Homework 1 - Solution

Steven Chiou

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 would be the drop-out due to the factor of interested and the censored event could be drop-out due to reasons other than the factor of interested.

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

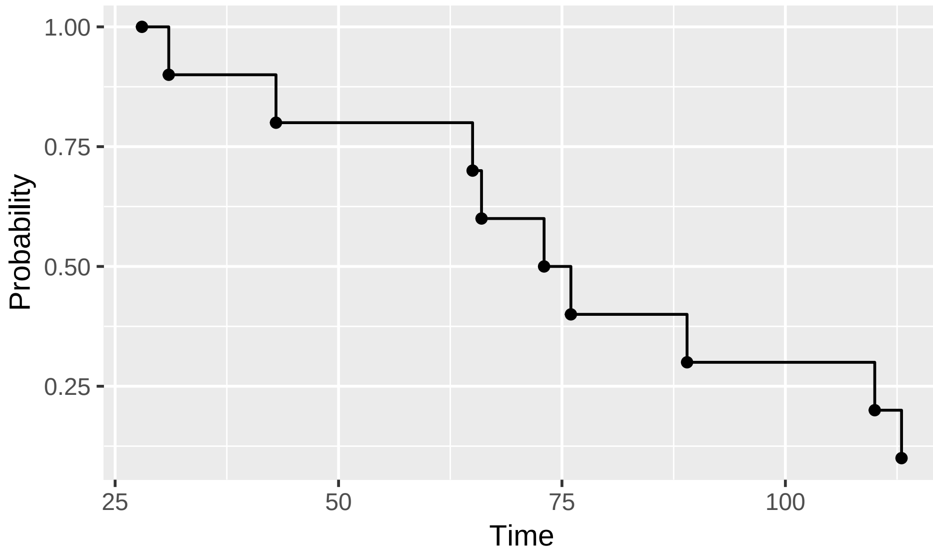
$$\int_0^\infty S(t)\,dt = \int_0^\infty \int_t^\infty f(x)\,dx\,dt = \int_0^\infty \int_0^x f(x)\,dt\,dx = \int_0^\infty xf(x)\,dx = E(T).$$

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.**

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

a. **Plot the empirical survival curve.**

```
> qplot(dat, rank(-dat) / 10) + geom_step() + ylab("Probability") + xlab("Time")
```



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

The expected survival time is `mean(dat)` = 69.4 since there is no censoring. Many approaches are available to directly compute the area under the empirical survival curve. Here is one

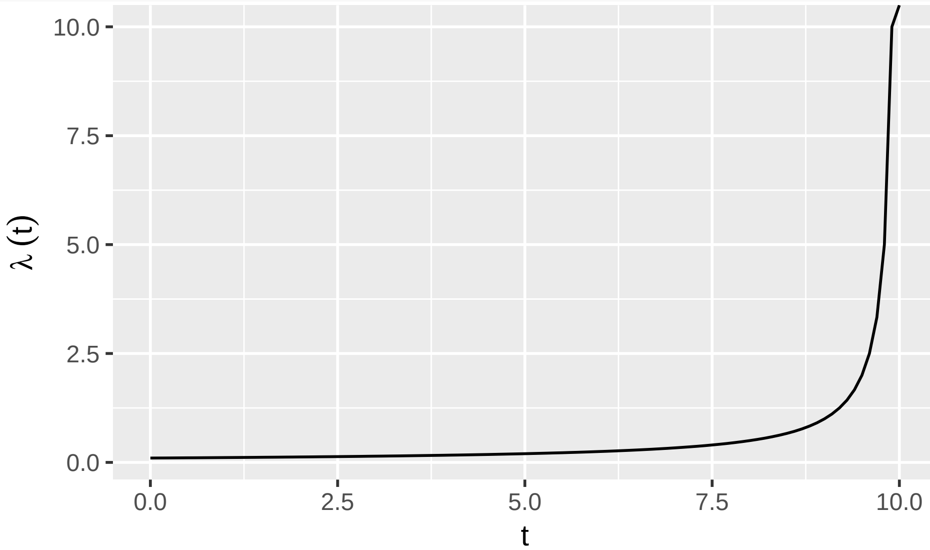
```
> max(dat) - integrate(ecdf(dat), 0, max(dat))$value
```

```
[1] 69.4004
```

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

   a. **Plot the hazard function.**

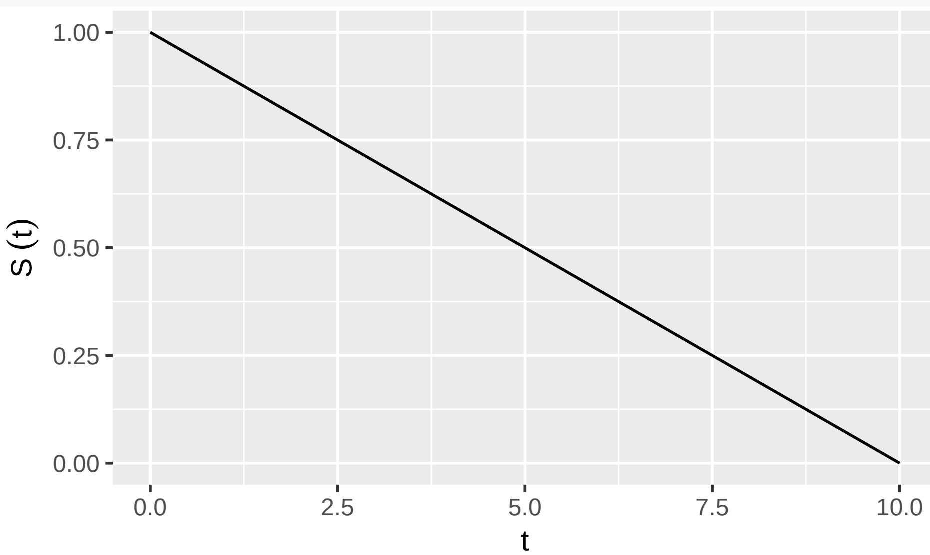
```
> ggplot(tibble(x = c(0, 10)), aes(x)) +
+   stat_function(fun = function(x) 1 / (10 - x)) +
+   xlab("t") + ylab(expression(lambda~(t)))
```



   b. **Plot the survival function.**

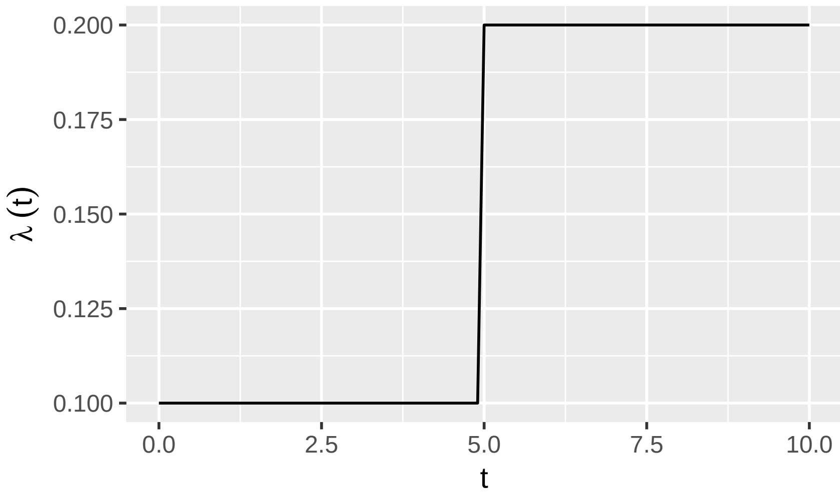
The cumulative hazard function is $\Lambda(t) = \int_0^t \frac{1}{10-x}\,dx = \log\left(\frac{10}{10-t}\right)$. This further implies $S(t) = \frac{10-t}{10}$.

```
> ggplot(tibble(x = c(0, 10)), aes(x)) +
+   stat_function(fun = function(x) .1 * (10 - x)) +
+   xlab("t") + ylab(expression(S~(t)))
```



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(tibble(x = c(0, 10)), aes(x)) +
+   stat_function(fun = function(x) 0.1 * (x < 5) + .2 * (x >= 5)) +
+   xlab("t") + ylab(expression(lambda~(t)))
```



b. **Plot the survival function.**

The cumulative hazard function is

$$\Lambda(t) = \begin{cases} 0.1t & \text{if } t < 5 \\ 0.2t - 0.5 & \text{if } t \geq 5 \end{cases}.$$

This then implies

$$S(t) = \begin{cases} e^{-0.1t} & \text{if } t < 5 \\ e^{-0.2t+0.5} & \text{if } t \geq 5 \end{cases}.$$

```
> ggplot(tibble(x = c(0, 10)), aes(x)) +
+   stat_function(fun = function(x) exp(ifelse(x < 5, -0.1 * x, -.2 * x + .5))) +
+   xlab("t") + ylab(expression(S~(t)))
```