Simple techniques for comparing survival functions with interval-censored data

Jinheum Kim, joint with Chung Mo Nam

jinhkim@suwon.ac.kr

Department of Applied Statistics

University of Suwon

Contents

- Introduction
- Review
- Proposed tests
- Simulation studies
- Application to real examples
- Summary

Interval censored data

- Failure time: time to an event of interest, denoted by $T > 0$
- T may not be known exactly, but is known only to lie in a subinterval of the real line, $(L, R]$
- T is right-censored if $R = \infty$; exactly observed if $L = R$; and interval-censored if $L \neq R$
- Right-censored data: T is either right-censored or exactly observed

First example

- Breast cancer cosmesis data taken from Finkelstein and Wolfe (1985, BCS)
- A retrospective study with a periodic follow-up to compare early breast cancer patients who had been treated with radiation therapy followed by chemotherapy to those treated with radiotherapy alone
- Follow-up intervals for those patients who were geographically remote were often longer with increasing time after irradiation therapy

Data set of first example

- Observe time until the appearance of breast retraction (monitored every 4 or 6 months)

RT only		RT+Chemo	
Patient #	$(L_i, R_i]$	Patient #	$(L_i, R_i]$
1	(45,_]	1	(8,12]
2	(25,37]	2	(0,5]
⋮		⋮	
46	(46,_]	48	(48,_]

Second example

- Lung cancer post-operative treatment data taken from Yonsei Cancer Center (Study period: Sep/1990-Sep/1994)
- A retrospective study to compare the patients who had been treated with radiation therapy followed by chemotherapy to those treated with radiotherapy alone after lung resection

Data set of second example

- Observe time to onset of relapse after resection

RT only		RT+Chemo	
Patient #	$(L_i, R_i]$	Patient #	$(L_i, R_i]$
1	(12,14]	1	(48,_]
2	(2,4]	2	(0,7]
⋮		⋮	
30	(20,_]	28	(13,16]

Goal

- Let $S_l(t) = \text{Pr}(T > t),\ l = 1, \ldots, p,$ denote the survival function of T for the lth group
- To test
  $H_0:\ S_1(t) = \cdots = S_p(t),\ \forall t \in (0, \infty)$

  vs.

  $H_a:$ Not all survival functions are equal at t
- In two examples, focus on comparing treatment effect rates between two groups

Notations & assumptions

- Data: $(\mathbf{x}_i', A_i = (L_i, R_i]),\ i = 1, \ldots, n$
- $\mathbf{x}_i' = (x_{i1}, \ldots, x_{i,p-1}) = (1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$ or $(0, \ldots, 0)$ for the individuals who come from the $1\text{st}, \ldots, (p-1)\text{th}, p\text{th}$ group
- $\delta_i = 0$ or 1 depending on whether the ith individual is right-censored or not
- The distinct endpoints, L_i and R_i, of A_i are ordered and labelled $0 = s_0 < s_1 < \cdots < s_m = \infty$
- For $i = 1, \ldots, n; j = 1, \ldots, m$, $\alpha_{ij} = 1$ if $L_i < s_j \leq R_i$; $0,$ otherwise
- Interval-censoring mechanism is independent of the failure time and covariates

Tests for treatment comparison

- Finkelstein & Wolfe (1985, BCS)
- Finkelstein (1986, BCS)
- Sun (1996, 1997, StatMed)
- Pan (2000, StatMed)
- Sun (2001. LDA)
- Zhao & Sun (2004, StatMed)

Finkelstein (1986, BCS)

- Likelihood
  $L = \prod_i \mathrm{Pr}(T_i \in A_i|\mathbf{x}_i) = \prod_i \{S(L_i|\mathbf{x}_i) - S(R_i|\mathbf{x}_i)\}$
- Introduce the Cox model for

$$S(s|\mathbf{x}) = \{S(s)\}^{\text{exp}(\mathbf{x}'\boldsymbol{\beta})}$$

$\rightarrow L$ is a function of $\boldsymbol{\beta}$ and $\gamma = \text{log}\{-\text{log}S(s)\}$

- Propose a score test for $\boldsymbol{\beta} = \mathbf{0}$, based on
  $\mathbf{U} = \frac{\partial \text{log} L}{\partial \boldsymbol{\beta}}$ evaluated at $\boldsymbol{\beta} = \mathbf{0}$ and $\hat{\gamma}(\mathbf{0})$

Finkelstein (1986, BCS): remarks

- Resemble the usual logrank test for right-censored data in two sample case, *i.e.*

$$U = \sum_j (d_{1j} - d_j n_{1j}/n_j),$$

  where

$$n_j = \sum_i w_{rij},\ d_j = \sum_i w_{dij};\ n_{1j} = \sum_i x_i w_{rij},\ d_{1j} = \sum_i x_i w_{dij}$$

$$w_{rij} = \sum_{k \geq j} \alpha_{ik}\hat{p}_k / \sum_k \alpha_{ik}\hat{p}_k, \quad w_{dij} = \alpha_{ij}\hat{p}_j / \sum_k \alpha_{ik}\hat{p}_k$$

- The observed Fisher information matrix corresponding to γ, which need be inverted, could have too many zero off-diagonal entries

Sun (1996, StatMed)

- An interval-censoring version of the usual logrank test for discrete interval-censored data
- Plug-in $d_j, n_j, d_{1j},$ and n_{1j} of Finkelstein (1986) into the usual logrank test
- There was a chance to overestimate these quantities and his test may not reduce to the usual logrank test while right-censored data are available (Zhao & Sun, 2004)

Goal

- Develop a nonparametric test not to require the EM algorithm
- Develop better estimates of the numbers of failures and of individuals in risk set
- Always be accessible to estimated null variance regardless of the number of distinct observed times

Notations

- For exact or right-censored observations, re-define L_i as the largest among s_j's less than L_i and R_i as L_i
- $m_i(\geq 1)$: total number of s_j's included in A_i, *i.e.* $m_i = \sum_j \alpha_{ij}$
- $A_i = I_{i_1} \cup I_{i_1+1} \cup \cdots \cup I_{i_1+m_i-1}$, where $I_j = (s_{j-1}, s_j]$ and $i_1 \in \{1, \ldots, m\}$
- $\mathcal{R}_j$: a pseudo risk set of all individuals who have a nonzero probability of being at risk in I_j
- $\mathcal{D}_j$: a pseudo failure set of all individuals who have a nonzero probability of failing in I_j

Uniform assumption and weights

- Under H_0, the true failure time of the ith individual, given A_i, is uniformly distributed over $\{s_{i_1}, s_{i_1+1}, \ldots, s_{i_1+m_i-1}\}$, *i.e.*

$$\text{Pr}(T_i = s_k | A_i) = 1/m_i,\ k = i_1, i_1+1, \ldots, i_1+m_i-1$$

- Under the model,

$$w_{rij} = \Pr(T_i \geq s_j | A_i,\ i = 1, \ldots, n) = \sum_{k \geq j} \alpha_{ik} / \sum_k \alpha_{ik} :$$

conditional prob. of individual i being at risk in I_j

$$w_{dij} = \Pr(T_i \in I_j | A_i,\ i = 1, \ldots, n) = \delta_i \alpha_{ij} / \sum_k \alpha_{ik} :$$

conditional prob. of individual i failing in I_j

Illustration: artificial data

- Data
  - Interval-censored:(1,2], (3,5], (4,6]
  - Right-censored: (3,∞)$\equiv$(2,3]
  - Exact: (4,4]$\equiv$(3,4]
- s_j's: s_1=0,1,2,3,4,5,6,$s_7 = \infty$
- Covering
  - Interval-censored:(1,2]=I_2, (3,5]=$I_4 \cup I_5$, (4,6]=$I_5 \cup I_6$
  - Right-censored: (3,∞)=I_3
  - Exact: (4,4]=I_4

Illustration: interval-censored

		s_j							
i		1	2	3	4	5	6	∞	m_i
1	α_{ij}	0	1	0	0	0	0	0	1
	w_{rij}	1	1	0	0	0	0	0	-
	w_{dij}	0	1	0	0	0	0	0	-
2	α_{ij}	0	0	0	1	1	0	0	2
	w_{rij}	1	1	1	1	0.5	0	0	-
	w_{dij}	0	0	0	0.5	0.5	0	0	-
3	α_{ij}	0	0	0	0	1	1	0	2
	w_{rij}	1	1	1	1	1	0.5	0	-
	w_{dij}	0	0	0	0	0.5	0.5	0	-

Illustration: right-censored & exact

i		s_j 1	2	3	4	5	6	∞	m_i
4	α_{ij}	0	0	1	0	0	0	0	1
	w_{rij}	1	1	1	0	0	0	0	-
	w_{dij}	0	0	0	0	0	0	0	-
5	α_{ij}	0	0	0	1	0	0	0	1
	w_{rij}	1	1	1	1	0	0	0	-
	w_{dij}	0	0	0	1	0	0	0	-
	n_j	5	5	4	3	1.5	0.5	0	-
	d_j	0	1	0	1.5	1	0.5	0	-

Risk set and death set

- As in Sun (1996)' nonparametric test, plug-in $d_j, n_j, d_{jl},$ and n_{jl} into the usual logrank test
  $n_j = \sum_{i=1}^{n} w_{rij}$: counts of $\mathcal{R}_j$ in I_j
  $d_j = \sum_{i=1}^{n} w_{dij}$: counts of $\mathcal{D}_j$ in I_j
  $n_{jl} = \sum_{i}^{l} w_{rij}$: pesudo-counts of individuals at risk from population l in I_j
  $d_{jl} = \sum_{i}^{l} w_{rij}$: pesudo-counts of failures from population l in I_j
- Remark: For right-censored data, $d_j, n_j, d_{jl},$ and n_{jl} reduce to corresponding values in the usual logrank test

Proposed test

- Propose a logrank-type statistic

$$\mathbf{U} = (U_1, \ldots, U_{p-1})',$$

where $U_l = \sum_{j=1}^{m}(d_{jl} - d_j n_{jl}/n_j),\ l = 1, \ldots, p-1$

- To test H_0, propose a logrank-type test based on $\mathbf{U}$,

$$P = \mathbf{U}'\hat{\mathbf{\Sigma}}^{-1}\mathbf{U},$$

where $\hat{\mathbf{\Sigma}}$ is an estimated covariance of $\mathbf{U}$

- Use $P \sim \chi^2(p-1)$ approximately under H_0

Alternatives for $\hat{\Sigma}$

- An ad-hoc version of right-censored data by plugging $d_j, n_j, d_{jl},$ and n_{jl} into covariance matrix of usual logrank test statistic
- Applying multiple imputation method to impute T for a subject with $\delta = 1$ (Pan, 2000; Sun, 2001)

Simulation studies: design pars

- $T:$ generated from Weibull(β, λ) for Group 1; Weibull$(\beta, r_h\lambda)$ for Group 2 with β=0.5, 1.0, 2.0 and r_h=1.0, 1.8, 3.0
- $\lambda:$ determined to satisfy mean failure time of 12
- Interval-censored data:
  $L =$ max$(0, T - u_1), R = T + u_2$ with $u_1, u_2:$ discrete uniform over $\{1, \ldots, D\}; D$=2, 3, 5
- $C:$ censoring indicator from Bernoulli$(c_p); c_p$=0, .3, .5
- If C=0, set R=∞
- Sample size: n=50, 100
- Replications: 3000 (with SE=0.004), M=50

Results: level when n=50

	D	2				5			
β	c_p	LR	U	P_{na}	P_{mi}	LR	U	P_{na}	P_{mi}
.5	.0	.056	.071	.055	.058	.059	.070	.053	.049
	.3	.055	.066	.051	.053	.050	.062	.047	.046
	.5	.052	.068	.050	.052	.047	.061	.044	.042
1	.0	.057	.067	.056	.058	.059	.067	.053	.052
	.3	.063	.077	.060	.067	.053	.062	.054	.053
	.5	.050	.062	.050	.052	.058	.076	.057	.057
2	.0	.058	.070	.053	.053	.051	.066	.038	.032
	.3	.057	.073	.053	.056	.053	.066	.044	.038
	.5	.052	.069	.048	.051	.055	.072	.052	.045

Results: power when n=50

		D	2				5			
β	c_p	r_h	LR	U	P_{na}	P_{mi}	LR	U	P_{na}	P_{mi}
.5	.0	1.8	.519	.544	.499	.492	.530	.544	.474	.430
		3.0	.959	.956	.945	.931	.953	.945	.911	.832
	.3	1.8	.381	.429	.378	.370	.396	.448	.383	.341
		3.0	.839	.881	.849	.816	.864	.883	.841	.712
	.5	1.8	.285	.349	.290	.278	.293	.355	.294	.257
		3.0	.714	.810	.730	.669	.704	.806	.709	.533
1	.0	1.8	.532	.565	.528	.535	.520	.550	.497	.486
		3.0	.960	.968	.959	.960	.954	.961	.945	.934
	.3	1.8	.389	.424	.386	.392	.375	.411	.364	.355
		3.0	.851	.873	.848	.848	.841	.871	.829	.798
	.5	1.8	.281	.332	.285	.289	.303	.353	.299	.285
		3.0	.707	.749	.702	.696	.701	.755	.695	.653
2	.0	1.8	.526	.559	.517	.520	.526	.542	.458	.405
		3.0	.956	.964	.952	.953	.958	.960	.933	.912
	.3	1.8	.392	.426	.377	.379	.402	.434	.364	.317
		3.0	.848	.873	.841	.838	.854	.868	.823	.772
	.5	1.8	.294	.345	.289	.283	.291	.348	.272	.227
		3.0	.709	.757	.706	.693	.704	.759	.673	.599

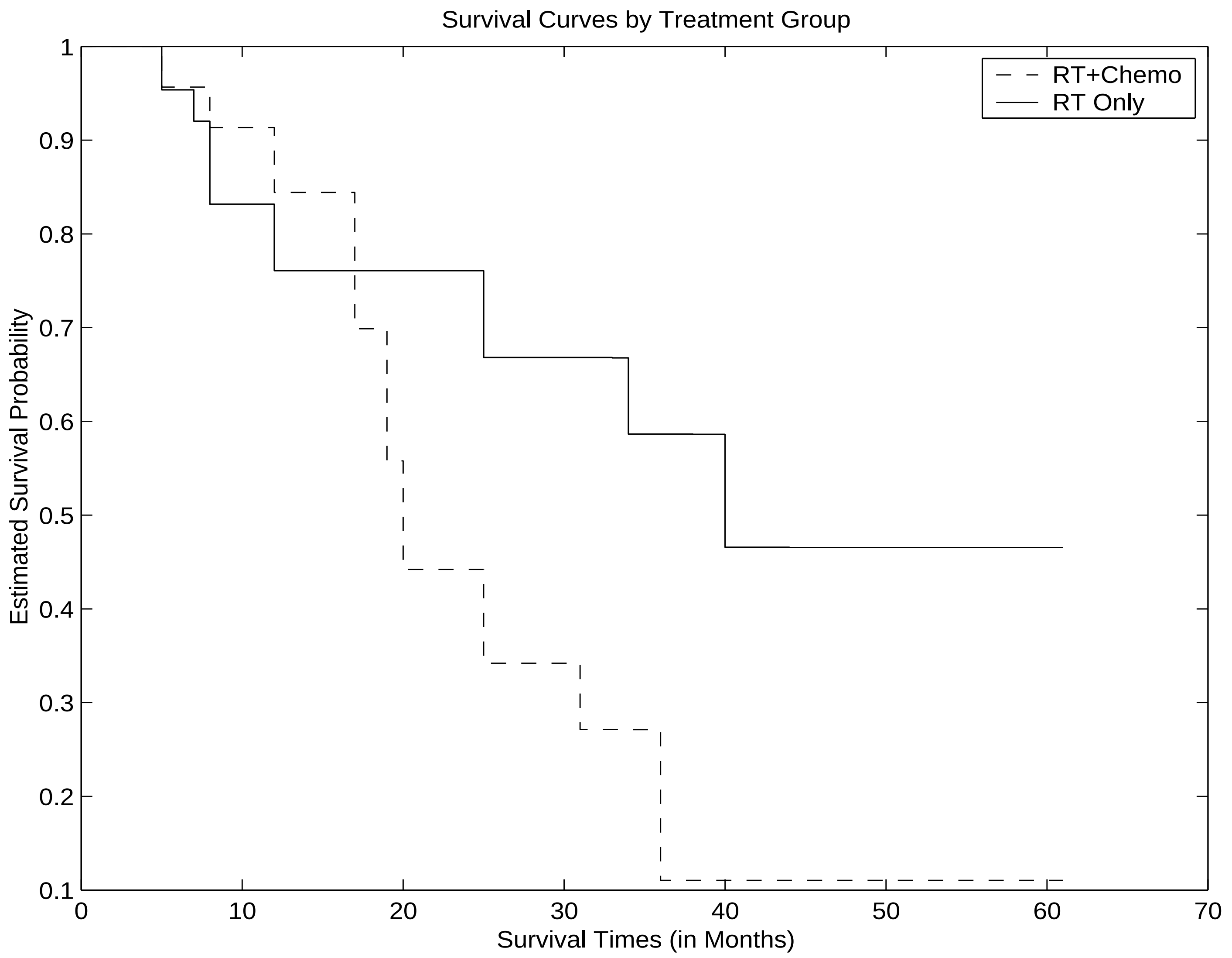
First example: revisit

- Breast cosmesis data

Treatment	n	Right Cen.(%)
RT only	46	54
RT+Chemo	48	27
Total	94	40

Test	Statistic	$p-$value
U	8.05	0.0046
P_{na}	10.12	0.0015
P_{mi}	9.34	0.0022

First example: survival curves



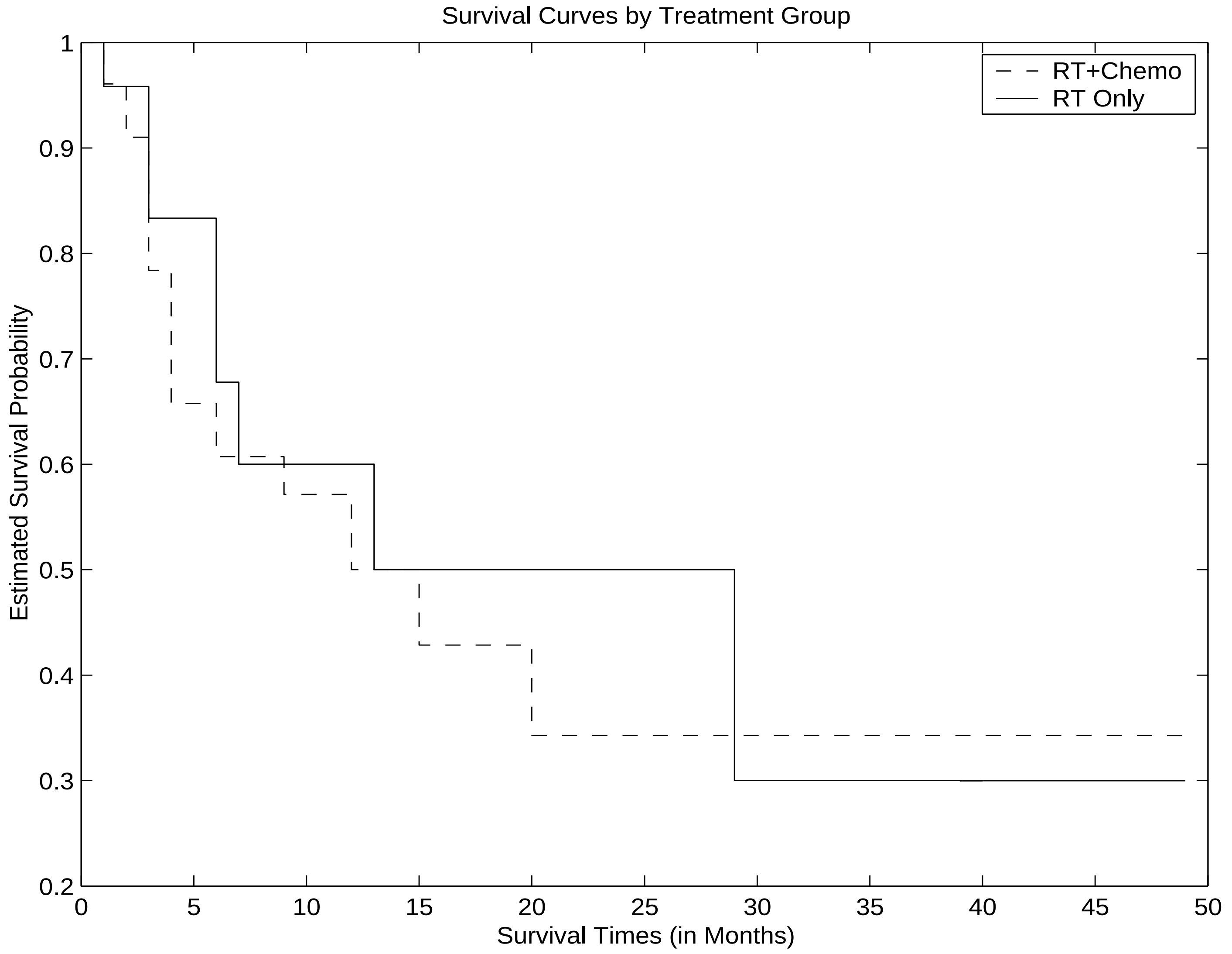
Second example: revisit

- Lung cancer post-operative data

Treatment	n	Right Cen.(%)
RT only	30	43
RT+Chemo	28	36
Total	58	40

Test	Statistic	$p-$value
Score	0.52	0.472
P_{na}	0.48	0.488
P_{mi}	0.49	0.484

Second example: survival curves



Concluding remarks

- Proposed a nonparametric test for comparing survival functions with interval-censored data
- The size of proposed test is well controlled; its power is comparable to that of efficient logrank test under proportional hazards
- Extend to more complicated censoring scheme, for example, truncated and doubly interval-censored (e.g., incubation time in AIDS study)

Thank you