Homework 3

Steven Chiou

Due date: Tuesday, November 13

1. **Textbook problem 2.1** Listed below are values of survival time in years for 6 males and 6 females from the *WHAS100* study. Right-censored times are denoted by a "+" as a superscript.

```
Males: 1.2, 3.4, 5.0<sup>+</sup>, 5.1, 6.1, 7.1
Females: 0.4, 1.2, 4.3, 4.9, 5.0, 5.1<sup>+</sup>
```

Using these data, compute the following without a software (survival) package:

- a. (2 pts) The Kaplan-Meier estimate of the survival function for each gender.
- b. (2 pts) Pointwise 95 % confidence intervals for the survival functions estimated in problem (1a).
- c. (2 pts) Pointwise 95 % confidence interval estimates of the 50th percentile of the survival time distribution for each gender.
- d. (2 pts) The estimated mean survival time for each gender using all available times, upto 7.1.
- e. (2 pts) A graph of the estimated survival functions for each gender computed in problem (1a) along with the point wise and overall 95 % limits computed in problem (1b).

Solution

First prepare the data:

```
# A tibble: 12 x 3
            cen gender
    Time
   <dbl> <dbl> <fct>
     1.2
 1
              1 male
 2
     3.4
              1 male
 3
     5
              0 male
 4
     5.1
              1 male
 5
     6.1
              1 male
 6
     7.1
              1 male
 7
     0.4
              1 female
8
              1 female
     1.2
9
     4.3
              1 female
10
     4.9
              1 female
11
     5
              1 female
12
     5.1
              0 female
```

a. There are no ties in each group, so the Kaplan-Meier estimates are easy to obtain.

```
> dat1 <- subset(dat, gender == "male") %>% mutate(km = cumprod((6:1 - cen) / 6:1))
> dat2 <- subset(dat, gender == "female") %>% mutate(km = cumprod((6:1 - cen) / 6:1))
```

We will display the result in the next part. b. First, compute the standard error, construct the confidence intervals endpoints on log-scale then transform the endpoints back.

Check with survfit

```
> dat1 %>% filter(cen > 0) %>% as.data.frame()
  Time cen gender
                                           lower upper
                         km
                                   se
  1.2
         1
             male 0.8333333 0.1521452 0.5826548
  3.4
             male 0.6666667 0.1924501 0.3786065
         1
                                                     1
  5.1
            male 0.4444444 0.2222222 0.1668079
                                                     1
  6.1
             male 0.2222222 0.1924501 0.0407029
         1
                                                     1
             male 0.0000000
  7.1
         1
                                  NaN
                                             NaN
                                                   NaN
> dat2 %>% filter(cen > 0) %>% as.data.frame()
  Time cen gender
                         km
                                   se
                                            lower
                                                     upper
  0.4
         1 female 0.8333333 0.1521452 0.58265480 1.000000
         1 female 0.6666667 0.1924501 0.37860646 1.000000
  1.2
  4.3
         1 female 0.5000000 0.2041241 0.22463035 1.000000
4
  4.9
         1 female 0.3333333 0.1924501 0.10750714 1.000000
  5.0
         1 female 0.1666667 0.1521452 0.02784913 0.997438
> summary(survfit(Surv(Time, cen) ~ gender, data = dat))
```

Call: survfit(formula = Surv(Time, cen) ~ gender, data = dat)

gender=male

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%	CI
1.2	6	1	0.833	0.152		0.5827			1
3.4	5	1	0.667	0.192		0.3786			1
5.1	3	1	0.444	0.222		0.1668			1
6.1	2	1	0.222	0.192		0.0407			1
7.1	1	1	0.000	NaN		NA			NA

gender=female

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0.4	6	1	0.833	0.152		0.5827		1.000
1.2	5	1	0.667	0.192		0.3786		1.000
4.3	4	1	0.500	0.204		0.2246		1.000
4.9	3	1	0.333	0.192		0.1075		1.000
5.0	2	1	0.167	0.152		0.0278		0.997

c. The median survival times and the corresponding 95% confidence intervals are 5.1 (3.4, NA) and 4.3 (1.2, NA) for males and females, respectively.

d. The estimated mean survival times for the two groups are

```
Male:
```

```
> with(dat1, integrate(f = approxfun(Time, km, "constant", yleft = 1, yright = min(km)), 0, 7.1))
```

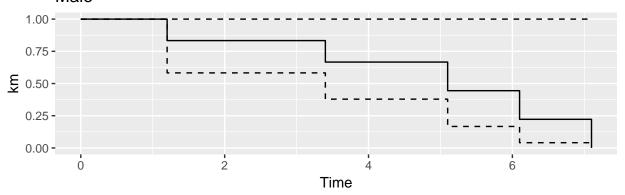
4.833335 with absolute error < 0.00048

Female:

```
> with(dat2, integrate(f = approxfun(Time, km, "constant", yleft = 1, yright = min(km)), 0, 7.1))
```

3.816665 with absolute error < 0.00038

Male



Female

