Survival Analysis: Time To Event Modelling

Sudipta Das

Assistant Professor,
Department of Computer Science,
Ramakrishna Mission Vivekananda Educational & Research Institute

Outline I

1. Censoring and Truncation
   - Censoring
     - Right Censoring
     - Left Censoring
     - Interval Censoring
   - Truncation
   - Likelihood Construction for Censored and Truncated Data

Introduction

- Survival data or Time-to-event data present themselves in different ways which create special problems in analyzing such data.
- There are two peculiar features one can see in the survival data
  - Censoring
  - Truncation

Censoring

- Censored data arises when an individual's life length is known to occur only in a certain period of time.
  - Equivalently, some lifetimes are known to have occurred only within certain intervals.
- There are, mainly, three types of censoring
  - Right censoring
  - Left censoring
  - Interval censoring

Right Censoring

- Right censoring
  - Here, it is known that the individual is still alive up to a certain time
- We shall discuss three types of *right censoring*
  - Type I censoring
  - Type II censoring
  - Competing risks censoring.

Right Censoring: Type *I* censoring I

- Type I censoring: Here, the event is observed only if it occurs prior to some prespecified time.
  - These censoring times may be fixed for all the individuals or it may vary from individual to individual.
- The data from Type I censoring can be conveniently represented by pairs of random variables

$$(T, \delta),$$

  where δ indicates
  - whether the lifetime X corresponds to an event observed ($\delta = 1$) or is censored ($\delta = 0$), and
  - T is equal to X, if the lifetime is observed, and to C_r if it is censored, i.e., $T = \min(X, C_r)$.

Right Censoring: Type I censoring II

- Mathematical representation of right censored data: $(T_1, \delta_1), (T_2, \delta_2), \ldots, (T_n, \delta_n)$, where $\delta_1, \delta_2, \ldots, \delta_n$ are censoring indicators,

$$\delta_i = \begin{cases} 1 & \text{if } T_i \text{ is observed failure,} \\ 0 & \text{if } T_i \text{ is censoring time.} \end{cases}$$

Right Censoring: Type *I* censoring III

- We shall discuss two types of Type I censoring
  - Progressive Type I censoring:
    - Censoring times are may be different, however, fixed for all the individuals
  - Generalized Type I censoring:
    - Censoring time is random, however, known at the time of arrival of the individual

Right Censoring: Progressive Type I censoring I

- Progressive Type I censoring
  - Example 3.1:- Consider a large scale animal experiment conducted at the National Center for Toxicological Research (NCTR) in which mice were fed a particular dose of a carcinogen. The goal of the experiment was to assess the effect of the carcinogen on survival. Toward this end, mice were followed from the beginning of the experiment until death or until a prespecified censoring time was reached, when all those still alive were sacrificed (censored).

Right Censoring: Progressive Type I censoring II

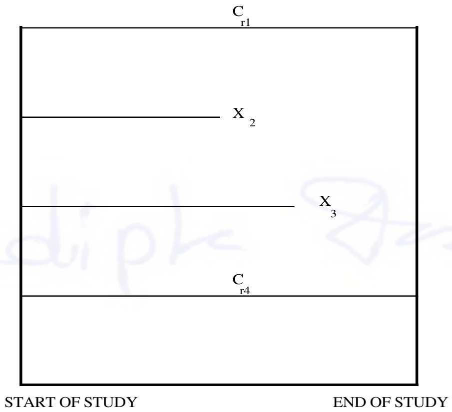

Figure 3.1 *Example of Type I censoring*

Right Censoring: Progressive Type I censoring III

- Progressive Type I censoring
  - Example 3.2:- Consider a mouse study where 200 mice were given a medical dose and each mouse was followed until death or until a prespecified sacrifice time (42 or 104 weeks) was reached (see Figure 3.2 for a schematic of this trial for one gender and one dose level). The two sacrifice times were chosen to reduce the cost of maintaining the animals while allowing for limited information on the survival experience of longer lived mice.

Right Censoring: Progressive Type I censoring IV

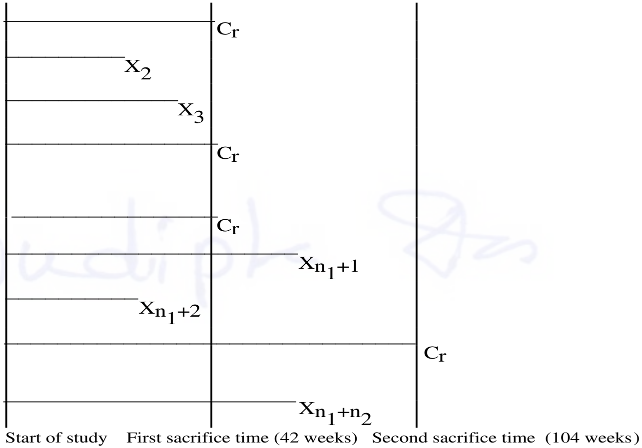

Figure 3.2 *Type I censoring with two different sacrifice times*

Right Censoring: Generalized Type I censoring I

- Generalized Type I censoring
  - In this case individuals enter the study at different times and the terminal point of the study is predetermined by the investigator, so that the censoring times are known when an individual is entered into the study.
  - In such studies (see Figure 3.3 for a hypothetical study with only four subjects), individuals have their own specific, fixed, censoring time.

Right Censoring: Generalized Type I censoring II

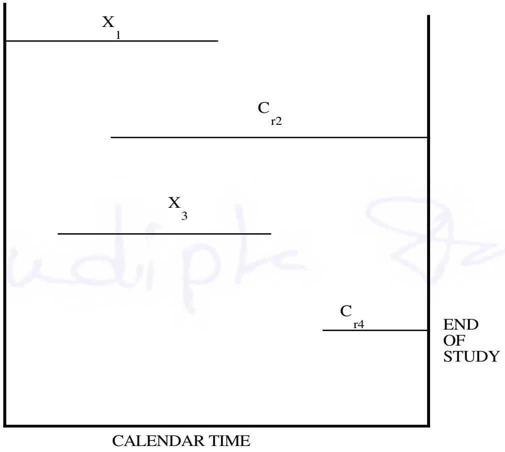

Figure 3.3 *Generalized Type I censoring when each individual has a different starting time*

- Equivalent representations

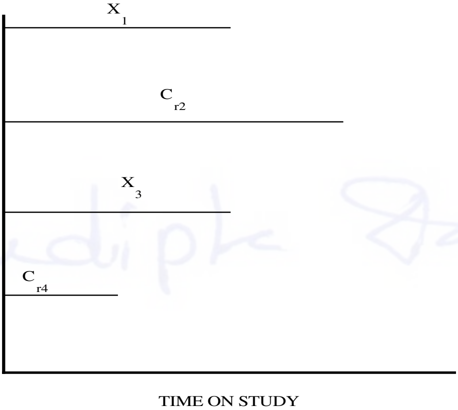

Figure 3.4 *Generalized Type I censoring for the four individuals in Figure 3.3 with each individuals starting time backed up to 0. $T_1 = X_1$ (death time for first individual) ($\delta_1 = 1$); $T_2 = C_{r2}$ (right censored time for second individual) ($\delta_2 = 0$); $T_3 = X_3$ (death time for third individual) ($\delta_3 = 1$); $T_4 = C_{r4}$ (right censored time for fourth individual) ($\delta_4 = 0$).*

Right Censoring: Generalized Type I censoring IV

- Lexis diagram

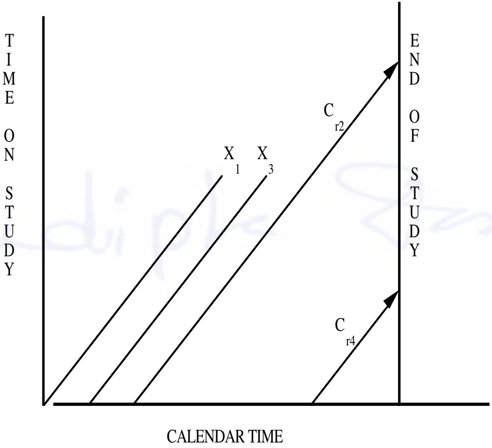

Figure 3.5 *Lexis diagram for generalized Type I censoring in Figure 3.3*

Right Censoring: Type *II* censoring I

- In Type II censoring, the study continues until the failure of the first r individuals, where r is some predetermined integer ($r < n$).
  - Experiments involving Type II censoring are often used in testing of equipment life.
  - Here, all items are put on test at the same time, and the test is terminated when r of the n items have failed.
- Such an experiment may save time and money because it could take a very long time for all items to fail.

Right Censoring: Type *II* censoring II

- It is also true that the statistical treatment of Type II censored data is simpler because the data consists of the r smallest lifetimes in a random sample of n lifetimes,
  - so that the theory of order statistics is directly applicable to determining the likelihood and any inferential technique employed.
- Here, it should be noted that
  - r the number of failures and
  - $n - r$ the number of censored observations are fixed integers and
  - the censoring time $T_{(r)}$, the rth ordered lifetime is random

Right Censoring: Progressive Type *II* censoring

- It is a generalization of Type II censoring.
- In this type of censoring,
  - the first r_1 failures (an integer chosen prior to the start of the study) in a sample of n items (or animals) are noted and recorded,
  - then $n_1 - r_1$ of the remaining $n - r_1$ unfailed items (or animals) are removed (or sacrificed) from the experiment, leaving $n - n_1$ items (or animals) on study,
  - after that, when the next r_2 items (another integer chosen prior to the start of the study) fail, $n_2 - r_2$ of the unfailed items are removed (or animals sacrificed).
  - This process continues until some predecided series of repetitions is completed.
- Again, r_i and $n_i (i = 1, 2)$ are fixed integers and the two censoring times, $T_{(r_1)}$ and $T_{(n_1+r_2)}$, are random.

Right Censoring: Competing risks censoring

- A third type of right censoring is competing risks censoring.
- A special case of competing risks censoring is random censoring.
- This type of censoring arises when we are interested in estimation of the marginal distribution of some event but some individuals under study may experience some competing event which causes them to be removed from the study.
- In such cases, the event of interest is not observable for those who experience the competing event and these subjects are random right censored at that time.

Left Censoring I

- A lifetime X associated with a specific individual in a study is considered to be *left censored* if it is less than a censoring time C_l (C_l for “left" censoring time),
  - that is, the event of interest has already occurred for the individual before that person is observed in the study at time C_l.
- For such individuals, we know that they have experienced the event sometime before time C_l, but their exact event time is unknown.
- The exact lifetime X will be known if, and only if, X is greater than or equal to C_l.

Left Censoring II

- The data from a left-censored sampling scheme can be represented by pairs of random variables

$$(T, \epsilon),$$

as in the previous section, where T is equal to X if the lifetime is observed and ϵ indicates whether the exact lifetime X is observed ($\epsilon = 1$) or not ($\epsilon = 0$).
- Note that, for left censoring as contrasted with right censoring,

$$T = \max(X, C_l).$$

Left Censoring III

- Example (Section 1.17):- In a study to determine the distribution of the time until first marijuana use among high school boys in California, the question was asked,
  - When did you you first use marijuana?
  - One of the responses was “I have used it but can not recall just when the first time was."
    - A boy who chose this response is indicating that the event had occurred prior to the boy’s age at interview but the exact age at which he started using marijuana is unknown.

Doubly Censoring I

- Often, if left censoring occurs in a study, right censoring may also occur, and the lifetimes are considered doubly censored
- Again, the data can be represented by a pair of variables (T, δ), where

$$T = \max[\min(X, C_r), C_l]$$

is the on study time;

  - δ is 1 if T is a death time, 0 if T is a right-censored time, and -1 if T is a left-censored time.
  - Here C_l is the time before which some individuals experience the event and C_r is the time after which some individuals experience the event.
  - X will be known exactly if it is less than or equal to C_r and greater than or equal to C_l.

Doubly Censoring II

- Example (Section 1.17):- Continuing the last example,
  - An additional possible response to the question “When did you first use marijuana?" was “I never used it" which indicates a right-censored observation.
- Thus, this is a doubly censored sampling scheme.

Interval Censoring I

- Interval censoring occurs when the lifetime is only known to occur within an interval.
- Such interval censoring occurs when patients in a clinical trial or longitudinal study have periodic follow-up and the patient's event time is only known to fall in an interval $(L_i, R_i]$
  - L for left endpoint and R for right end point of the censoring interval.
- This type of censoring may also occur in industrial experiments where there is periodic inspection for proper functioning of equipment items.

Interval Censoring II

- Interval censoring is a generalization of left and right censoring because,
  - when the left end point is 0 and the right end point is C_l we have left censoring and,
  - when the left end point is C_r and the right end point is infinite, we have right censoring.
- In addition, note that any combination of left, right, or interval censoring may occur in a study.
- Example (Section 1.18)

Truncation

- Truncation of survival data occurs when only those individuals whose event time lies within a certain observational window (Y_L, Y_R) are observed.
- **An individual whose event time is not in this interval is not observed and no information on this subject is available to the investigator.**
  - This is in contrast to censoring where there is at least partial information on each subject.
- Because we are only aware of individuals with event times in the observational window, the inference for truncated data is restricted to conditional estimation.

Left Truncation I

- When Y_R is infinite then we have left truncation.
- Here we only observe those individuals whose event time X exceeds the truncation time Y_L.
  - That is we observe X if and only if $Y_L < X$.
- In this case, individuals are selected and followed prospectively until failure or censoring, but their current lifetime at selection is not $t = 0$, but some value $u > 0$.
  - i.e., subjects may not come under actual observation until after the beginning time point has passed.
- It is called as *delayed entry*, also.

Left Truncation II

- Example (Section 1.16):-
  - In a survival study of residents of the Channing House retirement center located in California, ages at death (in months) are recorded, as well as ages at which individuals entered the retirement community (the truncation event).
  - Since an individual must survive to a sufficient age to enter the retirement center, all individuals who died earlier will not enter the center and thus are out of the investigator's cognizance;
    - i.e., such individuals have no chance to be in the study and are considered left truncated.
  - A survival analysis of this data set needs to account for this feature.

Right Truncation I

- When Y_L is zero then we have right truncation.
- Here we only observe those individuals whose event time X precedes the truncation time Y_R.
  - That is we observe X if and only if $X < Y_R$.
- This type of incomplete observation occurs when the entire study population has experienced the event of interest before the study begins
  - i.e., subjects have been selected because they have experienced the event of interest.

Right Truncation II

- Example (Section 1.19):-
  - Consider a AIDS study, where patients with transfusion-induced AIDS were sampled.
  - Retrospective determination of the transfusion times were used to estimate the waiting time from infection at transfusion to clinical onset of AIDS.
  - The registry was sampled on June 30, 1986, so only those whose waiting time from transfusion to AIDS was less than the time from transfusion to June 30, 1986, were available for observation.
  - Patients transfused prior to June 30, 1986, who developed AIDS after June 30, 1986, were not observed and are right truncated.

Likelihood Construction for Censored and Truncated Data I

- In constructing a likelihood function for censored or truncated data we need to consider carefully what information each observation gives us.
- We assume that the lifetimes and censoring times are independent.
- Exact lifetimes: An observation corresponding to an exact event time provides information on the probability that the event's occurring at this time, which is approximately equal to the density function of X at this time.
  - Likelihood contribution:- $f(x)$

Likelihood Construction for Censored and Truncated Data II

- Right-censored observations: We know that the event time is larger than this time, so the information is the survival function evaluated at the on study time.
    - Likelihood contribution:- $S(C_r)$
- Left-censored observations: We know that the event has already occurred, so the contribution to the likelihood is the cumulative distribution function evaluated at the on study time.
    - Likelihood contribution:- $1 - S(C_l)$
- Interval-censored observations: We know that the event occurred within the interval, so the information is the probability that the event time is in this interval.
    - Likelihood contribution:- $S(L) - S(R)$

Likelihood Construction for Censored and Truncated Data III

- For truncated data these probabilities are replaced by the appropriate conditional probabilities.
  - Likelihood contribution for left-truncated observations left truncated at time Y_L
  $$\frac{f(x)}{S(Y_L)}$$
  - Likelihood contribution for right-truncated observations right truncated at time Y_R
  $$\frac{f(x)}{1 - S(Y_R)}$$
  - Likelihood contribution for interval-truncated observations in the interval $[Y_L, Y_R]$
  $$\frac{f(x)}{S(Y_L) - S(Y_R)}$$

Overall Likelihood for Censored and Truncated Data I

- The overall likelihood function can be constructed by putting together the component parts as

$$L \propto \prod_{i \in D} f_i(x_i) \prod_{i \in R} S_i(C_r) \prod_{i \in L} [1 - S_i(C_l)] \prod_{i \in I} [S_i(L) - S_i(R)]$$

  where

  - D is the set of death times,
  - R the set of right-censored observations,
  - L the set of left-censored observations, and
  - I the set of interval- censored observations.

Overall Likelihood for Censored and Truncated Data II

- For left-truncated data, with truncation interval (Y_{L_i}, Y_{R_i}) independent from the death time, we replace
  - $f_i(x_i)$ by $\dfrac{f_i(x_i)}{S_i(Y_{L_i}) - S_i(Y_{R_i})}$ and
  - $S_i(C_i)$ by $\dfrac{S_i(C_i)}{S_i(Y_{L_i}) - S_i(Y_{R_i})}$
- For right-truncated data, only deaths are observed, therefore we *only* replace

$$f_i(x_i) \text{ by } \frac{f_i(x_i)}{1 - S_i(x_i)}$$

- For identical failure time distribution
  - $f_i(\cdot)$ and $S_i(\cdot)$ are replaced by $f(\cdot)$ and $S(\cdot)$, respectively.

Likelihood construction for Type I censoring I

- Data from experiments involving Type I censoring can be conveniently represented by pairs of random variables

$$(T, \delta),$$

  - where δ indicates whether the lifetime X is observed ($\delta = 1$) or not ($\delta = 0$),
  - and T is equal to X if the lifetime is observed and to C_r if it is right-censored, i.e.,
$$T = \min(X, C_r).$$
  - We assume X and C_r are independent random variables,
    - with f and g being their respective probability density functions
    - with S and G being their respective survival functions

Likelihood construction for Type *I* censoring II

- For $\delta = 0$,

$$\begin{aligned} Pr[T = t, \delta = 0] &= Pr[C_r = t, C_r < X] \\ &= Pr[C_r < X | C_r = t] Pr[C_r = t] \\ &= S(t)g(t). \end{aligned}$$

- For $\delta = 1$,

$$\begin{aligned} Pr[T = t, \delta = 1] &= Pr[X = t, X < C_r] \\ &= Pr(Cr > X | X = t) Pr(X = t) \\ &= G(t)f(t). \end{aligned}$$

- Combining into a single expression

$$\begin{aligned} Pr(t, \delta) &= [f(t)]^{\delta}[S(t)]^{1-\delta}[G(t)]^{\delta}[g(t)]^{1-\delta} \\ &\propto [f(t)]^{\delta}[S(t)]^{1-\delta} \end{aligned}$$

Likelihood construction for Type I censoring III

- If we have a i.i.d. random sample of pairs $(t_i, \delta_i), i = 1, \ldots, n$, the likelihood function is

$$
\begin{aligned}
L &= \prod_{i=1}^{n} Pr(t_i, \delta_i) \\
&= \prod_{i=1}^{n} [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i} \\
&= \prod_{i=1}^{n} [h(t_i)]^{\delta_i} S(t_i) \\
&= \prod_{i=1}^{n} [h(t_i)]^{\delta_i} e^{-H(t_i)}
\end{aligned}
$$

Likelihood construction for Type *II* censoring I

- Data from experiments involving Type II censoring consist of the rth smallest lifetimes

$$T_{(1)}, T_{(2)}, \ldots, T_{(r)}$$

  out of a random sample of n lifetimes $X_1, \ldots, X_n$ from the assumed life distribution.

- Assuming $T_1, \ldots, T_n$ are i.i.d. and have a continuous distribution with p.d.f. $f(t)$ and survival function $S(t)$, it follows that the joint p.d.f. of $T_{(1)}, T_{(2)}, \ldots, T_{(r)}$ is

$$\begin{aligned} L &= \frac{n!}{(n-r)!}\left[\prod_{i=1}^{r} f(t_{(i)})\right]\left[S(t_{(r)})\right]^{n-r} \\ &\propto \left[\prod_{i=1}^{r} f(t_{(i)})\right]\left[S(t_{(r)})\right]^{n-r} \end{aligned}$$

Likelihood construction for Exponential Failure Distribution I

- Suppose that lifetimes T_i are independent and follow an exponential distribution with p.d.f. $f(t) = \lambda e^{-\lambda t}$ and survival function $S(t) = e^{-\lambda t}$.
- Thus, the likelihood

$$L(\lambda) \quad \propto \quad \begin{cases} \lambda^r e^{-\lambda\left[\sum_{i=1}^{n} t_i\right]}, & \text{Type I censoring} \\ \lambda^r e^{-\lambda\left[\sum_{i=1}^{r} t_{(i)} + (n-r)t_{(r)}\right]}, & \text{Type II censoring} \end{cases}$$

Likelihood construction for Exponential Failure Distribution II

- Hence, *m.l.e.* of λ is

$$\hat{\lambda} = \begin{cases} r/\left[\sum_{i=1}^{n} t_i\right], & \text{Type I censoring} \\ r/\left[\sum_{i=1}^{r} t_{(i)} + (n-r)t_{(r)}\right], & \text{Type II censoring} \end{cases}$$

- Note that
  - For Type I censoring, the exact distribution of $\hat{\lambda}$ is intractable.
  - For Type II censoring,

$$2r\lambda/\hat{\lambda} \sim \chi^2_{2r}.$$

(*Kalbfleisch*:55)