Biostat/Stat 576

Statistical Methods for Survival Analysis

Lecture 3

April 7, 2009

Chapter 2. Parametric methods

1. Parametric distributions

2. Likelihood functions and MLE

- Parameter estimation: maximum likelihood estimation (MLE)

  - likelihood function:

$$
\begin{aligned}
\mathscr{L}(\theta) &= \prod_{i=1}^{n} f(x_i \mid Z_i;\theta)^{\delta_i} S(x_i \mid Z_i;\theta)^{1-\delta_i} \\
&= \prod_{i=1}^{n} \lambda(x_i \mid Z_i;\theta)^{\delta_i} S(x_i \mid Z_i;\theta)
\end{aligned}
$$

  - parameter estimation: $\widehat{\theta} = \arg\max_\theta \mathscr{L}(\theta) = \arg\max_\theta l(\theta)$

$$
\begin{aligned}
l(\theta) &= \log \mathscr{L}(\theta) \\
&= \sum_{i=1}^{n} \left[\delta_i \log \lambda(x_i \mid Z_i;\theta) + \log S(x_i \mid Z_i;\theta)\right] \\
&= \sum_{i=1}^{n} \left[\delta_i \log \lambda(x_i \mid Z_i;\theta) - \Lambda(x_i \mid Z_i;\theta)\right] \\
&= \sum_{i=1}^{n} \left[\delta_i \log \lambda(x_i \mid Z_i;\theta) - \int_0^{x_i} \lambda(u \mid Z_i;\theta)du\right]
\end{aligned}
$$

– score function

$$l'(\theta) = \sum_{i=1}^{n} \left[\delta_i \frac{\lambda'_\theta(x_i \mid Z_i; \theta)}{\lambda(x_i \mid Z_i; \theta)} - \int_0^\infty I(x_i \geq u) \lambda'_\theta(u \mid Z_i; \theta) du \right]$$

$$= \sum_{i=1}^{n} \left[\int_0^\infty \frac{\lambda'_\theta(u \mid Z_i; \theta)}{\lambda(u \mid Z_i; \theta)} dN_i(u) - \int_0^\infty Y_i(u) \lambda'_\theta(u \mid Z_i; \theta) du \right]$$

$$= \sum_{i=1}^{n} \int_0^\infty \frac{\lambda'_\theta(u \mid Z_i; \theta)}{\lambda(u \mid Z_i; \theta)} [dN_i(u) - Y_i(u) \lambda(u \mid Z_i; \theta) du]$$

$$\Longrightarrow \widehat{\theta}_{\text{MLE}} = \text{arg}_\theta l'(\theta) = 0$$

- Calculation of MLE's error bound
  - θ_0 is the true value of parameter θ
  - $n^{1/2}(\widehat{\theta}_{\mathsf{MLE}} - \theta_0) \sim_A \mathscr{N}(0, nI^{-1})$; $I = -El''(\theta_0)$ is the Fisher information
  - estimate asymptotic standard error by $s.e. = [\widehat{I}^{-1}(\widehat{\theta}_{\mathsf{MLE}})]^{1/2}$
  - 95% confidence interval: $\widehat{\theta}_{\mathsf{MLE}} \pm 1.96 \times s.e.$

- Example: One-sample estimation in $T_i \sim \text{exp}(\lambda)$
  - data observed: $\{(x_i, \delta_i); i = 1, 2, \ldots, n\}$
  - likelihood function: $\mathscr{L}(\lambda) = \prod_i (\lambda e^{-\lambda x_i})^{\delta_i} (e^{-\lambda x_i})^{1-\delta_i}$
  - score function: $l'(\lambda) = \sum_i (\delta_i / \lambda - x_i)$
  - MLE: $\widehat{\lambda} = \sum_i \delta_i / \sum_i x_i = d / \sum_i x_i$; $(d = \sum_i d_i)$
  - variance: $l''(\lambda) = -d/\lambda^2 \Rightarrow \text{var}(\widehat{\lambda}) = [-El''(\lambda)]^{-1} = \lambda^2/d$
  - $\widehat{s.e.}(\widehat{\lambda}) = \widehat{\lambda}/\sqrt{d}$
  - an application: estimate mean or median survival in exponential model

- Asymptotics on MLE
  - score function: $U(\theta) = l'(\theta) = [\log \mathscr{L}(\theta)]'$
    - unbiased estimating equation: $EU(\theta) = 0$
    - efficient: $I(\theta) = EU(\theta)^2 = El''(\theta)$
  - MLE $\widehat{\theta}$: $U(\widehat{\theta}) = 0$
  - Delta-method

$$
\begin{aligned}
&U(\widehat{\theta}) - U(\theta_0) \approx U'(\theta_0)(\widehat{\theta} - \theta_0) \\
\Rightarrow &n^{1/2}(\widehat{\theta} - \theta_0) = -[U'(\theta_0)/n]^{-1} \cdot n^{-1/2} U(\theta_0)
\end{aligned}
$$

    - $n^{-1}U'(\theta_0)$ is consistent, goes to $I(\theta_0)$
    - $n^{-1/2}U(\theta_0)$ is asymptotically normal $\mathscr{N}(0, nI(\theta_0))$
  - $\widehat{\theta} - \theta_0 \sim \mathscr{N}(0, I(\theta_0)^{-1})$

- Hypothesis testing of $H_0 : \theta = \theta_0$, $p = \text{dim}(\theta)$
  - score test: $U(\theta_0)^{\mathsf{T}} I(\theta_0)^{-1} U(\theta_0) \sim \chi^2_p$
  - Wald's test: $(\widehat{\theta} - \theta_0)^{\mathsf{T}} I(\theta_0)(\widehat{\theta} - \theta_0) \sim \chi^2_p$
  - likelihood ratio test (LRT): $-2\log[\mathscr{L}(\theta_0)/\mathscr{L}(\widehat{\theta})] \sim \chi^2_p$

- Hypothesis testing on a subset of parameters $\mathrm{H}_0 : \theta_1 = \theta_{10}$
  - $p_1 = \dim(\theta_1) < p$
  - $\theta = (\theta_1, \theta_2)$
  - example: test adjusted treatment effect (θ_1) for covariates (θ_2)
  - score test
    * $U(\theta_1, \theta_2) = [U_1(\theta_1, \theta_2), U_2(\theta_1, \theta_2)]$
    * solve $U_2(\theta_{10}, \theta_2) = 0 \Rightarrow \widehat{\theta}_2 = \widehat{\theta}_2(\theta_{10})$
    * define $\widehat{U}_1(\theta_{10}) = U_1(\theta_{10}, \widehat{\theta}_2(\theta_{10}))$
    * define $\widehat{I}_{11}(\theta_{10}) = I_{11}(\theta_{10}, \widehat{\theta}_2(\theta_{10}))$
    * $\widehat{U}_1(\theta_{10})^{\mathsf{T}} \widehat{I}_{11}(\theta_{10}) \widehat{U}_1(\theta_{10}) \sim \chi^2_{p_1}$

- likelihood ratio test
  - $-2\log[\mathscr{L}(\theta_{10}, \widehat{\theta}_2(\theta_{10}))/\mathscr{L}(\widehat{\theta})] \sim \chi^2_{p1}$

Chapter 3. Kaplan-Meier Curve and NPMLE

1. Identifiablity: noninformative versus independent censoring

2. Nonparametric estimation: Nelson-Aalen and Kaplan-Meier estimators

3. Asymptotic theory

1. Identifiability

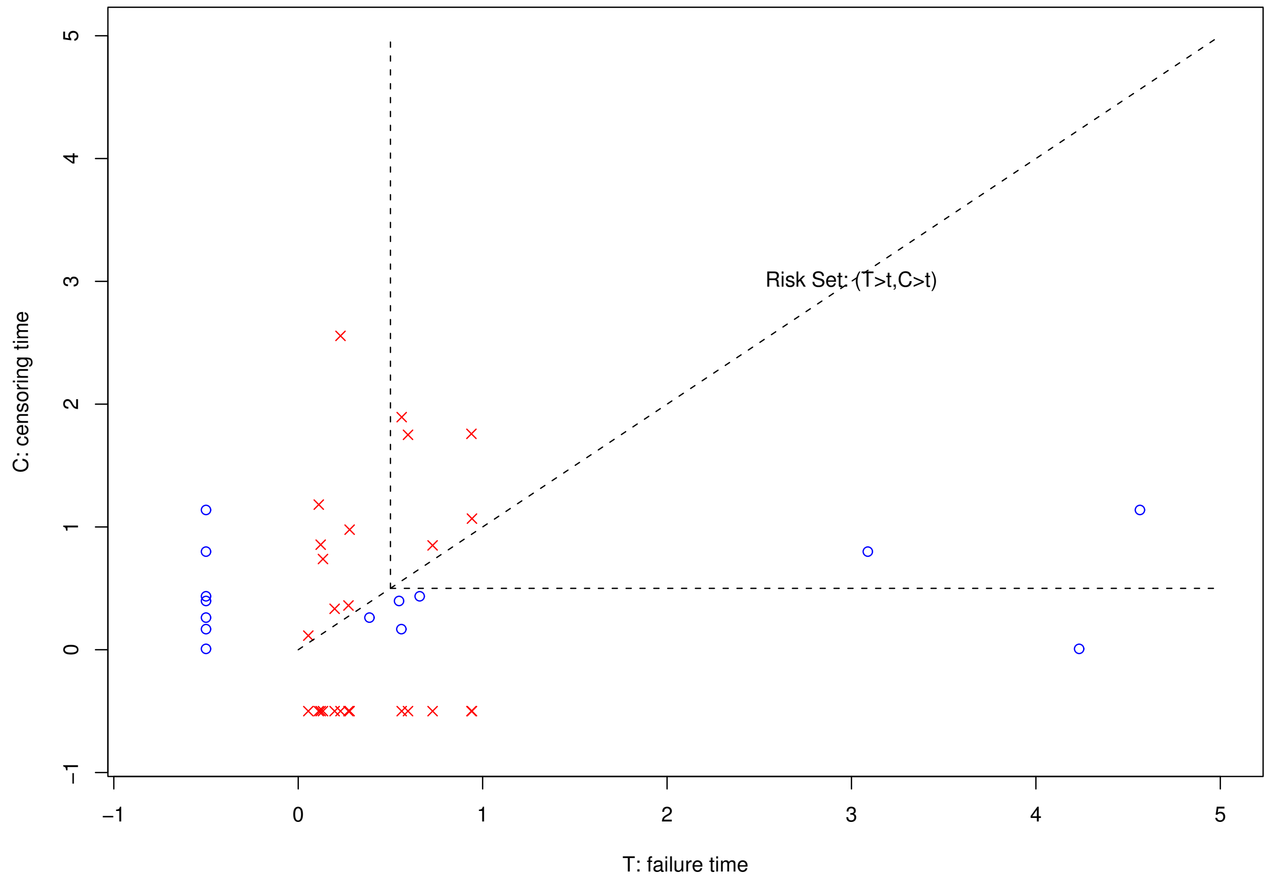
- Goal
  - Estimate the population distribution from observed data
  - Specifically, how to estimate $S(t) = \Pr\{T \geq t\} = E[I(T \geq t)]$?

- Observed data
  1. no censoring: $\{T_i; i = 1, 2, \ldots, n\}$ randomly drawn from population
     - empirical/moment estimator

$$\widehat{S}(t) = \frac{1}{n} \sum_i I(T_i \geq t)$$

  2. censoring: $\{X_i = \min(T_i, C_i), \Delta_i = I(T_i \leq C_i)\}$
     - what is a good estimator?

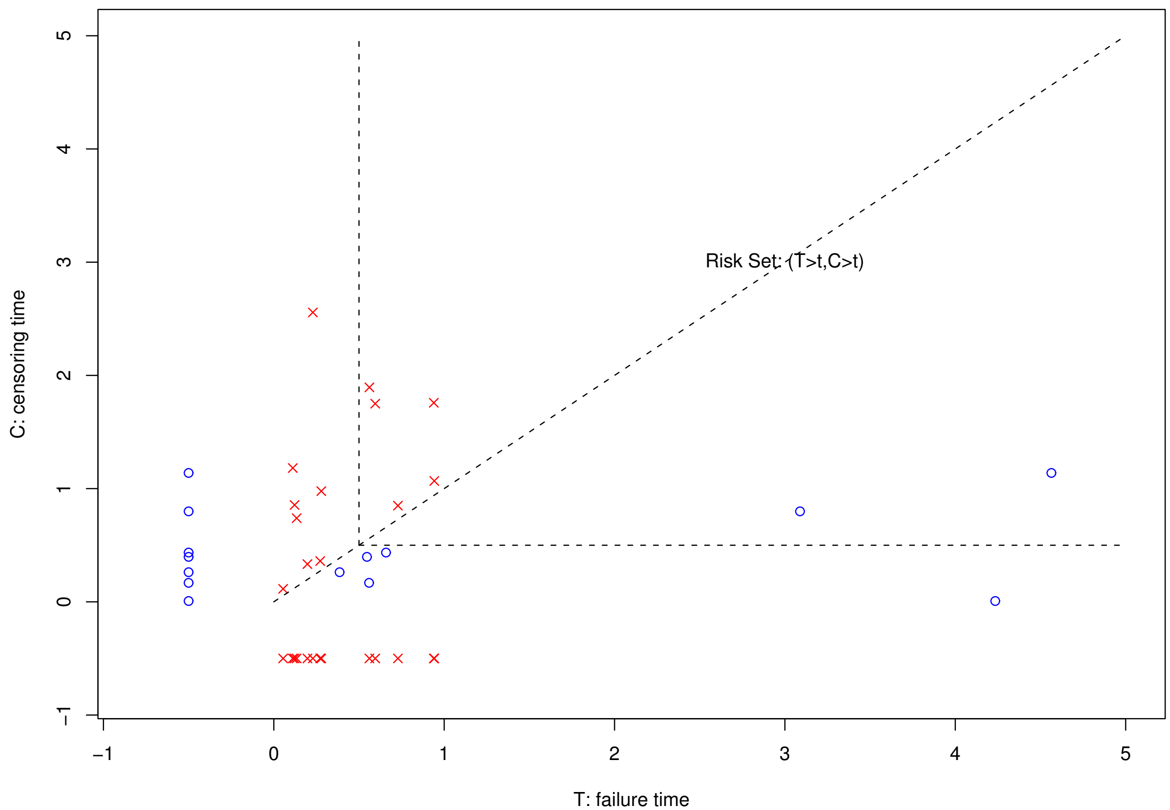
- What do we observe?



- $(T, C) \Rightarrow (X, \Delta)$
- $(X, \Delta) \Rightarrow (T, C)$?

- Conceptually

  1. $S(t) = \text{Pr}\{T \geq t\} = \int_0^\infty \text{Pr}\{T \geq t, C = u\}du$
  2. but for any given u, $\text{Pr}\{T \geq t, C = u\}$ is nowhere identifiable for any t



- What can we estimate?
  1. At-risk probability: $\Pr\{X \geq t\} = E[I(X \geq t)] = E[Y(t)]$
  2. sub-distribution functions

	$\Delta = 0$	$\Delta = 1$
$X \leq t$	$F_0^s(t) = \Pr\{X \leq t, \Delta = 0\}$	$F_1^s(t) = \Pr\{X \leq t, \Delta = 1\}$
$X \geq t$	$S_0^s(t) = \Pr\{X \geq t, \Delta = 0\}$	$S_1^s(t) = \Pr\{X \geq t, \Delta = 1\}$
	$\lambda_0^s(t) = dF_0^s(t)/S(t)dt$	$\lambda_1^s(t) = dF_1^s(t)/S(t)dt$

- noninformative censoring
  - $\lambda_1^s(t) = \lambda(t) \Rightarrow$

$$\exp\left\{-\int_0^t S(u)^{-1} dF_1^s(t)\right\} = \exp\left\{-\int_0^t \lambda_1^s(u) du\right\}$$
$$= \exp\left\{-\int_0^t \lambda(u) du\right\} = S(t)$$

- independent censoring
  - whenever $\Pr\{X > t\} > 0$

$$\begin{aligned}\lambda_1^s(t)dt &= \Pr\{t \le X \le t + dt, \Delta = 1 \mid X \ge t\} \\ &= \Pr\{t \le T \le t + dt, T \le C \mid T \ge t, C \ge t\} \\ &= \Pr\{t \le T \le t + dt, t \le C \mid T \ge t, C \ge t\} \\ &= \lambda_1(t)dt\end{aligned}$$

  - independent censoring implies noninformative censoring

- Example 1
  - (T, C)'s joint distribution

$$\Pr\{T > t, C > s\} = \exp(-\lambda t - \mu s - \theta ts), t \geq 0, s \geq 0$$

  - $\lambda_T^s(t) = \lambda + \theta t$
  - $\lambda_T(t) = \lambda$, $S_T(t) = \exp(-\lambda t)$
  - $\theta = 0$: (T, C) are independent

- Example 2
  - (T, C)'s joint distribution

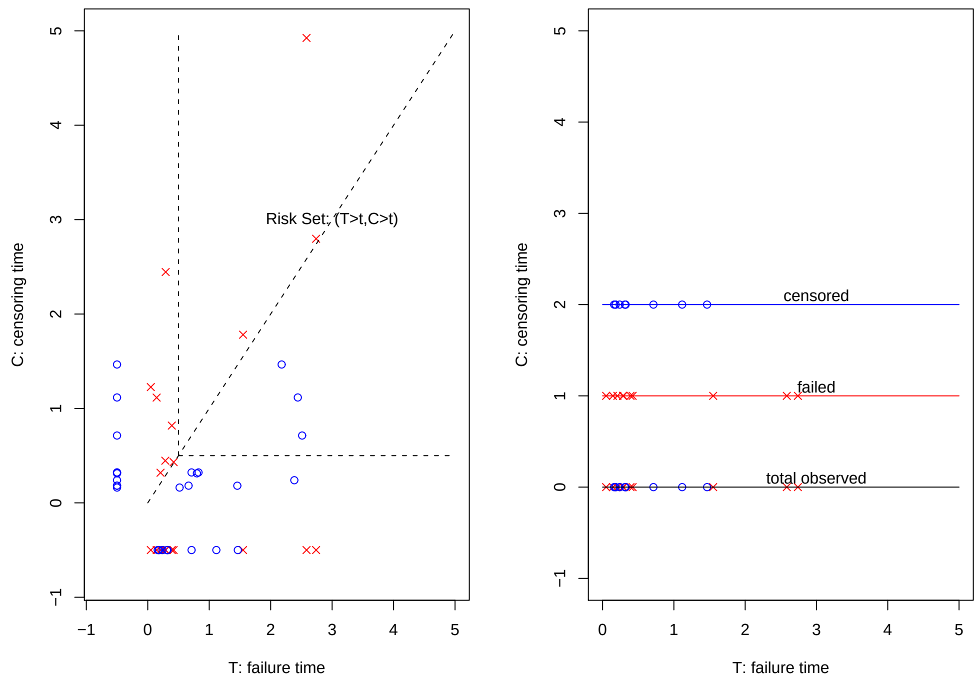
$$\Pr\{T > t, C > s\} = \exp[-\lambda t - \mu s - \theta(t^2 + s^2)/2], t \geq 0, s \geq 0$$

  - $\lambda_T^s(t) = \lambda + \theta t$
  - $\lambda_T(t) = \lambda + \theta t$, $S_T(t) = \exp(-\lambda t - \theta t^2/2)$
  - noninformative censoring but not independent for $\theta \neq 0$

- noninformative censoring is not verifiable if completely nonparametric

2. Nonparametric estimation

- Observed data



- At failure times $t_1 < t_2 < \ldots < t_k$, $t_0 = 0$ and $t_{k+1} = \infty$:
  1. n_j at-risk immediately prior to t_j
  2. d_j failed at t_j
  3. m_j cencored between $[t_j, t_{j+1})$

- Likelihood contribution

  1. failure t_j: $\text{Pr}\{T = t_j\} = \text{Pr}\{T \geq t_j\} - \text{Pr}\{T > t_j\} = S(t_j-) - S(t_j)$
  2. independent censoring t_{jl}: $\text{Pr}\{T > t_{jl}\} = S(t_{jl})$, $l = 1, \dots, m_j$
  3. total likelihood

$$\mathscr{L}(S) = \prod_{j=0}^{k} \left\{ [S(t_j-) - S(t_j)]^{d_j} \prod_{l=1}^{m_j} S(t_{jl}) \right\}$$

- Nonparametric maximum likelihood estimation (NPMLE)

  1. maximize $\mathscr{L}(S)$ with respect to S: $\widehat{S} = \arg\max_S \mathscr{L}(S)$
  2. $\widehat{S}(t)$ has to be discontinuous at t_j; otherwise $S(t_j) - S(t_j-) = 0$
  3. $t_{jl} \geq t_j$, $S(t_{jl}) \leq S(t_j) \Rightarrow \max \widehat{S}(t_{jl}) = \widehat{S}(t_j)$
  4. $\widehat{S}$ should be a step function with jumps at t_j

- How to choose jump sizes to maximize $\mathscr{L}(S)$?

1.

$$
\begin{aligned}
S(t_j) &= \Pr\{T > t_j\} = \Pr\{T \geq t_{j+1}\} \\
&= \frac{\Pr\{T \geq t_{j+1}\}}{\Pr\{T \geq t_j\}} \cdot \frac{\Pr\{T \geq t_j\}}{\Pr\{T \geq t_{j-1}\}} \cdot \cdots \cdot \frac{\Pr\{T \geq t_2\}}{\Pr\{T \geq t_1\}} \cdot \Pr\{T \geq t_1\} \\
&= \left[1 - \frac{\Pr\{T = t_j\}}{\Pr\{T \geq t_j\}}\right] \cdots \cdots \left[1 - \frac{\Pr\{T = t_1\}}{\Pr\{T \geq t_1\}}\right] \\
&= \prod_{l=1}^{j}(1 - \lambda_l)
\end{aligned}
$$

2. $S(t_j-) = \prod_{l=1}^{j-1}(1-\lambda_l) \Rightarrow [S(t_j-) - S(t_j)]^{d_j} = \lambda_j^{d_j} \prod_{l=1}^{j-1}(1-\lambda_l)^{d_j}$

3.

$$
\mathscr{L}(S) = \prod_{j}^{k} \left[\lambda_j^{d_j} \prod_{l=1}^{j-1}(1-\lambda_l)^{d_j} \prod_{l=1}^{j}(1-\lambda_l)^{m_j} \right] = \prod_{j}^{k} \lambda_j^{d_j}(1-\lambda_j)^{n_j - d_j}
$$

4. $\widehat{\lambda}_j = d_j/n_j \Rightarrow \widehat{S}(t) = \prod_{t_j \leq t}(1 - d_j/n_j)$ is the NPMLE.

- Error bound of Kaplan-Meier estimator

  1. $\log \widehat{S}(t) = \sum_{t_j \le t} \log(1 - \widehat{\lambda}_j)$
  2. $\mathrm{var}[\log \widehat{S}(t)] = \sum_{t_j \le t} (1 - \widehat{\lambda}_j)^{-2} \mathrm{var}(1 - \widehat{\lambda}_j) = \sum_{t_j \le t} d_j / [n_j(n_j - d_j)]$
  3. $\mathrm{var}[\widehat{S}(t)] = \widehat{S}(t)^2 \sum_{t_j \le t} d_j / [n_j(n_j - d_j)]$

- Example
  - Carcinogenesis data, p. 2
  - Kaplan-Meier estimate, p. 16-17
  - Kalbfleisch & Prentice

- Kaplan-Meier estimator

$$\widehat{S}(t) = \prod_{t_j \le t} \left(1 - \frac{d_j}{n_j}\right) \approx \prod_{t_j \le t} e^{-\frac{d_j}{n_j}} = e^{-\sum_{t_j \le t} \frac{d_j}{n_j}}$$

- Assume T is continuous $\Rightarrow d_j = 1$ mostly

  1. $d_j = N(t_j) - N(t_j-) = dN(t_j)$
  2. $n_j = Y(t_j) > 0$
  3. $\sum_{t_j \le t} \frac{d_j}{n_j} = \int_{u \le t} I(Y(u) > 0) dN(u)/Y(u)$

- Nelson-Aalen estimator

$$\widehat{\Lambda}(t) = -\log \widehat{S}(t) = \int_0^t \frac{I(Y(u) > 0) dN(u)}{Y(u)}$$