Biostat/Stat 576

Statistical Methods for Survival Analysis

Lecture 2

April 2, 2009

Chapter 1. Introduction

1. A brief history of time-to-event analysis
2. Time-to-event and censoring
3. Life tables
4. Counting processes and martingales

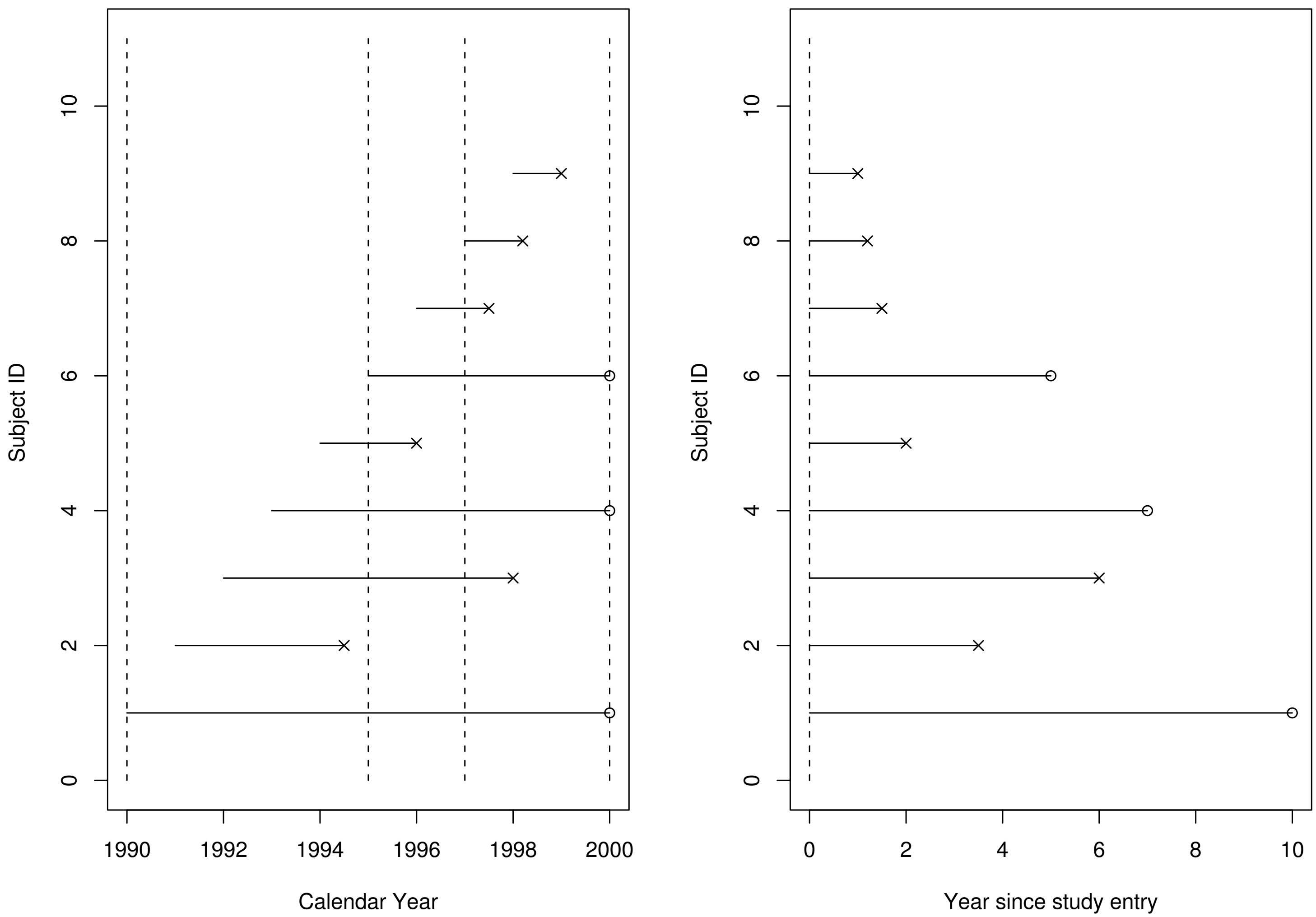
4. Counting Processes and Martingales

- T: time-to-event, failure time, survival time, nonnegative random variable
- Characterization
  1. $f(t)$: density function
  2. $F(t) = \int_0^t f(u)du$: (cumulative) distribution function, (cumulative) risk
  3. $S(t) = \bar{F}(t) = 1 - F(t)$: survival function
  4. $\Lambda(t) = -\log S(t)$: (cumulative) hazard function, (cumulative) risk
  5. $\lambda(t) = \Lambda'(t)$: hazard function, (instantaneous) risk
  6. $m(t) = E(T - t \mid T \geq t)$: mean residual life function

- Center piece: hazard function

1. $\lambda(t) = \lim_{\Delta t \to 0} \Pr\{t \le T \le t + \Delta t \mid T \ge t\}/\Delta t$
2. $\lambda(t)\Delta t \approx \Pr\{t \le T \le t + \Delta t \mid T \ge t\}$
3. $\lambda(t) = f(t)/S(t) = -S'(t)/S(t) = [-\log S(t)]'_t$
4. $\lambda(t) = [1 + m'(t)]/m(t)$

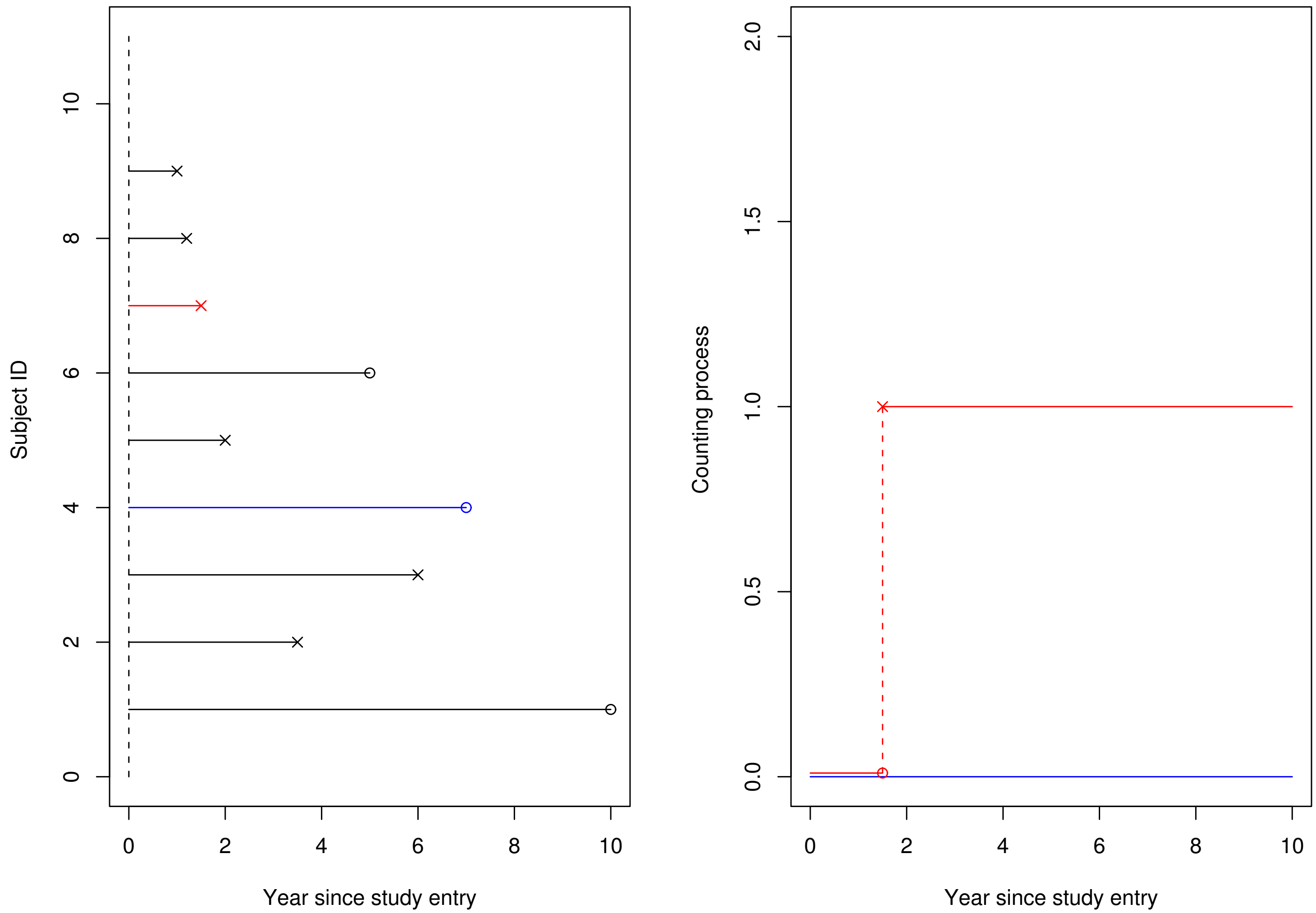
- Typical observed right-censored data



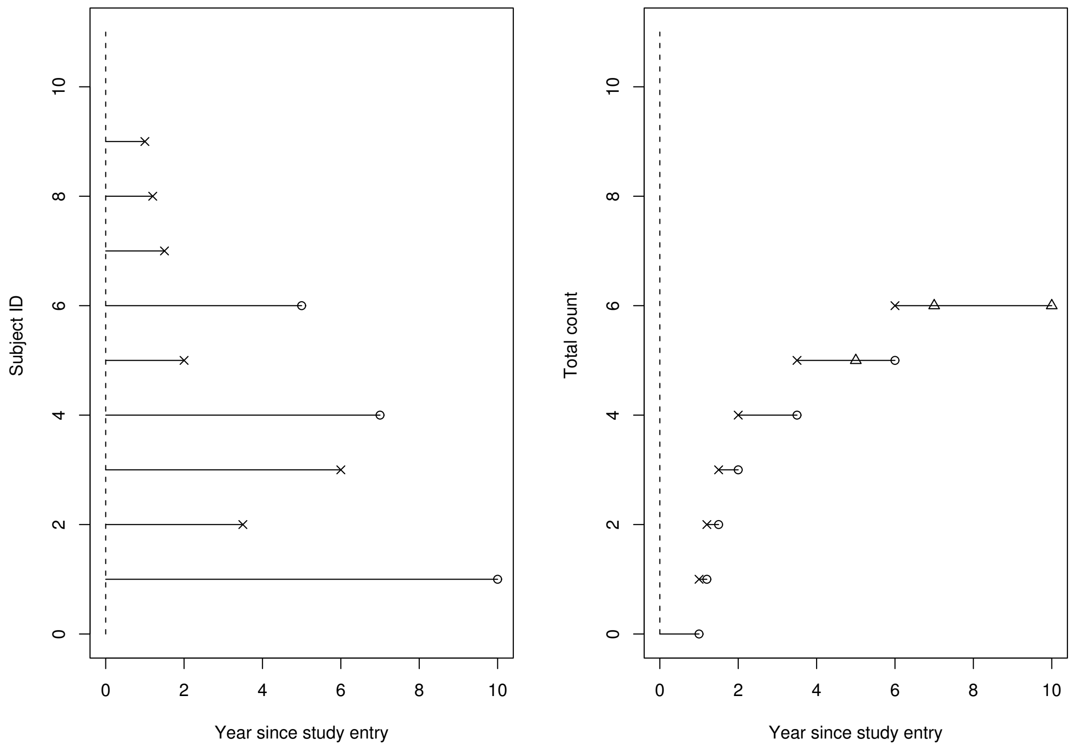
- Data formulation
  1. T: time-to-event
  2. C: censoring time, usually right-censored assumed
  3. $X = \min(T, C)$: (censored) survival time
  4. $\Delta = I(T \leq C)$: event/censoring indicator
  5. Z or $Z(t)$: (time-varying) covariates
  6. $i = 1, 2, \ldots, n$: subjects
  7. $\{(X_i, \Delta_i, Z_i), i = 1, 2, \ldots, n\}$: collected data
- Major goals of survival analysis
  1. one-sample: identify T's characterization
  2. two-sample: compare T's characterizations
  3. regression: identify association between Z and T's characterization

- Counting processes

  1. $N_i(t) = I(X_i \le t, \Delta_i = 1)$: count the event prior to time t



2. $N(t) = \sum_i N_i(t)$: total number of events prior to time t



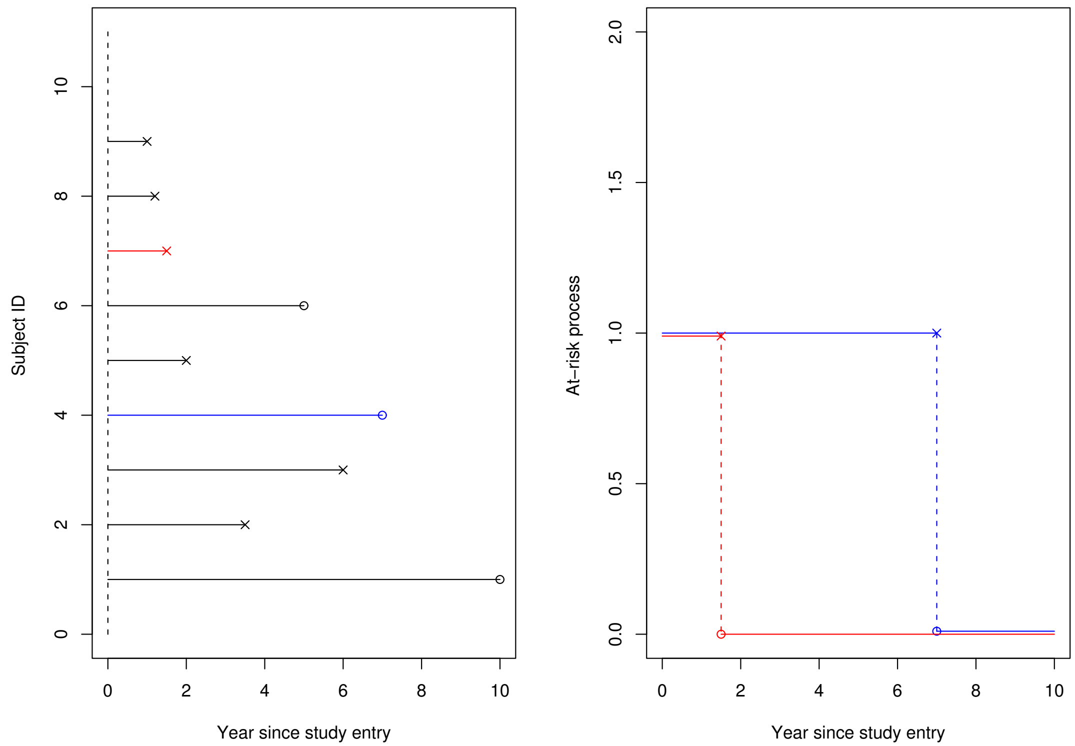
(a) stochastic processes

(b) non-decreasing sample paths

(c) step function taking jumps of size one

(d) cadlag: right-continuous with left-hand limits (*continué à droit, limité à gauche*)

- At-risk process: $Y_i(t) = I(X_i \geq t)$



1. $Y(t) = \sum_i Y_i(t)$: total number of subjects at risk
2. right-continuous with right-hand limit
3. “predictable”

- Steiltjes integral: $\int_0^t f(u)dG(u)$
    1. $G(\cdot)$: $[0, \infty) \mapsto \mathscr{R}$
    2. right-continuous, increasing and differentiable with derivative $g(\cdot)$ except at a countable set of points $\{0 < t_0 < t_1 < \ldots\}$
    3. $\int_0^t f(u)dG(u) = \sum_{t_i \le t} \int_{t_{i-1}}^{t_i} f(u)g(u)du + \sum_{t_i \le t} f(t_i)[G(t_i) - G(t_i-)]$

- Stochastic integrals: Steiltjes integral of one stochastic process with respect to another

- Example

  1. Data: $3 = (3, 1), 4^+ = (4, 0), 8 = (8, 1), 9^+ = (9, 0), 11 = (11, 1)$

  2. Nelson-Aalen estimator

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

t	$Y(t)$	$dN(t)$	$\widehat{\Lambda}(t)$
$[0, 3)$	5	0	0
$[3]$	5	1	1/5
$(3, 4]$	4	0	1/5+0/4
$(4, 8)$	3	0	1/5+0/4+0/3
$[8]$	3	1	1/5+1/3
$(8, 9]$	2	0	1/5+1/3+0/2
$(9, 11)$	1	0	1/5+1/3+0/2+0/1
$[11]$	1	1	1/5+1/3+1/1
$(11, \infty)$	0	0	1/5+1/3+1/1

- What’s so SPECIAL about time?
  1. progressive: data accumulation
  2. irreversible: uni-direction
  3. when things have happened, there is nothing we can do about them but learn from them!

- Filtration: $\mathscr{F}_t$
  1. entire historical information up to t
  2. data collected: $\{(N_i(t), Y_i(t), Z_i); i = 1, 2, \ldots, n\}$
  3. $\mathscr{F}_t = \sigma\{(N_i(u), Y_i(u), Z_i); u \leq t, i = 1, 2, \ldots, n\}$
  4. $\mathscr{F}_{t-}$: information accumulated right before t

- Martingale
  1. $\{M(t); t \geq 0\}$ is a stochastic process
  2. for any $s < t$, $E[M(t) \mid \mathscr{F}_s] = M(s)$
  3. $E[M(t) \mid \mathscr{F}_{t-}] = M(t-) \Longleftrightarrow E[dM(t) \mid \mathscr{F}_{t-}] = 0$

- Example: calculate $E[dN(t) \mid \mathscr{F}_{t-}]$

1. $dN(t) = 0$ or 1
2. $E[dN(t) \mid \mathscr{F}_{t-}] = \Pr\{dN(t) = 1 \mid \mathscr{F}_{t-}\}$
3. if $Y(t) = 0 \Longrightarrow Y(t-) = 0$, then $\Pr\{dN(t) = 1 \mid \mathscr{F}_{t-}\} = 0$
4. if $Y(t) = 1 \Longrightarrow Y(t-) = 1$, then

$$\Pr\{dN(t) = 1 \mid \mathscr{F}_{t-}\} = \Pr\{t \le X \le t + dt, \Delta = 1 \mid X \ge t\}$$

$X = \min(T, C) \ge t \Leftrightarrow T \ge t, C \ge t$, if T and C are independent, then

$$\Pr\{dN(t) = 1 \mid \mathscr{F}_{t-}\} = \Pr\{t \le T \le t + dt \mid T \ge t\} = \lambda(t)dt$$

5. $E[dN(t) \mid \mathscr{F}_{t-}] = Y(t)\lambda(t)dt$
6. $M(t) = N(t) - \int_0^t Y(u)\lambda(u)du$ is a martingale

Chapter 2. Parametric methods

1. Parametric distributions
2. Likelihood functions and MLE

1. Parametric distributions

- Density function: $f(t) = \text{lim}_{\Delta t \to 0} \text{Pr}\{t \le T \le T + \Delta t\}$
  1. $\int_0^\infty f(u)du = 1$
  2. $F(t) = \int_0^t f(u)du$, $S(t) = \int_t^\infty f(u)du$
  3. $\mu(t) = \int_0^t S(u)du$, $\mu = ET = \int_0^\infty S(u)du$
  4. Denominator: all the subjects

- Hazard function: $\lambda(t) = \text{lim}_{\Delta t \to 0} \text{Pr}\{t \le T \le T + \Delta t \mid T \ge t\}/\Delta t$
  1. $\lambda(t) = f(t)/S(t)$
  2. Denominator: subjects survival up to t

- Mean residual life function: $m(t) = E(T - t \mid T \ge t)$
  1. Residual time: $R_T(t) = (T - t)I(T \ge t) = (T - t)^+ = (T - t) \vee 0$
  2. $m(t) = \int_t^\infty S(u)du/S(t)$

- Example: exponential distribution
  1. Density function: $f(t) = \lambda e^{-\lambda t}$, $S(t) = e^{-\lambda t}$
  2. Constant hazard function: $\lambda(t) = \lambda$, $\Lambda(t) = \lambda t$
  3. Constant mean residual life: $m(t) = \mu = 1/\lambda$
  4. Lack of memory: $f_{R_T(t)}(u) = f_T(u)$
  5. Coefficient of variation: s.d./mean=1, for judging relative dispersion
  6. For any T, $\Lambda(T)$ is exponential

- Other parametric distributions: $\{f(t;\theta);\theta \in \Theta\}$
  1. Gamma
  2. Weibull
  3. Log normal
  4. log logistic
  5. Ref: Kalbfleisch & Prentice, §§2.2.1-2.2.8

2. Likelihood functions and MLE

- Statisitcal inferences
  1. dimension reduction: summarizing data by models with parameters
  2. parameter estimation: identifying parameters and providing error bound
  3. estimation efficiency: utilizing maximum information
  4. model checking: assessing model adequacy
  5. model-based prediction

- Likelihood inferences
  1. choose a parametric family of distributions: $f(t \mid Z_i; \theta)$
  2. if no censoring, $\{(T_i, Z_i); i = 1, 2, \ldots, n\}$
     - likelihood function based on observed $T_i = t_i$ is

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

3. if censored, survival data become $\{(X_i, \Delta_i, Z_i); i = 1, 2, \ldots, n\}$
   - likelihood contribution is $\Pr\{X_i = x_i, \Delta_i = \delta_i, Z_i = z_i\}$
   - $\Delta_i = 1 \Longrightarrow X_i = t_i$, likelihood contribution is $f_T(t_i \mid Z_i; \theta) S_C(t_i \mid Z_i)$
   - $\Delta_i = 0 \Longrightarrow X_i = c_i$, likelihood contribution is $f_C(c_i \mid Z_i) S_T(c_i \mid Z_i; \theta)$
   - likelihood function based on observed data is then

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

4. What are the key assumptions?
   - Z_i do not have information about θ
   - conditional independence: given Z_i, T_i and C_i are independent
   - examples
     * clinical trials
     * Z_i: treatment assignment
     * C_i: censoring due to side-effects

5. 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: $\widehat{\theta} = \arg\max_\theta \mathscr{L}(\theta) = \arg\max_\theta l(\theta)$

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

– score function

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

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

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

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