SOLUTION TO EXERCISES WEEK 39

Exercise 3.5

$X_1, X_2, \dots, X_n$ are iid nonnegative random variables with survival function $S(t) = P(X_i > t)$ and cumulative distribution function $F(t) = P(X_i \le t) = 1 - S(t)$.

a) When there are no censored observations, the number at risk $Y(t)$ takes the form

$$Y(t) = \sum_{i=1}^{n} I(X_i > t) = n - \sum_{i=1}^{n} I(X_i \le t)$$

In particular, if we let $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ be the ordered X_i's, we have

$$Y(X_{(1)}) = n$$

$$Y(X_{(2)}) = n-1$$

$$Y(X_{(3)}) = n-2,$$

etc

i.e $\quad Y(X_{(j)}) = n - j + 1$

For $X_{(j)} \le t < X_{(j+1)}$ we therefore have

$$\hat{S}(t) = \prod_{X_{(i)} \le t} \left(1 - \frac{1}{Y(X_{(i)})}\right)$$

$$= \prod_{X_{(i)} \le t} \frac{Y(X_{(i)}) - 1}{Y(X_{(i)})}$$

$$= \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdot \frac{n-3}{n-2} \cdots \frac{n-j+1}{n-j+2} \cdot \frac{n-j}{n-j+1}$$

$$= \frac{n-j}{n} = 1 - \frac{j}{n}$$

$$= 1 - \hat{F}(t)$$

Since $\hat{F}(t) = n^{-1} \sum_{i=1}^{n} I(X_i \le t) = j/n$

for $X_{(j)} \le t < X_{(j+1)}$.

b) Greenwood's formula takes the form

$$\tilde{\tau}^2(t) = \hat{S}(t)^2 \sum_{X_{(i)} \leq t} \frac{1}{Y(X_{(i)})(Y(X_{(i)})-1)}$$

In order to show that $\tilde{\tau}^2(t) = \dfrac{\hat{S}(t)(1-\hat{S}(t))}{n}$, it is sufficiently to show that

$$\frac{\tilde{\tau}^2(t)}{\hat{S}(t)^2} = \sum_{X_{(i)} \leq t} \frac{1}{Y(X_{(i)})(Y(X_{(i)})-1)} \qquad (*)$$

is equal to $(1-\hat{S}(t))/(n\hat{S}(t))$.

We will show this by noting that both expressions equal zero when $t=0$, by noting that both expressions are constant between the $X_{(j)}$'s, and by showing that the increments at the $X_{(j)}$'s are the same.

The increment of (*) at $X_{(j)}$ equals

$$\frac{1}{Y(X_{(j)})(Y(X_{(j)})-1)} = \frac{1}{(n-j+1)(n-j)},$$

while the increment of $(1-\hat{S}(t))/(n\hat{S}(t))$ at $t = X_{(j)}$ equals

$$\frac{1-\hat{S}(X_{(j)})}{n\hat{S}(X_{(j)})} - \frac{1-\hat{S}(X_{(j-1)})}{n\hat{S}(X_{(j-1)})}$$

$$= \frac{1-(1-\frac{j}{n})}{n(1-\frac{j}{n})} - \frac{1-(1-\frac{j-1}{n})}{n(1-\frac{j-1}{n})}$$

$$= \frac{j}{n(n-j)} - \frac{j-1}{n(n-j+1)}$$

$$= \frac{j(n-j+1)-(j-1)(n-j)}{n(n-j+1)(n-j)} = \frac{1}{(n-j+1)(n-j)}$$

This shows the result.

Exercise 3.6

We have that $\hat{S}(t)$ is approximately normally distributed with mean $S(t)$ and a variance that may be estimated by $\hat{\tau}^2(t)$.

By an argument similar to the one in exercise 3.3, we have that

$$g(\hat{S}(t)) \pm z_{1-\alpha/2}\, g'(\hat{S}(t))\, \hat{\tau}^2(t)$$

is an approximate $100(1-\alpha)\%$ confidence interval for $g(S(t))$. Here $g(x)$ is a strictly increasing continuously differentiable function.

For $g(x) = -\log(-\log x)$, we have $g'(x) = -1/(x \log x)$. [Note that $g'(x) > 0$ for $x \in (0,1)$.]

Thus a $100(1-\alpha)\%$ confidence interval for $-\log(-\log S(t))$ is ("$\pm$" means + for the upper limit and minus for the lower limit)

$$-\log(-\log \hat{S}(t)) \pm z_{1-\alpha/2} \left(\frac{-1}{\hat{S}(t) \log \hat{S}(t)}\right) \hat{\tau}(t)$$

It follows that a confidence interval for $\exp(-\log(-\log S(t))) = -1/\log S(t)$ is given by

$$\frac{-1}{\log \hat{S}(t)} \exp\left(\pm z_{1-\alpha/2} \frac{-\hat{\tau}(t)}{\hat{S}(t) \log \hat{S}(t)}\right)$$

Hence an interval for $\log S(t)$ is (note that $\log S(t) < 0$):

$$\log \hat{S}(t) \cdot \exp\left(\pm z_{1-\alpha/2} \frac{\hat{\tau}(t)}{\hat{S}(t) \log \hat{S}(t)}\right)$$

Finally this yields the following intervall for $S(t)$

$$\hat{S}(t)^{\exp\left(\pm z_{1-\alpha/2}\, \hat{\tau}(t)/(\hat{S}(t)\log \hat{S}(t))\right)}$$

cf. (3.30) in ABG.

Exercise 3.8

We have that $(\hat{S}(t) - S(t))/\hat{\tau}(t)$ is approximately standard normally distributed

a) Let ξ_p be the p-th fractile of the survival distribution. Then $S(\xi_p) = 1-p$. In order to test the hypothesis

$$H_0: \xi_p = \xi_p^0 \quad \text{vs} \quad H_A: \xi_p \neq \xi_p^0$$

We may consider the test statistic

$$Z = \frac{\hat{S}(\xi_p^0) - S(\xi_p^0)}{\hat{\tau}(\xi_p^0)} = \frac{\hat{S}(\xi_p^0) - (1-p)}{\hat{\tau}(\xi_p^0)}$$

which is approximately $N(0,1)$ distributed under H_0. We may therefore reject H_0 if $|Z| > z_{1-\alpha/2}$.

b) According to a general result, we get a $100(1-\alpha)\%$ confidence for ξ_p as all ξ_p^0-values that are not rejected by the test in a.

This confidence interval is given by

$$\left\{ \xi_p^0 : \left| \hat{S}(\xi_p^0) - (1-p) \right| / \hat{\tau}(\xi_p^0) \le z_{1-\alpha/2} \right\}$$

$$= \left\{ \xi_p^0 : \left| \hat{S}(\xi_p^0) - (1-p) \right| \le z_{1-\alpha/2}\, \hat{\tau}(\xi_p^0) \right\}$$

The confidence interval may be read of a plot of the Kaplan-Meier estimator with confidence limits as illustrated in Fig. 3.13 in ABG.

An explanation is given in the figure below: