

國立高雄大學 統計學研究所

2014 / 10 / 15

**An Algorithm for Estimating Survival under a
Copula-Based Dependent Truncation Model:
with Applications to Elderly Human's Survival**

Takeshi Emura

國立中央大學, 統計研究所

(Joint work with K. Murotani 名古屋大学医学部)

Outlines

- Truncated survival data - review
- Product-limit estimator - review
- Copula-based estimator - review
- Proposed algorithm
- Data analysis
- Conclusion

Survival of elderly human

Shigechiyo Isumi

(泉重千代, Japan)

120 Year old when he die

Record in:

The National Oldest-old
Survivors List and Population
Movement Statistics by the
Ministry of Health and Labor
in Japan

(全国高齢者名簿)

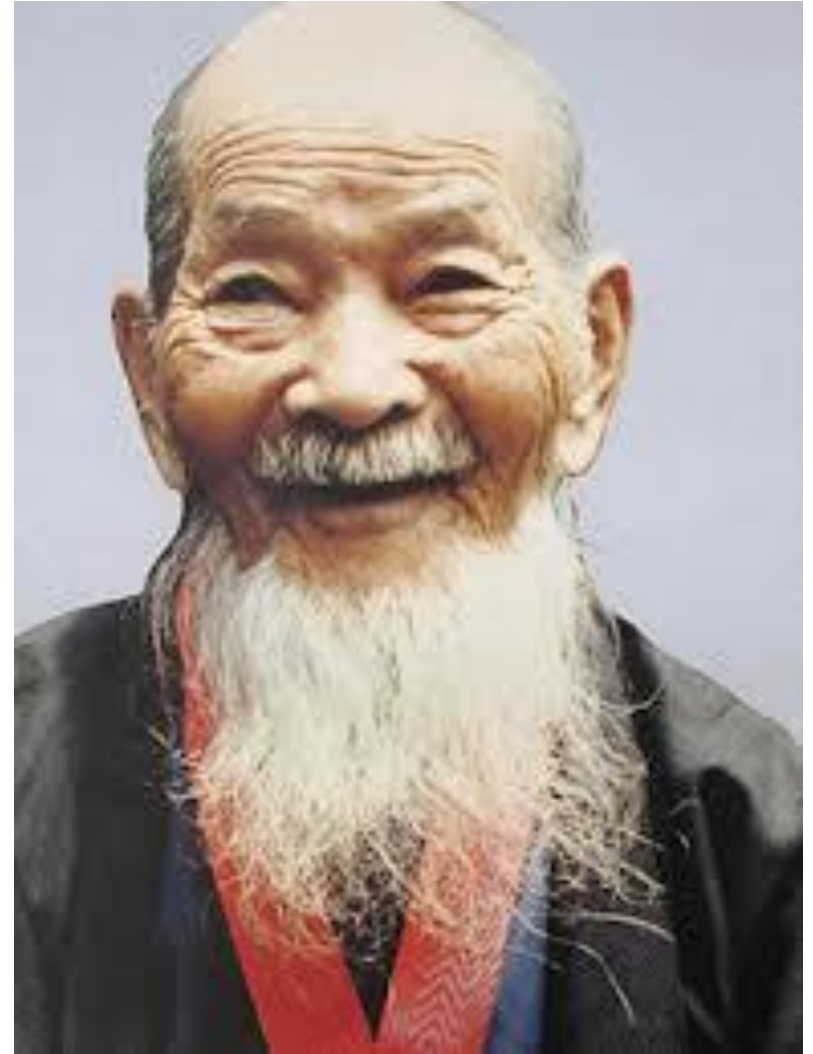


表 2. 年齢時代別 100 歳以上男性生存者数. 全国高齢者名簿より.

生存者数 (男)	100 歳	101 歳	102 歳	103 歳	104 歳	105 歳	106 歳	107 歳	108 歳	109 歳	110 歳
1964 年	19	1	5	5	-	-	-	-	-	-	-
1965 年	24	7	1	1	2	-	-	-	-	-	-
1966 年	22	15	4	1	1	2	-	-	-	-	-
1967 年	26	10	12	2	1	-	-	-	-	-	-
1968 年	37	14	6	7	2	-	-	-	-	-	-
1969 年	26	25	8	6	3	2	-	-	-	-	-
1970 年	22	15	15	4	4	1	1	-	-	-	-
1971 年	31	11	10	10	4	2	1	1	-	-	-
1972 年	41	15	6	5	6	1	3	1	-	-	-
1973 年	56	27	8	4	2	5	-	-	1	-	-
1974 年	44	28	16	1	2	2	2	-	-	1	-
1975 年	43	23	18	11	1	2	1	2	-	-	1
1976 年	57	22	12	11	8	1	1	-	-	-	-
1977 年	56	31	13	6	8	5	1	1	-	-	-
1978 年	68	31	16	5	4	2	4	-	1	-	-
1979 年	101	45	21	4	2	3	-	3	-	-	-
1980 年	70	52	26	12	1	-	2	-	1	-	-
1981 年	98	47	32	16	4	1	-	2	-	1	-
1982 年	108	60	27	23	9	2	1	-	2	-	-
1983 年	132	68	33	18	14	2	1	-	-	-	-
1984 年	178	81	48	20	10	8	-	1	-	-	-
1985 年	157	102	53	27	8	6	4	-	1	-	-
1986 年	168	97	50	26	12	4	3	1	-	-	-
1987 年	225	108	65	32	20	6	2	3	1	-	-
1988 年	298	134	63	31	18	11	5	1	-	1	-
1989 年	302	171	80	40	13	11	7	4	1	-	1
1990 年	314	194	93	46	19	6	3	4	1	-	-
1991 年	373	183	100	47	27	14	4	-	1	-	-
1992 年	388	223	101	60	26	13	8	3	-	-	-
1993 年	430	243	143	62	35	15	8	6	1	-	-
1994 年	509	282	148	84	30	20	12	4	3	1	-
1995 年	640	302	158	78	47	13	8	4	2	2	1
1996 年	619	412	188	89	47	26	8	6	2	2	-
1997 年	722	385	265	96	52	25	18	4	2	-	1
1998 年	837	457	257	153	54	25	17	7	2	2	-
1999 年	885	495	292	157	86	30	12	9	4	1	2
2000 年	945	566	305	169	93	49	14	7	6	3	-
2001 年	1173	623	361	181	108	53	25	9	2	3	2
2002 年	1265	747	401	229	106	65	34	16	5	1	3

	111 歳	112 歳	113 歳	114 歳	115 歳	116 歳	117 歳	118 歳	119 歳	120 歳
1964 年	-	1	-	-	-	-	-	-	-	-
1965 年	-	-	1	-	-	-	-	-	-	-
1966 年	-	-	-	1	-	-	-	-	-	-
1967 年	-	-	-	-	1	-	-	-	-	-
1968 年	-	-	-	-	-	1	-	-	-	-
1969 年	-	-	-	-	-	-	-	-	-	-
1970 年	-	-	-	-	-	-	-	-	-	-
1971 年	-	-	-	-	-	-	-	-	-	-
1972 年	-	-	-	-	-	-	-	-	-	-
1973 年	-	-	-	-	-	-	-	-	-	-
1974 年	-	-	-	-	-	-	-	-	-	-
1975 年	-	-	-	-	-	-	-	-	-	-
1976 年	1	-	-	-	-	-	-	-	-	-
1977 年	-	1	-	-	-	-	-	-	-	-
1978 年	-	-	1	-	-	-	-	-	-	-
1979 年	-	-	-	1	-	-	-	-	-	-
1980 年	-	-	-	-	1	-	-	-	-	-
1981 年	-	-	-	-	-	1	-	-	-	-
1982 年	-	-	-	-	-	-	1	-	-	-
1983 年	-	-	-	-	-	-	-	1	-	-
1984 年	-	-	-	-	-	-	-	-	1	-
1985 年	-	-	-	-	-	-	-	-	-	1
1986 年	-	-	-	-	-	-	-	-	-	-
1987 年	-	-	-	-	-	-	-	-	-	-
1988 年	-	-	-	-	-	-	-	-	-	-

Sibuya M, Hanayama N (2004)

Estimation of human longevity distribution based on tabulated statistics.

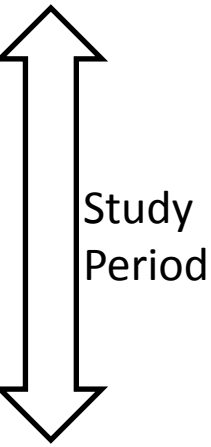
Proceeding of AISM 52: 117-34

Here is 泉重千代

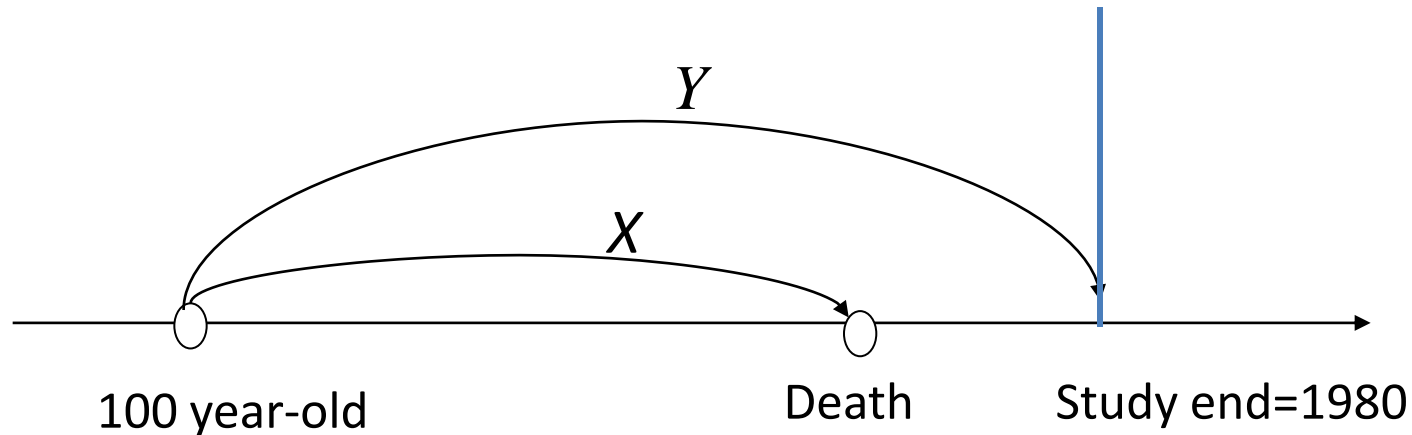
年齢時代区分データによる超高齢者寿命分布の推測

表 2. 年齢時代別 100 歳以上男性生存者数. 全国高齢者名簿より.

生存者数 (男)	100 歳	101 歳	102 歳	103 歳	104 歳	105 歳	106 歳	107 歳	108 歳	109 歳	110 歳
1964 年	19	1	5	5	-	-	-	-	-	-	-
1965 年	24	7	1	1	2	-	-	-	-	-	-
1966 年	22	15	4	1	1	2	-	-	-	-	-
1967 年	26	10	12	2	1	-	-	-	-	-	-
1968 年	37	14	6	7	2	-	-	-	-	-	-
1969 年	26	25	8	6	3	2	-	-	-	-	-
1970 年	22	15	15	4	4	1	1	-	-	-	-
1971 年	31	11	10	10	4	2	1	1	-	-	-
1972 年	41	15	6	5	6	1	3	1	-	-	-
1973 年	56	27	8	4	2	5	-	-	1	-	-
1974 年	44	28	16	1	2	2	2	-	-	1	-
1975 年	43	23	18	11	1	2	1	2	-	-	1
1976 年	57	22	12	11	8	1	1	-	-	-	-
1977 年	56	31	13	6	8	5	1	1	-	-	-
1978 年	68	31	16	5	4	2	4	-	1	-	-
1979 年	101	45	21	4	2	3	-	3	-	-	-
1980 年	70	52	26	12	1	-	2	-	1	-	-
1981 年	98	47	32	16	4	1	-	2	-	1	-
1982 年	108	60	27	23	9	2	1	-	2	-	-
1983 年	132	68	33	18	14	2	1	-	-	-	-
1984 年	178	81	48	20	10	8	-	1	-	-	-
1985 年	157	102	53	27	8	6	4	-	1	-	-
1986 年	168	97	50	26	12	4	3	1	-	-	-
1987 年	225	108	65	32	20	6	2	3	1	-	-
1988 年	298	134	63	31	18	11	5	1	-	1	-
1989 年	302	171	80	40	13	11	7	4	1	-	1
1990 年	314	194	93	46	19	6	3	4	1	-	-
1991 年	373	183	100	47	27	14	4	-	1	-	-
1992 年	388	223	101	60	26	13	8	3	-	-	-
1993 年	430	243	143	62	35	15	8	6	1	-	-
1994 年	509	282	148	84	30	20	12	4	3	1	-
1995 年	640	302	158	78	47	13	8	4	2	2	1
1996 年	619	412	188	89	47	26	8	6	2	2	-
1997 年	722	385	265	96	52	25	18	4	2	-	1
1998 年	837	457	257	153	54	25	17	7	2	2	-
1999 年	885	495	292	157	86	30	12	9	4	1	2
2000 年	945	566	305	169	93	49	14	7	6	3	-
2001 年	1173	623	361	181	108	53	25	9	2	3	2



Right-truncated survival data



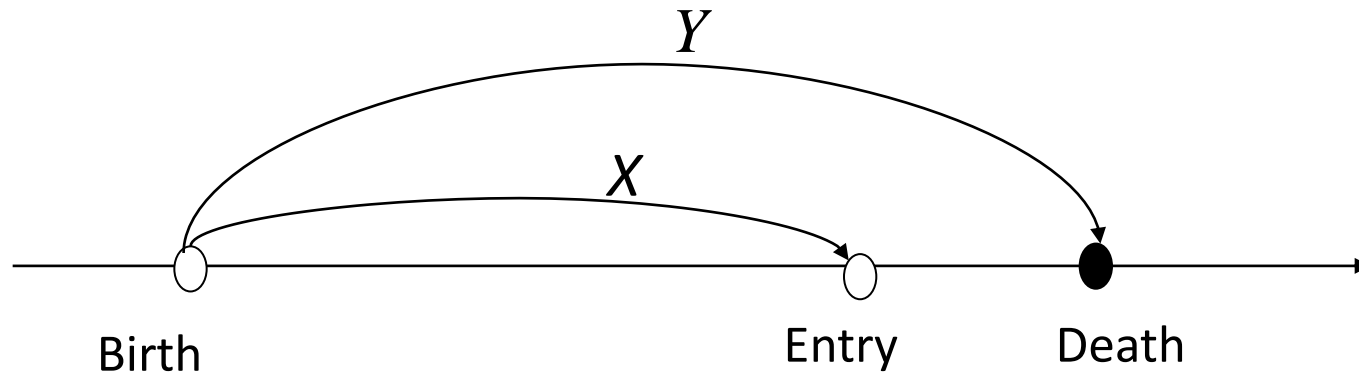
Japanese centenarians data: Sibuya M, Hanayama N (2004)

- $n = 662$ Japanese male centenarians ascertained before year 1980
- Age at death = X
- Age at study end = Y

Truncation criterion:

$$X \leq Y \quad \Rightarrow \quad \text{Sample is available}$$

Left-truncated survival data



Channing House data (Hyde, 1977, 1980)

Channing house is a retirement center in California

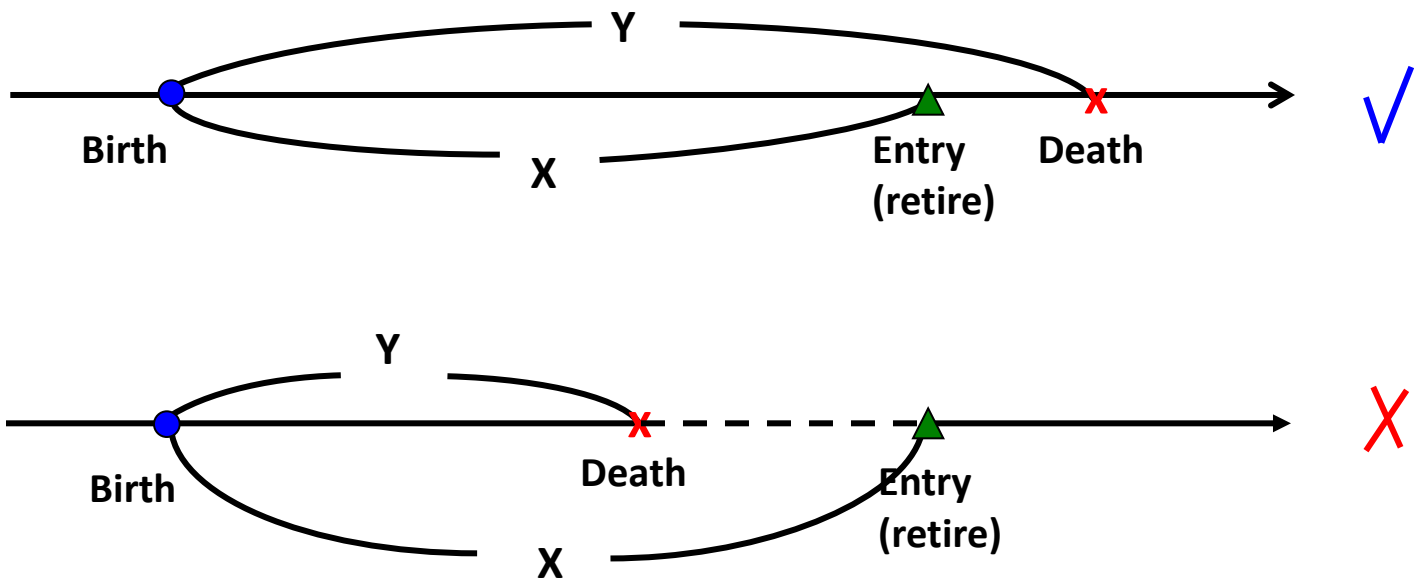
- $n = 97$ elderly residents in the Channing house
- Age at entry = X
- Age at death = Y (possibly right-censored)

Left-truncation criterion:

$$X \leq Y \quad \Rightarrow \quad \text{Sample is available}$$

Left-truncated survival data

- Lifetime $Y \rightarrow$ left-truncated by X



Linear interpolation in the 1958 Commissioners Standard Ordinary Mortality Table for Male Lives was used to generate the distribution F , and hence h_t , for each month. The table stops at age 100, so that the data must be artificially censored at 1200 months. This does not affect the data in Table 1, but it does mean that $E(\lambda^*)$ is finite.

The observed number of deaths was 46, and the expected number was 72.2. The estimated variance was 68.3. The value of the statistic is thus $(46 - 72.2)/\sqrt{68.3} = -3.16$, which indicates that the null hypothesis should be rejected in favour of a smaller hazard rate.

Entry Death

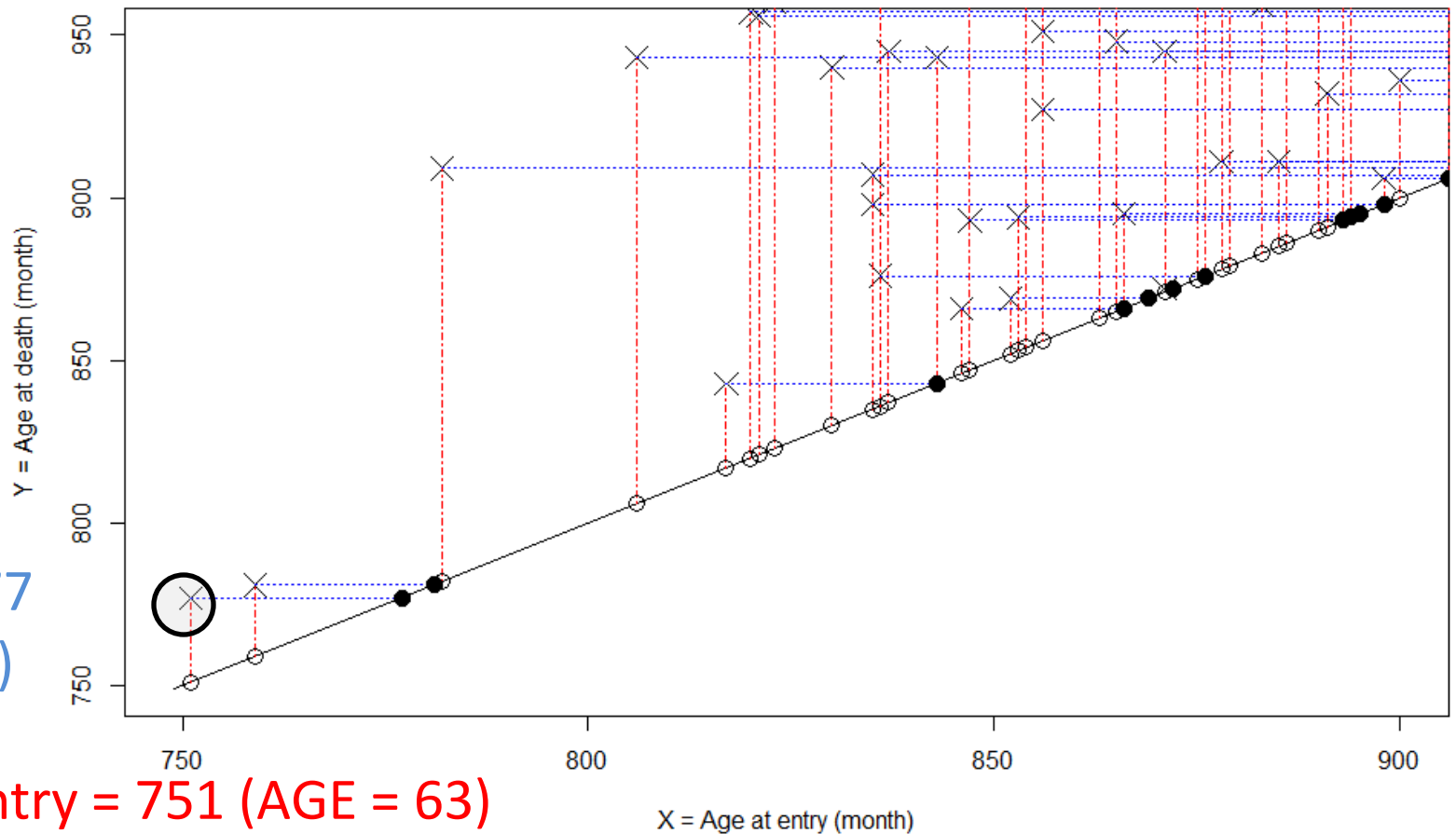
Table 1. *Survival data for males in a retirement community*

δ	$\nu+1$	λ^*	δ	$\nu+1$	λ^*	δ	$\nu+1$	λ^*	δ	$\nu+1$	λ^*	δ	$\nu+1$	λ^*
1	782	909	0	865	1002	1	854	989	0	1016	1153	0	959	972
1	1020	1128	0	953	1031	0	890	1027	0	969	1106	1	921	993
1	856	969	0	871	945	0	1041	1044	0	900	936	1	836	876
1	915	957	0	982	1006	0	978	1005	0	898	906	1	919	993
1	863	983	0	883	959	0	836	973	0	846	866	1	751	777
1	906	1012	0	817	843	0	938	1064	1	964	1029	1	906	966
1	955	1055	0	875	1012	0	886	1023	1	984	1053	1	835	907
1	943	1025	0	821	956	0	876	1013	1	1046	1080	1	946	1031
1	943	1043	0	936	1073	0	955	977	1	871	872	1	759	781
1	837	945	0	971	1107	0	960	1047	1	847	893	0	909	914
1	966	1009	0	830	940	0	843	943	0	962	966	1	962	998
1	936	971	0	885	911	0	856	951	1	853	894	1	984	1022
1	919	1033	0	894	1031	0	847	984	1	967	985	1	891	932
1	852	869	0	893	996	0	1027	1058	1	1063	1094	1	835	898
1	1073	1139	0	866	895	0	988	1045	1	856	927	1	1039	1060
1	925	1036	0	878	1015	0	953	953	1	865	948	1	1010	1044
1	967	1085	0	820	957	0	978	1018	1	1051	1059	0	823	960
0	806	943	0	1007	1043	0	981	1118	1	1010	1012			
0	969	1001	0	879	1016	0	926	970	1	878	911			
0	923	1060	0	956	1093	0	1036	1070	1	1021	1094			

Here $\delta = 1$ if subject died during study, $\delta = 0$ otherwise; $\nu+1$ is 1 + age in months at entry into study; λ^* is age in months when last seen in study.

If the discrete version is viewed as an approximation for the continuous case, and if it is assumed that the actual hazard rate is a multiple c of the hazard rate corresponding to F , then the ideas of § 4 can be applied. A 90 % confidence interval for c is (0.500, 0.812), and the approximate median unbiased estimate of c is 0.637.

I would like to thank Dr Rupert Miller for his support and sound advice, and Dr Bradley Efron for some helpful comments. Dr Walter Bortz kindly permitted me to use the data in



- Hyde (1980) assumed:
knowing the person's entry age will provide no additional information about prospects for survival
- This means $X \perp Y$
(X: Age at entry Y: Age at death)

Left-truncated survival data

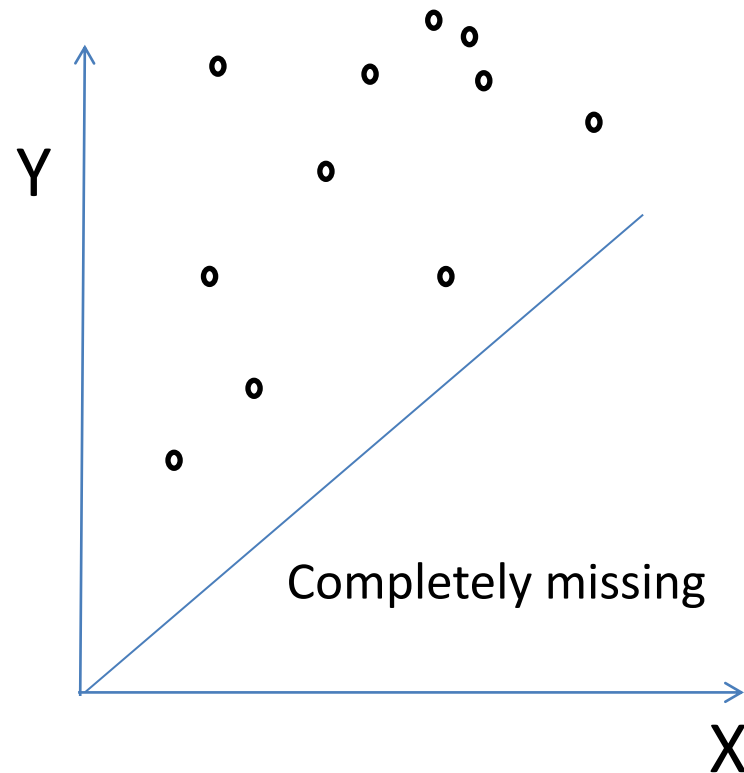
- Left - truncated data

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

subject to $X_j \leq Y_j$

↑↑

i.i.d. from $P(X \leq x, Y \leq y | X \leq Y)$



- Quasi - independence assumption (Tsai, 1990)

$$H_0 : \Pr(X = x, Y = y | X \leq Y) \propto dF_X(x) dF_Y(y)$$

Estimating survival

• Target : $S_Y(t) = P(Y > t)$

• Product - limit representation

$$S_Y(t) = \prod_{u \leq t} \{ 1 - P(Y = u | Y \geq u) \}$$

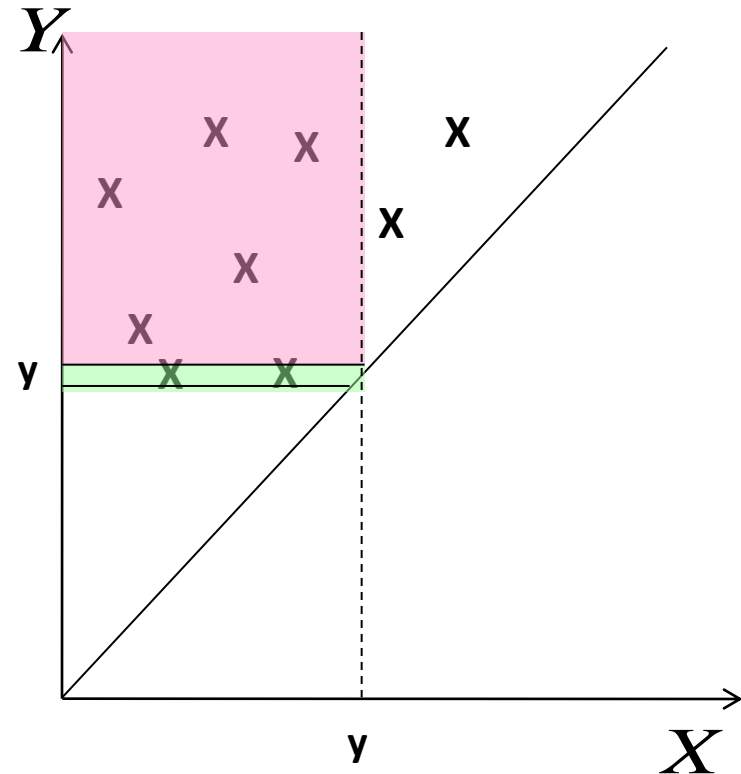
• Hazard function

$$P(Y = u | Y \geq u) = \frac{P(Y = u)}{P(Y \geq u)}$$

↪

$$= \frac{P(Y = u, X \leq u)}{P(Y \geq u, X \leq u)}$$

Quasi-independence



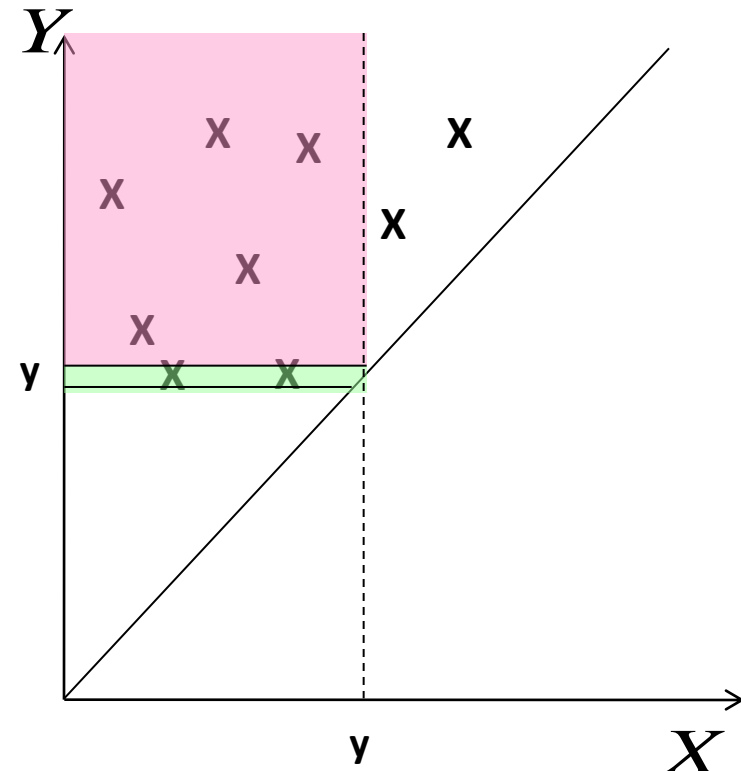
Estimating survival

- Product-limit estimator of $S_Y(y)$
(Lynden-Bell 1971)

$$\hat{S}_Y(y) = \prod_{u \leq y} \left\{ 1 - \frac{\sum_{i=1}^n I(X_i < u, Y_i = u)}{\sum_{i=1}^n I(X_i < u, Y_i \geq u)} \right\}$$

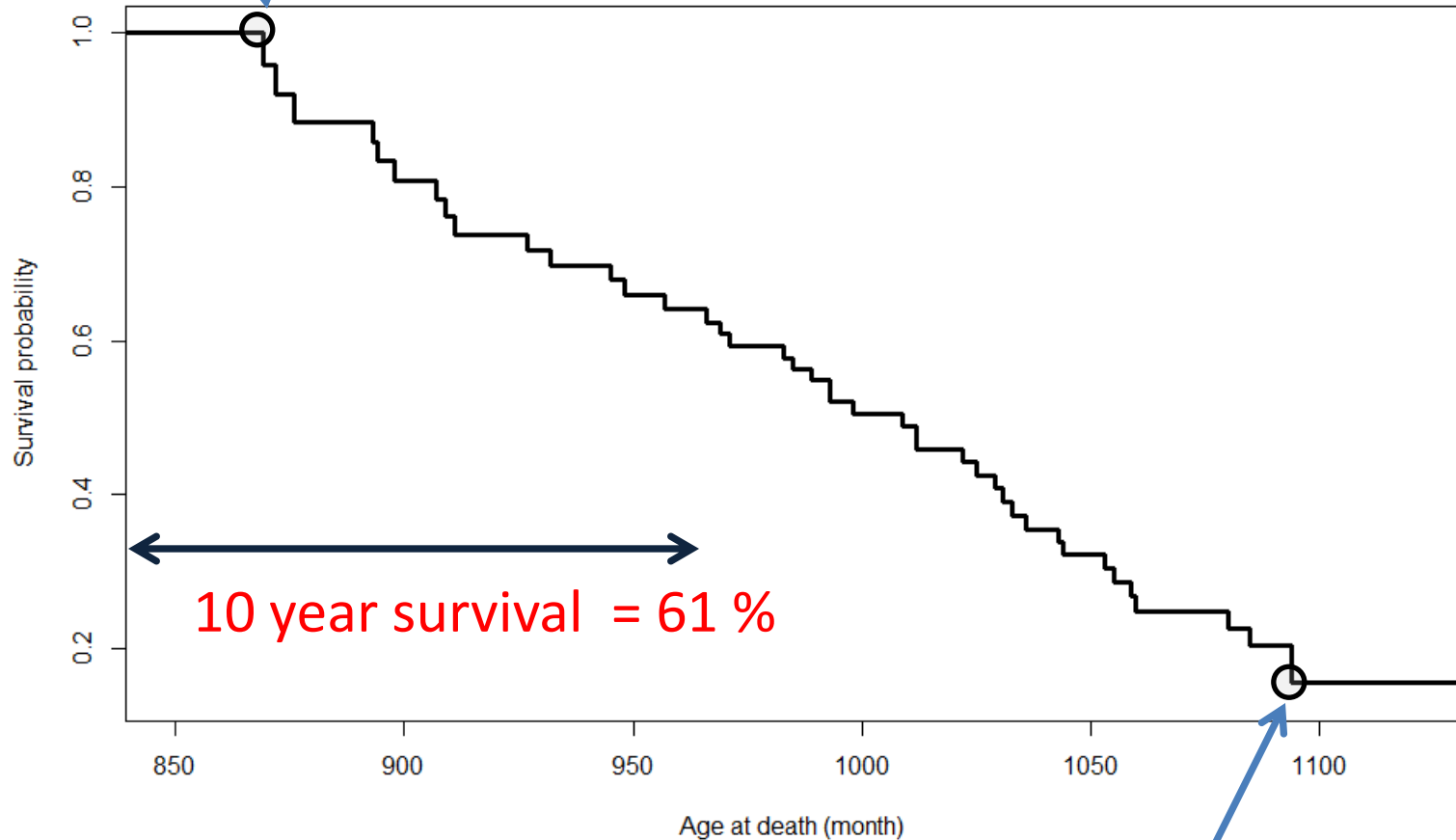


$$S_Y(t) = \prod_{u \leq t} \{ 1 - P(Y = u | Y \geq u) \}$$



Product-limit Estimates of $S_Y(y)$

869 month
(AGE = 72)



1094 month
(AGE = 91)

Testing quasi-independence

Testing quasi-independence

$$H_0 : \Pr(X = x, Y = y | X \leq Y) \propto dF_X(x)dF_Y(y)$$

Available test statistics:

1. [Chen et al. \(1996 JASA\)](#)

- Based on the conditional Pearson-correlation

2. [Tsai \(1990 Biometrika\)](#); [Martin & Betensky \(2005 JASA\)](#)

-Based on the conditional Kendall's tau

3. [Emura & Wang \(2010 JMVA\)](#) - Based on weighted-logrank test

(Optimal weight choice)

Testing quasi-independence

$$H_0 : \Pr(X = x, Y = y \mid X \leq Y) \propto dF_X(x)dF_Y(y)$$

is rejected at 5 % level

Table 4 of Emura and Wang (2010).

Tests of quasi-independence for the Channing House data.

	Logrank test	Tsai test	Marting & Betensky test
	$L_{\rho=1}$		
P-value	0.048	0.043	0.040

- Quasi-independence is questionable
 - Product-limit estimates of survival probability may be biased
- In Channing house data, the truncation (entry age) may be dependent on survival.
 - Motivate Copula modeling for dependent truncation

Chaieb, Rivest, Abdous (2006, Biometrika)

Beaudoin and Lakhal-Chaieb (2008 Stat. Med.),

Emura and Wang (2010, JMVA)

Emura, Wang and Hung (2011 Sinica)

Emura and Wang (2012, JMVA)

Ding (2012 Lifetime Data Analysis)

Copula model for dependent truncation

Let $\pi(x, y) \equiv \Pr(X \leq x, Y > y | X \leq Y)$

- Copula model (Chaieb et al, 2006):

$$\pi(x, y)$$

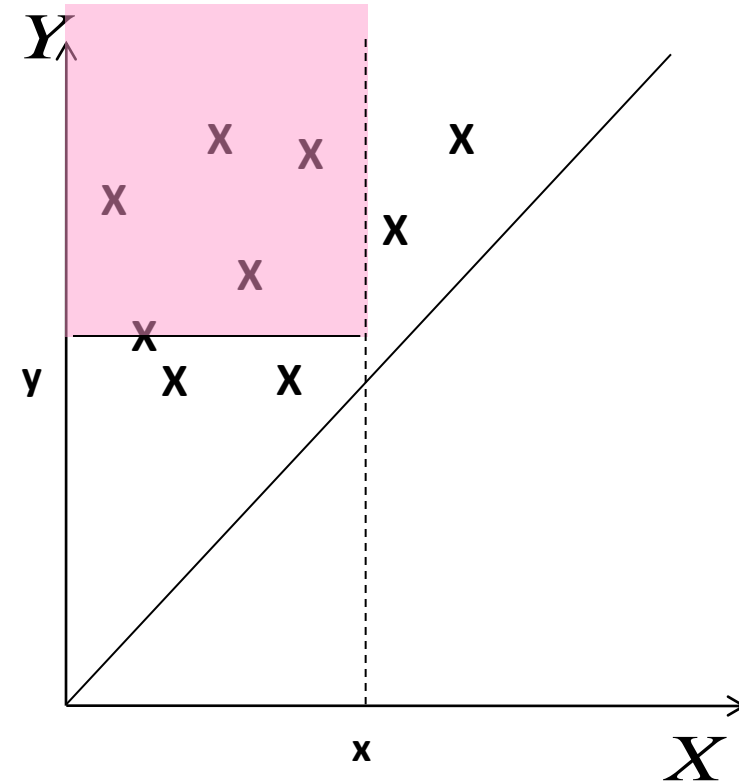
$$= \phi_\alpha^{-1} [\phi_\alpha \{F_X(x)\} + \phi_\alpha \{S_Y(y)\}] / c$$

- Clayton copula:

$$\phi_\alpha(t) = (t^{-(\alpha-1)} - 1) / (\alpha - 1)$$

$$\Rightarrow \Pr(X \leq x, Y > y | X \leq Y)$$

$$= (1/c) [F_X(x)^{-(\alpha-1)} + S_Y(y)^{-(\alpha-1)} - 1]^{-\frac{1}{\alpha-1}}$$



Copula model for dependent truncation

$$\pi(x, y) \equiv \Pr(X \leq x, Y > y | X \leq Y)$$

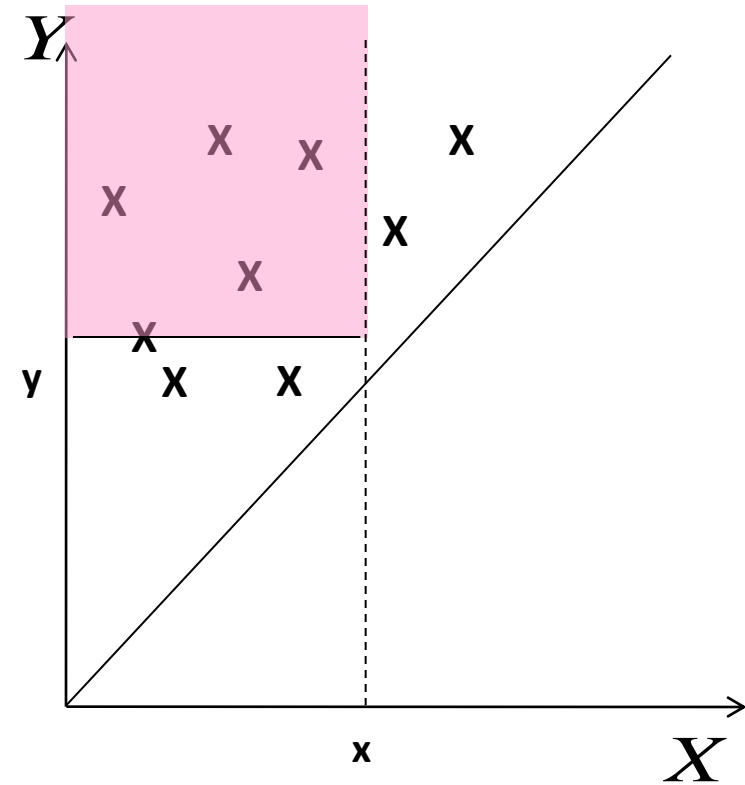
$$\pi(x, y)$$

$$= \phi_\alpha^{-1} [\phi_\alpha \{F_X(x)\} + \phi_\alpha \{S_Y(y)\}] / c$$

Chaieb et al (2006): Plug in

$$\hat{\pi}(x, y) \equiv \frac{1}{n} \sum_{j=1}^n \mathbf{I}(X_j \leq x, Y_j > y)$$

⇒ Get the estimator of (α, c, F_X, S_Y)



Estimating equation

$$\hat{\pi}(t, t) = \phi_\alpha^{-1} [\phi_\alpha \{F_X(t)\} + \phi_\alpha \{S_Y(t)\}] / c$$
$$\Leftrightarrow \phi_\alpha(c\hat{\pi}(t, t)) = \phi_\alpha(F_X(t)) + \phi_\alpha(S_Y(t))$$

Chaieb et al. (2006) use some algebraic techniques of Rivest and Wells (2001, JMVA) to get solutions:

$$\hat{S}_Y(t) = \phi_\alpha^{-1} \left(- \sum_{j: Y_j \leq t} \left[\phi_\alpha \left\{ c \frac{\tilde{R}(Y_j)}{n} \right\} - \phi_\alpha \left\{ c \frac{\tilde{R}(Y_j) - 1}{n} \right\} \right] \right)$$
$$\hat{F}_X(t) = \phi_\alpha^{-1} \left(- \sum_{j: X_j > t} \left[\phi_\alpha \left\{ c \frac{\tilde{R}(X_j)}{n} \right\} - \phi_\alpha \left\{ c \frac{\tilde{R}(X_j) - 1}{n} \right\} \right] \right)$$

Estimating equation

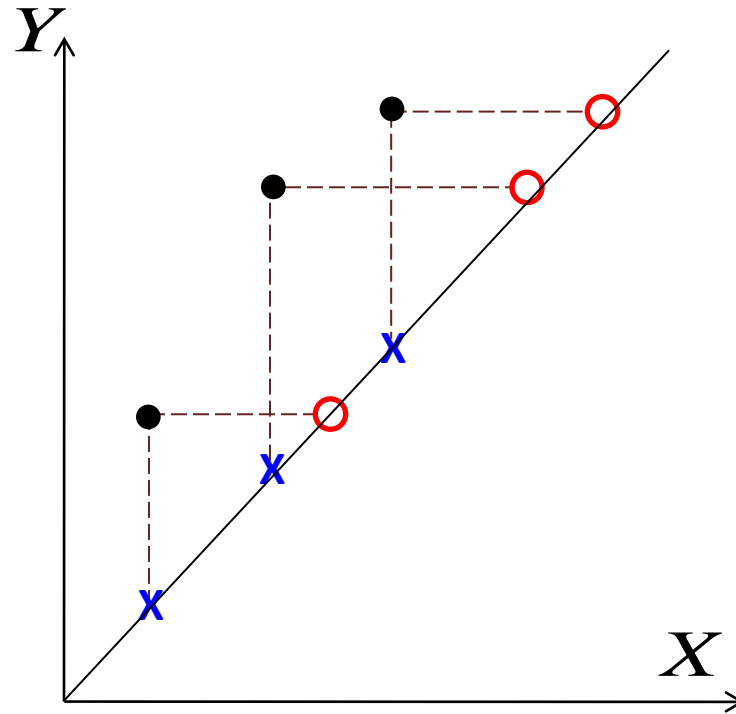
$$\hat{\pi}(t, t) = \phi_{\alpha}^{-1}[\phi_{\alpha}\{F_X(t)\} + \phi_{\alpha}\{S_Y(t)\}] / c$$
$$\Leftrightarrow \phi_{\alpha}(c\hat{\pi}(t, t)) = \phi_{\alpha}(F_X(t)) + \phi_{\alpha}(S_Y(t))$$

In this research:

I propose an new algorithm to solve the estimating equation

- New algorithm is easier to understand (straightforward derivation)
- New algorithm yields the same solution as Chaieb et al. (2006)

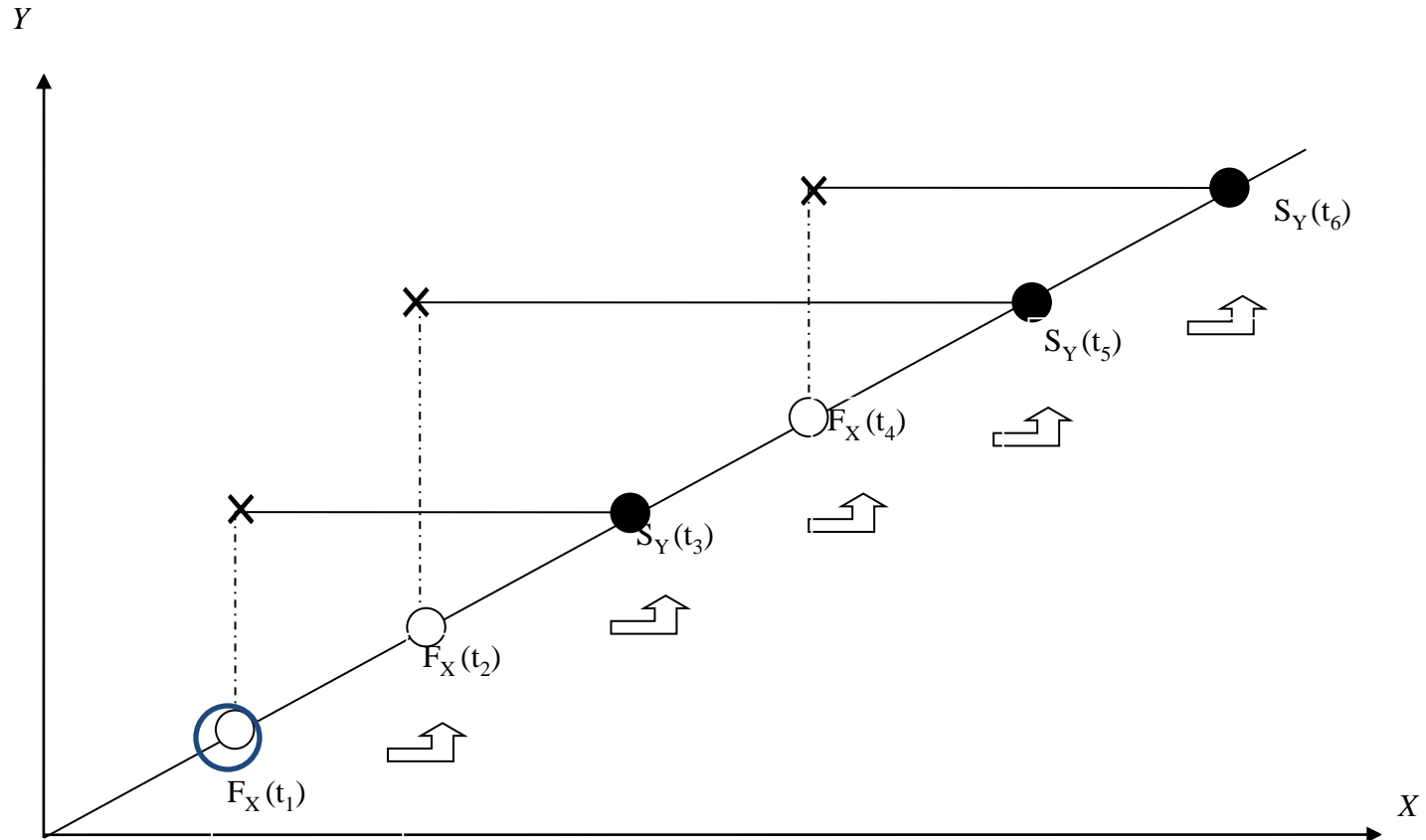
Proposed algorithm



$$(X_1, \dots, X_n, Y_1, \dots, Y_n) \Rightarrow \underbrace{t_1}_{1^{\text{st}} \text{ Entry}} < \dots < t_{2n-1} < \underbrace{t_{2n}}_{\text{Last die}}$$

1st Entry

Solving: $\phi_\alpha(c\hat{\pi}(t,t)) = \phi_\alpha(F_X(t)) + \phi_\alpha(S_Y(t))$



At $t_1 = 1^{\text{st}}$ Entry, nobody die $\rightarrow S_Y(t_1) = 1$

$$\rightarrow \phi_\alpha(F_X(t_1)) = \phi_\alpha(c\hat{\pi}(t_1, t_1)) - \phi_\alpha(S_Y(t_1)) = \phi_\alpha(c/3) - \phi_\alpha(1) = \phi_\alpha(c/3)$$

$$\therefore F_X(t_1) = c/3$$

Table 1: Results of performing Step 1 of the proposed algorithm for a small dataset:

$(X_1, Y_1) = (1, 3)$, $(X_2, Y_2) = (2, 5)$ and $(X_3, Y_3) = (4, 6)$.

1 st Entry	$\hat{\pi}(t_j, t_j)$	$\phi_\alpha\{c\hat{\pi}(t_j, t_j)\}$	$\phi_\alpha\{\hat{F}_X(t_j)\}$	$\phi_\alpha\{\hat{S}_Y(t_j)\}$
$t_1 = X_1 = 1$	$\frac{1}{3}$	$\phi_\alpha\left(\frac{c}{3}\right)$	$\phi_\alpha\left(\frac{c}{3}\right)$	0
$t_2 = X_2 = 2$	$\frac{2}{3}$	$\phi_\alpha\left(\frac{2c}{3}\right)$	$\phi_\alpha\left(\frac{2c}{3}\right)$	0
$t_3 = Y_1 = 3$	$\frac{1}{3}$	$\phi_\alpha\left(\frac{c}{3}\right)$	$\phi_\alpha\left(\frac{2c}{3}\right)$	$\phi_\alpha\left(\frac{c}{3}\right) - \phi_\alpha\left(\frac{2c}{3}\right)$
$t_4 = X_3 = 4$	$\frac{2}{3}$	$\phi_\alpha\left(\frac{2c}{3}\right)$	$2\phi_\alpha\left(\frac{2c}{3}\right) - \phi_\alpha\left(\frac{c}{3}\right)$	$\phi_\alpha\left(\frac{c}{3}\right) - \phi_\alpha\left(\frac{2c}{3}\right)$
$t_5 = Y_2 = 5$	$\frac{1}{3}$	$\phi_\alpha\left(\frac{c}{3}\right)$	$2\phi_\alpha\left(\frac{2c}{3}\right) - \phi_\alpha\left(\frac{c}{3}\right)$	$2\phi_\alpha\left(\frac{c}{3}\right) - 2\phi_\alpha\left(\frac{2c}{3}\right)$
Last die $t_6 = Y_3 = 6$	$\frac{0}{3}$	$\phi_\alpha(0)$	Undetermined	Undetermined

Solution of the proposed algorithm

- 1) Under the Clayton copula $\phi_\alpha(t) = (t^{-(\alpha-1)} - 1)/(\alpha - 1)$

$$\hat{F}_X(X_1) = 1/3, \quad \hat{F}_X(X_2) = 2/3, \quad \hat{F}_X(X_3) = 1,$$

$$\hat{S}_Y(Y_{(1)}) = 2/3, \quad \hat{S}_Y(Y_{(2)}) = 1/3, \quad \hat{S}_Y(Y_{(3)}) = \text{undetermined.}$$

- 2) Under the quasi-independence $\phi_\alpha(t) = -\log(t)$

$$\hat{F}_X(X_1) = 1/4, \quad \hat{F}_X(X_2) = 1/2, \quad \hat{F}_X(X_3) = 1,$$

$$\hat{S}_Y(Y_{(1)}) = 1/2, \quad \hat{S}_Y(Y_{(2)}) = 1/4, \quad \hat{S}_Y(Y_{(3)}) = \text{undetermined.}$$

Proposed method

Proposed algorithm → Explicit formula

$$\hat{S}_Y(t) = \phi_\alpha^{-1} \left(- \sum_{j: Y_j \leq t} \left[\phi_\alpha \left\{ c \frac{\tilde{R}(Y_j)}{n} \right\} - \phi_\alpha \left\{ c \frac{\tilde{R}(Y_j) - 1}{n} \right\} \right] \right)$$

$$\hat{F}_X(t) = \phi_\alpha^{-1} \left(\sum_{j: t_1 < X_j \leq t} \left[\phi_\alpha \left\{ c \frac{\tilde{R}(X_j)}{n} \right\} - \phi_\alpha \left\{ c \frac{\tilde{R}(X_j) - 1}{n} \right\} \right] + \phi_\alpha \left(\frac{c}{n} \right) \right)$$

Proposed vs. Chaieb et al.

1) Proposed

$$\hat{S}_Y(t) = \phi_\alpha^{-1} \left(- \sum_{j: Y_j \leq t} \left[\phi_\alpha \left\{ c \frac{\tilde{R}(Y_j)}{n} \right\} - \phi_\alpha \left\{ c \frac{\tilde{R}(Y_j) - 1}{n} \right\} \right] \right)$$

$$\hat{F}_X(t) = \phi_\alpha^{-1} \left(\sum_{j: t_1 < X_j \leq t} \left[\phi_\alpha \left\{ c \frac{\tilde{R}(X_j)}{n} \right\} - \phi_\alpha \left\{ c \frac{\tilde{R}(X_j) - 1}{n} \right\} \right] + \phi_\alpha \left(\frac{c}{n} \right) \right)$$

2) Chaieb et al. (2006)

$$\hat{S}_Y(t) = \phi_\alpha^{-1} \left(- \sum_{j: Y_j \leq t} \left[\phi_\alpha \left\{ c \frac{\tilde{R}(Y_j)}{n} \right\} - \phi_\alpha \left\{ c \frac{\tilde{R}(Y_j) - 1}{n} \right\} \right] \right)$$

$$\hat{F}_X(t) = \phi_\alpha^{-1} \left(- \sum_{j: X_j > t} \left[\phi_\alpha \left\{ c \frac{\tilde{R}(X_j)}{n} \right\} - \phi_\alpha \left\{ c \frac{\tilde{R}(X_j) - 1}{n} \right\} \right] \right)$$

Two estimators (Proposed vs. Chaieb et al.) are the same

Theorem 1: The proposed estimating equation (10) is equivalent to the estimating equation (6) of Chaieb et al. (2006) under $x_0 \in [X_{(n)}, t_{2n-1}]$.

Proof: Note that $\phi_\alpha\{\hat{F}_X(t_{2n-1})\} = \phi_\alpha\{\hat{F}_X(x_0)\}$ since there is no jump for X beyond $x_0 \in [X_{(n)}, t_{2n-1}]$. Thus, the estimating equation (10) becomes

$$\begin{aligned} U_c(\alpha, c) &= \phi_\alpha\{\hat{F}_X(x_0)\} \\ &= \phi_\alpha\left\{c \frac{R(x_0, x_0+)}{n}\right\} - \phi_\alpha\{S_Y(x_0)\} \\ &= \phi_\alpha\left\{c \frac{R(x_0, x_0+)}{n}\right\} + \sum_{j: Y_j \leq x_0} \left[\phi_\alpha\left\{c \frac{\tilde{R}(Y_j)}{n}\right\} - \phi_\alpha\left\{c \frac{\tilde{R}(Y_j) - 1}{n}\right\} \right] \\ &= \phi_\alpha\left\{c \frac{\tilde{R}(x_0)}{n}\right\} + \sum_{j: Y_j < x_0} \left[\phi_\alpha\left\{c \frac{\tilde{R}(Y_j)}{n}\right\} - \phi_\alpha\left\{c \frac{\tilde{R}(Y_j) - 1}{n}\right\} \right], \end{aligned}$$

where the last equation is equivalent to Equation (6). \square

Extension for right-censoring

- Left - truncation + Right - censoring

$Z_j = \min(Y_j, C_j) = \min(\text{Age at death}, \text{withdrawal})$

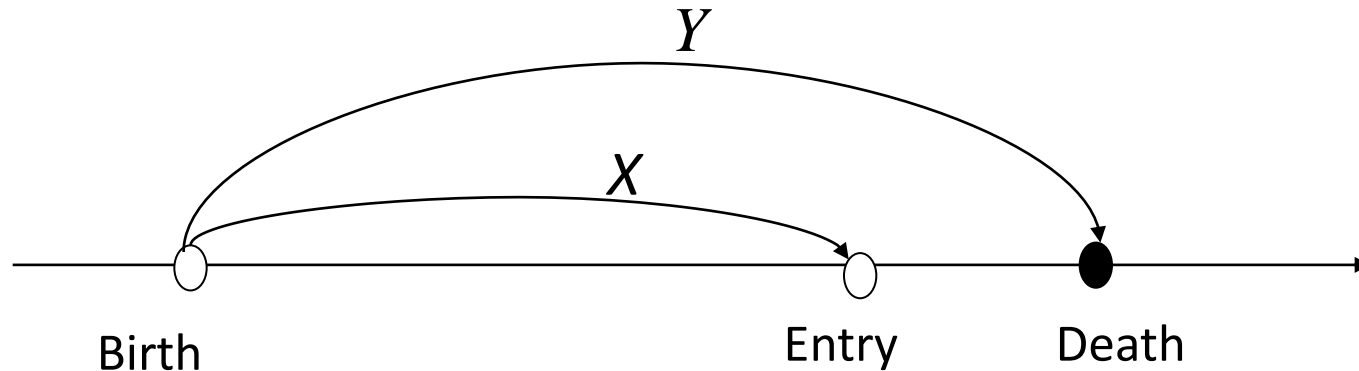
$X_j = \text{Entry age}$

$$\delta_j = \begin{cases} 1 & \text{die during the study : } Y_j \leq C_j \\ 0 & \text{withdraw from the study: } Y_j > C_j \end{cases}$$

Data: $\{(X_j, Z_j, \delta_j); j = 1, \dots, n\}$ subject to $X_j \leq Z_j$

$$\Rightarrow \hat{S}_Y(t) = \phi_\alpha^{-1} \left(- \sum_{j; z_j \leq t, \delta_j = 1} \left[\phi_\alpha \left\{ c^* \frac{\tilde{R}(Z_j)}{n\hat{S}_C(Z_j)} \right\} - \phi_\alpha \left\{ c^* \frac{\tilde{R}(Z_j) - 1}{n\hat{S}_C(Z_j)} \right\} \right] \right)$$

Data analysis (I)



Channing House data (Hyde, 1977, 1980)

$n = 97$ elderly residents in the Channing house

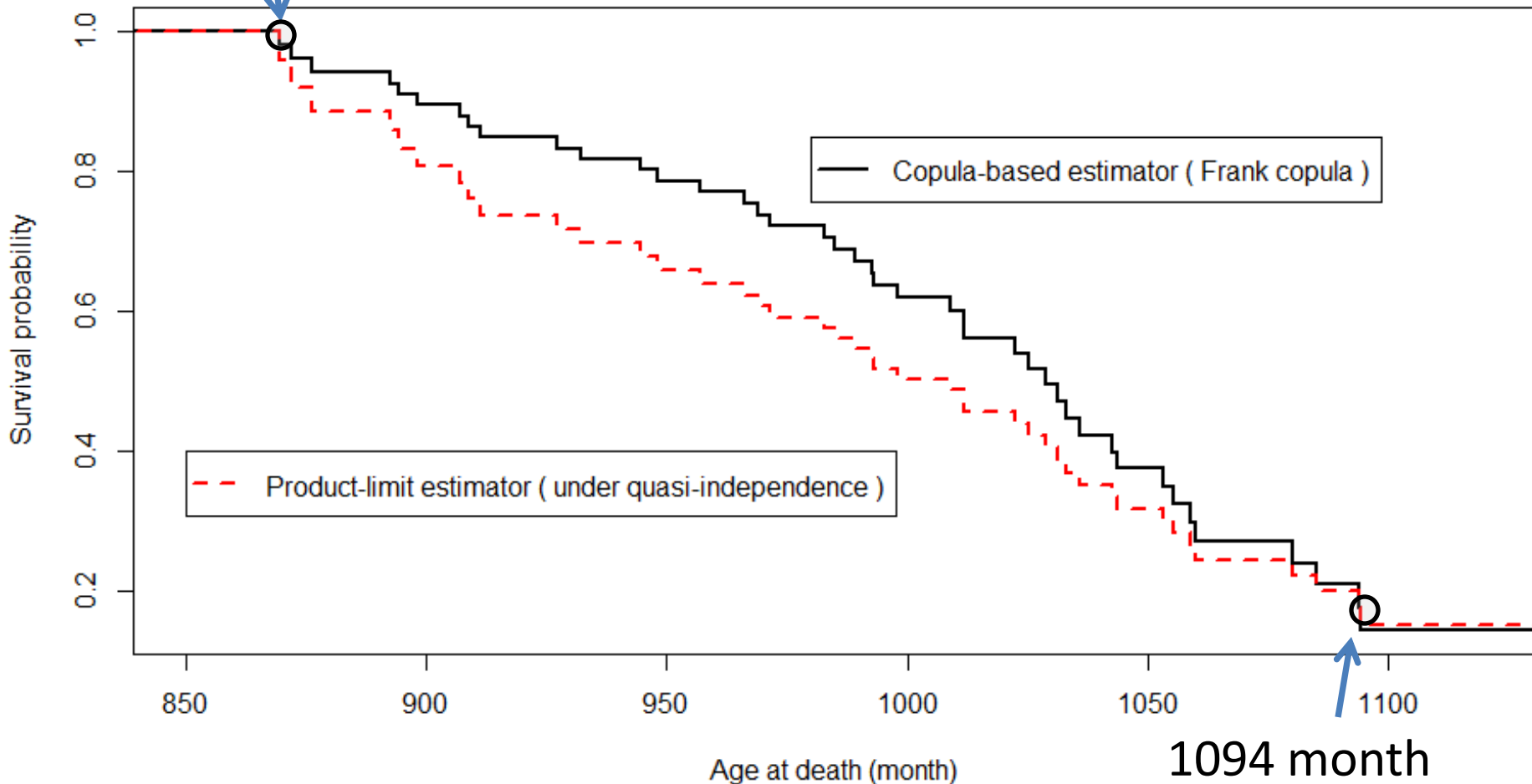
- Age at entry = X
- Age at death = $Z = \min(Y, C)$
(possibly right-censored by withdrawal)

Data:

$(X_j, Z_j, \delta_j); j = 1, \dots, n$ subject to $X_j \leq Z_j$

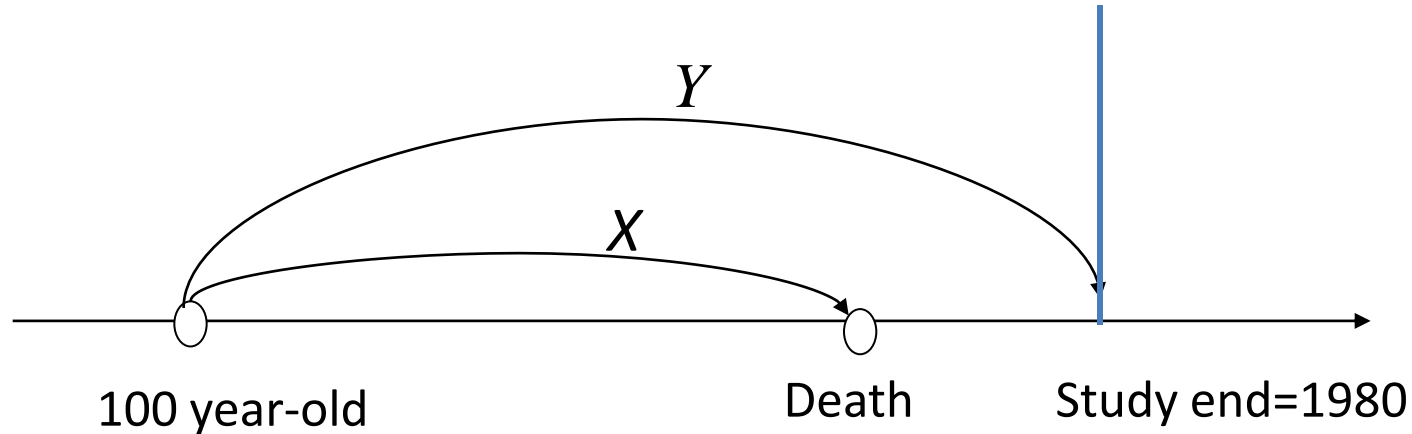
Estimated survival function

869 month
(AGE = 72)



1094 month
(AGE = 91)

Data analysis (II)



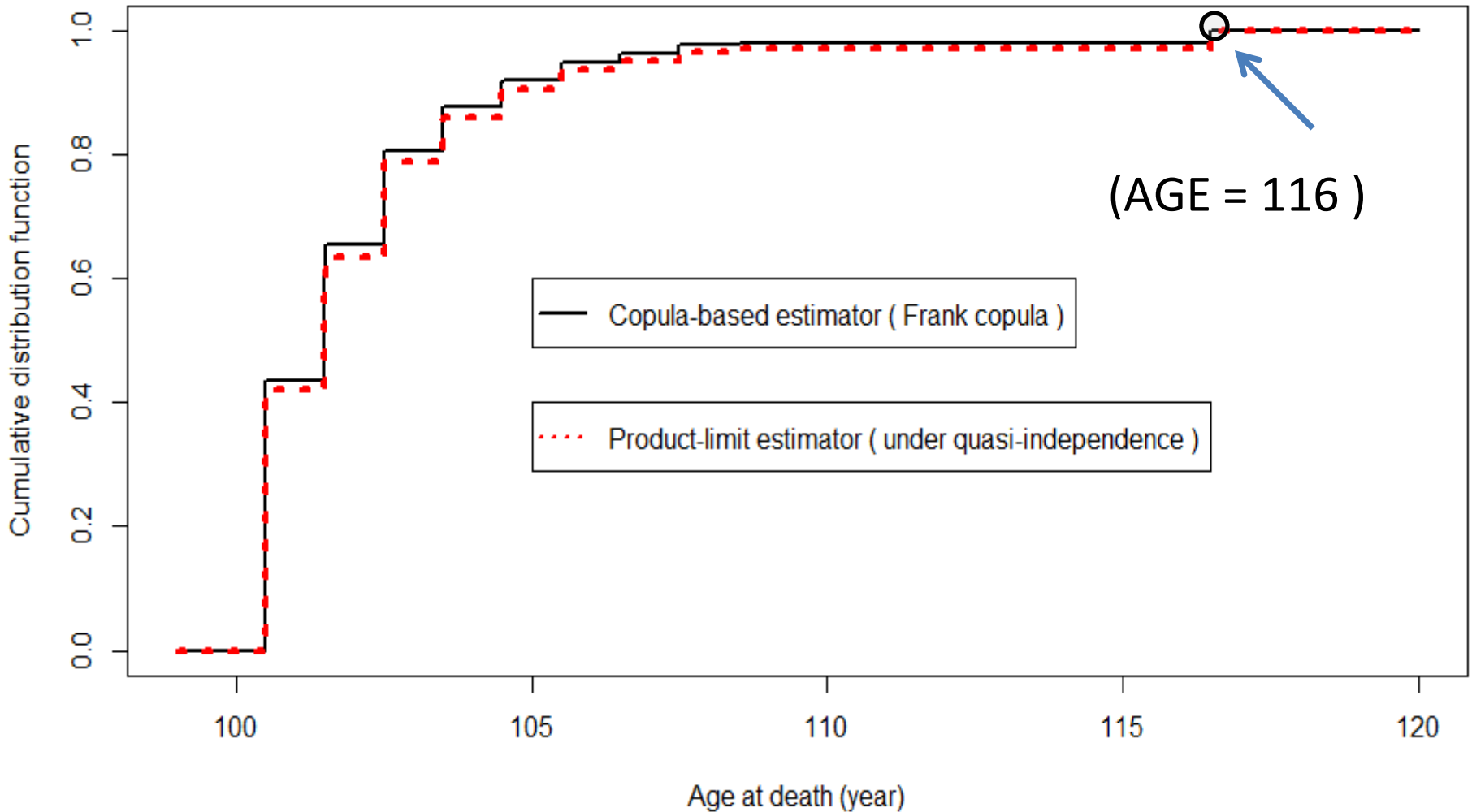
Japanese centenarians data: Sibuya M, Hanayama N (2004)

- $n = 662$ Japanese male centenarians ascertained before year 1980
- Age at death = X
- Age at study end = Y

Data structure:

$$(X_j, Y_j); j = 1, \dots, n \quad \text{subject to } X_j \leq Y_j$$

Estimated Distribution of Japanese male centenarians



Summary

- The product-limit estimator for survival function relies on the **quasi-independence** (Tsai 1990)
- Quasi-independence is rejected for the Channing house data (Emura and Wang 2010).
- Copula models relax the quasi-independence by introducing the dependence between survival and truncation (Chaieb et al., 2006).
- In this research, I propose a new algorithm to solve the estimating equation of Chaieb et al. (2006). R depend.truncation package <http://cran.r-project.org/web/packages/depend.truncation/index.html>
- Copula approaches yield high survival probability for elderly residents in the Channing house residents than the product limit estimator does.