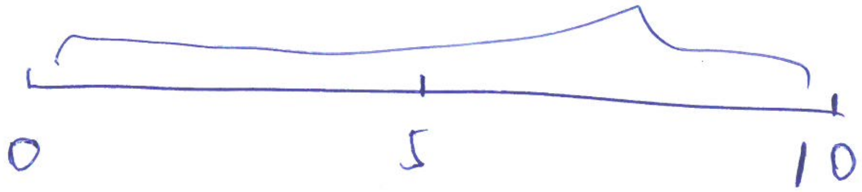
Lecture week 4

Censoring : 10 deaths



observe 100 ppl over 10 years.

10 deaths 10 left at time 5.

At time 10, you know the number of deaths is between 10 ~ 20 out of 100.

How to estimate survival function at time 10.

Remove the censored = loss of information

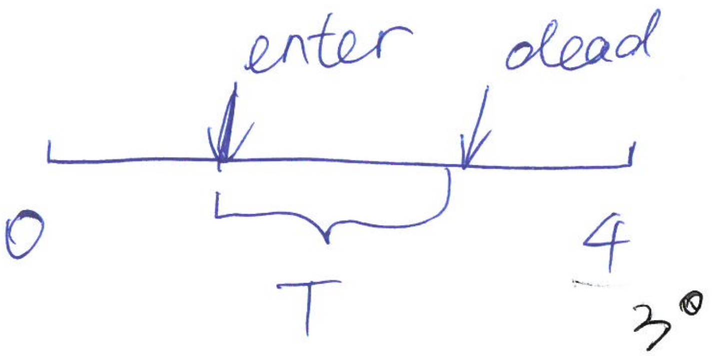
10 censored lives to 5 at least

Ignore the censoring : Assuming censored are all dead, underestimate survival function

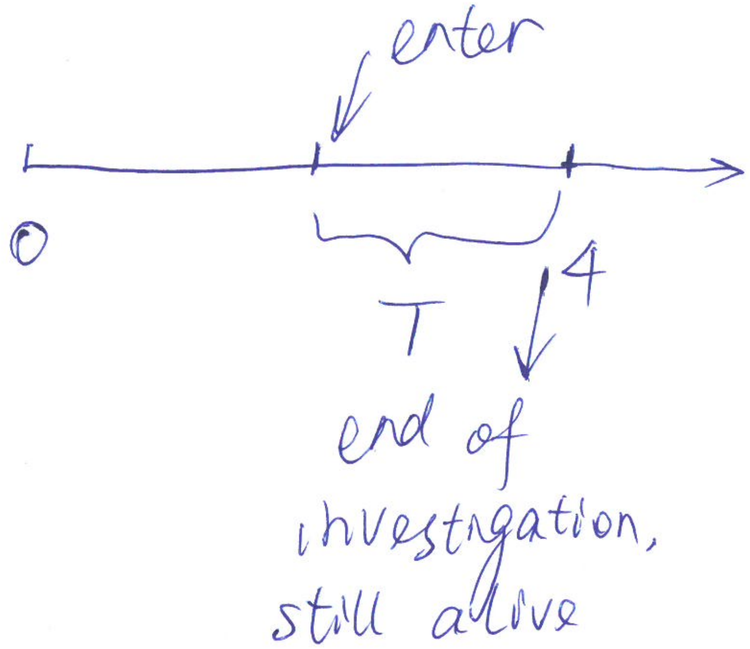
. . . . . .

R example

T: future lifetime observed

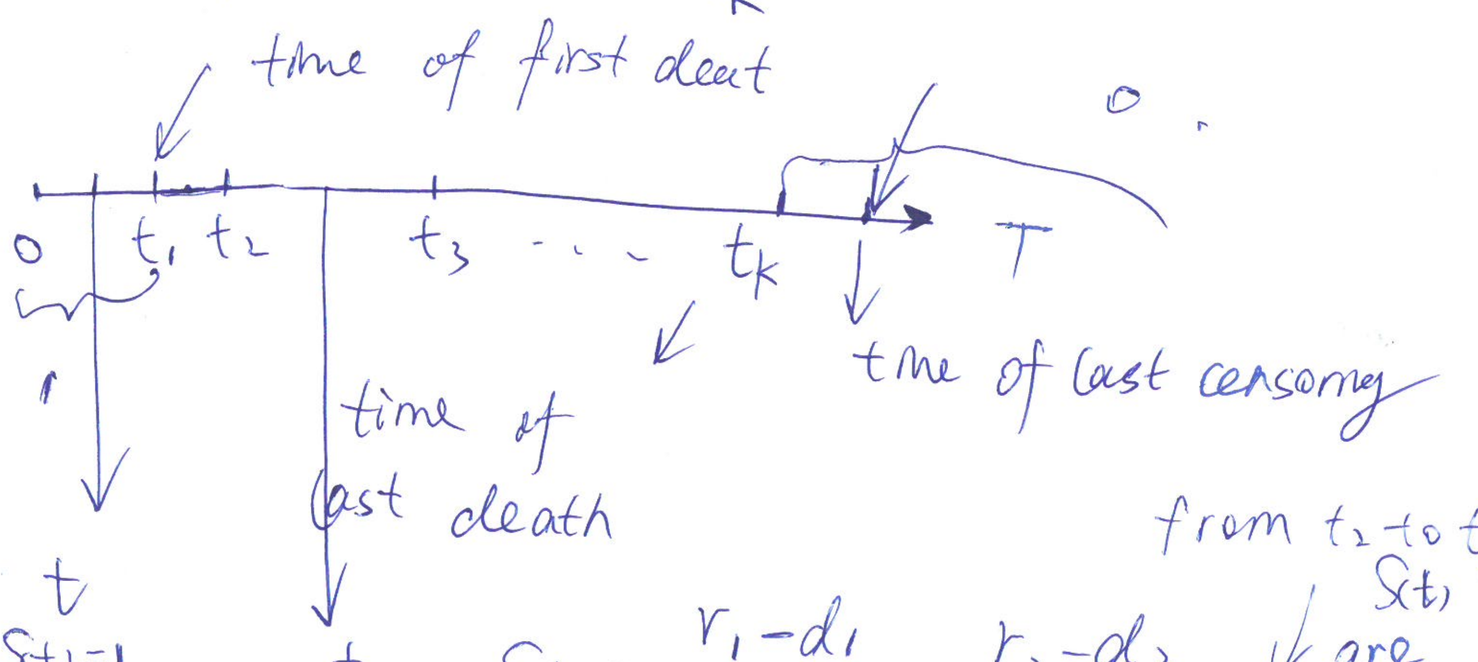


① uncensored



② censored.

$0 < t_1 < t_2 < \cdots < t_k$



$S(t)=1$

$S(t) = \frac{r_1-d_1}{r_1} \cdot \frac{r_2-d_2}{r_2}$

from t_2 to t_3 $S(t)$ are the same

Greenwood's formula:

$$\log[\hat{S}(t)] = \sum_{t_j \le t} \log \frac{r_j - d_j}{r_j} = \sum_{t_j \le t} \log(\hat{p}_j)$$

$$\hat{p}_j = \frac{r_j - d_j}{r_j} \qquad d_j \sim Bin(r_j, 1-p_j)$$

$$E(\hat{p}_j) = 1 - E\left(\frac{d_j}{r_j}\right) = p_j$$

$$Var(\hat{p}_j) = Var\left(\frac{d_j}{r_j}\right) = \frac{1}{r_j^2} \cdot r_j(1-p_j)p_j$$

$$\approx \frac{1}{r_j} \hat{p}_j (1-\hat{p}_j)$$

$$Var[\log(\hat{p}_j)] = \frac{1}{\hat{p}_j^2} \frac{1}{r_j} \hat{p}_j (1-\hat{p}_j)$$

$$= \frac{d_j}{r_j(r_j - d_j)} \quad \text{(delta method)}$$

$$Var[\log \hat{S}(t)] = \sum_{t_j \le t} Var[\log(\hat{p}_j)]$$

$$= \sum_{t_j \le t} \frac{d_j}{r_j(r_j - d_j)}$$

Using these results, we can get

$$Var(\hat{S}(t)) = Var\left(e^{\log(\hat{S}(t))}\right)$$

$$= (\hat{S}(t))^2 \sum_{t_j \le t} \frac{d_j}{r_j(r_j - d_j)}$$

Why we are interested in $\log \hat{S}(t)$

Constructing C.I for $S(t)$

① $\hat{S}(t) \pm 1.96\, s.d(\hat{S}(t))$?

based on normal $\hat{S}(t) \sim \text{Normal}$?

not proper because

$S(t) > 0$.

②

$\log \hat{S}(t) \pm 1.96\, s.d(\log \hat{S}(t))$

Then transform it back to $\hat{S}(t)$

$= \sum_{t_j \le t} \log(\hat{p}_j)$

Now Assume $\log \hat{S}(t) \sim \text{Normal}$. By CLT

- - - - - - - - - - -