Biostatistics Task 2

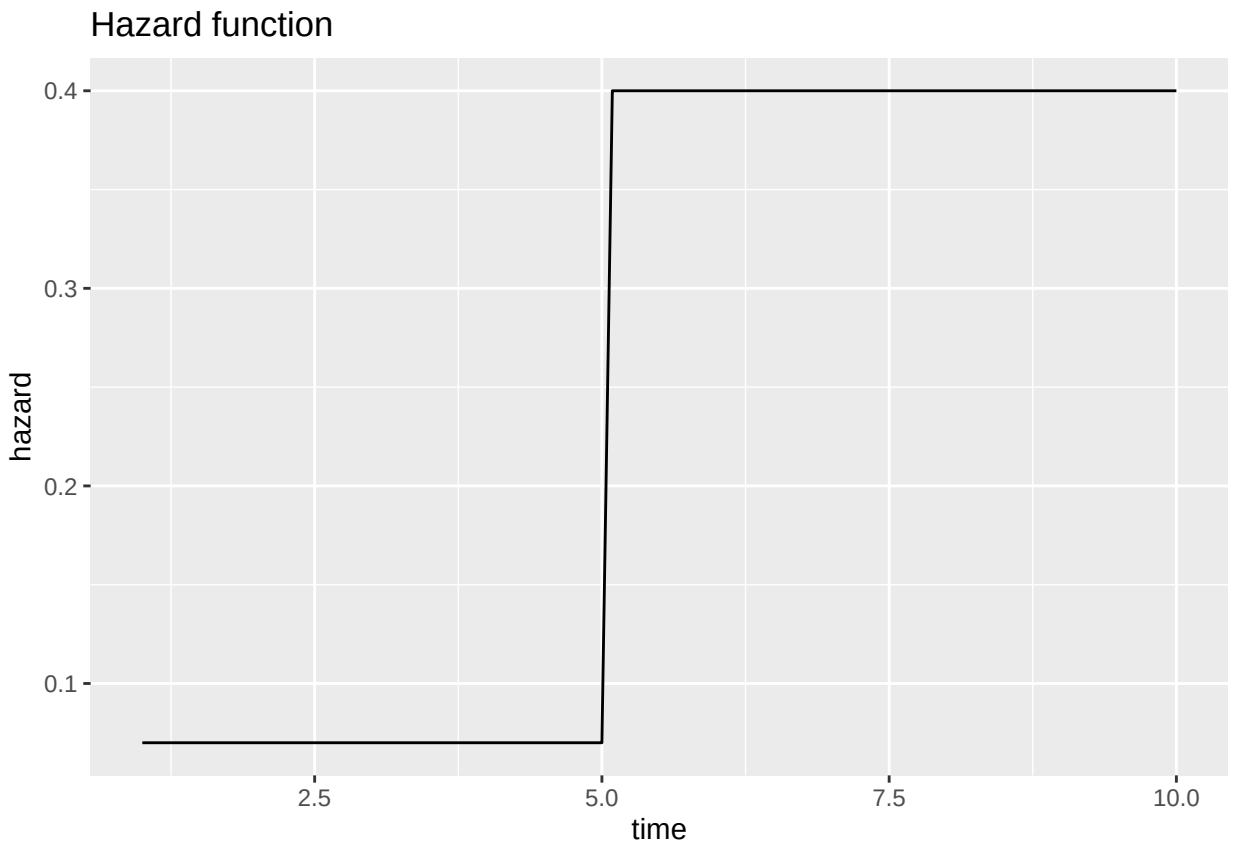
Danyu Zhang & Daniel Alonso

May 21st, 2021

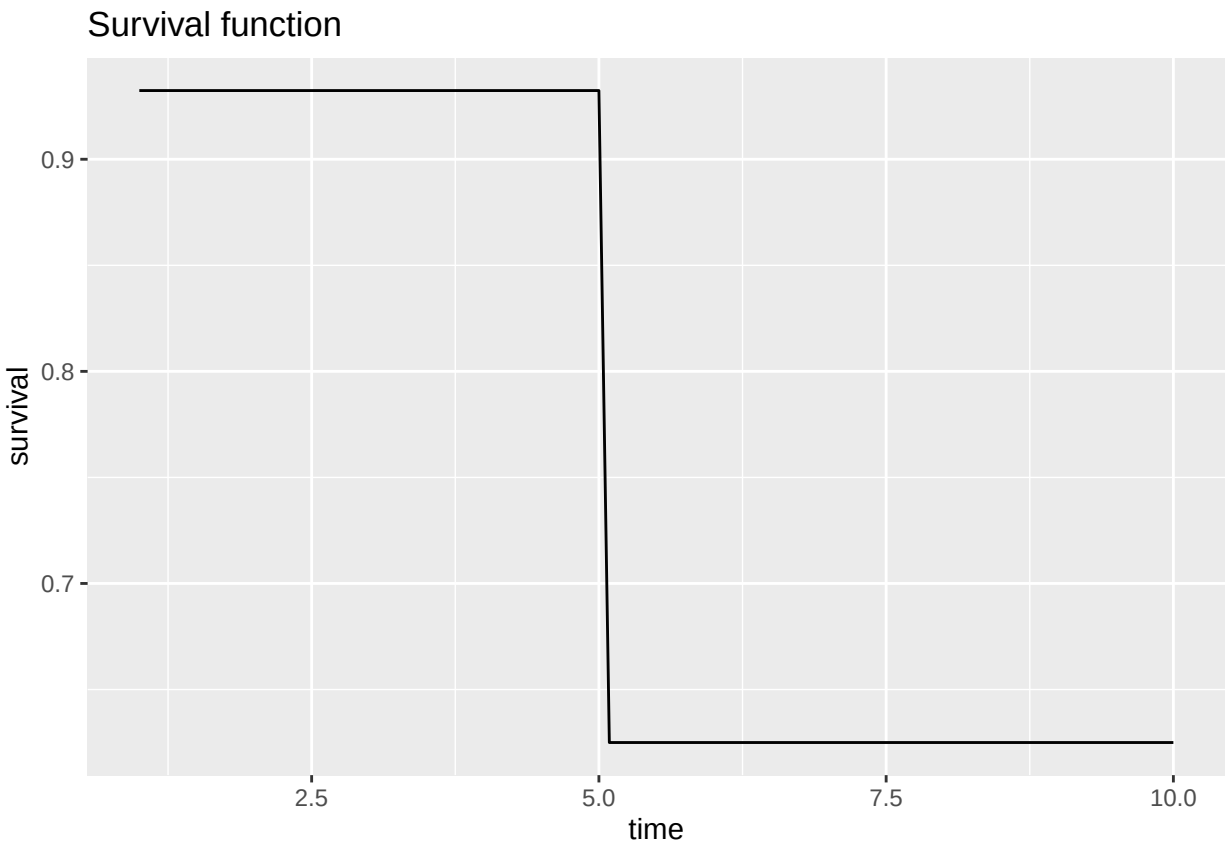
```
library(survival)
library(ggplot2)
```

Exercise 1

Hazard plot



Survival plot

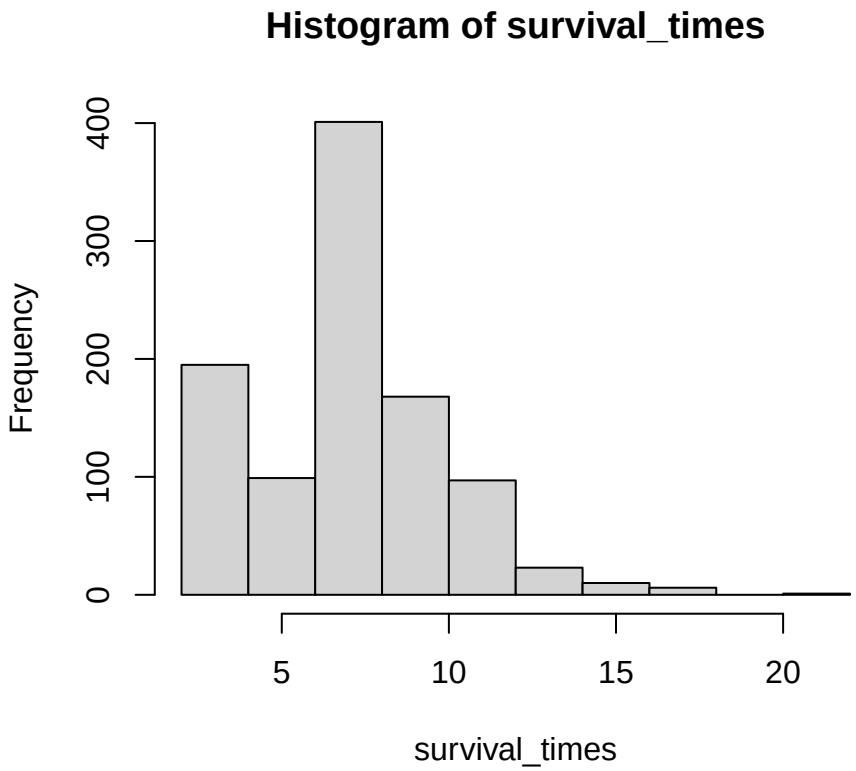


Survival times simulation

```
# number of trials
trials <- 1000
time_length <- 1000
survival_prob <- c(1, rep(0.9323938, 5), rep(0.6250023, time_length-6))

# t is in seconds
survival_times <- rep(0,trials)
for (i in 1:trials) {
    time = 1
    while (time < time_length) {
        if (runif(1,0,1) > survival_prob[time]) {
            break
        } else {
            time = time + 1
        }
    }
    survival_times[i] <- time
}
```

Histogram for survival times



Median survival time

```
#> [1] 7
```

Exercise 2

Given the following density function:

$$f(y) = (\lambda_0 + \lambda_1 y)e^{-\lambda_0 y - \frac{1}{2}\lambda_1 y^2}$$

We obtain the survival function as follows:

$$\begin{aligned} S(t) = P(T > t) &= \int_t^\infty (\lambda_0 + \lambda_1 y)e^{-\lambda_0 y - \frac{1}{2}\lambda_1 y^2} dy \\ &= \lim_{b\to\infty}[-e^{\frac{\lambda_1 b^2}{2} - \lambda_0 b}] + e^{\frac{-\lambda_1 t^2}{2} - \lambda_0 t} \\ &= 0 + e^{\frac{-\lambda_1 t^2}{2} - \lambda_0 t} \\ S(t) &= e^{\frac{-\lambda_1 t^2}{2} - \lambda_0 t},\ \lambda_1 \in \mathbb{R}, \lambda_0 > 0 \end{aligned}$$

We obtain the hazard function as follows:

$$h(t) = \frac{f(t)}{S(t)} = \frac{(\lambda_0 + \lambda_1 t)e^{-\lambda_0 t - \frac{1}{2}\lambda_1 t^2}}{e^{-\lambda_0 t - \frac{1}{2}\lambda_1 t^2}} = \lambda_0 + \lambda_1 t$$

$$h(t) = \lambda_0 + \lambda_1 t$$

And the cumulative hazard function:

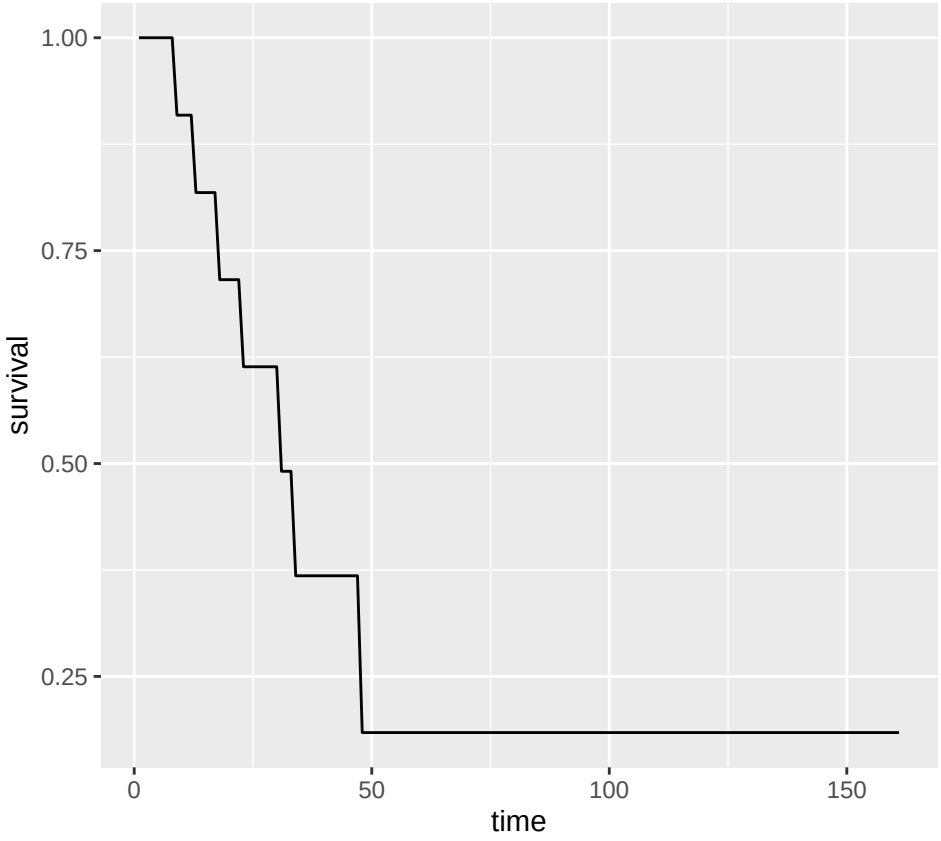
$$H(t) = -log(S(t)) = \frac{\lambda_1 t^2}{2} + \lambda_0 t$$

Exercise 3

KM estimator implementation of the survival function

Utilizing the function to obtain the survival function for the leukaemia dataset

The survival probabilities are as follows, and these change over time at the times displayed on the *time* column of this table:



Exercise 4

Exercise 5

Exercise 6