4장 개수과정과 마팅게일

2020년 가을학기

전북대학교 통계학과

Outline

개수과정

강도과정과 누적강도과정

마팅게일 (martingale)

개수과정으로 표현한 추정량

개수과정

중도절단을 포함한 생존데이터에서의 추론

- 우도함수 이용
- 개수과정(counting process) 이용 [Aalen(1971)]

생존데이터

- 각 개체에 대해 특정 사건 발생이 일어날 때까지의 시간 데이터
- 매 시점에서 사건발생 건수로 표현 가능 - 개수과정

개수과정의 활용

- Nelson-Aalen 비모수적 누적위험함수
- Kaplan-Meier 추정량과 로그순위 검정량
- Cox 모형

NOTE :개수과정 접근법 - 확률과 확률과정 이론에 근거함

개수과정 (counting process)

- 시간구간에 걸쳐 수집된 데이터 - 확률과정(stochastic process)으로 표현가능
- 시간이 지남에 따라 발생하는 사건의 개수에 대한 데이터 - 개수과정으로 표현

개수과정 $N = \{N(t),\ t \geq 0\}$

- $N(t)$= 구간 $(0, t]$에서 발생한 사건의 수
- $N(0) = 0,\ N(t) < \infty$

개수과정 (counting process)

NOTE: 4장부터는 생존시간을 Y_i 대신에 $\tilde{T}_i$로 표기

생존데이터의 개수과정 표현

생존데이터$\{(T_i, \delta_i), i = 1, 2, \ldots, n\} \Rightarrow$ 개수과정 $\{(N_i(t), Y_i(t)), i = 1, 2, \ldots, n\}$

$N_i(t) = I(T_i \le t, \delta = 1)$(구간 $[0, t]$에서 i번째 개체에 사건 발생 여부)

$\tilde{T}_i \le t, \tilde{T}_i \le C_i$인 경우 $: N_i(t) = 1$

점프크기가 1인 우연속(right continuous) 계단함수(step function)

$Y_i(t) = I(T_i \ge t)$(구간 $[0, t]$에 i번째 개체가 연구대상에 남아 있는지 여부)

$dN_i(t) = [t, t + dt)$(구간에서 N_i의 증가량)

한 개체에 여러 번의 사건이 발생하는 경우

예) 재발사건 - 여러 번 재발하는 경우

- $N_i(t)$=구간 $[0, t]$에서 i번째 개체에 발생한 사건의 개수
- $N_i(t) \ge 1$일 수 있음

개수과정 (counting process)

- $N(t) = \sum_{i=1}^{n} N_i(t) = \sum_{i=1}^{n} I(T_i \le t, \delta = 1) = \sum_{t_i \le t} \delta_i$
  (t시점까지의 사건발생수 개수과정)
- $Y(t) = \sum_{i=1}^{n} Y_i(t)$ (t시점에 위험에 놓인 개체수 개수과정)

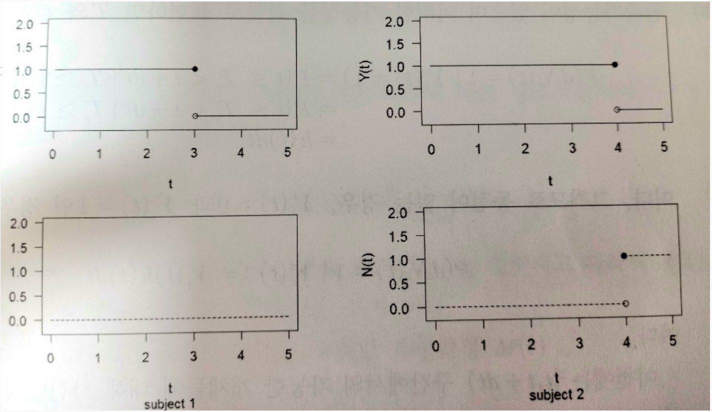

Figure 1: 예제 4.1) 2명의 생존데이터 (T, δ) : (3,0), (4,1)

강도과정과 누적강도과정

생존분석 모형 : $N(t) = \Lambda(t) + M(t)$

- DATA = MODEL+ERROR
- $\Lambda(t) = \int_0^t \lambda(s)ds$: 누적강도과정(cumulative intensity process)
- $\lambda(t) = Y(t)h(t)$: 강도과정(intensity process)
- $M(t)$: 개수과정 마팅게일(martingale)

- 일반 선형모형에서 오차에 대해 정규분포 가정 $\rightarrow$ 회귀모수(절편, 기울기 등)에 대한 통계적 추론 가능
- 생존분석과 연관된 모형의 통계적 추론 - 마팅게일의 성질이 중요한 역할

강도과정과 누적강도과정

용어

- 확률공간 $(\Omega, \mathcal{F}, P)$ - $\mathcal{F}$: σ-field, P:확률측도
- $\{\mathcal{F}_t, t \geq 0\}$:t시점까지의 역사(history)
  - $s \leq t$에 대하여 $\mathcal{F}_s \subset \mathcal{F}_t$
  - 자료와 연관된 모든 정보를 포함

성질

- $E[N(t)|\mathcal{F}_{t-}] = E[\Lambda(t)|\mathcal{F}_{t-}] = \Lambda(t)$
- $E[dM(t)|\mathcal{F}_{t-}] = E[dN(t) - d\Lambda(t)|\mathcal{F}_{t-}] = \lambda(t) - \lambda(t) = 0$

마팅게일 (martingale)

마팅게일 (martingale)

$(M(t), \mathcal{F}_t)$: 좌극한을 가진 우연속 (cadlag) 확률과정 마팅게일

- 마팅게일 증가량은 평균 0
  - 모든 $t \geq 0$에 대하여 $E|M(t)| < \infty$
  - 모든 $t > s \geq 0$에 대하여 $E(M(t)|\mathcal{F}_s) = M(s)$
  - 모든 $t \geq 0$에 대하여 $E(dM(t)|\mathcal{F}_{t-}) = 0$
- 마팅게일 증가량은 서로 상관되지 않음

  $t, u, s > 0$에 대하여

$$Cov[M(t), M(t+u) - M(t)] = 0$$

$$Cov[M(t) - M(t-s), M(t+u) - M(t)] = 0$$

개수과정 마팅게일 (counting process martingale)

개수과정은 마팅게일과 보상자(compensator)로 유일하게 분리됨 (by Doob-Meier decomposition)

$$\begin{aligned} N_i(t) &= \Lambda_i(t) + M_i(t) \quad \text{(counting process=compensator+martingale)} \\ &= \int_0^t Y_i(u)h(u)du + M_i(t) \end{aligned}$$

- $E[dN(t)|\mathcal{F}_{t-}] = \lambda(t)$
- $Var[N(t)|\mathcal{F}_{t-}] = \lambda(t)(1-\lambda(t))$

개수과정으로 표현한 추정량

개수과정으로 표현한 추정량

Nelson-Aalen 추정량

NOTE $H(s+u) - H(s) \approx h(s)u = P(\text{event in } (s, s+u] | \text{ at risk at } s)$

$$\hat{H}(t) = \int_0^t \frac{N(s+u) - N(s)}{Y(s)} du = \int_0^t \frac{dN(s)}{Y(s)} = \sum_{i:t_i \leq t} \frac{dN(t_i)}{Y(t_i)}$$

Kaplan-Meier 추정량

$$\hat{S}(t) = \prod_{s=0}^{t} \left[1 - d\hat{H}(t)\right] = \prod_{s=0}^{t} \left[1 - \frac{dN(s)}{Y(s)}\right]$$

- $dN(t) > 0$인 시점에서 점프하는 계단함수
- $\hat{S}(t)/S(t) - 1$:마팅게일 M에 의한 확률적분으로 표현되는 마팅게일
- 생존함수에 대한 신뢰구간 추론에 사용