Recitation 2

ABA Spring 2024
02/02/24

Agenda

- Kaplan Meier Example
- Likelihood functions and MLE
- MLE Example: Poisson

Reminders

- Quiz 2 due Feb 5, 11:59PM

Kaplan Meier Example

KM estimator

- The **Kaplan–Meier estimator** is a non-parametric statistic used to estimate the survival function (probability of a person surviving) from lifetime data.
- In medical research, it is often used to measure the fraction of patients living for a certain amount of time after treatment. For example, calculating the amount of time certain patient lived after he/she was diagnosed with the cancer or when his treatment starts. The estimator is named after **Edward L. Kaplan** and **Paul Meier**.
- Probability of survival is how many subject (patients) survive (do not perish) out of the total events (patients) at that time.
- The probability of survival at time t_i, $S(t_i)$, is calculated as

$$S(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i} = \frac{survive}{total}$$

Example

- A study involves 20 participants who are 65 years of age and older; they are enrolled over a 5-year period and are followed for up to 24 years until they die, the study ends, or they drop out of the study (lost to follow-up). [Note that if a participant enrolls after the study start, their maximum follow up time is less than 24 years. e.g., if a participant enrolls two years after the study start, their maximum follow up time is 22 years.]
- The data are shown. In the study, there are 6 deaths and 3 participants with complete follow-up (i.e., 24 years). The remaining 11 have fewer than 24 years of follow-up due to enrolling late or loss to follow-up.

participant	Year of Death	Year of Last Contact
1		24
2	3	
3		11
4		19
5		24
6		13
7	14	
8		2
9		18
10		17
11		24
12		21
13		12
14	1	
15		10
16	23	
17		6
18	5	
19		9
20	17	

- Survival probability is given by $S(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$
- We only take times when either the event or censoring happens
- 1, 2, 3, 5, 6, 9, 10, 11, 12, 13, 14, 17, 18, 19, 21, 23, 24
- The number at risk goes down only when di >0 which is only in years 1, 3, 5, 14, 17, 23
- So we write the survival probability as

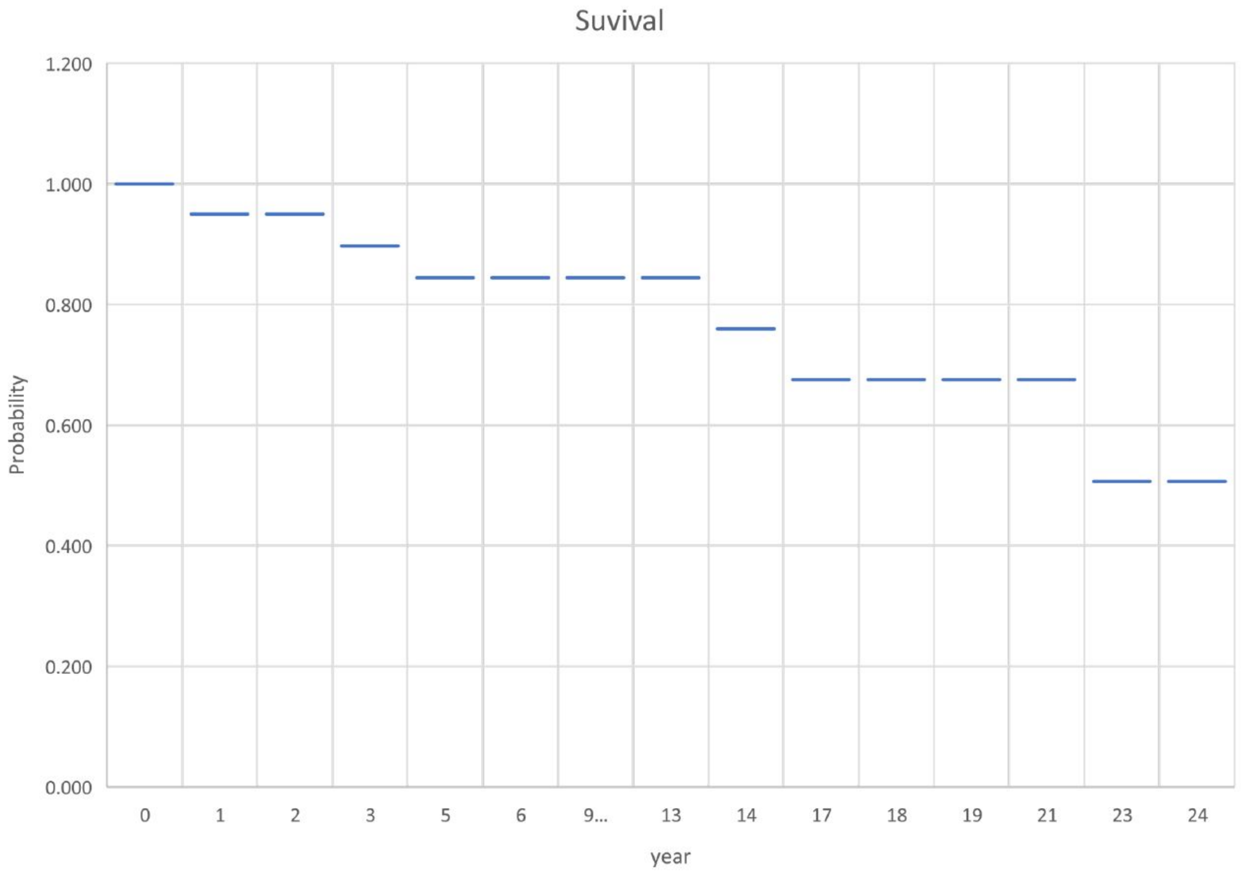
Example:
https://docs.google.com/spreadsheets/d/16qdtrjxo8sYUU0fnLTt1BjdE1rh8j0HFrZdri2x0DsM/edit?usp=sharing

KM table

- We use the same approach.
- Key to note is that censoring does not change the survival probability.

Time, Year	No at Risk - Nt	No of Deaths-Dt	No Censored - Ct	Survival Probability $S_{t+1} = S_t*(1-D_{t+1}/N_{t+1})$
0	20			1
1	20	1		=1*(1-1/20)=0.95
2	19		1	=0.95*(1-0/19)=0.95
3	18	1		=0.95*(1-1/18)=0.897
5	17	1		=0.897*(1-1/17) = 0.844
6	16		1	=0.844
9…	15…		1…	=0.844
13	11		1	=0.844
14	10	1		=0.844*(1-1/10) = 0.760
17	9	1	1	=0.760*(1-1/9) = 0.676
18	7		1	=0.676
19	6		1	=0.676
21	5		1	=0.676
23	4	1		=0.507
24	3		3	=0.507

KM Curve

- 

Suvival
1.200
1.000
0.800
0.600
0.400
0.200
0.000
Probability
0
1
2
3
5
6
9...
13
14
17
18
19
21
23
24
year

Maximum Likelihood

Motivation for Maximum likelihood estimation

- KM provided us with useful information about survival
- We need a “regression” like model to account for covariates
- We’ll use the widely studied Cox Proportional Hazard model (Cox Regression Model)
- Cox is a semi-parametric approach:

  We estimate the parameters (Betas) of the model using the partial likelihood.

Maximum likelihood estimation

- The goal of data analysis is to identify the population that is most likely to have generated the sample (we want to make inferences about the population)
- Each population is identified by a corresponding probability distribution (distribution of y, pdf/pmf).
- The desired parameters of the probability distribution are the ones that make the observed data “most likely”
- Likelihood: how likely the observed data is, given a set of parameters values. Formally: $L(w|y)$ equals the probability of the observed data, given the parameters.
- The MLE finds the parameters which make the distribution fit closest to the data. It does by maximizing the likelihood function.

Maximum likelihood estimation

- It turns out that likelihood function L(w|y) – where w is parameters and y is data - is proportional to the density function

$$L(w/y) \propto f(y/w)$$

- One simply maximizes L(.) to **recover parameter w.**
- With n **independent observations**, likelihood function is simply

$$L(w/y_1, y_2, \ldots.. y_n) = f(y_1, y_2, \ldots.. y_n/w) = \prod_n f(y_n/w)$$

  - Multiplication of n densities.
- Taking logs would simplify this (multiplication would turn to addition) and hence

$$\max_w LogL = \max_w \sum_n Log\left(f(y_n / w)\right)$$

MLE Example: Poisson

- How does one write a likelihood function if Y follows Poisson distribution?
- Since we know the pdf of Poisson,

$$f(y_i/\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$$

- λ is the parameter of Poisson distribution. For n sample, log likelihood (LL) function would be:

$$LL(\lambda|y) = \sum_{i=1}^{n} \text{Ln}(\frac{e^{-\lambda}\lambda^y}{y!})$$

- $= \sum_{i=1}^{n} -\lambda + yln(\lambda) - \ln(y!)$

- One maximizes this likelihood function to estimate parameter λ. This is simple maximize since it can be analytically solved

$$\frac{\partial LL}{\partial \lambda} = 0 \ or \ -\sum_{1}^{n} 1 + \sum_{i=1}^{n} \frac{y}{\lambda} = 0 \ \ or \ \lambda = \frac{1}{n}\sum y$$

- Or estimate λ is simply mean of y
- In many cases, this **can not be solved analytically**, and one relies on numerical methods (and we minimize the Negative LL)

Motivation for Maximum likelihood estimation

- KM provided us with useful information about survival
- We need a “regression” like model to account for covariates
- We’ll use the widely studied Cox Proportional Hazard model (Cox Regression)
- Cox is a semi-parametric approach:

We estimate the parameters (Betas) of the model using the partial likelihood.