Marginal Proportional Hazards Models for Multivariate Interval-Censored Data

Yangjianchen Xu, Donglin Zeng and D. Y. Lin

Department of Biostatistics, University of North Carolina at Chapel Hill

Background

- Multivariate interval-censored data
  - Multiple types of events or clusters of study subjects
  - Each event is only known to occur over a particular time interval
- Example: Atherosclerosis Risk in Communities Study

Data

- n clusters
- n_i subjects in ith cluster
- K types of events
- $X_{ijk}(\cdot)$: covariates for the kth event time for the jth subject of the ith cluster
- T_{ijk}: event time bracketed by $(L_{ijk}, R_{ijk}]$

Models and Pseudo-Likelihood

- Marginal hazard function for T_{ijk}:

$$\lambda_{ijk}(t) = \lambda_k(t) exp\{\beta_k^T X_{ijk}(t)\}$$

- Pseudo-Likelihood for β_k and Λ_k:

$$\prod_{i=1}^{n}\prod_{j=1}^{n_i}\left(\exp\left[-\int_0^{L_{ijk}} \exp\left\{\beta_k^{\mathrm{T}} X_{ijk}(t)\right\} \mathrm{d}\Lambda_k(t)\right] - \exp\left[-\int_0^{R_{ijk}} \exp\left\{\beta_k^{\mathrm{T}} X_{ijk}(t)\right\} \mathrm{d}\Lambda_k(t)\right]\right)$$

EM-type Algorithm

- Introduce latent (Poisson) random variables W_{ijkq}
- E-step

$$\hat{E}(W_{ijkq}) = \frac{I(L_{ijk} < t_{kq} \le R_{ijk} < \infty)\lambda_{kq} \exp\left(\beta_k^{\mathrm{T}} X_{ijkq}\right)}{1 - \exp\left\{-\sum_{L_{ijk} < t_{kq'} \le R_{ijk}} \lambda_{kq'} \exp\left(\beta_k^{\mathrm{T}} X_{ijkq'}\right)\right\}}$$

- M-step
  - Update β_k:

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{q=1}^{m_k} I(R_{ijk}^* \ge t_{kq})\hat{E}(W_{ijkq})\left\{X_{ijkq} - \frac{\sum_{i'=1}^{n}\sum_{j'=1}^{n_{i'}} I\left(R_{i'j'k}^* \ge t_{kq}\right)\exp\left(\beta_k^{\mathrm{T}} X_{i'j'kq}\right) X_{i'j'kq}}{\sum_{i'=1}^{n}\sum_{j'=1}^{n_{i'}} I\left(R_{i'j'k}^* \ge t_{kq}\right)\exp\left(\beta_k^{\mathrm{T}} X_{i'j'kq}\right)}\right\} = 0$$

  - Update λ_{kq}:

$$\lambda_{kq} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i} I(R_{ijk}^* \ge t_{kq})\hat{E}(W_{ijkq})}{\sum_{i=1}^{n}\sum_{j=1}^{n_i} I(R_{ijk}^* \ge t_{kq})\exp\left(\beta_k^{\mathrm{T}} X_{ijkq}\right)}$$

Asymptotic Properties

- Strong consistency of parameter estimators:

$$|\hat{\beta} - \beta_0| + \sum_{k=1}^{K} \sup_{t \in [0, \tau_k]} \left|\widehat{\Lambda_k}(t) - \Lambda_{0k}(t)\right| \to_{a.s.} 0$$

- Weak convergence: $n^{1/2}(\hat{\beta} - \beta_0) \rightharpoonup N(0, \Omega)$
- Sandwich variance estimator is consistent:

$$\{D_{h_n}^2 \mathrm{pl}_k(\widehat{\beta_k})\}^{-1} \sum_{i=1}^{n} D_{h_n}\mathrm{pl}_{ki}(\widehat{\beta_k}) D_{h_n}\mathrm{pl}_{li}(\widehat{\beta_l})^{\mathrm{T}} \{D_{h_n}^2 \mathrm{pl}_l(\widehat{\beta_l})\}^{-1}$$

Simulation Studies

- Clustered data: compare marginal model and random-effects model

Table 1. *Parameter estimation results for simulation studies with clustered data*

			Marginal model						Random-effects model			
ρ	No. of clusters	Parameter	Bias	SE	SEE	CP	SEEn	CPn	Bias	SE	SEE	CP
0.2	$n = 100$	$\beta_1 = 0.5$	0.005	0.202	0.211	95.5	0.182	92.0	0.060	0.222	0.220	94.2
		$\beta_2 = -0.5$	−0.012	0.314	0.314	94.8	0.279	91.9	−0.064	0.355	0.344	94.2
	$n = 200$	$\beta_1 = 0.5$	0.005	0.142	0.145	95.4	0.125	91.6	0.060	0.157	0.154	93.1
		$\beta_2 = -0.5$	−0.006	0.219	0.216	94.6	0.194	91.7	−0.069	0.244	0.241	94.2
	$n = 400$	$\beta_1 = 0.5$	0.003	0.097	0.100	95.7	0.087	92.3	0.059	0.110	0.109	91.7
		$\beta_2 = -0.5$	−0.004	0.154	0.152	94.5	0.136	91.5	−0.061	0.171	0.170	93.4
0.5	$n = 100$	$\beta_1 = 0.5$	0.007	0.238	0.247	95.7	0.182	86.8	0.224	0.335	0.326	89.4
		$\beta_2 = -0.5$	−0.013	0.369	0.359	94.1	0.279	86.5	−0.229	0.493	0.483	92.6
	$n = 200$	$\beta_1 = 0.5$	0.004	0.165	0.169	95.6	0.125	86.5	0.215	0.233	0.227	84.7
		$\beta_2 = -0.5$	−0.005	0.253	0.249	94.5	0.194	86.7	−0.210	0.340	0.336	90.7
	$n = 400$	$\beta_1 = 0.5$	0.002	0.114	0.118	95.7	0.087	86.5	0.209	0.162	0.160	74.6
		$\beta_2 = -0.5$	−0.002	0.178	0.174	94.6	0.136	87.0	−0.209	0.238	0.236	85.7
0.75	$n = 100$	$\beta_1 = 0.5$	0.013	0.266	0.278	95.9	0.182	82.2	0.515	0.521	0.493	82.2
		$\beta_2 = -0.5$	−0.013	0.410	0.399	93.9	0.279	82.2	−0.519	0.699	0.677	88.4
	$n = 200$	$\beta_1 = 0.5$	0.006	0.184	0.191	95.6	0.125	82.0	0.527	0.359	0.346	67.0
		$\beta_2 = -0.5$	−0.007	0.285	0.277	94.2	0.194	81.7	−0.524	0.491	0.473	80.0
	$n = 400$	$\beta_1 = 0.5$	0.004	0.130	0.132	95.4	0.087	81.4	0.516	0.254	0.239	42.8
		$\beta_2 = -0.5$	−0.001	0.197	0.194	94.4	0.136	82.6	−0.506	0.338	0.331	66.4

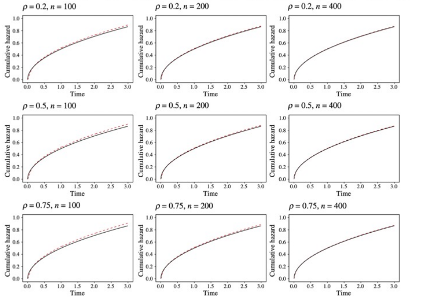

Fig. 1. Estimation of the cumulative baseline hazard function $\Lambda(t)$ for clustered data; the solid and dashed curves show the true values and averaged estimates, respectively, where each average is based on 10 000 replicates.

- Multiple-event data: estimate a common parameter

Table 2. *Parameter estimation results for simulation studies with multiple-event data*

		First event				Second event				Optimal combination			
ρ	No. of subjects	Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
0.2	$n = 100$	0.019	0.287	0.297	95.9	0.018	0.281	0.289	95.8	0.007	0.213	0.223	96.0
	$n = 200$	0.009	0.196	0.202	95.8	0.010	0.191	0.196	95.7	0.005	0.147	0.151	95.5
	$n = 400$	0.005	0.136	0.139	95.5	0.006	0.134	0.135	95.2	0.004	0.103	0.104	95.1
0.5	$n = 100$	0.020	0.285	0.297	96.2	0.026	0.280	0.289	96.0	0.011	0.232	0.248	96.3
	$n = 200$	0.010	0.195	0.202	95.9	0.014	0.191	0.196	95.8	0.008	0.160	0.167	96.1
	$n = 400$	0.007	0.138	0.139	95.4	0.007	0.133	0.135	95.7	0.005	0.112	0.115	95.7
0.75	$n = 100$	0.020	0.282	0.297	96.2	0.021	0.276	0.289	96.0	0.004	0.260	0.268	96.5
	$n = 200$	0.009	0.195	0.202	96.0	0.009	0.188	0.195	95.8	0.003	0.170	0.180	96.4
	$n = 400$	0.006	0.136	0.139	95.7	0.010	0.134	0.135	95.2	0.006	0.120	0.124	95.8

A Real Data Example

- Events: diabetes and hypertension
- Baseline risk factors: age, body mass index, glucose level, systolic blood pressure and diastolic blood pressure
- 8735 individuals, 2000 distinct interval endpoints

Table 3. *Regression analysis of the Atherosclerosis Risk in Communities Study data*

	Diabetes			Hypertension			Overall test		Difference		
Risk factor	Est	SE	p-value	Est	SE	p-value	Test	p-value	Est	SE	95% CI
Jackson	−0.145	0.149	0.332	−0.239	0.077	0.002	10.14	0.006	0.094	0.162	(−0.234, 0.413)
Minneapolis suburbs	−0.389	0.076	$<10^{-4}$	−0.100	0.046	0.031	29.17	$<10^{-4}$	−0.289	0.085	(−0.455, −0.122)
Washington County	0.115	0.073	0.114	0.078	0.048	0.103	4.68	0.096	0.037	0.083	(−0.125, 0.199)
Age	−0.014	0.005	0.007	0.013	0.003	$<10^{-4}$	26.17	$<10^{-4}$	−0.027	0.006	(−0.038, −0.016)
Male	−0.062	0.055	0.265	−0.238	0.034	$<10^{-4}$	49.34	$<10^{-4}$	0.176	0.062	(0.056, 0.297)
Caucasian	−0.451	0.160	0.005	−0.480	0.081	$<10^{-4}$	40.29	$<10^{-4}$	0.029	0.172	(−0.307, 0.366)
Body mass index (kg/m^2)	0.075	0.005	$<10^{-4}$	0.017	0.004	$<10^{-4}$	236.83	$<10^{-4}$	0.059	0.006	(0.047, 0.070)
Derived glucose value (mg/dl)	0.096	0.003	$<10^{-4}$	0.001	0.002	0.595	961.78	$<10^{-4}$	0.095	0.004	(0.088, 0.102)
Systolic blood pressure (mmHg)	0.005	0.003	0.096	0.058	0.002	$<10^{-4}$	914.31	$<10^{-4}$	−0.053	0.003	(−0.060, −0.046)
Diastolic blood pressure (mmHg)	0.005	0.004	0.310	0.011	0.003	$<10^{-4}$	17.48	0.0002	−0.007	0.005	(−0.016, 0.003)

Future Work

- Model checking
- Joint model: allow intermittent missing covariate values or measurement errors
- Transformation model: non-proportional hazard

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL