

## Summary on One-sample Nonparametric Estimation

- Two estimators
  1. Kaplan-Meier estimator
  2. Nelson-Aalen estimator
- Connection with life-tables
- Asymptotics
  1. counting processes & martingales
  2. martingale central limit theorem
  3. functional Delta method
- Yet to be covered:
  1. small-sample properties: (Fleming & Harrington)
  2. optimality: Wellner (1982, *Ann. Stat.*)

## Chapter 4. Comparing Survival Curves

1. Log-rank statistic
2. weighted Log-rank statistics
3. Sample size calculation

## 1. Log-rank statistic

- Goal: test equality of survival functions
- Review on Mantel-Haenszel method
  - study association between  $Y = 0/1$  and  $X = 1/2$  stratified by  $Z = 0/1$
  - multiple  $2 \times 2$  tables

$Z$	$X$	$Y = 1$	$Y = 0$
$Z = 0$	$X = 1$	$O_1(E_1)$	
	$X = 2$		
$Z = 1$	$X = 1$	$O_2(E_2)$	
	$X = 2$		

- prototype M-H test statistic

$$\frac{\{\sum(O - E)\}^2}{\sum \text{var}(O)} \sim \chi_1^2$$

- Two sample log-rank test statistic

$t$	$X$	$Y = 1$	$Y = 0$	$N$
$t = t_{(1)}$	$X = 1$	$D_{11}$	$Y_{11} - D_{11}$	$Y_{11}$
	$X = 2$	$D_{21}$	$Y_{21} - D_{21}$	$Y_{21}$
		$D_1$	$Y_1 - D_1$	$Y_1$
$t = t_{(k)}$	$X = 1$	$D_{1k}$	$Y_{1k} - D_{1k}$	$Y_{1k}$
	$X = 2$	$D_{2k}$	$Y_{2k} - D_{2k}$	$Y_{2k}$
		$D_k$	$Y_k - D_k$	$Y_k$

– test statistic

$$\frac{\{\sum_k (D_{1k} - E_{1k})\}^2}{\sum_k \text{var}(D_{1k})} \sim \chi_1^2$$

– assumptions

1.  $D_{1k} \sim B(Y_{1k}, \lambda_1)$  and  $D_{2k} \sim B(Y_{2k}, \lambda_2)$

2. Given  $D_k$ ,  $D_{1k}$  is hypergeometric

$$\binom{D_k}{D_{1k}} \binom{Y_k - D_k}{Y_{1k} - D_{1k}} / \binom{Y_k}{Y_{1k}}$$

3.  $ED_{1k} = D_k Y_{1k} / Y_k$ ,  $\text{var}(D_{1k}) = Y_{1k} Y_{2k} D_k (Y_k - D_k) / [Y_k^2 (Y_k - 1)]$

- example: Breast cancer prognosis (D. Collett)
- issues
  1. when would the log-rank test be most powerful?
  2. when would the log-rank test be least powerful?
  3. does it take into account the actual time?

- Extension to multiple samples

$t$	$X$	Events	No Events	$N$
$t = t_{(k)}$	$X = 0$	$D_{0k}$	$Y_{0k} - D_{0k}$	$Y_{0k}$
	$X = 1$	$D_{1k}$	$Y_{1k} - D_{1k}$	$Y_{1k}$
	...	...	...	...
	$X = j$	$D_{jk}$	$Y_{jk} - D_{jk}$	$Y_{jk}$
	...	...	...	...
	$X = m$	$D_{mk}$	$Y_{mk} - D_{mk}$	$Y_{mk}$
		$D_k$	$Y_k - D_k$	$Y_k$

- assumptions:

1.  $D_{jk} \sim B(Y_{jk}, \lambda_k), j = 0, 1, \dots, m$
2. Given  $D_k, (D_{0k}, \dots, D_{mk})$  is multivariate hypergeometric

$$\binom{Y_{0k}}{D_{0k}} \cdots \binom{Y_{mk}}{D_{mk}} / \binom{Y_k}{D_k}$$

$t$	$X$	Events	No Events	$N$
$t = t_{(k)}$	$X = 0$	$D_{0k}$	$Y_{0k} - D_{0k}$	$Y_{0k}$
	$X = 1$	$D_{1k}$	$Y_{1k} - D_{1k}$	$Y_{1k}$
	...	...	...	...
	$X = j$	$D_{jk}$	$Y_{jk} - D_{jk}$	$Y_{jk}$
	...	...	...	...
	$X = l$	$D_{lk}$	$Y_{lk} - D_{lk}$	$Y_{lk}$
	...	...	...	...
	$X = m$	$D_{mk}$	$Y_{mk} - D_{mk}$	$Y_{mk}$
		$D_k$	$Y_k - D_k$	$Y_k$

- $O_k - E_k = (D_{1k} - E_{1k}, \dots, D_{mk} - E_{mk})$  with

$$E_{jk} = D_k Y_{jk} / Y_k$$

- $\text{var}(O_k) = (\sigma_{jl,k}^2)_{m \times m}$

- $\text{var}(O_{jk}) = \sigma_{jj,k}^2 = Y_{jk}(Y_k - Y_{jk})D_k(Y_k - D_k) / [Y_k^2(Y_k - 1)]$

- $\text{cov}(O_{jk}, O_{lk}) = \sigma_{jl,k}^2 = -Y_{jk}Y_{lk}D_k(Y_k - D_k) / [Y_k^2(Y_k - 1)]$

- test statistics: define  $O - E = \sum_k (O_k - E_k)$ ,  $\text{var}(O) = \sum_k \text{var}(O_k)$

$$(O - E)\text{var}^{-1}(O)(O - E)^T \sim \chi_m^2$$

- Counting processes representation of  $O - E$ 
  - sample 1:  $j = 1, 2, \dots, n_1$ 
    1. data:  $X_{1j} = \min(T_{1j}, C_{1j}), \Delta_{1j} = I(T_{1j} \leq C_{1j})$
    2. counting processes:  $N_{1j}(t) = I(X_{1j} \leq t, \Delta_{1j} = 1), N_1(t) = \sum_j N_{1j}(t)$
    3. at-risk processes:  $Y_{1j}(t) = I(X_{1j} \geq t), Y_1(t) = \sum_j Y_{1j}(t)$
  - sample 2:  $j = 1, 2, \dots, n_2$ 
    1. data:  $X_{2j} = \min(T_{2j}, C_{2j}), \Delta_{2j} = I(T_{2j} \leq C_{2j})$
    2. counting processes:  $N_{2j}(t) = I(X_{2j} \leq t, \Delta_{2j} = 1), N_2(t) = \sum_j N_{2j}(t)$
    3. at-risk processes:  $Y_{2j}(t) = I(X_{2j} \geq t), Y_2(t) = \sum_j Y_{2j}(t)$
  - total sample:  $N(t) = N_1(t) + N_2(t), Y(t) = Y_1(t) + Y_2(t)$

$$- O_k = D_{1k} = \sum_{j=1}^{n_1} dN_{1j}(t_{(k)})$$

$$\Rightarrow O = \sum_{t(k)} O_k = \sum_{j=1}^{n_1} \int_0^\infty dN_{1j}(u)$$

$$- E_k = D_k Y_{1k} / Y_k = \sum_{i=1}^2 \sum_{j=1}^{n_i} dN_{ij}(t_{(k)}) Y_1(t_{(k)}) / Y(t_{(k)})$$

$$\Rightarrow E = \sum_{t(k)} E_k = \sum_{i=1}^2 \sum_{j=1}^{n_i} \int_0^\infty \frac{Y_1(u)}{Y(u)} dN_{ij}(u)$$

- Then

$$\begin{aligned} O - E &= \sum_{j=1}^{n_1} \int_0^\infty dN_{1j}(u) - \sum_{i=1}^2 \sum_{j=1}^{n_i} \int_0^\infty \frac{Y_1(u)}{Y(u)} dN_{ij}(u) \\ &= \sum_{j=1}^{n_1} \int_0^\infty dN_{1j}(u) - \sum_{j=1}^{n_1} \int_0^\infty \frac{Y_1(u)}{Y(u)} dN_{1j}(u) - \sum_{j=1}^{n_2} \int_0^\infty \frac{Y_1(u)}{Y(u)} dN_{2j}(u) \\ &= \sum_{j=1}^{n_1} \int_0^\infty \frac{Y_2(u)}{Y(u)} dN_{1j}(u) - \sum_{j=1}^{n_2} \int_0^\infty \frac{Y_1(u)}{Y(u)} dN_{2j}(u) \\ &= \int_0^\infty \frac{Y_2(u)}{Y(u)} dN_1(u) - \int_0^\infty \frac{Y_1(u)}{Y(u)} dN_2(u) \end{aligned}$$

– (continued)

$$\begin{aligned} O - E &= \int_0^\infty \frac{Y_2(u)}{Y(u)} dN_1(u) - \int_0^\infty \frac{Y_1(u)}{Y(u)} dN_2(u) \\ &= \int_0^\infty \frac{Y_1(u)Y_2(u)}{Y(u)} \left\{ \frac{dN_1(u)}{Y_1(u)} - \frac{dN_2(u)}{Y_2(u)} \right\} \\ &= \int_0^\infty \frac{Y_1(u)Y_2(u)}{Y(u)} \left\{ d\hat{\Lambda}_1(u) - d\hat{\Lambda}_2(u) \right\} \end{aligned}$$

– or alternatively,

$$O - E = \int_0^{\infty} \frac{Y_2(u)dN_1(u) - Y_1(u)dN_2(u)}{Y(u)}$$

where the numerator is indeed

$$dN_1(u)[Y_2(u) - dN_2(u)] - dN_2(u)[Y_1(u) - dN_1(u)]$$

– c.f. single  $2 \times 2$  table at a specific  $u$

$u \geq 0$	$X$	Events	No Events	At-Risk
$u$	$X = 1$	$a = dN_1(u)$	$b = Y_1(u) - dN_1(u)$	$Y_1(u)$
	$X = 2$	$c = dN_2(u)$	$d = Y_2(u) - dN_2(u)$	$Y_2(u)$

- view 1: log-rank test statistic is simply sum of  $(ad - bc)$  over  $2 \times 2$  tables

- Log-rank statistic

$$\int_0^{\infty} \frac{Y_1(u)Y_2(u)}{Y_1(u) + Y_2(u)} \left\{ d\hat{\Lambda}_1(u) - d\hat{\Lambda}_2(u) \right\}$$

- why  $\{Y_1(u)Y_2(u)\}/\{Y_1(u) + Y_2(u)\}$ ?
- harmonic mean

$$\propto \frac{Y_1(u)Y_2(u)}{Y_1(u) + Y_2(u)} = \frac{1}{\frac{1}{Y_1(u)} + \frac{1}{Y_2(u)}}$$

- variance of difference in two proportions under null

$$\text{var}(\hat{p}_1 - \hat{p}_2) = pq \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \propto \frac{1}{n_1} + \frac{1}{n_2}$$

- view 2: log-rank statistic is a cumulative difference between two rates weighted by the inverse of its variance

## 2. Weighted Log-rank statistics

- A general form

$$W = \int_0^{\infty} W(u) \frac{Y_1(u)Y_2(u)}{Y_1(u) + Y_2(u)} \left\{ d\hat{\Lambda}_1(u) - d\hat{\Lambda}_2(u) \right\}$$

- statistics of the class  $K$  (Gill, 1980)
- $W(\cdot)$ : weight function
- weighted differences in cumulative hazard functions

- Choices of weight function  $W(\cdot)$ 
  - $W(t) = 1$ : Log-rank
  - $W(t) = c \cdot Y(t)$ : Wilcoxon rank-sum, Gehan-Wilcoxon
  - $W(t) = \hat{S}(t-)$ : Prentice-wilcoxon
  - $W(t) = \hat{S}(t-)^{\rho}, \rho > 0$ :  $G^{\rho}$ -family
  - $W(t) = \hat{S}(t-)^{\rho}[1 - \hat{S}(t-)]^{\gamma}$ :  $G^{\rho, \gamma}$ -family
  - $W(t) = K(t)[n_1 n_2 / (n_1 + n_2)]^{1/2} [Y_1(t) + Y_2(t)] / Y(t)$ : the class  $K$

## 2. Weighted Log-rank statistics

- Weighted Log-rank statistic

$$W = \int_0^{\infty} W(u) \frac{Y_1(u)Y_2(u)}{Y_1(u) + Y_2(u)} \left\{ \frac{dN_1(u)}{Y_1(u)} - \frac{dN_2(u)}{Y_2(u)} \right\}$$

- How do we view it?
  1. Perspective 1: weighted difference in cumulative hazard functions
  2. Perspective 2: weighted sum of  $(ad - bc)$  of the  $2 \times 2$  tables over time

- An equivalent form:

$$\begin{aligned}
 W &= \int_0^\infty W(u) \frac{Y_0(u)Y_1(u)}{Y_0(u) + Y_1(u)} \left\{ \frac{dN_1(u)}{Y_1(u)} - \frac{dN_0(u)}{Y_0(u)} \right\} \\
 &= \int_0^\infty W(u) \left\{ \frac{Y_0(u)}{Y(u)} dN_1(u) - \frac{Y_1(u)}{Y(u)} dN_0(u) \right\}
 \end{aligned}$$

–  $Z_i = 0/1$  in group 0/1, respectively

–  $Y_1(u) = \sum_{i=1}^n Z_i Y_i(u)$

–

$$\frac{Y_1(u)}{Y(u)} = \frac{\sum_{i=1}^n Z_i Y_i(u)}{\sum_{i=1}^n Y_i(u)} = \bar{Z}(u)$$

–

$$\frac{Y_0(u)}{Y(u)} = 1 - \bar{Z}(u)$$

- After substitution,

$$\begin{aligned}
 W &= \int_0^\infty W(u) \left\{ \frac{Y_0(u)}{Y(u)} dN_1(u) - \frac{Y_1(u)}{Y(u)} dN_0(u) \right\} \\
 &= \int_0^\infty W(u) \{ [1 - \bar{Z}(u)] dN_1(u) + [0 - \bar{Z}(u)] dN_0(u) \} \\
 &= \int_0^\infty W(u) \sum_{i=1}^n \{ Z_i - \bar{Z}(u) \} dN_i(u) \\
 &= \sum_{i=1}^n \int_0^\infty W(u) \{ Z_i - \bar{Z}(u) \} dN_i(u)
 \end{aligned}$$

- Perspective 3: weighted difference on covariates, because

$$\bar{Z}(u) = \frac{\sum_{i=1}^n Z_i Y_i(u)}{\sum_{i=1}^n Y_i(u)} \rightarrow \frac{E[ZI(X \geq u)]}{\Pr(X \geq u)} = E(Z | X \geq u) = \mu_Z(u)$$

- Story continues:

$$\begin{aligned}
 W &= \sum_{i=1}^n \int_0^{\infty} W(u) \{Z_i - \bar{Z}(u)\} dN_i(u) \\
 &= \sum_{i=1}^n \int_0^{\infty} W(u) \{Z_i - \bar{Z}(u)\} [dN_i(u) - Y_i(u)\lambda(u)du] \\
 &\quad + \sum_{i=1}^n \int_0^{\infty} W(u) \{Z_i - \bar{Z}(u)\} Y_i(u)\lambda(u)du
 \end{aligned}$$

- the second sum is zero, because

$$\sum_{i=1}^n \{Z_i - \bar{Z}(u)\} Y_i(u) = \sum_{i=1}^n \left\{ Z_i - \frac{\sum_{i=1}^n Z_i Y_i(u)}{\sum_{i=1}^n Y_i(u)} \right\} Y_i(u) = 0$$

- Eureka!

$$W = \sum_{i=1}^n \int_0^{\infty} W(u) \{Z_i - \bar{Z}(u)\} dM_i(u)$$

- cf. linear regression model  $y_i = \beta_0 + \beta_1 x_i + e_i$
- LS estimation of the second equation w.r.t.  $\beta_1$ :

$$\sum_{i=1}^n x_i (y_i - \hat{y}_i) = \sum_{i=1}^n x_i \hat{e}_i = \sum_{i=1}^n (x_i - \bar{x}) \hat{e}_i$$

- Asymptotics under  $H_0 : \lambda(t | Z_i) = \lambda(t)$

–

$$U_n(t) = n^{-1/2} \sum_{i=1}^n \int_0^t W(u) \{Z_i - \bar{Z}(u)\} dM_i(u)$$

- weighted Log-rank statistic:  $W = n^{1/2}U_n(\tau)$

- $\mathcal{F}_t = \sigma\{N_i(u), Y_i(u), Z_i; i = 1, 2, \dots, n, 0 \leq u \leq t\}$

- $M_i(\cdot)$  are  $\mathcal{F}_t$ -martingales

- $H_i(u) = n^{-1/2}W(u) \{Z_i - \bar{Z}(u)\}$  are  $\mathcal{F}_t$ -predictable

- $U_n(t) = \sum_{i=1}^n \int_0^t H_i(u) dM_i(u)$

- Martingale CLT

- $U_n(t) = \sum_{i=1}^n \int_0^t n^{-1/2} W(u) \{Z_i - \bar{Z}(u)\} dM_i(u)$

- $\langle U_n, U_n \rangle (t)$  should be

$$\begin{aligned} & \sum_{i=1}^n \int_0^t \left[ n^{-1/2} W(u) \{Z_i - \bar{Z}(u)\} \right]^2 Y_i(u) \lambda(u) du \\ &= \int_0^t \frac{1}{n} \sum_{i=1}^n W(u)^2 \{Z_i - \bar{Z}(u)\}^2 Y_i(u) \lambda(u) du \end{aligned}$$

- Assume that  $W(u) \rightarrow w(u)$

$$\langle U_n, U_n \rangle (t)$$

$$\begin{aligned} & \rightarrow_P \int_0^t w(u)^2 E \left[ \{Z - \mu_Z(u)\}^2 I(X \geq u) \right] \lambda(u) du \\ & = \alpha(t) \end{aligned}$$

–  $\langle U_{n,\epsilon}, U_{n,\epsilon} \rangle (t) \rightarrow_P 0$ , because

$$\sum_{i=1}^n \int_0^t \left[ n^{-1/2} W(u) \{Z_i - \bar{Z}(u)\} \right]^2 \\ \times I \left\{ \left| n^{-1/2} W(u) \{Z_i - \bar{Z}(u)\} \right| \geq \epsilon \right\} Y_i(u) \lambda(u) du$$

– therefore,  $U_n \Rightarrow U$

$$\text{var}[U(t)] = \alpha(t)$$

$$= \int_0^t w(u)^2 E \left[ \{Z - \mu_Z(u)\}^2 I(X \geq u) \right] \lambda(u) du$$

$$= \int_0^t w(u)^2 \frac{E \left[ \{Z - \mu_Z(u)\}^2 I(X \geq u) \right]}{EI(X \geq u)} EI(X \geq u) \lambda(u) du$$

$$= \int_0^t w(u)^2 \text{var}(Z \mid X \geq u) EI(X \geq u) \lambda(u) du$$

- Weighted Log-rank statistic:  $W = n^{1/2}U_n(\tau)$

$$n^{-1/2}W \rightarrow_D \mathcal{N}(0, \alpha(\tau))$$

- Standardized weighted Log-rank test statistic:

$$\frac{n^{-1/2}W}{\sqrt{\alpha(\tau)}} \rightarrow_D \mathcal{N}(0, 1)$$

- How to estimate  $\alpha(\tau)$ ?

- We know

- $\alpha(t)$  equals

$$\int_0^t w(u)^2 \text{var}(Z | X \geq u) EI(X \geq u) \lambda(u) du$$

- $\hat{\lambda}(t)dt = d\hat{\Lambda}(t) = dN(t)/Y(t)$

- $\hat{EI}(X \geq u) = Y(u)/n$

- $\widehat{\text{var}}(Z | X \geq u) = \hat{p}\hat{q}$ , where

$$\hat{p} = \hat{E}(Z | X \geq u) = \bar{Z}(u)$$

- $\hat{\alpha}(\tau)$  is estimated by

$$n^{-1} \int_0^t W(u)^2 \bar{Z}(u) [1 - \bar{Z}(u)] dN(u)$$

- Standardized weighted Log-rank statistics

$$\frac{n^{-1/2}W}{\sqrt{\hat{\alpha}(\tau)}} = \frac{\sum_{i=1}^n \int_0^\tau W(u) \{Z_i - \bar{Z}(u)\} dN_i(u)}{\{\sum_{i=1}^n \int_0^\tau W(u)^2 \bar{Z}(u) [1 - \bar{Z}(u)] dN_i(u)\}^{1/2}}$$

goes to  $\mathcal{N}(0, 1)$

- Reject  $H_0$  when

$$\left| \frac{n^{-1/2}W}{\sqrt{\hat{\alpha}(\tau)}} \right| > 1.96$$

for type-I error of 5%

- What is weighted Log-rank test statistic anyway?

–

$$\frac{n^{-1/2}W}{\sqrt{\hat{\alpha}(\tau)}} = \frac{\sum_{i=1}^n \int_0^\tau W(u) \{Z_i - \bar{Z}(u)\} dN_i(u)}{\{\sum_{i=1}^n \int_0^\tau W(u)^2 \bar{Z}(u) [1 - \bar{Z}(u)] dN_i(u)\}^{1/2}}$$

– if  $\Delta_i = 0$ , then  $dN_i(u) = 0$

– if  $\Delta_i = 1$ ,

\*  $dN_i(t) = 1$  at  $t = X_i$  and 0 elsewhere

\*  $\int_0^\tau W(u) \{Z_i - \bar{Z}(u)\} dN_i(u) = W(X_i) \{Z_i - Y_1(X_i)/Y(X_i)\} = w_i \{Z_i - Y_{i1}/Y_i\}$

– Numerator is  $\sum_{i=1}^n w_i \Delta_i (Z_i - Y_{i1}/Y_i)$

– Denominator is  $\{\sum_{i=1}^n w_i^2 \Delta_i Y_{i1} Y_{i0} / Y_i^2\}^{1/2}$

- Numerator is  $\sum_{i=1}^n w_i \Delta_i (Z_i - Y_{i1}/Y_i)$
- Denominator is  $\{\sum_{i=1}^n w_i^2 \Delta_i Y_{i1} Y_{i0} / Y_i^2\}^{1/2}$
- $2 \times 2$  table for  $i$ th failure,  $\Delta_i = 1$

$t$	$Z$	$dN(t) = 1$	$Y(t) - dN(t)$	$Y(t)$
$X_i$	$Z_i = 1$	1	$Y_{i1} - 1$	$Y_{i1}$
	$Z_i = 0$	0	$Y_{i0}$	$Y_{i0}$
		1	$Y_i - 1$	$Y_i$
$X_i$	$Z_i = 1$	0	$Y_{i1}$	$Y_{i1}$
	$Z_i = 0$	1	$Y_{i0} - 1$	$Y_{i0}$
		1	$Y_i - 1$	$Y_i$

- $O_i = Z_i = 0/1, E_i = 1 * Y_{i1}/Y_i$
- $\text{var}(O_i) = 1 * (Y_i - 1) * Y_{i1} * Y_{i0} / [Y_i^2 (Y_i - 1)]$

- Power analysis of weighted Log-rank test statistics

1. type-I error:  $\alpha = 5\%$

2. power level

3. alternative hypothesis

4. error bound

- Under  $H_0 : \lambda_0(t) = \lambda_1(t) = \lambda(t)$ ,

$$n^{-1/2}W \sim \mathcal{N}(0, \alpha(\tau))$$

- Alternative hypothesis

- $H_1 : \lambda_1(t) = \lambda_0(t)e^{\beta_n \times \theta(t)}$

- $\log[\lambda_1(t | Z_i)/\lambda_0(t)] = \beta_n Z_i \times \theta(t)$

- $\theta(t)$ : take into account of nonproportionality

- $\beta_n$ : distance between the null and an alternative

- 1.  $n^{1/2}\beta_n \rightarrow \xi \in (0, \infty)$

- 2. local alternatives:  $\beta_n \rightarrow 0$

- Given a sample size  $n$ ,

$$\text{Power} = \Pr \left\{ \left| n^{-1/2}W / \sqrt{\hat{\alpha}(\tau)} \right| > z_{1-\alpha/2} \mid H_1 \right\}$$

- Asymptotic distribution of  $n^{-1/2}W$  under  $H_1$

- $n^{-1/2}W = n^{-1/2} \sum_{i=1}^n \int_0^\tau W(u) \{Z_i - \bar{Z}(u)\} dN_i(u)$

- under  $H_1$ ,

$$E[dN_i(u) \mid \mathcal{F}_{u-}] = Y_i(u) \lambda_i(u) du = Y_i(u) \lambda_0(u) e^{\beta_n Z_i \times \theta(u)} du$$

–

$$n^{-1/2}W = n^{-1/2} \sum_{i=1}^n \int_0^\tau W(u) \{Z_i - \bar{Z}(u)\} dM_i(u) \\ + n^{-1/2} \sum_{i=1}^n \int_0^\tau W(u) \{Z_i - \bar{Z}(u)\} Y_i(u) \lambda_0(u) e^{\beta_n Z_i \times \theta(u)} du$$

- apply MCLT

- Asymptotic distribution of  $n^{-1/2}W$  under  $H_1$

- $n^{-1/2}W = n^{-1/2} \sum_{i=1}^n \int_0^\tau W(u) \{Z_i - \bar{Z}(u)\} dN_i(u)$

- under  $H_1$ ,

$$E[dN_i(u) \mid \mathcal{F}_{u-}] = Y_i(u) \lambda_i(u) du = Y_i(u) \lambda_0(u) e^{\beta_n Z_i \times \theta(u)} du$$

–

$$\begin{aligned} n^{-1/2}W &= n^{-1/2} \sum_{i=1}^n \int_0^\tau W(u) \{Z_i - \bar{Z}(u)\} dM_i(u) \\ &+ n^{-1/2} \sum_{i=1}^n \int_0^\tau W(u) \{Z_i - \bar{Z}(u)\} Y_i(u) \lambda_0(u) e^{\beta_n Z_i \times \theta(u)} du \\ &= \text{Term I} + \text{Term II} \end{aligned}$$

- **Term I**: predictable variation

$$\int_0^\tau n^{-1} \sum_{i=1}^n W(u)^2 \{Z_i - \bar{Z}(u)\}^2 Y_i(u) \lambda_0(u) e^{\beta_n Z_i \times \theta(u)} du$$

- $\beta_n \rightarrow 0 \Rightarrow e^{\beta_n Z_i \times \theta(u)} \rightarrow 1$  and  $H_{1n} \rightarrow H_0$

- **Term I**  $\rightarrow \int_0^\tau w(u)^2 E[(Z_i - \mu_Z(u))^2 I(X \geq u)] \lambda_0(u) du$

- **Term I** asymptotically

$$\mathcal{N}\left(0, \int_0^\tau w(u)^2 E_{H_0}[(Z_i - \mu_Z(u))^2 I(X \geq u)] \lambda_0(u) du\right)$$

- Term II:

$$n^{-1/2} \sum_{i=1}^n \int_0^\tau W(u) \{Z_i - \bar{Z}(u)\} Y_i(u) \lambda_0(u) e^{\beta_n Z_i \times \theta(u)} du$$

- Taylor expansion:  $e^{\beta_n Z_i \times \theta(u)} = 1 + \beta_n Z_i \times \theta(u) + O(\beta_n^2)$
- $O(\beta_n^2)/\beta_n^2$  is bounded
- Term II = Term IIa + Term IIb + Term IIc

- Term IIa

$$n^{-1/2} \sum_{i=1}^n \int_0^\tau W(u) \{Z_i - \bar{Z}(u)\} Y_i(u) \lambda_0(u) du = 0$$

- Term IIb

$$\begin{aligned}
& n^{-1/2} \sum_{i=1}^n \int_0^\tau W(u) \{Z_i - \bar{Z}(u)\} Z_i Y_i(u) \beta_n \theta(u) \lambda_0(u) du \\
&= \int_0^\tau W(u) n^{-1} \sum_{i=1}^n \{Z_i - \bar{Z}(u)\} Z_i Y_i(u) \times n^{1/2} \beta_n \times \theta(u) \lambda_0(u) du \\
&= \int_0^\tau W(u) n^{-1} \sum_{i=1}^n \{Z_i - \bar{Z}(u)\}^2 Y_i(u) \times n^{1/2} \beta_n \times \theta(u) \lambda_0(u) du \\
&\rightarrow \int_0^\tau w(u) E_{H_0}[(Z_i - \mu_Z(u))^2 I(X \geq u)] \times \xi \times \theta(u) \lambda_0(u) du \\
&= \xi \int_0^\tau w(u) \theta(u) E_{H_0}[(Z_i - \mu_Z(u))^2 I(X \geq u)] \lambda_0(u) du
\end{aligned}$$

- Term IIc

- $|O(\beta_n^2)/\beta_n^2| < M \Rightarrow nO(\beta_n^2) = O(n\beta_n^2)$

- $n^{1/2}O(\beta_n^2) = n^{-1/2}O(n\beta_n^2) = o(n^{-1/2})$

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{Z}(u)\} Y_i(u) O(\beta_n^2) \lambda_0(u) du \\ &= \int_0^\tau W(u) n^{-1} \sum_{i=1}^n \{Z_i - \bar{Z}(u)\} Y_i(u) o(n^{-1/2}) \lambda_0(u) du \\ &\rightarrow 0 \end{aligned}$$

- **Term II** = Term IIa + Term IIb + Term IIc converges to

$$\xi \int_0^{\tau} w(u)\theta(u)E_{H_0}[(Z_i - \mu_Z(u))^2 I(X \geq u)]\lambda_0(u)du$$

- Recall on **Term I**

$$\mathcal{N}(0, \int_0^{\tau} w(u)^2 E_{H_0}[(Z_i - \mu_Z(u))^2 I(X \geq u)]\lambda_0(u)du)$$

- Under  $H_{1n}$ :

- $A(w^2) = \int_0^{\tau} w(u)^2 E_{H_0}[(Z_i - \mu_Z(u))^2 I(X \geq u)]\lambda_0(u)du$

- 

$$n^{-1/2}W \sim \mathcal{N}(\xi A(\theta w), A(w))$$