Analyzing the clustered and interval-censored data based on the semiparametric frailty model

Jinheum Kim[1] & Youn Nam Kim[2]

[1]Department of Applied Statistics, University of Suwon, [2]Yonsei University Graduate School of Public Health

July 2, 2012

Outline

- Clustered & interval-censored data: review
- Propose a Cox proportional shared frailty model
- Parameter inference: EM method
- Simulations
- Illustrative data analysis: DRS data
- Concluding remarks: discussion

Data setup

- $T_{ij}(i = 1, \ldots, n; j = 1, \ldots, n_i)$: survival time of the jth subject in the ith cluster (discrete or continuous)
- Observable data: $(L_{ij}, U_{ij}]$ instead of T_{ij} (note possibly $U_{ij} = \infty$ or $L_{ij} = 0$)
- Examples:
    - Goggins & Finkelstein(2000, BCS): Data from AIDS clinical trials on HIV infected individuals, urine & blood samples were supposed to be collected every 4 and 12 weeks, respectively, for testing in the presence of the opportunistic infection CMV
    - Goethals et al.(2009, JABES): Mastitis data, infection times of individual cow udder quarters with a bacterium, four udder quarters are clustered within a cow and udder quarters are sampled monthly
    - Diabetic Retinopathy Study(DRS) data, Twin study dat, Amalgam fillings data

Review

- Focus on the frailty-based methods:
- Bellamy et al.(2004, StatMed), Goethals et al.(2009, JABES), Ampe et al.(2012, Preventive Veterinary Medicine): parametric approach, using log-normal frailty(the first) and gamma frailty(the second & the third)
- Lam et al.(2010, StatMed): multiple imputation approach
- Duchateau and Janssen(2008): semi-parametric approach for the clustered and right-censored data
- Remarks: marginal model approach
    - Goggins & Finkelstein(2000, BCS), Kim & Sue(2002, StatMed)

Model

- u_i : unobservable frailty shared among the members of the ith cluster
- Assume that conditional on u_i, T_{ij}'s within the ith cluster are independent
- Consider a Cox proportional shared frailty model:

$$\lambda(t_{ij}|x_{ij}, u_i) = \lambda_0(t) u_i \exp(\beta' x_{ij})$$

with the marginal survival function of T_{ij} expressed as

$$S(t_{ij}|x_{ij}, u_i) = \exp\{-u_i \Lambda_0(t_{ij}) \exp(\beta' x_{ij})\},$$

where

$$\Lambda_0(t) = \int_0^t \lambda_0(s) ds$$

Remarks

- Assume $u_i \sim G(\theta^{-1}, \theta),\ \theta > 0$
- Conditional multivariate survival function of $T_{i1}, \ldots, T_{in_i}$ is

$$S(t_{i1}, \ldots, t_{in_i} | x_{ij}, u_i) = \prod_{j=1}^{n_i} S(t_{ij} | x_{ij}, u_i) = \exp\{-u_i \sum_{j=1}^{n_i} \Lambda_0(t_{ij}) \exp(\beta' x_{ij})\}$$

- Hence,

$$S(t_{i1}, \ldots, t_{in_i} | x_{ij}) = \{1 + \theta \sum_{j=1}^{n_i} \Lambda_0(t_{ij}) \exp(\beta' x_{ij})\}^{-\theta^{-1}},$$

- The association between cluster members is measured as Kendall's τ, i.e, $\tau = \theta/(\theta + 2)$

Likelihood construction

- $D_i = \{j | T_{ij} \in (L_{ij}, U_{ij}]\}$: set of individuals interval-censored within the ith cluster
- $R_i = \{j | T_{ij} \in (L_{ij}, \infty)\}$: set of individuals right-censored within the ith cluster
- d_i : the size of D_i
- Conditional on u_i, the likelihood of the ith cluster is given by

$$L_{c,i}(\lambda_0, \beta | x_{ij}, u_i) = \prod_{j \in R_i} \exp(-u_i \tilde{L}_{ij}) \prod_{j \in D_i} \{\exp(-u_i \tilde{L}_{ij}) - \exp(-u_i \tilde{U}_{ij})\},$$

where

$$\tilde{L}_{ij} = \Lambda_0(L_{ij})\exp(\beta' x_{ij}), \tilde{U}_{ij} = \Lambda_0(U_{ij})\exp(\beta' x_{ij}) :$$

independent of u_i

Representation using a Kronecker product

- a_{ik} $(k = 1, \ldots, 2^{d_i})$: the kth element of a_i, where

$$a_i = (\exp(-u_i \tilde{L}_{i1}), -\exp(-u_i \tilde{U}_{i1}))' \otimes \cdots \otimes (\exp(-u_i \tilde{L}_{id_i}), -\exp(-u_i \tilde{U}_{id_i}))'$$

- Then,

$$L_{c,i}(\lambda_0, \beta | x_{ij}, u_i) = \exp(-u_i \sum_{j \in R_i} \tilde{L}_{ij}) \sum_{k=1}^{2^{d_i}} a_{ik} = \exp(-u_i C_i) \sum_{k=1}^{2^{d_i}} a_{ik},$$

where

$$C_i = \sum_{j \in R_i} \tilde{L}_{ij}$$

Complete data-based likelihood

- Using a gamma frailty, $G(\theta^{-1}, \theta)$,

$$L_{f,i}(\lambda_0, \beta, \theta) = L_{c,i}(\lambda_0, \beta | x_{ij}, u_i) g(u_i; \theta)$$

$$= \sum_{k=1}^{2^{d_i}} a_{ik} \times \frac{u_i^{\theta^{-1}-1} \exp\{-u_i(\theta^{-1} + C_i)\}}{\Gamma(\theta^{-1})\theta^{\theta^{-1}}}$$

$$= \sum_{k=1}^{2^{d_i}} (-1)^{d_{ik}} \times \frac{u_i^{\theta^{-1}-1} \exp\{-u_i(\theta^{-1} + C_i + \log b_{ik})\}}{\Gamma(\theta^{-1})\theta^{\theta^{-1}}},$$

where b_{ik} $(k = 1, \ldots, d_i)$: the kth element of b_i,

$$b_i = (\exp(\tilde{L}_{i1}), \exp(\tilde{U}_{i1}))' \otimes \cdots \otimes (\exp(\tilde{L}_{id_i}), \exp(\tilde{U}_{id_i}))'$$

and d_{ik} : the number of the term $\exp(-u_i \tilde{U}_{ij})$ included in a_{ik}

Marginal likelihood

- Then,

$$L_{m,i}(\lambda_0, \beta, \theta) = \int_0^\infty L_{f,i}(\lambda_0, \beta, \theta) du_i$$
$$= \sum_{k=1}^{2^{d_i}} \frac{(-1)^{d_{ik}}}{(\theta^{-1} + C_i + \log b_{ik})^{\theta^{-1}} \theta^{\theta^{-1}}},$$

Posterior distribution & posterior mean

- Using the Bayes' rule,

$$f_{U_i}(u_i|\mathrm{Data}) = \sum_{k=1}^{2^{d_i}} w_{ik}\, G(\theta^{-1}, 1/(\theta^{-1} + C_i + \log b_{ik})),$$

where

$$w_{ik} = \frac{(-1)^{d_{ik}}/(\theta^{-1} + C_i + \log b_{ik})^{\theta^{-1}}}{\sum_{l=1}^{2^{d_i}}(-1)^{d_{il}}/(\theta^{-1} + C_i + \log b_{il})^{\theta^{-1}}}$$

- Then,

$$E(U_i|\mathrm{Data}) = \sum_{k=1}^{2^{d_i}} w_{ik} \frac{\theta^{-1}}{\theta^{-1} + C_i + \log b_{ik}} = u_i^*$$

and

$$E(\log U_i|\mathrm{Data}) = \sum_{k=1}^{2^{d_i}} w_{ik}\{\psi(\theta^{-1}) - \log(\theta^{-1} + C_i + \log b_{ik})\} = lu_i^*,$$

where $\psi(\cdot)$: di-gamma function

Likelihood

- Likelihood based on the complete data:

$$L_f(\lambda_0, \beta, \theta) = \prod_{i=1}^{n} L_{f,i}(\lambda_0, \beta, \theta)$$

- log-likelihood:

$$l_f(\lambda_0, \beta, \theta) = \sum_{i=1}^{n} \{ \sum_{j \in D_i} \log\{\exp(-u_i \tilde{L}_{ij}) - \exp(-u_i \tilde{U}_{ij})\} - u_i C_i \}$$

$$+ \sum_{i=1}^{n} \{(\theta^{-1} - 1)\log u_i - \theta^{-1} u_i\} - n\log\Gamma(\theta^{-1}) - n\theta^{-1}\log\theta$$

E-step

- Replace u_i by u_i^* in the first term and u_i and $\log u_i$ by u_i^* and lu_i^* in the second term

M-step: representation of log-likelihood

- $0 = s_0 < s_1 < \cdots < s_m < s_{m+1} = \infty$: midpoints of equivalence sets of $\{(L_{ij}, U_{ij}]\ , i = 1, \ldots, n;\ j = 1, \ldots, n_i\}$ (Lindsey & Ryan, 1998, StatMed)
- Assume that

$$S_0(s_q) = \exp\{-\sum_{k=0}^{q} \exp(\alpha_k)\}, q = 1, \ldots, m$$

- Then,

$$\Lambda_0(s_q) = \sum_{k=0}^{q} \exp(\alpha_k) = a_q, q = 1, \ldots, m$$

with $a_0 = 0, a_{m+1} = \infty$ (i.e., $\alpha_0 = -\infty, \alpha_{m+1} = \infty$)

M-step: representation of log-likelihood

- Letting $\alpha_{ijq} = I(s_q \in (L_{ij}, U_{ij}]),\ q = 1, \ldots, m+1,$

$$l_f^*(\alpha, \beta, \theta) = E\{l_f(\alpha, \beta, \theta)|\mathrm{Data}\}$$

$$= \sum_{i=1}^{n} \{ \sum_{j \in D_i} \log\{ \sum_{q=1}^{m+1} \alpha_{ijq} (\exp\{-u_i^* a_{q-1} \exp(\beta' x_{ij})\} - \exp\{-u_i^* a_q \exp(\beta' x_{ij})\})\}$$

$$-u_i^* \sum_{j \in R_i} \sum_{q=1}^{m+1} I(L_{ij} \in [s_{q-1}, s_q)) a_{q-1} \exp(\beta' x_{ij})\}$$

$$+ \sum_{i=1}^{n} \{(\theta^{-1} - 1) lu_i^* - \theta^{-1} u_i^*\} - n \log \Gamma(\theta^{-1}) - n\theta^{-1} \log\theta,$$

where $\alpha = (\alpha_1, \ldots, \alpha_m)'$

M-step: score functions

- Let

$$f^*_{ijq} = S^*(s_q|x_{ij})\log S^*(s_q|x_{ij}, u^*_i),$$

where

$$S^*(t|x_{ij}, u^*_i) = \exp\{-\Lambda_0(t)u^*_i \exp(\beta' x_{ij})\}$$

with $f^*_{ij0} = f^*_{ijm+1} = 0$,

$$b^*_{ijq} = u^*_i \exp(\alpha_q + \beta' x_{ij}),$$

$$c^*_{ijq} = \sum_{l=q}^{m+1}(\alpha_{ijl} - \alpha_{ijl+1})S^*(s_l|x_{ij}, u^*_i)$$

with $\alpha_{ijm+2} = 0$,

$$g^*_{ij} = \sum_{q=1}^{m+1}\alpha_{ijq}\{S^*(s_{q-1}|x_{ij}, u^*_i) - S^*(s_q|x_{ij}, u^*_i)\}$$

M-step: score functions

- Score functions:

$$U_{\beta}^{*} = \frac{\partial l_{f}^{*}}{\partial \beta} = \sum_{i=1}^{n} \{ \sum_{j \in D_i} x_{ij} \frac{\sum_{q=1}^{m+1} \alpha_{ijq}(f_{ijq-1}^{*} - f_{ijq}^{*})}{g_{ij}^{*}}$$

$$+ u_i \sum_{j \in R_i} x_{ij} \sum_{q=1}^{m+1} I(L_{ij} \in [s_{q-1}, s_q)) a_{q-1} \exp(\beta' x_{ij}) \},$$

$$U_{\alpha_q}^{*} = \frac{\partial l_{f}^{*}}{\partial \alpha_q} = \sum_{i=1}^{n} \{ \sum_{j \in D_i} \frac{b_{ijq}^{*} c_{ijq}^{*}}{g_{ij}^{*}} - \sum_{j \in R_i} b_{ijq}^{*} I(L_{ij} \in [s_q, \infty)) \}, \ q = 1, \ldots, m,$$

$$U_{\theta}^{*} = \frac{\partial l_{f}^{*}}{\partial \theta} = \theta^{-2} \{ \sum_{i=1}^{n} (u_i^{*} - lu_i^{*}) + n\psi(\theta^{-1}) + n \log\theta - n \}$$

M-step: observed information matrix

- Let

$$I_{11} = -\frac{\partial^2 l_f^*}{\partial\beta\partial\beta'} = \sum_{i=1}^{n}\{\sum_{j\in D_i} x_{ij}x_{ij}'\{(\frac{\sum_{q=1}^{m+1}\alpha_{ijq}(f_{ijq-1}^* - f_{ijq}^*)}{g_{ij}^*})^2 - \frac{\sum_{q=1}^{m+1}\alpha_{ijq}(h_{ijq-1}^* - h_{ijq}^*)}{g_{ij}^*}\}$$

$$+u_i\sum_{j\in R_i} x_{ij}x_{ij}'\sum_{q=1}^{m+1} I(L_{ij}\in[s_{q-1},s_q))a_{q-1}\exp(\beta' x_{ij})\},$$

$$I_{12} = I_{21}'$$

$$= -\frac{\partial^2 l_f^*}{\partial\alpha_q\partial\beta} = -\sum_{i=1}^{n}\{\sum_{j\in D_i} x_{ij}b_{ijq}^*\{\frac{c_{ijq}^* + \sum_{l=q}^{m+1}(\alpha_{ijl}-\alpha_{ijl+1})f_{ijl}^*}{g_{ij}^*} - \frac{c_{ijq}^*}{g_{ij}^{*2}}\sum_{q=1}^{m+1}\alpha_{ijq}(f_{ijq-1}^* - f_{ijq}^*)\}$$

$$-\sum_{j\in R_i} x_{ij}b_{ijq}^* I(L_{ij}\in[s_q,\infty))\},$$

$$I_{22} = -\frac{\partial^2 l_f^*}{\partial\alpha_q^2} = -\sum_{i=1}^{n}\{\sum_{j\in D_i} b_{ijq}^*c_{ijq}^*(\frac{1-b_{ijq}^*}{g_{ij}^*} - \frac{b_{ijq}^*c_{ijq}^*}{g_{ij}^{*2}}) - \sum_{j\in R_i} b_{ijq}^* I(L_{ij}\in[s_q,\infty))\},$$

$$I_{22} = -\frac{\partial^2 l_f^*}{\partial\alpha_q\partial\alpha_r} = \sum_{i=1}^{n}\sum_{j\in D_i}(\frac{b_{ijq}^*b_{ijr}^*c_{ijq}^*c_{ijr}^*}{g_{ij}^{*2}} + \frac{b_{ijq}^*b_{ijr}^*c_{ijr}^*}{g_{ij}^*})(q<r),$$

where

$$h_{ijq} = f_{ijq}^*(1+\log S^*(s_q|x_{ij},u_i^*))$$

with $h_{ij0} = h_{ijm+1} = 0$

M-step: observed information matrix

- Let

$$I_\theta = -\frac{\partial^2 l_f^*}{\partial\theta^2} = \frac{2}{\theta^3}\{\sum_{i=1}^{n}(u_i^* - lu_i^*) + n\psi(\theta^{-1}) + n\log\theta - \frac{3}{2}n + \frac{1}{2}\theta^{-1}n\psi'(\theta^{-1})\},$$

where

$$\psi'(x) = \Gamma''(x)/\Gamma(x) - \psi(x)^2$$

M-step: Newton-Raphson algorithm

- The sth-step solution for $(\alpha, \beta)'$ is obtained through

$$\begin{pmatrix} \beta^{(s)} \\ \alpha^{(s)} \end{pmatrix} = \begin{pmatrix} \beta^{(s-1)} \\ \alpha^{(s-1)} \end{pmatrix} + I^{-1} \begin{pmatrix} U^*_\beta \\ U^*_\alpha \end{pmatrix},$$

where

$$I^{-1} = \begin{pmatrix} I^{-1}_{11|2} & I_{12|2} \\ I_{21|1} & I_{22|1} \end{pmatrix},$$

with $I_{11|2} = I_{11} - I_{12} I^{-1}_{22} I_{21}$, $I_{12|2} = I'_{21|2} = -I^{-1}_{11|2} I_{12} I^{-1}_{22}$, $I_{22|1} = I^{-1}_{22} + I^{-1}_{22} I_{21} I^{-1}_{11|2} I_{12} I^{-1}_{22}$

- Also, the sth-step solution for $(\alpha, \beta)'$ is obtained through

$$\theta^{(s)} = \theta^{(s-1)} + I^{-1}_\theta U^*_\theta$$

- Proceed E- and M-step iteratively until

$$\max\{|\alpha^{(s)}_q - \alpha^{(s-1)}_q|, |\beta^{(s)}_r - \beta^{(s-1)}_r|, |\theta^{(s)} - \theta^{(s-1)}|\} < \epsilon$$

Simulation: setup

- baseline hazard rate: $\lambda_0(t)$=0.1
- association par.: θ=0.5,1.0 or 1.5, i.e., τ(Kendall's tau)=$\frac{1}{5}$, $\frac{1}{3}$, or $\frac{3}{7}$
- number of clusters: n =100
- number of members: n_i=3 or 5
- 0-1 binary covariate
- β= 0, 0.5, or 1

Procedure

- Generate x_{ij} from the Bernoulli distribution with a success probability of 0.5
- Generate $u_i \sim G(\theta^{-1}, \theta)$
- Generate p from $U(0,1)$. With x_{ij}, u_i, and $\lambda_0(t)$, generate T_{ij} through

$$S(t|x_{ij}, u_i) = \exp(-u_i \Lambda_0(t) \exp(\beta x_{ij})) = 1 - p$$

- Assume follow-up visits were scheduled at 1,2,. . . , 12, resulting in a total of 12 possible visits.
- Construct a vector, v_i, of 12 independent Bernoulli variables with a success probability of π, to indicate whether or not the subject made or missed the visit (eg, $v_i = (1,1,0,0,1,0,1,0,0,0,1,0)'$, possible intervals: (0,1], (1,2], (2,5], (5,7], (7,11], (11,∞))
- T_{ij} is censored to the times of the nearest visits made before and after the failure time (eg, If $t_{ij} = 8.5$, interval censored to (7,11])

Simulation: results

Table : Bias, standard deviation(SD), mean of se(SeM), 95% coverage rate(CP) of parameters, β and θ, based on 2,000 replications when n_i=3

True				β				θ			
π	β	θ	RC	Bias	SD	SeM	CP	Bias	SD	SeM	CP
0.5	0	0.5	42.7	-0.008	0.175	0.154	91.5	0.000	0.176	0.065	54.1
		1	48.6	0.000	0.187	0.164	91.6	0.000	0.265	0.124	65.5
		1.5	53.2	-0.004	0.210	0.172	89.2	-0.002	0.361	0.179	66.5
	0.5	0.5	35.5	-0.002	0.167	0.147	91.1	0.007	0.163	0.066	58.1
		1	42.3	-0.005	0.186	0.156	89.9	-0.016	0.246	0.122	67.2
		1.5	47.8	-0.009	0.199	0.165	90.0	-0.018	0.335	0.178	68.2
	1	0.5	29.7	-0.015	0.171	0.146	90.6	-0.016	0.147	0.064	59.9
		1	37.2	-0.029	0.181	0.154	89.8	-0.023	0.233	0.122	69.0
		1.5	43.0	-0.036	0.198	0.162	88.2	-0.049	0.318	0.174	70.8
0.8	0	0.5	39.9	-0.005	0.169	0.150	91.7	0.009	0.168	0.066	56.7
		1	45.9	0.005	0.185	0.158	90.9	-0.002	0.263	0.124	65.0
		1.5	51.0	0.006	0.207	0.167	89.1	0.034	0.373	0.183	67.3
	0.5	0.5	32.9	-0.003	0.163	0.143	91.2	0.005	0.154	0.066	60.6
		1	40.1	-0.003	0.188	0.152	89.0	0.008	0.249	0.125	68.5
		1.5	45.7	-0.003	0.200	0.160	88.4	0.013	0.338	0.181	70.5
	1	0.5	27.2	-0.007	0.167	0.142	91.0	-0.005	0.145	0.065	64.6
		1	34.9	0.002	0.186	0.150	88.5	0.002	0.231	0.124	71.5
		1.5	41.1	-0.007	0.199	0.157	88.6	0.012	0.324	0.181	73.9

Simulation: results

Table : Bias, standard deviation(SD), mean of se(SeM), 95% coverage rate(CP) of parameters, β and θ, based on 2,000 replications when n_i=5

True				β				θ			
π	β	θ	RC	Bias	SD	SeM	CP	Bias	SD	SeM	CP
0.5	0	0.5	42.2	-0.004	0.132	0.119	92.6	-0.005	0.125	0.065	68.9
		1	48.1	0.002	0.142	0.126	91.7	0.003	0.202	0.125	77.7
		1.5	52.6	0.003	0.148	0.132	92.7	-0.007	0.293	0.179	76.9
	0.5	0.5	35.1	-0.002	0.123	0.114	93.1	-0.004	0.117	0.065	73.1
		1	42.0	-0.004	0.138	0.121	91.3	-0.006	0.196	0.124	76.5
		1.5	47.3	-0.005	0.140	0.127	92.6	-0.018	0.266	0.178	80.1
	1	0.5	29.4	-0.010	0.125	0.113	92.1	-0.006	0.114	0.065	74.5
		1	36.7	-0.007	0.134	0.119	91.7	-0.027	0.188	0.121	77.7
		1.5	42.6	-0.014	0.148	0.125	89.6	-0.021	0.264	0.177	79.5
0.8	0	0.5	39.8	-0.001	0.126	0.116	93.3	0.004	0.125	0.066	69.0
		1	46.1	-0.004	0.134	0.123	92.5	0.010	0.203	0.125	78.3
		1.5	50.8	-0.005	0.146	0.129	92.2	-0.017	0.288	0.178	78.2
	0.5	0.5	32.8	-0.001	0.123	0.111	92.4	-0.003	0.115	0.065	73.1
		1	40.0	-0.001	0.136	0.118	90.7	-0.010	0.191	0.123	78.0
		1.5	45.6	-0.001	0.143	0.124	91.9	-0.002	0.277	0.179	79.3
	1	0.5	27.2	-0.001	0.124	0.110	91.9	-0.009	0.109	0.064	75.2
		1	34.9	-0.002	0.130	0.116	91.9	-0.011	0.186	0.123	80.1
		1.5	40.9	-0.012	0.141	0.121	91.3	-0.012	0.261	0.178	81.1

Illustrative analysis: DRS data

- Data from the Diabetic Retinopathy Study
  - to evaluate the effectiveness of laser photocoagulation in delaying or preventing the onset of blindness in individuals with diabetes associated with retinopathy
- Data collection
  - 197 diabetic patients who have a high risk of experiencing blindness in both eyes
  - one eye was randomly selected for treatment and the other eye went untreated
  - visual acuity was measured in both eyes before treatment and at 4-months intervals following treatment
  - time to blindness was defined as the first occurrence of visual acuity less than 5/200

Illustrative analysis: DRS data

- Covariates
  - type of diabetes: $x_1 = 0$ if the age at onset is < 20 and 1 o.w
  - presence/absence of treatment: $x_2 = 0$ if untreated and 1 if treated with laser photocoagulation
  - $x_3 = x_1 \times x_2$
- Ross & Moore(BCS, 1999): categorized into 16 intervals such as (0,6], (6,10], (10, 14], . . ., (54,58], (58,66], (66,83]

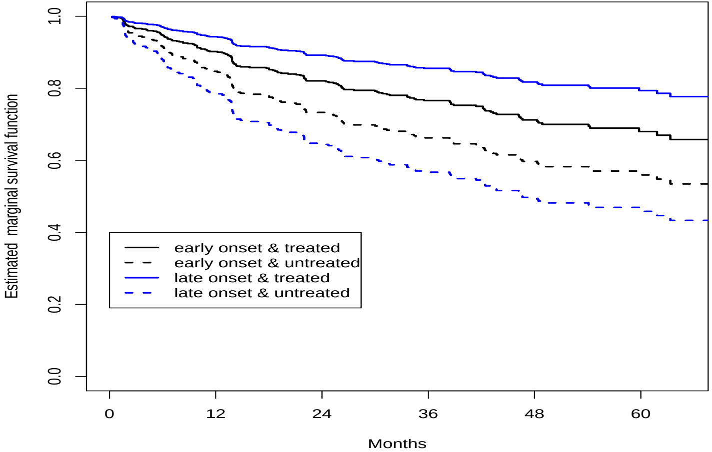
DRS data: results

- Laser photocoagulation appears to be effective (p=0.018) in delaying the occurrence of blindness, although there is also a significant treatment by diabetes type interaction effect (p=0.006)
- Laser photocoagulation is effective in delaying blindness for both types of diabetes (HR(late onset)=0.24, HR(early onset)=0.56)
- More effective for the adult-onset diabetes than for juvenile-onset diabetes(=2.46)
- The frailty effects are statistically significant (p=0.000)

Table : Parameter estimation for DRS data

	Type(x_1)		Treatment(x_2)		Interaction(x_3)		θ	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
				Grouped case				
Proposed	0.40	0.20	-0.52	0.22	-0.96	0.35	0.99	0.09
Lam et al.	0.41	0.26	-0.51	0.23	-0.99	0.36	0.93	0.31
				Ungrouped case				
Cox model(frailty)	0.40	0.26	-0.51	0.23	-0.99	0.36	0.93	0.24
Weibull(frailty)	0.42	0.27	-0.53	0.24	-1.03	0.37	1.10	0.37

Estimated marginal survival functions



Discussion

- Propose a semi-parametric model for analysing the clustered and interval-censored data and also plug-in a gamma frailty to the model to measure the association between members within a same cluster
- Propose an estimation procedure based on EM algorithm
- Simulation results showed that our estimation procedure may result in unbiased estimates, but the standard error is smaller than expected. It gave conservative results in estimating the coverage rate
- To overcome this conservativeness, we are trying to apply the MI method to estimating the standard error
- Additionally to investigate how robust our estimating procedure is to a misspecification of the frailty distribution and to compare our semi-parametric approach with the parametric approaches

Thank you!