Daeyoung Lim
Grad School
2016, Late

1. 표 1은 확률변수 (X, Y)의 결합밀도함수 p를 나타낸다. 이에 대해 다음 물음에 답하시오.
   (a) X의 주변밀도함수를 구하시오.
   (b) $X = 30$일 때 Y의 조건부밀도함수를 구하시오.
   (c) $X = 30$일 때 Y의 조건부평균을 구하시오.
   (d) $\mathrm{E}\,(Y-a)^2$을 최소로 하는 상수 a^*를 구하시오.
   (e) $\mathrm{E}\,(Y-\beta_0-\beta_1 X)^2$을 최소로 하는 2차원 벡터 $\beta^* = (\beta_0^*, \beta_1^*)$를 구하시오.
   (f) $X = 30$일 때 최적선형예측값 $\beta_0^* + \beta_1^* \times 30$을 구하고, $X = 30$일 때 Y의 조건부 평균값과 비교하시오.

Solution:

(a) 주변(marginal)분포이므로 가생이에 있는 걸 쓰면 된다.

$$\Pr(X=x) = \begin{cases} 0.45, & \text{if } x = 30 \\ 0.55, & \text{if } x = 50 \end{cases} \tag{1}$$

(b) $\Pr(X = 30) = 0.45$이므로

$$\Pr(Y = 4000 \mid X = 30) = \frac{8}{9} \tag{2}$$
$$\Pr(Y = 7000 \mid X = 30) = \frac{1}{9} \tag{3}$$

(c) 다음과 같다.

$$\mathrm{E}\,(Y \mid X = 30) = 4000 \times \frac{8}{9} + 7000 \times \frac{1}{9} = \frac{13000}{3} \tag{4}$$

(d) 다음과 같이 구할 수 있다.

$$\begin{aligned} \mathrm{E}\left((Y-a)^2\right) &= \mathrm{E}\left((Y-\mathrm{E}\,(Y)+\mathrm{E}\,(Y)-a)^2\right) && (5)\\ &= \mathrm{E}\left((Y-\mathrm{E}\,(Y))^2 + \cancel{2(\mathrm{E}\,(Y)-a)(Y-\mathrm{E}\,(Y))} + (\mathrm{E}\,(Y)-a)^2\right) && (6)\\ &= \mathrm{E}\left((Y-\mathrm{E}\,(Y))^2 + (\mathrm{E}\,(Y)-a)^2\right) && (7)\end{aligned}$$

따라서 제곱식 2개가 있을 때 변할 수 있는 값이 a뿐일 때 제곱식 하나를 0으로 만드는 것이 전체를 최소화하는 방법이므로 $a^* = \mathrm{E}\,(Y)$이다. 고로 본 문제에서는

$$a^* = \frac{1}{2}(4000 + 7000) = 5500 \tag{8}$$

(e) 선형식을 주었을 때는 다음과 같이 구한다. $f(\beta_0, \beta_1) = \mathrm{E}\left((Y - \beta_0 - \beta_1 X)^2\right)$라고 할 때

$$\frac{\partial}{\partial \beta_0} f(\beta_0, \beta_1) = -2\left(\mathrm{E}(Y) - \beta_0 - \beta_1 \mathrm{E}(X)\right) = 0 \tag{9}$$

$$\frac{\partial}{\partial \beta_1} f(\beta_0, \beta_1) = -2\left(\mathrm{E}(XY) - \beta_0 \mathrm{E}(X) - \beta_1 \mathrm{E}\left(X^2\right)\right) = 0 \tag{10}$$

(8)을 $\mathrm{E}(Y) = \beta_0 + \beta_1 \mathrm{E}(X)$로 정리하고 양변에 $\mathrm{E}(X)$을 곱해주면

$$\mathrm{E}(X)\,\mathrm{E}(Y) = \beta_0 \mathrm{E}(X) + \beta_1 \{\mathrm{E}(X)\}^2 \tag{11}$$

이 되고 (9)번 식도 정리하면 $\mathrm{E}(XY) = \beta_0 \mathrm{E}(X) + \beta_1 \mathrm{E}\left(X^2\right)$이 되므로 이 식에서 (10)을 빼주면

$$\mathrm{Cov}(X, Y) = \beta_1 \mathrm{Var}(X) \tag{12}$$

이 되어서

$$\beta_1^* = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} \tag{13}$$

가 된다. 그리고 (8)번 식에서

$$\beta_0^* = \mathrm{E}(Y) - \beta_1^* \mathrm{E}(X) \tag{14}$$

를 유도할 수 있다. 실제로 값을 구할 때 가장 문제가 되는 것은 $\mathrm{E}(XY)$인데 이는 다음과 같이 구할 수 있다.

$$\begin{aligned}
\mathrm{E}(XY) &= \sum_x \sum_y xy f_{X,Y}(x, y) && (15)\\
&= 30 \cdot 4000 \cdot 0.4 + 50 \cdot 4000 \cdot 0.1 + 30 \cdot 7000 \cdot 0.05 + 50 \cdot 7000 \cdot 0.45 && (16)\\
&= 236000 && (17)\\
\mathrm{E}(X) &= 30 \cdot 0.45 + 50 \cdot 0.55 && (18)\\
&= 41 && (19)\\
\mathrm{E}\left(X^2\right) &= 30^2 \cdot 0.45 + 50^2 \cdot 0.55 \quad \text{(By the law of unconscious statistician)} && (20)\\
&= 1780 && (21)\\
\mathrm{Var}(X) &= 99 && (22)\\
\beta_1^* &= \frac{10500}{99} && (23)\\
\beta_0^* &= \frac{38000}{33} && (24)
\end{aligned}$$

(f) 그냥 대입하자.

$$\beta_0^* + \beta_1^* \times 30 = \frac{38000}{33} + \frac{10500 \times 30}{99} = \frac{13000}{3} \tag{25}$$

$$\mathrm{E}\left(Y \mid X = 30\right) = 4000 \times \frac{8}{9} + 7000 \times \frac{1}{9} = \frac{13000}{3} \tag{26}$$

똑같다.

2. Suppose that a random sample of size n is drawn from each of the following distributions. Obtain the maximum likelihood estimate (MLE) for each of (a) and (b).

(a) $f\left(y;\mu,\sigma^2\right) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\dfrac{\left(y-\mu\right)^2}{2\sigma^2}\right)$

(b) $f\left(y;\pi\right) = \pi^y \left(1-\pi\right)^{1-y} I_{[0,1]}\left(y\right)$, where $0 \le \pi \le 1$.

In addition, sketch the likelihood function for $n = 3$ in (b), and mark the MLE on your figure.

Solution:

(a) 정규분포 MLE는 다 알 것으로 간주...

$$\widehat{\mu} = \overline{Y}_n \tag{27}$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}_n\right)^2 \tag{28}$$

(b) 문제가 좀 잘못됐다. indicator function의 아랫부분이 $[0,1]$이 아니라 집합이어야 한다. 즉, $\{0,1\}$이어야 한다. 그러면 베르누이 변수들이므로 우리가 아는 것처럼

$$\widehat{\pi} = \overline{Y}_n \tag{29}$$

이 된다. 개형은 알아서...

3. $X_1, \ldots, X_n$이 다음의 확률밀도함수를 갖는 임의표본이라고 하자.

$$f\left(x\right) = \frac{\theta^3 e^{-\theta/x}}{2x^4}, \qquad x > 0.$$

(a) $H_0 : \theta = \theta_0$ vs $H_a : \theta > \theta_0$의 가설검정에 대하여, 유의수준 α의 균일최강력검정 (Uniformly Most Powerful Test)의 기각역을 구하시오.

(b) $n = 10$이고 $\theta_0 = 1$, $\alpha = 0.05$라 하자. $\theta = 2$에서 (a)에서 구한 균일최강력검정의 검정력을 구하시오. 첨부된 카이제곱 임계값을 이용하시오.

Solution:

(a) 먼저 가능도함수를 구하자.

$$L\left(\theta \mid \{X_i\}_{i=1}^n\right) = \frac{\theta^{3n}}{2^n\left(X_1X_2\cdots X_n\right)^4}\exp\left(-\theta\left(\frac{1}{X_1}+\cdots+\frac{1}{X_n}\right)\right) \tag{30}$$

따라서 $\theta_1 > \theta_0$인 임의의 θ_1를 정해놓고 가능도비를 구하면

$$\frac{L_0}{L_1} = \left(\frac{\theta_0}{\theta_1}\right)^{3n}\exp\left(\left(\theta_1-\theta_0\right)\sum_{i=1}^n \frac{1}{X_i}\right) \leq k \tag{31}$$

이므로 *Karlin-Rubin theorem*에 따라

$$\text{RR} = \left\{ \{X_i\}_{i=1}^n \,\middle|\, \sum_{i=1}^n \frac{1}{X_i} \leq k' \text{ for some } k' \right\} \tag{32}$$

사실 $X_i \sim \text{InvGam}\left(3,\theta\right)$이다. 따라서 $X_i^{-1} \sim \text{Ga}\left(3,\theta\right)$이다. 그러므로

$$\sum_{i=1}^n \frac{1}{X_i} \sim \text{Ga}\left(3n,\theta\right) \tag{33}$$

이 된다. 이를 카이제곱 변수로 만들기 위해서는 다음과 같이 해야 한다.

$$2\theta\sum_{i=1}^n \frac{1}{X_i} \sim \text{Ga}\left(\frac{6n}{2},\frac{1}{2}\right) \overset{d}{\equiv} \chi^2\left(6n\right) \tag{34}$$

이를 이용하여

$$\Pr\left(2\theta_0\sum_{i=1}^n \frac{1}{X_i} \leq \chi^2_{6n}\left(1-\alpha\right)\right) = \alpha \tag{35}$$

임을 알 수 있고 따라서 기각역에서 k'는 다음과 같다.

$$k' = \frac{1}{2\theta_0}\chi^2_{6n}\left(1-\alpha\right) \tag{36}$$

(b) 주어진 조건에 의하면

$$\sum_{i=1}^{10} \frac{1}{X_i} \sim \text{Ga}\left(30,1\right) \tag{37}$$

이고 검정력은 다음과 같다.

$$\Pr\left(\sum_{i=1}^{10} \frac{1}{X_i} \leq \frac{1}{4}\chi^2_{60}\left(0.95\right)\right) \tag{38}$$

우을 구하기 위해 R의 다음 코드, `qchisq(0.95,60,lower.tail=FALSE)/4`를 이용하면 `10.79699`를 얻을 수 있다. 따라서 검정력을 다음 R 코드를 통해 구한다.

`pgamma(10.79699,30,1)=1.171414e-06`. 그러므로 $\alpha = 0.05$ 유의수준에서 귀무가설은 기각된다.

4. 체중(x)과 신장(y)의 선형관계를 살펴보기 위하여 남녀 각각 3명을 랜덤하게 추출하였다. 남녀 간의 평균 신장 및 평균 체중의 차이를 반영하기 위하여 아래의 회귀모형을 고려하였다. 여기서 $\overline{x}_{i\cdot}$는 남자와 여자의 체중의 표본평균을 나타낸다.

$$y_{ij} = \mu_i + \left(x_{ij} - \overline{x}_{i\cdot}\right)\beta + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right), \ \ i = \text{남자, 여자}, \ \ j = 1,2,3.$$

(a) μ_i의 최소제곱추정량을 구하시오.

(b) $\widehat{\beta}$을 β의 최소제곱추정량이라 하자. (a)의 답과 $\widehat{\beta}$을 이용하여 σ^2의 비편향 추정량을 구하시오.

Solution:

(a) 이를 어떻게든 회귀모형의 행렬꼴로 바꿔야 한다. 즉 $Y = \mathbf{X}\boldsymbol{\theta} + \epsilon$. 다음과 같이 바꾸자.

$$Y = \begin{bmatrix} y_{11} & y_{12} & y_{13} & y_{21} & y_{22} & y_{23} \end{bmatrix}' \tag{39}$$

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & (x_{11} - \overline{x}_{1\cdot}) \\ 1 & 0 & (x_{12} - \overline{x}_{1\cdot}) \\ 1 & 0 & (x_{13} - \overline{x}_{1\cdot}) \\ 0 & 1 & (x_{21} - \overline{x}_{2\cdot}) \\ 0 & 1 & (x_{22} - \overline{x}_{2\cdot}) \\ 0 & 1 & (x_{23} - \overline{x}_{2\cdot}) \end{bmatrix} \tag{40}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \mu_1 & \mu_2 & \beta \end{bmatrix}' \tag{41}$$

$$\epsilon = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} & \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \end{bmatrix}' \tag{42}$$

나머지는 회귀분석과 동일하다.

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'Y \tag{43}$$

하... 구해보자.

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 3 & 0 & \sum_{j=1}^{3}\left(x_{1j}-\overline{x}_{1\cdot}\right) \\ 0 & 3 & \sum_{j=1}^{3}\left(x_{2j}-\overline{x}_{2\cdot}\right) \\ \sum_{j=1}^{3}\left(x_{1j}-\overline{x}_{1\cdot}\right) & \sum_{j=1}^{3}\left(x_{2j}-\overline{x}_{2\cdot}\right) & \sum_{i=1}^{2}\sum_{j=1}^{3}\left(x_{ij}-\overline{x}_{i\cdot}\right)^2 \end{bmatrix} \tag{44}$$

위 행렬은 대각행렬임을 알 수 있다. 따라서

$$\mathbf{X}'Y = \begin{bmatrix} y_{11}+y_{12}+y_{13} \\ y_{21}+y_{22}+y_{23} \\ \sum_{i=1}^{2}\sum_{j=1}^{3}\left(x_{ij}-\overline{x}_{i\cdot}\right)y_{ij} \end{bmatrix} \tag{45}$$

그러므로

$$\boldsymbol{\theta} = \begin{bmatrix} \frac{1}{3}\left(y_{11}+y_{12}+y_{13}\right) \\ \frac{1}{3}\left(y_{21}+y_{22}+y_{23}\right) \\ \left(\sum_{i=1}^{2}\sum_{j=1}^{3}\left(x_{ij}-\overline{x}_{i\cdot}\right)y_{ij}\right)\Big/\left(\sum_{i=1}^{2}\sum_{j=1}^{3}\left(x_{ij}-\overline{x}_{i\cdot}\right)^2\right) \end{bmatrix} \tag{46}$$

(b) 회귀분석과 똑같다. 비편향 추정량은 다음과 같다.

$$\widehat{\sigma}^2 = \frac{\left(Y-\widehat{Y}\right)'\left(Y-\widehat{Y}\right)}{n-3} \tag{47}$$

여기서 3은 이미 추정한 모수의 개수이다. 비편향 추정량인지 확인하려면 다음처럼 하면

된다.

$$
\begin{aligned}
\mathrm{E}\left(\widehat{\sigma}^2\right) &= \frac{1}{n-3}\mathrm{E}\left(\left(Y-\widehat{Y}\right)'\left(Y-\widehat{Y}\right)\right) && (48)\\
&= \frac{1}{n-3}\mathrm{E}\left(Y'\left(\mathrm{I}_n-\mathbf{H}\right)Y\right) \qquad \left(\text{여기서 } \mathbf{H}=\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right) && (49)\\
&= \frac{1}{n-3}\mathrm{E}\left(\left(Y-\mathbf{X}\boldsymbol{\theta}\right)'\left(\mathrm{I}_n-\mathrm{H}\right)\left(Y-\mathbf{X}\boldsymbol{\theta}\right)\right) \qquad (\text{전개하면 소거되고 위와 같다}) && (50)\\
&= \frac{1}{n-3}\mathrm{E}\left(\epsilon'\left(\mathbf{I}_n-\mathbf{H}\right)\epsilon\right) && (51)\\
&= \frac{1}{n-3}\mathrm{E}\left(\mathrm{Tr}\left(\epsilon'\left(\mathbf{I}_n-\mathbf{H}\right)\epsilon\right)\right) && (52)\\
&= \frac{1}{n-3}\mathrm{E}\left(\mathrm{Tr}\left(\left(\mathbf{I}_n-\mathbf{H}\right)\epsilon\epsilon'\right)\right) \qquad (\text{Tr 연산자는 permutable}) && (53)\\
&= \frac{1}{n-3}\,\mathrm{Tr}\left(\left(\mathbf{I}_n-\mathbf{H}\right)\mathrm{E}\left(\epsilon\epsilon'\right)\right) \qquad (\text{Tr 연산자는 linear하므로 E가 들어갈 수 있다.}) && (54)\\
&= \frac{\sigma^2}{n-3}\,\mathrm{Tr}\left(\mathbf{I}_n-\mathbf{H}\right) && (55)\\
&= \sigma^2 && (56)
\end{aligned}
$$

따라서 비편향 추정량이 된다.