Homework 1

Xinyu Zhang

Due date: Thursday, September 20

1. **Textbook problem 1.3** The investigator of a large clinical trial would like to assess factors that might be associated with drop-out over the course of the trial. Describe what would be the event and which observations would be considered censored for such a study.

   The event is the participant drops out of the trial before it ends. The observations that the trial ends before the participatant drops out for certain reasons should be considered as censored.

2. Let T be a positive continuous random variable, show $E(T) = \int_0^\infty S(t)\, dt$.

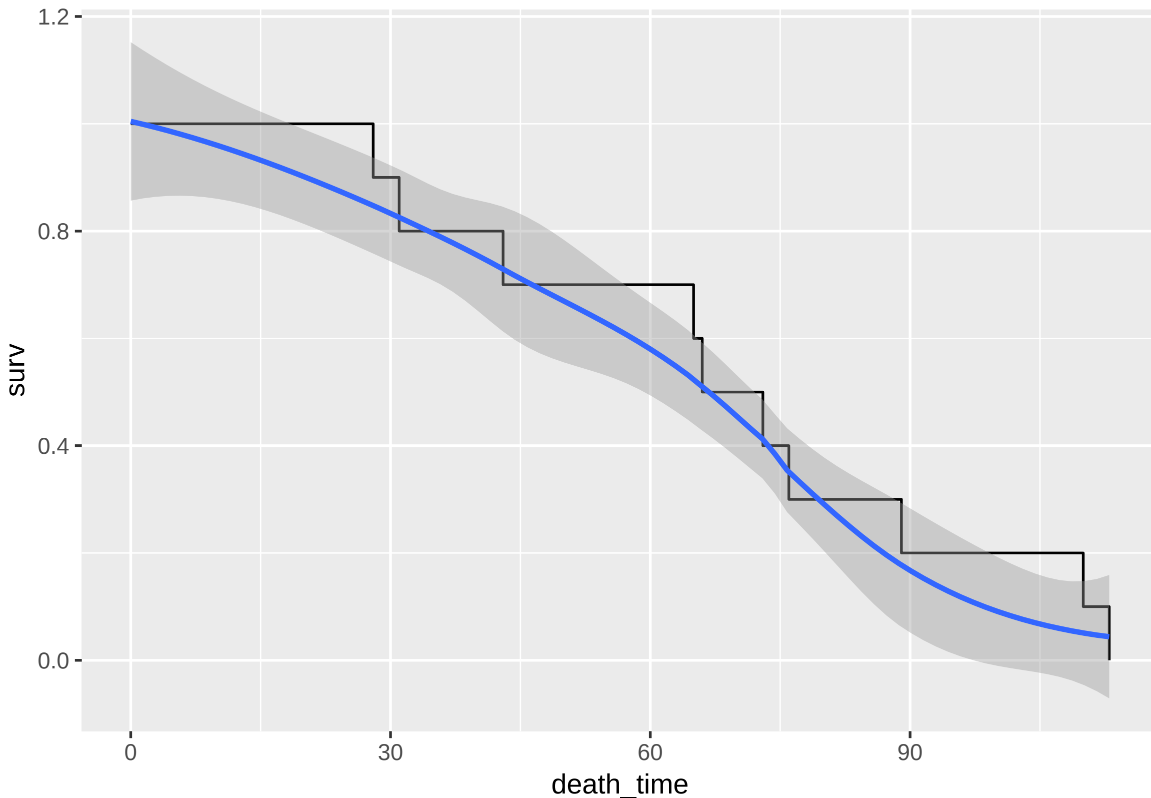
   $E(T) = \int_{-\infty}^{\infty} tf(t)dt = \int_0^\infty tf(t)dt$
   $Let\ u = t\ and\ dv = f(t)dt,\ then\ du = dt\ and\ v = F(t) = -S(t)$
   $uv\Big|_0^\infty - \int_0^\infty vdu = t(-S(t))\Big|_0^\infty - \int_0^\infty -S(t)dt$
   $= 0 + \int_0^\infty S(t)\, dt = \int_0^\infty S(t)\, dt$

3. Question 2 suggests that the area under the survival curve can be interpreted as the expected survival time. Consider the following hypothetical data set with 10 death times.

```
> dat <- c(43, 110, 113, 28, 73, 31, 89, 65, 66, 76)
```

   a. Plot the empirical survival curve.

```
> matrix(dat, nrow=length(dat), dimnames = list(NULL, c("death_time")))%>%
+ as.data.frame() %>% mutate(surv = 1 - ecdf(death_time)(death_time)) %>%
+ add_row(death_time = 0, surv = 1,.before = 1) %>%
+ ggplot(aes(death_time, surv)) + geom_step() + geom_smooth()
```



b. Find the expected survival time for the hypothetical data set.

```
> matrix(dat, nrow=length(dat), dimnames = list(NULL, c("death_time"))) %>%
+ as.data.frame() %>%
+ mutate(surv = 1 - ecdf(death_time)(death_time)) %>%
+ add_row(death_time = 0, surv = 1,.before = 1) %>%
+ arrange(desc(surv)) %>%
+ mutate(time_diff=lead(death_time,default = 0)-death_time) %>%
+ summarise(expected_lifetime=sum(time_diff*surv))
```

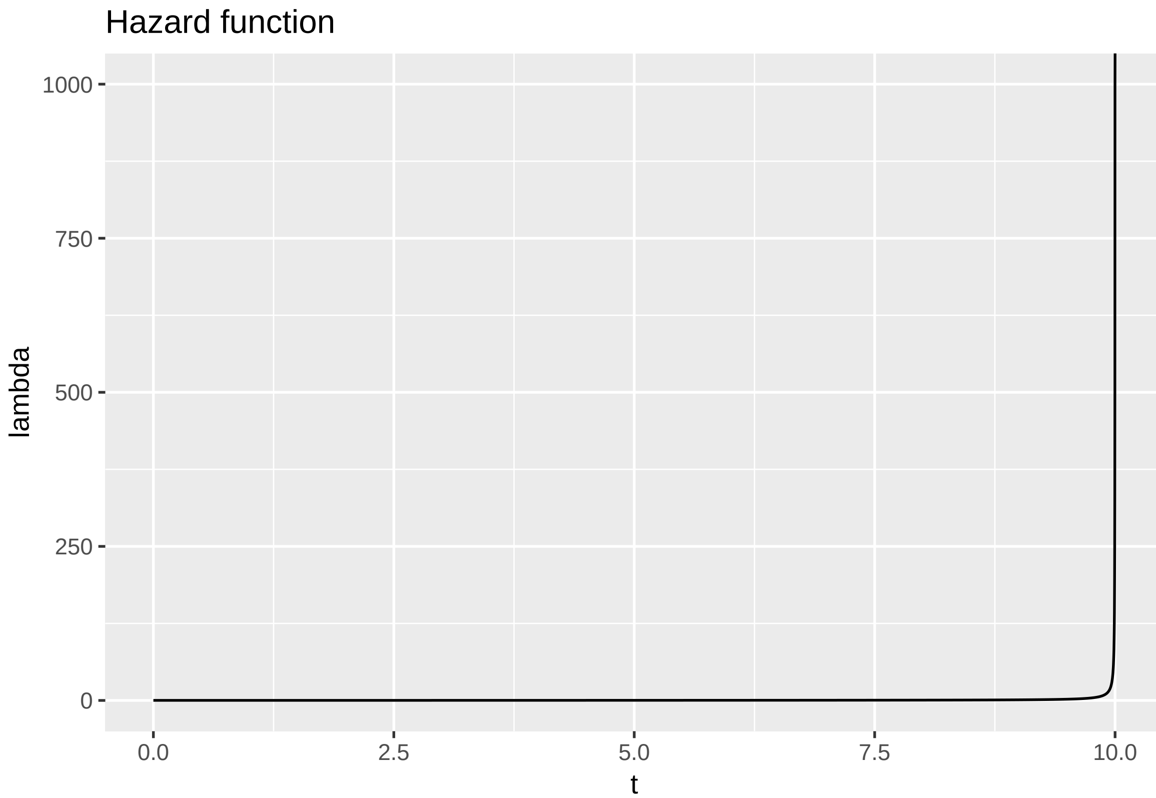
```
  expected_lifetime
1              69.4
```

The expected lifetime is 69.4.

4. Consider a survival time random variable with hazard $\lambda(t) = \frac{1}{10-t}$ in $[0, 10)$.

a. Plot the hazard function.

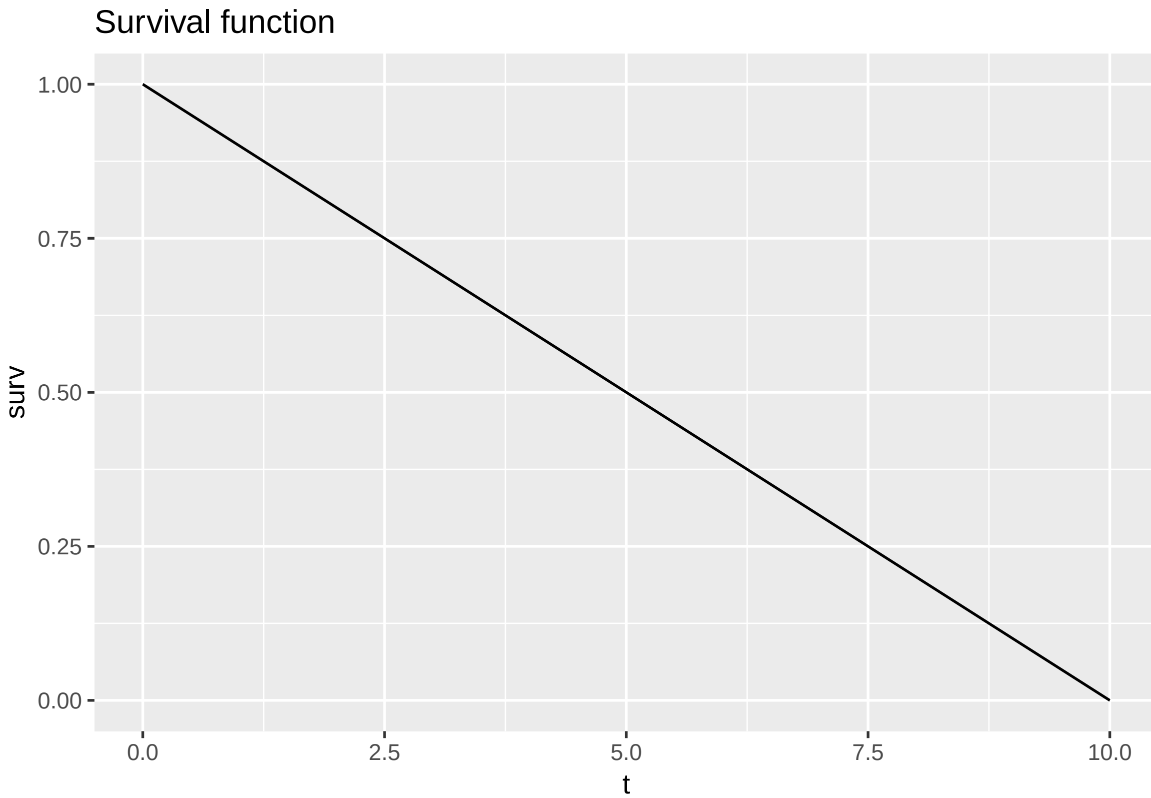
```
> matrix(seq(0, 10, length.out = 10000), dimnames = list(NULL, c("t")))%>%
+ as.data.frame() %>%
+ mutate(lambda = 1/(10-t)) %>%
+ ggplot(aes(t,lambda))+geom_line()+ggtitle("Hazard function")
```



b. Plot the survival function.

$S(t) = \exp(-\int_0^t \frac{1}{10-x}dx) = \exp(-(-\ln(10-x)\Big|_0^t)) = \exp(\ln(10-t) - \ln(10)) = \frac{10-t}{10} = 1 - \frac{t}{10}$

```
> matrix(seq(0, 10, length.out = 10000), dimnames = list(NULL, c("t"))) %>%
+ as.data.frame() %>%
+ mutate(surv = 1-t/10) %>%
+ ggplot(aes(t,surv))+geom_line()+ggtitle("Survival function")
```

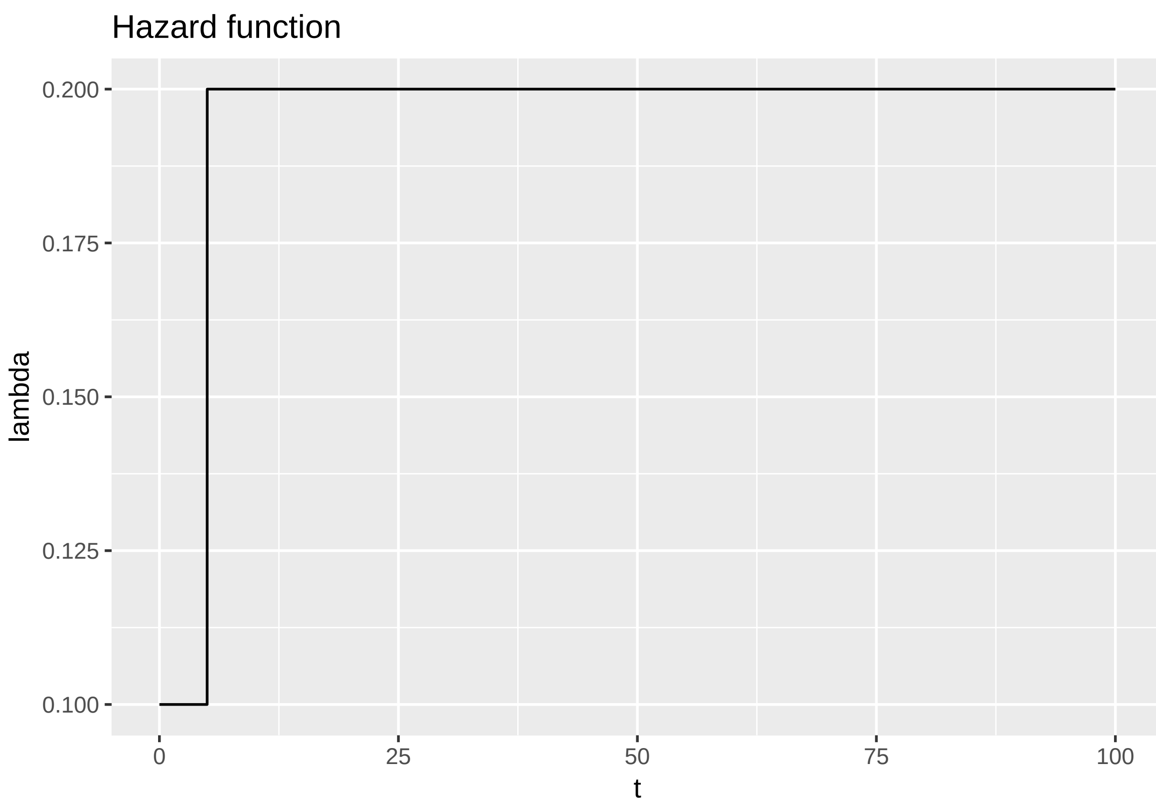


The distribution of future lifetime is uniform(0,10).

5. Consider a survival time random variable with constant hazard $\lambda = 0.1$ in $[0, 5)$, and $\lambda = 0.2$ in $[5, \infty)$. This is known as a piece-wise constant hazard.

    a. Plot the hazard function.

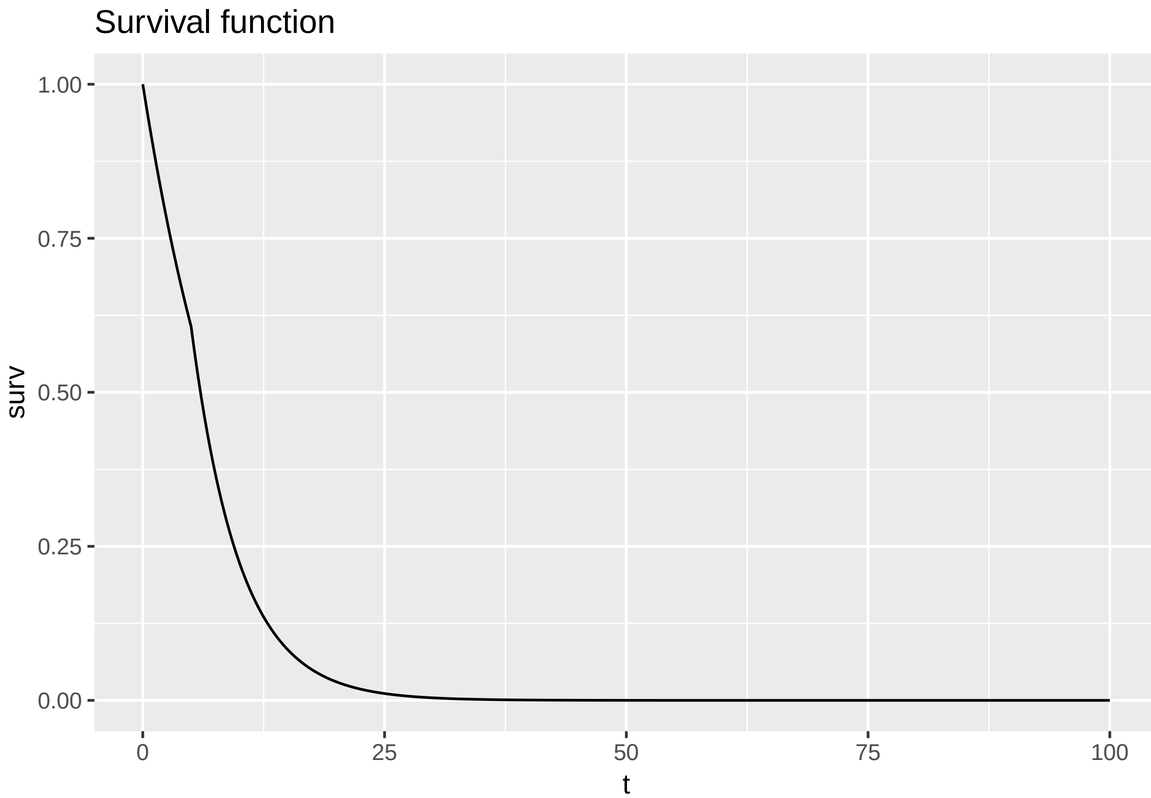
```
> matrix(seq(0, 100, length.out = 10000), dimnames = list(NULL, c("t"))) %>%
+ as.data.frame() %>%
+ mutate(lambda = 0.1*(t<5)+0.2*(t>=5)) %>%
+ ggplot(aes(t,lambda))+geom_line()+ggtitle("Hazard function")
```



b. Plot the survival function.

$$S(t) = \begin{cases} \exp(-\int_0^t 0.1dx) = \exp(-0.1t) & t \in [0,5) \\ \\ \exp(-0.1 \times 5) \cdot \exp(-\int_5^{t-5} 0.2dx) \\ = \exp(-0.5) \cdot \exp(1 - 0..2t) = \exp(0.5 - 0.2t) & t \in [5, \infty) \end{cases}$$

```
> matrix(seq(0, 100, length.out = 10000), dimnames = list(NULL, c("t"))) %>%
+ as.data.frame() %>%
+ mutate(surv = exp(-0.1*t)*(t<5)+exp(0.5-0.2*t)*(t>=5)) %>%
+ ggplot(aes(t,surv))+geom_line()+ggtitle("Survival function")
```


Survival function
surv
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
t