Daeyoung Lim

Grad School

Year of 2016, late

1. X는 자연수의 값만 갖는 이산확률변수로서, $p_k = \Pr(X=k)$는 $\beta > 0, r > -1\ (r \neq 0)$에 대해 다음의 점화식을 만족한다.

$$\frac{p_k}{p_{k-1}} = \frac{\beta}{\beta+1}\left(1+\frac{r-1}{k}\right), \quad k = 2, 3, 4, \ldots$$

   (a) $p_k \ (k = 1, 2, \cdots)$를 r, β, k의 식으로 표현하시오. 이때, 다음의 결과를 이용할 수 있다.

$$\forall x \text{ s.t. } |x| < 1, \forall r \in \mathbb{R}, \ (1-x)^{-r} = \sum_{k=0}^{\infty} \binom{-r}{k} (-x)^k = \sum_{k=0}^{\infty} \binom{r+k-1}{k} x^k$$

   (b) 이 분포의 확률생성함수(probability generating function)를 구하고, 이것을 사용하여 $\mathrm{E}(X(X-1))$을 구하시오.

Solution:

(a) 점화식을 풀기 위해 다음과 같이 곱한다.

$$\frac{\cancel{p_2}}{p_1} \times \frac{\cancel{p_3}}{\cancel{p_2}} \times \cdots \times \frac{p_k}{\cancel{p_{k-1}}} = \prod_{x=2}^{k} \frac{\beta}{\beta+1}\left(1+\frac{r-1}{x}\right) \tag{1}$$

상수인 베타의 분수 항을 제외하고 실제로 변하는 x가 들어있는 항을 정리하면 다음과 같다.

$$\prod_{x=2}^{k}\left(1+\frac{r-1}{x}\right) = \frac{1}{r}\left(\frac{r}{1} \times \frac{r+1}{2} \times \frac{r+2}{3} \times \cdots \times \frac{k+r-1}{k}\right) \tag{2}$$

$$= \frac{1}{r}\binom{r+k-1}{k} \tag{3}$$

그러면 (1)은 다음과 같이 정리된다.

$$\frac{p_k}{p_1} = \frac{1}{r}\binom{r+k-1}{k}\left(\frac{\beta}{\beta+1}\right)^{k-1} \tag{4}$$

p_k는 확률질량함수이므로 $k = 1, 2, \ldots$에 대해 다 더하면 1이 되어야 한다. 따라서

$$\sum_{k=1}^{\infty} \frac{p_k}{p_1} = \sum_{k=1}^{\infty} \frac{1}{r} \binom{r+k-1}{k} \left(\frac{\beta}{\beta+1}\right)^{k-1} \tag{5}$$

$$= \sum_{k=1}^{\infty} \frac{1}{r} \binom{r+k-1}{k} \left(\frac{\beta}{\beta+1}\right)^{k} \cdot \frac{\beta+1}{\beta} \tag{6}$$

$$= \frac{1}{r} \frac{\beta+1}{\beta} \left(\sum_{k=0}^{\infty} \binom{r+k-1}{k} \left(\frac{\beta}{\beta+1}\right)^{k} - 1\right) \tag{7}$$

$$\frac{1}{p_1} = \frac{1}{r} \frac{\beta+1}{\beta} \left(\left(1 - \frac{\beta}{\beta+1}\right)^{-r} - 1\right) \tag{8}$$

$$p_k = \binom{r+k-1}{k} \left(\frac{\beta}{\beta+1}\right)^{k} \Bigg/ \left(\left(\frac{1}{\beta+1}\right)^{-r} - 1\right) \tag{9}$$

(b) 확률생성함수는 오직 '이산확률변수'에 대해서만 정의할 수 있으며 해석학에서 나왔던, 확률질량함수의 *'power series'*에 해당된다. 그러므로 다음과 같다.

$$G(z) = \mathrm{E}\left(z^X\right) = \sum_{x=0}^{\infty} p(x)\, z^x \tag{10}$$

*'Power series'*는 $|z| \le 1$인 복소수 $z \in \mathbb{C}$에 대해서 절대수렴(*absolute convergence*)한다. 본 문제의 확률질량함수에 대한 확률생성함수를 구해보자. 편의상 (a)의 마지막 p_k의 분모에 있는 정규화 상수(*normalizing constant*)를 C라고 놓자.

$$\sum_{x=1}^{\infty} p_x z^x = \frac{1}{C} \sum_{x=1}^{\infty} \binom{r+x-1}{x} \left(\frac{\beta}{\beta+1}\right)^{x} z^x \tag{11}$$

$$= \frac{1}{C} \sum_{x=1}^{\infty} \binom{r+x-1}{x} \left(\frac{\beta z}{\beta+1}\right)^{x} \tag{12}$$

$$= \frac{1}{C} \left(\sum_{x=0}^{\infty} \binom{r+x-1}{x} \left(\frac{\beta z}{\beta+1}\right)^{x} - 1\right) \tag{13}$$

$$= \frac{1}{C} \left(\left(1 - \frac{\beta z}{\beta+1}\right)^{-r} - 1\right) \tag{14}$$

확률생성함수의 특징은 'factorial moment'를 구할 수 있다는 것이다. 즉,

$$\left.\frac{d^k}{dz^k} G(z)\right|_{z=1} = \mathrm{E}\left(X(X-1)\cdots(X-k+1)\right) \tag{15}$$

직접 전개해서 미분해보면 알 수 있다. 따라서 문제에서 요구한 'second order factorial moment'는 두번 미분해서 $z = 1$를 대입하면 되므로

$$\frac{d^2}{dz^2} G(z) = \frac{r(r+1)}{C} \left(\frac{\beta}{\beta+1}\right)^2 \left(1 - \frac{\beta z}{\beta+1}\right)^{-r-2} \tag{16}$$

$z = 1$ 대입하고 C와 함께 예쁘게 정리하면 다음과 같다.

$$\mathrm{E}(X(X-1)) = \frac{r(r+1)\beta^2}{(\beta+1)^r} \tag{17}$$

2. Let $X_1, X_2, \ldots, X^n \overset{iid}{\sim} \text{Unif}(0, \theta)$.

   (a) Show that the sample maximum, $X_{(n)} = \max(X_1, \ldots, X_n)$ is the maximum likelihood estimator (MLE) of θ.

   (b) Compute $\mathrm{E}\left(X_{(n)}\right)$ and $\mathrm{Var}\left(X_{(n)}\right)$.

   (c) Show that $X_{(n)}$ is a consistent estimator of θ. That is, show that for an arbitrary given $\epsilon > 0$,

   $$\lim_{n\to\infty} \Pr\left(\left|X_{(n)} - \theta\right| < \epsilon\right) = 1.$$

   (d) It is known that $X_{(n)}$ is a complete and sufficient statistic for θ. Find the minimum variance unbiased estimator (MVUE) of θ.

   (e) Let $\widehat{\theta}_1$ and $\widehat{\theta}_2$ denote the MLE and MVUE of θ respectively. Compare the mean squared error (MSE) of $\widehat{\theta}_1$ and $\widehat{\theta}_2$. Which one is better in terms of the MSE?

Solution:

(a) 가능도함수를 지표함수(*indicator function*)과 함께 쓰면 다음과 같다.

$$L\left(\theta \mid \{X_i\}_{i=1}^n\right) = \left(\frac{1}{\theta}\right)^n I_{\left(X_{(n)}, \infty\right)}(\theta) \tag{18}$$

그러므로 단조감소함수인 θ^{-n}의 최댓값은 가장 작은 값인 $X_{(n)}$이다.

(b) CDF부터 시작하면

$$\begin{aligned}
\Pr\left(X_{(n)} \le x\right) &= \Pr(X_1 \le x) \Pr(X_2 \le 2) \cdots \Pr(X_n \le x) && (19)\\
&= \left(\Pr(X_1 \le x)\right)^n && (20)\\
&= \left(\frac{x}{\theta}\right)^n && (21)
\end{aligned}$$

그러므로 $X_{(n)}$의 PDF는 다음과 같다.

$$f_{X_{(n)}}(x) = nx^{n-1}\theta^{-n}, \quad 0 < x < \theta \tag{22}$$

이를 통해 1,2차 적률을 통해 분산을 구하면

$$\begin{aligned}
\mathrm{E}\left(X_{(n)}\right) &= n\theta^{-n}\int_0^\theta x^n\,dx &(23)\\
&= \frac{n}{n+1}\theta &(24)\\
\mathrm{E}\left(X_{(n)}^2\right) &= n\theta^{-n}\int_0^\theta x^{n+1}\,dx &(25)\\
&= \frac{n}{n+1}\theta^2 &(26)\\
\mathrm{Var}\left(X_{(n)}\right) &= \frac{n}{(n+1)^2(n+2)}\theta^2 &(27)
\end{aligned}$$

(c) 마코프 부등식(*Markov's inequality*)를 이용하면 쉽게 풀 수 있다.

$$\Pr\left(\left|X_{(n)}-\theta\right| \geq \epsilon\right) \leq \frac{\mathrm{E}\left|X_{(n)}-\theta\right|}{\epsilon} \tag{28}$$

$X_{(n)} < \theta$이므로 우변의 분자는 $(-)$가 곱해져서 나온다. 고로

$$\Pr\left(\left|X_{(n)}-\theta\right| \geq \epsilon\right) \leq \frac{1}{\epsilon}\left(\theta - \frac{n}{n+1}\theta\right) \tag{29}$$

극한을 취하면

$$\lim_{n\to\infty}\Pr\left(\left|X_{(n)}-\theta\right| \geq \epsilon\right) \leq \lim_{n\to\infty}\frac{\theta}{\epsilon(n+1)} \tag{30}$$

샌드위치 정리에 의해 우변이 0으로 간다. 확률측도에 대해서 다음을 정의하자. 표본공간(전체집합)을 Ω라 했을 때 $\omega \in \Omega$에 대해서

$$A = \left\{\omega \,\middle|\, \left|X_{(n)}(\omega)-\theta\right| \geq \epsilon, \quad \text{for some } \epsilon > 0\right\} \tag{31}$$

인데 문제에서 주어진 것은 A^{c}의 측도(*measure*)이므로 σ-algebra의 성질을 이용하여 다음과 같이 구할 수 있다.

$$\Pr\left(A^{\mathsf{c}}\right) = \Pr\left(\Omega \setminus A\right) \tag{32}$$

그런데 다음에 의해

$$A \cup A^{\mathsf{c}} = \Omega \implies \Pr(A) + \Pr\left(A^{\mathsf{c}}\right) = \Pr(\Omega) \tag{33}$$

문제에 주어진 A^c의 측도는 우리가 구한 A에 대해 다음과 같은 성질을 만족한다. 이는 유한 측도 공간(*finite measure space*)이기 때문에 가능하다.

$$\lim_{n\to\infty} \Pr\left(A^c\right) = \lim_{n\to\infty}\left(1 - \Pr\left(A\right)\right) \tag{34}$$

따라서 다음이 성립한다.

$$\lim_{n\to\infty} \Pr\left(\left|X_{(n)} - \theta\right| < \epsilon\right) = 1 \tag{35}$$

(d) *Lehmann Scheffé lemma*에 의해 완비충분통계량이 존재한다면 최소분산비편향 추정량은 완비충분통계량의 어떤 함수꼴로 표현된다. 따라서 $X_{(n)}$가 완비충분통계량이라면 이를 비편향 추정량으로만 만들면 최소분산이 된다. 고로

$$\frac{n+1}{n}X_{(n)} \tag{36}$$

가 최소분산비편향 추정량이 된다.

(e) *Variance-bias decomposition*에 의해 어떤 추정량 $\widehat{\theta}$의 MSE는 다음과 같이 분해된다.

$$\text{MSE}\left(\widehat{\theta}\right) = \text{Var}\left(\widehat{\theta}\right) + \text{Bias}^2\left(\widehat{\theta}\right) \tag{37}$$

우선 MLE는 비편향추정량이 아니므로 편향이 없어지지 않는다. 따라서 그대로 계산하면

$$\text{MSE}\left(\widehat{\theta}_1\right) = \frac{2}{(n+1)(n+2)}\theta^2 \tag{38}$$

그리고 비편향추정량인 $\widehat{\theta}_2$의 MSE는 분산과 같다. 고로

$$\text{MSE}\left(\widehat{\theta}_2\right) = \text{Var}\left(\frac{n+1}{n}X_{(n)}\right) = \left(\frac{n+1}{n}\right)^2 \text{Var}\left(X_{(n)}\right) = \frac{n}{n^2(n+2)}\theta^2 \tag{39}$$

둘을 비교하면 최소분산비편향 추정량의 MSE가 더 작다.

3. Suppose a box contains four marbles, θ white ones and $4-\theta$ black ones. Let two marbles be drawn from this box without replacement. Define a random variable X as

$$X = \begin{cases} 1, & \text{if both selected marbles are same color} \\ 0, & \text{otherwise} \end{cases}$$

Test $H_0 : \theta = 2$ against $H_1 : \theta \neq 2$ as follows; reject H_0 if both selected marbles are the same color.

(a) Derive the probability mass function of the discrete random variable X. That is, erive the probabilities $\Pr(X=1)$ and $\Pr(X=0)$ as functions of θ.

(b) Derive the power function, which is the probability of rejecting $H_0 : \theta = 2$, of this test.

(c) What is the significance level α of this test?

Solution:

(a) 흰공 두개가 나오는 사건을 WW라 표기하고 검은공 두개가 나오는 사건을 BB라 표기하자. 그러면 그 확률은 다음과 같다.

$$\Pr(\mathrm{WW}) = \frac{\theta(\theta-1)}{12} \tag{40}$$

$$\Pr(\mathrm{BB}) = \left(1-\frac{\theta}{4}\right)\left(1-\frac{\theta}{3}\right) \tag{41}$$

$X=1$은 WW 혹은 BB이 일어날 사건이므로

$$\Pr(X=1) = \frac{\theta^2-4\theta+6}{6} \tag{42}$$

$$\Pr(X=0) = \frac{\theta(4-\theta)}{6} \tag{43}$$

(b) $X=1$일 때 기각한다고 했으므로 (a)에서 (42)번 식이 검정력함수이다.

(c) 검정력함수에 귀무가설에서 상정한 모수값을 집어넣은 것이 유의수준이다. 따라서

$$\alpha = 1/3 \tag{44}$$

4. n개의 자료가 단순선형회귀모형

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n \tag{45}$$

을 따른다고 가정하자. 여기서 y_i는 i번째 개체의 반응변수값, x_i는 i번째 개체의 설명변수값, 그리고 β_0와 β_1의 회귀계수들을 의미하며 ϵ_i은 평균이 0, 분산이 x_i의 값이 증가함에 따라 증가하는 분포를 따른다고 한다.

(a) β_0과 β_1의 가중최소제곱 추정량(weighted least squares estimator) $\widehat{\beta}_0$와 $\widehat{\beta}_1$을 각각 구하시오.

(b) (a)에서 구한 $\widehat{\beta}_0$와 $\widehat{\beta}_1$이 각각 β_0과 β_1의 비편향 추정량(unbiased estimator)인지 아닌지 보이시오.

Solution:

(a) GLS에서 오차항의 공분산행렬을 $\sigma^2\mathbf{V}$라고 가정하고 그것을 단위행렬로 돌리기 위해서 *square-root matrix*를 곱해주었듯이, WLSE에서도 오차항의 공분산행렬을 다음과 같이 정의한다.

$$\mathbf{W} = \operatorname{diag}\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2\right) \tag{46}$$

그리고 다음과 같이 변환해준다.

$$\mathbf{W}^{-1/2}Y = \mathbf{W}^{-1/2}\mathbf{X}\beta + \mathbf{W}^{-1/2}\epsilon \tag{47}$$

그러면 $\mathbf{W}^{-1/2}\epsilon$의 공분산행렬은 $\mathbf{I}_n$이 된다. 이제 선형회귀모형의 등분산성(*homoskedasticity*)를 만족하므로 일반적인 최소제곱법을 써서 추정할 수 있다.

$$\widehat{\beta}^{\text{WLSE}} = \operatorname*{argmin}_{\beta}\left(Y - \mathbf{X}\beta\right)'\mathbf{W}^{-1}\left(Y - \mathbf{X}\beta\right) \tag{48}$$

$$= \left(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}^{-1}Y \tag{49}$$

본 문제는 단순선형회귀모형이므로 손으로 하기 귀찮긴 하지만 못하진 않는다. 다음 두 등식을 이용하면 편하다. $\mathbf{x}_i$를 $\mathbf{X}$행렬의 i번째 행을 열벡터(*column vector*)로 세운 것을 의미한다고 할 때

$$\mathbf{X}'\mathbf{W}^{-1}\mathbf{X} = \sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i'}{\sigma_i^2} \tag{50}$$

$$\mathbf{X}'\mathbf{W}^{-1}Y = \sum_{i=1}^{n}\frac{y_i\mathbf{x}_i}{\sigma_i^2} \tag{51}$$

이를 이용해서 구해보면

$$\mathbf{x}_i\mathbf{x}' = \begin{bmatrix} 1 \\ x_i \end{bmatrix} \begin{bmatrix} 1 & x_i \end{bmatrix} \tag{52}$$

$$= \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix} \tag{53}$$

$$\sum_{i=1}^{n} \frac{\mathbf{x}_i\mathbf{x}_i'}{\sigma_i^2} = \begin{bmatrix} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} & \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} \\ \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} & \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} \end{bmatrix} \tag{54}$$

$$\sum_{i=1}^{n} \frac{y_i\mathbf{x}_i}{\sigma_i^2} = \begin{bmatrix} \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2} \\ \sum_{i=1}^{n} \frac{x_i y_i}{\sigma_i^2} \end{bmatrix} \tag{55}$$

그리고 (54)의 역행렬을 구하기 위해 행렬식(*determinant*)를 구하자.

$$\det\left(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\right) = \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\right)\left(\sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2}\right) - \left(\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}\right)^2 \tag{56}$$

그러면 $\widehat{\beta}_0, \widehat{\beta}_1$은 다음과 같다.

$$\widehat{\beta}_0^{\text{WLSE}} = \frac{1}{\det\left(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\right)}\left(\sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2}\sum_{i=1}^{n} \frac{y_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}\sum_{i=1}^{n} \frac{x_i y_i}{\sigma_i^2}\right) \tag{57}$$

$$\widehat{\beta}_1^{\text{WLSE}} = \frac{1}{\det\left(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\right)}\left(\sum_{i=1}^{n} \frac{x_i y_i}{\sigma_i^2}\sum_{i=1}^{n} \frac{1}{\sigma_i^2} - \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}\sum_{i=1}^{n} \frac{y_i}{\sigma_i^2}\right) \tag{58}$$

(b) 행렬로 계산하면 쉽다.

$$\mathrm{E}\left(\widehat{\beta}^{\text{WLSE}}\right) = \mathrm{E}\left(\left(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}^{-1}Y\right) \tag{59}$$

$$= \left(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathrm{E}\left(Y\right) \tag{60}$$

$$= \left(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\beta \tag{61}$$

$$= \beta \tag{62}$$

따라서 비편향 추정량이다.