The Analysis of Spontaneous Abortion with Left Truncation, Partly Interval Censoring and Cure Rate

Yuan Wu, Ph.D. and Ronghui (Lily) Xu, Ph.D.
Duke University and University of California San Diego

Spontaneous Abortion (SAB)

- Definition: When happening no later than 20 weeks of gestational age; Otherwise it's called 'still birth'.
- Timing of SAB is of interest and therefore it is useful to model the SAB hazard over time.
- Related clinical question: Should pregnant women take any medication or vaccine?
  - H1N1 flu, Asthma.
  - Lack of data: Clinical trials typically exclude pregnant women.
- An example: autoimmune disease database.
  - Data source: Organization of Teratology Information Specialists (OTIS).
  - Sample size 963, 74 observed SAB events, 21 lost to follow-up, 10 interval censored, 868 observed cured.
  - Left truncation: Some women would have SAB before they have a chance to be referred to OTIS.
  - Main exposure of interest: autoimmune disease drugs.

Notation for SAB data

- Let $i = 1, \dots n$ denote the subjects in the study.
- Let Z_i be the vector of covariates, including a '1' for intercept.
- Among subjects experiences SAB, let T_i be the gestational age when SAB occurs.
- Let Q_i be the gestational age at the entry to study.
- Let $[U_i, V_i]$ be the gestational window including the exact date of SAB for some subjects.

Semiparametric Sieve Estimation

- Non-mixture cure survival model:

$$S(T|Z) = \exp[-e^{\beta^T Z} F(T)],$$

where T is random event time, Z is random covariate vector including '1', F is a distribution function.

- Let Q be random left truncation time, $[U, V]$ be the censored interval following Q. We also assume that Q, U and V are all independent of T conditional on Z. $\delta^{(1)} = 1_{[T \le U]}$, $\delta^{(2)} = 1_{[U < T \le V]}$, $\delta^{(3)} = 1_{[T > V]}$.
- Likelihood function:

$$l(\beta, F) = \log\left[\left\{\frac{\frac{\partial F(T|Z)}{\partial T}}{S(Q|Z)}\right\}^{\delta^{(1)}} \left\{\frac{S(U|Z) - S(V|Z)}{S(Q|Z)}\right\}^{\delta^{(2)}} \left\{\frac{S(V|Z)}{S(Q|Z)}\right\}^{\delta^{(3)}}\right].$$

- B-spline to I-spline: $I_k = \sum_{j \ge k} B_j$.
- I-spline approximation for F finite support $[a, b]$:

$$F_I(t) = \sum_k \alpha_k I_k(t),$$

for $\sum_k \alpha_k = 1$ and $\alpha_k \ge 0$.

- Constrained MLE problem:

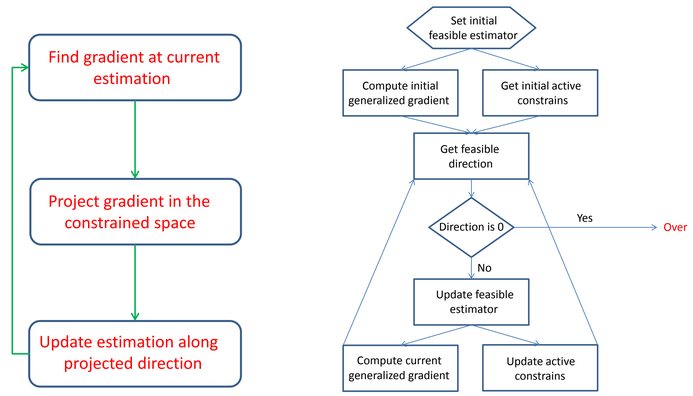
$$\max \sum_i^n l_i(\beta, F_I),$$

with $\sum_k \alpha_k = 1$ and $\alpha_k \ge 0$.

Estimation Algorithm

- Generalized gradient projection method.

Figure : Algorithm flowchart



Inference for Parametric Part

- General semiparametric information theory:

$$n^{1/2}(\hat{\beta}_n - \beta_0) \to_d N(0, I^{-1}(\beta_0)).$$

- Observed information:

$$\dot{l}_\beta(\hat{\beta}_n, \hat{F}_I) = \{\dot{l}_{\beta_1}(\hat{\beta}_n, \hat{F}_I), \dot{l}_{\beta_2}(\hat{\beta}_n, \hat{F}_I), \cdots, \dot{l}_{\beta_{number(Z)}}(\hat{\beta}_n, \hat{F}_I)\}^T.$$

$$\dot{l}_F(\hat{\beta}_n, \hat{F}_I)[B] = \{\dot{l}_F(\hat{\beta}_n, \hat{F}_I)[B_1], \dot{l}_F(\hat{\beta}_n, \hat{F}_I)[B_2], \cdots, \dot{l}_F(\hat{\beta}_n, \hat{F}_I)[B_{number(I)}]\}^T.$$

$$A_{11} = \mathbb{P}_n\{\dot{l}_\beta(\hat{\beta}_n, \hat{F}_I)^{\otimes 2}\}, \quad A_{12} = \mathbb{P}_n\{\dot{l}_\beta(\hat{\beta}_n, \hat{F}_I)\dot{l}_F^T(\hat{\beta}_n, \hat{F}_I)[B]\},$$

$$A_{21} = A_{12}^T, \quad A_{22} = \mathbb{P}_n\{(\dot{l}_F^T(\hat{\beta}_n, \hat{F}_I)[B^l])^{\otimes 2}\}.$$

$$\hat{\mathcal{O}}_n = A_{11} - A_{12}A_{22}^{-1}A_{21}.$$

- $\hat{\mathcal{O}}_n$ is consistent estimator for $I(\beta_0)$.

Simulation Study

- Data generating.
  - $F(t) = (1 - e^{-t})/(1 - e^{-4})$, $\beta_0 = (0.5, 0.5)^T$, $Z = (1, Z_1)$ with Z_1 follows $N(0, 1)$.
  - Q follows uniform $[0, 1]$, U and V both follow uniform $[1, 3.95]$.
  - Sample size 200, repetition 500.
- Simulation results.

Table : Parametric estimation

	True value	Estimate	SD	SE	95% coverage probability
β_{01}	0.5	0.55	0.206	0.203	94.2%
β_{02}	0.5	0.51	0.105	0.109	95.2%

SAB Analysis Results

- Covariates effect.

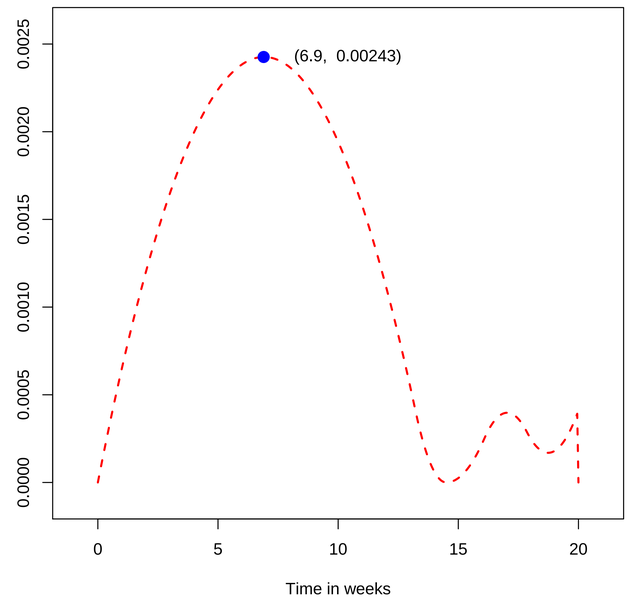
Table : Parametric estimation

	Estimate	SE	p-value
Intercept	-3.740	0.887	$<$0.0001
Maternal age	0.077	0.024	0.0012
Prior Tab	-0.299	0.416	0.473
Smoking	0.565	0.361	0.117
Healthy control	-0.544	0.476	0.252
Diseased control	0.116	0.265	0.662

Note: healthy control and diseased control are both in terms of exposure to autoimmune disease drugs.

- Baseline hazard estimation.

Figure : Baseline hazard plot



- Interpretation: Exposure to autoimmune disease drugs dose not significantly change the risk of experiencing SAB; Older pregnant women are at higher risk of experiencing SAB.

Final Remarks

- The proposed model can handle very complex survival data with the following features.
  - Left truncation.
  - Partly interval censoring (right censoring and interval censoring are both special cases).
  - Cure rate.

yuan.wu@duke.edu and rxu@ucsd.edu