Survival Function Estimation with Recurrent Events: Case of Retinal Neural Firing

Eric Kawaguchi
Katrina Harper

The University of Iowa

July 19 2012

Introduction

Outline:

1. Introduction to Survival Analysis

Introduction

Outline:

1. Introduction to Survival Analysis
2. Survival Analysis with a Single Event

Introduction

Outline:

1. Introduction to Survival Analysis
2. Survival Analysis with a Single Event
3. How to Estimate the Survival Function of Recurrent Events

Introduction

Outline:

1. Introduction to Survival Analysis
2. Survival Analysis with a Single Event
3. How to Estimate the Survival Function of Recurrent Events
4. Analysis of our Project Data

Introduction

Outline:

1. Introduction to Survival Analysis
2. Survival Analysis with a Single Event
3. How to Estimate the Survival Function of Recurrent Events
4. Analysis of our Project Data
5. Future Work

What is Survival Analysis?

- Survival analysis is a field of statistics that analyzes and models time-to-event data.

What is Survival Analysis?

- Survival analysis is a field of statistics that analyzes and models time-to-event data.
- This area of statistics is widely used in medical research, economics, and reliability.

What is Survival Analysis?

- Survival analysis is a field of statistics that analyzes and models time-to-event data.
- This area of statistics is widely used in medical research, economics, and reliability.
- Examples:
  - Time from cancer remission to relapse.

What is Survival Analysis?

- Survival analysis is a field of statistics that analyzes and models time-to-event data.
- This area of statistics is widely used in medical research, economics, and reliability.
- Examples:
  - Time from cancer remission to relapse.
  - Duration between administering a treatment and recovery.

Special Features of Survival Data

- One problem we encounter with time-to-event data is **censoring**.

Special Features of Survival Data

- One problem we encounter with time-to-event data is **censoring**.
- Censoring occurs when an individual's end-point of interest has not been observed.

Special Features of Survival Data

- One problem we encounter with time-to-event data is **censoring**.
- Censoring occurs when an individual's end-point of interest has not been observed.
- This can be the result of different circumstances.

Special Features of Survival Data

- One problem we encounter with time-to-event data is **censoring**.
- Censoring occurs when an individual's end-point of interest has not been observed.
- This can be the result of different circumstances.
- Examples:
  - Data from a study is to be analyzed at a point of time when some individuals' events have not occurred yet.

Special Features of Survival Data

- One problem we encounter with time-to-event data is **censoring**.
- Censoring occurs when an individual's end-point of interest has not been observed.
- This can be the result of different circumstances.
- Examples:
  - Data from a study is to be analyzed at a point of time when some individuals' events have not occurred yet.
  - Individuals whose survival times can not be analyzed because they have been lost to follow-up.

The University of Iowa

Special Features of Survival Data

- One problem we encounter with time-to-event data is **censoring**.
- Censoring occurs when an individual's end-point of interest has not been observed.
- This can be the result of different circumstances.
- Examples:
  - Data from a study is to be analyzed at a point of time when some individuals' events have not occurred yet.
  - Individuals whose survival times can not be analyzed because they have been lost to follow-up.
  - Occurance of an event due to a cause other than one of interest.

Terminology

Suppose there are n individuals with observed survival times $t_1, t_2, t_3, \dots, t_n$.

Terminology

Suppose there are n individuals with observed survival times $t_1, t_2, t_3, \ldots, t_n$.

- It is possible that some of these individuals have censored observation times and that some may have the same observed survival time.

Terminology

Suppose there are n individuals with observed survival times $t_1, t_2, t_3, \dots, t_n$.

- It is possible that some of these individuals have censored observation times and that some may have the same observed survival time.
- Let r be the number of event times amongst the individuals, so that $r \leq n$, where r does not include censored times and duplicate times are considered as one time.

Terminology

Suppose there are n individuals with observed survival times $t_1, t_2, t_3, \ldots, t_n$.

- It is possible that some of these individuals have censored observation times and that some may have the same observed survival time.
- Let r be the number of event times amongst the individuals, so that $r \leq n$, where r does not include censored times and duplicate times are considered as one time.
- Now let $t_{(j)}$, for $j = 1, 2, 3, \ldots, r$, be the r ordered event times such that $t_{(1)} < t_{(2)} < \ldots t_{(r)}$.

Terminology

- Let n_j be the number of individuals at risk just before $t_{(j)}$.

Terminology

- Let n_j be the number of individuals at risk just before $t_{(j)}$.
- Let d_j denote the number of events at $t_{(j)}$.

Terminology

- Let n_j be the number of individuals at risk just before $t_{(j)}$.
- Let d_j denote the number of events at $t_{(j)}$.
- The **survival function**, $S(t)$, is the probability that the time of event is later than some specified time t.

$$S(t) = \Pr(T > t)$$

Terminology

- Let n_j be the number of individuals at risk just before $t_{(j)}$.
- Let d_j denote the number of events at $t_{(j)}$.
- The **survival function**, $S(t)$, is the probability that the time of event is later than some specified time t.

$$S(t) = \Pr(T > t)$$

- A common way of estimating the survival function is the Kaplan-Meier estimator. (Kaplan, E.L & Meier P., 1958)

The Kaplan-Meier Estimator

- **Kaplan-Meier (Product Limit) Estimate:**

$$\widehat{S}(t) = \prod_{j=1}^{r} \left\{1 - \frac{d_j}{n_j}\right\}$$

$$\text{SE } \widehat{S}(t) \approx \widehat{S}(t) \left\{\sum_{j=1}^{r} \frac{d_j}{n_j(n_j - d_j)}\right\}^{\frac{1}{2}}$$

The Kaplan-Meier Estimator

- **Kaplan-Meier (Product Limit) Estimate:**

$$\widehat{S}(t) = \prod_{j=1}^{r} \left\{1 - \frac{d_j}{n_j}\right\}$$

$$\text{SE } \widehat{S}(t) \approx \widehat{S}(t) \left\{\sum_{j=1}^{r} \frac{d_j}{n_j(n_j - d_j)}\right\}^{\frac{1}{2}}$$

- $(1-\alpha) * 100\%$ **Confidence Intervals of** $\widehat{S}(t)$**:**

$$\widehat{S}(t) \pm z_{\alpha/2} * \text{ SE } \widehat{S}(t)$$

THE UNIVERSITY OF IOWA

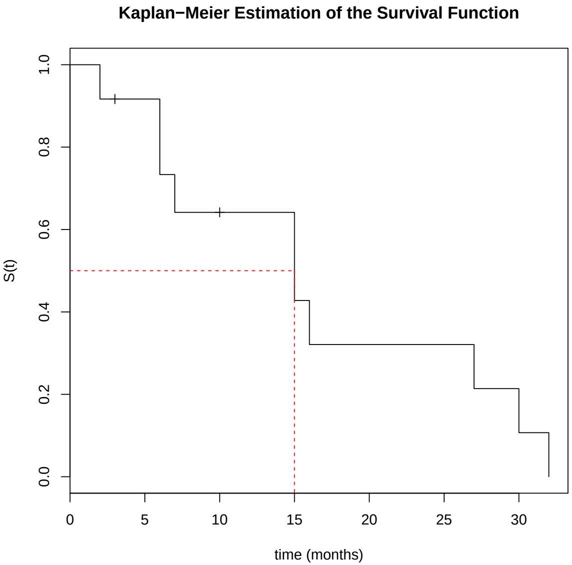
Illustrative Example of Kaplan-Meier

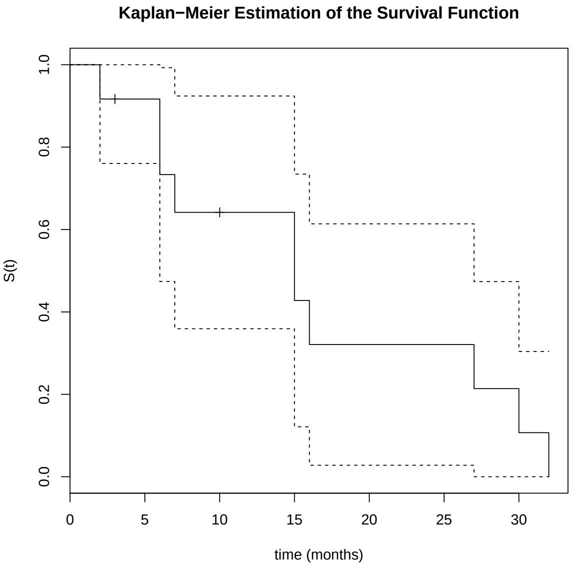
Example

Suppose there is a sample of 12 hemophiliacs under the age of 40 with HIV seroconversion.
Instead of using time intervals, exact times at which failures occured are used.
These times, in months, are listed below:

Patient	1	2	3	4	5	6	7	8	9	10	11	12
Months	2	3*	6	6	7	10*	15	15	16	27	30	32

NOTE: * denotes censored values.


Kaplan−Meier Estimation of the Survival Function
S(t)
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
time (months)


Kaplan−Meier Estimation of the Survival Function
S(t)
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
time (months)

Survival Function Estimation with Recurrent Events:

The purpose of our project is to estimate the survival function of recurrent event data.

Survival Function Estimation with Recurrent Events:

The purpose of our project is to estimate the survival function of recurrent event data.

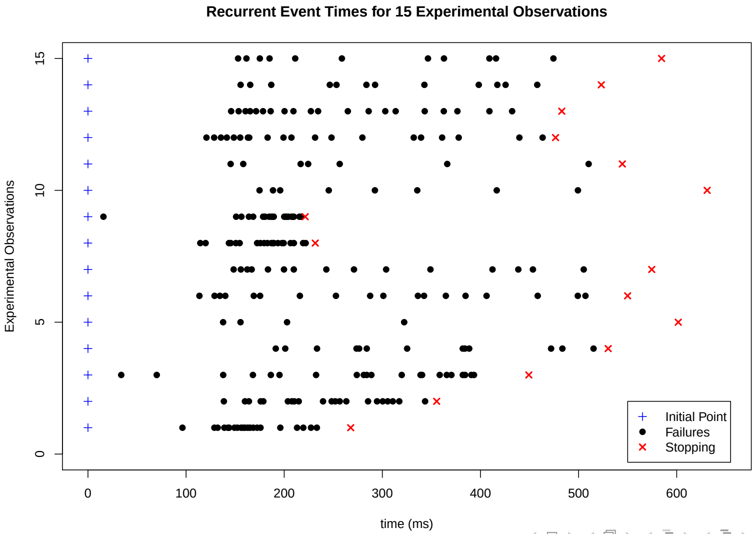
- When multiple events occur within the same subject, they are known as **recurrent events**.

Survival Function Estimation with Recurrent Events:

The purpose of our project is to estimate the survival function of recurrent event data.

- When multiple events occur within the same subject, they are known as **recurrent events**.
- Examples include migraines, seizures, heart attacks, strokes etc.

Time to Event Plot



Terminology

Wang and Chang (1999) developed a non-parametric estimator of the survival function for recurrent events.

Terminology

Wang and Chang (1999) developed a non-parametric estimator of the survival function for recurrent events.

- Let i be the index for an individual or subject.

Terminology

Wang and Chang (1999) developed a non-parametric estimator of the survival function for recurrent events.

- Let i be the index for an individual or subject.
- Let j be the index for an event.
  $j = 0$ denotes the index for the initial event.

Terminology

Wang and Chang (1999) developed a non-parametric estimator of the survival function for recurrent events.

- Let i be the index for an individual or subject.
- Let j be the index for an event.
  $j = 0$ denotes the index for the initial event.
- Let T_{ij} be the time from the $(j-1)^{th}$ to the j^{th} event for subject i.

THE UNIVERSITY OF IOWA

Terminology

Wang and Chang (1999) developed a non-parametric estimator of the survival function for recurrent events.

- Let i be the index for an individual or subject.
- Let j be the index for an event.
  $j = 0$ denotes the index for the initial event.
- Let T_{ij} be the time from the $(j-1)^{th}$ to the j^{th} event for subject i.
- Let the censoring time, C_i, be the time between the initial event and the end of follow-up for subject i.

Terminology

- Let m_i denote the index satisfying:

$$\sum_{j=1}^{m_i - 1} T_{ij} \leq C_i$$

and

$$\sum_{j=1}^{m_i} T_{ij} > C_i$$

Terminology

- Let m_i denote the index satisfying:

$$\sum_{j=1}^{m_i-1} T_{ij} \leq C_i$$

and

$$\sum_{j=1}^{m_i} T_{ij} > C_i$$

- Let m_i be the number of recurrent events for subject i.
m_i^* is the number of uncensored recurrent events for subject i.

$$m_i^* = \begin{cases} 1 & \text{if} \quad m_i = 1 \\ m_i - 1 & \text{if} \quad m_i \geq 2 \end{cases}$$

THE UNIVERSITY OF IOWA

Terminology

- Let m_i denote the index satisfying:

$$\sum_{j=1}^{m_i-1} T_{ij} \leq C_i$$

and

$$\sum_{j=1}^{m_i} T_{ij} > C_i$$

- Let m_i be the number of recurrent events for subject i.
  m_i^* is the number of uncensored recurrent events for subject i.

$$m_i^* = \begin{cases} 1 & \text{if} \quad m_i = 1 \\ m_i - 1 & \text{if} \quad m_i \geq 2 \end{cases}$$

- Let y_{ij} be our observed recurrence times defined by:

$$y_{ij} = \begin{cases} t_{ij} & \text{if} \quad j = 1, \ldots, m_i - 1 \\ t_{i,m_i}^{+} & \text{if} \quad j = m_i \end{cases}$$

Terminology

- Let $R^*(t)$ be the total mass of the risk set at time t be calculated as:

$$R^*(t) = \sum_{i=1}^{n} \left[\frac{a_i}{m_i^*} \sum_{j=1}^{m_i^*} I(y_{ij} \geq t) \right]$$

The University of Iowa

Terminology

- Let $R^*(t)$ be the total mass of the risk set at time t be calculated as:

$$R^*(t) = \sum_{i=1}^{n} \left[\frac{a_i}{m_i^*} \sum_{j=1}^{m_i^*} I(y_{ij} \geq t) \right]$$

  - Let a_i be defined as a positive-valued function of the censored value subject to the constraint $E(a_i^2) < \infty$.

THE UNIVERSITY OF IOWA

Terminology

- Let $R^*(t)$ be the total mass of the risk set at time t be calculated as:

$$R^*(t) = \sum_{i=1}^{n} \left[\frac{a_i}{m_i^*} \sum_{j=1}^{m_i^*} I(y_{ij} \geq t) \right]$$

  - Let a_i be defined as a positive-valued function of the censored value subject to the constraint $E(a_i^2) < \infty$.
  - The indicator function, $I(y_{ij} \geq t)$, is a binary operator with values 0 if $y_{ij} < t$ and 1 if $y_{ij} \geq t$.

Terminology

- Let $R^*(t)$ be the total mass of the risk set at time t be calculated as:

$$R^*(t) = \sum_{i=1}^{n} \left[\frac{a_i}{m_i^*} \sum_{j=1}^{m_i^*} I(y_{ij} \geq t) \right]$$

  - Let a_i be defined as a positive-valued function of the censored value subject to the constraint $E(a_i^2) < \infty$.
  - The indicator function, $I(y_{ij} \geq t)$, is a binary operator with values 0 if $y_{ij} < t$ and 1 if $y_{ij} \geq t$.
  - $R^*(t)$ is the summation of the weighted average of the total number of observed uncensored recurrent times for a subject that are greater than or equal to t.

THE UNIVERSITY OF IOWA

Terminology

- Let the mass evaluated at time t be:

$$d^*(t) = \sum_{i=1}^{n} \left[\frac{a_i I(m_i \geq 2)}{m_i^*} \sum_{j=1}^{m_i^*} I(y_{ij} = t) \right]$$

Terminology

- Let the mass evaluated at time t be:

$$d^*(t) = \sum_{i=1}^{n} \left[\frac{a_i I(m_i \geq 2)}{m_i^*} \sum_{j=1}^{m_i^*} I(y_{ij} = t) \right]$$

  - $d^*(t)$ is the summation of the weighted average of the total number of observed uncensored recurrent times for a subject that are equal to t.

Wang and Chang Product-Limit Estimation (1999)

- Letting $y_1^*, y_2^*, \ldots, y_K^*$ be the ordered, and distinct uncensored times, Wang and Chang created a Kaplan-Meier type estimator:

$$\widehat{S}_n(t) = \prod_{y_i^* \leq t} \left\{ 1 - \frac{d^*(y_i^*)}{R^*(y_i^*)} \right\}$$

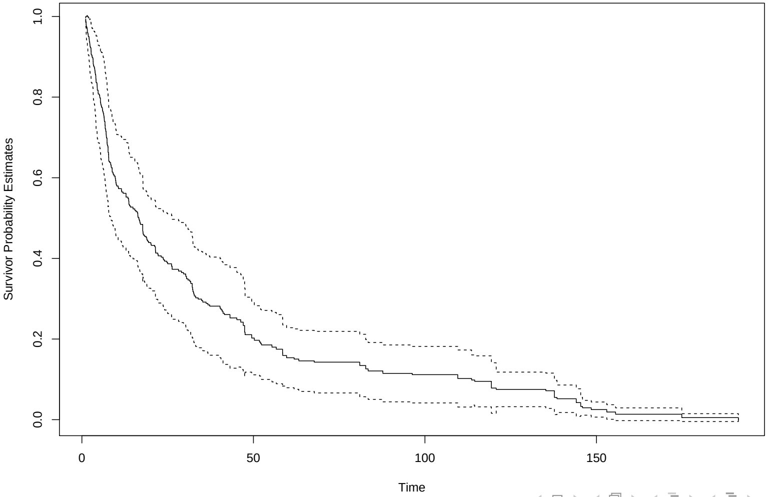
Wang and Chang Product-Limit Estimation (1999)

- Letting $y_1^*, y_2^*, \ldots, y_K^*$ be the ordered, and distinct uncensored times, Wang and Chang created a Kaplan-Meier type estimator:

$$\widehat{S}_n(t) = \prod_{y_i^* \leq t} \left\{1 - \frac{d^*(y_i^*)}{R^*(y_i^*)}\right\}$$

- This estimator also sets a_i equal to 1, giving every estimate equal weight.

Product-Limit Estimate



Analysis

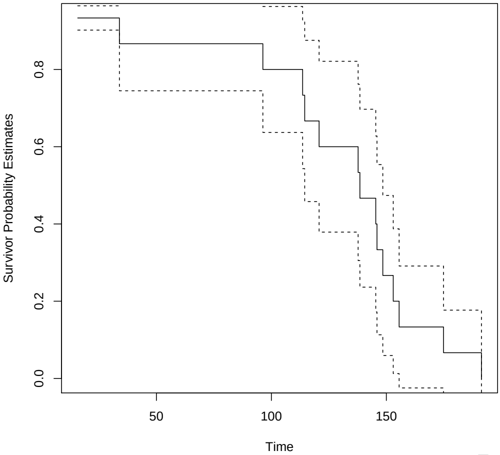
- Using the R package **survrec**, we found the median survival time, $t_{.5}$, to be 16.6 ms.

Analysis

- Using the R package **survrec**, we found the median survival time, $t_{.5}$, to be 16.6 ms.
- The 95% confidence interval around the median survival time is $(0.4693057, 32.7306943)$.
  - The standard error was found through Greenwood's formula. (Greenwood M., 1926)

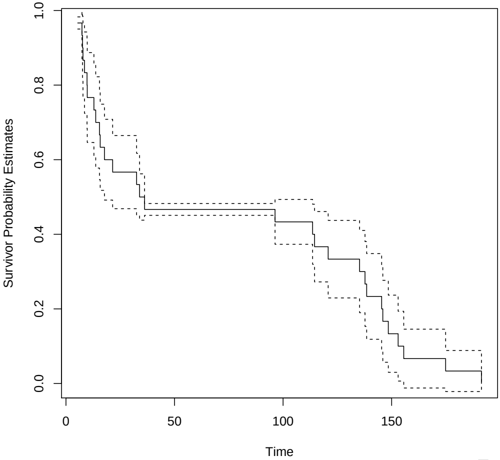
WC Plot for 2 Inter-Event Times:

Median Survival Time: 138 ms



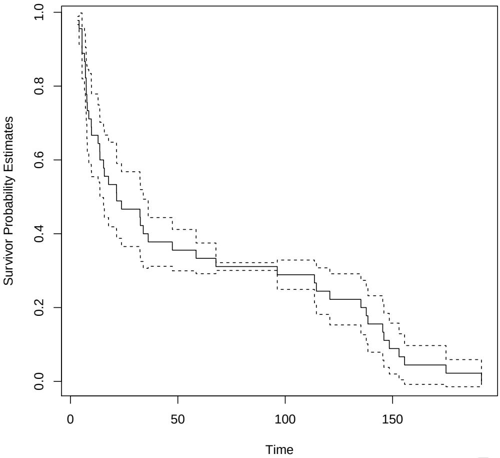
WC Plot for 3 Inter-Event Times:

Median Survival Time: 35 ms



WC Plot for 4 Inter-Event Times:

Median Survival Time: 21.5 ms



Future Work

- Latency

Future Work

- Latency
- Correlation

Future Work

- Latency
- Correlation
- More Subjects

Acknowledgements

Special Thanks To:

- Dr. Gideon K.D. Zamba, PhD.
- University of Iowa, Department of Biostatistics
- University of Iowa, Department of Ophthamology and Neurology
- National Heart Lung and Blood Institute (NHLBI)
- National Center for Research Resources (NCRR) of the National Institutes of Health (NIH)

Partly supported by the University of Iowa Department of Biostatistics NHLBI-NCRR-NIH training grant (T15-HL097622-01 NHLBI)

References

- Greenwood, M. 1946. The Statistical Study of Infectious Diseases. *Journal of the Royal Statistical Society* 109(2):85-110
- Kaplan, E.L., Meier, P. 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53:457–481
- Wang, M.C., Chang, S.H. 1999. Nonparametric Estimation of a Recurrent Survival Function. *Journal of the American Statistical Association* 94:146–153.