To get the NPMLE $(\hat{\alpha}, \hat{H}_X, \hat{\Lambda}_Y)$

  we apply a numerical maximization to

$$l_n(\alpha, H_X, \Lambda_Y) =$$

$$\sum_{j=1}^{n} \log \eta_\alpha[H_X(X_j), \Lambda_Y(Y_j-)] + \log\{-dH_X(X_j)\} + \log d\Lambda_Y(Y_j) - \log c(\alpha, H_X, \Lambda_Y)$$

  for parameter $(\alpha, -dH_X(X_1), \ldots, -dH_X(X_n), d\Lambda_Y(Y_1), \ldots, d\Lambda_Y(Y_n))$

  subject to $-dH_X(X_{(1)}) = d\Lambda_Y(Y_{(n)}) = 1$

(e.g. use “nlm” program in R)

- $l_n(\alpha, H_X, \Lambda_Y)$: twice differentiable & convex

Proposed method

- The NPMLE $(\hat{\alpha}, \hat{H}_X, \hat{\Lambda}_Y)$ is consistent & asymptotic normal
- Observed Fisher information

  = minus of the Hessian of $l_n(\alpha, H_X, \Lambda_Y)$

$$\hat{i}_n(\hat{\alpha}, \hat{H}_X, \hat{\Lambda}_Y) = \begin{bmatrix} \hat{i}_{n,11} & \hat{i}'_{n,12} \\ \hat{i}_{n,12} & \hat{i}_{n,22} \end{bmatrix}$$

- Consistent variance estimator

$$\hat{V}_n(\hat{\alpha}) \approx (\hat{i}_{n,11} - \hat{i}'_{n,12}\hat{i}^{-1}_{n,22}\hat{i}_{n,12})^{-1}$$

Proposed method

Simulation setting (I):

- Plackett copula (not Archimedean family)

$$C_\alpha[u,v] = \frac{1}{2(\alpha-1)} + \frac{u+v}{2} - \frac{[\{1+(\alpha-1)(u+v)\}^2 - 4uv\alpha(\alpha-1)]^{1/2}}{2(\alpha-1)}$$

$\alpha = 1/2.51,\ 1/5.11,\ 2.51,\ 5.11$

(s.t. Spearmen's rho $= 0.25,\ 0.5, -0.25, -0.5$)

- Exponential margins

$$H_X(x) = -\log(1-e^{-1.5x})$$

$$\Lambda_Y(y) = 0.5y$$

- Data generation

If $X_j \le Y_j$ then included in the sample. Otherwise truncated.

Repeat until we get *n* (=125 or 250) pair of (X_j, Y_j)

Simulation results (positive dependence)

Parameter		*Mean* (*Bias*)	*SE*	*SEE*	95%*Cov*
Spearman's ρ = 0.25 (α =1/2.15, $\Pr(X \leq Y) = 0.79$)					
$\log(\alpha)$=-0.765	n =125	-0.778 (-0.013)	0.407	0.407	0.945
	n =250	-0.697 (0.068)	0.311	0.296	0.965
$H_X(t) = 0.693$	n =125	0.736 (0.043)	0.123	0.121	0.955
	n =250	0.733 (0.040)	0.090	0.086	0.970
$\Lambda_Y(t) = 0.693$	n =125	0.710 (0.017)	0.144	0.139	0.960
	n =250	0.725 (0.032)	0.104	0.102	0.970
Spearman's ρ = 0.50 (α =1/5.11, $\Pr(X \leq Y) = 0.84$)					
$\log(\alpha)$=-1.631	n =125	-1.642 (-0.011)	0.323	0.319	0.965
	n =250	-1.652 (-0.021)	0.231	0.222	0.940
$H_X(t) = 0.693$	n =125	0.726 (0.033)	0.101	0.092	0.910
	n =250	0.716 (0.023)	0.067	0.064	0.920
$\Lambda_Y(t) = 0.693$	n =125	0.704 (0.011)	0.110	0.102	0.960
	n =250	0.701 (0.008)	0.068	0.069	0.950

Simulation results (negative dependence)

Parameter		*Mean* (*Bias*)	*SE*	*SEE*	95%*Cov*
Spearman's ρ = -0.25 (α =2.15, $\Pr(X \le Y) = 0.72$)					
$\log(\alpha)$=0.765	n=125	0.859 (0.094)	0.598	0.554	0.960
	n=250	0.717 (-0.048)	0.342	0.359	0.930
$H_X(t) = 0.693$	n=125	0.809 (0.116)	0.313	0.244	0.960
	n=250	0.717 (0.024)	0.139	0.138	0.935
$\Lambda_Y(t) = 0.693$	n=125	0.793 (0.100)	0.363	0.267	0.960
	n=250	0.699 (0.006)	0.139	0.137	0.930
Spearman's ρ = -0.50 (α =5.11, $\Pr(X \le Y) = 0.70$)					
$\log(\alpha)$= 1.631	n=125	1.758 (0.127)	0.818	0.598	0.915
	n=250	1.708 (0.077)	0.534	0.386	0.955
$H_X(t) = 0.693$	n=125	0.883 (0.190)	0.582	0.343	0.925
	n=250	0.787 (0.094)	0.374	0.196	0.960
$\Lambda_Y(t) = 0.693$	n=125	0.862 (0.169)	0.624	0.354	0.885
	n=250	0.775 (0.082)	0.404	0.207	0.955

Proposed method

Simulation setting (II):

- Frank copula (Archimedean family)

$$C_\alpha[u, v] = \log_{\alpha^{-1}}\left\{1 + \frac{(\alpha^{-u} - 1)(\alpha^{-v} - 1)}{(\alpha^{-1} - 1)}\right\},$$

$$\log(\alpha) = 2.38,\ 5.746,\ -2.38,\ -5.746$$

$$(\text{s.t. Kendall's tau} = 0.25,\ 0.5,\ -0.25,\ -0.5)$$

- Exponential margins

$$H_X(x) = -\log(1 - e^{-1.5x})$$

$$\Lambda_Y(y) = 0.5y$$

- Compare with estimator of Chaieb et al. (2006) and Emura et al. (2011)

Simulation results (positive dependence)

Parameter	*n*	*NPMLE*	*Emura et al.*	*Chaeib et al.*
		Kendall's τ =0.25		
$\log(\alpha)$ =2.38	125	0.0661 (0.8767)	-0.0075 (0.8546)	-0.0051 (0.8569)
	250	-0.1113 (0.5707)	-0.1326 (0.5636)	-0.1325 (0.5650)
$F_X(t) = 0.50$	125	-0.0067 (0.0509)	-0.0034 (0.0518)	-0.0034 (0.0518)
	250	-0.0115 (0.0409)	-0.0097 (0.0412)	-0.0098 (0.0413)
$S_Y(t) = 0.50$	125	-0.0033 (0.0585)	-0.0045 (0.0594)	-0.0045 (0.0595)
	250	-0.0057 (0.0437)	-0.0057 (0.0440)	-0.0057 (0.0441)
		Kendall's τ =0.5		
$\log(\alpha)$ =5.746	125	-0.0008 (1.1621)	-0.2696 (0.9674)	-0.2673 (0.9735)
	250	0.0580 (0.7594)	-0.0684 (0.6817)	-0.0675 (0.6835)
$F_X(t) = 0.50$	125	-0.0122 (0.0460)	-0.0092 (0.0426)	-0.0092 (0.0426)
	250	-0.0028 (0.0347)	-0.0022 (0.0325)	-0.0022 (0.0325)
$S_Y(t) = 0.50$	125	0.0006 (0.0470)	-0.0005 (0.0443)	-0.0005 (0.0443)
	250	-0.0029 (0.0386)	-0.0029 (0.0364)	-0.0028 (0.0364)

Each cell contains the average bias ($\times 10^{-2}$) and standard deviation ($\times 10^{-2}$) (in parenthesis) based on 200 runs.

Simulation results (negative dependence)

Parameter	*n*	*NPMLE*	*Emura et al.*	*Chaeib et al.*
		Kendall's τ =-0.25		
$\log(\alpha)$=-2.38	125	-0.1158 (1.1836)	0.3508 (1.0770)	0.3501 (1.0670)
	250	-0.0225 (1.0129)	0.0540 (1.0019)	0.0187 (1.0668)
$F_X(t) = 0.50$	125	-0.0175 (0.1063)	0.0404 (0.1142)	0.0403 (0.1140)
	250	-0.0207 (0.0831)	0.0141 (0.0939)	0.0114 (0.0944)
$S_Y(t) = 0.50$	125	-0.0161 (0.1112)	0.0421 (0.1132)	0.0419 (0.1129)
	250	-0.0136 (0.0918)	0.0211 (0.0948)	0.0183 (0.0953)
		Kendall's τ =-0.5		
$\log(\alpha)$=-5.746	125	0.6819 (0.9779)	2.4216 (2.0641)	2.3795 (2.1017)
	250	0.5088 (0.9816)	2.0827 (2.1311)	2.0571 (2.1160)
$F_X(t) = 0.50$	125	0.0485 (0.0882)	0.2099 (0.1676)	0.2070 (0.1695)
	250	0.0303 (0.0752)	0.1728 (0.1758)	0.1704 (0.1755)
$S_Y(t) = 0.50$	125	0.0508 (0.0862)	0.2090 (0.1695)	0.2061 (0.1715)
	250	0.0338 (0.0800)	0.1757 (0.1728)	0.1732 (0.1728)

Each cell contains the average bias ($\times 10^{-2}$) and standard deviation ($\times 10^{-2}$) (in parenthesis) based on 200 runs.

Proposed method

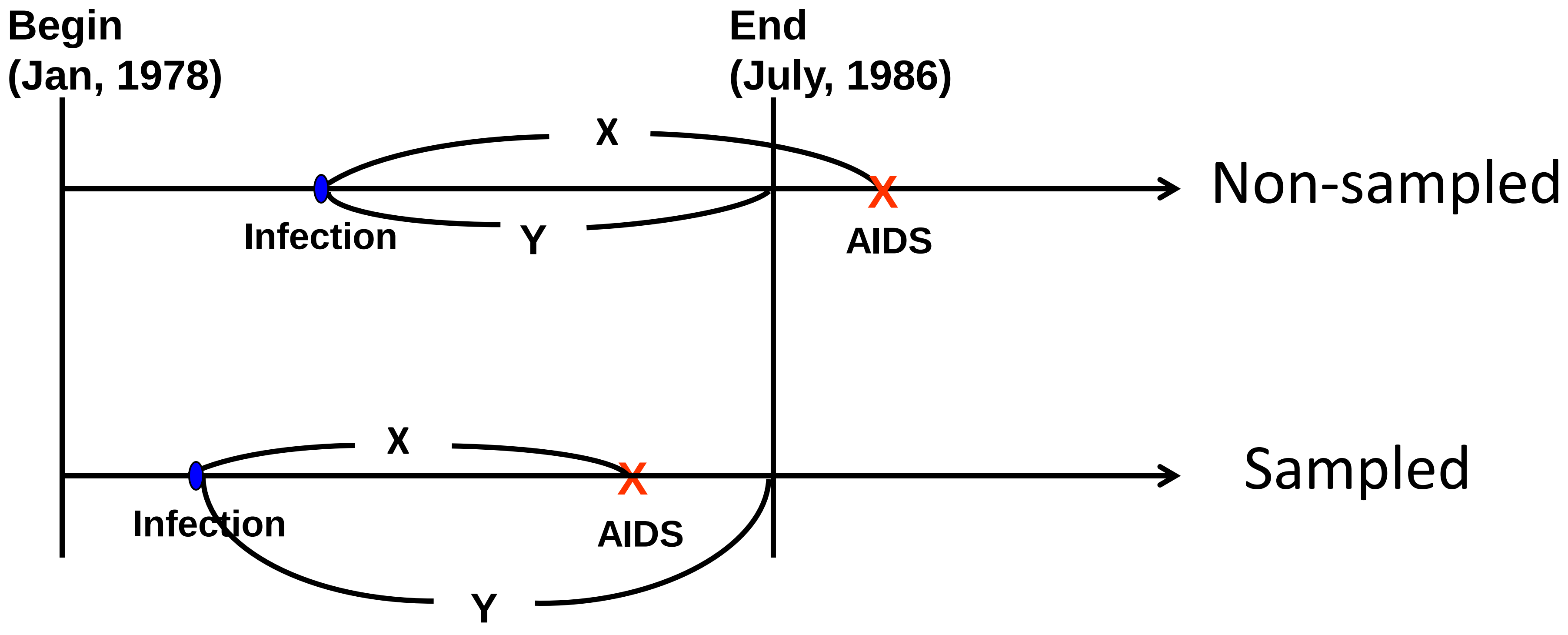
Data analysis

- Transfusion-related AIDS (Kalbfleisch & Lawless, 1989, JASA)

  X : Time from infection to AIDS (month)

  Y : 102 - time of infection (month)

  n : sample size = 293



Proposed method

Proposed method

- Model selection with (*K+1*) different copulas

$$\begin{cases} C^{(0)}[u,v] = uv \\ C_\alpha^{(k)}[u,v], \quad k = 1,\ldots,K \end{cases}$$

where $\lim_{\alpha \to 1} C_\alpha^{(k)}[u,v] = uv$

- Deviance

$$2\{l_n(\hat{\alpha}, \hat{H}_X, \hat{\Lambda}_Y) - l_n(1, \hat{H}_X^{\alpha=1}, \hat{\Lambda}_Y^{\alpha=1})\} \quad \sim \quad \chi^2_{df=1}$$

Step 1: Calculate deviances for *K* copulas

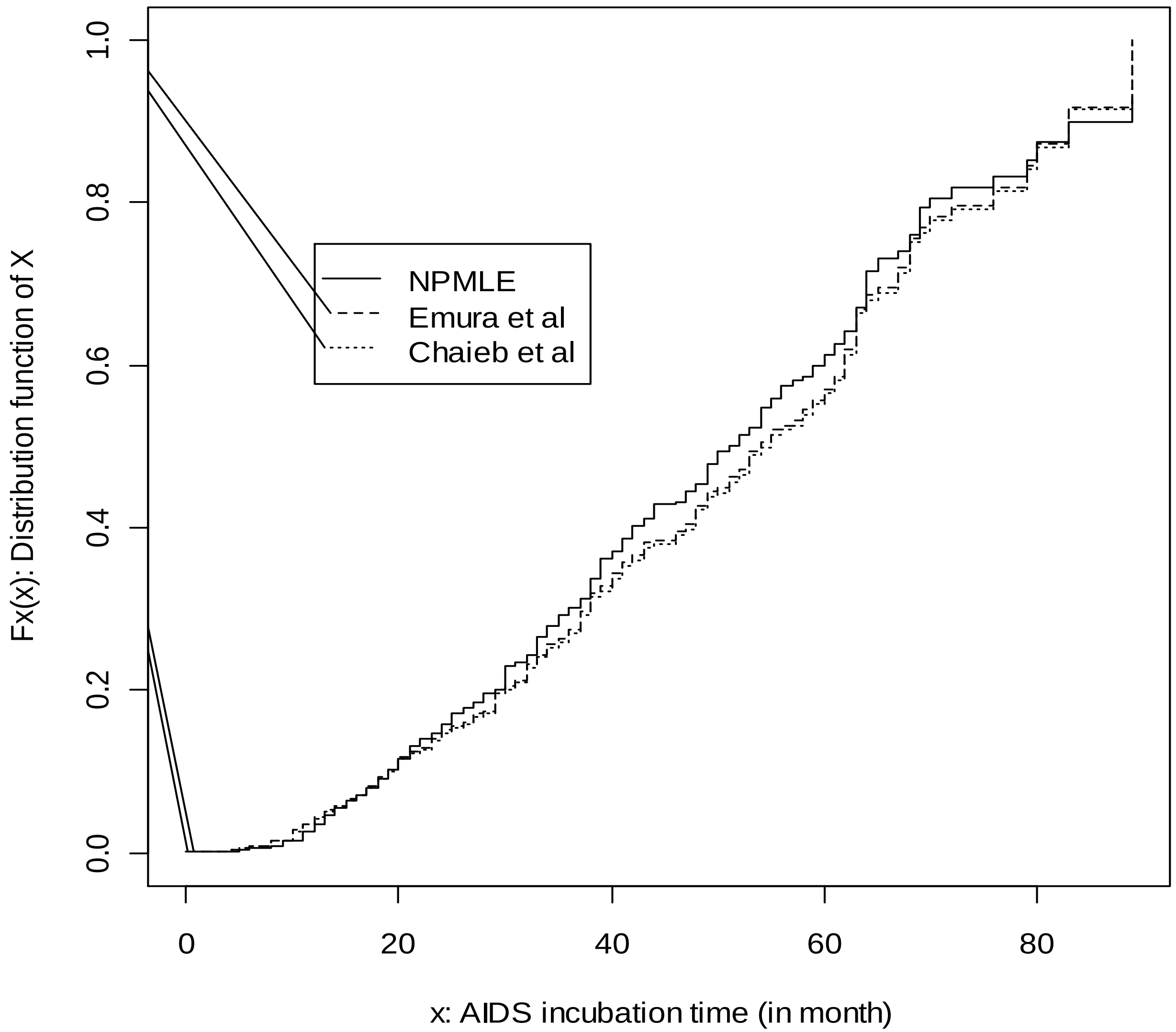
Step 2: Choose the copula with largest deviance & p-value <0.05

Table 1: Analysis of the Transfusion-related AIDS data

opula Type	Form*	Association parameter $\hat{\alpha}$ (SE)	Kendall's $\tau(\hat{\alpha})$	95% CI for α	Deviance (p-value)
Clayton	R	1.521 (0.172)	0.207	(1.218, 1.898)	8.568 (0.003)
ppendix C.1)	S	1.645 (0.233)	0.244	(1.246, 2.171)	5.228 (0.022)
	SS	0.763 (0.033)	0.134	(0.701, 0.831)	19.028 (0.000)
Frank	R, S	55.725 (42.704)	0.390	(12.41, 250.24)	10.828 (0.001)
ppendix C.2)	SS	0.018 (0.014)	0.390	(0.004, 0.081)	10.828 (0.001)
Plackett	R, S	5.293 (1.390)	0.356	(3.164, 8.856)	8.068 (0.005)
ppendix C.4)	SS	0.189 (0.050)	0.356	(0.113, 0.316)	8.068 (0.005)
Gumbel	R	1.459 (0.136)	0.315	(1.257, 1.821)	7.868 (0.005)
ppendix C.3)	S	1.340 (0.120)	0.254	(1.170, 1.678)	6.368 (0.012)
vo-parameter	R	$\hat{\alpha}$:1.521 (0.400) $\hat{\beta}$:1.000 (**)	0.207	α :(1.116, 3.348) β : (**)	8.588 (0.003)
ppendix C.5)	S	$\hat{\alpha}$:1.344 (0.264) $\hat{\beta}$:1.235 (0.140)	0.309	α :(1.076, 2.551) β :(1.073, 1.756)	7.928 (0.005)

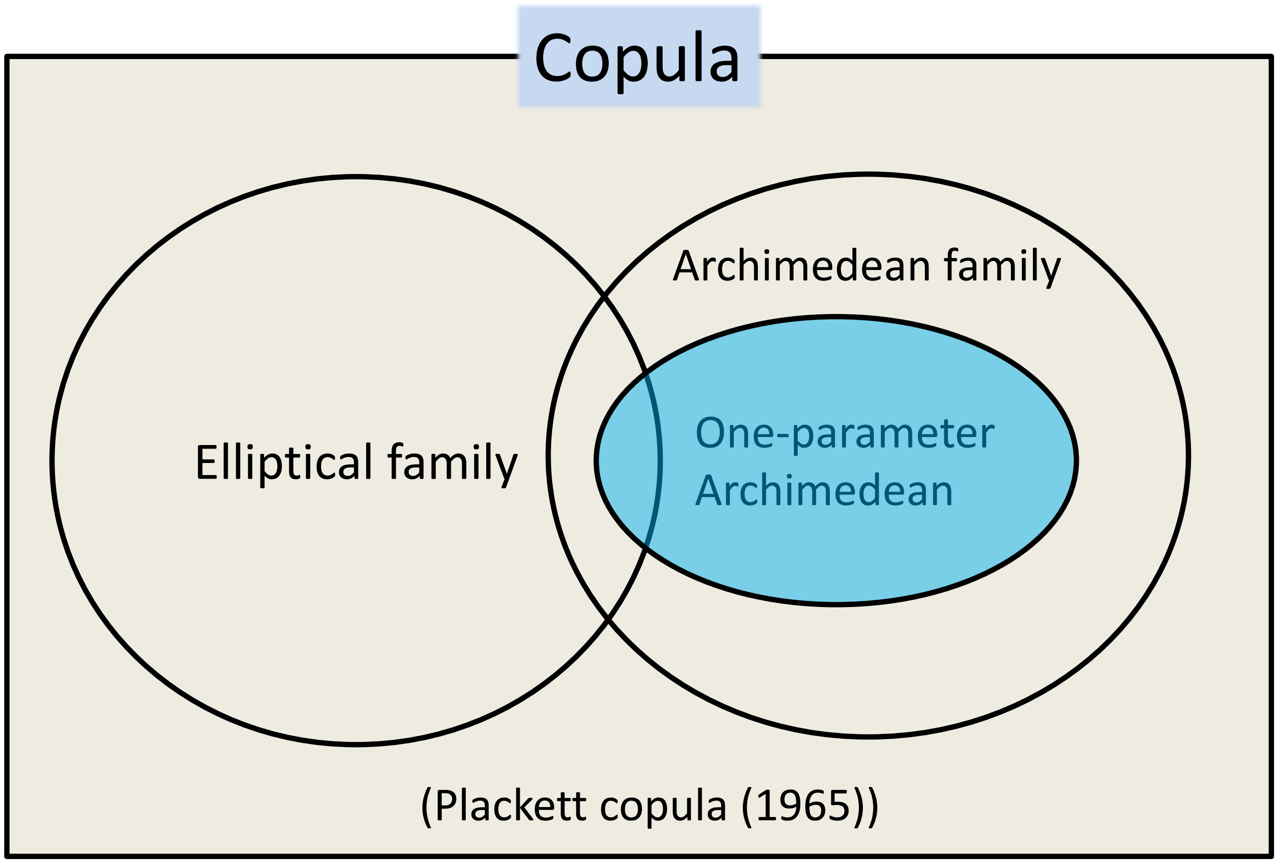
Under Clayton copula

$\hat{F}_X(x) = e^{-\hat{H}_X(x)}$: Time from infection to AIDS (month)



Proposed method

Major advantage: NPMLE can fit more copulas not restricted within one-parameter Aarchimedean family; Model selection among broader copulas



Summary: Proposed method

- We proposed NPMLE for dependent truncation data
- The application of the reverse-time hazard function and semi-survival copula is the key to have a Breslow-Aalen type formula
- The NPMLE can fit broad class of copula and easily adjust for ties
- SD can be estimated by the inverse Fisher information, which is confirmed by simulations
- Under negative correlation, NPMLE worked better than the Emura et al. (2011) and Chaieb et al. (2006)
- NPMLE is computationally demanding (drawback)

Further research

- Theory & simulations for the proposed model selection

$$2\{l_n(\hat{\alpha}, \hat{H}_X, \hat{\Lambda}_Y) - l_n(1, \hat{H}_X^{\alpha=1}, \hat{\Lambda}_Y^{\alpha=1})\} \quad \sim \quad \chi^2_{df=1}$$

1. Proof
2. Numerical comparison with the model selection proposed by Beaudoin & Lakhal-Chaieb (2008, stat. in med.)

- Computational demands for elliptical copulas

Numerical techniques to non-closed form copulas ?

- Regression under dependent truncation

I am currently working on the accelerated failure time regression of the form

$$Y = \beta' Z + \gamma X + \varepsilon$$

Thank you for your kind attention