9장 모수적 회귀모형을 이용한 생존분석

2020년 가을학기

전북대학교 통계학과

Outline

모수적 비례위험 모형

가속실패모형

모수적 비례위험 모형

모수적 방법

Basic Idea
생존시간은 확률분포를 따름

Goal
확률분포의 모수 추정

$\Rightarrow$ 완전한 형태가 주어진 모형

$\Rightarrow$ 시간 분위수 추정

모수적 비례위험 모형

모수적 비례위험 모형	Cox 비례위험모형
$h(t), S(t)$에 대해 완전한 형태가 주어짐	$S(t)$의 형태를 모름
이론적 $S(t)$와 더 일치성을 가짐	이론적 $S(t)$와 일치성을 덜 가짐
시간-분위수 추정 가능	분포 가정에 의존하지 않음
생존시간에 대한 분포를 알아야 함	위험률(HR) 추정을 위해 기저위험률 필요 없음

모수적 비례위험 모형과 Cox 비례위험모형과의 비교

모수적 비례위험 모형

생존시간 $T \sim$ 확률밀도함수 $f(t)$

- 생존함수 $S(t) = P(T > t) = \int_t^{\infty} f(u)du$
- 위험함수 $h(t) = -\frac{\frac{d}{dt}S(t)}{S(t)}$ 누적위험함수 $H(t) = \int_0^t h(u)du$
- $S(t) = \exp(-H(t)), \quad f(t) = h(t)S(t)$

분포이름	확률밀도함수 $f(t)$	위험함수 $h(t)$	생존함수 $S(t)$	공변량 Z가 있는 경우 생존함수
지수분포 $\lambda > 0 t \geq 0$	$\lambda \exp(-\lambda t)$	λ	$\exp(-\lambda t)$	$\exp(-te^{\beta_0+\beta_1 Z})$
와이블분포 $\kappa > 0, \lambda > 0$ $t \geq 0$	$\kappa\lambda^\kappa t^{\kappa-1} \cdot \exp(-\lambda t)$	$\kappa\lambda^\kappa t^{\kappa-1}$	$\exp(-\lambda^\kappa t^\kappa)$	$\exp(-t^\kappa[e^{\beta_0+\beta_1 Z}]^\kappa)$
로그-로지스틱 분포 $\alpha > 0, \lambda > 0$	$\frac{\lambda\alpha t^{\alpha-1}}{(1+\lambda t^\alpha)^2}$	$\frac{\lambda\alpha t^{\alpha-1}}{(1+\lambda t^\alpha)}$	$\frac{1}{1+\lambda t^\alpha}$	$\frac{1}{1+t^\alpha e^{\beta_0+\beta_1 Z}}$

모수회귀에 많이 사용되는 분포

와이블 비례위험 모형

모수 κ, λ를 갖는 와이블분포 하에서 위험함수

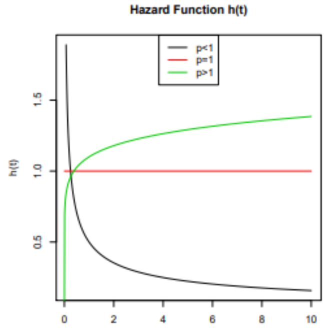
- 공변량이 없는 경우 위험함수 $h(t) = \kappa\lambda^{\kappa} t^{\kappa-1}$
- 공변량 $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_p)'$이 주어진 경우 $h(t|\boldsymbol{Z}) = h_0(t)g(\boldsymbol{Z}'\boldsymbol{\beta})$
  - 연결함수 $g(\cdot)$의 예 : $g_1(x) = e^x, \;\; g_2(x) = e^{-x}$

Weibull 분포의 성질 : $\log t$와 $\log(-\log S(t))$는 직선관계

$\Rightarrow$ $(\log t, \log(-\log \hat{S}(t))$ 그림으로 와이블모형 타당성 확인 ($\hat{S}(t)$: K-M 추정량)

κ : shape parameter

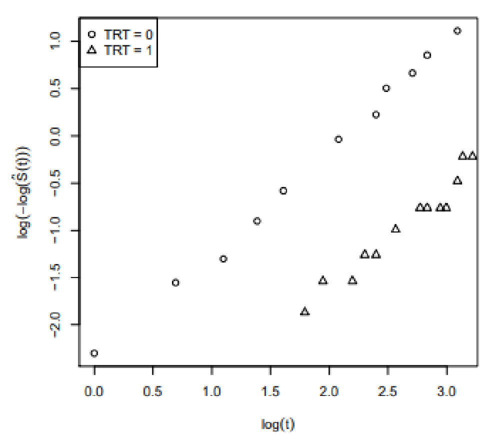
- $\kappa > 1$인 위험함수: 증가함수
- $\kappa = 1$인 위험함수: 상수함수
- $\kappa < 1$인 위험함수: 감소함수



Example: Remission Data

42명의 폐렴 환자의 회복시간 자료: MASS패키지의 gehan data

- 6-MP 처리를 받은 21명의 환자
- control 21명의 환자



- 선형성 ⇒ 와이블
- 동일한 기울기⇒ 비례위험

와이블 비례위험 모형

Recall $h(t) = \kappa\lambda^{\kappa} t^{\kappa-1}$

Weibull PH 모형
`eha` 패키지의 `phreg` 함수에서의 λ모형

- λ의 모형화

$$\lambda^{\kappa} = \exp(\beta_0 + \beta_1 \mathit{treat})$$

- 위험률 (treat=1 vs. treat=0)

$$HR = \frac{\exp[\beta_0 + \beta_1]\kappa t^{\kappa-1}}{\exp[\beta_0]\kappa t^{\kappa-1}} = \exp(\beta_1)$$

  PH가정 만족

$$\log h(t|\mathit{treat}) = (\kappa - 1)\log t + \beta_0 + \beta_1 \mathit{treat} + \log\kappa$$

$\log t$에 대해 절편 $\beta_0 + \beta_1 \mathit{treat} + \log\kappa$, 기울기 $\kappa - 1$인 직선

지수 비례위험 모형

지수분포: $\kappa = 1$인 와이블분포

- 공변량 treat이 주어진 경우 $h(t|treat) = \exp(\beta_0 + \beta_1 treat)$
- eha 패키지의 phreg 함수 이용

```
> phreg(Surv(time,cens) ~ treat, data=gehan2, dist='weibull', shape =1)
Call:
phreg(formula = Surv(time, cens) ~ treat, data = gehan2, dist = "weibull"
    shape = 1)

Covariate           W.mean      Coef Exp(Coef)  se(Coef)    Wald p
treat
         control     0.336     0         1          (reference)
            6-MP     0.664    -1.527     0.217     0.398     0.000

log(scale)                     2.159               0.218     0.000

 Shape is fixed at  1
```

추정된 위험률 (HR)

$$\hat{HR}(\text{treat=6-MP vs.treat=control}) = \exp(\hat{\beta}_1) = \exp(-1.527) = 0.22$$

해석 : 6-MP그룹의 위험률이 control그룹의 0.22배 ($p - value < 0.05$), 즉 treatment가 효과가 있음

가속실패모형

가속실패모형 (Accelerated Failure Time model: AFT 모형)

- 모수적인 접근 방법
- 위험률을 공변량 효과의 함수로 표현되는 비례위험모형 대안
- 생존시간 자체에 대한 공변량의 효과를 다룸

두 집단의 생존함수 $S_1(t),\ S_2(t)$ 비교

$$S_1(t) = S_2(\gamma t), \quad t \geq 0$$

- $\gamma > 0$: 과속화 인수 (acceleration factor)
- 모집단 1의 노화비율 (aging rate)이 모집단 2의 노화비율보다 γ배임을 의미

Example

- 개(dog)의 수명은 10∼15년
- 일반적으로 개의 1년은 사람의 7년에 해당
- $S_1(t)$: 사람의 생존함수, $S_2(t)$:개의 생존함수:

가속실패모형 (Accelerated Failure Time model: AFT 모형)

AFT 모형
생존시간의 확장 또는 축소를 공변량의 함수로 모형화

Smokers vs. nonsmokers
흡연자와 비흡연자의 생존함수를 각각 $S_S(t)$, $S_{NS}(t)$

AFT 가정

- 생존함수로 표현: $S_{NS}(t) = S_S(\gamma t), \ t \geq 0$
- 생존시간으로 표현 : $\gamma T_{NS} \stackrel{d}{=} T_S$

  여기에서 T_S, T_{NS}는 흡연자와 비흡연자의 생존시간

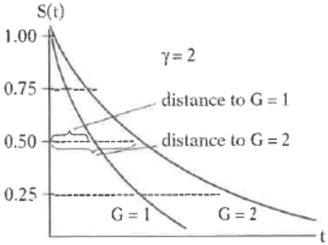
가속실패모형 (Accelerated Failure Time model: AFT 모형)

- μ_i : 모집단 i의 생존시간의 평균

$$\text{모집단 2의 평균}\mu_2 = \int_0^\infty S_2(t)dt = \gamma \int_0^\infty S_2(\gamma u)du, \quad (t = \gamma u)$$
$$= \gamma \int_0^\infty S_1(u)du = \gamma\mu_1$$

- ψ_i : 모집단 i의 θ 분위수(quantile), 즉 $S_i(\psi_i) = \theta$

$$S_2(\psi_2) = \theta = S_1(\psi_1) = S_2(\gamma\psi_1) \Rightarrow \psi_2 = \gamma\psi_1$$



[그림 9.1] 두 모집단에 대한 가속실패시간모형을 만족하는 생존함수

가속화 인수 γ의 해석

Acceleration factor

$$\gamma > 1 \Rightarrow \text{exposure benefits survival}$$

$$\gamma < 1 \Rightarrow \text{exposure harmful to survival}$$

Note Hazard ratio

$$HR > 1 \Rightarrow \text{exposure harmful to survival}$$

$$HR < 1 \Rightarrow \text{exposure benefits survival}$$

$$\gamma = HR = 1 \Rightarrow \text{no effect from exposure}$$

가속실패모형 (Accelerated Failure Time model: AFT 모형)

공변량 X를 가진 AFT 모형

$$\log(T) = \alpha_0 + \alpha_1 X + \epsilon$$

여기에서 ϵ는 오차

T	$\log T$
Exponential	Extreme value
Weibull	Extreme value
Log-logistic	Logistic
Lognormal	Normal

T가 와이블 분포를 따르는 경우 AFT 모형

$$\log(T) = \alpha_0 + \alpha_1 X + \sigma\epsilon$$

여기에서 $\sigma = 1/\kappa$에 해당하는 모수

$$T = \exp(\alpha_0 + \alpha_1 X + \sigma\epsilon) = \exp(\alpha_0)\exp(\alpha_1 X)\exp(\sigma\epsilon)$$

Weibull AFT 모형

- gehan data 사용하여 공변량이 treat인 Weibull AFT 모형 적합
- AFT 모형 가정 : 6-MP그룹과 control 그룹의 시간-분위수 비는 상수 γ

시간-분위수 (time-quantiles)의 표현

$$S(t) = \exp(-\lambda^\kappa t^\kappa) \Leftrightarrow -\log S(t) = (\lambda t)^\kappa$$
$$\Leftrightarrow t = \frac{(-\log S(t))^{1/\kappa}}{\lambda}$$

survival패키지의 survreg함수에서 λ 모형

- λ의 모형화 : $\lambda = \exp(-\alpha_0 - \alpha_1 treat)$

$$t = (-\log S(t))^{1/\kappa} \exp(\alpha_0 + \alpha_1 treat)$$

- q-분위수 : $S(t_q) = q \Leftrightarrow t_q = (-\log q)^{1/\kappa} \exp(\alpha_0 + \alpha_1 treat)$

$$\gamma = \gamma(\text{treat=6-MP vs. treat=control}) = \frac{(-\log q)^{1/\kappa} \exp(\alpha_0 + \alpha_1)}{(-\log q)^{1/\kappa} \exp(\alpha_0)} = \exp(\alpha_1)$$

R Code and R Output

패키지의 `survreg()`함수

- `dist` 옵션 : "weibull", "exponential", "gaussian", "logistic","lognormal", "loglogistic".

```
> weib.aft <- survreg(Surv(time,cens) ~ treat, data=gehan2, dist='weibull')
> summary(weib.aft)

Call:
survreg(formula = Surv(time, cens) ~ treat, data = gehan2, dist = "weibull")
             Value Std. Error     z       p
(Intercept)  2.248      0.166 13.55 < 2e-16
treat6-MP    1.267      0.311  4.08 4.5e-05
Log(scale)  -0.312      0.147 -2.12   0.034

Scale= 0.732

Weibull distribution
Loglik(model)= -106.6   Loglik(intercept only)= -116.4
        Chisq= 19.65 on 1 degrees of freedom, p= 9.3e-06
Number of Newton-Raphson Iterations: 5
n= 42
```

Weibull AFT 모형: Acceleration factor

가속화 인수 γ

$$\gamma = \gamma(\text{treat=6-MP vs. treat=control}) = \frac{(-\log q)^{1/\kappa}\exp(\alpha_0+\alpha_1)}{(-\log q)^{1/\kappa}\exp(\alpha_0)} = \exp(\alpha_1)$$

대조 그룹에 대해 처리 그룹을 비교할 때 추정된 가속화 인수 $\hat{\gamma}$

$$\hat{\gamma} = \exp(\hat{\alpha}_1)$$

```
> exp(weib.aft$coefficients[2])
treat6-MP
 3.551374
```

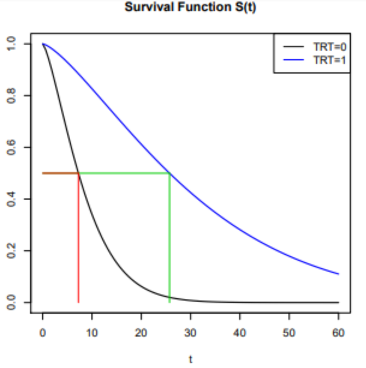
해석: 처리 그룹의 생존 시간은 대조 그룹에 비하여 인수 3.55로 증가$\Rightarrow$ 6-MP는 positive 효과가 있다.

생존시간의 중앙값, 즉 0.5-분위수

$$S(t) = 0.5 \;\Rightarrow\; \hat{t}_{0.5} = (-\log 0.5)^{1/\hat{\kappa}} \exp(\hat{\alpha}_0 + \hat{\alpha}_1 \, treat)$$

중앙값의 추정

```
> median <-predict(weib.aft,
+                newdata=list(treat=c("contro
+                type='quantile',p=0.5)
> median
       1         2
 7.242697 25.721526
> median[2]/median[1]
      2
3.551374
```



Weibull AFT와 PH 계수의 관계

- AFT : $\lambda = \exp(-\alpha_0 - \alpha_1 treat) \Rightarrow \log\lambda = -\alpha_0 - \alpha_1 treat$
- PH : $\lambda^\kappa = \exp(\beta_0 + \beta_1 treat) \Rightarrow \log\lambda = (\beta_0 + \beta_1 treat)/\kappa$

`eha` 패키지의 `phreg()`함수

- `dist` 옵션 : "weibull", "gompertz", "pch", "lognormal", "loglogistic"

  exponential은 `dist="weibull", scale=1`

R Output

```
> weib.ph <- phreg(Surv(time,cens) ~ treat, data=gehan2, dist='weibull')
> summary(weib.ph)
Call:
phreg(formula = Surv(time, cens) ~ treat, data = gehan2, dist = "weibull")

Covariate          W.mean      Coef Exp(Coef)  se(Coef)    Wald p
treat
        control     0.336      0         1          (reference)
           6-MP     0.664    -1.731      0.177      0.413      0.000

log(scale)                    2.248                 0.166      0.000
log(shape)                    0.312                 0.147      0.034
```

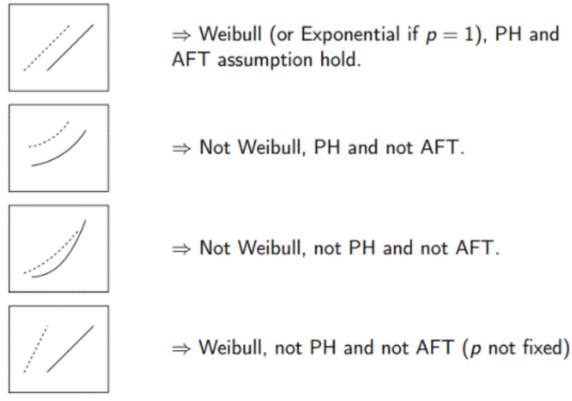
- $\beta_1/\kappa = -\alpha_1 \Rightarrow \beta_1 = -\alpha_1\kappa$
- `survreg` 결과에서 `Scale`$=1/\kappa \Rightarrow 1.731 = 1.267/0.732$
- AFT 가정 $\Leftrightarrow$ PH 가정

그림을 이용한 모형 적합성에 대한 평가

Kaplan-Meier 생존함수 $\hat{S}(t)$ 그림으로 평가

- $(t, -\log \hat{S}(t))$ 그림
  직선 $\Rightarrow$ 지수분포
- $(\log t, \log[-\log \hat{S}(t)])$ 그림
  직선 $\Rightarrow$ 와이블분포

범주형 공변량의 경우 각 공변량 범주에 대해 Kaplan-Meier 생존함수를 그림으로 평가



$(\log t, \log(-\log \hat{S}(t))$ 그림

모형적합도 통계량

- AIC (Akaike Information Criterion) $=$ $-2 \times \log(\text{maximum likelihood}) + 2 \times p$
    - $p =$ 적합모형의 모수 개수
    - AIC는 모수가 많을수록 AIC 값이 커지는 벌점(penalty)효과를 가짐
    - AIC를 최소로 하는 모형이 더 나은 모형으로 추천됨
- BIC (Bayesian Information Criterion) $=$ $-2 \times \log(\text{maximum likelihood}) + (\log n) \times p$
    - BIC를 최소로 하는 모형이 더 나음 모형으로 추천됨