통계세미나- 주제: 생존분석

2장 생존분석의 기초적인 개념과 통계량

2020년 가을학기

전북대학교 통계학과

생존 분석 (survival analysis)

생존시간 (survival time): 어떤 정의된 시점부터 특정한 사건(event)이 관측될 때까지의 시간 (time-to-event)

생존분석의 주 관심 문제

- 생존함수 (survival function) 추정 : 예) 1년 생존율, 3년 생존율, 환자의 절반이 생존하는 시간, 두 처리군의 생존분포 비교
- 생존함수 또는 위험함수 (hazard function)에 영향을 주는 공변량 (covariate) 또는 예측변수를 찾아내어 연관성 표현과 각 요인의 효과 추정

주요 용어와 기본 개념

특정 사건까지의 발생시간 T

- 생존시간 변수, $T > 0$인 확률변수
- 확률밀도함수 $f(t)$, 분포함수 $F(t)$

T의 분포 정보를 주는 함수들

- 생존함수 (survival function)
- 위험함수 (hazard function)
- 누적위험함수 (cumulative hazard function)
- 평균잔여수명(mean residual life)

NOTE: 일반적인 통계분석에서 요약통계량으로 관측값의 평균, 분산 등이 이용되나 생존분석에서는 중도절단자료로 인해 다른 형태의 통계량 필요

$\Rightarrow$ 생존함수, 위험함수, 누적위험함수, 평균잔여수명

생존함수 $S(t)$

생존함수(survival function)

t 시점 이후 사건 발생할 확률

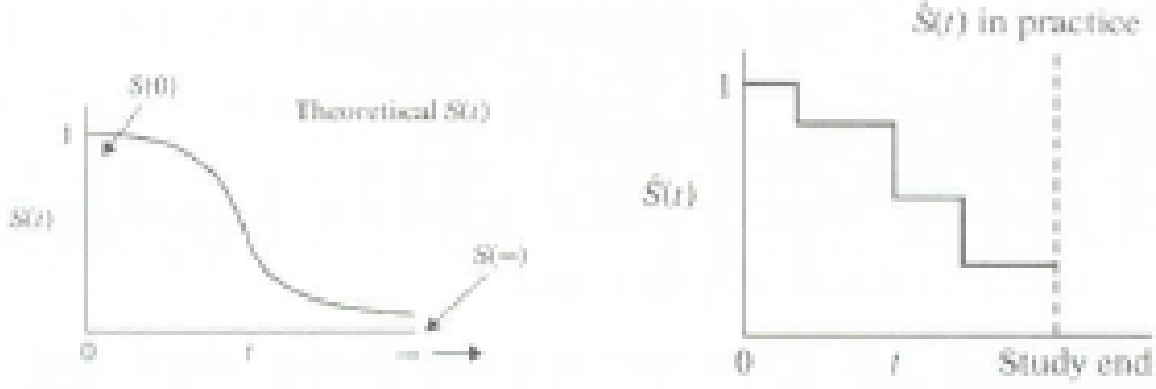
$$S(t) = P(T > t) = \int_t^{\infty} f(x) = 1 - P(T \leq t) = 1 - F(t)$$

- 비증가함수 (non-increasing function)
- $S(0) = 1, S(\infty) = 0, 0 \leq S(t) \leq 1$
- $P(T > t)$: 적어도 t시점까지 생존할 확률, t시점 이후에 사건이 일어날 확률
- $P(T \leq t)$: t시점 이전에 사건이 일어날 확률
- 신뢰성함수 (reliability function)라고도 불림

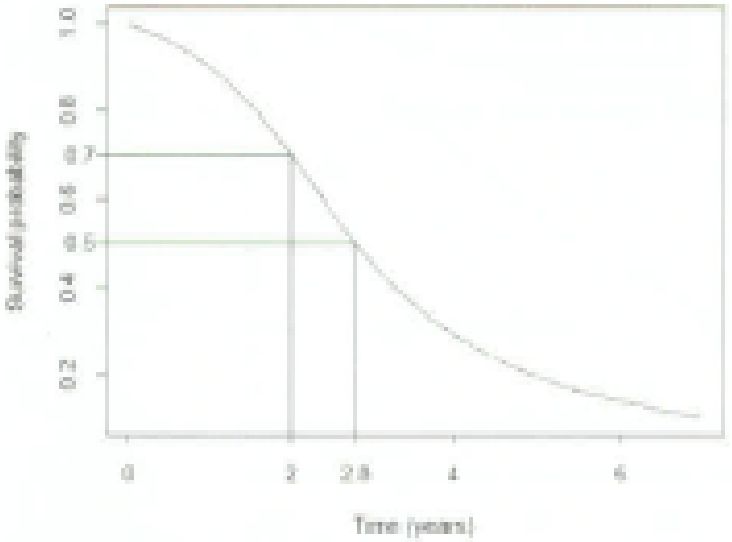
생존함수를 이용한 확률밀도함수의 표현

$$f(t) = -\frac{dS(t)}{dt}$$

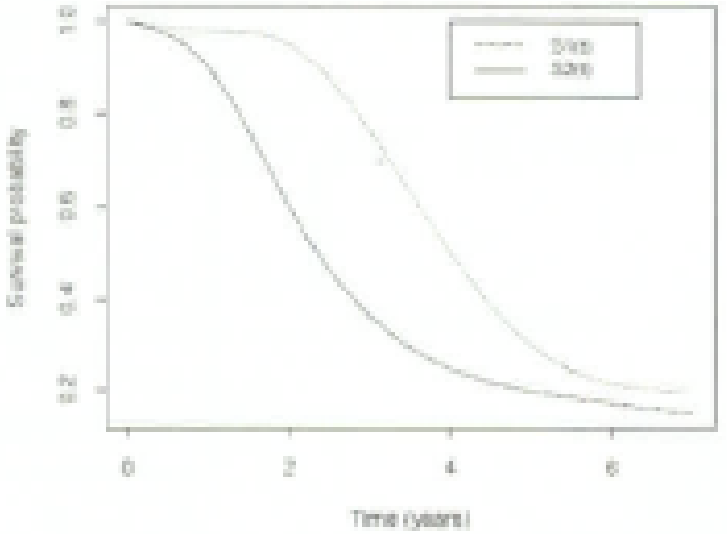
생존함수 $S(t)$



[그림 2.1] 이론적 생존함수와 추정된 생존함수



[그림 2.2] 생존함수와 생존확률



[그림 2.3] 두 개의 생존함수 비교

위험함수 $h(t)$

위험함수 (hazard function) 또는 위험률함수 (hazard rate function)

t 시점에서 생존한 조건 하에서 사건이 발생할 확률 (조건부 확률)

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t) | T \ge t)}{\Delta t}$$
$$= \frac{f(t)}{S(t)} = \frac{dF(t)/dt}{S(t)} = \frac{d(1 - S(t))/dt}{S(t)} = \frac{-S'(t)}{S(t)} = -\frac{d \log[S(t)]}{dt}$$

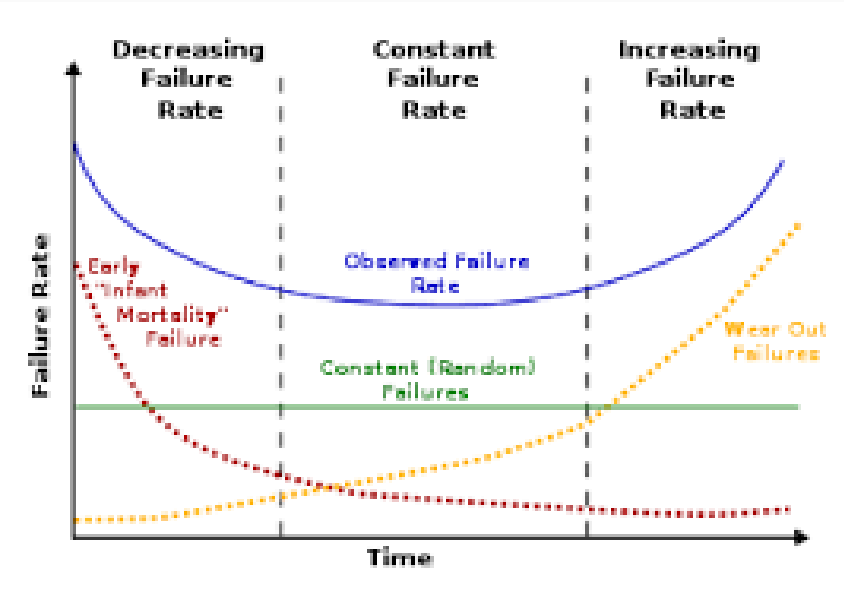
NOTE: 확률밀도함수

$$f(t) = \lim_{h \to 0} \frac{P(t \le T \le t + h)}{h}$$

위험함수 $h(t)$

해석

- $h(t) \geq 0$
- 확률로 표현되므로 일반적으로 관측불가능한 양 $\Rightarrow$ 데이터에 근거하여 추정



위험함수 형태의 예

위험률의 예

- 위험률 증가의 예: 나이에 따른 체력, 시간에 따른 기계의 마모율
- 장기이식 환자들은 이식 후 초기에 위험률 증가하나 일정기간 후 안정기에 들어 위험률 감소

누적위험함수 $H(t)$

누적위험함수 (cumulative hazard function)

t 시점까지의 누적 위험률

$$H(t) = \int_0^t h(u)du = \int_0^t \frac{\frac{d(1-S(u))}{du}}{S(u)} du = -\log S(t)$$

NOTE: 누적위험함수를 이용한 생존함수 표현

$$S(t) = \exp[-H(t)] = \exp\left[-\int_0^t h(u)du\right]$$

예제 2.1 와 2.2 : 직장암 (rectal cancer)

의사와 환자의 **관심**: 시간에 따른 생존 가능성과 위험률

1. 사건 : 암 수술 후 사망
2. 반응변수 : 암 수술 후 사망까지의 시간
3. 생존함수 : t 시점 이후 사망할 확률
4. 위험함수 : t 시점까지 생존한 상황에서 t 시점 순간에 사망할 위험률

의사와 환자의 **관심**: 시간에 따른 재발 가능성과 위험률

1. 사건 : 암 재발 사건
2. 반응변수 : 암 재발까지의 시간
3. 생존함수 : t 시점 이후 재발할 확률
4. 위험함수 : t 시점까지 재발이 없는 상황에서 t 시점 순간에 재발할 위험률

예제 2.3 : 디젤 엔진의 환풍기(diesel generator fan) 고장

엔지니어의 **관심**: 환풍기가 고장나기 전 교체 시점 결정

1. 사건 : 디젤 엔진의 환풍기 고장 사건
2. 반응변수 : 디젤 엔진의 환풍기가 고장날 때까지의 시간
3. 생존함수 : t 시점 이후 고장 날 확률
4. 위험함수 : t 시점까지 고장 없는 상황에서 t 시점 순간에 고장 날 위험률

평균잔여수명 (mean residual life function)

x 시점 이후 평균잔여수명 : x 시점까지 생존한 조건하에서 x 시점 이후 생존 가능한 시간의 기대값

$$mrl(x) = E(T - x | T > x) = \frac{\int_x^\infty (t-x) f(t) dt}{S(x)} = \frac{\int_x^\infty S(t) dt}{S(x)}$$

참고 2.2 적분식 활용 (교재)

참고 2.3 기댓값과 분산을 생존함수로 표현

$$\mu = E(T) = \int_0^\infty t f(t) dt = \int_0^\infty S(t) dt$$

$$\sigma^2 = Var(T) = 2\int_0^\infty t S(t) dt - \left[\int_0^\infty S(t) dt\right]^2$$

참고 2.4 평균잔여수명의 추정

$$\widehat{mrl}(x) = \frac{\int_x^\infty \hat{S}(t) dt}{\hat{S}(x)}$$

평균잔여수명과 중간수명

중간수명 (median life) 생존확률이 0.5인 시점 $x_{0.5}$: $S(x_{0.5}) = 0.5$

분위수 q-백분위수 x_q : $F(x_q) = q$ 또는 $S(x_q) = 1 - q$

분위수 추정 $\hat{x}_q = \min\{t_i : \hat{S}(t_i) \leq 1 - q\}$

예제 2.4
평균이 θ인 지수분포를 따르는 경우 q-분위수

예제 2.5
확률변수 T가 평균 $1/\lambda$지수분포를 따르는 경우

- 생존함수
- 위험함수
- 누적위험함수
- a시점에서의 평균잔여수명
- 중간수명

생존데이터에 대한 모수적 분포

생존분석의 모형 - 비모수적모형, 준모수적모형, **모수적모형**

생존분석에 자주 사용되는 모수적 분포들

분포	확률밀도함수 $f(t)$	생존함수 $S(t)$	위험함수 $h(t)$	평균 $E(T)$
지수	$\lambda e^{-\lambda t}$	$e^{-\lambda t}$	λ	$1/\lambda$
와이블	$\alpha\lambda(\lambda t)^{\alpha-1}e^{-(\lambda t)^{\alpha}}$	$e^{-(\lambda t)^{\alpha}}$	$\alpha\lambda(\lambda t)^{\alpha-1}$	$\Gamma(1+\frac{1}{\alpha})/\lambda$
감마	$\frac{1}{\Gamma(\alpha)}\lambda^{\alpha}t^{\alpha-1}e^{-\lambda t}$	$1-\int_0^{\lambda t}\frac{u^{\alpha-1}e^{-u}}{\Gamma(\alpha)}du$	$\frac{f(t)}{S(t)}$	$\frac{\alpha}{\lambda}$

NOTE:

- $(t, H(t)) = (t, -\log S(t))$ 그래프가 직선 $\Rightarrow$ 지수분포
- $(\log t, \log[-\log(S(t))])$ 그래프가 직선 $\Rightarrow$ 와이블분포