Daeyoung Lim

Grad School

Year of 2013, early

1. (25점) Suppose that X and Y have the following joint probability density function

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{2}(x+y)e^{-(x+y)}, & x,y \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

(a) Find the moment generating function of $Z = X + Y$.

(b) Find the expected value of $W = X/(X+Y)$.

(c) Are Z and W independent?

Solution:

(a) $Z = X + Y$, $X = X$로 놓고 변수변환을 하게 되면

$$J = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \tag{2}$$

$$|J| = 1 \tag{3}$$

$$f_{X,Z}(x,z) = \frac{1}{2}ze^{-z}, \quad 0 \leq x \leq z \tag{4}$$

따라서 적분해서 x를 소거하면

$$f_Z(z) = \int_0^z \frac{1}{2}ze^{-z}\,dx = \frac{1}{2}z^2e^{-z} \tag{5}$$

이 되어 $Z \sim \mathrm{Ga}(3,1)$임을 알 수 있다. MGF를 직접 정의를 이용해 구해도 상관없지만 이미 꼴을 알고 있다면 그냥 감마분포임을 이용해 구해도 무방하다. 정의를 통해 구해보자.

$$\mathrm{E}\left(e^{tZ}\right) = \int_0^\infty \frac{1}{2}z^2e^{-(1-t)z}\,dz \tag{6}$$

$$= \frac{1}{\cancel{2}}\frac{\cancel{\Gamma(3)}}{(1-t)^3} \tag{7}$$

$$= (1-t)^{-3} \tag{8}$$

(b) *The law of unconscious statistician*에 따라 다음과 같이 구할 수 있다.

$$\mathrm{E}\left(\frac{X}{X+Y}\right) = \int_0^\infty \int_0^\infty \frac{x}{x+y}\frac{1}{2}\left(x+y\right)e^{-(x+y)}\,dx\,dy \tag{9}$$
$$= \int_0^\infty \frac{1}{2}e^{-y}\underbrace{\int_0^\infty xe^{-x}\,dx}_{=1}\,dy \tag{10}$$
$$= \frac{1}{2}\underbrace{\int_0^\infty e^{-y}\,dy}_{=1} \tag{11}$$
$$= \frac{1}{2} \tag{12}$$

(c) 둘의 결합밀도함수를 구해보자. $Z = X+Y$, $W = X/\left(X+Y\right)$이므로

$$|J| = z \tag{13}$$
$$f_{W,Z}\left(w,z\right) = f_{X,Y}\left(wz,\left(1-w\right)z\right)\cdot z \tag{14}$$
$$= \frac{1}{2}z^2e^{-z} \tag{15}$$

범위가 중요한데 $x = wz$, $y = z - wz$이고 $x > 0, y > 0$이므로 $wz > 0, z > wz$가 되어 $z > 0, 1 > w > 0$이 나온다. 즉 W, Z는 서로 의존하지 않고 결합밀도함수 역시 주변밀도함수의 곱으로 이뤄짐을 알 수 있다. $Z \sim \mathrm{Ga}\left(3,1\right)$, $W \sim \mathrm{Unif}\left(0,1\right)$이다. 이는 (b)에서 구했던 W의 기댓값과도 일치한다. 따라서 W, Z는 서로 독립이다.

2. (25점) $X_1, X_2, \ldots, X_n$을 평균이 μ이고 분산이 $\sigma^2(<\infty)$인 i.i.d한 표본이라고 할 때

(a) 모평균 μ에 대한 적률추정치(moment estimator)와 최소제곱추정치(least squared estimator) 모두가 표본 평균임을 보이시오.

(b) 표본평균이 일치추정량(consistent estimator)임을 보이시오.

Solution:

(a) 모평균의 적률추정치가 1차 표본적률(표본평균)이 되는 것은 당연하다. 최소제곱추정치는 대표값으로 '평균'을 사용하는 것을 정당화해준다. 다시 말해, 손실함수(loss function)을 오차의 제곱합으로 잡을 경우 그것의 최소화하는 것은 평균이 된다. 먼저 표본이 아닌 모집단으로 생각해보자.

$$\hat{c} = \underset{c}{\operatorname{argmin}}\,\mathrm{E}\left(\left(X-c\right)^2\right) \tag{16}$$

MSE(mean squared error)의 관점에서 가장 좋은 추정량은 $\hat{c} = \mathrm{E}(X)$이 된다.

PROOF:

$$\begin{aligned}
\mathrm{E}\left((X - \mathrm{E}(X) + \mathrm{E}(X) - c)^2\right) &= \mathrm{E}\Big[(X - \mathrm{E}(X))^2 &(17)\\
&\quad \cancel{+2(X - \mathrm{E}(X))(\mathrm{E}(X) - c)} + (\mathrm{E}(X) - c)^2\Big] &(18)\\
&= \mathrm{E}\left[(X - \mathrm{E}(X))^2 + (\mathrm{E}(X) - c)^2\right] &(19)\\
\hat{c} &= \mathrm{E}(X) &(20)
\end{aligned}$$

QED

다시 표본의 입장에서 미분을 통해 증명하면

PROOF:

$$\begin{aligned}
f(c) &= \sum_{i=1}^{n} (X_i - c)^2 &(21)\\
&= \sum_{i=1}^{n} X_i^2 - 2c\sum_{i=1}^{n} X_i + nc^2 &(22)\\
\frac{d}{dc} f(c) &= -2\sum_{i=1}^{n} X_i + 2nc = 0 &(23)\\
\hat{c} &= \frac{1}{n}\sum_{i=1}^{n} X_i &(24)
\end{aligned}$$

QED

따라서

$$\sum_{i=1}^{n} (X_i - c^{\mathrm{sam}})^2 \xrightarrow{p} \mathrm{E}\left((X - c^{\mathrm{pop}})^2\right) \tag{25}$$

이므로 $\hat{c}^{\mathrm{sam}} \xrightarrow{p} \hat{c}^{\mathrm{pop}}$이다. 즉, 평균의 최소제곱추정량이 $\hat{c}^{\mathrm{sam}}$이 되어 표본평균이 된다.

(b) 약대수의 법칙(*the weak law of large numbers*)를 증명하는 것인데 이는 *Chebyshev's inequality*를 이용하면 쉽다.

Lemma 1 *(Chebyshev's inequality) Suppose* $0 \le t < \|f\|_{\infty}$. *Define*

$$A = \{x \in X \mid |f(x)| \ge t\}.$$

Then,

$$\mu(A) \le \left(\frac{\|f\|_p}{t}\right)^p.$$

PROOF: L^∞ norm의 정의에 따라 $\mu(A) > 0$로 A는 μ-null set이 아니다. 따라서

$$\|f\|_p \geq \left(\int_A |f|^p \, d\mu\right)^{1/p} \tag{26}$$
$$\geq \left(t^p \mu(A)\right)^{1/p} \tag{27}$$
$$= t\mu(A)^{1/p} \tag{28}$$

QED

따라서 확률론적으로 우리가 아는 평균에 대한 부등식으로 표현하면 다음과 같아진다.

$$\Pr\left(\left|\overline{X}_n - \mathrm{E}(X)\right| \geq \lambda\right) \leq \frac{\mathrm{E}\left|X - \mathrm{E}(X)\right|^2}{n\lambda^2} \tag{29}$$

원래는 부등식의 우변에 $\overline{X}_n$에 대한 식이 와야 하는데 다음의 항등식으로 인해 위와 같이 쓸 수 있다.

$$\mathrm{E}\left|\overline{X}_n - \mathrm{E}(X)\right|^2 = \frac{1}{n}\mathrm{E}\left|X - \mathrm{E}(X)\right|^2, \left(\mathrm{Var}\left(\overline{X}_n\right) = \frac{1}{n}\mathrm{Var}(X)\right)$$

따라서 $n \to \infty$함에 따라

$$\lim_{n\to\infty} \Pr\left(\left|\overline{X}_n - \mathrm{E}(X)\right| \geq \lambda\right) = 0 \tag{30}$$

고로 $\overline{X}_n$는 $\mathrm{E}(X)$에 대한 일치추정량이다. 그냥 *Chebyshev's inequality* 증명 없이 바로 써도 상관없다.

3. (25점) We want to develop general methods for testing hypotheses and apply those methods to some common problems.

   (a) Briefly explain what a uniformly most powerful (UMP) test is. In particular, what is the meaning of "uniformly"?

   (b) Let $X_1, \ldots, X_n$ be a random sample from $f(x) = \theta e^{-\theta x} I_{(0,\infty)}(x)$. Find a UMP test of $H_0, \theta = \theta_0$ versus $H_1 : \theta > \theta_0$.

Solution:

(a) (*Casella-Berger*에 따르면) $\mathcal{C}$이 어떤 모수 θ에 대해 다음을 두 가설에 대한 검정들의 집합

이라 하자.

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_0^c \tag{31}$$

만약 $\mathcal{C}$에 속하는 어떤 검정의 검정력함수 $\beta(\theta)$이 그 어떤 $\theta \in \Theta_0^c$와 그 어떤 $\beta'(\theta)$을 택해도 $\beta(\theta) \geq \beta'(\theta)$일 때 우리는 그 검정을 균일최강력검정(UMP test)라 부른다.

우리가 평소에 자주 쓰는 가능도비는 균일최강력검정을 유도하기 위한 방법으로 그 정당성은 *Neyman-Pearson lemma*(simple hypotheses), *Karlin-Rubin theorem*(composite hypotheses)에 있다. 여기서 '균일(uniform)'이란 말이 붙은 것은 원래 simple alternative hypothesis를 어떻게 잡아도 최강력검정이 되기 때문에 '균일하게' 다 발라버린다는 의미로 쓰인 것인데, simple vs simple일 때도 대립가설의 모수공간이 singleton이기 때문에 다른 simple을 잡을 수 없을 뿐이지 균일하게 최강력이긴 하다. 따라서 구분하지 않아도 무방하지만 그래도 simple hypotheses일 때는 '균일'하다는 말을 붙이는 것이 무의미하긴 하다.

(b) *Karlin-Rubin theorem*에 따라 가능도비를 구하고 *monotone likelihood ratio*가 되는지를 확인해야 한다. $\theta_1 > \theta_0$인 대립가설의 모수 θ_1 하나를 임의로 택하자.

$$\frac{L_1}{L_0} = \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left((\theta_1 - \theta_0)\sum_{i=1}^{n} X_i\right) \leq c_1 \tag{32}$$

$$\sum_{i=1}^{n} X_i \leq c_2 \tag{33}$$

따라서 기각역은 어떤 상수 k에 대해 $\sum_{i=1}^{n} X_i \leq k$일 때이다. 다음을 이용하여 k를 구한다.

$$2\theta \sum_{i=1}^{n} X_i \sim \chi^2(2n) \tag{34}$$

그러므로

$$\Pr\left(2\theta \sum_{i=1}^{n} X_i \leq \chi^2_{2n}(1-\alpha) \,\middle|\, H_0\right) = \alpha \tag{35}$$

4. (25점) Consider the regression model

$$\mathrm{E}(Y_i) = \beta_0 - 2\beta_2 + \beta_1 x_i + 3\beta_2 x_i^2, \quad \mathrm{Var}(Y_i) = \sigma^2 \text{ for } i = 1, 2, 3. \tag{36}$$

Suppose that we observe $x_1 = -1, x_2 = 0$, and $x_3 = 1$.

(a) Find the least square estimates of the parameters β_0, β_1, and β_2.

(b) Find the covariance of the least squares estimate of β_1 and the least squares estimate of β_2.

Solution:

(a) 모형을 행렬표현으로 바꾸면

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} 1 & x_1 & 3x_1^2 - 2 \\ 1 & x_2 & 3x_2^2 - 2 \\ 1 & x_3 & 3x_3^2 - 2 \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}}_{\epsilon} \tag{37}$$

그리고 관측치가 $x_1 = -1, x_2 = 0, x_3 = 1$이므로 계산이 단순해질 것을 기대하고 대입해보자.

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{bmatrix} \tag{38}$$

$$= \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{bmatrix} \tag{39}$$

$$\mathbf{X}'y = \begin{bmatrix} y_1 + y_2 + y_3 \\ y_3 - y_1 \\ y_1 - 2y_2 + y_3 \end{bmatrix} \tag{40}$$

$$\widehat{\beta} = \begin{bmatrix} \frac{1}{3}\left(y_1 + y_2 + y_3\right) \\ \frac{1}{2}\left(y_3 - y_1\right) \\ \frac{1}{6}\left(y_1 - 2y_2 + y_3\right) \end{bmatrix} \tag{41}$$

(b) Y_i가 서로 독립이라는 것과 공분산 연산자의 bilinearity를 이용해서 전개해서 구해도 되겠지만, 일반적으로

$$\mathrm{Var}\left(\widehat{\beta}\right) = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} \tag{42}$$

임을 이용하면 더 쉽다. 왜냐하면 대각원소를 제외한 상삼각행렬/하삼각행렬이 모두 0이기 때문에

$$\mathrm{Cov}\left(\widehat{\beta}_i, \widehat{\beta}_j\right) = \begin{cases} \sigma^2 \mathbf{C}_{(i+1)(i+1)}, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \quad \text{for } i, j = 0, 1, 2 \tag{43}$$

이고 여기서 $\mathbf{C} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}$이다. 따라서 $\mathrm{Cov}\left(\widehat{\beta}_1, \widehat{\beta}_2\right) = 0$이다.