Homework 1 - Solution

Tian Jiang

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.

- Event: Subject drops out over the course of the trial
- Censoring: clinical trial ends, lost to follow-up, or the symptom of interest in this clinical trial occurs

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

$$
\begin{aligned}
\int_0^\infty S(t)\,dt &= \int_0^\infty P(T > t)\,dt = \int_0^\infty P(T \geq t)\,dt = \int_0^\infty E(I(T > t))\,dt \\
&= \int_0^\infty \int I(T > t)\,dP\,dt = \int_0^\infty \int_{T>t} dP\,dt \\
&= \int_0^\infty \int_0^T dt\,dP = \int_0^\infty T\,dP \\
&= E(T)
\end{aligned}
$$

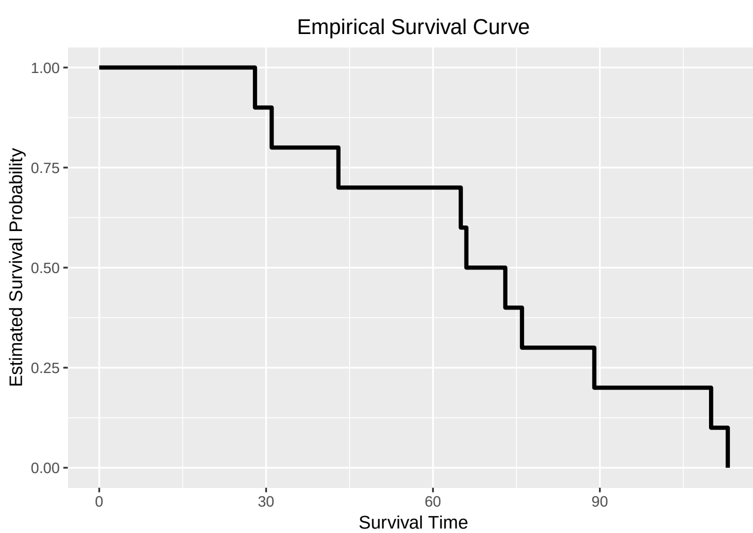
3. Question 2 suggests that the area under the survival curve can be interpreted as the expected survival time. Consider the following hypothetical data set with 10 death times.

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

a. Plot the empirical survival curve.

Insert codes here

```
> dat %>% mutate(surv = 1 - ecdf(T)(T)) %>% add_row(T = 0, surv = 1,.before = 1) %>%
+   ggplot(aes(T, surv)) + geom_step(size = 1.2) + xlab("Survival Time") +
+   ylab("Estimated Survival Probability") + ggtitle("Empirical Survival Curve") +
+   theme(plot.title = element_text(hjust = 0.5))
```



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

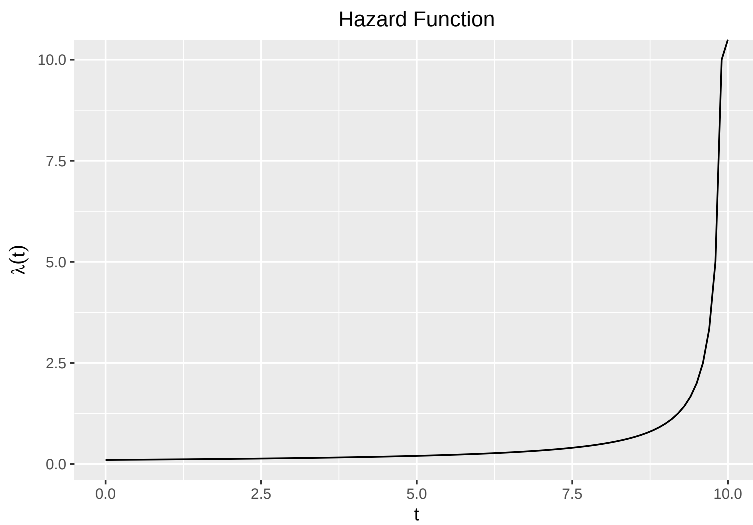
```
> dat %>% mutate(surv = 1 - ecdf(T)(T)) %>% add_row(T = 0, surv = 1,.before = 1) %>%
+   arrange(T) %>% mutate(diff = lead(T) -T) %>% mutate(prod = cumsum(surv*diff)) %>%
+   summarise('Expected Survival Time' = max(prod, na.rm = TRUE))
```

```
# A tibble: 1 x 1
  `Expected Survival Time`
                     <dbl>
1                     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.

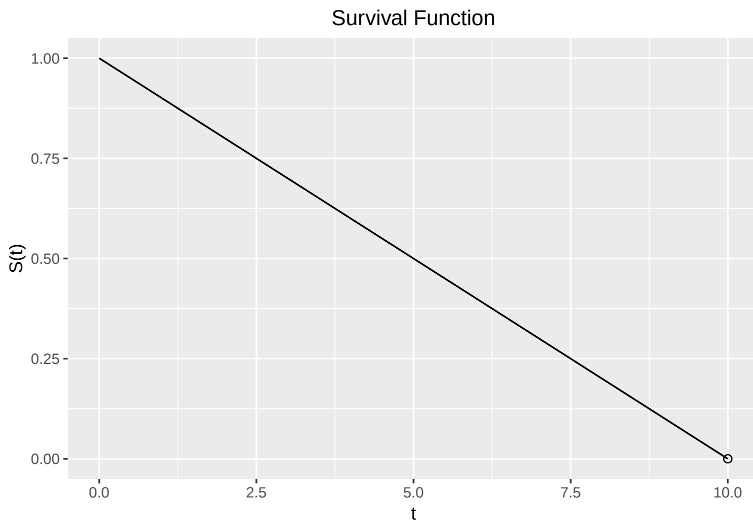
```
> ggplot(data.frame(x = c(0, 10)), aes(x)) + stat_function(fun=function(x) 1/(10-x)) +
+   xlab("t") + ylab(expression(lambda(t))) + ggtitle("Hazard Function") +
+   theme(plot.title = element_text(hjust = 0.5))
```



b. Plot the survival function.

$$S(t) = e^{-\int_0^t \lambda(u)\,du} = e^{-\int_0^t \frac{1}{10-u}\,du} = e^{\ln(10-t)-\ln(10)} = \frac{10-t}{10},\ t \in [0, 10)$$

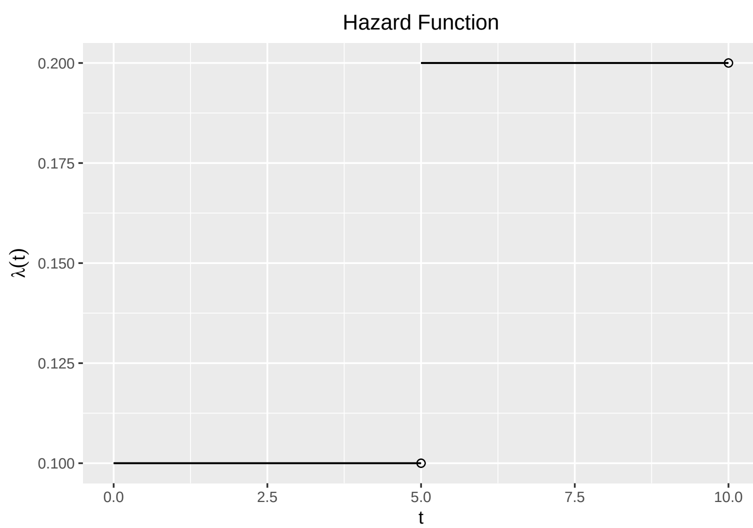
```
> ggplot(data.frame(x = c(0, 10)), aes(x)) + stat_function(fun=function(x) (10-x)/10) +
+   geom_point(aes(x = 10, y = 0), shape = 1, size = 2) + xlab("t") + ylab("S(t)") +
+   ggtitle("Survival Function") + theme(plot.title = element_text(hjust = 0.5))
```



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.

```
> ggplot(data.frame(x = c(0, 10)), aes(x)) +
+   geom_segment(aes(x = 0, xend = 5, y = 0.1, yend = 0.1)) +
+   geom_point(aes(x = 5, y = 0.1), shape = 1, size = 2) +
+   geom_segment(aes(x = 5, xend = 10, y = 0.2, yend = 0.2)) +
+   geom_point(aes(x = 10, y = 0.2), shape = 1, size = 2) + xlab("t") +
+   ylab(expression(lambda(t))) + ggtitle("Hazard Function") +
+   theme(plot.title = element_text(hjust = 0.5))
```



   b. Plot the survival function.

$$S(t) = e^{-\int_0^t \lambda(u)\,du} = \begin{cases} e^{-\int_0^t 0.1\,du} = e^{-0.1t}, & 0 \le t < 5 \\ e^{-0.5-\int_0^{t-5} 0.2\,du} = e^{0.5-0.2t}, & 5 \le t < 10 \end{cases}$$

```r
> ggplot(data.frame(x = c(0, 10)), aes(x)) +
+   stat_function(fun = function(x) (x<5)*exp(-0.1*x)+(x>=5)*exp(0.5-0.2*x)) +
+   geom_point(aes(x = 5, y = exp(-0.5)), size = 2) +
+   geom_point(aes(x = 10, y = exp(-1.5)), shape = 1, size = 2) +
+   xlab("t") + ylab("S(t)") + ggtitle("Survival Function") +
+   theme(plot.title = element_text(hjust = 0.5))
```