ch4. 피셔 추론과 최대 우도 추정

Ji Seung Ryu

7월 10, 2019

4.1 우도와 최대 우도

Likelihood(우도)

- $(X_1, X_2, X_3, ..., X_n) \sim f(x_1, x_2, x_3, ..., x_n|\theta), \theta \in \Omega$
- $L(\theta|x_1, x_2, x_3, ..., x_n) = f(x_1, x_2, x_3, ..., x_n|\theta), \theta \in \Omega$

  $= L_X(\theta)$ for brevity, $\theta \in \Omega$
- $l_X(\theta) = log\{f_\theta(x)\}$

관측치 x가 고정된 상태에서 모수 θ가 변화된다.

단일 값일 경우에는 θ를 사용하고 모수 벡터일 경우에는 μ를 사용한다.

확률 밀도 함수를 랜덤변수의 값 x 의 함수가 아닌 파라미터 θ 의 함수로 보는 것

확률 분포로부터 특정한 샘플 값 x 가 발생하였을 때, 이 샘플 값 x 가 나오게 하는 파라미터 θ 의 가능성

확률 분포로부터 특정한 샘플 값 x 가 발생하였을 때, 샘플 값 x 와 변수 θ 에서의는 확률(밀도함수)

확률 밀도 함수

$$f_X(x; \theta)$$

- θ 값을 이미 알고 있음
- θ 는 상수, x 는 변수
- θ 가 이미 정해져 있는 상황에서의 x 값의 상대적 가능성

Likelihood(우도)

$$L(\theta) = f_X(x|\theta)$$

- x 가 이미 발생. 값을 이미 알고 있음
- x 는 상수, θ 는 변수
- x 가 이미 정해져 있는 상황에서의 θ 값의 상대적 가능성

MLE(최대 우도 추정)

로그 우도 함수 = $logL_\theta(x)$

보통 계산하기 간편한 로그 우도함수를 많이 사용한다.

$$MLE : \hat{\theta} = argmax_{\theta \in \Omega} l_x(\theta)$$

MLE란 모수 공간 Ω에서 $l_x(\theta)$를 최대화 하는 θ 값이다.

1. 최대 우도 추정은 전통적 응용 추정에서 대세적인 기법이며, 새로운 현상을 접할 때 가장 먼저 적용해보는 기법이다.

- MLE알고리즘은 자동화되어 있다.
- MLE는 빈도주의가 가진 뛰어난 성질을 잘 활용한다.(대규모 표본 : 불편 추정치, 최소한의 분산 / 소규모 표본 : 효율적, 최상위급 성능)
- MLE는 베이즈주의 정당성을 가지고 있다. (베이즈 방식에서 사전함수가 균등, 상수일 경우)

2. 최대 우도 추정에서는 개별적으로 최대 우도는 잘 구하지만 다차원일 경우에는 잘못 구해질 수 있다.

- 전립선 예제

4.2 피셔 정보와 MLE

단일 모수 밀도 패밀리

$$MLE : \hat{\theta} = argmax_{\theta \in \Omega} \{l_x(\theta)\}$$

$F = \{f_\theta(x), \theta \in \Omega, x \in X\}$

스코어 함수 $\dot{l}_x(\theta) = \frac{\partial}{\partial \theta} log f_\theta(x) = \frac{\dot{f}_\theta(x)}{f_\theta(x)}$ (점 표기는 미분을 뜻한다.)

$E(\dot{l}_x(\theta)) = 0, Var(\dot{l}_x(\theta)) = I_\theta$ (정의)

즉, $\dot{l}_x(\theta)$ ~ $(0, I_\theta)$ => MLE $\hat{\theta} \dot{\sim} N(\theta, 1/I_\theta)$

$\hat{\theta}$의 분산이 $1/I_\theta$에 근사하는 것 증명 - The asymptotic distribution of MLE (http://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/Fisher_info.pdf)

θ는 거의 불편향이고 큰 피셔정보라는 것은 더 작은 MLE분산을 암시한다.

$\ddot{l}_x(\theta) = \frac{\partial^2}{\partial \theta^2} log f_\theta(x) = \frac{\ddot{f}_\theta(x)}{f_\theta(x)} - (\frac{\dot{f}_\theta(x)}{f_\theta(x)})^2$

$E_\theta\{\ddot{l}_x(\theta)\} = -I_\theta, Var_\theta\{\ddot{l}_x(\theta)\} = J_\theta$

즉, $\ddot{l}_x(\theta)$ ~ (I_θ, J_θ)

iid 표본의 예

$x = (x_1, x_1, \dots, x_n) \sim^{iid} f_\theta(x)$

$\dot{l}_x(\theta) = \sum_{i=1}^{n} \dot{l}_{x_i}(\theta)$

$-\ddot{l}_x(\theta) = \sum_{i=1}^{n} -\ddot{l}_{x_i}(\theta)$

전체 표본 x에 기반한 MLE θ는 최대화 조건 $\dot{l}_x(\hat{\theta}) = 0$을 만족한다.

일차 테일러 급수 : $\hat{\theta} \doteq \theta + \frac{\dot{l}_x(\theta)/n}{-\ddot{l}_x(\theta)/n}$

중심극한정리 : $\dot{l}_x(\theta)/n \dot{\sim} N(0, I_\theta/n)$

$Var(\sum_{i=1}^{n} \dot{l}_x(\theta)/n) = \frac{1}{n^2} nVar(\dot{l}_x(\theta)) = \frac{1}{n} Var(\dot{l}_x(\theta)) = I_\theta/n$

대수의 법칙 : $-\ddot{l}_x(\theta)/n \dot{\sim} I_\theta$

$E(\sum_{i=1}^{n} -\ddot{l}_x(\theta)/n) = \frac{1}{n} nE(-\ddot{l}_x(\theta)) = E(-\ddot{l}_x(\theta)) = I_\theta$

즉, iid 대규모 표본에서는 $\hat{\theta} \dot{\sim} N(\theta, 1/(nI_\theta))$ <= iid 표본의 전체 피셔 정보는 nI_θ

정규 표본의 예

i = 1, 2, …, n에 대해 $x_i \sim^{iid} N(\theta, \sigma^2)$

만약 σ^2 를 알 수 있다면, $l_x(\theta) = -\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \theta)^2}{\sigma^2} - \frac{n}{2} log(2\pi\sigma^2)$

$\dot{l}_x(\theta) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \theta)$이고 $- \ddot{l}_x(\theta) = \frac{n}{\sigma^2}$

즉, $\hat{\theta} \sim N(\theta, \sigma^2/n)$ <= $I_\theta = 1/\sigma^2$

- 피셔의 주된 전술은 논리적으로 복잡한 추론 문제를 간단한 형태로 줄여서 누구에게나 명백한 답을 찾는 것이었다.
- 단일 스칼라 관측치 $\hat{\theta}$가 미지의 관심 대상인 모수인 θ 근처를 알려진 분산 σ^2/n으로 정규분포하도록 만드는 것이다.

크레이머-라오 경계

$$Var_\theta\{\tilde{\theta}\} \geq 1/(nI_\theta)$$

코시 슈바르츠 부등식에 의해 증명됨

MLE는 적어도 θ의 최적 불편 추정만큼 작은 분산을 가진다.

더 자세히 알아봅시다 (http://m.blog.daum.net/gongdjn/122?np_nil_b=2)

4.3 조건부 추론

간단한 예제

다음의 $x_i \sim^{iid} N(0, 1), i = 1, 2, \ldots, n$ iid 표본이 추정치 $\hat{\theta} = \bar{x}$를 생성했다.

표본크기 n에 대한 의견에 서로 동의하지 못해서 동전을 던져서 결정하기로 했다.

$$n = \begin{cases} 25 & \text{확률 } 1/2 \\ 100 & \text{확률 } 1/2 \end{cases}$$

동전 던지기 결과 n = 25가 선정.

여기서! $\bar{x}$의 표준편차는 얼마일까?

조건부 추론의 추종자 : $1/(25)^{0.5} = 0.2$

빈도주의 추종자 : $barx$는 동일한 확률로 N(0,1/100) 또는 N(0,1/25)가 될 수 있기 때문에 표준편차는 $[(0.01 + 0.04)/2]^{0.5}$ = 0.158

피셔가 조건부 추론을 주장하는 근거 : 좀 더 연관된 추론, 더 간단한 추론

- 최대 우도의 정확도 : $nI_{\hat{\theta}}$보다 $I(x)$를 이용하는 것이 $\hat{\theta}$에 더 정확한 결론을 안겨준다. ex) 코시밀도 시뮬레이션(조건화 + 경험적 분산)

4.4 순열과 랜덤화(Permutation Test 순열검정법)

- 피셔가 직면한 비판 : 정규분포 표본이라는 가정에 지나치게 의존하고 있다.
- 피셔는 순열과 랜덤화를 주장

ex)유전자 폐렴 예제

귀무가설 : A와 B집단이 차이가 없다.

아이디어 : 만약, A와 B집단에 차이가 없다면 A집단과 B집단을 무작위로 섞어도 차이가 없을 것이다.

[R 예시]https://bioinformaticsandme.tistory.com/10 (https://bioinformaticsandme.tistory.com/10)