

1. (a) 양의 확률변수(positive random variable)  $X$ 의 1차 적률(평균)이 존재한다고 가정할 때, 다음의 부등식이 성립함을 보이시오.

$$\Pr(X \geq 1) \leq E(X) \quad (1)$$

- (b)  $Z_1, Z_2, \dots$ 이 다음과 같은 확률밀도함수를 갖는 연속형 확률분포를 만족하는 분포에서 나온 임의표본이라고 가정하고

$$f(z) = \begin{cases} \lambda e^{-\lambda z}, & z > 0, \lambda > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$N$ 을 다음과 같은 확률질량함수를 갖는 이산형 확률변수라고 가정하고

$$\Pr(N = n) = \beta(1 - \beta)^{n-1}, \quad n = 1, 2, \dots, \quad 0 < \beta < 1 \quad (3)$$

$Z_1, Z_2, \dots$ 와  $N$ 이 서로 독립이라고 가정하자.

이때, (a)의 결과를 이용하여  $X = \sum_{i=1}^N Z_i$ 가 1보다 크거나 같을 확률의 상한을 구하시오. 또한  $X$ 의 분산을 구하시오.

**Solution:**

- (a) 마코프 부등식(Markov's inequality)는 증명하는 방법이 꽤 여러가지다. 가장 대표적인 3가지를 소개한다.

- 일반적으로  $a > 0$ 이라 가정할 때

$$aI_{(X \geq a)} = \begin{cases} a, & \text{if } X \geq a \\ 0, & \text{if } X < a \end{cases} \quad (4)$$

따라서  $aI_{(X \geq a)} \leq X$ 가 성립함을 알 수 있다. 기댓값 연산자(expectation operator)  $E$ 은 본질적으로 적분이므로 부등호의 방향이 바뀌지 않는다. 그러므로

$$aE(I_{(X \geq a)}) \leq E(X) \quad (5)$$

이 성립하고 indicator variable의 평균은 그 집합의 확률이 되므로

$$\Pr(X \geq a) \leq \frac{E(X)}{a} \quad (6)$$

이 된다. 따라서 주어진 문제의  $a = 1$ 인 경우도 성립한다.

- 두 번째 증명은 기댓값을 적분으로 표현하여 증명한 것이다.

$$E(X) = \int_0^{\infty} x f_X(x) dx \quad (7)$$

$$= \int_0^a x f_X(x) dx + \int_a^{\infty} x f_X(x) dx \quad (8)$$

$$\geq \int_0^a 0 \cdot f_X(x) dx + \int_a^{\infty} a f_X(x) dx \quad (9)$$

$$= a \Pr(X \geq a) \quad (10)$$

(9)에서 두 번째 적분은 적분하는  $x$ 의 범위가  $a < x < \infty$ 이므로 하한인  $x = a$ 로 고정하게 되면 당연히 (8)의 두 번째 적분보다 작아지게 된다. 따라서 똑같이 나온다.

- 마지막은 Lebesgue theory를 이용해서 증명하는 방법이 있다. 일반적인 르벡 측도 (Lebesgue measure)  $\mu$ 가 있을 때 함수  $f \geq 0$ 에 대하여 다음과 같은 simple function을 생각할 수 있다.

$$s(x) = \begin{cases} a, & \text{if } f(x) \geq a \\ 0, & \text{if } f(x) < a \end{cases} \quad (11)$$

그러면  $0 \leq s(x) \leq f(x)$ 이 되므로 르벡 적분(Lebesgue integral)의 성질에 따라 전체 집합  $\Omega$ 에 대해

$$\int_{\Omega} f d\mu \geq \int_{\Omega} s d\mu \quad (12)$$

가 성립하고 simple function의 르벡 적분은

$$\int_{\Omega} s d\mu = a\mu(\{\omega \in \Omega \mid f(\omega) \geq a\}) \quad (13)$$

이 되어

$$\mu(\{\omega \in \Omega \mid f(\omega) \geq a\}) \leq \frac{1}{a} \int_{\Omega} f d\mu \quad (14)$$

가 된다. 원래 확률변수는 표본공간에서 실수 집합  $\mathbb{R}$ 로의 대응관계(mapping)이므로 위 증명에서 정의된 함수  $f$ 가 된다.

- (b) 먼저 마코프 부등식을 이용하여 상한을 구하려면 기댓값을 알아야 한다. 이 기댓값을 구하는 것도 크게 두 가지 방법이 있다.

- 먼저 정의를 그대로 이용하는 방법이다.

$$E(X) = E\left(\sum_{i=1}^N Z_i\right) \quad (15)$$

$$= E(N Z_1) \quad (16)$$

$$= E(N) E(Z_1) \quad (17)$$

$$= \frac{1}{\beta\lambda} \quad (18)$$

가 된다. 왜냐하면  $N \sim \text{Geo}(\beta)$ 이고  $Z_i \sim \text{Exp}(\lambda^{-1})$ 이기 때문이다. 다시 여기서 지수분포는 rate parameter로 표기되었다.

- 두 번째는 조건부로 표현하는 것이다. 다시 말해, 원래는  $X | N = \sum_{i=1}^N Z_i$ 이기 때문에 온전히  $X$ 만의 기댓값을 구하기 위해서는 이중기댓값의 법칙(the law of double expectation)을 이용할 수 있다.  $X | N \sim \text{Ga}(N, \lambda)$ 이다. 고로

$$E(E(X | N)) = E\left(\frac{N}{\lambda}\right) \quad (19)$$

$$= \frac{1}{\lambda} E(N) \quad (20)$$

$$= \frac{1}{\beta\lambda} \quad (21)$$

가 된다.

아무튼 이렇게 해서 마코프 부등식은 다음과 같이 표현된다.

$$\Pr(X \geq 1) \leq \frac{1}{\beta\lambda} \quad (22)$$

- 또한  $X$ 의 분산을 구하기 위해서는 역시 분산의 분해(variance decomposition)를 이용하는 편이 편리하다.

$$\text{Var}(X) = \text{Var}(E(X | N)) + E(\text{Var}(X | N)) \quad (23)$$

$$= \text{Var}\left(\frac{N}{\lambda}\right) + E\left(\frac{N}{\lambda^2}\right) \quad (24)$$

$$= \frac{1}{\lambda^2} \left( \frac{1-\beta}{\beta^2} + \frac{\beta}{\beta^2} \right) \quad (25)$$

$$= \frac{1}{\beta^2\lambda^2} \quad (26)$$

이다. 여기서 기하분포의 평균은 모수의 역수이고 분산은  $(1-\beta)/\beta^2$ 이 됨을 기억하자. 까먹었으면 직접 구해도 된다. 귀찮을 뿐.

- 직접 구해도 된다. 즉  $\text{Var}(X) = E(X^2) - E(X)^2$ 이므로

$$E(X^2) = E\left(\sum_{i=1}^N Z_i^2 + \sum_{i \neq j} Z_i Z_j\right) \quad (27)$$

$$= E(NZ_1) + E\left(\frac{N(N+1)}{2} Z_1 Z_2\right) \quad (28)$$

$$= E(N)E(Z_1) + \frac{1}{2}(E(N^2) + E(N))E(Z_1)E(Z_2) \quad (29)$$

$$= E(N)E(Z_1) + \frac{1}{2}(E(N^2) + E(N))\{E(Z_1)\}^2 \quad (30)$$

알아서 잘 하면 첫 번째와 똑같이 나온다. 귀찮아서 못하겠다.

2.  $X_1, \dots, X_n$ 이 다음과 같은 확률질량함수(probability mass function)를 갖는 이산형 확률분포로부터의 임의 표본이라고 할 때

$$\Pr(X = x) = \begin{cases} \theta, & x = -1, 0 < \theta < 1 \\ (1 - \theta)^2 \theta^x, & x = 0, 1, 2, \dots \end{cases} \quad (31)$$

모수  $\theta$ 에 관하여 다음의 두 가지 추정량을 고려하자.

$$T_1(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n I(X_i = -1), \quad T_2(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i I(X_i \geq 0). \quad (32)$$

- (a)  $n^{-1} \sum_{i=1}^n X_i = T_2 - T_1$ 이 됨을 보이고, 이를 이용하여 두 추정량이 각각  $\theta$ 의 비편향 추정량(unbiased estimator)임을 보이시오.
- (b) 두 추정량의 MSE(mean square error)를 구하고 비교하시오.
- (c) 모두  $\theta$ 의 최대가능도 추정량은(maximum likelihood estimator) 다음과 같이 주어짐을 보이고

$$\hat{\theta}_n = \frac{2 \sum_{i=1}^n I(X_i = -1) + \sum_{i=1}^n X_i}{2n + \sum_{i=1}^n X_i} \quad (33)$$

$\hat{\theta}_n^2$ 이  $\theta^2$ 의 일치추정량(consistent estimator)이 됨을 보이시오.

**Solution:**

- (a)  $X_i$ 가 취할 수 있는 값은 0부터가 아니라  $-1, 0, 1, 2, \dots$ 이므로 0보다 큰  $X_i$ 들을 모두 더했을

때  $T_2$ 가 된다. 왜냐하면  $I(X_i \geq 0)$ 항은  $X_i = -1$ 일 때만 0이 되고 그 이외에서는 1이 되기 때문에 0보다 크거나 같은 확률변수만 남겨두는 효과를 낳게 된다. 따라서 모두 더한 값은  $X_i = -1$ 인 횟수만큼  $T_2$ 에서 빼주어야 하는데 그 ‘횟수’가 바로  $T_1$ 이다. 횟수는 양수인데 반해  $X_i = -1$ 은 음수여야 하므로  $T_2$ 에서  $T_1$ 을 빼주어야 그냥 다 더해서 나눈 값이 된다. 수식으로 말하자면

$$-T_1 = \frac{1}{n} \sum_{i=1}^n X_i I(X_i = -1) \quad (34)$$

이므로

$$\underbrace{\frac{1}{n} \sum_{i=1}^n X_i I(X_i \geq 0)}_{T_2} + \underbrace{\frac{1}{n} \sum_{i=1}^n X_i I(X_i = -1)}_{-T_1} = \frac{1}{n} \sum_{i=1}^n X_i \underbrace{(I(X_i = -1) + I(X_i \geq 0))}_{\text{서로소인 집합이므로 항상 1}} \quad (35)$$

이 되어  $\bar{X}_n = T_2 - T_1$ 이 된다. 각각의 기댓값은

$$E(T_1) = \frac{1}{n} \sum_{i=1}^n E(I(X_i = -1)) \quad (36)$$

$$= \frac{1}{n} \sum_{i=1}^n \Pr(X_i = -1) \quad (37)$$

$$= \frac{1}{n} \cdot n\theta = \theta \quad (38)$$

$$E(T_2) = \frac{1}{n} \sum_{i=1}^n E(X_i I(X_i \geq 0)) \quad (39)$$

$$= \frac{1}{n} \sum_{i=1}^n \sum_{x=0}^{\infty} x(1-\theta)^2 \theta^x, \quad \text{먹급수 계산은 알아서...} \quad (40)$$

$$= \frac{1}{n} \sum_{i=1}^n \theta = \theta \quad (41)$$

(b) 둘 모두 비편향 추정량이므로 *variance-bias decomposition*에 의해 MSE를 비교하는 것은 분산을 비교하는 것과 같아진다. 둘의 분산을 구하면 다음과 같다.

$$\text{Var}(T_1) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n I(X_i = -1)\right) \quad (42)$$

$$= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(I(X_i = -1)) \quad (43)$$

$$= \frac{1}{n^2} \sum_{i=1}^n \left( E([I(X_i = -1)]^2) - [E(I(X_i = -1))]^2 \right) \quad (44)$$

이산형 확률변수의 함수의 기댓값은 예전에 고등학교 때  $X$ 와  $\Pr(X)$  표를 그려서 하나

하나 곱해서 더하던 것을 생각해보면 편하다. 함수꼴의 기댓값 역시 *the law of unconscious statistician*에 의해 그냥 원래 확률을 그대로 써도 된다. 고로  $[I(X_i = -1)]^2$ 은 언제나 1 이므로 그 기댓값은  $\theta$ 가 되고  $I(X_i = -1)$ 의 기댓값은 그 자체로도  $\theta$ 이므로

$$\text{Var}(T_1) = \frac{1}{n^2} n (\theta - \theta^2) \quad (45)$$

$$= \frac{1}{n} \theta (1 - \theta) \quad (46)$$

이다.  $T_2$ 의 분산 역시 같은 방식으로 하면

$$E(X_i I(X_i \geq 0)) = \sum_{x=0}^{\infty} x (1 - \theta)^2 \theta^x = \theta \quad (47)$$

$$E([X_i I(X_i \geq 0)]^2) = \sum_{x=0}^{\infty} x^2 (1 - \theta)^2 \theta^x = \frac{\theta(\theta + 1)}{1 - \theta} \quad (48)$$

$$\text{Var}(T_2) = \frac{\theta(\theta + 1)}{1 - \theta} - \theta^2 \quad (49)$$

$$= \frac{\theta(\theta^2 + 1)}{1 - \theta} \quad (50)$$

둘의 분산을 비교해 보았을 때  $T_1$ 의 분산이 더 작으므로  $T_1$ 이 MSE의 측면에서 더 좋은 추정량이다.

(c)  $n$ 개의 확률변수 중  $X_i = -1$ 인 개수를  $m$ 개라 하면  $\sum_{i=1}^n I(X_i = -1) = m$ 이 되고 가능도 함수는 다음과 같다.

$$L(\theta | \{X_i\}_{i=1}^n) = \theta^m (1 - \theta)^{2(n-m)} \theta^{\sum_{i=1}^n X_i I(X_i \geq 0)} \quad (51)$$

$$\ell(\theta | \{X_i\}_{i=1}^n) = \left( m + \sum_{i=1}^n X_i I(X_i \geq 0) \right) \ln \theta + 2(n - m) \ln(1 - \theta) \quad (52)$$

$$\frac{d}{d\theta} \ell(\theta | \{X_i\}_{i=1}^n) = \frac{m + \sum_{i=1}^n X_i I(X_i \geq 0)}{\theta} - \frac{2(n - m)}{1 - \theta} \quad (53)$$

$$= \frac{m + \sum_{i=1}^n X_i I(X_i \geq 0) - \theta \left( m + \sum_{i=1}^n X_i I(X_i \geq 0) + 2(n - m) \right)}{\theta(1 - \theta)} \quad (54)$$

고로 이를 0으로 놓고  $\theta$ 에 대해 정리한 뒤  $m$ 을 원래대로 복원하면

$$\hat{\theta}_n = \frac{\sum_{i=1}^n I(X_i = -1) + \sum_{i=1}^n X_i I(X_i \geq 0)}{2n + \sum_{i=1}^n X_i I(X_i \geq 0) - \sum_{i=1}^n I(X_i = -1)} \quad (55)$$

$$= \frac{-\sum_{i=1}^n (-1) I(X_i = -1) + \sum_{i=1}^n X_i I(X_i \geq 0)}{2n + \sum_{i=1}^n X_i I(X_i \geq 0) + \sum_{i=1}^n (-1) I(X_i = -1)} \quad (56)$$

$$= \frac{-2 \sum_{i=1}^n (-1) I(X_i = -1) + \sum_{i=1}^n X_i I(X_i \geq 0) + \sum_{i=1}^n (-1) I(X_i = -1)}{2n + \sum_{i=1}^n X_i} \quad (57)$$

$$= \frac{2 \sum_{i=1}^n I(X_i = -1) + \sum_{i=1}^n X_i}{2n + \sum_{i=1}^n X_i} \quad (58)$$

마지막으로  $\hat{\theta}_n$ 은  $\theta$ 의 일치추정량이다. 왜냐하면 대수의 법칙에 의해

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} E(X_1) \quad (59)$$

인데

$$E(X_1) = -\theta + \sum_{x=0}^{\infty} x(1-\theta)^2 \theta^x = -\theta + \theta = 0 \quad (60)$$

이므로  $\bar{X}_n \xrightarrow{p} 0$ 이며

$$\frac{1}{n} I(X_i = -1) \xrightarrow{p} E(I(X_1 = -1)) \quad (61)$$

이므로  $\theta$ 에 수렴하게 된다. 따라서 *Slutsky's theorem*에 의해서

$$\hat{\theta}_n \xrightarrow{p} \frac{2\theta + 0}{2 + 0} = \theta \quad (62)$$

이므로 MLE의 불변성(invariance property)에 의해  $\hat{\theta}_n^2$  역시  $\theta^2$ 의 일치추정량이 된다.

3.  $X_1, \dots, X_n$ 이  $\mathcal{N}(\mu, \sigma^2)$ 로부터의 랜덤포본이라 할 때, 모평균에 대한 가설  $H: \mu = \mu_0$  vs  $K: \mu \neq \mu_0$ 에 대한  $t$  검정법을 가능도비를 사용하여 유도하고, 유도된 검정법의 검정력 및 성질에 대해 논하라. ( $\sigma^2$ 는 알려지지 않았음.)

**Solution:**  $\mu$ 의 MLE는  $\bar{X}_n$ 임이 자명하므로 다시 구하지 않겠다. 가능도비를 구하면

$$\frac{L(\mu_0)}{L(\hat{\mu}^{\text{MLE}})} = \frac{\exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu_0)^2\right)}{\exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2\right)} \quad (63)$$

$$= \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (X_i - \mu_0)^2\right) - \sum_{i=1}^n (X_i - \bar{X}_n)^2\right) \quad (64)$$

$$= \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (X_i - \bar{X}_n)^2 + n(\bar{X}_n - \mu_0)^2 - \sum_{i=1}^n (X_i - \bar{X}_n)^2\right)\right) \quad (65)$$

$$= \exp\left(-\frac{1}{2\sigma^2} (\bar{X}_n - \mu_0)^2\right) \leq k \quad (66)$$

일 때 귀무가설이 기각된다. 따라서  $(\bar{X}_n - \mu_0)^2 \geq k_2$ 일 때, 다시 고치면

$$\left(\frac{\bar{X}_n - \mu_0}{S}\right)^2 \geq \frac{k_2}{S^2} \quad (67)$$

$$\underbrace{\left(\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{S}\right)^2}_{T^2} \geq n \frac{k_2}{S^2} \quad (68)$$

이므로  $T^2$ 가 충분히 클 때 귀무가설은 기각된다. 다시 말해서  $T$ 가 충분히 크거나 충분히 작을 때 귀무가설이 기각된다는 의미이다. 그리고  $T$ 는 자유도가  $n-1$ 인 t-분포를 따르므로  $T \leq t_{n-1}(1-\alpha)$  이거나  $T \geq t_{n-1}(\alpha)$ 일 때  $\alpha$  신뢰수준에서 귀무가설은 기각된다.



4. 단순회귀모형(simple regression model)  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 을 고려하자. 여기서  $\epsilon_i$ 는 기본회귀모형 가정을 만족하고  $i = 1, 2, \dots, n$ 이다.

(a) 위 모형을 행렬 형식으로 자세히 표현하고, 행렬 형태로 회귀 모수  $\beta = (\beta_0, \beta_1)'$ 의 최소제곱추정치(least square estimator; LSE)를 구하시오.

(b) LSE  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)'$ 의 분포를 구하시오.

(c)  $\hat{\beta}$ 와  $S^2$ 가 독립임을 보이시오. 여기서  $S^2 = \text{SSE}/(n-2) = (Y - \mathbf{X}\hat{\beta})'(Y - \mathbf{X}\hat{\beta})$ 이다.

**Solution:**

(a) 위 모형은 행렬식으로 표현하면

$$\underbrace{\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}}_Y = \underbrace{\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}}_X \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{\epsilon} \quad (69)$$

이 되고 LSE는  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'Y$ 이다.

(b)  $\hat{\beta}$ 는  $Y$ 의 선형변환이고  $Y \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$ 이므로 정규분포의 선형불변성(affine property)에 의해 어떠한 선형변환꼴도 다시 정규분포를 따르게 되어 있다. 따라서

$$E(\hat{\beta}) = E((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'Y) \quad (70)$$

$$= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}\beta \quad (71)$$

$$= \beta \quad (72)$$

$$\text{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\text{Var}(Y) \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \quad (73)$$

$$= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \cancel{\mathbf{X}'\mathbf{X}} (\mathbf{X}'\mathbf{X})^{-1} \quad (74)$$

$$= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \quad (75)$$

$$\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}) \quad (76)$$

(c) 먼저  $\text{Cov}(\hat{\beta}, Y - \mathbf{X}\hat{\beta})$ 을 구하면

$$\text{Cov}(\hat{\beta}, Y - \mathbf{X}\hat{\beta}) = \text{Cov}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y, (\mathbf{I}_n - \mathbf{H})Y) \quad (77)$$

$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Cov}(Y)(\mathbf{I}_n - \mathbf{H})' \quad (78)$$

$$= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_n - \mathbf{H}) \quad (79)$$

$$= \mathbf{0} \quad (80)$$

정규분포의 경우 공분산이 0이면 서로 독립이므로  $\hat{\beta}$ 와  $Y - \mathbf{X}\hat{\beta}$ 는 서로 독립이며 그 함수꼴 역시 서로 독립이므로  $\hat{\beta}$ 와  $S^2$  역시 서로 독립이다.