Daeyoung Lim

Grad School

Year of 2009, early

1. X와 Y가 독립적으로 구간 $(0,1)$에서 균일분포(uniform distribution)를 따른다고 하자. 이로부터 U와 V가 다음과 같이 정의된다.

$$U = XY, \;\; V = \frac{X}{Y}$$

   1. (U,V)의 결합밀도(joint density) 함수를 구하여라.
   2. U와 V 각각의 주변밀도(marginal density) 함수를 구하여라.

Solution:

1. 1개 이상의 변수가 존재하고, 그것들의 변환(transformation)으로 새로운 변수가 정의되었을 때 원래 알던 결합밀도함수에 자코비언만 곱하면 된다. 단, 원래 알던 결합밀도함수의 변수에 새로 정의된 변수를 대입해야 한다. 즉 $U = XY$이고 $V = X/Y$라면 U, V의 결합밀도함수는 다음과 같이 정의된다.

$$f_{U,V}\left(u,v\right) = f_{X,Y}\left(x,y\right) \cdot \left|\det\left(J\right)\right|$$

원래 알던 변수 X, Y를 새로 정의된 변수 U, V로 표현하면 $X = U^{1/2}V^{1/2}$, $Y = U^{1/2}V^{-1/2}$이다. 고로 자코비언은

$$J = \begin{bmatrix} \dfrac{\sqrt{v}}{2\sqrt{u}} & \dfrac{\sqrt{u}}{2\sqrt{v}} \\ \dfrac{1}{2\sqrt{uv}} & -\dfrac{\sqrt{u}}{2}v^{3/2} \end{bmatrix}$$

이고 $\left|\det\left(J\right)\right| = 2^{-1}v^{-1}$이다.

$$0 < \sqrt{uv} < 1 \quad \Longrightarrow \quad 0 < u < \frac{1}{v} \tag{1}$$

$$0 < \sqrt{\frac{u}{v}} < 1 \quad \Longrightarrow \quad 0 < u < v \tag{2}$$

중요한 것은 범위인데 우리가 알고 있는 X와 Y의 범위를 이용하면 된다. 그리고 u를 세로축 v를 가로축으로 하여 부등식의 범위를 빗금으로 칠하면 그것이 U와 V가 이루고 있는 범위가 된다. 이 범위는 이중 적분할 때와 같이 한 부등식은 하나가 다른 하나에 의존할 것이고, 나머지 한 부등식은 그 변수가 독립적일 것이다. 그림은 생략하고 그러면 범위가 다음과 같다.

$$u < v < \frac{1}{u} \tag{3}$$

$$0 < u < 1 \tag{4}$$

그러므로 (U, V)의 결합밀도함수는

$$f_{U,V}(u, v) = \frac{1}{2v}, \;\; u < v < \frac{1}{u}, \; 0 < u < 1$$

이다.

2. 위에서 구한 범위로 적분해 내면 된다. 다만 독립적이었던 변수를 의존적으로 바꿔주어야 한다. $0 < u < v^{-1}$과 $0 < u < v$ 둘이 있는데 이에 따라 v의 범위가 달라진다. 그림을 그리면 쉽게 알 수 있다.

$$\begin{aligned} f_U(u) &= \int_u^{1/u} \frac{1}{2v}\, dv \qquad (5)\\ &= -\ln u, \;\; 0 < u < 1 \qquad (6) \end{aligned}$$

$$f_V(v) = \begin{cases} \displaystyle\int_0^v \frac{1}{2v}\, du = \frac{1}{2}, & 0 < v < 1 \\ \displaystyle\int_0^{1/v} \frac{1}{2v}\, du = \frac{1}{2v^2}, & 1 \le v < \infty \end{cases} \qquad (7)$$

2. Let X take on the values 0 and 1 with probabilities p and $1-p$, respectively. It is known that $1/3 \le p \le 2/3$.

   1. Find the MLE $\hat{p}$ of p.
   2. Find the expected squared error, $\mathrm{E}(\hat{p} - p)^2$ of the MLE.
   3. Show that the expected squared error of the MLE is uniformly larger than that of $\widetilde{p} \equiv 1/2$. That is, $\mathrm{E}(\hat{p} - p)^2 > \mathrm{E}\left(\widetilde{p} - p\right)^2$ for all p.

Solution:

1. MLE라는 것이 내가 관측된 샘플을 가장 그럴듯하게 만들어주는 모수의 값을 고르는 것이므로, 만약 $X = 1$이라면 $1 - p$가 가장 높은 값을 $\hat{p}$로 정할 것이다. 그 값은 알려진 바에 따르면 1/3일 것이므로 $\hat{p} = 1/3$이다. 반대로 관측된 것이 $X = 0$이라면 p를 가장 높에 해야 하므로 그 값에 해당하는 2/3가 $\hat{p}$일 것이다. 정리하면

$$\hat{p} = \begin{cases} 1/3, & \text{if } X = 1 \text{ (with prob}=1-p) \\ 2/3, & \text{if } X = 0 \text{ (with prob}=p) \end{cases}$$

2. 위에서 구한 것을 바탕으로

$$\begin{aligned} \mathrm{E}\left(\hat{p}-p\right)^2 &= \mathrm{E}\left(\hat{p}^2 - 2p\hat{p} + p^2\right) && (8)\\ &= \frac{1}{9}\left(1-p\right) + \frac{4}{9}p - 2p\left(\frac{1}{3}\left(1-p\right) + \frac{2}{3}p\right) + p^2 && (9)\\ &= \frac{1}{2}\left(p^2 - p + \frac{1}{3}\right) && (10) \end{aligned}$$

3. 2에서 이미 구했으므로 $\widetilde{p} = 1/2$을 넣어 일단 부등식을 정리하자.

$$\mathrm{E}\left(\hat{p}-p\right)^2 > \mathrm{E}\left(\widetilde{p}-p\right)^2 \quad \Longrightarrow \quad -\frac{1}{36}\left(24p^2 - 24p + 5\right) > 0$$

p의 범위가 1/3과 2/3 사이이므로

$$f\left(p\right) = -\frac{1}{36}\left(24p^2 - 24p + 5\right)$$

이 최대가 되게 하는 정의역 $p = 1/2$를 사이에 두고 양옆으로 걸쳐있다. 즉 1/3과 2/3에서만 $f\left(p\right)$이 양수면 된다.

$$\begin{aligned} f\left(\frac{1}{3}\right) &= \frac{1}{108} && (11)\\ f\left(\frac{2}{3}\right) &= \frac{1}{108} && (12) \end{aligned}$$

사실 위로 볼록이 포물선에서 1/2는 1/3과 2/3의 중간에 위치하므로 값이 같아지는 건 당연하다. 어쨌든 둘 다 양수이므로 증명 끝.

3. 다음의 확률 밀도 함수 (probability density function)를 갖는 확률변수를 고려하자.

$$f\left(x;\theta\right) = \theta x^{\theta-1}, \quad 0 < x < 1.$$

1. 크기가 1 ($n = 1$)인 표본을 이용하여 다음의 가설을 검정하기 위한 $\alpha = 0.05$인 최강력 검정법 (the most powerful test)을 정의하시오.

$$H_0 : \theta = 1 \quad \text{vs} \quad H_1 : \theta = 2$$

2. $\theta = 2$인 경우, 1에서 정의된 검정법의 검정력(power)를 계산하시오.

Solution:

1. X는 $\text{Be}(\theta, 1)$를 따른다. 가능도비를 계산하면

$$\frac{L_0}{L_1} = \frac{1}{2x} < k$$

이므로 기각역이 $c < X$이다. 상수 c는 $\Pr(X > c \mid H_0) = \alpha$를 통해 알 수 있다.

$$\Pr(X > c \mid H_0) = \int_c^1 1\,dx \tag{13}$$
$$= 1 - c = 0.05 \tag{14}$$
$$c = 0.95 \tag{15}$$

따라서 $X > 0.95$일 때 귀무가설을 기각하고 $X \leq 0.95$일 때는 귀무가설을 기각하지 않는 것이 최강력 검정법이다.

2. 검정력은 대립가설이 참일 때 귀무가설을 기각할 확률이므로 $\Pr(X > 0.95 \mid H_1)$이다.

$$\Pr(X > 0.95 \mid H_1) = \int_{0.95}^1 2x\,dx \tag{16}$$
$$= 1 - 0.95^2 \tag{17}$$
$$= 0.0975 \tag{18}$$

4. 다음과 같은 다중선형회귀모형

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

을 고려하자. 여기서, Y는 독립변수를, X_1, X_2, X_3는 설명변수들을, 그리고 $\beta_0, \beta_1, \beta_2, \beta_3$는 회귀계수를 의미하며 ϵ은 평균이 0, 분산이 σ^2인 정규분포를 따르는 오차항을 의미한다. 모수에 대한 가설 $H_0 : \beta_1 = \beta_3$ 대 $H_1 : \beta_1 \neq \beta_3$를 고려하자. 유의수준 5% 하에서 이 가설을 검정하는 방법을 자세히 기술하시오.

Solution: 자세히 기술하라고 했으므로 정말 자세히 기술하도록 한다. 위와 같은 문제는 회귀계수에 제약조건을 주고 그것이 참인지 가설검정을 하는 법을 묻고 있다. 즉 $H : \mathbf{A}\beta = \mathbf{c}$와 같은 가설을 세운 것이다. 여기서 $\mathbf{A}$
$beta$의 각 행이 하나의 제약조건을 구성한다. 검정을 하기 위한 통계량은 당연히 $\mathbf{A}\widehat{\beta}$이 될 것이고 그 값이 $\mathbf{c}$와 너무 다르면 가설을 기각하게 된다. 여기서 너무 '다르면'의 기준을 정해야 하는데

그걸 제공해 주는 것이 바로 '거리' 개념이다. 수학에서 거리 개념은 *distance function* 혹은 *metric* 을 통해 계산하게 되는데 집합의 두 원소를 아무렇게나 잡아도 거리를 계산해낼 수 있는 공간을 *metric space*라 부른다. 그리고 두 원소 사이의 거리뿐만 아니라 개별 원소의 '크기'도 계산해낼 수 있으려면 그 공간에 *norm*이 주어져야 한다. 여기서 $\mathbf{A}\widehat{\beta}-\mathbf{c}$의 *norm*을 계산하면 그 거리가 될 것이다. 일반적으로 우리가 쓰는 크기의 개념은 ℓ_2 norm이라 부르고 유한한 벡터공간(finite vector space)에서는 '제곱합'꼴로 표현된다. 즉 벡터로 말하면 *quadratic form*이 되는 것이다.

따라서 $\mathbf{A}\beta-\mathbf{c}$의 제곱합은 $(\mathbf{A}\beta-\mathbf{c})'(\mathbf{A}\beta-\mathbf{c})$일 텐데 선형회귀에서는 각 추정량의 분산을 고려해주어야 한다. 즉 $\widehat{\beta}$가 결정된 값이 아니라 관측 이전의 값이므로 각각의 원소가 $\mathbf{c}$와의 거리를 계산할 때 얼마나 변하는지 고려하는 것이 상식적일 것이다. 따라서 우리가 정한 통계량과 검정하고 싶은 값 사이의 거리를 계산하기 위해 도출된 식은 다음과 같다.

$$\left(\mathbf{A}\widehat{\beta}-\mathbf{c}\right)'\left(\mathrm{Var}\left(\mathbf{A}\widehat{\beta}\right)\right)^{-1}\left(\mathbf{A}\widehat{\beta}-\mathbf{c}\right)$$

그리고 다음의 과정을 거쳐서 다시 계산된다.

- $\mathrm{Var}\left(\mathbf{A}X\right)=\mathbf{A}\mathrm{Var}\left(X\right)\mathbf{A'}$ (여기서 X는 임의의 random vector)
- $\mathrm{Var}\left(\widehat{\beta}\right)=\mathrm{Var}\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'y\right)=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\sigma^2\mathbf{I}_n\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)'=\sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$
- 실제로 $\widehat{\beta}$의 분산을 구할 때는 σ^2를 모르므로 σ^2의 비편향추정량인 $S^2=\mathrm{SSE}/\left(n-p-1\right)$로 대체한다. 여기서 p는 '*predictor*'의 앞글자로 설명변수의 개수를 의미한다. 절편까지 이미 추정한 모수의 개수가 $p+1$이므로 자유도에서 그만큼 빠진다.

이렇게 되어 우리가 쓰게 될 검정통계량의 초기형태는 다음과 같다.

$$\frac{1}{S^2}\left(\mathbf{A}\widehat{\beta}-\mathbf{c}\right)'\left(\mathbf{A}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\beta}-\mathbf{c}\right)$$

H가 참일 때 구한 SSE을 SSE_H이라 표시하고 다변량 정규분포의 특성을 이용하면 다음과 같은 사실이 구해진다.

- $\mathbf{A}\widehat{\beta}\sim\mathcal{N}_q\left(\mathbf{c},\sigma^2\mathbf{A}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{A}'\right)$ (여기서 q는 $\mathbf{A}$의 행수)
- 위에서 구한 $\mathrm{Var}\left(\mathbf{A}\widehat{\beta}\right)^{-1/2}$를 $\mathbf{A}\widehat{\beta}-\mathbf{c}$ 앞에 곱해주면 $\mathcal{N}_q\left(\mathbf{0},\mathbf{I}_q\right)$이 된다.
- 즉, 카이제곱분포가 표준정규분포를 따르는 확률변수의 제곱합꼴로 표시된다는 점을 상기해볼 때
$$\left(\mathbf{A}\widehat{\beta}-\mathbf{c}\right)'\left(\mathrm{Var}\left(\mathbf{A}\widehat{\beta}\right)\right)^{-1}\left(\mathbf{A}\widehat{\beta}-\mathbf{c}\right)\sim\chi^2\left(q\right)\tag{19}$$
  임을 알 수 있다.

- 다음과 같은 사실이 성립한다. 증명은 Appendix로 뺀다.

$$\frac{\mathrm{SSE}_H - \mathrm{SSE}}{\sigma^2} = \left(\mathbf{A}\widehat{\beta} - \mathbf{c}\right)' \left(\mathrm{Var}\left(\mathbf{A}\widehat{\beta}\right)\right)^{-1} \left(\mathbf{A}\widehat{\beta} - \mathbf{c}\right) \tag{20}$$

- 따라서 $\mathrm{SSE}/\sigma^2 \sim \chi^2\left(n-p-1\right)$이므로

$$F = \frac{\left(\mathrm{SSE}_H - \mathrm{SSE}\right)/\left(\sigma^2 q\right)}{\mathrm{SSE}/\left(\sigma^2\left(n-p-1\right)\right)} \sim F_{q,n-p-1}$$

이제 문제로 다시 돌아가보자. 제약조건은 $\beta_1 = \beta_3$로 한 개이다. 따라서 $\mathbf{A}$의 행은 1개, 즉 $q = 1$이다. 따라서

$$\mathbf{A}\beta = \begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = 0$$

이 된다. 그리고

$$F = \frac{\mathrm{SSE}_H - \mathrm{SSE}/1}{\mathrm{SSE}/\left(n-4\right)} \sim F_{1,n-4}$$

를 따르고 $F > F_{1,n-4}^{-1}\left(0.05\right)$일 때 0.05 신뢰수준에서 귀무가설을 기각한다.

A SSE$_H$ 유도

통계의 많은 문제는 최댓값/최솟값을 찾는 수치최적화 문제로 치환된다. 일반적인 선형회귀 모형은

$$Y = \mathbf{X}\beta + \epsilon$$

이며 위 모형을 적합할 때는

$$\widehat{\beta} = \underset{\beta}{\operatorname{argmin}} \left(Y - \mathbf{X}\beta\right)' \left(Y - \mathbf{X}\beta\right)$$

와 같이 최적화 문제로 바뀐다. 일반 선형회귀 모형을 적합하는 것은 제약조건이 없는 경우이지만 $\mathbf{A}\beta = \mathbf{c}$ 와 같은 제약조건 하에서 최소화하는 β를 찾는 문제는 라그랑지 승수를 써서 해결한다. 제약조건이 있는 최적화 문제는 다음과 같이 표기한다.

$$\min_{\text{subject to } \mathbf{A}\beta = \mathbf{c}} \left(Y - \mathbf{X}\beta\right)' \left(Y - \mathbf{X}\beta\right)$$

라그랑지 승수를 써서 다시 문제를 쓰면 아래와 같다.

$$\mathbf{r} = \left(Y - \mathbf{X}\beta\right)' \left(Y - \mathbf{X}\beta\right) + \boldsymbol{\lambda}' \left(\mathbf{A}\beta - \mathbf{c}\right) \tag{21}$$

$$\widehat{\beta}_H = \underset{\beta}{\operatorname{argmin}}\, \mathbf{r} \tag{22}$$

라그랑지 승수로 표현된 식을 최소화하는 방법은 정형화되어 있다. 먼저 정의역으로 정의된 변수로 미분을 하고 제약조건에 대입한 뒤에 나온 라그랑지 승수의 해를 원래 정의역으로 정의된 변수 식에 대입하는 것! 그러므로 $\mathbf{r}$을 β에 대해 미분하면

$$\frac{\partial \mathbf{r}}{\partial \beta} = -2\mathbf{X}'Y + 2\mathbf{X}'\mathbf{X}\beta + \mathbf{A}'\boldsymbol{\lambda} = \mathbf{0}$$

이 되고 $\widehat{\beta}_H = \widehat{\beta} - \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{A}'\widehat{\boldsymbol{\lambda}}/2$이 도출된다. 이를 원래 제약 조건에 대입하면

$$\widehat{\boldsymbol{\lambda}} = -2\left(\mathbf{A}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{A}'\right)^{-1}\left(\mathbf{c} - \mathbf{A}\widehat{\beta}\right)$$

이 된다. 이제 $\operatorname{Var}\left(\mathbf{A}\widehat{\beta}\right)^{-1}$꼴이 등장한다. 이를 다시 $\widehat{\beta}_H$의 식에 대입하면

$$\widehat{\beta}_H = \widehat{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{A}'\left(\mathbf{A}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{A}'\right)^{-1}\left(\mathbf{c} - \mathbf{A}\widehat{\beta}\right) \tag{23}$$

이 나온다. 이제부터 식이 좀 복잡해지는데 다시 $\mathrm{SSE}_H - \mathrm{SSE}$를 구하기 위해서는 다음과 같이 시작한다.

$$\left(Y-\mathbf{X}\widehat{\beta}_H\right)'\left(Y-\mathbf{X}\widehat{\beta}_H\right) = \left(Y-\mathbf{X}\widehat{\beta}+\mathbf{X}\widehat{\beta}-\mathbf{X}\widehat{\beta}_H\right)'\left(Y-\mathbf{X}\widehat{\beta}+\mathbf{X}\widehat{\beta}-\mathbf{X}\widehat{\beta}_H\right) \tag{24}$$

$$= \underbrace{\left(Y-\mathbf{X}\widehat{\beta}\right)'\left(Y-\mathbf{X}\widehat{\beta}\right)}_{\mathrm{SSE}} + \left(\widehat{\beta}-\widehat{\beta}_H\right)'\mathbf{X}'\mathbf{X}\left(\widehat{\beta}-\widehat{\beta}_H\right) + 2\left(Y-\mathbf{X}\widehat{\beta}\right)'\left(\mathbf{X}\widehat{\beta}-\mathbf{X}\widehat{\beta}_H\right) \tag{25}$$

$$= \mathrm{SSE} + \left(\widehat{\beta}-\widehat{\beta}_H\right)'\mathbf{X}'\mathbf{X}\left(\widehat{\beta}-\widehat{\beta}_H\right) + 2\underbrace{\left(Y-\mathbf{X}\widehat{\beta}\right)'\mathbf{X}\left(\widehat{\beta}-\widehat{\beta}_H\right)}_{=0} \tag{26}$$

$$= \mathrm{SSE} + \left(\widehat{\beta}-\widehat{\beta}_H\right)'\mathbf{X}'\mathbf{X}\left(\widehat{\beta}-\widehat{\beta}_H\right) \tag{27}$$

여기서 저게 왜 0이 되냐하면 $\mathbf{X}'Y = \mathbf{X}'\mathbf{X}\widehat{\beta}$이기 때문. 이제 (23)의 $\widehat{\beta}$를 넘겨서 (27)에 대입하고 정리하면 샤샤샤 소거되고 (20)이 남는다.

B SSE $\sim \chi^2(n-p-1)$

$Z \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$이고 어떤 행렬 $\mathbf{A}$가 $\mathbf{A}^2 = \mathbf{A}$를 만족한다고 하자(idempotent). 그러면 다음과 같은 성질을 만족한다.

- $\mathbf{A}$의 고유값(eigenvalue)는 0 또는 1이다.
- $Z'\mathbf{A}Z \sim \chi^2(\mathrm{Tr}(\mathbf{A}))$이다.

먼저 고유값이 0 또는 1인 사실은 다음을 통해 간단히 알 수 있다.

$$\mathbf{A}\lambda = \lambda v \tag{28}$$

$$\mathbf{A}^2\lambda = \lambda\mathbf{A}v \tag{29}$$

$$(\mathrm{LHS}) = \mathbf{A}\lambda = \lambda v \tag{30}$$

$$(\mathrm{RHS}) = \lambda^2 v \tag{31}$$

$$\lambda^2 v = \lambda v \tag{32}$$

$$\therefore \lambda = 0 \text{ or } 1 \tag{33}$$