# Lec 5: Survival Analysis & Cox Model

Yanjun Han

Oct. 3, 2023

# Life table.

Example dataset of an insurance company, where:

| Age | n | d |
|---|---|---|
| 30 | $n_{30}$ | $d_{30}$ |
| 31 | $n_{31}$ | $d_{31}$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| 89 | $n_{89}$ | $d_{89}$ |

- $n_i$ : # of policy holders at age $i$
- $d_i$ : # of deaths

<div style="border: 1px solid red; padding: 4px;">

Target: estimate the "survival function" $S_i = \mathbb{P}(X \geq i)$

</div>

Notations:
- $X$ : a typical lifetime of a population (random variable)
- $S_i = \mathbb{P}(X \geq i)$ : survival function
- $h_i$ : hazard rate, defined as

$$h_i = \frac{\mathbb{P}(X = i)}{\mathbb{P}(X \geq i)} = \frac{S_i - S_{i+1}}{S_i} .$$

## MLE.

Model: for $i \in \{30, \cdots, 89\}$, a sample of size $n_i$ is drawn from a conditional population with $X \geq i$, where $d_i$ of them die within one year.

Log-likelihood: $\ell(h_{30}, \cdots, h_{89}) = \sum_{i=30}^{89} \left( d_i \log h_i + (n_i - d_i) \log(1 - h_i) \right)$

MLE for $h_i$: $\frac{\partial \ell}{\partial h_i} = 0 \implies \frac{d_i}{\hat{h}_i} - \frac{n_i - d_i}{1 - \hat{h}_i} = 0 \implies \hat{h}_i = \frac{d_i}{n_i} .$

MLE for $S_i$ : $\qquad \hat{S}_i = \prod\limits_{j=30}^{i-1} (1 - \hat{h}_j)$.

Similarly, for $i < j$, one can estimate
$$\hat{\mathbb{P}}(X \geq j \mid X \geq i) = \prod\limits_{k=i}^{j-1} (1 - \hat{h}_k).$$

> One year's data suffices to learn the survival functions

## Censored data.

An example survival data after          Transform into a lifetable:
a clinical trial :

$\Longrightarrow$

| days | n | d | $\ell$ |
|------|------|------|------|
| $t_1$ | $n_1$ | $d_1$ | $\ell_1$ |
| $t_2$ | $n_2$ | $d_2$ | $\ell_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $t_m$ | $n_m$ | $d_m$ | $\ell_m$ |

$\{$ 64, 73+, 160, 160 , 185+,

110l, 1412+ , $\cdots$ $\}$

(a+ . still alive after a days)

Notations:
- $d_i$ : # of observed deaths at day $t_i$ after the trial
- $\ell_i$ : # of lost followups at day $t_i$ after the trial
- $n_i$ : # of individuals known to have survived at the beginning of day $t_i$
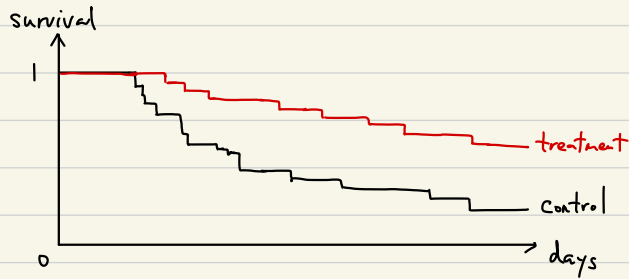$$n_i = \sum\limits_{j \geq i} (d_j + \ell_j)$$
( for convenience, we assume that deaths & lost followups cannot happen
simultaneously ; i.e. for every i, either $d_i = 0$, or $\ell_i = 0$          )

Target: estimate the survival function $S(t) = \mathbb{P}(X \geq t)$

<u>Kaplan - Meier estimator</u>.

$$\hat{h}_i = \hat{P}(X = t_i \mid X \geq t_i) = \frac{d_i}{n_i}$$
$$\hat{S}(t) = \prod_{i: t_i < t} (1 - \hat{h}_i) = \prod_{i: t_i < t} (1 - \frac{d_i}{n_i})$$

A Kaplan - Meier survival curve:



Derivation: empirical likelihood.

Let $f$ be the pdf/pmf of $X$, and $S(t) = P(X \geq t)$

$$\text{log-likelihood}(S) = \sum_{i=1}^{m} \left( d_i \log(f(t_i)) + \ell_i \log(S(t_i)) \right)$$

likelihood of dying at $t_i$     likelihood of surviving until $t_i$

( An implicit assumption: probability of getting censored does not depend on the survival function )

Maximum value of $f(t_i)$ given $S$:    $f(t_i) \leq S(t_i) - S(t_{i+1})$

Empirical likelihood: assume that $S$ has jumps only at $\{t_1, \cdots, t_n\}$

empirical log-likelihood $(S) = \sum_{i=1}^{m} \left( d_i \log (S_i - S_{i+1}) + \ell_i \log S_i \right)$

<span style="color:blue">( Write $S_i := S(t_i)$ )</span>

Maximizing over $1 = S_1 \geq S_2 \geq \cdots \geq S_{m+1} \geq 0$

$\Longleftrightarrow$ Maximizing over $h_1, \cdots, h_m \in [0,1)$, where

$$h_i = 1 - \frac{S_{i+1}}{S_i} \iff S_i = \prod_{j=1}^{i-1} (1 - h_j).$$

Now empirical log-likelihood $(h)$

$$= \sum_{i=1}^{m} \left( d_i \log \left( h_i \prod_{j=1}^{i-1} (1-h_j) \right) + \ell_i \log \prod_{j=1}^{i-1} (1-h_j) \right)$$

$$= \sum_{i=1}^{m} \left( d_i \log h_i + \sum_{j=1}^{i-1} (d_i + \ell_i) \log (1 - h_j) \right)$$

$$= \sum_{i=1}^{m} \left( d_i \log h_i + \sum_{k > i} (d_k + \ell_k) \log (1 - h_i) \right)$$

<span style="color:blue">( swapping the sum $\sum_i \sum_{j < i} = \sum_j \sum_{i > j}$ )</span>

$$= \sum_{i=1}^{m} \left( d_i \log h_i + (n_i - d_i - \ell_i) \log (1 - h_i) \right)$$

<span style="color:blue">( recall that $n_i = \sum_{k \geq i} (d_k + \ell_k)$ )</span>

F.O.C. for $h_i \Rightarrow \hat{h}_i = \dfrac{d_i}{n_i - \ell_i} = \dfrac{d_i}{n_i}$

<span style="color:blue">( because either $d_i = 0$, or $\ell_i = 0$ )</span>

$$\Rightarrow \hat{S}_i = \prod_{j < i} (1 - \hat{h}_j) = \prod_{j < i} \left( 1 - \frac{d_j}{n_j} \right)$$

$$\Rightarrow \hat{S}(t) = \max_{i:\, t_i < t} \hat{S}_{i+1} = \prod_{i:\, t_i < t} \left( 1 - \frac{d_i}{n_i} \right).$$

<span style="color:red">Note: empirical likelihood is a special case of the nonparametric maximum likelihood (NPMLE).</span>

# Proportional hazards model (Cox model).

> Question: what if different individuals have different features?

Data: a collection of $\{(t_i, \Delta_i, x_i)\}$ with hidden $\{(d_i, c_i)\}$.
- $d_i$ : lifetime of individual $i$
- $c_i$ : censored time of individual $i$
- $t_i = \min\{c_i, d_i\}$ : death/censored time, whichever is earlier
  (right censoring)
- $\Delta_i = \mathbb{1}(d_i \leq c_i)$ . 1 if not censored, 0 if censored
  (true death)
- $x_i \in \mathbb{R}^p$ : feature vector of individual $i$.

## Continuous-time hazard rate

$$h(t) = \text{density of } (X = t \mid X \geq t) = \frac{f(t)}{S(t)} \leftarrow \text{unconditional density of } X$$

$$\Rightarrow \frac{d}{dt} \log S(t) = \frac{S'(t)}{S(t)} = -\frac{f(t)}{S(t)} = h(t)$$

$$\Rightarrow S(t) = \exp\left(-\int_0^t h(s)\, ds\right).$$

## Proportional hazards model

$$h(t \mid x) = e^{\beta^T x}\, h(t)$$

baseline hazard

log ratio : $\log \dfrac{h(t \mid x_1)}{h(t \mid x_2)} = \beta^T (x_1 - x_2)$

Target: estimate $\beta \in \mathbb{R}^p$.

## Partial likelihood.

The Cox model is solved by maximizing the following partial likelihood:

$$L(\beta) = \prod_{i:\Delta_i=1} \left( \frac{e^{x_i^T \beta}}{\sum_{j \in R_i} e^{x_j^T \beta}} \right)$$

where:
- $\{i : \Delta_i = 1\}$ represents the occurrences of true deaths
- $R_i$ : the set of individuals at risk when $i$ dies, i.e.
$$R_i = \{j : t_j \geq t_i\}$$
- each term represents the probability of "$i$ first dies among all individuals in the risk set $R_i$"
- no baseline hazard $h(t)$ in partial likelihood

## Derivation: profile likelihood

### Complete likelihood

$$L(\beta, h) \propto \prod_{i=1}^{n} \begin{cases} S(t_i | x_i) h(t_i | x_i), & \text{if } \Delta_i = 1 \\ S(t_i | x_i), & \text{if } \Delta_i = 0 \end{cases}$$

target    nuisance

( Implicit assumption: $c_i$ and $d_i$ are independent
conditioning on $x_i$ )

$$= \prod_{i=1}^{n} \left[ \exp(-e^{x_i^T \beta} H(t_i)) \left( e^{x_i^T \beta} h(t_i) \right)^{\Delta_i} \right]$$

$$\left( H(t) := \int_0^t h(s) \, ds \right)$$

### Profile likelihood.

$$\boxed{pL(\beta) = \sup_h L(\beta, h)}$$

Computation of profile likelihood in Cox model:

1. Using empirical likelihood to assume that
   $h$ is supported on $\{t_1, \cdots, t_n\}$ & $H(t_i) = \sum_{j: t_j \leq t_i} h(t_j)$

   $\Rightarrow L(\beta, h) = \prod_{i=1}^{n} \left[ \exp\left(-e^{x_i^T \beta} \sum_{j: t_j \leq t_i} h(t_j)\right) \left(e^{x_i^T \beta} h(t_i)\right)^{\Delta_i} \right]$

2. F.O.C. for $h(t_i)$:

   $$\frac{\partial(\log L)}{\partial h(t_i)} = \frac{\Delta_i}{h(t_i)} - \sum_{k: t_k \geq t_i} e^{x_k^T \beta} \begin{cases} = 0 & \text{if } h(t_i) > 0 \\ \leq 0 & \text{if } h(t_i) = 0 \end{cases}$$

   $$\Rightarrow h(t_i) = \begin{cases} 0 & \text{if } \Delta_i = 0 \\ \left(\sum_{k: t_k \geq t_i} e^{x_k^T \beta}\right)^{-1}, & \text{if } \Delta_i = 1 \end{cases}$$

3. A crucial identity: for above $h$,

   $$\prod_{i=1}^{n} \exp\left(-e^{x_i^T \beta} \sum_{j: t_j \leq t_i} h(t_j)\right)$$

   $$= \exp\left(-\sum_{i=1}^{n} e^{x_i^T \beta} \sum_{j: t_j \leq t_i} \frac{\Delta_j}{\sum_{k: t_k \geq t_j} e^{x_k^T \beta}}\right)$$

   $$= \exp\left(-\sum_{j=1}^{n} \Delta_j \underbrace{\sum_{i: t_i \geq t_j} \frac{e^{x_i^T \beta}}{\sum_{k: t_k \geq t_j} e^{x_k^T \beta}}}_{= 1}\right)$$

   $$= \prod_{j=1}^{n} \left(\frac{1}{e}\right)^{\Delta_j}$$

4. Plug back to $L(\beta, h)$:

$$pL(\beta) = \sup_h L(\beta, h)$$

$$= \prod_{i=1}^{\hat{n}} \left( \frac{1}{e} \frac{e^{x_i^T \beta}}{\sum_{k:\, t_k \geq t_i} e^{x_k^T \beta}} \right)^{\Delta_i}$$

$$\propto \prod_{i:\, \Delta_i = 1} \left( \frac{e^{x_i^T \beta}}{\sum_{k \in R_i} e^{x_k^T \beta}} \right),$$

agreeing with the partial likelihood.