Daeyoung Lim
Grad School
Year of 2014, late

1. Let X be a random variable defined on $[0,\theta]$ with the density function

$$f(x) = 1/\theta, \qquad 0 \le x \le \theta,\ \ 1 < \theta \le 2.$$

Suppose that we only observe $Y = X$ if $0 \le X \le 1$ and $Y = 0$ if $X > 1$,

(a) Obtain the probability density function (PDF) of Y and verify that it is a PDF.

(b) Obtain the cumulative distribution function (CDF) of Y and verify that it is a CDF. Also sketch the cdf.

Solution:

(a) 흔히 연속형 변수 혹은 이산형 변수만 보았는데 여기서 Y가 혼합형 변수이다. 즉

$$Y = \begin{cases} X, & \text{if } 0 \le X \le 1 \\ 0, & \text{if } X > 1 \end{cases} \tag{1}$$

인데 다시 말하면

$$f_Y(y) = \begin{cases} \Pr(0 \le X \le 1), & \text{if } 0 \le y \le 1 \\ \Pr(X > 1), & \text{if } y = 0 \end{cases} \tag{2}$$

이다. $0 \le y \le 1$일 때는 연속형 변수이고, $y = 0$일 때는 이산형이다. 연속형일 때도 $y = 0$일 수 있어서 범위가 잘못된 것처럼 보일 수 있지만 어차피 연속형일 때 $\Pr(Y \in \{0\}) = 0$이므로 null set이 되어 상관없다. 따라서 이것이 PDF임을 보이려면 전체 정의역에서 확률이 1이 되는지 보아야 한다. 우선 확률을 구하자.

$$\Pr(0 \le X \le 1) = \int_0^1 \frac{1}{\theta}\,dx \tag{3}$$

$$= \frac{1}{\theta} \tag{4}$$

$$\Pr(X > 1) = 1 - \frac{1}{\theta} \tag{5}$$

$$f_Y(y) = \begin{cases} \dfrac{1}{\theta}, & \text{if } 0 \le y \le 1 \\ 1 - \dfrac{1}{\theta}, & \text{if } y = 0 \end{cases} \tag{6}$$

따라서

$$\Pr(0 \le Y \le 1) + \Pr(Y = 0) = \int_0^1 \frac{1}{\theta}\, dy + 1 - \frac{1}{\theta} \tag{7}$$

$$= 1 \tag{8}$$

(b) CDF도 마찬가지이다.

$$F_Y(y) = \begin{cases} 1 - \dfrac{1}{\theta}, & \text{if } y = 0 \\ \dfrac{y}{\theta} + \left(1 - \dfrac{1}{\theta}\right), & \text{if } 0 < y \le 1 \end{cases} \tag{9}$$

개형은 각자.

2. 랜덤표본 $X_1, \dots, X_n$의 누적분포(CDF)에 대한 $F(x) = \Pr(X_1 \le x)$의 추정량으로 경험적 확률분포(empirical distribution function)인 $F_n(x) = (\# \text{ of } X_i \le x)/n$을 고려할 때, $F_n(x)$의 일치성(consistency)을 보여라.

Solution: 일치성은 확률수렴(convergence in probability; convergence in measure)이므로 대수의 법칙에 의해 증명할 수 있다. 만약 이것이 *almost-sure convergence*였다면 이를 *Glivenko–Cantelli theorem*이라 부르고 증명이 매우 어려워진다. 아무튼 대수의 법칙에 의해 다음을 알고 있다.

$$\frac{1}{n}\sum_{i=1}^{n} \delta_A(X_i) \xrightarrow{p} \Pr(X_1 \in A) \tag{10}$$

우리가 일반적으로 알고 있는 대수의 법칙과 똑같다. 단지 *dirac-delta function*은 indicator function의 기능을 함으로써 기댓값을 취하면 확률이 될 뿐이다. 즉,

$$\delta_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases} \tag{11}$$

따라서 경험적 확률분포를 다시 쓰면 $A = \{X \mid X \le x \text{ for some } x\}$이라 할 때

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_A(X_i) \tag{12}$$

이 된다. 고로 똑같이

$$\frac{1}{n}\sum_{i=1}^{n}\delta_A\left(X_i\right) \xrightarrow{p} \Pr\left(X_1 \in A\right) \quad \left(=\Pr\left(X_1 \le x\right)\right) \tag{13}$$

따라서 $F_n\left(x\right) \xrightarrow{p} F\left(x\right)$.

3. Let $X_1, \ldots, X_n$ be a random sample from a distribution with probability density function $f\left(x;\theta\right) = 1/\theta$ if $0 \le x \le \theta$ and zero otherwise. Derive the likelihood ratio test of size α of $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.

Solution: MLE는 $X_{(n)}$이므로 가능도비는

$$\Lambda = \begin{cases} \left(\dfrac{X_{(n)}}{\theta_0}\right)^n, & \text{if } X_{(n)} \le \theta_0 \\ 0, & \text{if } X_{(n)} > \theta_0 \end{cases} \tag{14}$$

이는 당연한 것이 모수의 참값이 θ_0라면 관측치 중 가장 큰 값 $X_{(n)}$이 θ_0보다 절대 커지면 안 된다. 만약 크다면 귀무가설을 기각할 수밖에 없다. 따라서 상식에도 부합한다. 엄밀하게 계산하고 싶으면 indicator function을 이용해서 구하면 된다. 따라서 $\Lambda \le k$일 때 기각역이므로

$$\text{RR} = \left\{X_{(n)} \,|\, X_{(n)} \le \theta_0 k^{1/n} \quad \text{or} \quad X_{(n)} > \theta_0\right\} \tag{15}$$

이 된다.

4. 다음과 같은 단순선형회귀모형

$$Y = \beta_0 + \beta_1 X + \epsilon$$

을 고려하자. 여기서, Y는 반응변수를, X는 설명 변수를, 그리고 β_0, β_1는 회귀 계들을 의미하며 ϵ은 평균이 0, 분산이 σ^2인 정규분포를 따른다고 가정하자.

(a) β_1의 최소제곱추정량(LSE) $\widehat{\beta}_1$은 Y와 X 사이의 표본상관계수(correlation coefficient) r과 동일한 부호를 가진다는 것을 보여라.

(b) 가설 $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$을 검정하기 위한 t-test 검정통계량은

$$t = \frac{\sqrt{n-1}r}{\sqrt{1-r^2}}$$

로 표현될 수 있음을 보여라.

Solution:

(a) 단순회귀일 때 다음과 같은 관계식이 성립한다.

$$r_{xy} = \widehat{\beta}_1 \frac{s_x}{s_y} \tag{16}$$

표준편차는 항상 0보다 크므로 부호에 영향을 주지 않는다. 따라서 부호가 같다.

(b) 모상관계수를 ρ라고 할 때 $\beta_1 = 0$을 검정하는 것은 $\rho = 0$을 검정하는 것과 같다. 따라서 $\rho = 0$라는 귀무가설 하에서

$$f(r) = \frac{1}{\Gamma(1/2)} \frac{\Gamma((n-1)/2)}{\Gamma((n-2)/2)} \left(1 - r^2\right)^{(n-4)/2} \tag{17}$$

이라는 사실이 알려져 있다. (*Mathematical Methods of Statistics, Harald Cramer* p.400 참조) 이를 변수 변환하면

$$t = \frac{\sqrt{n-1}r}{\sqrt{1-r^2}} \sim t(n-2) \tag{18}$$

이다.