

Chapter 2

(AST405) Lifetime data analysis

Md Rasel Biswas

Lecture Outline

1 2. Observed Schemes, Censoring, and Likelihood

- 2.1 Introduction
- 2.2 Right censoring and maximum likelihood
- 2.3 Other type of incomplete data

Section 1

2. Observed Schemes, Censoring, and Likelihood

Subsection 1

2.1 Introduction

Preliminary discussion of likelihood;

- Suppose that the probability distribution of potentially observable data in a study specified up to the parameter vector θ
- A likelihood function for θ is a function of θ and is proportional to the probability of data that were observed

$$L(\theta) \propto Pr(\text{Data}; \theta) \quad (2.1.1)$$

- ▶ Data \rightarrow observed data
- ▶ Pr \rightarrow probability density or mass function from which the data are assumed to arise
- ▶ $L(\theta; \text{Data})$ \rightarrow is a more formal notation for likelihood function

Asymptotic results and large sample methods

- Assume that the data consist of a random sample y_1, \dots, y_n from a distribution with probability density function $f(y; \theta)$
- $\theta = (\theta_1, \dots, \theta_k)'$ \rightarrow a vector of unknown parameters, $\theta \in \Omega$
- y_i 's can be vectors, but for simplicity, they are considered scalars
- The likelihood function for θ

$$L(\theta) = \prod_{i=1}^n f(y_i; \theta)$$

- If y_i 's are independent but not identically distributed then $f(y_i; \theta)$ is replaced by $f_i(y_i; \theta)$ in the definition of likelihood function

Asymptotic results and large sample methods

- Let $\hat{\theta}$ be a point in Ω at which $L(\theta)$ is maximized
 - ▶ $\hat{\theta} \rightarrow$ maximum likelihood estimator (m.l.e.) of θ
 - ▶ In most simple settings $\hat{\theta}$ exists and unique
- It is often convenient to work with the log-likelihood function, which is also maximized at $\hat{\theta}$

$$\ell(\theta) = \log L(\theta) = \sum_i \log f(y_i; \theta)$$

Asymptotic results and large sample methods

- The m.l.e. $\hat{\theta}$ can be found by solving

$$U_j(\theta) = \mathbf{0} \quad (j = 1, \dots, k)$$

- ▶ $U_j(\theta) = \frac{\partial \ell(\theta)}{\partial \theta_j} \rightarrow jth$ score or score function
- ▶ $\mathbf{U}(\theta) = [U_1(\theta), \dots, U_k(\theta)]' \rightarrow$ score vector of order $k \times 1$

Asymptotic results and large sample methods

- Score vector $U(\theta)$ is asymptotically (k -variate) normally distributed with mean $\mathbf{0}$ and variance-covariance matrix $\mathcal{J}(\theta)$ with the (j, j') th entry of $\mathcal{J}(\theta)$

$$\mathcal{J}_{jj'}(\theta) = E \left(\frac{-\partial^2 \ell(\theta)}{\partial \theta_j \partial \theta_{j'}} \right) \quad j, j' = 1, \dots, k$$

- ▶ $\mathcal{J}(\theta) \rightarrow$ Fisher or expected information matrix
- ▶ $I(\hat{\theta}) \rightarrow$ observed information matrix
- Under mild regularity conditions
 - ▶ $\hat{\theta}$ is a consistent estimator of θ
 - ▶ $I(\hat{\theta})/n$ is a consistent estimator of $\mathcal{J}(\theta)/n$

Optimization methods for maximum likelihood

- Maximum likelihood estimator $\hat{\theta}$ correspond to the maximum of the log-likelihood function $\ell(\theta)$, i.e.

$$\hat{\theta} = \arg \max_{\theta \in \Omega} \ell(\theta)$$

Optimization methods for maximum likelihood

- There are different numerical approaches for optimizing the multiparameter log-likelihood function $\ell(\theta)$ for $\theta \in \Omega$
- Many likelihood functions have a unique maximum at $\hat{\theta}$, which is a stationary point satisfying $\partial\ell(\theta)/\partial\theta = \mathbf{0}$
- Numerical approaches involve a starting point θ_0 and an iterative procedure designed to give a sequence of points $\theta_1, \theta_2, \dots$ converging to $\hat{\theta}$ (i.e. when $\theta_j \simeq \theta_{j-1}$)

Optimization methods for maximum likelihood

- Three types of optimization methods are available in statistical software
 - ▶ Methods that do not use derivatives (e.g., simplex algorithm, such as Nelder-Mead method)
 - ▶ Methods that use only first derivatives $U(\theta) = \partial\ell(\theta)/\partial\theta$ (e.g. steepest ascent, quasi-Newton, and conjugate gradient method)
 - ▶ Methods that use both first derivative $U(\theta)$ and second derivative $H(\theta) = \partial^2\ell(\theta)/\partial\theta\partial\theta'$ (e.g. Newton-Raphson method)

$H(\theta)$ is known as Hessian matrix and observed information matrix
 $I(\hat{\theta}) = -H(\hat{\theta})$

Optimization methods for maximum likelihood

- Newton-Raphson method is commonly used for optimizing $\ell(\theta)$, which is based on the iteration scheme for the j th step ($j = 1, 2, \dots$) as

$$\theta_j = \theta_{j-1} - [H(\theta_{j-1})]^{-1} U(\theta_{j-1})$$

- ▶ θ_{j-1} \rightarrow value of θ at the $(j - 1)$ th iteration
-

- Using Taylor series expansion, expanding $U(\theta)$ at θ_j

$$U(\theta) = U(\theta_j) + H(\theta_j)(\theta - \theta_j)$$

- ▶ Then

$$U(\theta) = \mathbf{0} \Rightarrow \theta = \theta_j - [H(\theta_j)]^{-1} U(\theta_j)$$

Optimization methods for maximum likelihood

- In many situations, finding the maximum likelihood estimator may be challenging, e.g. it may be on the boundary of Ω
- Likelihood function may also possess multiple stationary points, and optimization techniques are designed to obtain local maxima, so it may not converge to global maxima
- It is important to understand the shape of $\ell(\theta)$ before applying an optimization method

Example 2.1.1

- Suppose that the lifetimes of individuals in some population follow a distribution with density function $f(t)$ and distribution function $F(t)$, and that the lifetimes t_1, \dots, t_n for a random sample of n individuals are observed.
- In format of Eq. 2.1.1, $Data = (t_1, \dots, t_n)$ and

$$Pr(Data) = \prod_{i=1}^n f(t_i) \quad (2.1.2)$$

Parametric approach

- Assume that $f(t)$ has a specific parametric form $f(t; \theta)$
- Likelihood function

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta)$$

- The maximum likelihood estimator $\hat{\theta}$ can be obtained by maximizing $L(\theta)$ or $\ell(\theta)$ and consequently an estimate of $F(t; \hat{\theta})$, the distribution function
- For example, if $T \sim \text{Exp}(\theta)$ then $\hat{\theta} = \bar{t}$ and $F(t; \hat{\theta}) = \exp(-t/\hat{\theta})$

Nonparametric approach

- Assume $F(t)$ is discrete with unspecified probabilities
 $f(t) = F(t) - F(t - 1)$ at the jump points $t = 1, 2, \dots$
- The model parameters are $\mathbf{f} = (f(1), f(2), \dots)$ and the likelihood function is

$$L = \prod_{i=1}^n f(i)$$

- ▶ Restrictions: $f(t) \geq 0 \quad \forall t$ and $\sum_{s=1}^n f(s) = 1$

Nonparametric approach

- The maximum likelihood estimators can be obtained by maximizing the corresponding likelihood function

$$\hat{f}(t) = \frac{1}{n} \sum_{i=1}^n I(t_i = t)$$

- ▶ $I(A)$ is an indicator function

$$I(A) = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

Nonparametric approach

- Estimate of the distribution function $F(t)$

$$\begin{aligned}\hat{F}(t) &= \sum_{t_i \leq t} \hat{f}(t_i) \\ &= \frac{1}{n} \sum_{i=1}^n I(t_i \leq t)\end{aligned}$$

- ▶ $\hat{F}(t) \rightarrow$ *empirical distribution function*

Likelihood for a truncated sample

- Suppose that t_1, \dots, t_n are not from an unrestricted random sample of individuals, but a random sample of those with lifetimes one year or less
 - ▶ No information is available for those whose lifetimes are greater than one year
- The likelihood function for this truncated sample is given by

$$\prod_{i=1}^n Pr(t_i | T_i \leq 1) = \prod_{i=1}^n \frac{f(t_i)}{F(1)}$$

rather than (2.1.2)

Likelihood based inferences

Score test

- Score vector $U(\theta)$ asymptotically follows k -variate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathcal{J}(\theta)$
- Under the null hypothesis $H_0 : \theta = \theta_0$

$$W(\theta_0) = U(\theta_0)' \mathcal{J}(\theta_0)^{-1} U(\theta_0)$$

asymptotically follows $\chi_{(k)}^2$ distribution.

- The statistic $W(\theta_0)$ can also be used to obtain confidence intervals for θ

Likelihood based inferences

Wald test

- The m.l.e. $\hat{\theta}$ follows a k -dimensional normal distribution with mean θ and variance-covariance matrix $\mathcal{J}(\theta)^{-1}$
- In other words, $\sqrt{n}(\hat{\theta} - \theta)$ follows a k -dimensional normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $n\mathcal{J}(\theta)^{-1}$
- Under $H_0 : \theta = \theta_0$

$$(\hat{\theta} - \theta_0)' \mathcal{J}(\theta_0) (\hat{\theta} - \theta_0)$$

asymptotically follows $\chi_{(k)}^2$.

- Since $I(\hat{\theta})/n$ is a consistent estimator of $\mathcal{J}(\theta_0)$, we can replace $\mathcal{J}(\theta_0)$ by $\mathbf{I}(\hat{\theta})$ in the test statistic.

Likelihood based inferences

Likelihood ratio test

- Under $H_0 : \theta = \theta_0$

$$\begin{aligned}\Lambda(\theta_0) &= -2 \log \left[\frac{L(\theta_0)}{L(\hat{\theta})} \right] \\ &= 2\ell(\hat{\theta}) - 2\ell(\theta_0)\end{aligned}$$

asymptotically follows $\chi_{(k)}^2$

Subsection 2

2.2 Right censoring and maximum likelihood

Right censoring and maximum likelihood

- For *right censored data*, only the lower bounds on lifetime are available for some individuals
- Right censored lifetimes are observed for various reasons, such as termination of the study, lost-to-follow-up, etc.
- Contribution to the likelihood function would be different for right censored and complete lifetimes
- Construction of the likelihood function could differ for different types of censoring, such as left censoring, interval censoring, etc.

Right censoring and maximum likelihood

- Let the random variables T_1, \dots, T_n represent the lifetimes of n individuals
 - ▶ Let C_1, \dots, C_n be the corresponding right censoring times
- For the i th individual, we observe $\{(t_i, \delta_i), i = 1, \dots, n\}$
 - ▶ t_i is a sample realization of $X_i = \min\{T_i, C_i\}$
 - ▶ $\delta_i = I(T_i \leq C_i)$ is known as censoring or status indicator, i.e.

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

Right censoring and maximum likelihood

- There are three types of right censoring mechanism
 - ▶ Type I censoring
 - ▶ Independent random censoring
 - ▶ Type II censoring

Type I censoring

- In Type I censoring, potential censoring time C_i is assumed to be fixed for each individual
- Type I censoring often arises when a study is conducted over a specified period of time
 - ▶ For example, if termination of life test on electrical insulation specimens after 180 minutes, then $C_i = 180 \quad \forall i$

Type I censoring

- Likelihood function for the observed Type I censored sample $\{(t_i, \delta_i), i = 1, \dots, n\}$

$$L = \prod_{i=1}^n Pr(X_i = t_i, \delta_i)$$

Type I censoring

- Likelihood function for the observed Type I censored sample $\{(t_i, \delta_i), i = 1, \dots, n\}$

$$L = \prod_{i=1}^n Pr(X_i = t_i, \delta_i)$$

$$L = \prod_{i=1}^n [Pr(X_i = t_i, \delta_i = 1)]^{\delta_i} [Pr(X_i = t_i, \delta_i = 0)]^{1-\delta_i}$$

Type I censoring

- Likelihood function for the observed Type I censored sample $\{(t_i, \delta_i), i = 1, \dots, n\}$

$$L = \prod_{i=1}^n [Pr(X_i = t_i, \delta_i = 1)]^{\delta_i} [Pr(X_i = t_i, \delta_i = 0)]^{1-\delta_i}$$

- We can show that

$$\begin{aligned} Pr(X_i = t_i, \delta_i = 1) & \\ &= Pr(\min\{T_i, C_i\} = t_i, \delta_i = 1) \\ &= Pr(T_i = t_i) \\ &= f(t_i) \end{aligned}$$

$$\begin{aligned} Pr(X_i = t_i, \delta_i = 0) & \\ &= Pr(\min\{T_i, C_i\} = t_i, \delta_i = 0) \\ &= Pr(T_i > C_i, C_i = t_i) \\ &= Pr(T_i > t_i) \\ &= S(t_i+) \end{aligned}$$

Type I censoring

- Likelihood function for the observed Type I censored sample $\{(t_i, \delta_i), i = 1, \dots, n\}$

$$\begin{aligned} L &= \prod_{i=1}^n Pr(X_i = t_i, \delta_i) \\ &= \prod_{i=1}^n [Pr(X_i = t_i, \delta_i = 1)]^{\delta_i} [Pr(X_i = t_i, \delta_i = 0)]^{1-\delta_i} \\ &= \prod_{i=1}^n [f(t_i)]^{\delta_i} [S(t_i+)]^{1-\delta_i} \end{aligned}$$

- ▶ If $S(t)$ is continuous at t_i , then $S(t_i+) = S(t_i)$

Example 2.2.1

- Suppose that lifetimes T_i are independent and follow an exponential distribution with the p.d.f. $f(t) = \lambda e^{-\lambda t}$.
- Let $\{(t_i, \delta_i), i = 1, \dots, n\}$ be a random sample (right censored, Type I) from the exponential distribution.
- Obtain the expression of the likelihood function for the given sample.

Example 2.2.1

- Given

$$f(t) = \lambda e^{-\lambda t} \Rightarrow S(t) = e^{-\lambda t}$$

- The likelihood function

$$\begin{aligned} L &= \prod_{i=1}^n [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i} \\ &= \prod_{i=1}^n [\lambda e^{-\lambda t_i}]^{\delta_i} [e^{-\lambda t_i}]^{1-\delta_i} \\ &= \left[\lambda^{\sum_{i=1}^n \delta_i} \right] \left[e^{-\lambda \sum_{i=1}^n t_i} \right] \\ &= \left[\lambda^r \right] \left[e^{-\lambda \sum_{i=1}^n t_i} \right] \end{aligned}$$

► $r = \sum_{i=1}^n \delta_i$

Independent Random Censoring

- Censoring time C is assumed to be continuous random variable with survivor function $G(t)$ and density function $g(t)$
- Lifetime T is also continuous random variable with survivor function $S(t)$ and density function $f(t)$
- Assumptions:
 - ▶ T and C are independent
 - ▶ $G(t)$ does not depend on any of the parameters of $S(t)$

Independent Random Censoring

- Likelihood function for the observed independent random censored sample $\{(t_i, \delta_i), i = 1, \dots, n\}$

$$\begin{aligned} L &= \prod_{i=1}^n Pr(X_i = t_i, \delta_i) \\ &= \prod_{i=1}^n [Pr(X_i = t_i, \delta_i = 1)]^{\delta_i} [Pr(X_i = t_i, \delta_i = 0)]^{1-\delta_i} \end{aligned}$$

Independent Random Censoring

- We can show

$$\begin{aligned}Pr(X_i = t_i, \delta_i = 1) &= Pr(T_i < C_i, T_i = t_i) \\&= Pr(T_i < C_i | T_i = t_i) Pr(T_i = t_i) \\&= Pr(C_i > t_i) Pr(T_i = t_i) \\&= G(t_i+) f(t_i)\end{aligned}$$

Independent Random Censoring

- We can show

$$\begin{aligned}Pr(X_i = t_i, \delta_i = 0) &= Pr(T_i > C_i, C_i = t_i) \\&= Pr(T_i > C_i | C_i = t_i) Pr(C_i = t_i) \\&= Pr(T_i > t_i) Pr(C_i = t_i) \\&= S(t_i+) g(t_i)\end{aligned}$$

Independent Random Censoring

- Likelihood function for the observed independent random censored sample $\{(t_i, \delta_i), i = 1, \dots, n\}$

$$\begin{aligned}L &= \prod_{i=1}^n Pr(X_i = t_i, \delta_i) \\&= \prod_{i=1}^n [Pr(X_i = t_i, \delta_i = 1)]^{\delta_i} [Pr(X_i = t_i, \delta_i = 0)]^{1-\delta_i} \\&= \prod_{i=1}^n [G(t_i+) f(t_i)]^{\delta_i} [S(t_i+) g(t_i)]^{1-\delta_i}\end{aligned}$$

- Since $G(t)$ and $g(t)$ don't involve any parameters of $f(t)$

$$L = \prod_{i=1}^n [f(t_i)]^{\delta_i} [S(t_i+)]^{1-\delta_i}$$

Type II censoring

- In Type II censoring, lifetest starts with n units and it stops when r number of failures are observed
 - ▶ So r smallest lifetimes $t_{(1)} \leq \dots \leq t_{(r)}$ in a random sample of n are observed
 - ▶ $r \rightarrow$ a specified integer that lies between 1 and n
 - ▶ The remaining $(n - r)$ units are considered as censored at the time $t_{(r)}$

Type II censoring

- For Type II censoring, the likelihood function is the probability of observing r smallest lifetimes $t_{(1)} \leq \dots \leq t_{(r)}$ out of n lifetimes

$$L = \binom{n}{r} \left\{ \prod_{i=1}^r f(t_{(i)}) \right\} \left[S(t_{(r)}) \right]^{n-r}$$

- This expression is similar to the expression obtained for Type I and random independent censoring with all the censoring times equal to $t_{(r)}$

Example 2.2.2

- Let $t_{(1)}, \dots, t_{(r)}$ be a Type II random sample of n lifetimes T_1, \dots, T_n , where T_i follows an exponential distribution with rate λ
- The likelihood function

$$\begin{aligned}L &= \binom{n}{r} \left\{ \prod_{i=1}^r f(t_{(i)}) \right\} \left[S(t_{(r)}) \right]^{n-r} \\ &= \binom{n}{r} \left\{ \prod_{i=1}^r \lambda e^{-\lambda t_{(i)}} \right\} \left[e^{-\lambda t_{(r)}} \right]^{n-r} \\ &= \binom{n}{r} \lambda^r e^{-\lambda W}\end{aligned}$$

► $W = \sum_{i=1}^r t_{(i)} + (n-r)t_{(r)}$

A general formulation of right censoring

- The censoring process is often not any of the types discussed so far, and may be sufficiently complicated to make modeling it impossible.
- For example, a decision to terminate a life test or clinical trial at time t , or to withdraw certain individuals, might be based on failure information prior to time t ,
- Fortunately it can be shown that under rather general conditions the observed likelihood is of the form (2.2.3) and can be used in the normal way to make inferences about the lifetime distribution under study.

$$L = \prod_{i=1}^n [f(t_i)]^{\delta_i} [S(t_i+)]^{1-\delta_i}$$

- Read Section 2.2.2 of the textbook for details.

Subsection 3

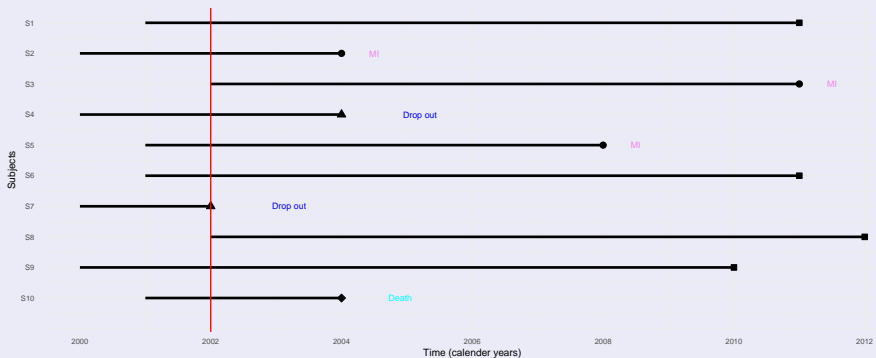
A Hypothetical Study

A hypothetical study

- A small prospective study was run, where 10 participants were recruited to follow
- The event of interest was **the development of myocardial infarction** (MI, or heart attack) over a period of 10 years (follow-up period)
- Participants are recruited into the study over a period of two years and were then followed for up to 10 years

A hypothetical study

Study in calendar years

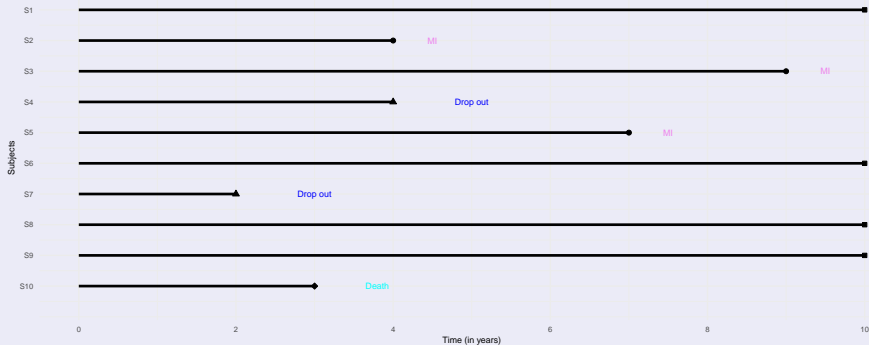


A hypothetical study

- {S2, S3, S5} → experienced MI
- {S4, S7} → dropped out
- {S10} → died from other causes
- {S1, S6, S8, S9} → completed 10-year follow-up without MI

A hypothetical study

Study in years



A hypothetical study

subject	t	outcome	status	δ
S1	10	alive	censored	0
S2	4	MI	failure	1
S3	9	MI	failure	1
S4	4	drop-out	censored	0
S5	7	MI	failure	1
S6	10	alive	censored	0
S7	2	drop-out	censored	0
S8	10	alive	censored	0
S9	10	alive	censored	0
S10	3	death	censored	0

A hypothetical study

- Times for the subjects who did not experience MI by the end of 10-year follow-up or dropped-out or died from causes not related to MI are known as **censored times**
- Time-to-MI (time to the event of interest) are known as **failure time** (or survival time)
- For survival data, **the pair (time, status)**, (t, δ) , is considered as the response
- A sample of survival data can also be expressed as following, where $+$ sign indicates censored observations

$10^+, 4, 9, 4^+, 7, 10^+, 2^+, 10^+, 10^+, 3^+$

Subsection 4

2.3 Other type of incomplete data