Machine Learning for Healthcare

6.871, HST.956

Lecture 16: Survival modeling

David Sontag

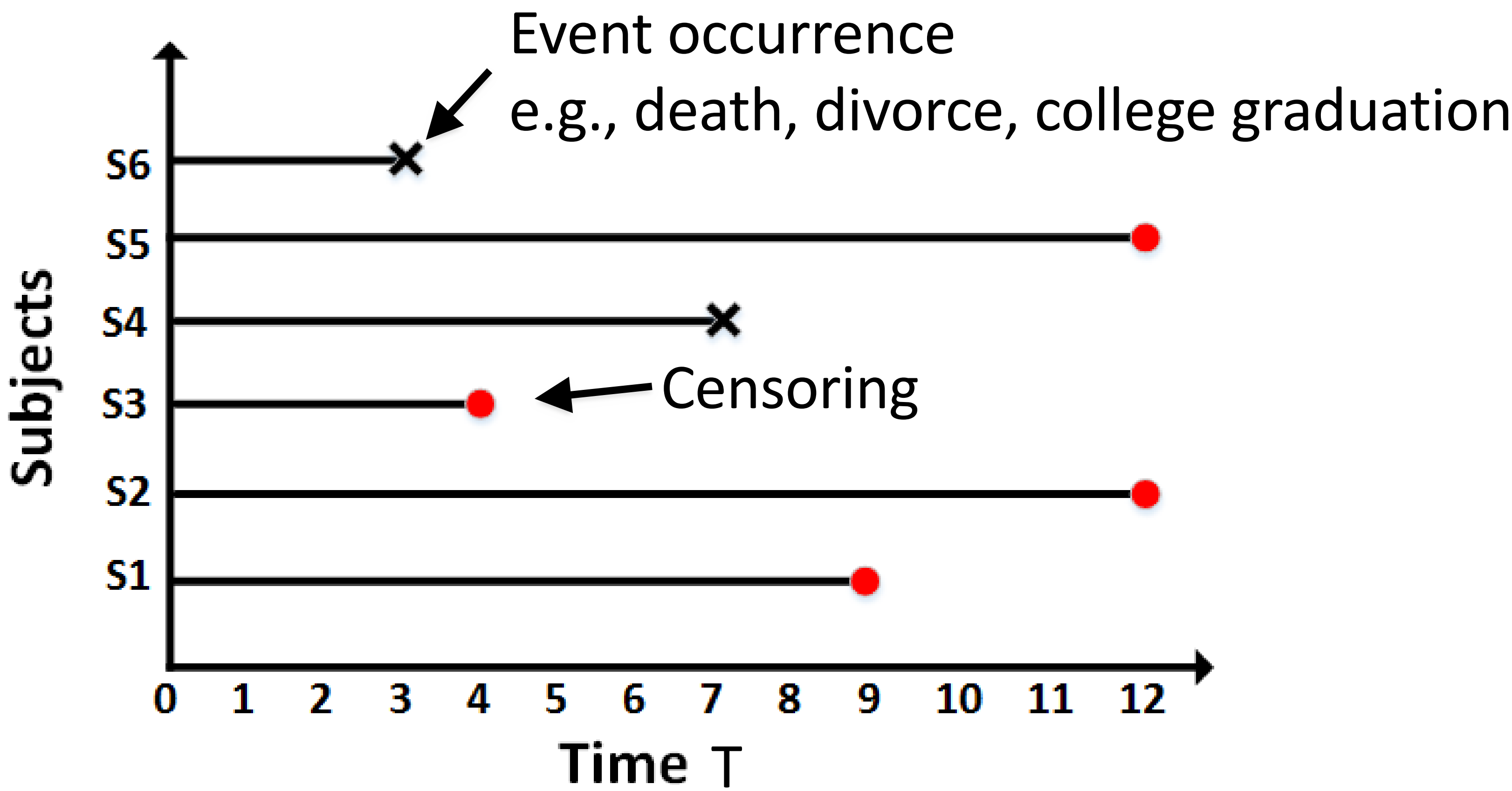
CSAIL

INSTITUTE FOR MEDICAL ENGINEERING & SCIENCE

HEALTH SCIENCES & TECHNOLOGY

Survival modeling

- Regression (i.e., predict time to event) with (potentially) right-censored data



[Wang, Li, Reddy. Machine Learning for Survival Analysis: A Survey. 2017]

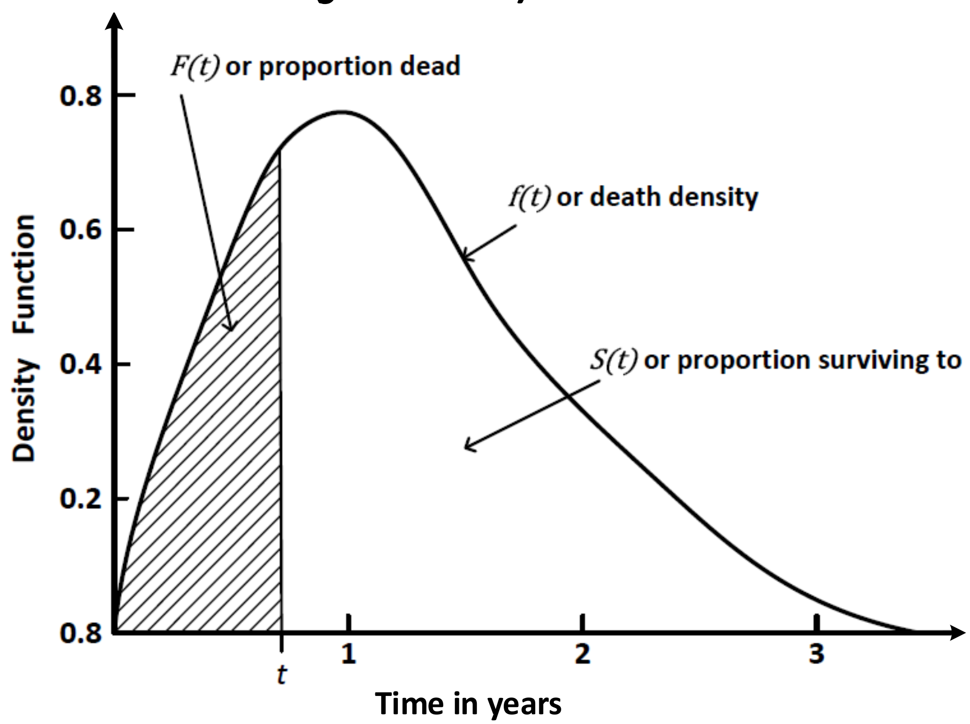
Why might censorship occur?

- Person does not experience event before **study ends**
- Person **lost to follow-up** during study period
- Person **withdraws from the study** because of death (if death is not event of interest) or some other reason (e.g. adverse drug reaction)

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

Notation and formalization

- *f(t|x)*, the probability of death/failure at time t, conditioned on **x**
- *Survival function* is 1 – (*f*'s CDF): $S(t \mid x) = P(T > t \mid x) = \int_{u=t}^{\infty} f(u \mid x)du$



$$S(0) = 1$$

$$S(\infty) = 0$$

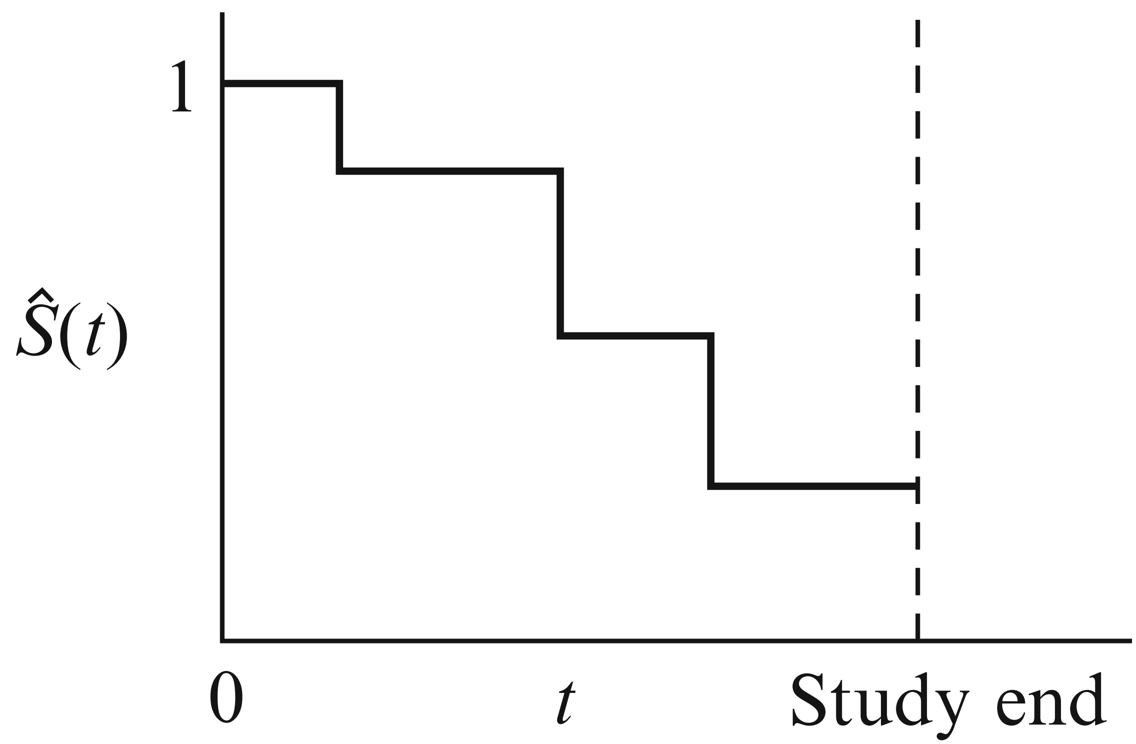
Fig. 2: Relationship among different entities $f(t)$, $F(t)$ and $S(t)$.

[Wang, Li, Reddy. Machine Learning for Survival Analysis: A Survey. 2017]
[Ha, Jeong, Lee. Statistical Modeling of Survival Data with Random Effects. Springer 2017]

Kaplan-Meier estimator of survival function S(t)=P(T > t)

- Example of a non-parametric method; good for unconditional density estimation

Estimated survival probability



[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

Kaplan-Meier estimator of survival function S(t)=P(T > t)

General Data Layout:

Indiv. #	t	d	X_1	$X_2 \ldots X_p$
1	t_1	d_1	X_{11}	$X_{12} \ldots X_{1p}$
2	t_2	d_2	X_{21}	$X_{22} \ldots X_{2p}$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n	t_n	d_n	X_{n1}	$X_{n2} \ldots X_{np}$

$$d = (0,\ 1) \text{ random variable}$$
$$= \begin{cases} 1 & \text{if failure} \\ 0 & \text{if censored} \end{cases}$$

Alternative (ordered) data layout:

Ordered failure times, $t_{(f)}$	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f
$t_{(0)} = 0$	$m_0 = 0$	q_0
$t_{(1)}$	m_1	q_1
$t_{(2)}$	m_2	q_2
.	.	.
.	.	.
.	.	.
$t_{(k)}$	m_k	q_k

n_f

Number of individuals who survived at least to this time (including those failing & censored in this interval)

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

EXAMPLE

The data: remission times (weeks) for two groups of leukemia patients

Group 1 (n = 21) treatment	Group 2 (n = 21) placebo
6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+,	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

Note: + denotes censored

	# failed	# censored	Total
Group 1	9	12	21
Group 2	21	0	21

failure times, $t_{(f)}$ — # at risk n_f — # of failures m_f — # censored in $[t_{(f)}, t_{(f+1)})$, q_f

Group 2 (placebo)

$t_{(f)}$	n_f	m_f	q_f
0	21	0	0

By default, we always include t=0 with n_f = n

[Freireich et al. The Effect of 6-Mercaptopurine on the Duration of Steroid-Induced Remissions in Acute Leukemia: A Model for Evaluation of Other Potentially Useful Therapy. *Blood*, **21**: 699-716, 1963]

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

EXAMPLE

The data: remission times (weeks) for two groups of leukemia patients

Group 1 (n = 21) treatment	Group 2 (n = 21) placebo
6, 6, 6, 7, 10,	1, 1, 2, 2, 3,
13, 16, 22, 23,	4, 4, 5, 5,
6+, 9+, 10+, 11+,	8, 8, 8, 8,
17+, 19+, 20+,	11, 11, 12, 12,
25+, 32+, 32+,	15, 17, 22, 23
34+, 35+,	

Note: + denotes censored

	# failed	# censored	Total
Group 1	9	12	21
Group 2	21	0	21

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f

Group 2 (placebo)

$t_{(f)}$	n_f	m_f	q_f
0	21	0	0
1	21	2	0

At t=1 there are 2 failure events (and no censoring events)

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

EXAMPLE

The data: remission times (weeks) for two groups of leukemia patients

Group 1 (n = 21) treatment	Group 2 (n = 21) placebo
6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+,	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

Note: + denotes censored

	# failed	# censored	Total
Group 1	9	12	21
Group 2	21	0	21

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f

Group 2 (placebo)

$t_{(f)}$	n_f	m_f	q_f
0	21	0	0
1	21	2	0
2	19	2	0

Beginning at t=2 there are 19 subjects still in the study, and there are 2 failure events (and no censoring events)

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

EXAMPLE

The data: remission times (weeks) for two groups of leukemia patients

Group 1 ($n = 21$) treatment	Group 2 ($n = 21$) placebo
6, 6, 6, 7, 10,	1, 1, 2, 2, 3,
13, 16, 22, 23,	4, 4, 5, 5,
6+, 9+, 10+, 11+,	8, 8, 8, 8,
17+, 19+, 20+,	11, 11, 12, 12,
25+, 32+, 32+,	15, 17, 22, 23
34+, 35+,	

Note: + denotes censored

	# failed	# censored	Total
Group 1	9	12	21
Group 2	21	0	21

→

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f

Group 2 (placebo)

$t_{(f)}$	n_f	m_f	q_f
0	21	0	0
1	21	2	0
2	19	2	0
3			

What are n_f, m_f, and q_f for $t_{(f)}$=3?

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

EXAMPLE

The data: remission times (weeks) for two groups of leukemia patients

Group 1 (n = 21) treatment	Group 2 (n = 21) placebo
6, 6, 6, 7, 10,	1, 1, 2, 2, 3,
13, 16, 22, 23,	4, 4, 5, 5,
6+, 9+, 10+, 11+,	8, 8, 8, 8,
17+, 19+, 20+,	11, 11, 12, 12,
25+, 32+, 32+,	15, 17, 22, 23
34+, 35+,	

Note: + denotes censored

	# failed	# censored	Total
Group 1	9	12	21
Group 2	21	0	21

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f

Group 2 (placebo)

$t_{(f)}$	n_f	m_f	q_f
0	21	0	0
1	21	2	0
2	19	2	0
3	17	1	0

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

EXAMPLE

The data: remission times (weeks) for two groups of leukemia patients

Group 1 ($n = 21$) treatment	Group 2 ($n = 21$) placebo
6, 6, 6, 7, 10,	1, 1, 2, 2, 3,
13, 16, 22, 23,	4, 4, 5, 5,
6+, 9+, 10+, 11+,	8, 8, 8, 8,
17+, 19+, 20+,	11, 11, 12, 12,
25+, 32+, 32+,	15, 17, 22, 23
34+, 35+,	

Note: + denotes censored

	# failed	# censored	Total
Group 1	9	12	21
Group 2	21	0	21

→

Group 2 (placebo)

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f
0	21	0	0
1	21	2	0
2	19	2	0
3	17	1	0
4	16	2	0
5	14	2	0
8	12	4	0
11	8	2	0
12	6	2	0
15	4	1	0
17	3	1	0
22	2	1	0
23	1	1	0

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f

Group 2: no censored subjects
Group 2 (placebo)

$t_{(f)}$	n_f	m_f	q_f	$\hat{S}(t_{(f)})$
0	21	0	0	1
1	21	2	0	
2	19	2	0	
3	17	1	0	
4	16	2	0	
5	14	2	0	
8	12	4	0	
11	8	2	0	
12	6	2	0	
15	4	1	0	
17	3	1	0	
22	2	1	0	
23	1	1	0	

$$\hat{S}(t_{(f)}) = \frac{\#\text{ surviving past } t_{(f)}}{n}$$

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

failure times, $t_{(f)}$ | # at risk n_f | # of failures m_f | # censored in $[t_{(f)}, t_{(f+1)})$, q_f

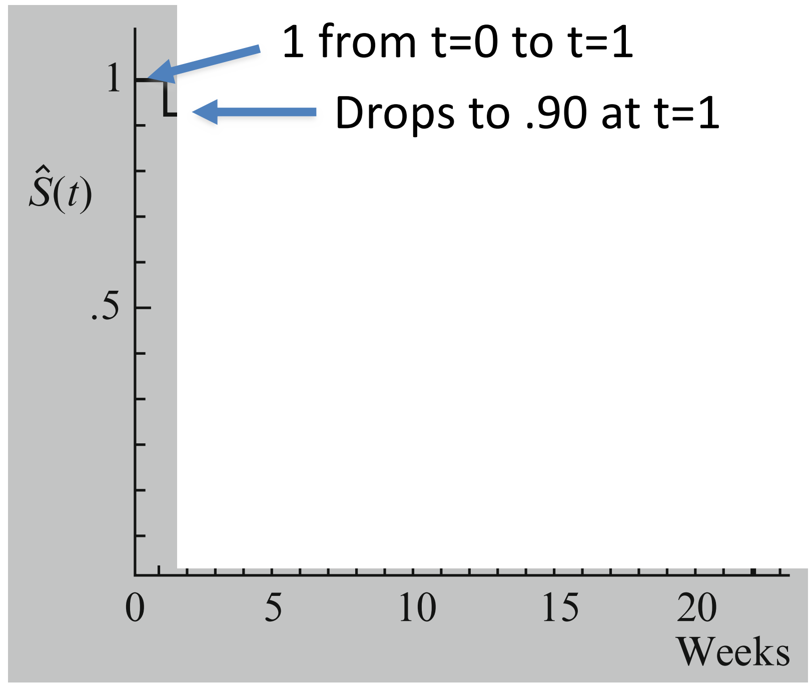
Group 2: no censored subjects

Group 2 (placebo)

$t_{(f)}$	n_f	m_f	q_f	$\hat{S}(t_{(f)})$
0	21	0	0	1
1	21	2	0	19/21 = .90
2	19	2	0	
3	17	1	0	
4	16	2	0	
5	14	2	0	
8	12	4	0	
11	8	2	0	
12	6	2	0	
15	4	1	0	
17	3	1	0	
22	2	1	0	
23	1	1	0	

$$\hat{S}(t_{(f)}) = \frac{\#\text{ surviving past } t_{(f)}}{n}$$

Kaplan-Meier curve



[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f

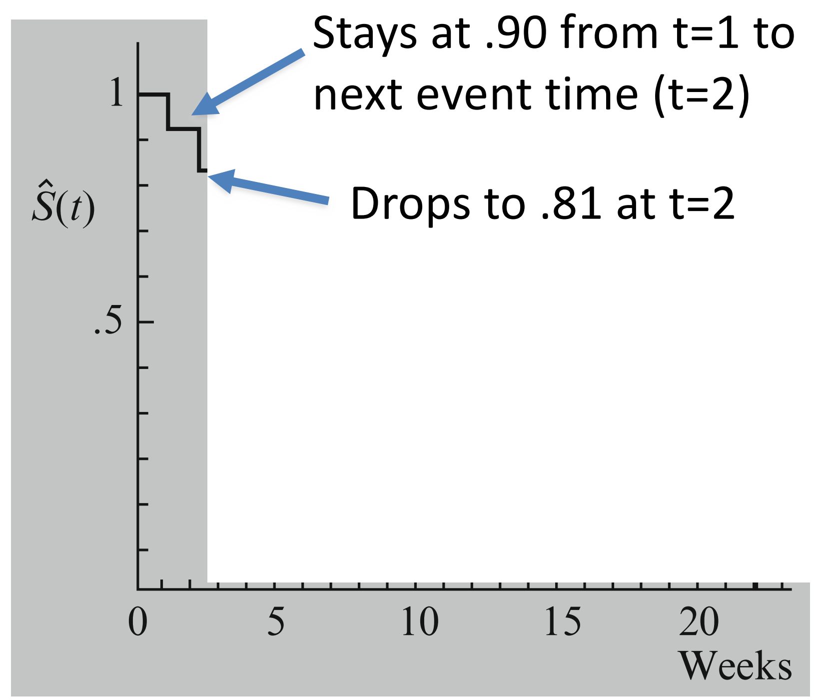
Group 2: no censored subjects

Group 2 (placebo)

$t_{(f)}$	n_f	m_f	q_f	$\hat{S}(t_{(f)})$
0	21	0	0	1
1	21	2	0	19/21 = .90
2	19	2	0	17/21 = .81
3	17	1	0	
4	16	2	0	
5	14	2	0	
8	12	4	0	
11	8	2	0	
12	6	2	0	
15	4	1	0	
17	3	1	0	
22	2	1	0	
23	1	1	0	

$$\hat{S}(t_{(f)}) = \frac{\#\ \text{surviving past } t_{(f)}}{n}$$

Kaplan-Meier curve



[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f	

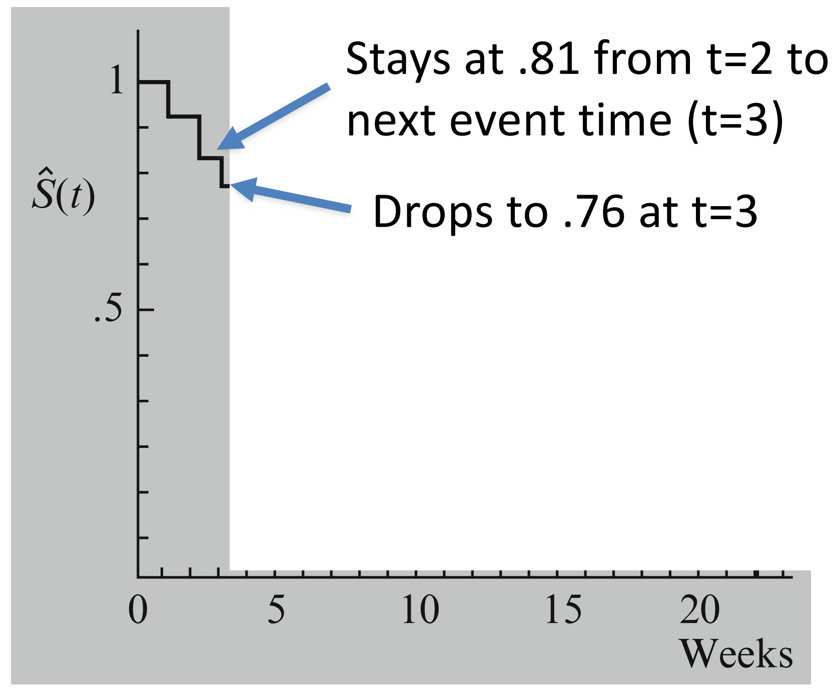
Group 2: no censored subjects

Group 2 (placebo)

$t_{(f)}$	n_f	m_f	q_f	$\hat{S}\,(t_{(f)})$
0	21	0	0	1
1	21	2	0	19/21 = .90
2	19	2	0	17/21 = .81
3	17	1	0	16/21 = .76
4	16	2	0	
5	14	2	0	
8	12	4	0	
11	8	2	0	
12	6	2	0	
15	4	1	0	
17	3	1	0	
22	2	1	0	
23	1	1	0	

$$\hat{S}(t_{(f)}) = \frac{\#\text{ surviving past } t_{(f)}}{n}$$

Kaplan-Meier curve



[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

failure times, $t_{(f)}$ | # at risk n_f | # of failures m_f | # censored in $[t_{(f)}, t_{(f+1)})$, q_f

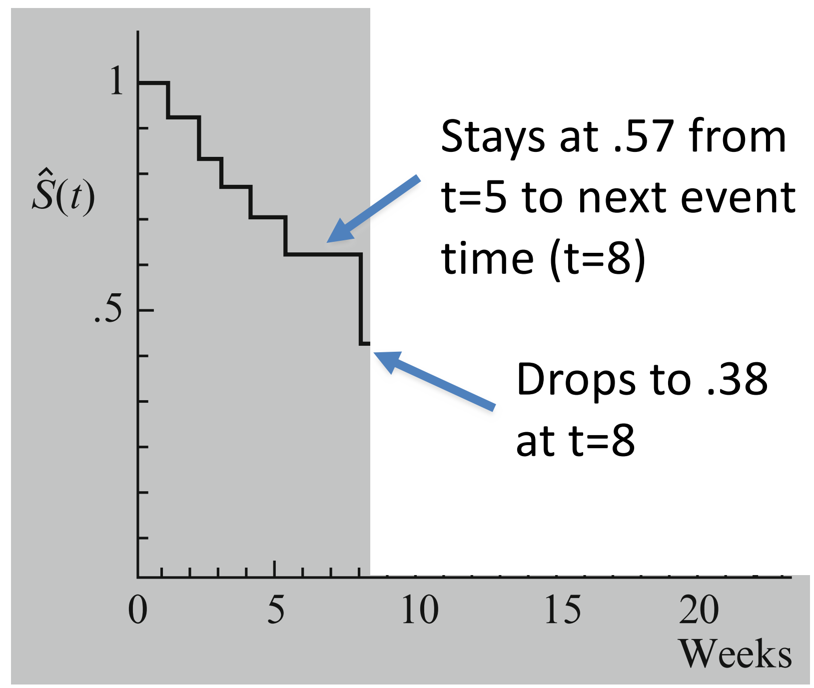
Group 2: no censored subjects

Group 2 (placebo)

$t_{(f)}$	n_f	m_f	q_f	$\hat{S}(t_{(f)})$
0	21	0	0	1
1	21	2	0	19/21 = .90
2	19	2	0	17/21 = .81
3	17	1	0	16/21 = .76
4	16	2	0	14/21 = .67
5	14	2	0	12/21 = .57
8	12	4	0	8/21 = .38
11	8	2	0	
12	6	2	0	
15	4	1	0	
17	3	1	0	
22	2	1	0	
23	1	1	0	

$$\hat{S}(t_{(f)}) = \frac{\#\text{ surviving past } t_{(f)}}{n}$$

Kaplan-Meier curve



[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

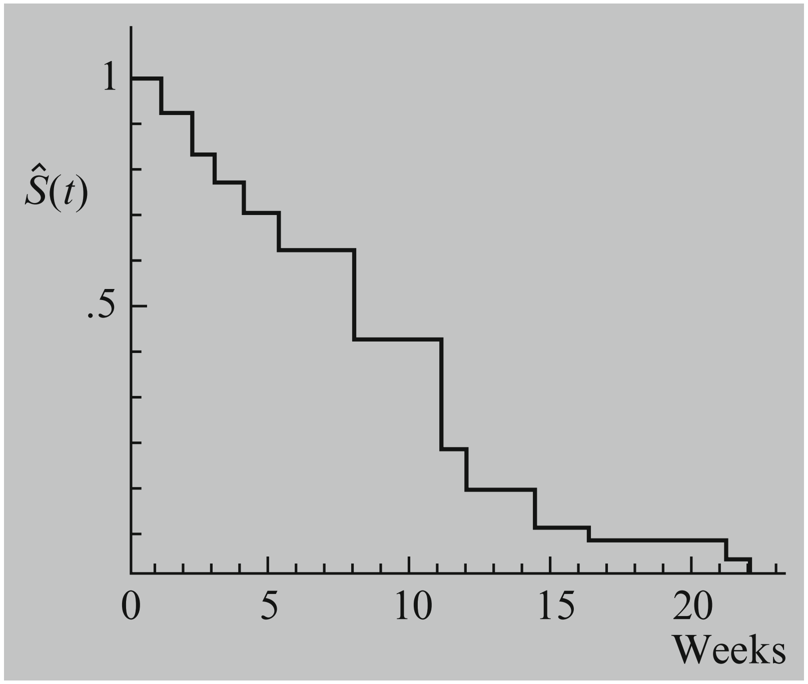
failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f

Group 2: no censored subjects
Group 2 (placebo)

$t_{(f)}$	n_f	m_f	q_f	$\hat{S}(t_{(f)})$
0	21	0	0	1
1	21	2	0	19/21 = .90
2	19	2	0	17/21 = .81
3	17	1	0	16/21 = .76
4	16	2	0	14/21 = .67
5	14	2	0	12/21 = .57
8	12	4	0	8/21 = .38
11	8	2	0	6/21 = .29
12	6	2	0	4/21 = .19
15	4	1	0	3/21 = .14
17	3	1	0	2/21 = .10
22	2	1	0	1/21 = .05
23	1	1	0	0/21 = .00

$$\hat{S}(t_{(f)}) = \frac{\#\text{ surviving past } t_{(f)}}{n}$$

Kaplan-Meier curve



[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

EXAMPLE

The data: remission times (weeks) for two groups of leukemia patients

Group 1 ($n = 21$) treatment	Group 2 ($n = 21$) placebo
6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+,	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

Note: + denotes censored

	# failed	# censored	Total
Group 1	9	12	21
Group 2	21	0	21

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f
Group 1 (treatment)			
$t_{(f)}$	n_f	m_f	q_f
0	21	0	0
6	21	3	1

At t=6 there are 3 failure events and 1 censoring event

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

EXAMPLE

The data: remission times (weeks) for two groups of leukemia patients

Group 1 ($n = 21$) treatment	Group 2 ($n = 21$) placebo
6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+,	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

Note: + denotes censored

	# failed	# censored	Total
Group 1	9	12	21
Group 2	21	0	21

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f
Group 1 (treatment)			
$t_{(f)}$	n_f	m_f	q_f
0	21	0	0
6	21	3	1
7	17	1	1
10			

Beginning at t=7 there are 17 subjects still in the study

Between t=7 and t=10 there is 1 failure event and 1 censoring event

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

EXAMPLE

The data: remission times (weeks) for two groups of leukemia patients

Group 1 (n = 21) treatment	Group 2 (n = 21) placebo
6, 6, 6, 7, 10,	1, 1, 2, 2, 3,
13, 16, 22, 23,	4, 4, 5, 5,
6+, 9+, 10+, 11+,	8, 8, 8, 8,
17+, 19+, 20+,	11, 11, 12, 12,
25+, 32+, 32+,	15, 17, 22, 23
34+, 35+,	

Note: + denotes censored

	# failed	# censored	Total
Group 1	9	12	21
Group 2	21	0	21

→

Group 1 (treatment)

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f
0	21	0	0
6	21	3	1
7	17	1	1
10	15	1	2
13			
16			

What are n_f, m_f, and q_f for $t_{(f)}$=13?

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

EXAMPLE

The data: remission times (weeks) for two groups of leukemia patients

Group 1 (n = 21) treatment	Group 2 (n = 21) placebo
6, 6, 6, 7, 10,	1, 1, 2, 2, 3,
13, 16, 22, 23,	4, 4, 5, 5,
6+, 9+, 10+, 11+,	8, 8, 8, 8,
17+, 19+, 20+,	11, 11, 12, 12,
25+, 32+, 32+,	15, 17, 22, 23
34+, 35+,	

Note: + denotes censored

	# failed	# censored	Total
Group 1	9	12	21
Group 2	21	0	21

→

Group 1 (treatment)

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f
0	21	0	0
6	21	3	1
7	17	1	1
10	15	1	2
13	12	1	0
16			

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

EXAMPLE

The data: remission times (weeks) for two groups of leukemia patients

Group 1 (n = 21) treatment	Group 2 (n = 21) placebo
6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+,	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

Note: + denotes censored

	# failed	# censored	Total
Group 1	9	12	21
Group 2	21	0	21

→

failure times, $t_{(f)}$ | # at risk n_f | # of failures m_f | # censored in $[t_{(f)}, t_{(f+1)})$, q_f

Group 1 (treatment)

$t_{(f)}$	n_f	m_f	q_f
0	21	0	0
6	21	3	1
7	17	1	1
10	15	1	2
13	12	1	0
16	11	1	3
22	7	1	0
23	6	1	5
>23	—	—	—

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f	
Group 1 (treatment)				
$t_{(f)}$	n_f	m_f	q_f	$\hat{S}(t_{(f)})$
0	21	0	0	1
6	21	3	1	$1 \times \frac{18}{21} = .8571$
7	17	1	1	
10	15	1	2	
13	12	1	0	
16	11	1	3	
22	7	1	0	
23	6	1	5	

Fraction at $t_{(f)}$: $\Pr(T > t_{(f)} \mid T \geq t_{(f)})$

Pr(surviving to time t) = Pr(surviving to time t-1)
x Pr(surviving to time t | survived to time t-1)

$$\hat{S}(t_{(f)}) = \hat{S}(t_{(f-1)}) \times \hat{Pr}(T > t_{(f)} \mid T \geq t_{(f)})$$

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f

Group 1 (treatment)

$t_{(f)}$	n_f	m_f	q_f	$\hat{S}\,(t_{(f)})$
0	21	0	0	1
6	21	3	1	$1 \times \frac{18}{21} = .8571$
7	17	1	1	$.8571 \times \frac{16}{17} = .8067$
10	15	1	2	**?**
13	12	1	0	
16	11	1	3	
22	7	1	0	
23	6	1	5	

Fraction at $t_{(f)}$: $\Pr(T > t_{(f)} \mid T \geq t_{(f)})$

Pr(surviving to time t) = Pr(surviving to time t-1)
x Pr(surviving to time t | survived to time t-1)

$$\hat{S}(t_{(f)}) = \hat{S}(t_{(f-1)}) \times \hat{Pr}(T > t_{(f)} \mid T \geq t_{(f)})$$

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f

Group 1 (treatment)

$t_{(f)}$	n_f	m_f	q_f	$\hat{S}(t_{(f)})$
0	21	0	0	1
6	21	3	1	$1 \times \frac{18}{21} = .8571$
7	17	1	1	$.8571 \times \frac{16}{17} = .8067$
10	15	1	2	$.8067 \times \frac{14}{15} = .7529$
13	12	1	0	
16	11	1	3	
22	7	1	0	
23	6	1	5	

Fraction at $t_{(f)}$: $\Pr(T > t_{(f)} \mid T \geq t_{(f)})$

Pr(surviving to time t) = Pr(surviving to time t-1)
x Pr(surviving to time t | survived to time t-1)

$$\hat{S}(t_{(f)}) = \hat{S}(t_{(f-1)}) \times \hat{Pr}(T > t_{(f)} \mid T \geq t_{(f)})$$

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

failure times, $t_{(f)}$	# at risk n_f	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f	

Group 1 (treatment)

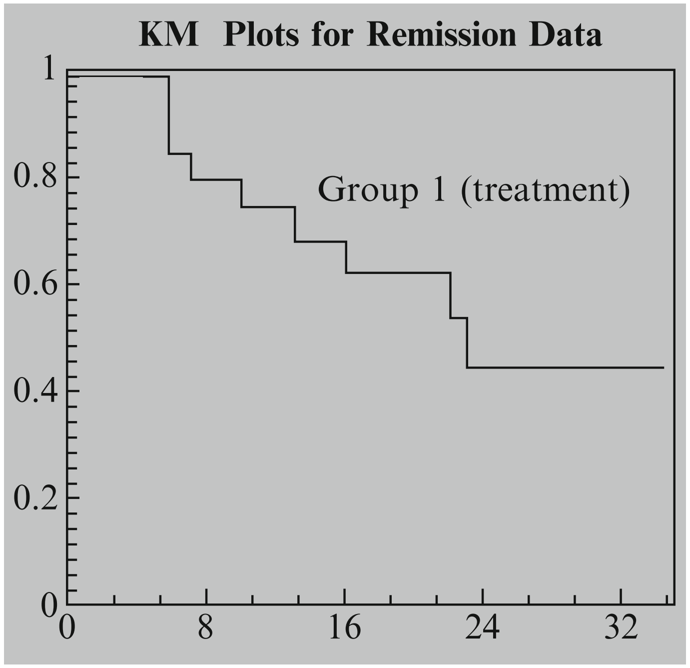
$t_{(f)}$	n_f	m_f	q_f	$\hat{S}(t_{(f)})$
0	21	0	0	1
6	21	3	1	$1 \times \frac{18}{21} = .8571$
7	17	1	1	$.8571 \times \frac{16}{17} = .8067$
10	15	1	2	$.8067 \times \frac{14}{15} = .7529$
13	12	1	0	$.7529 \times \frac{11}{12} = .6902$
16	11	1	3	$.6902 \times \frac{10}{11} = .6275$
22	7	1	0	$.6275 \times \frac{6}{7} = .5378$
23	6	1	5	$.5378 \times \frac{5}{6} = .4482$

Fraction at $t_{(f)}$: $\Pr(T > t_{(f)} \mid T \geq t_{(f)})$

Pr(surviving to time t) = Pr(surviving to time t-1)
x Pr(surviving to time t | survived to time t-1)

$$\hat{S}(t_{(f)}) = \hat{S}(t_{(f-1)}) \times \hat{Pr}(T > t_{(f)} \mid T \geq t_{(f)})$$

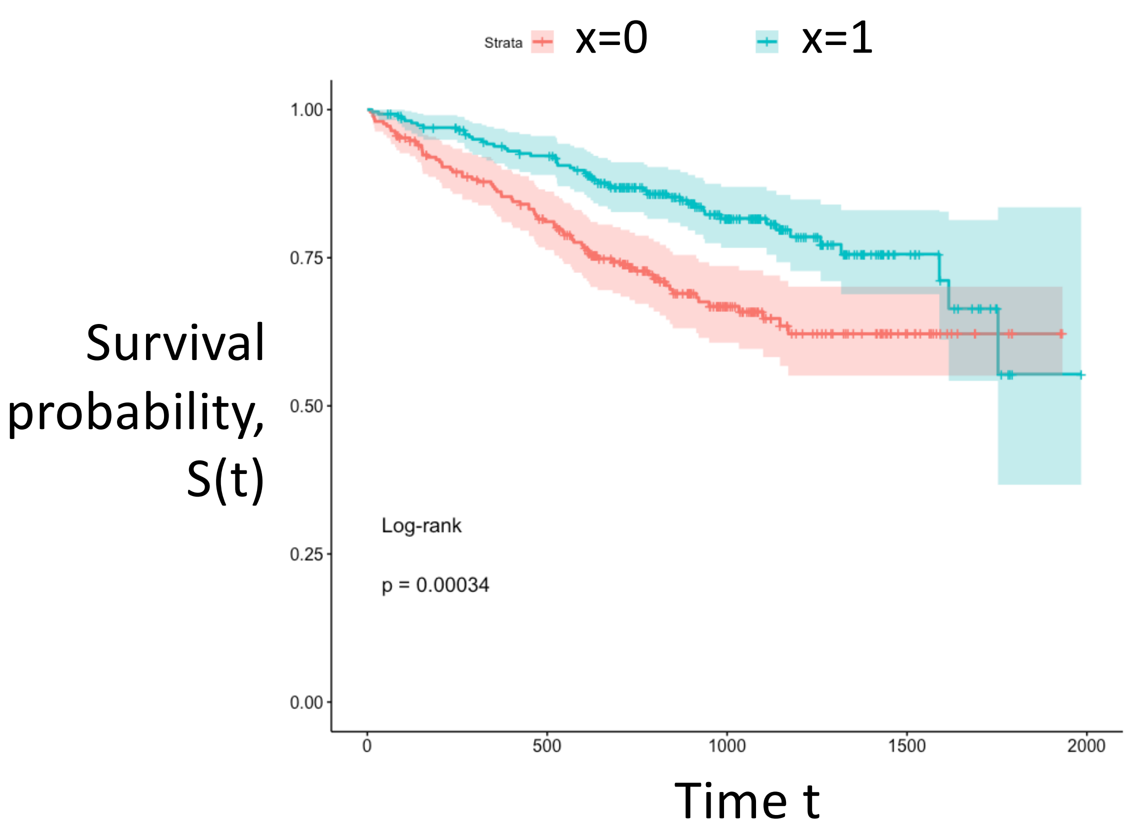
Kaplan-Meier curve



[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

Kaplan-Meier estimator of survival function S(t)=P(T > t)

- Example of a non-parametric method; good for unconditional density estimation



[Figure credit: Rebecca Boiarsky]

How do we compute confidence intervals for KM curves?

→ Use Greenwood's formula (see Ch. 2, VII, pgs 78-79)

Are these two curves statistically significantly different?

→ Use log-rank test (see Ch. 2, IV, pgs 67-73)

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

Relationship between probability density, hazard, and survival functions

Recall $S(t) = \int_{u=t}^{\infty} f(u)du$

The *hazard function* h(t) is:

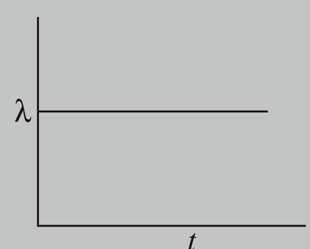
$$h(t) = \frac{-d[S(t)]/dt}{S(t)}$$

EXAMPLE

①

Constant hazard
(exponential model)

$h(t)$ for healthy persons



$h(t) = \lambda$ if and only if

$$S(t) = e^{-\lambda t}$$

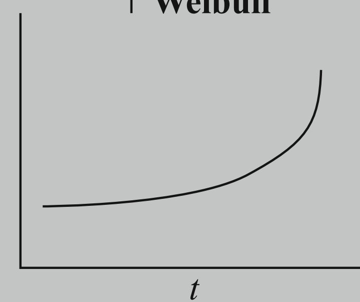
[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

EXAMPLE: (continued)

②

↑ Weibull

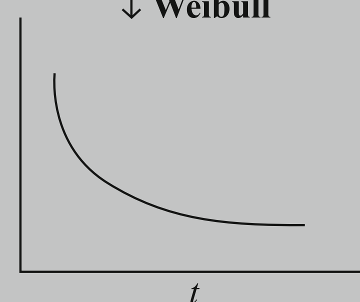
$h(t)$ for leukemia patients



③

↓ Weibull

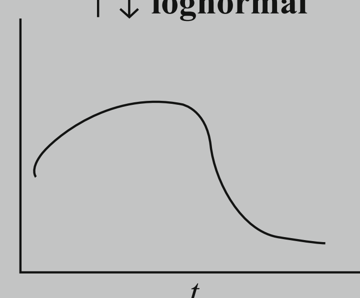
$h(t)$ for Persons recovering from surgery



④

↑↓ lognormal

$h(t)$ for TB patients



Commonly used parametric survival models

Table 2.1 Useful parametric distributions for survival analysis

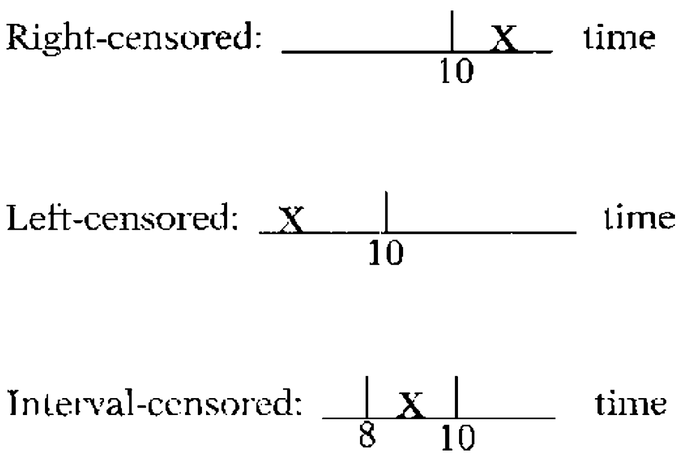
Distribution	Hazard rate $h(t)$	Survival function $S(t)$	Density function $f(t)$
Exponential ($\lambda > 0$)	λ	$\exp(-\lambda t)$	$\lambda \exp(-\lambda t)$
Weibull ($\lambda, \phi > 0$)	$\lambda\phi t^{\phi-1}$	$\exp(-\lambda t^{\phi})$	$\lambda\phi t^{\phi-1}\exp(-\lambda t^{\phi})$
Log-normal ($\sigma > 0, \mu \in R$)	$f(t)/S(t)$	$1 - \Phi\{(\ln t - \mu)/\sigma\}$	$\varphi\{(\ln t - \mu)/\sigma\}(\sigma t)^{-1}$
Log-logistic ($\lambda > 0, \phi > 0$)	$(\lambda\phi t^{\phi-1})/(1 + \lambda t^{\phi})$	$1/(1 + \lambda t^{\phi})$	$(\lambda\phi t^{\phi-1})/(1 + \lambda t^{\phi})^2$
Gamma ($\lambda, \phi > 0$)	$f(t)/S(t)$	$1 - I(\lambda t, \phi)$	$\{\lambda^{\phi}/\Gamma(\phi)\}t^{\phi-1}\exp(-\lambda t)$
Gompertz ($\lambda, \phi > 0$)	$\lambda e^{\phi t}$	$\exp\{\frac{\lambda}{\phi}(1 - e^{\phi t})\}$	$\lambda e^{\phi t}\exp\{\frac{\lambda}{\phi}(1 - e^{\phi t})\}$

We obtain **conditional** models $f(t \mid \boldsymbol{x}; \beta)$ by letting, e.g., $\lambda = \exp(\boldsymbol{\beta} \cdot \boldsymbol{x})$

[Ha, Jeong, Lee. Statistical Modeling of Survival Data with Random Effects. Springer 2017]

Likelihood function

Examples of Censored Subjects:



Barry, Gary, Larry, . . ., Outcome Distribution $f(t)$

Subject	Event Time
Barry	$t = 2$
Gary	$t > 8$ (right-censored)
Harry	$t = 6$
Carrie	$t < 2$ (left-censored)
Larry	$4 < t < 8$ (interval-censored)

Likelihood of observations:

$$L = f(2)x \int_{8}^{\infty} f(t)dt \times f(6) \times \int_0^2 f(t)dt \times \int_0^8 f(t)dt$$

(Barry × Gary × Harry × Carrie × Larry)

For **right-censored observations**, the corresponding integral is the survival function:

e.g., $\int_8^\infty f(t)dt = S(8)$

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

Maximum likelihood estimation

- Random variables T_i, C_i, X_i
  - C_i: censoring time of i'th individual
  - T_i: event time of i'th individual
  - X_i: features of i'th individual

- Observed data are $\{(t_i, d_i, \mathbf{x}_i)\}$, where $\mathbf{x}_i$ are the features and d_i is the indicator of whether the outcome is censored for the i'th individual
  - If $d_i=1$, then time t is the time of the event occurrence
  - If $d_i=0$, then time t is the time of censoring
  - Thus, $d_i=1[T_i < C_i]$ and $t_i = d_iT_i + (1-d_i)C_i$

- Formally, we assume **(a)** $C_i \perp T_i \mid X_i$, i.e. censoring time is (conditionally) independent of event time, and **(b)** all individuals' are independent

Maximum likelihood estimation

- Two kinds of observations: censored and right uncensored
- Putting the two together, we get the log-likelihood is, where n=# data points:

$$\sum_{i=1}^{n} [d_i \log f(t_i \mid \mathbf{x}_i; \beta) + (1 - d_i) \log S(t_i \mid \mathbf{x}_i; \beta)]$$

- Maximize via gradient or stochastic gradient ascent!

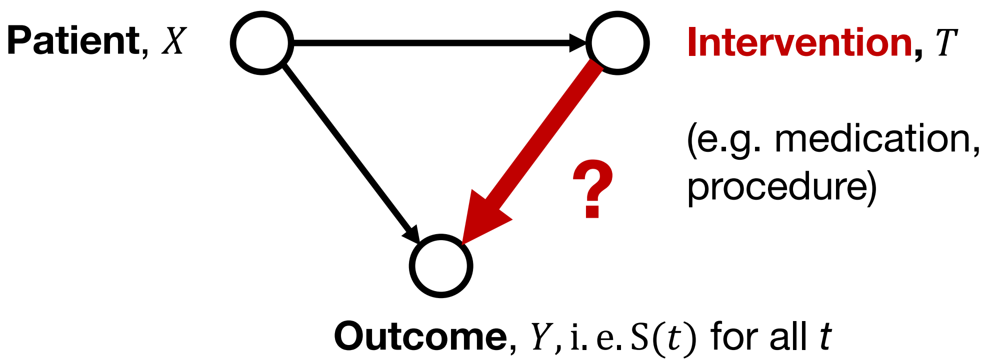
Distribution	Hazard rate $h(t)$	Survival function $S(t)$	Density function $f(t)$
Exponential ($\lambda > 0$)	λ	$\exp(-\lambda t)$	$\lambda \exp(-\lambda t)$

Suppose $\lambda = \exp(\boldsymbol{\beta} \cdot \boldsymbol{x_i})$. Then: $f(t_i \mid \mathbf{x}_i; \beta) = \exp(\beta \cdot \mathbf{x}_i) \exp(-\exp(\beta \cdot \mathbf{x}_i) t_i)$

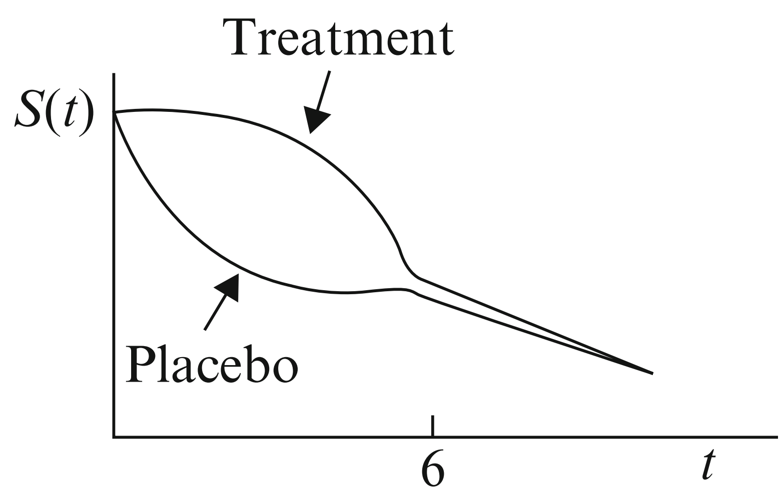
$$S(t_i \mid \mathbf{x}_i; \beta) = \exp(-\exp(\beta \cdot \mathbf{x}_i) t_i)$$

Example: estimating (heterogeneous) treatment effects

One can perform *covariate adjustment* using survival models



Goal: predict individual's survival curve



Must include all confounding factors in *X* (needed for adjustment formula to hold)

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

Example: estimating (heterogeneous) treatment effects

T=0		T=1	
Group 1		Group 2	
t (weeks)	log WBC	t (weeks)	log WBC
6	2.31	1	2.80
6	4.06	1	5.00
6	3.28	2	4.91
7	4.43	2	4.48
10	2.96	3	4.01
13	2.88	4	4.36
16	3.60	4	2.42
22	2.32	5	3.49
23	2.57	5	3.97
6+	3.20	8	3.52
9+	2.80	8	3.05
10+	2.70	8	2.32
11+	2.60	8	3.26
17+	2.16	11	3.49
19+	2.05	11	2.12
20+	2.01	12	1.50
25+	1.78	12	3.06
32+	2.20	15	2.30
32+	2.53	17	2.95
34+	1.47	22	2.73
35+	1.45	23	1.97

Same leukemia data as before, from Freireich et. al. *Blood*, **21**: 699-716, 1963.

[Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer, 2005]

Example: estimating (heterogeneous) treatment effects

T=0		T=1	
Group 1		Group 2	
t (weeks)	log WBC	*t* (weeks)	log WBC
6	2.31	1	2.80
6	4.06	1	5.00
6	3.28	2	4.91
7	4.43	2	4.48
10	2.96	3	4.01
13	2.88	4	4.36
16	3.60	4	2.42
22	2.32	5	3.49
23	2.57	5	3.97
6+	3.20	8	3.52
9+	2.80	8	3.05
10+	2.70	8	2.32
11+	2.60	8	3.26
17+	2.16	11	3.49
19+	2.05	11	2.12
20+	2.01	12	1.50
25+	1.78	12	3.06
32+	2.20	15	2.30
32+	2.53	17	2.95
34+	1.47	22	2.73
35+	1.45	23	1.97

Treatment indicator

	Individual #	*t* (weeks)	d	X_1 (Group)	X_2 (log WBC)
Group 1	1	6	1	1	2.31
	2	6	1	1	4.06
	3	6	1	1	3.28
	4	7	1	1	4.43
	5	10	1	1	2.96
	6	13	1	1	2.88
	7	16	1	1	3.60
	8	22	1	1	2.32
	9	23	1	1	2.57
	10	6	0	1	3.20
	11	9	0	1	2.80
	12	10	0	1	2.70
	13	11	0	1	2.60
	14	17	0	1	2.16
	15	19	0	1	2.05
	16	20	0	1	2.01
	17	25	0	1	1.78
	18	32	0	1	2.20
	19	32	0	1	2.53
	20	34	0	1	1.47
	21	35	0	1	1.45
Group 2	22	1	1	0	2.80
	23	1	1	0	5.00
	24	2	1	0	4.91
	25	2	1	0	4.48
	26	3	1	0	4.01
	27	4	1	0	4.36
	28	4	1	0	2.42
	29	5	1	0	3.49
	30	5	1	0	3.97
	31	8	1	0	3.52
	32	8	1	0	3.05
	33	8	1	0	2.32
	34	8	1	0	3.26
	35	11	1	0	3.49
	36	11	1	0	2.12
	37	12	1	0	1.50
	38	12	1	0	3.06
	39	15	1	0	2.30
	40	17	1	0	2.95
	41	22	1	0	2.73
	42	23	1	0	1.97

Evaluation for survival modeling

- Concordance-index (also called C-statistic): look at model's ability to predict *relative* survival times:

$$\hat{c} = \frac{1}{n_c} \sum_{i:d_i=1} \sum_{t_i<t_j} 1[\hat{y}_i < \hat{y}_j] \qquad n_c = \sum_{i:d_i=1} \sum_{t_i<t_j} 1 \qquad \hat{y}_i = \mathbb{E}_{f(T|\mathbf{x}_i;\beta)}[T]$$

Example:

Distribution	Hazard rate $h(t)$	Survival function $S(t)$	Density function $f(t)$
Exponential ($\lambda > 0$)	λ	$\exp(-\lambda t)$	$\lambda \exp(-\lambda t)$

The mean of an exponential distribution is $1/\lambda$.
Suppose we parameterize with $\lambda = \exp(\boldsymbol{\beta} \cdot \boldsymbol{x})$. Then $\hat{y}_i = \exp(-\boldsymbol{\beta} \cdot \boldsymbol{x}_i)$.

[Wang, Li, Reddy. Machine Learning for Survival Analysis: A Survey. 2017]

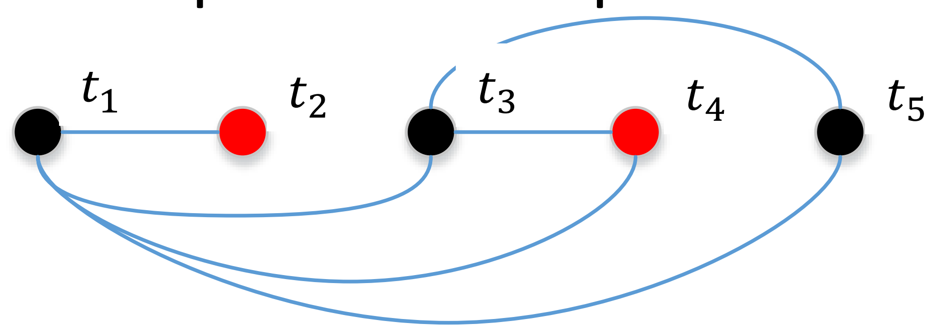
Evaluation for survival modeling

- Concordance-index (also called C-statistic): look at model's ability to predict *relative* survival times:

$$\hat{c} = \frac{1}{n_c} \sum_{i:d_i=1} \sum_{t_i<t_j} 1[\hat{y}_i < \hat{y}_j] \qquad n_c = \sum_{i:d_i=1} \sum_{t_i<t_j} 1 \qquad \hat{y}_i = \mathbb{E}_{f(T|\mathbf{x}_i;\beta)}[T]$$

- Illustration – blue lines denote pairwise comparisons:

Black = uncensored ($d_i = 1$)
Red = censored ($d_i = 0$)



- Equivalent to AUC for binary variables and no censoring

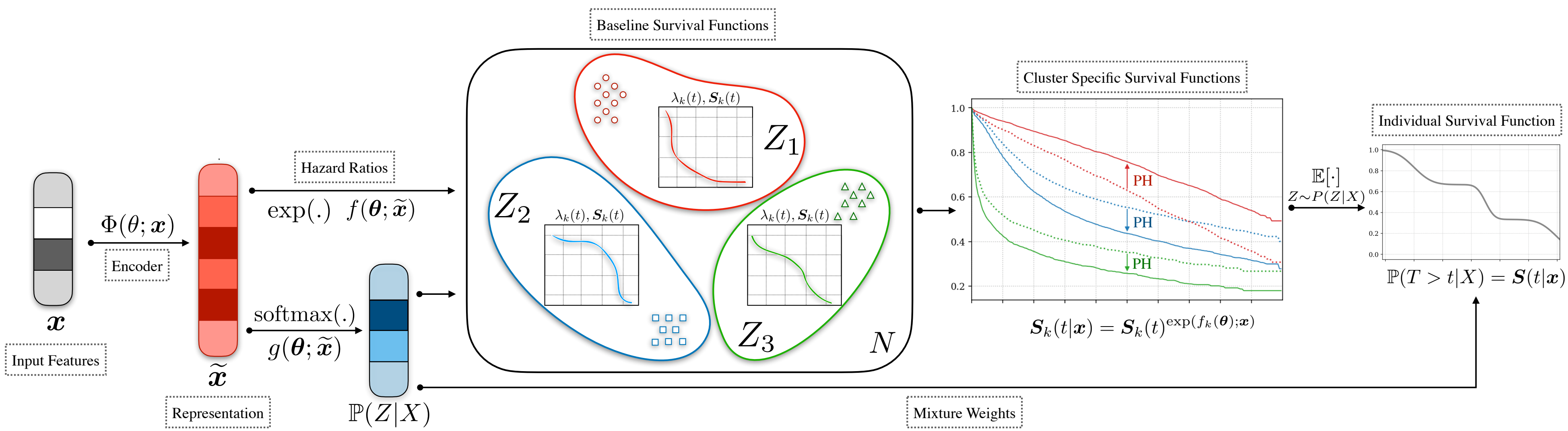
[Wang, Li, Reddy. Machine Learning for Survival Analysis: A Survey. 2017]

Comments on survival modeling

- Could also evaluate:
  - Mean-squared error for uncensored individuals
  - Held-out (censored) likelihood
  - Derive binary classifier from learned model and check calibration
- *Partial likelihood* estimators (e.g. for cox-proportional hazards models) can be much more data efficient (see Ch. 3 of book)

$$\mathscr{P}\mathscr{L}(\boldsymbol{\theta}) = \prod_{i:\delta_i=1} \frac{\exp\big(f(\boldsymbol{\theta};\boldsymbol{x}_i)\big)}{\sum_{j\in\mathscr{R}(t_i)} \exp\big(f(\boldsymbol{\theta};\boldsymbol{x}_j)\big)}$$

Deep Cox mixtures for survival regression



$$\mathscr{L}(\boldsymbol{\theta}, \boldsymbol{\Lambda}_k) = \prod_{i=1}^{|\mathscr{D}|} \int_Z (\boldsymbol{\lambda}(u_i|\boldsymbol{x}_i))^{\delta_i} \boldsymbol{S}_k(u_i|\boldsymbol{x}_i) \mathbb{P}(Z = k|\boldsymbol{x}_i).$$

$$\text{where, } \boldsymbol{\lambda}(u_i|\boldsymbol{x}_i) = \boldsymbol{\lambda}_k(u_i) \exp\left(f_k(\boldsymbol{\theta}, \boldsymbol{x}_i)\right), \quad \boldsymbol{S}_k(u_i|\boldsymbol{x}_i) = \boldsymbol{S}_k(u_i)^{\exp\left(f_k(\boldsymbol{\theta}; \boldsymbol{x}_i)\right)}$$

$$\text{and, } \mathbb{P}(Z = k|X = \boldsymbol{x}_i) = \text{softmax}\big(g(\boldsymbol{\theta}; \boldsymbol{x}_i)\big)$$

[Nagpal et al., Conference on Machine Learning for Healthcare, 2021]

Conclusion

- Last lecture and this, we tackled two challenges that commonly arise in supervised learning in health care
  1. Classification with noisy labels
  2. Regression with censored labels
- Strong assumptions allowed us to develop simple solutions
  - $X \perp \tilde{Y} \mid Y$ (noise rate constant for all examples)
  - $C \perp T \mid X$ (censoring time independent of survival time)
- Can we relax these assumptions? Can we do survival modeling with noisy labels?

References

- *Recommended starting place:* Kleinbaum & Klein. Survival Analysis: A Self-Learning Text. Springer Statistics for biology and Health, 2005
- *Additional detail:* Kalbfleisch & Prentice, The Statistical Analysis of Failure Time Data, Wiley 2002 [MIT proxy]
- Ishwaran et al., Random Survival Forests. The Annals of Applied Statistics, 2008
- Alaa and van der Schaar. Deep multi-task gaussian processes for survival analysis with competing risks. NeurIPS, 2017