[12강]

가설검정(2)

정보통계학과 이긍희 교수

학습목표

1. 가능도비검정을 이해한다.
2. 분할표에 대한 카이제곱검정을 이해한다.
3. 유의성검정을 이해한다.

[12강]

가설검정(2)

1 가능도검정

1.1 최강력 검정

- **단순가설** : $H_0 : \theta = \theta_0 \;\; vs \; H_1 : \theta = \theta_1$
- 최강력검정의 기각역(R)

$$R = \left\{ \boldsymbol{x} : \frac{f(\boldsymbol{x}|\theta_1)}{f(\boldsymbol{x}|\theta_0)} > k \right\} \leftrightarrow R = \left\{ \boldsymbol{x} : \frac{f(\boldsymbol{x}|\theta_0)}{f(\boldsymbol{x}|\theta_1)} < k' \right\}$$

1.2 복합가설 하의 가설검정

- $H_0: \theta \in \Omega_0 \;\; vs \;\; H_1: \theta \in \Omega_1$
  - Ω_0와 Ω_1은 서로 배반인 모수 θ의 집합
- $N(\theta, 1)$의 가설 $\rightarrow H_0: \theta = 0 \;\; vs \;\; H_1: \theta \neq 0$
  - $\Omega_0 = \{\theta : \theta = 0\}, \; \Omega_1 = \{\theta : \theta \neq 0\}$

1.2 복합가설 하의 가설검정

- 복합가설 하의 확률밀도함수가 하나로 결정되지 않음

→ 확률밀도함수 비를 바탕으로 한 최강력검정의 사용 제약

- 하나로 결정되지 않는 확률밀도함수 대신 각 가설 하 최대가능도를 이용

1.3 최대가능도비

- 귀무가설 하 θ의 최대가능도추정량 $\hat{\theta}_0$

$$f\left(\boldsymbol{x}\middle|\hat{\theta}_0\right) = \max_{\theta\in\Omega_0} f(\boldsymbol{x}|\theta)$$

- 대립가설 하 θ의 최대가능도추정량 $\hat{\theta}_1$

$$f\left(\boldsymbol{x}\middle|\hat{\theta}_1\right) = \max_{\theta\in\Omega_1} f(\boldsymbol{x}|\theta)$$

1.4 가능도비검정

- 가능도비검정(likelihood ratio test) : 귀무가설 하 최대가능도와 모수 전체에서 구한 최대가능도의 비에 의하여 기각역이 정해지는 검정
- 최대가능도 비를 사용한 기각역 R

$$R = \left\{ \boldsymbol{x} : \frac{f\left(\boldsymbol{x} \middle| \hat{\theta}_0\right)}{f\left(\boldsymbol{x} \middle| \hat{\theta}_1\right)} < k' \right\}$$

1.4 가능도비검정

- 기각역의 재표현($k' < 1$)

$$R = \left\{ \boldsymbol{x} : \frac{f(\boldsymbol{x}|\hat{\theta}_0)}{f(\boldsymbol{x}|\hat{\theta}_1)} < k' \right\}$$

- 상수 k' : 주어진 유의수준 α에 따라 결정

1.4 가능도비검정

- 가능도비 검정 : 귀무가설 $(H_0: \theta \in \Omega_0)$ 하 최대가능도와 모수 전체 $(\Omega = \Omega_0 \cup \Omega_1)$의 최대가능도의 비에 의하여 기각역이 정해지는 검정

$$R = \left\{ \boldsymbol{x} : \frac{\max_{\theta \in \Omega_0} f(\boldsymbol{x}|\theta)}{\max_{\theta \in \Omega_0 \cup \Omega_1} f(\boldsymbol{x}|\theta)} < k \right\}$$

1.4 가능도비검정

예 7.6

$X_1, \cdots, X_n \sim N(\theta, 1)$확률표본

$H_0: \theta = \theta_0 \;\; vs \; H_1: \theta \neq \theta_0$에 대한 유의수준 α에서 가능도비검정을 구하라.

1.4 가능도비검정

예 7.6

$X_1, \cdots, X_n \sim N(\theta, 1)$확률표본

$H_0 : \theta = \theta_0 \;\; vs \; H_1 : \theta \neq \theta_0$에 대한 유의수준 α에서 가능도비검정을 구하라.

1.4 가능도비검정

예 7.7

$X_1, \cdots, X_n \sim N(\mu, \sigma^2)$ 확률표본
$H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$에 대한 유의수준 α에서
가능도비검정을 구하라.

1.4 가능도비검정

예 7.7

$H_0: \mu = \mu_0 \;\; vs \; H_1: \mu \neq \mu_0$에 대한 유의수준 α에서 가능도비검정을 구하라.

1.4 가능도비검정

예 7.7

$H_0: \mu = \mu_0 \;\; vs \; H_1: \mu \neq \mu_0$에 대한 유의수준 α에서 가능도비검정을 구하라.

1.5 가능도비 검정의 기각역

- 귀무가설이 참일 때 가능도비의 로그변환된 식의 근사적 분포 : 카이제곱 (χ^2)분포

$$-2\log\frac{f\left(\boldsymbol{X}\middle|\hat{\theta}_0\right)}{f\left(\boldsymbol{X}\middle|\hat{\theta}\right)} \sim \chi^2(d.f.)$$

  - 자유도 $d.f.$: (모수 전체의 영역에서 추정하는 모수의 수)-(귀무가설이 참인 영역에서 추정하는 모수의 수)

1.5 가능도비 검정의 기각역

- 귀무가설 하 가능도비 식의 정확한 확률분포를 알지 못하는 경우 : 카이제곱(χ^2)분포를 이용하여 가능도비검정의 기각역을 구할 수 있음

$$R = \{\boldsymbol{x} | -2\log\frac{f(\boldsymbol{x}|\hat{\theta}_0)}{f(\boldsymbol{x}|\hat{\theta})} > \chi^2_\alpha(d.f.)\}$$

[12강]

가설검정(2)

2 카이제곱검정

2.1 분할표 검정

- m개 범주 빈도수 : $N_1, \cdots, N_m$

	범주1	범주2	⋯	범주m	합계
X	N_1	N_2	⋯	N_m	N
확률	p_1	p_2	⋯	p_m	1

2.1 분할표 검정

- 전체 빈도수 N, 빈도수의 분포 : 다항분포

$$f(n_1, \cdots, n_m | p_1, \cdots, p_m) = \frac{n!}{n_1! \cdots n_m!} p_1^{n_1} \cdots p_m^{n_m}$$

2.1 분할표 검정

- $H_0: (p_1, \cdots, p_n) \in \Omega_0 \;\; vs \;\; H_1: (p_1, \cdots, p_n) \in \Omega_1$

  모수 전체 영역 : p_i의 최대가능도추정량 $\hat{p}_i = \frac{n_i}{n}$

2.1 분할표 검정

- 가능도비 : $\dfrac{f(n_1,\cdots,n_m|\hat{p}_{10},\cdots,\hat{p}_{m0})}{f(n_1,\cdots,n_m|\hat{p}_1,\cdots,\hat{p}_m)}$

$$= \left(\frac{\hat{p}_{10}}{\hat{p}_1}\right)^{n_1} \cdots \left(\frac{\hat{p}_{m0}}{\hat{p}_m}\right)^{n_m}$$

$$-2\log\frac{f(n_1,\cdots,n_m|\hat{p}_{10},\cdots,\hat{p}_{m0})}{f(n_1,\cdots,n_m|\hat{p}_1,\cdots,\hat{p}_m)} = 2\sum_{i=1}^{m} n_i \log\left(\frac{n_i}{N\hat{p}_{i0}}\right)$$

2.1 분할표 검정

- 유의수준 α 가능도비검정의 기각역 :

$$2\sum_{i=1}^{m} n_i \log\left(\frac{N_i}{n\hat{p}_{i0}}\right) > \chi_\alpha^2(d.f.)$$

- 테일러 급수를 이용한 근사

$$\sum_{i=1}^{m} \frac{(N_i - n\hat{p}_{i0})^2}{n\hat{p}_{i0}} > \chi_\alpha^2(d.f.)$$

2.2 카이제곱검정

- $H_0: (p_1, \cdots, p_m) \in \Omega_0 \;\; vs \;\; H_1: (p_1, \cdots, p_m) \in \Omega_1$
- 테일러 급수 근사 : $\sum_{i=1}^{m} \frac{(N_i - n\hat{p}_{io})^2}{n\hat{p}_{i0}} > \chi_\alpha^2(d.f.)$
  - N_i : i번째 범주 관측빈도수
  - $n\hat{p}_{i0}$: 귀무가설이 참일 때 기대빈도수

2.3 적합도 검정

- $H_0 : p_i = p_{i0} \;\; vs \;\; H_1 : \sim H_0$

  N_i : 관측빈도수, np_{i0} :기대빈도수

- 검정통계량 $\chi^2 = \sum_{i=1}^{m} \frac{(N_i - np_{i0})^2}{np_{i0}} \sim \chi^2(m-1)$,

  $np_{i0} \geq 5$ 인 경우 사용 가능

- 검정법 : $\chi^2 > c$ 면 H_0 기각, $\Pr(\chi^2 > c | H_0) = \alpha$

2.4 독립성 검정

▪ 분할표

		Y				합계
		범주 1	범주 2	⋯	범주 c	
X	범주 1	N_{11}	N_{12}	⋯	N_{1c}	N_{1+}
	범주 2	N_{21}	N_{22}	⋯	N_{2c}	N_{2+}
	⋮	⋮	⋮	⋮	⋮	⋮
	범주 r	N_{r1}	N_{r2}	⋯	N_{rc}	N_{r+}
합계		N_{+1}	N_{+2}	⋯	N_{+c}	N_{++}

2.4 독립성 검정

- $r \times c$ 분할표, 각 셀의 확률 p_{ij}
  - $p_{i+} = \sum_j p_{ij}\,,\; p_{+j} = \sum_i p_{ij}\,,$
  - $N_{i+} = \sum_j N_{ij}\,,\; N_{+j} = \sum_i N_{ij}\,,\; N = \sum_i \sum_j N_{ij}$

2.4 독립성 검정

- 귀무가설 : 행변수와 열변수가 서로 독립

$$\Leftrightarrow H_0: p_{ij} = p_{i+} \cdot p_{+j},\ \ H_1: \sim H_0$$

- $\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(N_{ij} - \hat{B}_{ij})^2}{\hat{B}_{ij}} \sim \chi^2((r-1)(c-1))$,

$$\hat{B}_{ij} = n \cdot \hat{p}_{i+} \cdot \hat{p}_{+j} = \frac{N_{i+} \cdot N_{+j}}{N}$$

2.4 독립성 검정

- $\chi^2 > \chi^2_\alpha\big((r-1)(c-1)\big)$이면 H_0 기각

[12강]

가설검정(2)

3 유의성검정

3.1 통계적 가설검정의 역사

- 피셔의 유의성 검정 : 귀무가설에 대한 p-값 이용
- 네이만과 피어슨의 가설검정 : 귀무가설과 대립가설에 대하여 제1종 오류를 범할 확률과 제2종 오류를 범할 확률에 기초한 방법

3.2 피셔의 유의성 검정

- p-값 : 귀무가설 하에서 주어진 관측값보다 더 극단적인 값을 얻을 확률 → 귀무가설 반대 증거
- 귀무가설만 설정, 주어진 관측값이 이 가설에 얼마나 부합하는지 알아보는 것

3.3 네이만과 피어슨의 검정

- 귀무가설과 대립가설을 설정
  - 제1종 오류를 범할 확률 α, 제 2종 오류를 범할 확률 β와 검정력에 대한 개념 도입
- 주어진 α에 대하여 대립가설을 고려하여 최적의 기각역을 구하는 방법

[12강]

가설검정(2)

4 검정함수

4.1 검정함수

- $$\delta(x) = \begin{cases} 1, & \text{귀무가설 기각} \\ 0, & \text{귀무가설 기각하지 못함} \end{cases}$$

4.1 검정함수

- 제1종 오류를 범할 확률

$$\alpha = P(X \in R|H_0) = E[\delta(X)|H_0]$$

- 검정력 :$1 - \beta = E[\delta(X)|H_1]$

4.1 검정함수

- 연속형 확률분포 : 제1종 오류를 범할 확률이 주어진 유의수준 α와 일치하는 기각역을 가지는 검정함수를 정의
- 이산형 확률분포 : 검정함수를 일반화하여 정의

$$\delta(X) = \begin{cases} 1, & X = 0,1 \\ 0.5, & X = 2 \\ 0, & X = 3,4 \end{cases}$$

[12강]

가설검정(2)

5 정리하기

정리하기

1. 가능도비검정(likelihood ratio test)은 귀무가설 하에서의 최대가능도와 모수 전체에서 구한 최대가능도의 비에 의하여 기각역이 정해지는 검정이다.
2. 가능도비검정을 이용하여 각 범주에서 빈도수로 주어지는 분할표에 대한 적합도검정, 동질성검정, 독립성 검정에 대한 카이제곱 (χ^2) 검정을 유도할 수 있다.

정리하기

3. 피셔(R.A. Fisher)의 유의성검정은 귀무가설만 설정하고, 주어진 관측값이 이 가설에 얼마나 부합하는지 알아보고자 한 검정이다.

다음 시간 안내 ▼

13강. 구간추정

수고하셨습니다.