STAT6390 HW1

Cong Zhang

September 11, 2018

Question 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.

Answer The event is drop-off and an obs is considered as censored (right-censor) if the participant does not drop off before he/she leaves the study or the study is completed.

Question 2

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

Proof

$$E(T) = \int_0^\infty tf(t)dt = \int_0^\infty (\int_0^t 1ds)f(t)dt = \int_0^\infty (\int_s^\infty f(t)dt)ds = \int_0^\infty S(s)ds = \int_0^\infty S(t)dt$$

Question 3

Question 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.

a. Plot the empirical survival curve.

```
library(survMisc)
library(tidyverse)
dat3 <- data.frame(len = c(43, 110, 113, 28, 73, 31, 89, 65, 66, 76))
dat3 <- dat3 %>% mutate(surv = 1-ecdf(len)(len))
dat3 %>% mutate(surv = 1 - ecdf(len)(len)) %>%
    add_row(len = 0, surv = 1,.before = 1) %>%
    ggplot(aes(len, surv)) + geom_step() + theme_bw() +
    scale_x_continuous(name = "time", breaks = seq(0, 120, 20), limits = c(0, 120)) +
    scale_y_continuous(name = "empirical survival function")
```

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

```
a <- survfit(Surv(dat3$len, rep(1, 10)) ~1)
b <- print(a, print.rmean=TRUE)
```

```
## Call: survfit(formula = Surv(dat3$len, rep(1, 10)) ~ 1)
##
##          n     events     *rmean *se(rmean)     median    0.95LCL
##      10.00      10.00      69.40       8.89      69.50      43.00
##    0.95UCL
##         NA
##     * restricted mean with upper limit =  113
```

Graph of survival function was shown in Figure 1 and according to Question #2, the AUC of S(t) curve is the expected survival time, which is 69.40 here.

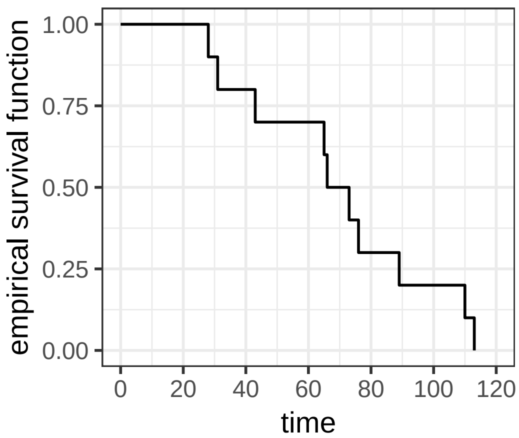

Figure 1: Empirical Survival Curve

Question 4

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

Hazard and survival curves were shown in Figure 2.

```
hzd <- function(x) {1/(10-x)}
h <- hzd
## H(t) cumulative hazard function: h(t) integrated from time = 0 to time = t
## Vectorize to enable use with a vector
H <- Vectorize(function(t) {
    res <- integrate(h, lower = 0, upper = t)
    res$value
})
## S(t) survivor function: Derived from H(t) = -logS(t)
S <- function(t) {
    exp(-1 * H(t))
}

xlim = c(0,9.999999)
ylim = c(0,10)

ggplot(data = data.frame(x = xlim), aes(x)) +
    stat_function(fun = h, aes(color = "h"), size = 1) +
    stat_function(fun = S, aes(color = "S"), size = 1, linetype = "dashed") +
    scale_x_continuous(name = "time",  limit = xlim) +
    scale_y_continuous(name = "value", limit = ylim,
                       breaks =c(seq(ylim[1], ylim[2], length.out = 5), 1)) +
    scale_color_manual(name = "functions",
                       values = c("h" = "skyblue", "S" = "mistyrose"),
                       breaks = c("h","S"),
                       labels = c("hazard h(t)","survival S(t)")) +
    theme_bw()
```

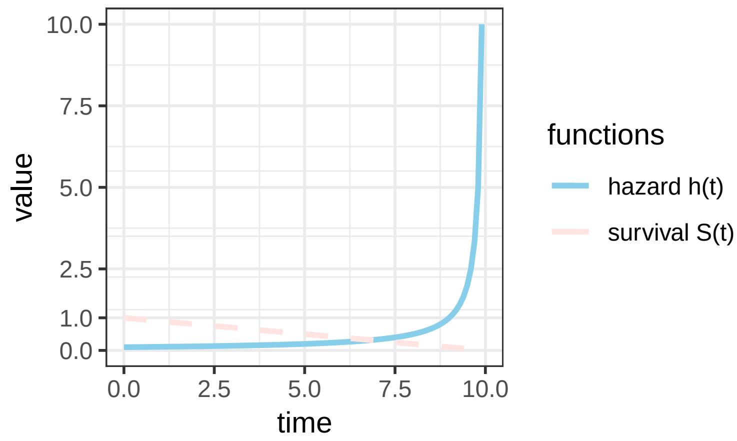

Figure 2: Harzard function and survival Curve (Uniform)

Question 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.

Hazard and survival curves were shown in Figure 3.

```
#define harzard fn
h.constant <- function(t) {0.1 + 0.1 * (t >= 5)}
#get survival fun
h <- h.constant
H <- Vectorize(function(t) {
    res <- integrate(h, lower = 0, upper = t)
    res$value
})

S <- function(t) {
    exp(-1 * H(t))
}
#plot
xlim = c(0,35)
ylim = c(0,1)
x <- seq(xlim[1], xlim[2], length.out = 99)
st <- data.frame(x,  y = h.constant(x))

ggplot(data = data.frame(x = xlim), aes(x)) +
    stat_function(fun = S, aes(color = "S"), size = 1, linetype = "dashed") +
    geom_step(aes(x, y, color = "h"), data = st, size = 1) +
    scale_x_continuous(name = "time",  limit = xlim, breaks = seq(0, 35, 5)) +
    scale_y_continuous(name = "value", limit = ylim, breaks = seq(0, 1, 0.1)) +
    scale_color_manual(name = "functions",
                       values = c("h" = "skyblue", "S" = "mistyrose"),
                       breaks = c("h","S"),
                       labels = c("hazard h(t)","survival S(t)")) +
    theme_bw()
```

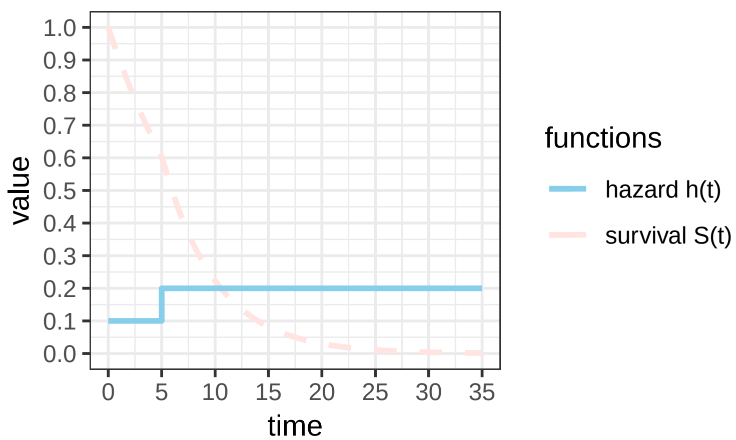

Figure 3: Harzard function and survival Curve (Constant)