

1. 선형모형

$$y_1 = \theta + \epsilon_1, \quad (1)$$

$$y_2 = 2\theta - \tau + \epsilon_2, \quad (2)$$

$$y_3 = \theta + 2\tau - \epsilon_3, \quad (3)$$

을 고려하자. 여기서, ϵ_1, ϵ_2 , 그리고 ϵ_3 는 서로 독립이고 평균이 0, 분산이 σ^2 인 정규분포를 따른다고 가정하자.

(a) $y_1 = 0, y_2 = 0$, 그리고 $y_3 = 1$ 으로 측정되었을 때, 모수 θ 의 최소제곱추정량(least squares estimator)을 구하시오. (15점)

(b) 귀무가설 $H_0 : \theta = \tau$ 을 검정하는 방법을 설명하시오. (10점)

Solution:

(a) $-\epsilon_3$ 은 ϵ_3 과 같은 분포를 가진다. 따라서 (3)은 $-\epsilon_3$ 을 ϵ_3 로 바꿔도 무방하다. 위 모형을 행렬꼴로 바꾸자.

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}}_y = \underbrace{\begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \theta \\ \tau \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}}_{\epsilon} \quad (4)$$

이제 다음의 일반적인 최소제곱법을 풀면

$$\hat{\beta}^{\text{LSE}} = \underset{\beta}{\operatorname{argmin}} (y - \mathbf{X}\beta)' (y - \mathbf{X}\beta) \quad (5)$$

모두가 알고 있는 정규식이 나온다. $\hat{\beta}^{\text{LSE}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'y$. 이를 대입하면

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{bmatrix} \quad (6)$$

$$= \begin{bmatrix} 6 & 0 \\ 0 & 5 \end{bmatrix} \quad (7)$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1/6 & 0 \\ 0 & 1/5 \end{bmatrix} \quad (8)$$

$$\mathbf{X}'y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (9)$$

$$= \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad (10)$$

$$\hat{\beta}^{\text{LSE}} = \begin{bmatrix} 1/6 \\ 2/5 \end{bmatrix} \quad (11)$$

(b) 귀무가설이 참이라고 가정할 때 모형은

$$y_1 = \theta + \epsilon_1 \quad (12)$$

$$y_2 = \theta + \epsilon_2 \quad (13)$$

$$y_3 = 3\theta + \epsilon_3 \quad (14)$$

이 되고 (12)와 (13)은 동일하므로 둘 중 하나를 없앤다. 이렇게 적합한 모형의 RSS(Residual Sum of Squares; SSE라고도 함)와 원래 모형의 RSS를 비교한다. 자세한 사항은 2009년 전기 참조.

2. P 가 확률측도(probability measure)가 되기 위한 3가지 조건을 제시하고, 이 3가지 조건을 이용하여 표본공간의 임의 집합 A 에 대해 $0 \leq P(A) \leq 1$ 임을 증명하라.

Solution: 일반적으로 임의의 집합에 대한 함수(set function) μ 가 ‘측도(measure)’가 되기 위해 서는 다음의 세 조건을 만족해야 한다.

- Non-negativity: 정의된 σ -algebra \mathcal{M} 에 대해 모든 $A \in \mathcal{M}$ 는 $\mu(A) \geq 0$ 여야 한다.

- Null empty set: 공집합의 측도는 0이어야 한다. $\mu(\emptyset) = 0$
- Countable additivity: σ -algebra \mathcal{M} 에 속하는 서로소(pairwise disjoint)인 집합 A_1, A_2, \dots 에 대해 다음이 만족해야 한다.

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k) \quad (15)$$

확률측도 역시 ‘측도’이므로 위 세가지 조건을 만족해야만 한다. 하지만 확률측도는 임의의 집합 X 로부터 만들어진 σ -algebra위에 정의된 것이 아니라 전체집합 X 의 측도가 1이 되도록 설계된 것이므로 그 조건이 추가되어야 한다. 이를 20세기 초 러시아의 수학자 *Andrey Kolmogorov*는 다음과 같이 정의했고 이를 우리는 ‘확률공리(probability axioms)’라 부른다.

- 측도공간(measure space) (Ω, \mathcal{F}, P) 이 다음의 세 조건을 만족하면 이를 ‘확률공간(probability space)’ 혹은 *probability triple*이라 부른다. 첫 번째 공리는 마찬가지로 non-negativity 이다. $\forall A \in \mathcal{F} \ P(A) \geq 0$

영국의 수학자 *Paul Dirac*은 negative probability를 제시했으나 양자역학을 제외한 분야에서는 널리 쓰이지 않는다.

- 공집합의 확률이 0이라는 것 대신 전체집합 Ω 의 확률이 1임을 주게 되면 공집합의 확률이 0임은 하나의 정리로 증명할 수 있다. 고로 확률공리는 전체집합을 상정한다. $P(\Omega) = 1$
- 마지막은 countable additivity로 동일하다. σ -field \mathcal{F} 에 속하는 서로소인 집합 A_1, A_2, \dots 에 대해

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) \quad (16)$$

여야 한다.

모든 집합의 확률이 $[0, 1]$ 사이에 있다는 것을 증명하기 위해서는 몇 번의 단계를 거쳐야 한다. 직관적으로 보기에는 전체집합(표본공간)의 확률이 1이고 모든 확률이 0보다 크니 당연해 보이지만 공리를 통해 증명해야 한다. 그 첫 번째는 여집합의 확률이다.

- (1) $A \in \mathcal{F}$ 인 집합에 대해 그 여집합인 A^c 의 확률은 다음의 관계로부터 얻어진다.

$$A \cup A^c = \Omega \quad (17)$$

그러면 세 번째 공리로부터

$$P(A) + P(A^c) = 1 \quad (18)$$

이 되고 $P(A^c) = 1 - P(A)$ 을 얻을 수 있다.

- (2) 그 다음으로는 확률의 대소 관계이다. $A \subset B \subset \Omega$ 라면 $B = A \cup (B \setminus A)$ 라는 관계가 성립하고 또 다시 세 번째 공리에 의해

$$P(B) = P(A) + P(B \setminus A) \quad (19)$$

가 된다. 모든 확률은 0보다 크거나 같은 첫 번째 공리에 의해 $P(B) \geq P(A)$ 이 된다.

사실 이 성질은 확률측도에만 해당되는 것은 아니고 일반적인 *Lebesgue measure*의 성질이며 이를 ‘측도의 단조성(monotonicity of measures)’이라 부르기도 한다.

- (3) 고로 σ -field \mathcal{F} 에 속하는 임의의 집합 A 에 대해서 $A \subset \Omega$ 이므로 바로 전에 증명한 대소관계를 통해

$$P(A) \leq 1 \quad (20)$$

이 얻어진다. 따라서 $0 \leq P(A) \leq 1$ 이 된다. 증명끝.

3. Let X_1, X_2, \dots, X_n be a random sample with a Gamma distribution with parameters α and β , whose probability density function is given by

$$f(x | \alpha, \beta) = \frac{1}{\Gamma(\alpha) \beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), \quad x > 0. \quad (21)$$

- (a) Find a distribution of $\bar{X} = n^{-1} \sum_{i=1}^n X_i$.
- (b) Suppose that $\alpha = 1$. Determine a constant c such that $c\bar{X}$ has a χ^2 distribution with degrees of freedom n , that is, $c\bar{X} \sim \chi^2(2n)$.
- (c) Suppose that $\alpha = 2$. Find the MLE(maximum likelihood estimator) $\hat{\beta}_n$ of β , and derive the asymptotic variance of $\hat{\beta}_n$ using the Fisher information.

Solution:

- (a) 해당 문제에서 쓰인 표기방식은 scale parameter이기 때문에 지금까지 써왔던 rate parameter 표기법과는 조금 다르다. 약간 헷갈릴 수도 있겠지만 본 문제의 표기법을 따르겠다. 모든 확률변수 $X_i \sim \text{Ga}(\alpha, \beta)$ 라면 $\sum_{i=1}^n X_i \sim \text{Ga}(n\alpha, \beta)$ 이 되며 n^{-1} 을 곱해주면 scale parameter 표기법으로 하면 뒤의 scale parameter인 β 에 똑같이 곱해지므로

$$\bar{X} \sim \text{Ga}(n\alpha, n^{-1}\beta) \quad (22)$$

이 된다.

(b) $\chi^2(2n)$ 분포는 $\text{Ga}(n, 2)$ 와 동일하므로

$$c\bar{X} \sim \text{Ga}\left(n, \frac{c}{n}\beta\right) \quad (23)$$

를 통해

$$c = \frac{2n}{\beta} \quad (24)$$

가 된다.

(c) 귀찮다.

$$L(\beta | \{X_i\}_{i=1}^n) = \left(\frac{1}{\Gamma(2)\beta^2}\right)^n \left(\prod_{i=1}^n X_i\right) \exp\left(-\frac{1}{\beta} \sum_{i=1}^n X_i\right) \quad (25)$$

$$\ell(\beta | \{X_i\}_{i=1}^n) = -n \ln \Gamma(2) - 2n \ln \beta + \sum_{i=1}^n \ln X_i - \frac{1}{\beta} \sum_{i=1}^n X_i \quad (26)$$

$$\frac{d}{d\beta} \ell(\beta | \{X_i\}_{i=1}^n) = -\frac{2n}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^n X_i = 0 \quad (27)$$

$$\hat{\beta}^{\text{MLE}} = \frac{1}{2n} \sum_{i=1}^n X_i \quad (28)$$

피셔의 정보량은 두번 미분해야 하므로

$$\frac{d^2}{d\beta^2} \ell(\beta | \{X_i\}_{i=1}^n) = \frac{2n}{\beta^2} - \frac{2}{\beta^3} \sum_{i=1}^n X_i \quad (29)$$

$$-E \left[\frac{d^2}{d\beta^2} \ell(\beta | \{X_i\}_{i=1}^n) \right] = -\frac{2n}{\beta^2} + \frac{2}{\beta^3} \cdot 2n\beta \quad (30)$$

$$= \frac{2n}{\beta^2} \quad (31)$$

따라서 최대가능도추정량의 점근분포는 다음과 같다.

$$\sqrt{n} \left(\beta - \hat{\beta}^{\text{MLE}} \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{\beta^2}{2n} \right) \quad (32)$$

4. Let X_1, X_2, \dots, X_n be an i.i.d. sample from $\text{Unif}(0, \theta)$, $\theta > 0$.

(a) Find the MLE $\hat{\theta}_n$ of θ , and compute the MSE (mean squared error) of $\hat{\theta}_n$.

(b) Let $T_n = 2\bar{X}_n$. Show that it is an unbiased estimator of θ .

(c) Which one would you like better between T_n and $\hat{\theta}_n$ as a point estimator of θ ? Give your reasoning.

Solution:

- (a) 균일분포의 최대값에 대한 추정량이므로 묻고 따지지도 않고 MLE는 최대값이 될 것이지만 정석대로 하자면

$$L(\theta | \{X_i\}_{i=1}^n) = \frac{1}{\theta^n} I_{(0,\theta)}(X_1) \cdots I_{(0,\theta)}(X_n) \quad (33)$$

$$= \frac{1}{\theta^n} I_{(X_{(n)}, \infty)}(\theta) \quad (34)$$

$$\hat{\theta}^{\text{MLE}} = X_{(n)} \quad (35)$$

- (b) $E(X_1) = \theta/2$ 이므로 $E(T_n) = \theta$ 이다.

- (c) 당연히 $\hat{\theta}^{\text{MLE}}$ 가 낫다. 왜냐하면 X_1, \dots, X_n 이 모두 $(\theta/2, \theta)$ 사이에서 발생하면 $T_n > \theta$ 인 상황이 발생한다. 분포의 support를 벗어나는 점추정량은 좋지 않다. 또한 MSE의 관점에서도 MLE가 더 낫다.

5. 아래 그림에서 보듯이 구슬을 위에서 아래로 흘려보낸다고 하자. 각 단계에서 구슬은 오른쪽으로 갈 확률이 p 이고 왼쪽으로 갈 확률이 $1 - p$ 이다 ($0 < p < 1$).

- (1) 가장 아래 단계에 있는 각 칸에 구슬의 수를 확률변수 $(X_1, X_2, X_3, X_4, X_5)$ 라고 하자. 그 확률변수의 분포를 구하시오.
- (2) 확률변수 $(X_1, X_2, X_3, X_4, X_5)$ 를 이용하여 다음 가설의 균일최강력(uniformly most powerful) 검정을 유도하시오.

$$H_0 : p = 1/2 \quad \text{vs} \quad H_1 : p > 1/2 \quad (36)$$

- (3) 각 칸에 구슬이 $(0, 0, 2, 4, 2)$ 인 경우에 균일최강력(uniformly most powerful) 검정의 p -값을 구하시오.

Solution:

- (a) 이것은 다항분포(multinomial distribution)이다. X_1 부터 X_5 까지 각 범주에 해당될 확률은 다음과 같이 정의된다.

$$\Pr(X_k) = \binom{4}{k-1} p^{k-1} (1-p)^{5-k} \quad (37)$$

다항분포의 그 분포식은 다음과 같다.

$$f(X_1, \dots, X_5 | p) = \frac{(X_1 + \dots + X_5)!}{X_1! \cdots X_5!} 4^{X_2+X_6} 6^{X_3} p^{X_2+2X_3+3X_4+4X_5} (1-p)^{4X_1+3X_2+2X_3+X_4} \quad (38)$$

전체 시행횟수 $(X_1 + \cdots + X_5)$ 를 n 으로 놓으면 $(1-p)$ 의 지수부분은

$$4n - X_2 - 2X_3 - 3X_4 - 4X_5 \quad (39)$$

로 바뀐다. 원래 제약조건이 하나 있으면 자유도가 하나 깎이듯 다항분포에서도 모두 더하면 n 이라는 제약 조건이 있기 때문에 항목이 5개여도 확률변수는 5개가 필요하지 않다. 이항 분포(binomial distribution)의 경우에서 2개 범주이지만 n 이 정해져있으므로 확률변수 한 개 X 가 이항분포를 따른다고 하는 것과 마찬가지이다.

(b) $p_1 > 1/2$ 로 놓고, 편의를 위해

$$\mathbf{X} = X_2 + 2X_3 + 3X_4 + 4X_5 \quad (40)$$

라 하자. 가능도비를 구하면

$$\frac{L_0}{L_1} = \frac{1}{2^{4n}} \left(\frac{1}{p_1} \right)^{\mathbf{X}} \left(\frac{1}{1-p_1} \right)^{4n-\mathbf{X}} < k \quad (41)$$

$$-\mathbf{X} \ln p_1 - (4n - \mathbf{X}) \ln (1 - p_1) < c_1 \quad (42)$$

$$-\mathbf{X} \ln p_1 + \mathbf{X} \ln (1 - p_1) < c_2 \quad (43)$$

$$\mathbf{X} \ln \left(\frac{1-p_1}{p_1} \right) < c_3 \quad (44)$$

$p_1 > 1/2$ 이므로

$$\frac{1-p_1}{p_1} < 1 \quad (45)$$

이고 따라서 균일최강력검정은 $\mathbf{X} > c_4$ 일 때이다.

(c) 귀무가설을 참이라 할 때 각각의 확률변수들이 따르는 분포는 다음과 같다.

- $X_2 \sim \text{Bin} \left(n, \frac{4}{16} \right)$
- $X_3 \sim \text{Bin} \left(n, \frac{6}{16} \right)$
- $X_4 \sim \text{Bin} \left(n, \frac{4}{16} \right)$
- $X_5 \sim \text{Bin} \left(n, \frac{1}{16} \right)$

그렇기 때문에 \mathbf{X} 의 분포를 정확히 계산해 내기란 어렵다. p -값을 써보면

$$\Pr \left(X_2 + 2X_3 + 3X_4 + 4X_5 \geq 24 \mid n = 8, p = \frac{1}{2} \right) \quad (46)$$

이므로 대수의 법칙에 의해

$$\frac{1}{M} \sum_{k=1}^M \delta(\mathbf{X}^{(k)} \geq 24) \xrightarrow{a.s.} \Pr\left(\mathbf{X} \geq 24 \mid n=8, p=\frac{1}{2}\right) \quad (47)$$

하므로 몬테칼로 방법으로 근사시킬 수 있다. δ 는 *dirac-delta function*으로 $\mathbf{X}^{(k)} \geq 24$ 이면 1, 아니면 0이다. 실제로 R로 시뮬레이션 해본 결과 p -값 ≈ 0.092 이다. 간략한 R코드는 마지막에 실어놓았다.

A Monte-Carlo Simulation

Listing 1: Monte-Carlo algorithm

```
mc <- function(m) {  
  X2 <- rbinom(m,8,0.25)  
  X3 <- rbinom(m,8,0.375)  
  X4 <- rbinom(m,8,0.25)  
  X5 <- rbinom(m,8,0.0625)  
  mean(X2+2*X3+3*X4+4*X5 >= 24)  
}  
  
mean(sapply(1:200,function(i,...) mc(100000)))
```
