BST 222 - Homework 2

Xiaowei Zeng

October 19, 2023

Problem 1

The SSA standard mortality table for 2019 is found at Mortality Table. Compute the net present value of a series of two payments of \$1000 at the end of year 1 and at the end of year 2, each payment contingent on at least one of two people being alive at the time of the scheduled payment. Individual 1 is male aged 80 and individual 2 is female aged 75. You can assume that the chance of surviving a year at a given age is as given in the table. Use a discount rate of 5%.

Table 1: Mortality Table (Extracted)

Individual	Gender	Age	Death probability
1	Male	80	0.056237
		81	0.062360
2	Female	75	0.024080
		76	0.026831

Solution: Pension calculation:

$$V = \sum_{t=1}^{\infty} \frac{M}{(1+i)^t} \prod_{s=0}^{t-1} P(A+S)$$

where M is the annual amount paid at the end of the year, i is the discount rate, A is the initial age, and $P(x)$ is the probability of surviving until the end of year at age x.

a. At the end of year 1: $M = 1000, i = 0.05, t = 1$.

$$\begin{aligned} P(\text{at least one survives}) &= 1 - P(\text{both die}) \\ &= 1 - P(1 \text{ dies}) \times P(2 \text{ dies}) \\ &= 1 - 0.056237 \times 0.024080 \\ &= 0.9986458 \end{aligned}$$

$$V_1 = \frac{1000}{1+0.05} * 0.9986458 = 951.0912$$

$\therefore$ the net present value is \$951.0912 at the end of year 1.

b. At the end of year 2: $M = 1000, i = 0.05, t = 2$.

Case 1: Both survive 2 years.

$$\begin{aligned} P(\text{Case 1}) = &\ P(\text{1 survives in year 1}) \times P(\text{1 survives in year 2}) \\ &\times P(\text{2 survives in year 1}) \times P(\text{2 survives in year 2}) \\ = &\ (1-0.056237) \times (1-0.062360) \times (1-0.024080) \times (1-0.026831) \\ = &\ 0.84043 \end{aligned}$$

Case 2: Individual 1 survives 2 years, individual 2 dies in year 1.

$$\begin{aligned} P(\text{Case 2}) = &\ P(\text{1 survives in year 1}) \times P(\text{1 survives in year 2}) \\ &\times P(\text{2 dies in year 1}) \\ = &\ (1-0.056237) \times (1-0.062360) \times 0.024080 \\ = &\ 0.02130863 \end{aligned}$$

Case 3: Individual 1 survives 2 years, individual 2 dies in year 2.

$$\begin{aligned} P(\text{Case 3}) = &\ P(\text{1 survives in year 1}) \times P(\text{1 survives in year 2}) \\ &\times P(\text{2 survives in year 1}) \times P(\text{2 dies in year 2}) \\ = &\ (1-0.056237) \times (1-0.062360) \times (1-0.024080) \times 0.026831 \\ = &\ 0.02317129 \end{aligned}$$

Case 4: Individual 2 survives 2 years, individual 1 dies in year 1.

$$\begin{aligned} P(\text{Case 4}) = &\ P(\text{2 survives in year 1}) \times P(\text{2 survives in year 2}) \\ &\times P(\text{1 dies in year 1}) \\ = &\ (1-0.024080) \times (1-0.026831) \times 0.056237 \\ = &\ 0.05341025 \end{aligned}$$

Case 5: Individual 2 survives 2 years, individual 1 dies in year 2.

$$\begin{aligned} P(\text{Case 1}) = &\ P(\text{2 survives in year 1}) \times P(\text{2 survives in year 2}) \\ &\times P(\text{1 survives in year 1}) \times P(\text{1 dies in year 2}) \\ = &\ (1-0.024080) \times (1-0.026831) \times (1-0.056237) \times 0.062360 \\ = &\ 0.05589482 \end{aligned}$$

Therefore, we have

$$
\begin{aligned}
P(\text{at least one survives}) &= \sum_{i=1}^{5} P(\text{Case } i) \\
&= 0.84043 + 0.02130863 + 0.02317129 \\
&\quad + 0.05341025 + 0.05589482 \\
&= 0.994215
\end{aligned}
$$

$$
\begin{aligned}
V_2 &= \frac{1000}{(1+0.05)^2} * 0.994215 = 901.7823 \\
V_1 + V_2 &= 951.0912 + 901.7823 = 1852.8735
\end{aligned}
$$

$\therefore$ the net present value is \$1852.8735 at the end of year 2.

Problem 2

The data set `tongue` from `KMsurv` is described in KM section 1.11.

```
Description:
        The 'tongue' data frame has 80 rows and 3 columns.
Format:
        This data frame contains the following columns:
        1. type: Tumor DNA profile (1=Aneuploid Tumor, 2=Diploid Tumor)
        2. time: Time to death or on-study time, weeks
        3. delta: Death indicator (0=alive, 1=dead)
```

(a) Construct and plot the Kaplan-Meier survival function estimates for the two types of tumors, first without the confidence limits and then with the confidence limits.

(b) Use `survdiff` to test the hypothesis that the two true survival curves are the same. Interpret the results.

(c) Compare the two Kaplan-Meier curves to the Nelson-Aalen estimates graphically. Interpret the results.

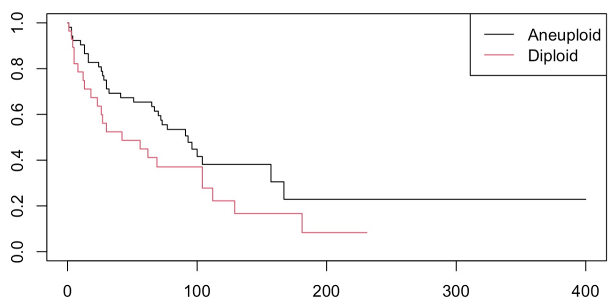
(d) To investigate the proportionality of the two hazard curves, compute and plot the Nelson-Aalen cumulative hazards, the ratio of the cumulative hazards, and the smoothed hazards using `muhaz`. Does it look as if the hazards are proportional?

Solution: Load the packages and data in R.

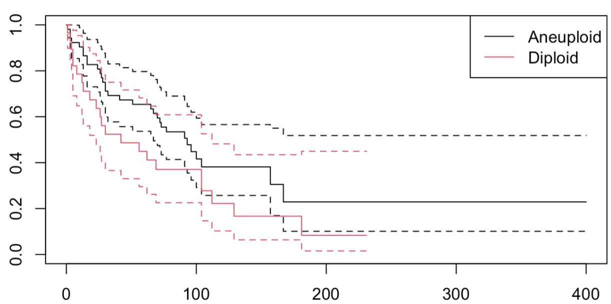
```r
library(KMsurv)
library(survival)
library(muhaz)
data('tongue')
dfsurv = Surv(tongue$time, tongue$delta) # create a survival object
```

(a) Plot the Kaplan-Meier curve for the two types of tumor.

```r
# Without CI.
plot(survfit(dfsurv ~ type, data = tongue), col = 1:2)
legend('topright', c("Aneuploid", 'Diploid Tumor'), col = 1:2)
```




```r
# With CI.
plot(survfit(dfsurv ~ type, data = tongue), col = 1:2, conf.int = T)
legend('topright', c("Aneuploid", 'Diploid Tumor'), col = 1:2)
```



(b) Suppose that the significance level α is 0.05.

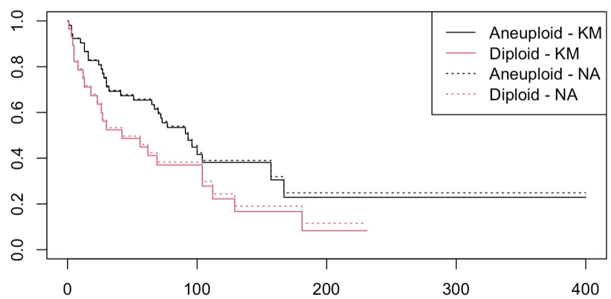
```r
survdiff(dfsurv ~ type, data = tongue)
```

```
Call:
survdiff(formula = dfsurv ~ type, data = tongue)
         N    Observed    Expected    (O-E)^2/E    (O-E)^2/V
type=1   52   31          36.6        0.843        2.79
type=2   28   22          16.4        1.873        2.79
Chisq= 2.8 on 1 degrees of freedom, p= 0.09
```

The p-value of the test is 0.09 > 0.05, so we **cannot reject** the null hypothesis that the two true survival curves are the same.

(c) We can use the option `type = 'fleming-harrington'` in `survfit` to get the Nelson-Aalen estimates. Note that Fleming Harrington method will reduce to Nelson-Aalen method when the data is unweighted.

```
plot(survfit(dfsurv ~ type, data = tongue), col = 1:2)
lines(survfit(dfsurv ~ type, data = tongue, type = 'fleming-harrington'),
      col = 1:2, lty = 3)
legend('topright', c("Aneuploid - KM", 'Diploid - KM',
                     "Aneuploid - NA", 'Diploid - NA'),
      col = c(1:2, 1:2), lty = c(1, 1, 3, 3), lwd = 1)
```



From the graph above, we can learn that, the Nelson-Aalen estimates for the survival function are a bit larger than the Kaplan-Meier estimates for each type of tumor. This is because

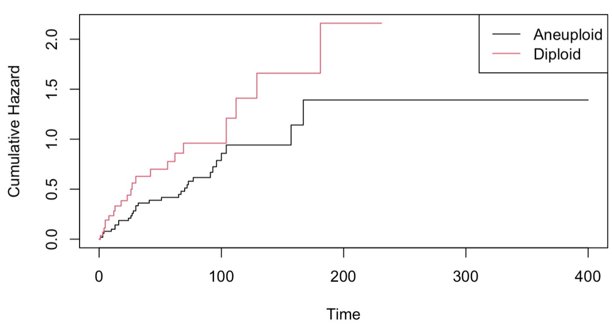
$$\hat{S}_{KM}(t) = \prod_{t_i<t}\left[1 - \frac{d_i}{Y_i}\right], \hat{S}_{NA}(t) = \prod_{t_i<t}\exp\left(-\frac{d_i}{Y_i}\right).$$

$$\because \exp(-x) - (1-x) > 0, \forall x \in (0,1),$$

$$\therefore \hat{S}_{NA}(t) > \hat{S}_{KM}(t), \forall x \in (0,1).$$

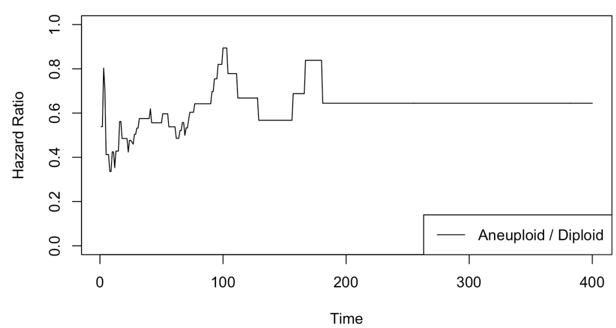
(d) Nelson-Aalen cumulative hazards:

```
plot(survfit(dfsurv ~ type, data = tongue, type = 'fleming-harrington'),
     col = 1:2, fun = 'cumhaz', xlab = 'Time', ylab = 'Cumulative Hazard')
legend('topright', c("Aneuploid", 'Diploid'), col = 1:2, lwd = 1)
```



Ratio of the cumulative hazards:

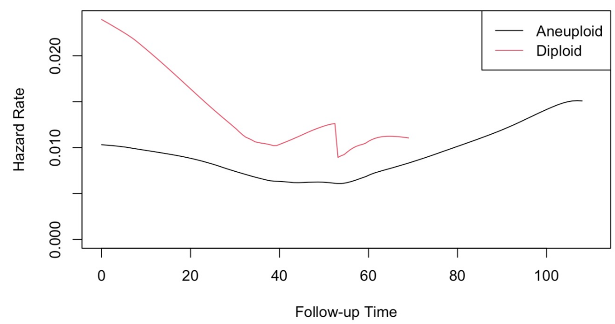
```
naft = survfit(dfsurv ~ type, data = tongue, type = 'fleming-harrington')
timevec = 1:400
sf1 = stepfun(naft[1]$time, c(1, naft[1]$surv))
sf2 = stepfun(naft[2]$time, c(1, naft[2]$surv))
cumhaz1 = - log(sf1(timevec))
cumhaz2 = - log(sf2(timevec))
plot(timevec, cumhaz1 / cumhaz2, type = 'l', ylim = c(0, 1),
      xlab = 'Time', ylab = 'Hazard Ratio')
legend('bottomright', 'Aneuploid / Diploid', lwd = 1)
```



The graph shows that the ratio of Aneuploid / Diploid fluctuates slightly between 0.4 and 0.8 and is always below 1, so we can assume that the hazards are proportional.

Smoothed hazards:

```
plot(muhaz(tongue$time, tongue$delta, tongue$type == 2), col = 2,
      xlim = c(0, 110))
lines(muhaz(tongue$time, tongue$delta, tongue$type == 1), col = 1)
legend('topright', c("Aneuploid", 'Diploid'), col = 1:2, lwd = 1)
```



According to the graph, the two hazard curves have similar trends and are roughly paralleled, so we can assume that the hazards are proportional.